Open access • Journal Article • DOI:10.1111/1541-0420.00035

Bayesian inference on order-constrained parameters in generalized linear models.

David B. Dunson, Brian Neelon

Institutions: National Institutes of Health

Published on: 01 Jun 2003 - Biometrics (Blackwell Publishing)

Topics: Bayesian linear regression, Gibbs sampling, Bayesian inference, Bayesian statistics and Bayes factor

Bayesian Inference on Order Constrained Parameters in Generalized Linear Models

David B. Dunson[1,*] and Brian Neelon[2]

July 31, 2002

[1] Biostatistics Branch, National Institute of Environmental Health Sciences, MD A3-03, P.O. Box 12233, Research Triangle Park, NC 27709

[2] Department of Biostatistics, University of North Carolina, Chapel Hill, NC

[*] *email*: dunson1@niehs.nih.gov

SUMMARY. In biomedical studies, there is often interest in assessing the association between one or more ordered categorical predictors and an outcome variable, adjusting for covariates. For a k-level predictor, one typically uses either a $k-1$ degree of freedom (df) test or a single df trend test, which requires scores for the different levels of the predictor. In the absence of knowledge of a parametric form for the response function, one can incorporate monotonicity constraints to improve the efficiency of tests of association. This article proposes a general Bayesian approach for inference on order constrained parameters in generalized linear models. Instead of choosing a prior distribution with support on the constrained space, which can result in major computational difficulties, we propose to map draws from an unconstrained posterior density using an isotonic regression transformation. This approach allows flat regions over which increases in the level of a predictor have no effect. Bayes factors for assessing ordered trends can be computed based on the output from a Gibbs sampling algorithm. Results from a simulation study are presented and the approach is applied to data from a time to pregnancy study.

KEY WORDS: Bayes factor; Constrained estimation; Categorical covariates; Gibbs sampling; Isotonic regression; Monotonicity; Simple ordering; Trend test.

1. Introduction

In analyzing data from epidemiologic studies investigating the relationship between a k-level ordered categorical predictor and an outcome variable, one typically uses a generalized linear model that includes indicator variables for the different levels of the predictor and adjusts for covariates. Maximum likelihood estimates of the regression coefficients can be obtained, and a $k-1$ degree of freedom test can be used to assess overall evidence of an association between the predictor and the response. When following this strategy, analysts are often faced with noisy estimates of the category-specific regression coefficients. This noise can lead to unreasonable patterns in the regression coefficients corresponding to different levels of the predictor, and to low power for the $k-1$ degree of freedom test and for one degree of freedom contrasts. In order to improve efficiency of overall tests for an association, one can take advantage of known orderings in the regression coefficients. For example, in many applications, it is reasonable to assume that the regression coefficients for the different levels of a predictor follow a simple non-decreasing order.

This article proposes a Bayesian approach for incorporating order constraints on the regression coefficients in a generalized linear model. In the Bayesian paradigm, order constraints are typically imposed by choosing a prior distribution that has support on a restricted space. Compared with the vast literature on classical approaches for order restricted inference (cf., Robertson, Wright and Dykstra, 1988), few Bayesian methods have been proposed. However, as noted by Gelfand, Smith and Lee (1992), certain types of parameter constraints can easily be incorporated in Bayesian analyses through the use of truncated conjugate priors. In particular, it is straightforward to accommodate strict constraints (e.g., $\theta_1 < \theta_2$) in an analysis that uses Gibbs sampling for posterior computation by first choosing a prior density without considering the constraint (e.g., normal) and then discarding draws inconsistent with the constraint. Although such an approach is very useful in certain applications, it

does not allow for equalities in the parameters (e.g., $\theta_1 = \theta_2$), and hence is not appropriate for inferences on uncertain orderings. For example, one may be interested in assessing evidence of an increase in the probability of disease associated with increases in the level of an exposure.

To allow for flat regions over which increases in the level of a predictor have no effect on the response distribution, one could potentially choose a prior distribution for the category-specific regression coefficients that allocates positive probabilities to equalities, such as $\theta_1 = \theta_2$. However, such priors typically have non-conjugate structures and posterior computation may be difficult, particularly in high dimensions. In fact, even in low dimensional problems, we have noted substantial difficulties in posterior computation. Markov chain Monte Carlo (MCMC) algorithms (cf., Chen, Shao and Ibrahim, 2000), such as the Metropolis-Hastings algorithm, tend to have high autocorrelation and low efficiency due to the constraints on the parameters. Motivated by these difficulties and by frequentist approaches to order restricted inference, which apply an isotonic regression transformation to the unconstrained parameter estimates, this article proposes an alternative approach.

Following Gelfand et al. (1992), we first choose a prior for the regression coefficients without considering the constraint. We then update the prior to obtain an unconstrained posterior density, which is available in closed form for conjugate models and can otherwise be estimated via Gibbs sampling (Gelfand and Smith, 1990). We consider the constrained parameters as functions of the unconstrained parameters, calculated by applying the isotonic regression transformation. Under this framework, samples from the constrained posterior density can be obtained by transforming draws from the unconstrained posterior density. Hence, existing Gibbs sampling algorithms for posterior computation of GLMs can be used directly.

