

# Re-considering the variance parameterization in multiple precision models

BY YI HE, JAMES S. HODGES AND BRADLEY P. CARLIN

*Division of Biostatistics, School of Public Health, University of Minnesota,  
MMC 303, Minneapolis, Minnesota 55455, U.S.A.*

Correspondence author: James S. Hodges

telephone: (612) 626-9626

fax: (612) 626-9054

email: [hodges@ccbr.umn.edu](mailto:hodges@ccbr.umn.edu)

October 5, 2006

## Abstract

Recent developments in Bayesian computing allow accurate estimation of integrals, making advanced Bayesian analysis feasible. However, some problems remain difficult, such as estimating posterior distributions for variance parameters. For models with three or more variances, this paper proposes a simplex parameterization for the variance structure, which has appealing properties and eases the related burden of specifying a reference prior. This parameterization can be profitably used in several multiple-precision models, including crossed random-effect models, many linear mixed models, smoothed ANOVA, and the conditionally autoregressive (CAR) model with two classes of neighbor relations, often useful for spatial data. The simplex parameterization has at least two attractive features. First, it typically leads to simple MCMC algorithms with good mixing properties regardless of the parameterization used to specify the model's reference prior. Thus, a Bayesian analysis can take computational advantage of the simplex parameterization even if its prior was specified using another parameterization. Second, the simplex parameterization suggests a natural reference prior that is proper, invariant under multiplication of the data by a constant, and which appears to reduce the posterior correlation of smoothing parameters with the error precision. We use simulations to compare the simplex parameterization, with its reference prior, to other parameterizations with their reference priors, according to bias and mean-squared error of point estimates and coverage of posterior 95% credible intervals. The results suggest advantages for the simplex approach, particularly when the error precision is small. We offer results in the context of two real data sets from the fields of periodontics and prosthodontics.

# 1 Introduction

Recent developments in Bayesian computing have made it possible to analyze many previously intractable models, but some problems remain difficult, such as estimating posterior distributions for variance parameters. This paper considers the class of multiple-precision linear models, having linear mean structure, normal errors, and at least three precision parameters. This class includes the conditionally autoregressive (CAR) model with two types of neighbor relations (2NRCAR; Besag & Higdon 1999, Reich et al 2004), crossed random-effects models (Box & Tiao 1992, Chapter 5), some dynamic linear models (West & Harrison 1999, Chapter 4), smoothed analysis of variance (Gelman 2005a, Hodges et al 2005), some spatio-temporal models with 1 or 2 spatial neighbor relations and 1 temporal relation (2NRCAR or 3NRCAR), several linear mixed models (Zhao et al 2006), e.g., additive mixed models and bivariate smoothing, and, finally, many problem-specific models (e.g., Gelman & Huang 2006). To make this discussion concrete, we use the 2NRCAR model applied to periodontal data, as follows.

In periodontics, attachment loss is used to assess cumulative damage to a patient’s periodontium and to monitor disease progression (Darby & Walsh 1995). Attachment loss is measured at six sites on each tooth; Figure 1 shows one patient’s data. Each measurement site is indicated by a small circle whose shade of grey indicates measured attachment loss, with darker shade indicating larger (worse) attachment loss. Excluding the four “wisdom teeth” (third molars), a full mouth of 28 teeth gives 168 measurements. If the two jaws are treated as isolated from each other, this spatial structure has at least 2 “islands”, i.e., disconnected groups of measurement sites.

Attachment loss measurements are spatially correlated, but the correlation may not simply be a function of distance. Instead of using point-data (geostatistical) methods, it is practical and intuitive to model attachment loss as measurements on a lattice, which suggests conditionally autoregressive (CAR) models. However, the 168 measurement sites have a complex topography, so more than one smoothing parameter may be needed for adequate fidelity. We consider CAR models with two classes of neighbor relations. Pairs of neighboring sites come in four types (Figure 2): direct neighbor (Type a), same- side neighbors crossing

the gap between teeth (Type b), opposite-side neighbors on the same tooth (Type c), and opposite-side neighbors crossing the gap between teeth (Type d). These four types of neighbor pairs can be grouped into two classes in various ways (Reich et al 2004). This paper considers the classes shown in Figure 2, with solid and dashed lines for class 1 and 2 pairs respectively (Grid A in Reich et al., 2004).

Figure 1 summarizes one patient’s data, to which we fit the 2NRCAR model, as follows. Let  $\mathbf{y} = (y_1, \dots, y_n)^T$  denote the attachment loss measurements, where the subscript indexes measurement sites, and specify this 2NRCAR model:

$$\begin{aligned} \mathbf{y}|\boldsymbol{\theta}, \tau_0 &\sim N(\boldsymbol{\theta}, \tau_0 I_n) \\ \boldsymbol{\theta}|\tau_1, \tau_2 &\propto c(\tau_1, \tau_2)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}'\{\tau_1 Q_1 + \tau_2 Q_2\}\boldsymbol{\theta}\right), \end{aligned} \tag{1}$$

where  $\tau_0, \tau_1$ , and  $\tau_2$  are precisions and  $Q_1$  and  $Q_2$  specify the spatial neighbor relations smoothed by  $\tau_1$  and  $\tau_2$  respectively.  $Q_k, k = 1, 2$ , is  $n \times n$  with off-diagonal entries  $q_{k,ij} = -1$  if sites  $i$  and  $j$  are class- $k$  neighbors and 0 otherwise, and diagonal entries  $q_{k,ii}$  the number of site  $i$ ’s class- $k$  neighbors.

Models are often reparameterized to improve computing or interpretation, e.g., a density with long, narrow contours can be transformed to have more circular contours. This paper proposes an alternative parameterization for variance-structure parameters, the simplex parameterization (Besag & Higdon 1999), and a slice sampler for MCMC draws in this parameterization. The simplex parameterization and its associated reference prior are then compared to other parameterizations and their reference priors. Often, the posterior for variance-structure parameters is sensitive to the prior because the data give little information about them, e.g., because of the spatial structure (Reich et al 2004). Thus, reference priors for variance-structure parameters are an active research area (e.g., Browne & Draper 2004; Gelman 2005b).

Section 2 illustrates some problems that can arise in the posterior distributions of variance-structure parameters, motivating the simplex parameterization. Section 3 develops the new parameterization and a slice sampler for it. Section 4 uses effective sample size to compare the computing performance of MCMC algorithms arising from the simplex parameterizations and three competing parameterizations: precisions

with gamma priors; standard deviations with flat priors (Gelman 2005b); and log precision ratios (defined below; Reich et al 2004) with flat priors. Our MCMC routine on the simplex parameterization generally outperforms MCMC routines on other parameterizations, even for reference priors specified on those other parameterizations. Section 5 uses simulation studies to explore statistical properties of the reference priors associated with each parameterization. Section 6 summarizes our findings. The computer code (in R) used for the simplex parameterization is available at <http://www.biostat.umn.edu/~brad/software.html>.

## 2 Problems with commonly-used parameterizations

For Bayesian analysis of multiple-precision models, several parameterizations have been proposed for the variance structure, including precisions  $\tau_k$ , standard deviations  $\sigma_k = \tau_k^{-1/2}$  (Gelman 2004), precision ratios  $r_k = \tau_k/\tau_0$ ,  $k = 1, 2, \dots$ , and log precision ratios  $z_k = \log r_k$  (Reich et al 2004). These parameterizations are often associated with specific reference priors. For the precision parameterization, the standard “vague” prior is  $\tau_k \sim \text{Gamma}(\epsilon, \epsilon)$  for  $\epsilon = 0.01$  or  $0.001$ . For the standard deviation parameterization, Gelman (2005b) proposed  $\sigma_k \sim \text{Unif}(0, L)$  for a suitable upper bound  $L$ . The precision ratios,  $r_k$ , are positive and somewhat like precisions, which suggests  $r_k \sim \text{Gamma}(\epsilon, \epsilon)$  as a “vague” prior. Finally, the log precision ratios take values anywhere in the real line, which suggests  $z_k \sim \text{Unif}(-L, L)$  for a suitable  $L$ .

These parameterizations are all subject to problems that we illustrate using the 2NRCAR model (1) and Figure 1’s data. Figure 3 suggests how the problems arise. Specifically, for each panel in Figure 3, we re-parameterized model (1) in terms of that panel’s parameterization, applied the reference prior described above, derived the exact marginal posterior distribution of the smoothing parameters, and plotted its contours. For the precision ratios  $(r_1, r_2)$  and log precision ratios  $(z_1, z_2)$ , Figure 3’s panels c and d respectively show contours of the log marginal posterior after integrating all other parameters out of the posterior. Panels a and b show the log conditional posterior for the precisions  $(\tau_1, \tau_2)$  and standard deviations  $(\sigma_1, \sigma_2)$  after integrating  $\boldsymbol{\theta}$  out of the posterior and fixing  $\tau_0 = 1$  and  $\sigma_0 = 1$ , respectively.

The contours for  $(\tau_1, \tau_2)$ ,  $(\sigma_1, \sigma_2)$ , and  $(r_1, r_2)$  (panels a, b, and c, respectively) are L-shaped with

two long arms and modes pressed tightly against one or both coordinate axes. While each plot assumes a particular reference prior, the same qualitative problems are present for other reference priors. The contours of  $(z_1, z_2)$ 's posterior are long and narrow here (panel d) but are distinctly L-shaped for other periodontal datasets (Reich et al 2004). Bimodal posteriors have been observed in the  $(r_1, r_2)$  and  $(z_1, z_2)$  parameterizations (Reich et al 2004), and indeed bimodality occurs readily even in the simplest hierarchical models (Liu & Hodges 2003).

Posterior distributions like these create predictable difficulties. First, standard MCMC approaches tend to give chains with high lagged autocorrelations and small effective sample sizes. For example, for the parameterizations in Figure 3 a, b, c, the autocorrelations at lag 10 are 0.2 to 0.4. Second, the parameters can be poorly identified, that is, either they are highly correlated *a posteriori*, or the posterior has a large flat mode indicating poor ability to distinguish between possible parameter values. Reich et al (2004) showed that for a variety of 2NRCAR spatial structures, posteriors for the precision parameters are either very flat or have pronounced ridges, inducing bad MCMC convergence and mixing (Gelfand et al 1995).

Different problems affect other aspects of Bayesian analysis. The posterior correlation between the error precision and the smoothing precisions is often high because the error precision in effect specifies the data's scale, and the data generally give much more information about this precision than about higher-level precisions. The variance, precision and standard deviation parameterizations are scale-dependent, so for example if the measurement unit is changed from centimeters to millimeters, these parameters are multiplied by 100, 0.01, and 10 respectively. This affects interpretation of hyperparameters and makes it difficult to specify a reference prior. The precision ratio and log precision ratio parameters  $r_k$  and  $z_k$  are scale-invariant, i.e., invariant if the data are multiplied by a constant, but as mentioned are prone to bimodality and highly autocorrelated MCMC draws. Sections 4.2 and 4.3 illustrate the latter point in detail. The simplex parameterization (Besag & Higdon 1999), which we now introduce, appears to avoid or mitigate these difficulties.

### 3 The simplex parameterization and associated methods

#### 3.1 Definition of the simplex parameterization

For a multiple-precision model with precisions  $(\tau_0, \tau_1, \dots, \tau_m)$ , define the total relative precision

$$\lambda = \sum_{k=1}^m r_k = \frac{1}{\tau_0} \sum_{k=1}^m \tau_k,$$

where  $r_k = \tau_k/\tau_0$ . Define the allocation of total relative precision as  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$ , where

$$\beta_k = \frac{r_k}{\lambda} = \frac{r_k}{\sum_{j=1}^m r_j} = \frac{\tau_k}{\sum_{j=1}^m \tau_j};$$

$\sum_{k=1}^m \beta_k = 1$ , and  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_m)$  takes values in the  $m$ -dimensional simplex. The 2, 3, and 4-dimensional simplices are a line segment, equilateral triangle, and tetrahedron, respectively.

This parameterization has two *a priori* attractive features. First, it is scale-invariant, that is, it does not change when the data are multiplied by a constant. Also, the simplex parameter  $\boldsymbol{\beta}$  lies in a bounded space, so a natural reference prior, the flat prior, is proper and exchangeable. The rest of this paper uses a flat prior on  $\boldsymbol{\beta}$  and gamma priors on  $\lambda$  and  $\tau_0$ .

#### 3.2 Computing strategy for the simplex parameterization

For a multiple-precision model like (1), the vector of unknown parameters is  $(\boldsymbol{\theta}, \tau_0, \lambda, \boldsymbol{\beta})$ , where  $\boldsymbol{\theta}$  is the mean-structure parameters,  $\tau_0$  the error precision,  $\lambda$  the total relative precision, and  $\boldsymbol{\beta}$  the allocation of total relative precision. To avoid MCMC sampling variation, we analytically integrate  $\boldsymbol{\theta}$  and  $\tau_0$  out of the joint posterior and run a slice sampler on the marginal posterior of  $(\lambda, \boldsymbol{\beta})$ . Posterior summaries for  $\boldsymbol{\theta}$  and  $\tau_0$  are then obtained by Rao-Blackwellizing.

Suppose the precision parameters in the 2NRCAR model (1) have prior  $p(\tau_0, \tau_1, \tau_2)$ . Then the joint posterior of all the unknowns is

$$p(\boldsymbol{\theta}, \tau_0, \tau_1, \tau_2 | \mathbf{y}) \propto p(\tau_0, \tau_1, \tau_2) p(\mathbf{y} | \boldsymbol{\theta}, \tau_0) p(\boldsymbol{\theta} | \tau_1, \tau_2)$$

$$\begin{aligned} &\propto p(\tau_0, \tau_1, \tau_2) \tau_0^{n/2} \exp\left(-\frac{\tau_0}{2} \sum (y_i - \theta_i)^2\right) \\ &\quad \times \prod_{j=1}^{n-G} (\tau_1 d_{1j} + \tau_2 d_{2j})^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{\theta}' (\tau_1 Q_1 + \tau_2 Q_2) \boldsymbol{\theta}\right), \end{aligned} \quad (2)$$

where  $G$  is the number of islands in the spatial map and  $d_{kj}$  is defined as follows. Simultaneously diagonalize the two positive semi-definite matrices  $Q_k$  as  $B'D_k B$ , where  $B$  is nonsingular (Newcomb 1961), and let  $D_k$  have  $j^{th}$  diagonal element  $d_{kj}$ . It is easy to see  $\boldsymbol{\theta}|\mathbf{y}, \tau_0, r_1, r_2 \sim N((Q_r + I_n)^{-1} X' \mathbf{y}, \tau_0(Q_r + I_n))$ , where  $Q_r = r_1 Q_1 + r_2 Q_2$  and  $r_k = \tau_k/\tau_0$ . After integrating out  $\boldsymbol{\theta}$ ,

$$\begin{aligned} p(\tau_0, r_1, r_2 | \mathbf{y}) &\propto p(\tau_0, r_1, r_2) \tau_0^{\frac{n-G}{2}} |Q_r + I_n|^{-\frac{1}{2}} \prod_{j=1}^{n-G} (r_1 d_{1j} + r_2 d_{2j})^{1/2} \\ &\quad \times \exp\left(-\frac{\tau_0}{2} [\mathbf{y}' \mathbf{y} - \mathbf{y}' (Q_r + I_n)^{-1} \mathbf{y}]\right). \end{aligned}$$

Then if  $\tau_0$ 's prior is  $Gamma(a_0, b_0)$ , with mean  $\frac{a_0}{b_0}$ , integrate out  $\tau_0$  to give

$$p(r_1, r_2 | \mathbf{y}) \propto p(r_1, r_2) \prod_{j=1}^{n-G} (r_1 d_{1j} + r_2 d_{2j})^{1/2} |Q_r + I_n|^{-\frac{1}{2}} R^{-b},$$

where  $R = b_0 + \frac{1}{2} [\mathbf{y}' \mathbf{y} - \mathbf{y}' (Q_r + I_n)^{-1} \mathbf{y}]$ , and  $b = a_0 + \frac{n-G}{2}$ . Now change to the simplex parameterization  $\lambda = r_1 + r_2$  and  $\beta = r_1/\lambda$ , giving

$$p(\lambda, \beta | \mathbf{y}) \propto p(\lambda, \beta) \lambda^{\frac{n-G}{2}} |I + \lambda Q_\beta|^{-\frac{1}{2}} \prod_{j=1}^{n-G} (\beta(d_{1j} - d_{2j}) + d_{2j})^{\frac{1}{2}} R^{-b},$$

where  $R = b_0 + \frac{1}{2} (\mathbf{y}' \mathbf{y} - \mathbf{y}' (I + \lambda Q_\beta)^{-1} \mathbf{y})$ ,  $Q_\beta = \beta Q_1 + (1 - \beta) Q_2$ , and the Jacobian is implicit in the change of variables in the prior, from  $p(r_1, r_2)$  to  $p(\lambda, \beta)$ .  $B$  is orthogonal if and only if  $Q_1 Q_2$  is symmetric, in which case

$$p(\lambda, \beta | \mathbf{y}) \propto p(\lambda, \beta) \prod_{j=1}^{n-G} \left(\frac{\lambda \gamma_j}{1 + \lambda \gamma_j}\right)^{\frac{1}{2}} \left[b_0 + \frac{1}{2} \left(\sum_j \frac{\lambda \gamma_j}{1 + \lambda \gamma_j} \mathbf{y}_j^{*2}\right)\right]^{-b}, \quad (3)$$

where  $\mathbf{y}^* = B \mathbf{y}$  and  $\gamma_j = \beta(d_{1j} - d_{2j}) + d_{2j}$ , so (3) depends on  $\lambda$  and  $\gamma_j$  only through  $\frac{\lambda \gamma_j}{1 + \lambda \gamma_j}$ .

For this problem, we propose a slice sampler with one auxiliary variable. A slice sampler can be more efficient than an ordinary Metropolis-Hastings algorithm, e.g., Neal (1997, 2003), Tierney & Mira (1999). Generally, the slice sampler can be described as follows (Damien et al 1999). Suppose an MCMC

has stationary distribution  $\pi(\lambda, \boldsymbol{\beta}) \propto p(\lambda, \boldsymbol{\beta})l(\lambda, \boldsymbol{\beta})$ . Introduce an auxiliary random variable  $U$  with a conditional uniform distribution  $U|\lambda, \boldsymbol{\beta} \sim \text{Unif}(0, l(\lambda, \boldsymbol{\beta}))$ . Then  $(\lambda, \boldsymbol{\beta}, U)$  has joint distribution

$$f(\lambda, \boldsymbol{\beta}, u) \propto p(\lambda, \boldsymbol{\beta})I_{\{u < l(\lambda, \boldsymbol{\beta})\}}(\lambda, \boldsymbol{\beta}, u).$$

The slice sampler is then a special case of the Gibbs sampler:

1. Initialize  $\boldsymbol{\beta}^{(0)}, \lambda^{(0)}$ ;
2. Generate  $U|\lambda, \boldsymbol{\beta}$  from a uniform distribution:  $U^t|\lambda^{t-1}, \boldsymbol{\beta}^{t-1} \propto \text{Unif}(0, l(\lambda^{t-1}, \boldsymbol{\beta}^{t-1}))$ .
3. Generate  $\boldsymbol{\beta}|u, \lambda$  from  $p(\lambda, \boldsymbol{\beta})$  restricted to  $l(\lambda, \boldsymbol{\beta}) > u$ :  $\boldsymbol{\beta}^t|\lambda^{t-1}, U^t \propto p(\lambda, \boldsymbol{\beta})I(l(\lambda^{t-1}, \boldsymbol{\beta}) > U^t)$ .
4. Generate  $\lambda|u, \boldsymbol{\beta}$  from  $p(\lambda, \boldsymbol{\beta})$  restricted to  $l(\lambda, \boldsymbol{\beta}) > u$ :  $\lambda^t|\boldsymbol{\beta}^t, U^t \propto p(\lambda, \boldsymbol{\beta})I(l(\lambda, \boldsymbol{\beta}^t) > U^t)$ .

Repeat steps 2-4; after convergence,  $(\boldsymbol{\beta}^t, \lambda^t)$  are samples from the stationary distribution  $\pi(\lambda, \boldsymbol{\beta})$ .

A natural  $p(\lambda, \boldsymbol{\beta})$  is  $p(\lambda, \boldsymbol{\beta}) = p_1(\lambda)p_2(\boldsymbol{\beta})$ , where  $p_1$  is a gamma density and  $p_2$  is uniform on the simplex, a special case of the Dirichlet distribution. With this choice, candidate  $\beta_j$  can be generated as  $X_j / \sum_{j=1}^m X_j$ , where  $X_1, \dots, X_m$  are independent exponential variates. An informative prior for  $\boldsymbol{\beta}$  can be  $\text{Dirichlet}(\alpha_1, \dots, \alpha_m)$ , from which samples can also be generated using draws from gamma distributions. For the 2NRCAR model,  $l(\lambda, \boldsymbol{\beta}) = \lambda^{\frac{n-G}{2}} \prod_{j=1}^{n-G} (\beta d_{1j} + (1-\beta)d_{2j})^{\frac{1}{2}} |I + \lambda Q_{\boldsymbol{\beta}}|^{-\frac{1}{2}} R^{-b}$  and  $p(\lambda, \boldsymbol{\beta}) = \frac{1}{\Gamma(a_{\lambda})} \lambda^{a_{\lambda}-1} e^{-b_{\lambda}\lambda} I(\boldsymbol{\beta} \in [0, 1])$ .

The posterior distributions of  $\boldsymbol{\theta}$  and  $\tau_0$  can be estimated by Rao-Blackwellizing (Casella & Robert 1996). For posterior samples  $(\lambda^t, \boldsymbol{\beta}^t)$ ,  $t = 1, 2, \dots, M$ ,  $\boldsymbol{\theta}$ 's posterior density can be estimated as

$$p(\boldsymbol{\theta}|\mathbf{y}) = \int p(\boldsymbol{\theta}|\lambda, \boldsymbol{\beta}, \mathbf{y})p(\lambda, \boldsymbol{\beta}|\mathbf{y})d\lambda d\boldsymbol{\beta} \approx \frac{1}{M} \sum_{t=1}^M p(\boldsymbol{\theta}|\lambda^t, \boldsymbol{\beta}^t, \mathbf{y}), \quad (4)$$

where  $p(\boldsymbol{\theta}|\lambda^t, \boldsymbol{\beta}^t, \mathbf{y})$  is  $\boldsymbol{\theta}$ 's conditional posterior given  $(\lambda^t, \boldsymbol{\beta}^t)$ . For the normal-error model (1),  $\boldsymbol{\theta}|\lambda, \boldsymbol{\beta}, \mathbf{y}$  is multivariate- $t$  with center  $(P^t)^{-1}\mathbf{y}$ , scale  $(P^t)^{-1}R^t/b$  and  $2b$  degrees of freedom, where  $R^t = b_0 + \frac{1}{2}[\mathbf{y}'\mathbf{y} - \mathbf{y}'(P^t)^{-1}\mathbf{y}]$ , and  $P^t = \lambda^t B'(\beta^t D_1 + (1-\beta^t)D_2)B + I_n$ . Thus  $\boldsymbol{\theta}$ 's posterior mean and variance are

estimated by

$$\begin{aligned}
E(\boldsymbol{\theta}|\mathbf{y}) &= E(E(\boldsymbol{\theta}|\lambda, \boldsymbol{\beta}, \mathbf{y})) \approx \frac{1}{M} \sum_{t=1}^M E(\boldsymbol{\theta}|\lambda^t, \boldsymbol{\beta}^t, \mathbf{y}) = \frac{1}{M} \sum_{t=1}^M \mu_{\boldsymbol{\theta}}^t = \bar{\mu}_{\boldsymbol{\theta}} \\
Var(\boldsymbol{\theta}|\mathbf{y}) &= E(Var(\boldsymbol{\theta}|\mathbf{y}, \lambda, \boldsymbol{\beta})) + Var(E(\boldsymbol{\theta}|\mathbf{y}, \lambda, \boldsymbol{\beta})) \\
&\approx \frac{1}{M} \left[ \sum_{t=1}^M \Sigma_{\boldsymbol{\theta}}^t + \sum_{t=1}^M (\mu_{\boldsymbol{\theta}}^t - \bar{\mu}_{\boldsymbol{\theta}})(\mu_{\boldsymbol{\theta}}^t - \bar{\mu}_{\boldsymbol{\theta}})' \right], \tag{5}
\end{aligned}$$

where  $\mu_{\boldsymbol{\theta}}^t$  and  $\Sigma_{\boldsymbol{\theta}}^t$  are the posterior mean and variance of  $p(\boldsymbol{\theta}|\lambda^t, \boldsymbol{\beta}^t, \mathbf{y})$ , respectively. Similarly,  $\tau_0|\lambda^t, \boldsymbol{\beta}^t, \mathbf{y}$  is gamma distributed with shape  $b$  and rate  $R^t$ , so posterior summaries for  $\tau_0$  can be obtained analogously.