This approach is a Bayesian alternative to likelihood ratio (LR) tests constructed using

restricted maximum likelihood estimates (RMLEs) obtained under an order-constrained alternative hypothesis (cf., Bartholomew, 1961; Robertson et al., 1988; Agresti and Coull, 1996). As noted by Mukerjee and Tu (1993), the important problem of order-restricted inferences on regression coefficients, which is our focus, has received relatively little attention. Methods have been proposed for incorporating monotonicity constraints in the setting of generalized additive models (cf., Bacchetti, 1989; Morton-Jones et al., 2000), and these methods can be used for testing hypotheses about increasing response functions. Our goal is a general methodology for inferences on arbitrary order constraints on parameters in hierarchical models, with the immediate emphasis on regression parameters in GLMs. By following a Bayesian approach, it becomes straightforward to incorporate prior information and to perform inferences that do not rely on asymptotic approximations.

Section 2 proposes an approach for assessing evidence of ordering in a vector of normal means. Section 3 extends the approach to assess ordered trends in regression parameters in the setting of linear regression, probit models for categorical data, and generalized linear models. Section 4 applies the approach to data from a time to pregnancy study, and Section 5 discusses the results.

2. Ordered Normal Means

2.1 *Isotonic Transformation Approach*

Let $\mathbf{y}_j$ denote an $n_j \times 1$ vector of iid $\mathrm{N}(\mu_j, \sigma^2)$ random variables, for $j = 1, \ldots, k$. Assuming a conjugate prior density,

$$\theta_j|\sigma^2 \sim \mathrm{N}(\theta_{0j}, \sigma^2/\kappa_{0j}), j = 1, \ldots, k, \quad \text{and} \quad \sigma^2 \propto (\sigma^2)^{-1},$$

the joint posterior density of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)'$ and σ^2 is the product of

$$\theta_j|\sigma^2, \mathbf{y} \sim \mathrm{N}(\widehat{\theta}_j, \widehat{\sigma}^2_j), j = 1, \ldots, k, \quad \text{and} \quad \sigma^2|\mathbf{y} \sim \text{scaled Inv-}\chi^2(\nu_n, \sigma^2_n), \tag{1}$$

where $\widehat{\theta}_j = (\kappa_{0j}\theta_{0j} + n_j\overline{y}_j)/(\kappa_{0j} + n_j)$, $\widehat{\sigma}_j^2 = \sigma^2/(\kappa_{0j} + n_j)$, $\nu_n = n$, and

$$\nu_n\sigma_n^2 = \sum_{j=1}^{k}(n_j - 1)s_j^2 + \frac{\kappa_{0j}n_j}{\kappa_{0j} + n_j}(\overline{y}_j - \theta_{0j})^2,$$

with s_j^2 denoting the empirical variance of $\mathbf{y}_j$.

We are interested in an order constrained functional of $\boldsymbol{\theta}$, which we denote $\boldsymbol{\theta}^* = g(\boldsymbol{\theta}) \in \boldsymbol{\Theta} \subset \Re^k$, where $g(\cdot)$ is the isotonic regression transformation. In particular, letting $\widehat{\boldsymbol{\Sigma}} = \text{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_k^2)$ denote the posterior covariance, we choose the following max-min formula (Robertson et al., 1988; Hwang and Peddada, 1994):

$$\theta_j^* = g_j(\boldsymbol{\theta}) = \min_{t \in U_j} \max_{s \in L_j} \left(\frac{\widehat{\boldsymbol{\Sigma}}_{[s:t]}^{-1}\boldsymbol{\theta}_{[s:t]}}{\widehat{\boldsymbol{\Sigma}}_{[s:t]}^{-1}\mathbf{1}_{t-s+1}} \right), \quad \text{for } j = 1, \ldots, k, \tag{2}$$

with L_j and U_j denoting subsets of $\{1, \ldots, k\}$ such that the ordering $\theta_{j'}^* \leq \theta_j^*$ is known for all $j' \in L_j$ and the ordering $\theta_{j'}^* \geq \theta_j^*$ is known for all $j' \in U_j$. The $[s:t]$ subscript indicates the submatrices and subvectors corresponding to elements $s, s+1, \ldots, t$. The isotonic regression transformation (2) sets θ_j^* equal to a weighted average of a subvector of $\boldsymbol{\theta}$, where the weights depend on the posterior covariance $\widehat{\boldsymbol{\Sigma}}$ and the elements of the subvector are chosen to satisfy the constraint.

The restricted maximum likelihood estimate (RMLE) is obtained by applying transformation (2) to the unrestricted maximum likelihood estimate, $(\overline{y}_1, \ldots, \overline{y}_k)'$, with $\kappa_{0j} = 0$ for $j = 1, \ldots, k$ (refer to Example 1.2.1 in Robertson et al., 1988). Formula (2) is a least squares projection from $\Re^k$ to the restricted space $\boldsymbol{\Theta}$. We obtain draws from the posterior density for the restricted parameter, $\boldsymbol{\theta}^*$, by transforming draws from the posterior density for the unrestricted parameter, $\boldsymbol{\theta}$, using a minimal distance mapping. In the limit as the sample size $\to \infty$, the posterior mode for $\boldsymbol{\theta}^*$ is equivalent to the RMLE. When the prior variance is large, we expect that the posterior mode will be similar to the RMLE even in small to moderate samples.