## 4 MCMC algorithm performance in the different parameterizations

### 4.1 Effective sample size (ESS)

Effective sample size (ESS) is commonly used to assess MCMC mixing (e.g., Carlin & Louis 2000, Chapter 5; Sargent et al 2000; Chen et al 2000; Ridgeway & Madigan 2003). The ESS of a sampled quantity is defined (Kass et al 1998) as

$$ESS = \frac{M}{1 + 2 \sum_{l=1}^{\infty} \rho_l}, \tag{6}$$

where  $M$  is the number of MCMC samples for that quantity and  $\rho_l$  is the estimated lag  $l$  autocorrelation of the samples. ESS can be interpreted as the size of an independent, identically distributed sample giving information equivalent to the autocorrelated MCMC sample. In practice  $\rho_l$  is estimated with error, and past a certain  $l$  the  $\hat{\rho}_l$  are dominated by noise (Gilks et al 1996; Chapter 3). To avoid summing noise, Geyer (1992) proposed the initial convex sequence estimator, which requires a sequence of empirical  $\Gamma_m$  estimates that are positive, monotone, and convex, where  $\Gamma_m$  is the sum of two lagged autocovariances  $\gamma_{2m}$  and  $\gamma_{2m+1}$ . The natural estimator of the lagged autocovariance is the empirical autocovariance  $\hat{\gamma}_l = \frac{1}{M} \sum_{t=1}^{M-l} (X_t - \bar{X})(X_{t+l} - \bar{X})$ , where  $\{X_t\}$  is the sequence of MCMC samples. Priestley (1981, p. 323) suggests using this “biased” estimate with divisor  $M$  rather than the “unbiased” estimate with divisor  $M - l$ . Define  $m^*$  as the largest integer such that  $\hat{\Gamma}_m$  is a positive, monotonely decreasing, and convex

sequence in  $m$ . Then the ESS in (6) sums only estimated autocorrelations  $\hat{\rho}_l$  for  $l \leq 2m^*$ .

## 4.2 Periodontal data analyzed using 2NRCAR

This section compares MCMC algorithms specified in each of four parameterizations, for the 2NRCAR model applied to Figure 1's data. For each parameterization, the data were analyzed three times, using three different prior distributions, each a reference prior for one of the parameterizations. This is an unusual simulation study design; the point is that one may prefer inferences using a reference prior specified on one parameterization, while it is advantageous to specify the MCMC algorithm on a different parameterization.

The four parameterizations are simplex, log precision ratios  $(z_1, z_2)$ , precisions  $(\tau_0, \tau_1, \tau_2)$ , and standard deviations  $(\sigma_0, \sigma_1, \sigma_2)$ . The three reference priors are as follows: for the simplex parameterization, we put a Gamma(0.01, 0.01) prior on  $\lambda$ , and on  $\beta$ , a uniform distribution on the unit interval; for the parameterization with three precisions, we gave each precision a Gamma(0.01, 0.01) prior; and for the parameterization with three standard deviations, we gave each standard deviation a uniform prior on the interval (0, 10). For each parameterization, for each prior, 10000 MCMC draws were made with 5000 discarded for burn-in. Table 1 describes the MCMC algorithm for each parameterization. Except for the simplex parameterization, the algorithms were Metropolis-Hastings with normal candidate draws for the working parameters (Table 1), centered on the current draw. For each working parameter, the sample standard deviation of the 5000 burn-in draws was used as the standard deviation of the candidate draws in the subsequent 5000 retained iterations. A dynamic search procedure (see the Appendix) was used to accelerate the slice sampler.

Table 2 shows effective sample size (ESS) for the four parameterizations and three priors. The simplex parameterization has the largest ESS for two priors, and roughly the same ESS as  $(z_1, z_2)$  for the flat prior on  $(\sigma_0, \sigma_1, \sigma_2)$ . The simplex parameterization's sample autocorrelations decrease quickly as lag increases and generally vanish by lag 10, while the alternatives have much larger autocorrelations at all lags (data not shown). As currently programmed, the simplex parameterization's slice sampler usually runs more slowly than the other algorithms, so it has a smaller advantage in ESS per second of run time (Table 3),

and is roughly tied with the log precision ratio parameterization  $(z_1, z_2)$ .

Section 2 suggested that the simplex parameters  $(\lambda, \beta)$  might have smaller posterior correlations with the error precision  $\tau_0$ , compared to other parameterizations' smoothing parameters. This was true for the present dataset, with the prior distribution having little effect. For each parameterization, we report the posterior correlation only for the parameter having the largest absolute correlation. In the simplex parameterization,  $\beta$  had the largest absolute posterior correlation with  $\tau_0$ , about 0.33 for all three priors. The analogous results for the other three parameterizations were:  $(z_1, z_2)$ , 0.53 for  $z_1$ ; precisions, 0.53 for  $\tau_1$ ; and standard deviations, 0.76 for  $\sigma_1$ . Contrary to our expectation,  $(z_1, z_2)$  — which, like the simplex parameterization, is invariant when the data are multiplied by a constant — gave the same maximum absolute posterior correlations as did the precision parameterization.

Figure 4 shows a contour plot of the log marginal posterior arising from the simplex parameterization and its reference prior. While this is not especially like a bivariate normal density, it does seem rather less irregular than the analogous contour plots for the other parameterizations (Figure 3).

### 4.3 Smoothed ANOVA (SANOVA) model

The smoothed ANOVA model used here was introduced by Sargent & Hodges (1997) and fully developed in Hodges et al (2005; see also Smith 1973, Gelman 2005a). Suppose the experimental design has one error term,  $c$  design cells, and  $n$  replications per cell. Parameterize each effect so the design matrix has orthogonal columns. Group the  $L$  columns for main effects, including the intercept, into a matrix  $A_1$ , and the  $N$  columns for interactions into a matrix  $A_2$ , and scale  $A_1$  and  $A_2$  so  $A_1' A_1 = cnI_L$  and  $A_2' A_2 = cnI_N$ ;  $A_1' A_2 = 0$ . The SANOVA model is

$$\mathbf{y} = A_1 \Theta_1 + A_2 \Theta_2 + \epsilon, \tag{7}$$

where  $\mathbf{y}$  is the  $cn$ -vector of observed outcomes,  $\epsilon \sim N(0, \Gamma_1)$ , the grand mean and main effects in  $\Theta_1$  have an improper flat prior, the interactions in  $\Theta_2$  have a  $N(0, \Gamma_2)$  prior,  $\epsilon$  and  $[\Theta_1 | \Theta_2]$  are independent *a priori*, and the two covariance matrices  $\Gamma_1$  and  $\Gamma_2$  are specified as  $\Gamma_1 = \frac{1}{\tau_0} I_{cn}$  and  $\Gamma_2^{-1} = \text{diag}(\phi_1, \dots, \phi_N)$ .

For a set of distinct smoothing precisions  $(\tau_1, \dots, \tau_s)$ ,  $s \leq N$ , define a deterministic assignment function  $j(k)$  that specifies groups of  $\phi_k$  within which  $\phi_k = \tau_{j(k)}$ , and let  $n_j$  be the number of  $\phi_k$  mapping to  $\tau_j$ .

The joint posterior after integrating out  $\Theta$  is

$$f(\tau_0, \mathbf{r} | \mathbf{Y}) \propto \pi(\tau_0, \mathbf{r}) \tau_0^{\frac{cn-L}{2}} \exp\left(-\frac{1}{2}\tau_0 W(\mathbf{r})\right) \prod_{j=1}^s \left(\frac{r_j}{r_j + cn}\right)^{n_j/2}, \quad (8)$$

where  $r_j = \frac{\tau_j}{\tau_0}$  and  $W(\mathbf{r}) = \mathbf{y}'\mathbf{y} - \frac{1}{cn}\mathbf{y}'A_1A_1'\mathbf{y} - \mathbf{y}'A_2diag((cn + r_{j(k)})^{-1})A_2'\mathbf{y}$ .

This model has  $s$  smoothing precisions  $\tau_1, \dots, \tau_s$ , so the simplex parameter  $\beta$  is  $s$ -dimensional. If  $\tau_0$  has a gamma prior  $G(a_0, b_0)$ , with mean  $\frac{a_0}{b_0}$ , then  $\tau_0$ 's full conditional posterior is also gamma. After integrating out  $\tau_0$ ,  $(\lambda, \beta)$  has marginal posterior

$$f(\lambda, \beta | \mathbf{Y}) \propto \pi(\lambda, \beta) \prod_{j=1}^s \left[1 + \frac{cn}{\lambda\beta_j}\right]^{-n_j/2} R^{-b}, \quad (9)$$

where  $R = b_0 + \frac{1}{2}\mathbf{y}'\mathbf{y} - \frac{1}{2cn}\mathbf{y}'A_1A_1'\mathbf{y} - \frac{1}{2}\mathbf{y}'A_2diag((cn + \lambda\beta_{j(k)})^{-1})A_2'\mathbf{y}$  and  $b = a_0 + \frac{cn-L}{2}$ .

Hodges & Sargent (2001, Section 6) applied smoothed ANOVA to a  $2^3$  factorial experiment testing a material's tensile strength (Lai & Hodges 1999). The three design factors were the type of mold, presence of pigment, and type of cure, with  $n = 6$  replications per cell. The dataset is in Hodges & Sargent (2001). We used this dataset to compare MCMC routines for different parameterizations and priors, as in Section 4.2's comparison for the 2NRCAR model, and using the same priors as in Section 4.2. For all three priors, the MCMC on the simplex parameterization has by far the largest ESS (Table 4) and the smallest autocorrelations (data not shown). The MCMC on the simplex parameterization also has the largest ESS/sec for two of the three priors (Table 5). Overall, the smoothed ANOVA results are consistent with the 2NRCAR results.

## 5 Statistical performance of each parameterization’s reference prior

### 5.1 2NRCAR model

To reduce computing time, we simulated periodontal measurements on upper and lower jaws with 5 teeth each, for 60 total measurements in one “patient”. The two neighbor classes are as in Figure 2. This simulation experiment’s design considered three factors: (1) true error precision  $\tau_0$ ; (2) the true degree of smoothness in the two classes of neighbor pairs,  $(\tau_1, \tau_2)$ ; and (3) the 4 parameterizations, each with its associated reference prior (Table 6). Table 7 gives the specific true values of  $(\tau_0, \tau_1, \tau_2)$ .

For each design cell, the 1000 simulated datasets were drawn as follows. By the spectral decomposition,  $\tau_1 Q_1 + \tau_2 Q_2 = \Gamma' \Lambda \Gamma$ , where  $\Gamma$  is an orthogonal matrix and  $\Lambda$  is diagonal. Then  $\boldsymbol{\theta}^* = \Gamma \boldsymbol{\theta}$  has density

$$p(\boldsymbol{\theta}^*) \propto \exp\left(-\frac{1}{2} \boldsymbol{\theta}^{*'} \Lambda \boldsymbol{\theta}^*\right) = \exp\left(-\frac{1}{2} \boldsymbol{\theta}_{n-G}^{*'} \Lambda_{n-G} \boldsymbol{\theta}_{n-G}^*\right)$$

where the subscript  $n - G$  indicates the first  $n - G$  rows and/or columns. Thus, the first  $n - G$  elements of  $\boldsymbol{\theta}^*$  were drawn from independent normal distributions, for  $G = 2$  islands in the “mouth”. The last 2 elements of  $\boldsymbol{\theta}^*$  have flat priors under  $p(\boldsymbol{\theta}^*)$  and were drawn from a uniform on  $(-10, 10)$ . Then the sample of true  $\boldsymbol{\theta}$  were obtained as  $\boldsymbol{\theta} = \Gamma' \boldsymbol{\theta}^*$ .

For the simplex and log precision ratio ( $Z$ ) parameterizations, MCMC samples were drawn from the marginal posterior after integrating out  $\boldsymbol{\theta}$  and  $\tau_0$ , and the posterior mean and interval coverage were estimated by Rao-Blackwellizing. For the precision and SD parameterizations, MCMC samples were drawn from the marginal posterior after integrating out only  $\boldsymbol{\theta}$ . For the simplex parameterization, we used the slice sampler (Section 3.2) with starting values  $\beta_k = \frac{1}{s}$ , where  $s$  is the number of smoothing precisions, and for the other parameterizations we used adaptive Metropolis algorithms as described in Section 4. Trace plots were checked for a sample of artificial datasets and in all cases indicated sampler convergence.

The parameterization/reference prior combinations (henceforth, “methods”) were compared according to their results on the standard deviation scale, i.e.,  $\sigma_k = 1/\sqrt{\tau_k}$ , the same scale as the data, using bias and MSE of posterior means as point estimates, and coverage of equal-tailed 95% credible intervals. (The

Appendix gives equations for Rao-Blackwellizing the  $Z$  and simplex parameters in the standard deviation scale.) To remove effects that obscure comparisons, we report bias as a percent of the true value and we scale MSE according to the true error variance.

Figure 5 displays scaled bias, scaled MSE, and 95% interval coverage for the four methods. All methods have small biases for the error standard deviation  $\sigma_0$  except the  $Z$  method in case 3, where the posterior mean overestimates  $\sigma_0$  by about 30%. By contrast, the  $Z$  method consistently underestimates  $\sigma_1$ , while the other methods have small biases. For  $\sigma_2$ , all methods have larger bias and the SD method performs worst, overestimating substantially in all cases. For all methods and cases, the MSEs for  $\sigma_0$  and  $\sigma_1$  are small. The  $Z$  method has the largest MSE for  $\sigma_1$ . For  $\sigma_2$ , all methods' MSEs vary a lot, but the simplex method consistently gives the smallest MSE and the SD method the largest. Finally, all methods give coverage close to 95% for  $\sigma_0$  and  $\sigma_1$  except for  $Z$ , which gives low coverage. For  $\sigma_2$ , the precision and simplex methods give coverage 95% or higher for all cases, while the  $Z$  and SD methods had quite low coverage for some cases.

## 5.2 SANOVA model

This simulation experiment used artificial data from a  $2^3$  design with  $n = 6$  replications per cell, as in Hodges et al's (2005, section 3) simulation study. The three design factors were: (1) the true error precision  $\tau_0$  (note that increasing  $n$  and  $\tau_0$  have the same effect); (2) the number of truly present interactions (1 or 3); and (3) the four parameterizations with associated reference priors, described in Table 6. Two further cases were simulated to examine the effect of multiplying the data by a constant. Table 7 gives the design values for the 8 cases considered.

We again generated 1000 simulated datasets for each "case". The design matrix for the  $2^3$  mean structure was orthogonal, so without loss of generality the true grand mean and main effects  $\theta_1, \theta_2, \theta_3, \theta_4$  were set to zero. If an interaction term was present, its  $\theta_k$  was set to 1, otherwise to zero. The interaction terms were *a priori* exchangeable and each was smoothed by its own smoothing precision, so as in Hodges

et al. (2005, Section 3), we need only consider how many interactions are truly present, not which ones.

The four methods were compared according to their performance for three groups of parameters: the four interaction  $\theta_k$ ,  $k = 5, \dots, 8$ ; the error precision  $\tau_0$ ; and the eight cell means  $c_j$ ,  $j = 1, \dots, 8$ . For each group of parameters, the methods were compared according to bias and MSE of posterior means as point estimates, and coverage probability of the 95% equal-tail credible interval, with one exception: cell-mean bias is a simple linear function of bias of the interaction  $\theta_k$  and is thus omitted. By design, all methods give identical bias and MSE for the grand mean and main effects, so they are not considered further. We follow Hodges et al (2005) in calling truly present interactions “target interactions” and truly absent interactions “null interactions”. By the simulation design’s exchangeability, all target interactions have the same true bias, MSE, and coverage for a given method, as do all null interactions, so we present average bias and MSE for the targets and for the nulls. For the interactions  $\theta_k$  and cell means  $c_j$ , we scaled bias and MSE as percents of the true error standard deviation  $\frac{1}{\sqrt{\tau_0}}$  and the true error variance  $\frac{1}{\tau_0}$ , respectively. Similarly, for the estimates of the error precision  $\tau_0$ , we report bias and square root of MSE as percents of  $\tau_0$ .

Figure 6 displays the bias and MSE of posterior mean estimates of the interaction  $\theta_k$ , and coverage of their 95% posterior intervals. For the target interactions, the number of truly present interactions has little effect on bias or MSE. Compared to the simplex method, the SD method has smaller bias (Figure 6a). In general, the SD method performs better than the precision method, which in turn performs better than the  $Z$  method. For the null interactions, all methods are essentially unbiased and the  $Z$  method has the smallest MSE (Figure 6b). As for 95% posterior intervals (Figure 6c,d), for the target interactions, the simplex and SD methods give coverage much closer to the nominal 95% than the  $Z$  and precision methods, which are too low for cases with small error precision. For the null interactions, the simplex and SD methods have about 95% coverage while coverage for the other two methods is too high. Broadly speaking, for the interaction  $\theta_k$ , the simplex method gives good performance that improves relative to the other methods as the error precision decreases.

Figure 7 shows scaled bias and MSE for the error precision  $\tau_0$  (panels a,b), and MSE and coverage

probability for the cell means (panels c,d). For  $\tau_0$ , the SD method outperforms the others in both bias and MSE (Figure 7a,b). The 95% CI coverage is close to the nominal 95% for all methods and cases (data not shown). For the cell means, Figure 7c,d show the scaled MSE (as a percent of  $\frac{1}{\tau_0}$ ) and 95% interval coverage averaged over the 8 cells. The simplex and SD methods perform similarly. When 1 target interaction is present, these methods have higher bias than the other two, but when 3 target interactions are present, they have smaller bias. Coverage of 95% credible intervals is close to the nominal 95% for all methods, except for the  $Z$  method for small error precisions when 3 target interactions are present.

### 5.3 Crossed random effect model

The crossed random effect model (10) has error precision  $\tau_0$  and two smoothing precisions  $\tau_1$  and  $\tau_2$  for rows and columns respectively in the two-way layout, as follows:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K, \quad (10)$$

where  $\alpha_i \sim N(0, \tau_1)$ ,  $\beta_j \sim N(0, \tau_2)$ , and  $\epsilon_{ijk} \sim N(0, \tau_0)$  for unknown  $\tau_0, \tau_1, \tau_2$ . This simulation experiment's design had three factors: (1) the true error precision  $\tau_0$ ; (2) the true  $\tau_1$  and  $\tau_2$ , considering equal and unequal smoothness in rows and columns; and (3) the four parameterizations with their reference priors, described in Table 6.

Each of the 1000 artificial datasets per simulation design cell had 5 row levels ( $\alpha_i$ ,  $i = 1, \dots, 5$ ), 5 column levels ( $\gamma_j$ ,  $j = 1, \dots, 5$ ), and 5 replicates ( $\epsilon_{ijk}$ ,  $k = 1, \dots, 5$ ). Without loss of generality, the grand mean  $\mu$  was set to zero. We generated artificial datasets as follows: Generate row effects  $\alpha_1, \dots, \alpha_5$ , column effects  $\gamma_1, \dots, \gamma_5$ , then in each of the 25 cells, add 5 random normal errors to give 125 total observations. The algorithms and outcome measures in this simulation study are the same as for the 2NRCAR simulation study (Section 5.1).

Figure 8 shows bias and MSE of posterior means as point estimates and 95% credible interval coverage, for the three standard deviations  $\sigma_0, \sigma_1$ , and  $\sigma_2$ . For the error standard deviation  $\sigma_0$ , all methods are essentially unbiased and have small MSE. However, bias is complex for the two smoothing standard

deviations  $\sigma_1$  and  $\sigma_2$ . The simplex method has much smaller bias than the SD method for most cases (Figure 8a), but otherwise it is difficult to generalize. For MSE (Figure 8b), the simplex method is lower than the alternatives except for cases 3 and 6 for  $\sigma_2$ . For coverage of 95% intervals (Figure 8c), all methods are consistently close to the nominal 95% for  $\sigma_0$ . For  $\sigma_1$  and  $\sigma_2$ , the simplex, precision, and SD methods perform similarly and fairly well, while the  $Z$  method performs worse, particularly for  $\sigma_2$ .

## 6 Discussion

We have developed a parameterization for multiple-precision models, first mentioned for 2NRCAR by Besag & Higdon (1999). Based on Sections 4 & 5, the simplex parameterization appears to have two advantages. First, it gives simple MCMC algorithms with good mixing properties for various reference priors. Thus Bayesian analyses may benefit from this parameterization even for priors specified in another parameterization. Second,  $\beta$  has a proper natural reference prior that is invariant when the data are multiplied by a constant;  $\lambda$  has the same invariance property. Section 5 showed that compared to other proposed reference priors, this prior yields posterior means with generally good bias and mean squared error, and 95% credible intervals with close to nominal coverage, for the range of cases considered. Its worst performance was for smoothed ANOVA in Section 5.2. If one were designing a software package solely to do smoothed ANOVA, these results suggest that the simplex parameterization — with the reference prior used here — might not be the best choice for a prior distribution. However, if one were seeking an all-purpose off-the-shelf prior, these results are not so discouraging: while the simplex parameterization was not the best prior for smoothed ANOVA, it did not lose badly to the other priors, while each of the other priors did perform poorly for at least one example.

The obvious question is: can we improve the statistical performance of the simplex parameterization? The first consideration in this vein is the reference prior. The allocation parameter  $\beta$  has a natural reference prior, but the total relative precision  $\lambda$  does not. Sections 4 & 5 used the conventional “vague” Gamma(0.01,0.01) prior, which, with 50th and 90th percentiles  $4 \times 10^{-29}$  and 0.0015 respectively, is in fact

quite informative. Other priors for  $\lambda$  may improve statistical or computing performance, though we do not yet have a firm basis for proposing an alternative. One simple alternative would be a log-normal prior. In preliminary results from a simulation study of smoothed ANOVA, giving  $\lambda$  a lognormal prior with a large variance seems to improve coverage of posterior 95% intervals compared to the gamma prior considered here, but otherwise the operating characteristics are similar.

It seems pertinent that  $\lambda$  is unitless or, put another way, that  $\lambda$  has the same scale for all problems. Thus, for the smoothed ANOVA and crossed random-effects models, it should be possible to determine universally-applicable large and small values of  $\lambda$ , and perhaps use that information to specify, say, a uniform prior for  $\lambda$ . The 2NRCAR example is more complicated in a manner that is beyond the present paper's scope, but it might be possible to extend this general idea.

Some literature on priors for hierarchical models (e.g., Daniels 1999; Gustafson et. al. 2006) suggests that a prior may be judged by the relative weight it gives to information arising from the data (governed by the error precision  $\tau_0$ ) and information arising from the model (governed by the smoothing parameters  $\tau_k$ ). One way to implement this idea is to consider, in our notation,  $\tau_k / \sum_{j=0}^s \tau_j$  for  $k = 0, \dots, s$ . The simplex parameterization lends itself readily to this suggestion. The error precision's fraction of total precision is easily shown to be  $1/(1 + \lambda)$ , which is readily computed in the context of MCMC. As for the smoothing precisions  $\tau_k$ ,  $k = 1, \dots, s$ , their aggregate fraction of total precision is  $\lambda/(1 + \lambda)$ , and  $\tau_k$ 's fraction of total precision is  $\beta_k \lambda / (1 + \lambda)$ , also easily computed using MCMC. A flat prior on  $\beta$  treats  $\tau_k$ ,  $k = 1, \dots, s$ , exchangeably; priors on  $\lambda$  might be compared according to how they weigh  $\tau_0$  against individual  $\tau_k$  or the ensemble of  $\tau_k$ s.