Suppose that interest focuses on assessing evidence to reject $H_0 : \theta_1^* = \theta_2^* = \ldots = \theta_k^*$ in favor of $H_1 : \theta_1^* \leq \theta_2^* \leq \ldots \leq \theta_k^*$ with at least one strict ordering. Note that H_0 and H_1 are equivalent to $H_0 : \theta_1^* = \theta_k^*$ and $H_1 : \theta_1^* < \theta_k^*$ under the constraint that $\boldsymbol{\theta}^* \in \boldsymbol{\Theta}$, with $\boldsymbol{\Theta} = \{\boldsymbol{\theta}^* : \theta_1^* \leq \ldots \leq \theta_k^*\}$. As a measure of weight of evidence in favor of H_1, we suggest the Bayes factor (cf., Kass and Raftery, 1995),

$$BF_{10} = \frac{\Pr(H_1 \,|\, \mathbf{y})/\Pr(H_1)}{\Pr(H_0 \,|\, \mathbf{y})/\Pr(H_0)} = \frac{\Pr(\theta_1^* < \theta_k^* \,|\, \mathbf{y})/\Pr(\theta_1^* < \theta_k^*)}{\Pr(\theta_1^* = \theta_k^* \,|\, \mathbf{y})/\Pr(\theta_1^* = \theta_k^*)}, \tag{3}$$

where $\Pr(\theta_1^* < \theta_k^*)$ is the prior probability of H_1 and $\Pr(\theta_1^* < \theta_k^* \,|\, \mathbf{y})$ is the posterior probability. Unlike the posterior probability, the Bayes factor is adjusted for the prior odds of H_1. Hence, BF_{10} is a more direct measure of the weight of evidence in the current data, $\mathbf{y}$, in favor of H_1. The quantities needed to calculate the Bayes factor are derived in Appendix A. It is straightforward to show that $BF_{10} \to \infty$ for large samples and finite prior precision under H_1.

2.2 *Simulation Study*

To study the behavior of the procedure, we conducted a small simulation study. In particular, we simulated data under 3 different patterns in $\boldsymbol{\theta}$:

1. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)' = \mathbf{0}$ [Null hypothesis, H_0],
2. $\boldsymbol{\theta} = (0, 1/(k-1), 2/(k-1), \ldots, 1)'$ [Linearly increasing, H_1],
3. $\boldsymbol{\theta} = (0, \ldots, 0, 1)'$ [Threshold, H_1].

For each pattern and for different choices of n_j (5,10,25), k (3,5) and $\kappa_{0j}(0.1, 1.0)$, we simulated 500 data sets. Then, for each simulated data set, we calculated BF_{10} assuming $\boldsymbol{\theta}_0 = \mathbf{0}$. To assess the frequentist operating characteristics, we used $BF_{10} > 0.95/0.05 = 19$ as the cutoff for significance. This choice should produce results comparable to an $\alpha = 0.05$-level test.

The simulation results are reported in Table 1, along with the results from a separate sim-

ulation evaluating the performance of the likelihood ratio (LR) test (cf., Robertson et al., 1988, p 63). Regardless of the values of k and n_j and the choice of κ_{0j}, the type I error rates were close to the 0.05 nominal level for both the proposed Bayesian test and the LR test, with the error rates of the Bayes procedure slightly lower on average. The power of the two test procedures was similar overall, with the LR test having higher power under the linear alternative and the Bayes test having higher power under the threshold alternative. The power of the Bayes procedure was slightly higher when a more diffuse prior was chosen (i.e., for $\kappa_{0j} = 0.1$).

3. Regression Models

3.1 *Linear Regression*

It is straightforward to generalize the procedure described in Section 2 to perform inferences on orderings in regression parameters. Suppose data for each study subject consist of a response, y_i, m ordered categorical predictors, $c_{i1}, \ldots, c_{im}$, and an additional p predictors that are either continuous or binary, $\mathbf{w}_i = (w_{i1}, \ldots, w_{ip})'$. For subject i $(i = 1, \ldots, n)$, let $\mathbf{u}_{il} = (1_{(c_{il}=2)}, \ldots, 1_{(c_{il}=k_l)})'$ denote a $(k_l - 1) \times 1$ vector of dichotomous indicators for the level of the lth categorical predictor, where $c_{il} \in \{1, \ldots, k_l\}$. We focus on the linear regression model,

$$y_i = \beta_0 + \sum_{l=1}^{m} \mathbf{u}_{il}'\boldsymbol{\beta}_l + \mathbf{w}_i'\boldsymbol{\alpha} + \epsilon_i = \mathbf{x}_i'\boldsymbol{\theta} + \epsilon_i, \tag{4}$$

where β_0 is an intercept parameter, $\boldsymbol{\beta}_l = (\beta_{l1}, \ldots, \beta_{l,k_l-1})'$ are coefficients for the different levels of c_{il}, $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}_1', \ldots, \boldsymbol{\beta}_m')'$, $\boldsymbol{\alpha}$ are coefficients for $\mathbf{w}_i$, $\mathbf{x}_i = (1, \mathbf{u}_{i1}', \ldots, \mathbf{u}_{im}', \mathbf{w}_i)'$, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$, and $\epsilon_i \sim \mathrm{N}(0, \tau^{-1})$. We assume conjugate priors for $\boldsymbol{\theta}$ and τ, $\boldsymbol{\theta} \sim \mathrm{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$ and $\tau \sim \mathrm{gamma}(a_0, b_0)$.

Let $g_{\boldsymbol{\Theta}_l}(\cdot)$ denote the isotonic regression transformation mapping from

$$\Re^{k_l-1} \rightarrow \boldsymbol{\Theta}_l = \{\boldsymbol{\beta}_l^* : 0 \leq \beta_{l1}^* \leq \ldots \leq \beta_{l,k_l-1}^*\},$$

and let $\boldsymbol{\beta}_l^* = g_{\boldsymbol{\Theta}_l}(\boldsymbol{\beta}_l)$. The transformation $g_{\boldsymbol{\Theta}_l}(\cdot)$ produces a $(k_l - 1) \times 1$ vector, with elements defined in expression (2), but with a different form for the posterior covariance (as described below).

Often, interest focuses on assessing evidence of an increasing trend in the mean response with increases in a categorical predictor, controlling for other factors. In particular, we may want to summarize evidence in the data for rejecting

$$\mathrm{H}_{0l} : 0 = \beta_{l1}^* = \ldots = \beta_{l,k_l-1}^* \text{ (equivalent to } \beta_{l,k_l-1}^* = 0 \text{ for } \boldsymbol{\beta}_l \in \boldsymbol{\Theta}_l)$$

in favor of the alternative hypothesis

$$\mathrm{H}_{1l} : 0 \leq \beta_{l1}^* \leq \ldots \leq \beta_{l,k_l-1}^* \text{ with at least one strict inequality}$$

(equivalent to $\beta_{l,k_l-1}^* > 0$ for $\boldsymbol{\beta}_l \in \boldsymbol{\Theta}_l$). Generalizing the procedure of Section 2, we recommend the Bayes factor,

$$\mathrm{BF}_l = \frac{\Pr(\mathrm{H}_{1l} \,|\, \mathbf{y}, \mathbf{X})/\Pr(\mathrm{H}_{1l})}{\Pr(\mathrm{H}_{0l} \,|\, \mathbf{y}, \mathbf{X})/\Pr(\mathrm{H}_{0l})} = \frac{\Pr(\beta_{l,k_l-1}^* > 0 \,|\, \mathbf{y}, \mathbf{X})/\Pr(\beta_{l,k_l-1}^* > 0)}{\Pr(\beta_{l,k_l-1}^* = 0 \,|\, \mathbf{y}, \mathbf{X})/\Pr(\beta_{l,k_l-1}^* = 0)}, \tag{5}$$

where $\mathbf{X} = (\mathbf{x}_1', \ldots, \mathbf{x}_n')'$ is the design matrix.

The Bayes factor, BF_l, can be estimated using the following procedure:

1. Draw samples from the posterior density of $(\boldsymbol{\theta}, \tau)$ using a Gibbs sampler, which alternately samples from the conditional densities,

   $$[\boldsymbol{\theta} \,|\, \tau, \mathbf{y}, \mathbf{X}] = \mathrm{N}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Sigma}}) \quad \text{and} \quad [\tau \,|\, \boldsymbol{\theta}, \mathbf{y}, \mathbf{X}] = \mathrm{gamma}\Big(a_0 + \frac{n}{2}, b_0 + \frac{1}{2}\sum_{i=1}^{n}(y_i - \mathbf{x}_i'\boldsymbol{\theta})^2\Big),$$

   where $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\Sigma}}(\boldsymbol{\Sigma}_0^{-1}\boldsymbol{\theta}_0 + \tau\mathbf{X}'\mathbf{y})$ and $\widehat{\boldsymbol{\Sigma}} = (\boldsymbol{\Sigma}_0^{-1} + \tau\mathbf{X}'\mathbf{X})^{-1}$.

2. Draw samples from the prior density of $\boldsymbol{\theta}$ by sampling from $\mathrm{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$.

3. Apply the isotonic regression transformation to each sample from steps 1 and 2 to obtain draws from the posterior and prior densities for $\boldsymbol{\beta}_l^*$, respectively $(l = 1, \ldots, m)$.

4. Estimate the posterior and prior probabilities in expression (5) as proportions of the draws for β^*_{l,k_l-1} equal to greater than 0, for $l = 1, \ldots, m$.

3.2 *Probit Models for Categorical Data*

Now suppose that $y_1, \ldots, y_n$ are independent Bernoulli distributed random variables, with

$$\Pr(y_i \,|\, \mathbf{x}_i, \boldsymbol{\theta}) = \Phi\Big(\beta_0 + \sum_{l=1}^{m} \mathbf{u}'_{il}\boldsymbol{\beta}_l + \mathbf{w}'_i\boldsymbol{\alpha}\Big) = \Phi(\mathbf{x}'_i\boldsymbol{\theta}). \tag{6}$$

As noted by Albert and Chib (1993), this is equivalent to assuming that $y_i = 1_{(z_i>0)}$, with $z_i \sim \mathrm{N}(\mathbf{x}'_i\boldsymbol{\theta}, 1)$ denoting independent and normally distributed random variables underlying y_i, for $i = 1, \ldots, n$. Under the conditionally conjugate prior density for $\boldsymbol{\theta}$ defined in subsection 3.1, posterior computation can proceed via a Gibbs sampler that alternates between (i) sampling from the conditional density of z_i,

$$[z_i \,|\, y_i, \mathbf{x}_i, \boldsymbol{\theta}] = \mathrm{N}(\mathbf{x}'_i\boldsymbol{\theta}, 1) \quad \text{truncated below (above) by 0 for } y_i = 1 \ (y_i = 0),$$

and (ii) sampling from the conditional density of $\boldsymbol{\theta}$, $[\boldsymbol{\theta} \,|\, \mathbf{z}, \mathbf{y}, \mathbf{X}] = \mathrm{N}(\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\Sigma}})$, where $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are as defined in subsection 3.1, with z_i substituted for y_i and $\tau = 1$ for identifiability. The algorithm outlined in subsection 3.1 can be used to estimate order-constrained parameters and to compute Bayes factors for order restricted inferences. It is straightforward to extend this approach for categorical y_i by following the approach of Albert and Chib (1993).