The simplex parameterization extends straightforwardly in two ways. First, it extends immediately if any of the models presented here is extended by adding one or more random effects parameterized by variances or precisions. For example, the 2NRCAR model (1) can be extended to a spatio-temporal model for multiple dental visits by adding a third class of neighbor pairs representing two consecutive observations at a given measurement site. This adds a third smoothing precision, which can be handled in the obvious

manner. A second extension is for models with many smoothing precisions that naturally fall into, say, two groups. In such a model, a separate simplex parameter pair  $(\lambda, \beta)$  can be used for each of the groups of smoothing precisions.

Although the simplex parameterization is applicable to a broad class of models (Section 1), extension to models with covariance matrices would be desirable. The approach of Barnard et al (2000), in which the covariance matrix is decomposed into standard deviations and correlations, is one possible extension, where the simplex parameterization would be applied to the vector of standard deviations.

## References

1. Barnard JD, McCulloch R, & Meng X-L (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, **10**, 1281-1311.
2. Besag J, Higdon D (1999). Bayesian inference for agricultural field experiments (with discussion). *J. Roy. Statist. Soc. B*, **61**, 691- 746.
3. Box GEP, Tiao GC (1992). *Bayesian Inference in Statistical Analysis*. Classic Edition. New York: Wiley.
4. Browne WJ, Draper D (2004). A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Nottingham Statistics Research Report 04-01*.
5. Carlin BP, Louis TA (2000). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall/CRC, Boca Raton FL, 2nd edition.
6. Casella G, Robert CP (1996). Rao-Blackwellisation of sampling schemes. *Biometrika*. **83**:81-94.
7. Chen L, Qin Z, Liu J (2000). Exploring hybrid Monte Carlo in Bayesian computation. *ISBA 2000, Proceedings*.

8. Damien P, Wakefield JC, Walker SG (1999). Gibbs sampling for Bayesian nonconjugate and hierarchical models by using auxiliary variables. *J. Roy. Statist. Soc. B.* **61**, 331-344.
9. Daniels MJ (1999). A prior for the variance in hierarchical models. *Can. J. Stat.*, **27**:567-578..
10. Darby ML, Walsh MM (1995). *Periodontal and Oral Hygiene Assessment. Dental Hygiene Theory and Practice*. Philadelphia: W.B. Saunders.
11. Gelfand AE, Sahu SK, Carlin BP (1995) Efficient parameterizations for normal linear mixed models. *Biometrika*, **82**, 479-488.
12. Gelman A (2004). Parameterization and Bayesian modeling. *J. Amer. Stat. Assoc.* **99**, 537-545.
13. Gelman A (2005a). Analysis of variance: why it is more important than ever (with discussion). *Ann. Stat.* **33**, 1-53.
14. Gelman A (2005b). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis.* **1**, 1-19.
15. Gelman A, Carlin J, Stern H, Rubin D (2004). *Bayesian Data Analysis*. CRC Press, Boca Raton FL, 2nd edition.
16. Gelman A, Huang Z (2006). Estimating incumbency advantage and its variation, as an example of a before/after study. *J. Amer. Stat. Assn.*, to appear.
17. Geyer CJ (1992). Practical Markov Chain Monte Carlo. *Stat. Science* **7**:473-483.
18. Gilks WR, Richardson S, and Spiegelhalter DJ (1996). *Markov Chain Monte Carlo in Practice*. New York: CRC press.
19. Gustafson P, Hossain S, MacNab Y (2006). Conservative prior distributions for variance parameters in hierarchical models. *Can. J. Stat*, in press.

20. Hodges JS, Carlin BP, Fan Q (2003). On the precision of the conditionally autoregressive prior in spatial models. *Biometrics*, **59**, 317-322.
21. Hodges JS, Cui Y, Sargent DJ, Carlin BP (2005). Smoothing balanced single-error-term analysis of variance. *Technometrics*, in press; <ftp://ftp.biostat.umn.edu/pub/2005/rr2005-018.pdf>
22. Hodges JS, Sargent DJ (2001). Counting degrees of freedom in hierarchical and other richly-parameterised models. *Biometrika*, **88**, 367-379.
23. Kass RE, Carlin BP, Gelman A, Neal R (1998). Markov chain Monte Carlo in practice: A roundtable discussion. *Amer. Stat.*, **52**, 93-100.
24. Lai JH, Hodges JS (1999). Effects of processing parameters on physical properties of the silicon maxillofacial prosthetic materials. *Dent. Mat.* **15**, 450-455.
25. Liu J, Hodges JS (2003). Posterior bimodality in the balanced one-way random-effects model. *J. Roy. Stat. Soc. B*, **65**, 247-255.
26. Neal RM (1997). Markov chain Monte Carlo methods based on “slicing” the density function. *Technical Report 9722*, Dept. Statistics, Univ. Toronto.
27. Neal RM (2003). Slice sampling. *Ann. Stat.* **31**, 705-767.
28. Newcomb RW (1961). On the simultaneous diagonalization of two semi-definite matrices. *Quart. Appl. Math.*, **19**, 144-146.
29. Priestley MB (1981). *Spectral Analysis and Time Series*. Academic, London.
30. Reich BJ, Hodges JS, Carlin BP (2004). Spatial analysis of periodontal data using conditional autoregressive priors having two types of neighbor relations. *J. Amer. Stat. Assoc.*, in press; <ftp://ftp.biostat.umn.edu/pub/2004/rr2004-004.ps.gz>.
31. Ridgeway G, Madigan D (2003). A sequential Monte Carlo method for Bayesian analysis of massive datasets. *Data Min. Knowl. Discov.* **7(3)**: 301-319.

32. Sargent DJ, Hodges JS (1997). Smoothed ANOVA with application to subgroup analysis. *Research Report RR97-002, University of Minnesota.*
33. Sargent DJ, Hodges JS, Carlin BP (2000). Structured Markov chain Monte Carlo. *J. Comp. Graph. Stat.*, **9(2)**:217-234.
34. Smith AFM (1973). Bayes estimates in one-way and two-way models. *Biometrika*, **60**, 319-329.
35. Tierney L, Mira A (1999). Some adaptive Monte Carlo methods for Bayesian inference. *Stat. in Med.* **18**, 2507-2515.
36. West M, Harrison J (1999). *Bayesian Forecasting and Dynamic Models*. 2nd edition. Springer-Verlag.
37. Zhao Y, Staudenmayer J, Coull BA, Wand MP (2006). General design Bayesian generalized linear mixed models. *Stat. Sci.*, **21**:35-51.

## Appendix

### 6.1 Rao-Blackwellizing on the standard deviation scale

In Section 5, the four parameterizations with their associated priors were compared according to point-estimate and interval-coverage performance on the standard deviation scale, with Rao-Blackwellizing done as follows. Suppose  $\tau_0|\lambda, \beta, \mathbf{y} \sim \text{Gamma}(b, R)$ , then  $p(\tau_0|\mathbf{y}) \approx \frac{1}{M} \sum_{t=1}^M \text{Gamma}(\tau_0|b^t, R^t)$ . Changing variables to  $\sigma_0 = \tau_0^{-1/2}$  and including the Jacobian,  $p(\sigma_0|\mathbf{y}) \approx \frac{1}{M} \sum_{t=1}^M 2\sigma_0^{-3} \text{Gamma}(\sigma_0^{-2}|b^t, R^t)$ , so

$$\begin{aligned}
 E(\sigma_0|\mathbf{y}) &\approx \frac{1}{M} \sum_{t=1}^M \int 2\sigma_0^{-2} \text{Gamma}(\sigma_0^{-2}|b^t, R^t) d\sigma_0 \\
 &= \frac{1}{M} \sum_{t=1}^M \int \tau_0^{-\frac{1}{2}} \text{Gamma}(\tau_0|b^t, R^t) d\tau_0 \\
 &= \frac{1}{M} \sum_{t=1}^M E(\tau_0^{-\frac{1}{2}}|b^t, R^t) = \frac{1}{M} \sum_{t=1}^M \frac{\Gamma(b^t - \frac{1}{2})}{\Gamma(b^t)} (R^t)^{\frac{1}{2}}
 \end{aligned}$$

Similarly, noting that  $\sigma_1 = r_1^{-\frac{1}{2}} \tau_0^{-\frac{1}{2}}$  and  $\sigma_2 = r_2^{-\frac{1}{2}} \tau_0^{-\frac{1}{2}}$ ,

$$E(\sigma_1|\lambda, \beta, \mathbf{y}) = r_1^{-\frac{1}{2}} E(\tau_0^{-\frac{1}{2}}|\lambda, \beta, \mathbf{y}) = r_1^{-\frac{1}{2}} \frac{\Gamma(b - \frac{1}{2})}{\Gamma(b)} (R)^{\frac{1}{2}}$$

$$\begin{aligned}
E(\sigma_1|\mathbf{y}) &\approx \frac{1}{M} \sum_{t=1}^M (r_1^t)^{-\frac{1}{2}} \frac{\Gamma(b^t - \frac{1}{2})}{\Gamma(b^t)} (R^t)^{\frac{1}{2}} \\
E(\sigma_2|\mathbf{y}) &\approx \frac{1}{M} \sum_{t=1}^M (r_2^t)^{-\frac{1}{2}} \frac{\Gamma(b^t - \frac{1}{2})}{\Gamma(b^t)} (R^t)^{\frac{1}{2}}
\end{aligned} \tag{11}$$

## 6.2 Dynamic search for the slice sampler

In the simplex parameterization's slice sampler (Section 3.2), to accept one sample, generally a large number of samples need to be drawn from  $p(\lambda, \beta)$ . The slice sampler can be accelerated by improving this acceptance rate. The following dynamic search is one approach for a low-dimensional parameter space; we show it for a scalar  $\beta$ .

1. Choose grid points for  $\lambda, \beta$  by a preliminary analysis, say,  $\lambda^1 < \dots < \lambda^\Omega$  and  $\beta^1 < \dots < \beta^\Pi$ .
2. Calculate  $l_{ij} = l(\lambda^i, \beta^j|\mathbf{y})$  at these grid points  $(\lambda^i, \beta^j)$ .
3. At the  $t^{\text{th}}$  MCMC cycle, given  $\lambda^t$  and  $U^t$ ,  $\beta$  is conditionally uniform on  $\{l(\lambda^t, \beta) > U^t\}$ . Thus,  $\beta$  can be generated from a uniform distribution on  $(a_\beta, b_\beta) \supset \{l(\lambda^t, \beta) > U^t\}$ , chosen as follows.
  - (a) From the pre-selected grid for  $\lambda$ , find the two  $\lambda^i$  that bracket  $\lambda^t$ . Call them  $L_\lambda$  and  $U_\lambda$ .
  - (b) Find the bounds of  $\beta$ ,  $(a_\beta^*, b_\beta^*)$  among  $(L_\lambda, \beta^j)$  and  $(U_\lambda, \beta^j)$  such that  $l(L_\lambda, \beta|\mathbf{y}) > U^t$  and  $l(U_\lambda, \beta|\mathbf{y}) > U^t$ .
  - (c) Extend both ends of the interval  $(a_\beta^*, b_\beta^*)$  until  $l(\lambda^t, a_\beta^*|\mathbf{y}) \leq U^t$  and  $l(\lambda^t, b_\beta^*|\mathbf{y}) \leq U^t$ , giving  $(a_\beta, b_\beta)$ .
4. Draw  $\beta$  from  $Unif(a_\beta, b_\beta)$ , until  $l(\lambda^t, \beta|\mathbf{y}) > U^t$ .

The pre-processing steps 1 and 2 are done before the MCMC draws. The interval  $(a_\beta, b_\beta)$  is in general much narrower than the original  $(0, 1)$ , so the acceptance rate is improved.

We present this accelerator as part of a proof of principle and do not claim it can be used generally. Obviously the efficiency of our slice sampler can and should be improved.