3.3 *Generalized Linear Models*

Suppose that $y_1, \ldots, y_n$ are independent random variables from a distribution in the exponential family, with τ denoting a scalar dispersion parameter (possibly known) and with

canonical parameter ψ_i related to covariates through the generalized linear model,

$$h(\psi_i) = \eta_i = \beta_0 + \sum_{l=1}^{m} \mathbf{u}_{il}'\boldsymbol{\beta}_l + \mathbf{w}_i'\boldsymbol{\alpha} = \mathbf{x}_i'\boldsymbol{\theta}. \tag{7}$$

Assuming a $\mathrm{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$ prior density for $\boldsymbol{\theta}$ and a log-concave prior for τ, posterior computation for $(\boldsymbol{\theta}, \tau)$ can proceed via an adaptive rejection Gibbs sampling algorithm (Dellaportas and Smith, 1993). For normal linear regression and probit models, the unconstrained posterior covariance used to weight the unconstrained parameters in expression (2) is known. However, for other generalized linear models, such as logistic or log-linear regression models, the posterior covariance must be estimated. We recommend first drawing a large number of samples from the posterior density of $\boldsymbol{\theta}$ and then setting $\widehat{\boldsymbol{\Sigma}}$ equal to the empirical covariance. The estimate $\widehat{\boldsymbol{\Sigma}}$ can then be used within expression (2) to weight the elements of $\boldsymbol{\theta}$ in mapping additional draws from the posterior density of $\boldsymbol{\theta}$ onto $\boldsymbol{\Theta}$.

4. Application to Discrete Time to Event Data

4.1 *Time to Pregnancy Data and Model*

We illustrate the methodology through application to time to pregnancy data from a study of dental assistants exposed to high levels of nitrous oxide (Rowland et al., 1992). Female dental assistants, aged 19 to 39, were randomly selected from the dental-assistants registry of the California Department of Consumer Affairs and invited to participate in the study if they met eligibility criteria. These criteria included having a planned pregnancy within the past four years and having worked full time during the six months prior to having non-contracepting intercourse. Of the 459 eligible women who completed the screening questionnaire, 428 provided detailed data on reproductive and contraceptive history, occupational exposures, and other factors related to fertility. Time to pregnancy was ascertained by calculating the number of menstrual cycles before the most recent pregnancy during which the woman was having non-contracepting sexual intercourse (cf., Rowland et al., 1992). Summary statistics

are provided in Table 2.

The goal of our analysis was to incorporate known order restrictions to improve efficiency in assessing covariate effects on fecundability, the probability of conception in a menstrual cycle. We were particularly interested in the effect of smoking, which was borderline significant [95% confidence interval = (.27,1.06)] in a previous unconstrained maximum likelihood analysis (Rowland et al., 1992). To this end, we modelled the number of menstrual cycles until conception using a discrete-time hazard model that included 23 parameters measuring (i) the baseline hazard of conception for menstrual cycles 1-13 (individuals not conceiving by cycle 13 were censored); (ii) age; (iii) intercourse frequency; (iv) cigarettes smoked per day; and (v) use of oral contraceptives in the cycle prior to beginning the pregnancy attempt.

Let $\mathbf{y} = (y_1, \ldots, y_n)'$ denote the conception outcomes for the $n = 1968$ non-contracepting menstrual cycles under study, where $y_i = 1$ if conception occurs in cycle i and $y_i = 0$ otherwise. We assume that

$$\Pr(y_i \,|\, \mathbf{x}_i, \boldsymbol{\beta}) = \Phi\Big(\beta_0 + \sum_{l=1}^{5} \mathbf{u}_{il}'\boldsymbol{\beta}_l\Big) = \Phi(\mathbf{x}_i'\boldsymbol{\theta}), \tag{8}$$

where $\mathbf{u}_{i1} = (u_{i11}, \ldots, u_{i1,12})'$ is a vector of 0/1 indicators of the cycle at risk for a women, $\mathbf{u}_{i2}, \mathbf{u}_{i3}, \mathbf{u}_{i4}$ and $\mathbf{u}_{i5}$ are vectors indicating, respectively, the category of age $(19-24, 25-29, >30)$, intercourse frequency $(<= 1, 1-3, 3-4, > 4$ times/week), cigarette smoking frequency (nonsmoker, $1-5, 6-10, 11-15, > 15$ cigs/day), and oral contraceptive use in the previous cycle (no, yes).

A Bayesian specification of the model is completed with a conditionally conjugate $\mathrm{N}(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$ prior density for the regression parameters. The prior mean for the parameters, $(\beta_0, \boldsymbol{\beta}_1)$, characterizing the baseline time to pregnancy distribution was set equal to the posterior mean from an analysis of the Tietze (1968) data using model (8). The prior covariance for these parameters was set equal to the estimated posterior covariance from the Tietze

analysis, with the variance inflated by a factor of 10 to accommodate possible differences between the study populations. The prior mean for the remaining parameters was set to 0, while the prior covariance matrix was chosen to be diag$(20, \ldots, 20)$.

We are interested in inferences on the order-constrained functional, $\boldsymbol{\theta}^* = g_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$, of the unconstrained regression coefficients, where

$$\boldsymbol{\Theta} = \{\boldsymbol{\beta}^* : 0 \geq \beta^*_{11} \geq \ldots \geq \beta^*_{1,12}, 0 \leq \beta^*_{31} \leq \beta^*_{32} \leq \beta^*_{33}, 0 \geq \beta^*_{41} \geq \ldots \geq \beta^*_{44}, \beta^*_5 \leq 0\}.$$

The ordering in $\boldsymbol{\beta}^*_1$ is consistent with published data from Tietze (1968), Wilcox et al. (1988) and Rowland et al. (1992). Non-increasing elements of $\boldsymbol{\beta}^*_1$ imply that the discrete hazard of conception is non-increasing with increases in the number of menstrual cycles at risk. This reflects the selection process by which women with relatively high fertility conceive rapidly and are removed from later risk sets. The constraint on $\boldsymbol{\beta}^*_3$ ensures that the hazard of conception is non-decreasing with increasing frequency of intercourse. Finally, the constraints on $\boldsymbol{\beta}^*_4$ and $\boldsymbol{\beta}^*_5$ restrict the hazard to be non-increasing with increasing number of cigarettes smoked per day and with recent pill use.

4.2 *Results*

Samples from the posterior density of the unconstrained regression coefficients, $\boldsymbol{\theta}$, were obtained using the Gibbs sampler of Albert and Chib (1993), as described in subsection 2.2. We ran the Gibbs sampler for 5,000 iterations, and discarded the first 500 as a burn-in. The isotonic regression transformation from $\Re^{23} \rightarrow \boldsymbol{\Theta}$ was applied to each sample after the burn-in to obtain samples from the posterior density of $\boldsymbol{\theta}^*$. Posterior summaries of the order-constrained baseline conception rates are plotted in Figure 1, along with posterior means for the unconstrained parameters and unconstrained maximum likelihood estimates. As expected, posterior means of the constrained parameters varied relatively smoothly with menstrual cycle number. There was no evidence of systematic differences between the es-

timates, suggesting that the order restriction is supported by the data. In addition, the posterior variance for the constrained parameters was 41% lower (on average) than that for the unconstrained parameters, suggesting an improvement in efficiency.

Table 3 presents posterior summaries of the regression coefficients characterizing the covariate effects. Table 3 also provides estimated Bayes factors for testing the hypotheses of interest; namely, decreases in the hazard of conception with increases in smoking frequency and with occurrence of recent pill use. For the highest category of cigarette smoking (> 15 cigarettes smoked / day) there was strong evidence of decreased fecundability relative to the non-smokers, with the Bayes factor equal to 28.76. This Bayes factor corresponds roughly to a p-value of 0.03 in terms of weight of evidence against the null hypothesis. In constrast, the p-value from the unconstrained maximum likelihood analysis presented by Rowland et al. (1992) was > 0.05 as was the Bayesian p-value (i.e., posterior probability) from the unconstrained analysis. These results suggest an improvement in power for our order-restricted Bayesian procedure, since an association with smoking and prolonged time to pregnancy has been shown repeatedly in other studies. Both the constrained and unconstrained analyses showed clear evidence of lower fecundability for recent pill users, as anticipated.

To assess the sensitivity of the results to the choice of hyperparameters, we repeated the analyses for different choices of prior. In particular, we considered priors for $\boldsymbol{\theta}$ with (i) precision 0.01 instead of 0.05; (ii) precision 0.1 instead of 0.05; and (iii) with prior mean

$$\boldsymbol{\theta}_0 = (0, \ldots, 0, -0.2, -0.5, 0.2, 0.4, 0.6, -0.1, -0.2, -0.3, -0.4, -0.1)'.$$

Essentially identical results were obtained in each of our sensitivity analyses, with the posterior means (standard deviations) all within $\pm 0.02(0.01)$ of the values reported in Table 3.

5. Discussion

This article has proposed a general and easy-to-implement approach for assessing orderings in regression parameters. Instead of defining a prior distribution with support on the constrained space, as in earlier methodology (cf., Chen and Shao, 1998; Gelfand and Kuo, 1991; Lavine and Mockus, 1995; Ramgopal, Laud and Smith, 1993), we consider the order-constrained parameters to be functions of parameters having support on $\Re^k$. A conjugate or conditionally-conjugate prior density is chosen and draws from the resulting unconstrained posterior density are mapped to the constrained space through use of an isotonic regression transformation. Posterior summaries of the order-constrained parameters and Bayes factors for testing of ordered trends can be calculated in closed form for the case of ordered normal means or directly from the output of Gibbs sampling algorithms for general regression problems. Although we have focused on the problem of assessing simple ordering in regression coefficients for different categories of an ordinal predictor, the approach can be applied directly for other types of orderings (e.g., simple tree) and parameters (e.g., scale parameters or random-effects).

For the simple problem of assessing ordering in normal means, the proposed Bayesian procedure yields very similar inferences to a likelihood ratio test when a diffuse prior is chosen. An advantage of the Bayesian procedure in this context is that exact posterior densities of the parameters can be easily estimated. In addition, it is straightforward to generalize the Bayesian procedure to arbitrary regression problems, while generalizations of the likelihood ratio test are limited by the difficulty of computing restricted maximum likelihood estimates in many cases.