	Simplex	$(z_1, z_2)$	$(\tau_0, \tau_1, \tau_2)$	$(\sigma_0, \sigma_1, \sigma_2)$
Algorithm	Slice sampler with 1 uniform auxiliary variable	adaptive Metropolis w/ Normal candidate	adaptive Metropolis w/ Normal candidate	adaptive Metropolis w/ Normal candidate
Working par.	$(\lambda, \beta)$	$(z_1, z_2)$	$(\log \tau_0, \log \tau_1, \log \tau_2)$	$(\log \sigma_0, \log \sigma_1, \log \sigma_2)$
Initial values	(4, 0.2)	(1, 1)	(1, 2, 1)	(1, 2, 1)
Initial tuning constants	—	0.5	0.2	0.2

Table 1: Description of algorithms for the 2NRCAR model

Prior	Parameterization used in MCMC algorithm			
	$(\lambda, \beta)$	$(r_1, r_2)$	$(\tau_1, \tau_2, \tau_0)$	$(\sigma_1, \sigma_2, \sigma_0)$
$\lambda \sim \text{Gamma}(0.01, 0.01)$ $\beta \sim \text{uniform on simplex}$	$\lambda$ : 573 $\beta$ : 1009	$\log(r_1)$ : 577 $\log(r_2)$ : 531	$\log(\tau_1)$ : 275 $\log(\tau_2)$ : 591 $\log(\tau_0)$ : 301	$\log(\sigma_1)$ : 379 $\log(\sigma_2)$ : 672 $\log(\sigma_0)$ : 359
$\text{Gamma}(0.01, 0.01)$ for $\tau_0, \tau_1$ and $\tau_2$	$\lambda$ : 1037 $\beta$ : 1035	$\log(r_1)$ : 640 $\log(r_2)$ : 713	$\log(\tau_1)$ : 261 $\log(\tau_2)$ : 697 $\log(\tau_0)$ : 248	$\log(\sigma_1)$ : 261 $\log(\sigma_2)$ : 253 $\log(\sigma_0)$ : 418
flat for SDs $\sigma_0, \sigma_1, \sigma_2$	$\lambda$ : 1389 $\beta$ : 860	$\log(r_1)$ : 648 $\log(r_2)$ : 651	$\log(\tau_1)$ : 175 $\log(\tau_2)$ : 406 $\log(\tau_0)$ : 215	$\log(\sigma_1)$ : 265 $\log(\sigma_2)$ : 156 $\log(\sigma_0)$ : 234

Table 2: Effective sample size (ESS) comparison of various parameterizations for the CAR model with two classes of neighbor relations.

Prior	Parameterization used in MCMC algorithm			
	$(\lambda, \beta)$	$(r_1, r_2)$	$(\tau_0, \tau_1, \tau_2)$	$(\sigma_0, \sigma_1, \sigma_2)$
$\lambda \sim \text{Gamma}(0.01, 0.01)$ $\beta \sim \text{uniform on simplex}$	$\lambda$ : 0.43 $\beta$ : 0.75	$\log(r_1)$ : 0.94 $\log(r_2)$ : 0.87	$\log(\tau_1)$ : 0.17 $\log(\tau_2)$ : 0.36 $\log(\tau_0)$ : 0.18	$\log(\sigma_1)$ : 0.24 $\log(\sigma_2)$ : 0.43 $\log(\sigma_0)$ : 0.23
$\text{Gamma}(0.01, 0.01)$ for $\tau_0, \tau_1$ and $\tau_2$	$\lambda$ : 0.59 $\beta$ : 0.59	$\log(r_1)$ : 1.02 $\log(r_2)$ : 1.14	$\log(\tau_1)$ : 0.17 $\log(\tau_2)$ : 0.44 $\log(\tau_0)$ : 0.16	$\log(\sigma_1)$ : 0.16 $\log(\sigma_2)$ : 0.26 $\log(\sigma_0)$ : 0.16
flat for SDs $\sigma_0, \sigma_1, \sigma_2$	$\lambda$ : 0.78 $\beta$ : 0.48	$\log(r_1)$ : 1.04 $\log(r_2)$ : 1.05	$\log(\tau_1)$ : 0.11 $\log(\tau_2)$ : 0.26 $\log(\tau_0)$ : 0.14	$\log(\sigma_1)$ : 0.29 $\log(\sigma_2)$ : 0.17 $\log(\sigma_0)$ : 0.26

Table 3: Effective sample size per second (ESS/sec) comparison of various parameterizations for the CAR model with two classes of neighbor relations.

Prior	Parameterization used in MCMC algorithm			
	$(\lambda, \beta)$	$\boldsymbol{r}$	$\boldsymbol{\tau}$	$\boldsymbol{\sigma}$
$\lambda \sim \text{Gamma}(0.01, 0.01)$ $\beta \sim \text{uniform on simplex}$	$\lambda$ : 1615	$\log(r_1)$ : 336	$\log(\tau_0)$ : 280	$\log(\sigma_0)$ : 313
	$\beta$ : 3965	$\log(r_2)$ : 231	$\log(\tau_1)$ : 244	$\log(\sigma_1)$ : 194
	4617	$\log(r_3)$ : 294	$\log(\tau_2)$ : 227	$\log(\sigma_2)$ : 169
	4971	$\log(r_4)$ : 331	$\log(\tau_3)$ : 219	$\log(\sigma_3)$ : 343
			$\log(\tau_4)$ : 230	$\log(\sigma_4)$ : 329
$\text{Gamma}(0.01, 0.01)$ for $\tau_0, \tau_1, \tau_2, \tau_3, \tau_4$	$\lambda$ : 2498	$\log(r_1)$ : 210	$\log(\tau_0)$ : 436	$\log(\sigma_0)$ : 287
	$\beta$ : 4110	$\log(r_2)$ : 365	$\log(\tau_1)$ : 336	$\log(\sigma_1)$ : 172
	5000	$\log(r_3)$ : 244	$\log(\tau_2)$ : 198	$\log(\sigma_2)$ : 240
	5000	$\log(r_4)$ : 370	$\log(\tau_3)$ : 321	$\log(\sigma_3)$ : 204
			$\log(\tau_4)$ : 149	$\log(\sigma_4)$ : 137
flat for SDs $\sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4$	$\lambda$ : 4614	$\log(r_1)$ : 638	$\log(\tau_0)$ : 484	$\log(\sigma_0)$ : 453
	$\beta$ : 4657	$\log(r_2)$ : 752	$\log(\tau_1)$ : 506	$\log(\sigma_1)$ : 591
	4576	$\log(r_3)$ : 626	$\log(\tau_2)$ : 503	$\log(\sigma_2)$ : 470
	5000	$\log(r_4)$ : 655	$\log(\tau_3)$ : 629	$\log(\sigma_3)$ : 516
			$\log(\tau_4)$ : 500	$\log(\sigma_4)$ : 199

Table 4: Comparison of effective sample size (ESS) in SANOVA model

Prior	Parameterization used in MCMC algorithm			
	$(\lambda, \beta)$	$\boldsymbol{r}$	$\boldsymbol{\tau}$	$\boldsymbol{\sigma}$
$\lambda \sim \text{Gamma}(0.01, 0.01)$ $\beta \sim \text{uniform on simplex}$	$\lambda$ : 152.5	$\log(r_1)$ : 60.1	$\log(\tau_0)$ : 46.0	$\log(\sigma_0)$ : 50.7
	$\beta$ : 374.4	$\log(r_2)$ : 41.3	$\log(\tau_1)$ : 40.1	$\log(\sigma_1)$ : 31.4
	436.0	$\log(r_3)$ : 52.6	$\log(\tau_2)$ : 37.3	$\log(\sigma_2)$ : 27.4
	469.4	$\log(r_4)$ : 59.2	$\log(\tau_3)$ : 36.0	$\log(\sigma_3)$ : 55.6
			$\log(\tau_4)$ : 37.8	$\log(\sigma_4)$ : 53.3
$\text{Gamma}(0.01, 0.01)$ for $\tau_0, \tau_1, \tau_2, \tau_3, \tau_4$	$\lambda$ : 215.3	$\log(r_1)$ : 36.6	$\log(\tau_0)$ : 75.2	$\log(\sigma_0)$ : 50.8
	$\beta$ : 354.3	$\log(r_2)$ : 63.6	$\log(\tau_1)$ : 57.9	$\log(\sigma_1)$ : 30.4
	431.0	$\log(r_3)$ : 42.5	$\log(\tau_2)$ : 34.1	$\log(\sigma_2)$ : 42.5
	431.0	$\log(r_4)$ : 64.5	$\log(\tau_3)$ : 55.3	$\log(\sigma_3)$ : 36.1
			$\log(\tau_4)$ : 25.7	$\log(\sigma_4)$ : 24.2
flat for SDs $\sigma_0, \sigma_1, \sigma_2, \sigma_3, \sigma_4$	$\lambda$ : 36.5	$\log(r_1)$ : 80.1	$\log(\tau_0)$ : 61.2	$\log(\sigma_0)$ : 57.9
	$\beta$ : 36.8	$\log(r_2)$ : 94.4	$\log(\tau_1)$ : 64.0	$\log(\sigma_1)$ : 75.5
	36.2	$\log(r_3)$ : 78.5	$\log(\tau_2)$ : 63.6	$\log(\sigma_2)$ : 60.0
	39.5	$\log(r_4)$ : 82.2	$\log(\tau_3)$ : 79.5	$\log(\sigma_3)$ : 65.9
			$\log(\tau_4)$ : 63.2	$\log(\sigma_4)$ : 25.4

Table 5: Comparison of effective sample size per second (ESS/sec) in SANOVA model

Table 6: Parameterization and associated reference priors

Method	Parameter	Prior	Integrate out $\tau_0$ ?
Simplex	$\beta_k = \frac{\tau_k}{\sum \tau_j}$ ;	$\lambda \sim \text{Gamma}(0.01, 0.01)$ ,	Yes
	$\lambda = \frac{\sum \tau_j}{\tau_0}$	$\beta \sim \text{Unif}$ on the simplex	
Precision	$\tau_0, \tau_1, \dots, \tau_s$	$\tau_k \sim \text{Gamma}(0.01, 0.01)$ , $k = 0, \dots, s$	No
SD	$\sigma_0 = \frac{1}{\sqrt{\tau_0}}$ ,	$\sigma_k \sim \text{Unif}(0, 100)$ , $k = 0, \dots, s$	No
	$\sigma_k = \frac{1}{\sqrt{\tau_k}}$	except SANOVA $\sigma_k \sim \text{Unif}(0, 10)$	
Z	$z_k = \log(\frac{\tau_k}{\tau_0})$	$z_k \sim \text{Unif}(-15, 15)$ , $k = 1, \dots, s$	Yes

Table 7: Design values in the simulation studies

Case	2NRCAR			SANOVA								Crossed RE			
	$\tau_0$	$\tau_1$	$\tau_2$	$\tau_0$	$\theta_1$	$\theta_2$	$\theta_3$	$\theta_4$	$\theta_5$	$\theta_6$	$\theta_7$	$\theta_8$	$\tau_0$	$\tau_1$	$\tau_2$
1	1	1	1	1	0	0	0	0	1	0	0	0	$\frac{1}{4}$	1	1
2	1	1	$\frac{1}{4}$	$\frac{1}{4}$	0	0	0	0	1	0	0	0	$\frac{1}{4}$	$\frac{1}{16}$	$\frac{1}{16}$
3	1	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{16}$	0	0	0	0	1	0	0	0	$\frac{1}{4}$	$\frac{1}{16}$	1
4	1	$\frac{1}{4}$	1	1	0	0	0	0	1	1	1	0	$\frac{1}{16}$	1	1
5	$\frac{1}{4}$	1	1	$\frac{1}{4}$	0	0	0	0	1	1	1	0	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
6	$\frac{1}{4}$	1	$\frac{1}{4}$	$\frac{1}{16}$	0	0	0	0	1	1	1	0	$\frac{1}{16}$	$\frac{1}{16}$	1
7	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{100}$	0	0	0	0	10	0	0	0	$\frac{1}{100}$	$\frac{1}{25}$	$\frac{1}{25}$
8	$\frac{1}{4}$	$\frac{1}{4}$	1	1	0	0	0	0	10	0	0	0	—	—	—

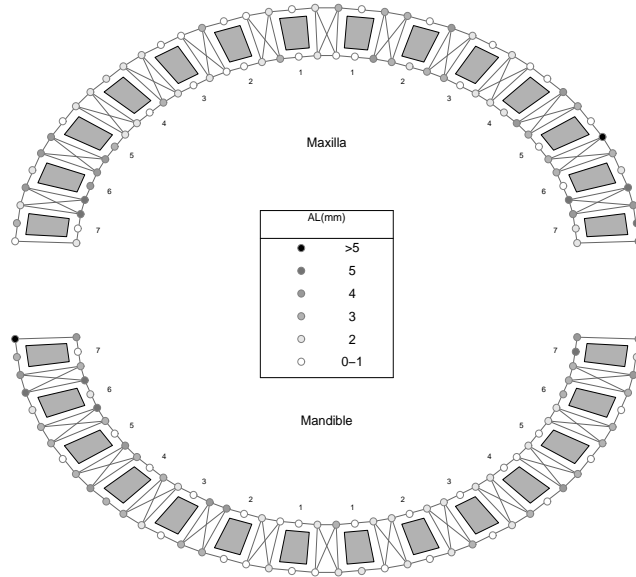


Figure 1: Attachment loss measurements for one patient. The maxilla is the upper jaw, the mandible is the lower jaw, the gray boxes are teeth, the small number counting from the center of each jaw is the tooth number. Small circles indicate the six measurement sites per tooth.