Our approach is motivated primarily by convenience in computational implementation for a wide variety of problems. For example, with minimal programming effort in S-PLUS (Mathsoft, Inc., 1999), Bayes factors and posterior summaries can be calculated directly from the

MCMC iterates output from WinBUGS (Lunn et al., 2000). Therefore, the procedure provides a useful black box approach for Bayesian inference on order constrained parameters in hierarchical models. It should be noted that, from a Bayesian perspective, it is somewhat more natural to incorporate order constraints through an appropriately defined prior instead of through a transformation. However, posterior computation is often prohibitively difficult when the prior density has constrained support, and general algorithms remain to be developed. As we have demonstrated through a simulation study and through application to reproductive epidemiology data, our approach is a reasonable alternative.

ACKNOWLEDGEMENTS

The authors would like to thank Zhen Chen, Sean O'Brien and Shyamal Peddada for helpful comments and discussions. Thanks also to Andrew Rowland for providing the data for the example.

Appendix A

Derivation of Bayes Factor for Testing Ordering in Normal Means

From the definition of the isotonic regression transformation, it is apparent that $\theta_1^* = \ldots = \theta_k^*$ if and only if $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_k$. Letting $\delta_j = \theta_j - \theta_{j+1}$, for $j = 1, \ldots, k-1$, the inequality $\delta_j \geq 0$ implies that $\theta_j \geq \theta_{j+1}$. The conditional posterior density of $\delta_j | \sigma^2$ is

$$\delta_j \,|\, \sigma^2, \mathbf{y} \sim \mathrm{N}(\widehat{\theta}_j - \widehat{\theta}_{j+1}, \widehat{\sigma}_j^2 + \widehat{\sigma}_{j+1}^2), \quad j = 1, \ldots, k-1,$$

where $\widehat{\sigma}_j^2 = \sigma^2/(\kappa_{0j} + n_j)$. Hence, conditional on σ^2, the posterior probability of $\delta_j \geq 0$ is

$$\Pr(\delta_j \geq 0 \,|\, \sigma^2, \mathbf{y}) = \Pr(\theta_j \geq \theta_{j+1} \,|\, \sigma^2, \mathbf{y}) = \Phi\left(\frac{\widehat{\theta}_j - \widehat{\theta}_{j+1}}{\sqrt{\widehat{\sigma}_j^2 + \widehat{\sigma}_{j+1}^2}} \right),$$

where $\Phi(\cdot)$ is the standard normal distribution function. Taking one minus the product across the different elements of $\{\delta_1, \ldots, \delta_{k-1}\}$, we obtain

$$1 - \Pr(\theta_1 \geq \ldots \geq \theta_k \,|\, \sigma^2, \mathbf{y}) = \Pr(\mathrm{H}_1 \,|\, \sigma, \mathbf{y}) = 1 - \prod_{j=1}^{k-1} \Phi\left(\frac{\widehat{\theta}_j - \widehat{\theta}_{j+1}}{\sqrt{\widehat{\sigma}_j^2 + \widehat{\sigma}_{j+1}^2}}\right).$$

A parallel approach can be used to derive the prior probability, $\Pr(\mathrm{H}_1|\sigma^2)$. The Bayes factor can be calculated by plugging in the prior and posterior probabilities, conditional on σ^2, into expression (3) and then removing the conditioning by numerical integration across the scaled-Inv-$\chi^2(\nu_n, \sigma_n^2)$ marginal posterior density for σ^2.

References

Agresti, A. and Coull, B.A. (1996). Order-restricted tests for stratified comparisons of binomial proportions. *Biometrics* **52**, 1103-1111.

Albert, J.H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**, 669-679.

Bartholomew, D.J. (1961). A test of homogeneity of means under restricted alternatives (with discussion). *Journal of the Royal Statistical Society B* **23**, 239-281.

Chen, M.-H. and Shao, Q.-M. (1998) Monte Carlo methods on Bayesian analysis of constrained parameter problems. *Biometrika* **85**, 73-87.

Chen, M.-H., Shao, Q.-M. and Ibrahim, J.G. (2000) *Monte Carlo Methods in Bayesian Computation*. New York: Springer.

Dellaportas, P. and Smith, A.F.M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *Applied Statistics* **42**, 443-459.

Gelfand, A.E. and Kuo, L. (1991) Nonparametric Bayesian bioassay including ordered polytomous response. *Biometrika* **78**, 657-666.

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

Gelfand, A.E., Smith, A.F.M., and Lee, T.M. (1992) Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association* **87**, 523-532.

Hwang, J.T.G. and Peddada, S.D. (1994). Confidence interval estimation subject to order restrictions. *Annals of Statistics* **22**, 67-93.

Kass and Raftery (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.

Lavine, M. and Mockus, A. (1995). A nonparametric Bayes method for isotonic regression. *Journal of Statistical Planning and Inference* **46**, 235-248.

Lunn, D.J., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WINBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing* **10**, 325-337.

Mathsoft, Inc. (1999). *S-PLUS 5 for UNIX Guide to Statistics*, Data Analysis Products Division, Mathsoft, Seattle.

McDermott, M.P. and Mudholkar, G.S. (1993). A simple approach to testing homogeneity of order-constrained means. *Journal of the American Statistical Association* **88**, 1371-1379.

Morton-Jones, T., Diggle, P., Parker, L., Dickinson, H.O. and Binks, K. (2000). Additive isotonic regression models in epidemiology. *Statistics in Medicine* **19**, 849-859.

Mukerjee, H. and Tu, R. (1995). Order-restricted inferences in linear regression. *Journal of the American Statistical Association* **90**, 717-728.