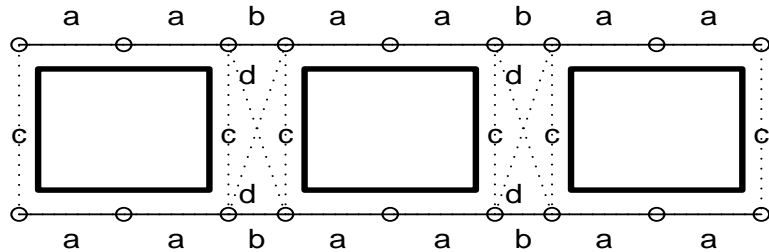


Figure 2: Neighbor types in periodontal measurements. Letters a-d specify neighbor types. Solid and dotted lines indicate the two classes of neighbors considered in this paper.

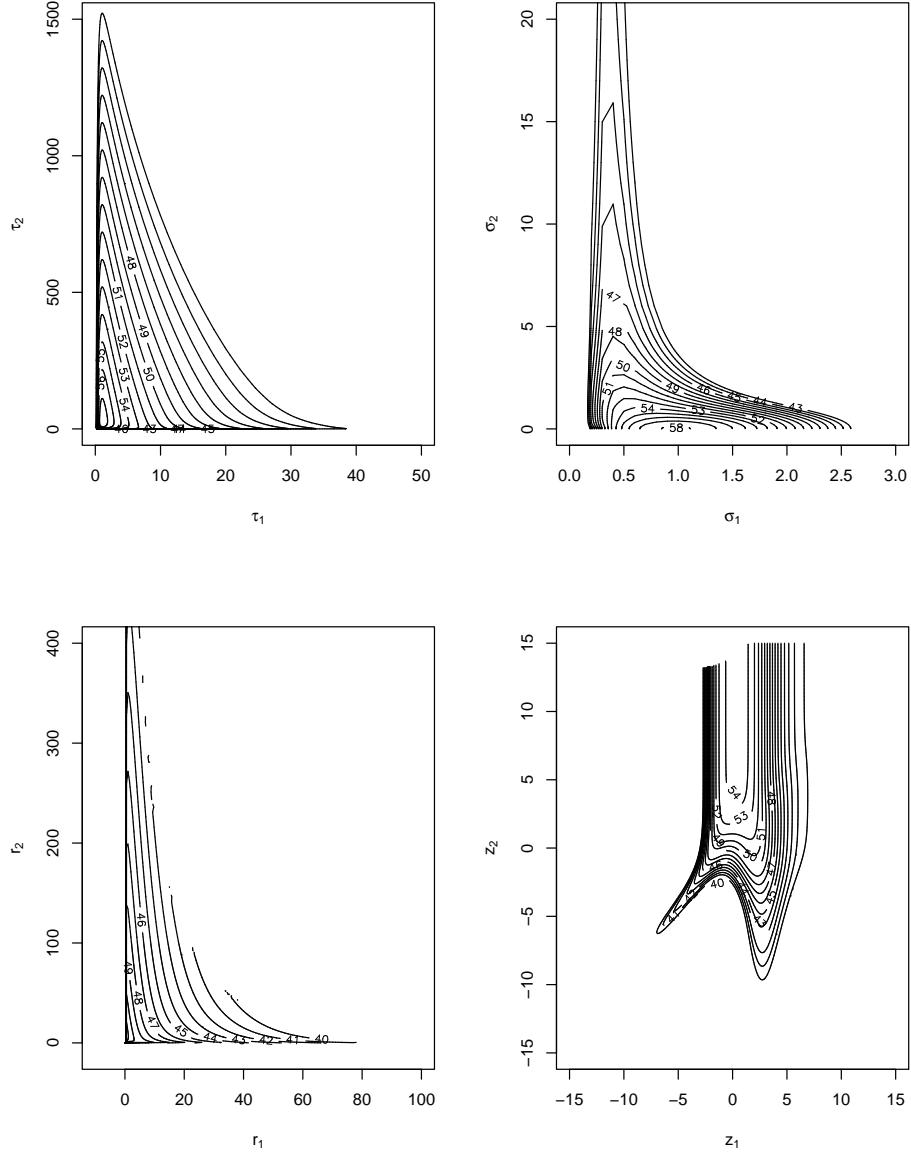


Figure 3: 2NRCAR model: Logarithm posterior contour plots with contours at 1 log intervals for four parameterizations with their own reference priors:  $\tau_0, \tau_1, \tau_2 \sim \text{Gamma}(0.01, 0.01)$ ,  $\sigma_0, \sigma_1, \sigma_2 \sim \text{Unif}(0, L)$ ,  $r_1, r_2 \sim \text{Gamma}(0.01, 0.01)$ ,  $z_1, z_2 \sim \text{Unif}(-15, 15)$ . The contours for  $(\tau_0, \tau_1, \tau_2)$  and  $(\sigma_0, \sigma_1, \sigma_2)$  are drawn for the slice  $\tau_0 = 1$  and  $\sigma_0 = 1$ , respectively.

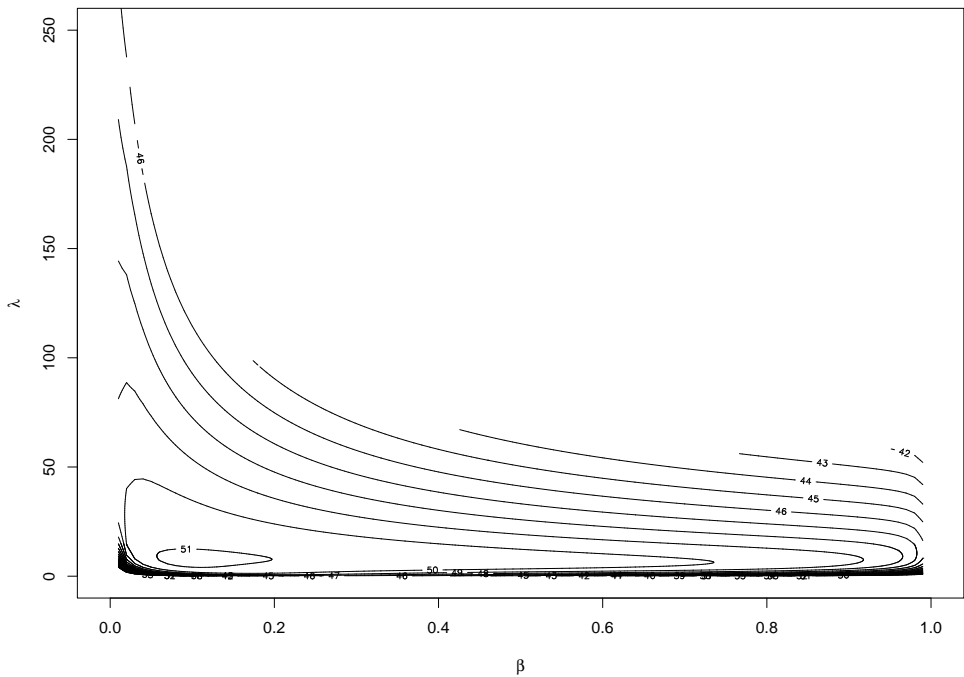


Figure 4: 2NRCAR model: Log posterior contour plot with contours at 1 log intervals, for the simplex parameterization with its reference prior.

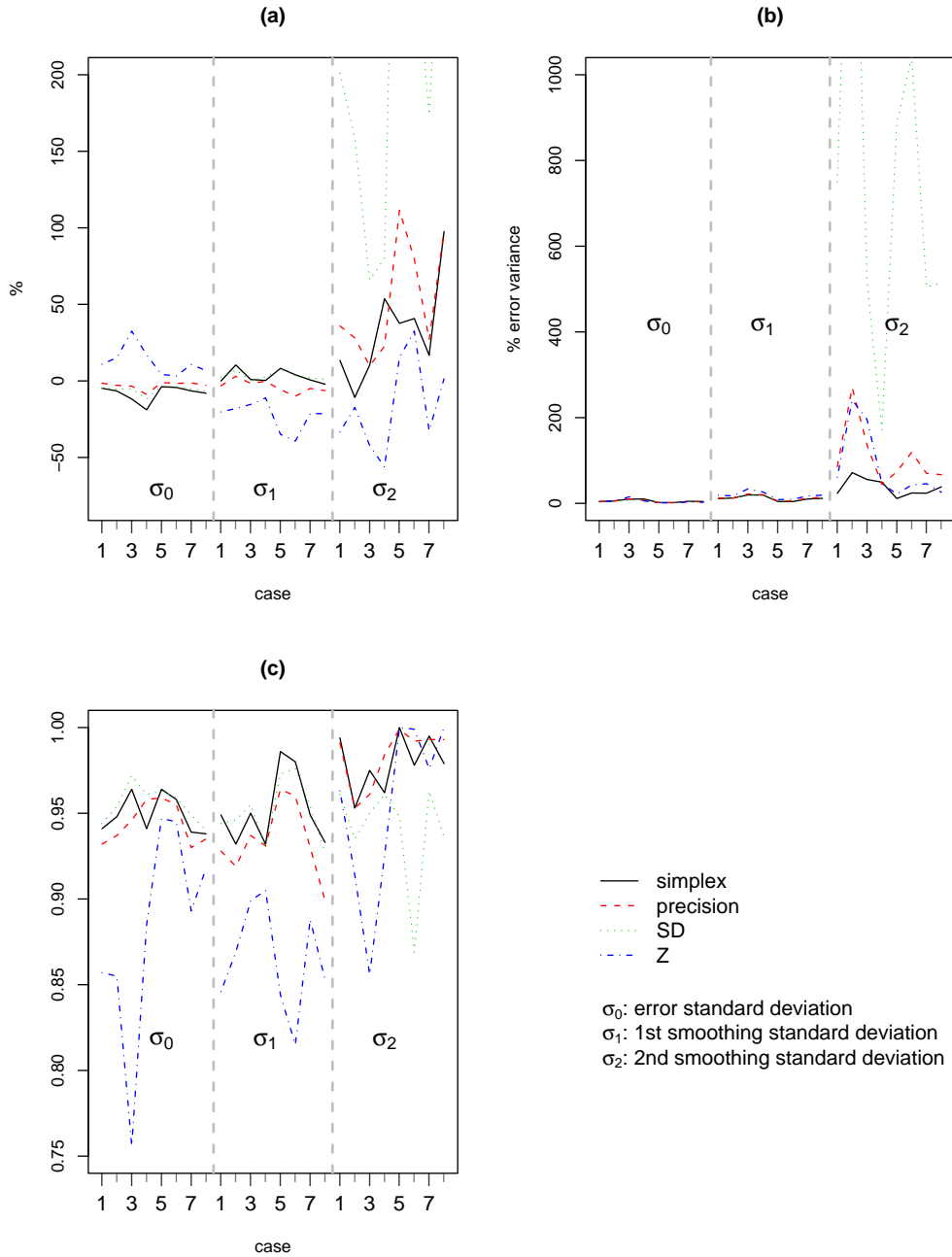


Figure 5: 2NRCAR simulation: Standard deviation bias (as a percent of true standard deviation) and MSE (divided by the true error variance  $\frac{1}{\tau_0}$ ). (a) scaled bias for  $\sigma_0$ ,  $\sigma_1$ , and  $\sigma_2$ ; (b) scaled MSE for  $\sigma_0$ ,  $\sigma_1$ , and  $\sigma_2$ ; (c) 95% interval coverage for  $\sigma_0$ ,  $\sigma_1$ , and  $\sigma_2$ .

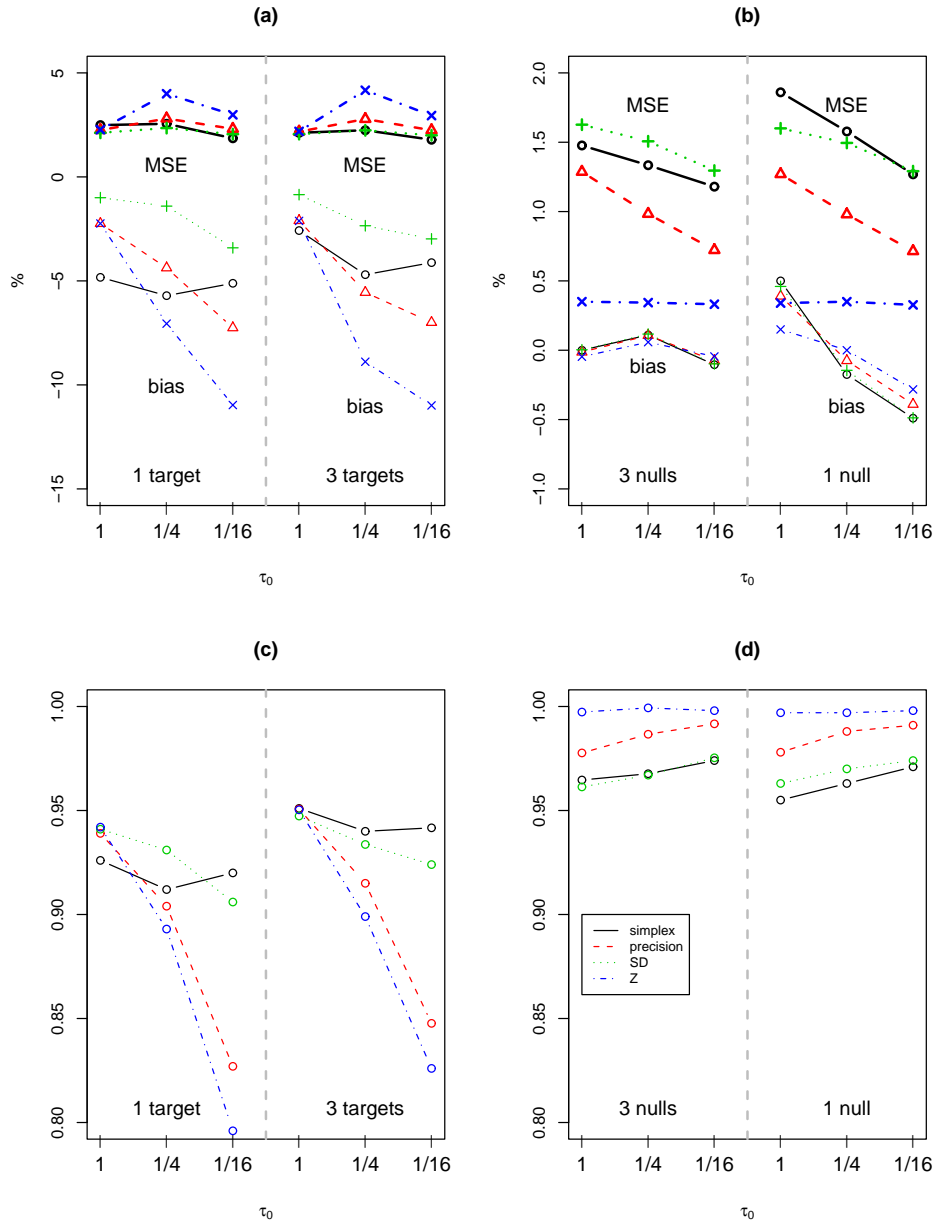


Figure 6: SANOVA simulation: Average bias and MSE as percents of  $\frac{1}{\sqrt{\tau_0}}$  and  $\frac{1}{\tau_0}$  respectively, for  $\theta_k$  for truly present interactions (a), and truly absent interactions (b). Within each figure, the upper curves are MSE and the lower curves are biases. 95% interval coverage probability for  $\theta_k$  for truly present interactions (c), and truly absent interactions (d).

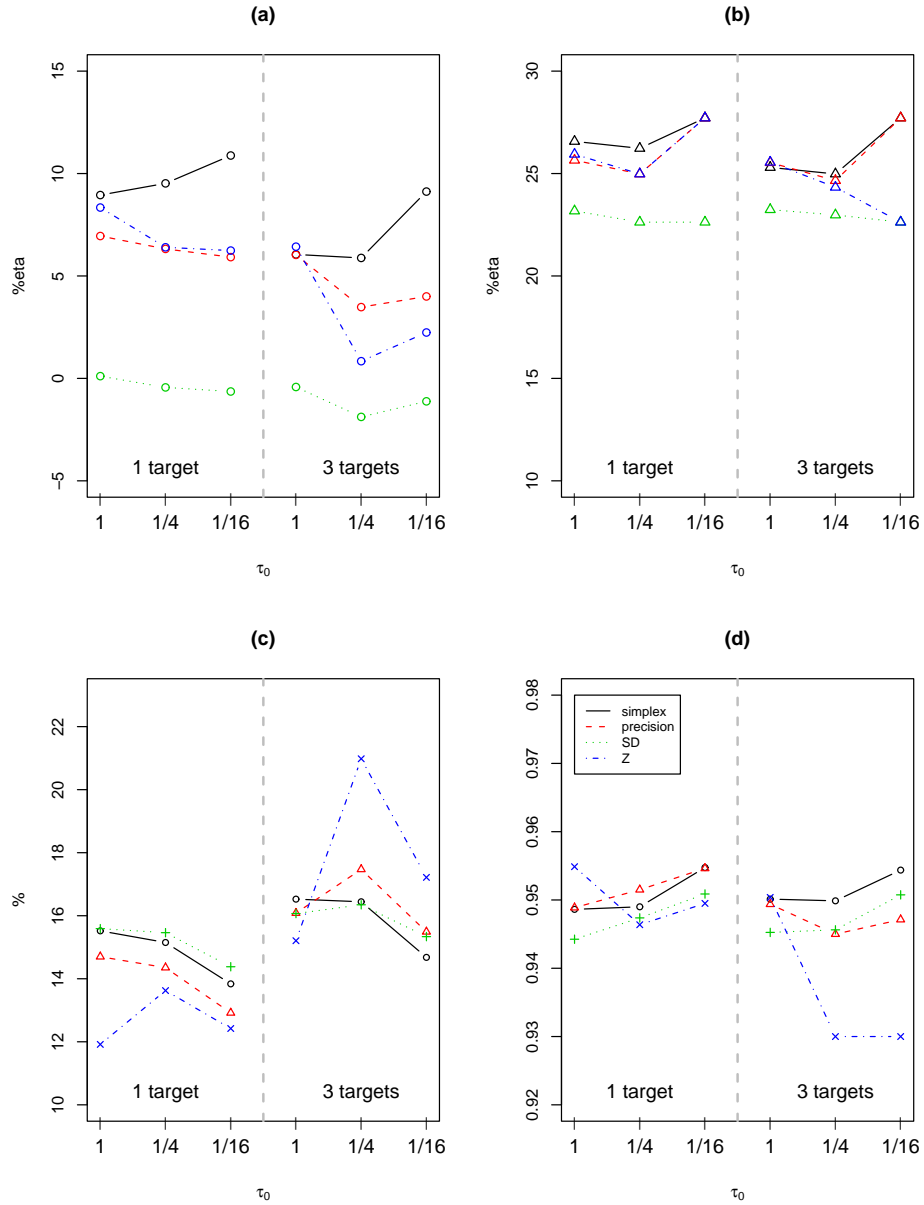


Figure 7: SANOVA simulation: (a) error precision bias as a percent of  $\tau_0$ ; (b) square root of MSE as a percent of  $\tau_0$ ; (c) average cell mean MSE (as a percent of  $\frac{1}{\tau_0}$ ); (d) average cell mean 95% interval coverage probability.

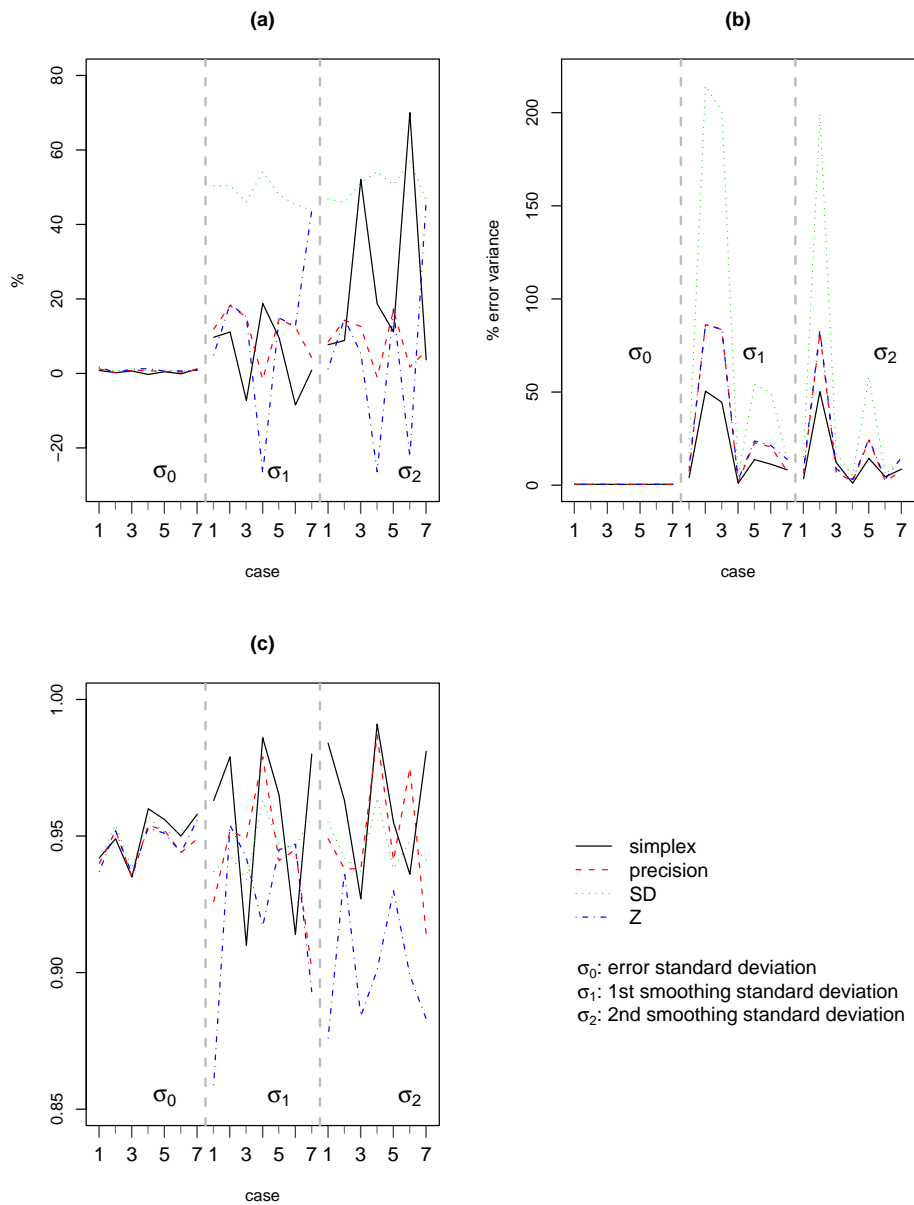


Figure 8: Crossed RE simulation: standard deviation bias (as a percent of true standard deviation) and MSE (divided by the true error variance  $\frac{1}{\tau_0}$ ). (a) scaled bias for  $\sigma_0$ ,  $\sigma_1$ , and  $\sigma_2$ ; (b) scaled MSE for  $\sigma_0$ ,  $\sigma_1$ , and  $\sigma_2$ ; (c) 95% interval coverage for standard deviations  $\sigma_0$ ,  $\sigma_1$ , and  $\sigma_2$ .