Ramgopal, P., Laud, P.W. and Smith, A.F.M. (1993). Nonparametric Bayesian bioassay with prior constraints on the shape of the potency curve. *Biometrika* **80**, 489-498.

Robertson, T., Wright, F. and Dykstra, R. (1988). *Order Restricted Statistical Inference*. New York: Wiley.

Rowland, A.S., Baird, D.D., Weinberg, C.R., Shore, D.L., Shy, C.M., and Wilcox, A.J. (1992). Reduced fertility among women employed as dental assistants exposed to high levels of nitrous oxide. *New England Journal of Medicine* **327**, 993-997.

Wilcox, A.J., Weinberg, C.R., O'Connor, J.F., Baird, D.D., Shlatterer, J.P., Canfield, R.E. et al. (1988). Incidence of early loss of pregnancy. *New England Journal of Medicine* **319**, 189-194.

Tietze, C. (1968) Fertility after discontinuation of intrauterine and oral contraception. *International Journal of Fertility* **13**, 385-389.

Table 1

Results from simulation study of the frequentist operating characteristics of the proposed Bayes Factor approach for assessing ordering in normal means.

						Rejection Rate	
Hypothesis	Pattern	k	n_j	κ_{0j}	$\widehat{BF}_{10}$	Bayes[†]	LR test[‡]
H_0	1	3	5	0.1	1.57	0.07	0.07
H_0	1	3	5	1.0	1.54	0.05	
H_0	1	3	10	0.1	1.44	0.04	0.05
H_0	1	3	10	1.0	1.46	0.04	
H_0	1	3	25	0.1	1.60	0.03	0.05
H_0	1	3	25	1.0	1.49	0.03	
H_0	1	5	5	0.1	1.81	0.05	0.04
H_0	1	5	5	1.0	1.68	0.02	
H_0	1	5	10	0.1	1.77	0.03	0.04
H_0	1	5	10	1.0	1.80	0.03	
H_0	1	5	25	0.1	1.93	0.04	0.06
H_0	1	5	25	1.0	1.71	0.01	
H_1	2	3	5	0.1	17.2	0.48	0.42
H_1	2	3	5	1.0	10.9	0.40	
H_1	2	3	10	0.1	52.9	0.68	0.65
H_1	2	3	10	1.0	46.6	0.65	
H_1	2	3	25	0.1	596.7	0.95	0.95
H_1	2	3	25	1.0	506.9	0.94	
H_1	2	5	5	0.1	14.1	0.44	0.44
H_1	2	5	5	1.0	8.55	0.28	
H_1	2	5	10	0.1	30.2	0.60	0.77
H_1	2	5	10	1.0	21.5	0.54	
H_1	2	5	25	0.1	183.4	0.92	0.99
H_1	2	5	25	1.0	135.5	0.90	
H_1	3	3	5	0.1	31.2	0.59	0.43
H_1	3	3	5	1.0	16.5	0.48	
H_1	3	3	10	0.1	129.0	0.75	0.74
H_1	3	3	10	1.0	72.8	0.70	
H_1	3	3	25	0.1	5774	0.98	0.98
H_1	3	3	25	1.0	6725	0.98	
H_1	3	5	5	0.1	19.8	0.53	0.46
H_1	3	5	5	1.0	12.8	0.40	
H_1	3	5	10	0.1	66.4	0.72	0.74
H_1	3	5	10	1.0	51.3	0.69	
H_1	3	5	25	0.1	2857	0.99	0.99
H_1	3	5	25	1.0	2092	0.98	

[†] proportion of simulations with $BF_{10} > 19$

[‡] proportion of simulations with LR test p-value < 0.05

Table 2

Descriptive statistics for the time to pregnancy example.

Number of women	428
Number of menstrual cycles	1968
Number of pregnancies	372
Average age	26
Number of smokers	38
Average cigs smoked by smokers	10/day
Average freq of intercourse	2-3/week
Number of women using pill in cycle 0	84

Table 3

Posterior summaries of the unconstrained and constrained regression coefficients for the time to pregnancy example.

Description	Parameter	Constrained Mean (sd)	Unconstrained Mean (sd)	Bayes Factor†
aged 25-29	β_{21}	0.10 (0.08)$^+$	0.10 (0.08)	–
aged > 30	β_{22}	0.21 (0.10)$^+$	0.21 (0.10)	–
1-3 times/wk	β_{31}	0.06 (0.08)	0.03 (0.12)	3.39
3-4 times/wk	β_{32}	0.15 (0.10)	0.15 (0.11)	10.18
> 4 times/wk	β_{33}	0.51 (0.15)	0.51 (0.15)	673.46‡
1-5 cigs/day	β_{41}	-0.06 (0.09)	-0.06 (0.18)	2.86
6-10 cigs/day	β_{42}	-0.14 (0.13)	-0.18 (0.19)	4.67
11-15 cigs/day	β_{43}	-0.28 (0.22)	-0.26 (0.34)	4.39
> 15 cigs/day	β_{44}	-0.46 (0.22)	-0.38 (0.25)	28.76‡
recent pill use	β_5	-0.55 (0.19)	-0.55 (0.19)	2315.95‡

$^+$ Coefficients for age are not restricted

† For assessing increases relative to baseline under monotonicity constraint

‡ Considered a significant increase under > 19 criterion

Figure 1. Estimated baseline conception rates in menstrual cycles 1-13 for the time to pregnancy example. The dotted lines represent 90% credible intervals.