INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

Bell & Howell Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

WASHINGTON UNIVERSITY

Department of Economics

Dissertation Committee:
Edward Greenberg, Chair
Robert S. Woodward
Fredric Q. Raines

A Box-Cox Random Coefficients Model: Bayesian Analysis and Applications

by

Christopher S. Hollenbeak

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2000

St. Louis, Missouri

Abstract

As a well-known generalization of the linear model, the Box-Cox (BC) model has been used extensively in applied econometrics. There are several reasons for its widespread use. First, the BC model nests as special cases the linear and semi-log models, as well as an infinite number of nonlinear models. The BC model permits a researcher to let the data determine the most appropriate functional form for a model rather than impose linearity or log-linearity arbitrarily. When theory implies some functional form for a regression equation the BC model may allow testing of the theory. More often, however, theory gives no guidelines for functional form, and the BC model provides some statistical justification for a dependent variable transformation that may otherwise be *ad hoc*. Second, maximum likelihood estimates for the model are easily obtained in cross-sectional or time-series settings using a grid search procedure.

There are currently very few BC models appropriate for analysis of clustered

data. This dissertation develops a BC model with random coefficients (BCRC) that is appropriate for clustered data and shows how Bayesian estimates may be obtained for the model parameters. I discuss how the posterior distribution may be simulated by Markov chain Monte Carlo (MCMC) methods and how marginal likelihoods and Bayes factors may be computed from the simulation output so that Bayesian model selection may be performed. The model is fit to a set of simulated data to test the performance of the sampler and the estimates.

I then use the BCRC model to risk-adjust the cost of providing inpatient hospital care to patients who undergo coronary artery bypass graft procedures. Clinical and cost data from four Midwestern hospitals are used to risk-adjust the cost of inpatient hospital care and compare the performance of the hospitals. Several important issues regarding risk adjustment are addressed, including 1) determining the most appropriate transformation of costs for the risk adjustment process, and 2) whether the hospital-level transformations are similar enough to be treated as equal. Results indicate that in these data the most appropriate transformation is inverse, and not natural log, and that the transformation is not equal across hospitals. Finally, the implication for ranking hospitals based on risk-adjusted costs is considered.

Acknowledgements

I have incurred many debts on the way to completing this thesis. To my parents, Keith and Christine Hollenbeak, who love learning and have always encouraged me to pursue knowledge and creativity. To Sue Baxter, Harvey Howeth, and Morgan Mannen, who took an interest in me and my intellectual development and have provided more inspiration than they know.

In my development as an economist I am particularly indebted to Ed Greenberg, who has been an exceptional advisor. I am very thankful to Ed for sparking my interest in Bayesian methods, helping me develop my ideas, and for patiently reading the many iterations of this thesis. To Bob Woodward, who introduced me to health economics and has given me the opportunity and support to do applied empirical research. Bob has been an exceptional mentor. To Fredric Raines, who has been a valuable mentor and deserves many thanks for his encouragement and support. To many others in the Economics department who have provided me

with guidance and support, particularly Bob Parks, Bruce Petersen, and Karen Rensing.

To the faculty and staff in the Health Administration Program who have given me a home for the last two years, and to Clay Dunagan, who allowed me to use data from the BJC Health System in this dissertation.

To Sid Chib, who taught me about hierarchical models, Markov chain Monte Carlo methods, and initially suggested the curious lack of Box-Cox models for clustered data. His pioneering work in Bayesian statistics has had a obvious and profound impact on me and on this work.

Finally, my deepest gratitude goes to my children, Madeline and Asher, who have been praying that I would finally finish this "emmertation," and to my wife, Kelly, who has supported my aspirations, understood my need to complete this work, and patiently endured my absence in body and mind. This work is dedicated to Kelly.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

As a well-known generalization of the linear model, the Box-Cox (BC) model has been used extensively in applied econometrics [5, 59, 62, 68, 74]. The BC model has been widely used for two reasons. First, it nests as special cases the linear and semi-log models—the most ubiquitous functional forms in econometrics—as well as an infinite number of nonlinear models, and it permits a researcher to let the data determine the most appropriate functional form for a model rather than impose linearity or log-linearity arbitrarily. When economic theory implies a specific functional form for an empirical model, the BC model may allow testing of the theory. For example, Mincer's [51] theory of wage determination implies that the natural log of wages is a linear function of years of schooling. Subsequently, Heckman and Polachek [33] and Bushinsky [7] used forms of the BC model to test

the validity of Mincer's theory.

More often, however, theory gives no guidelines for functional form, and the BC model is used to provide some statistical justification for a dependent variable transformation that is otherwise *ad hoc*.

The second reason the BC model has been used extensively is that maximum likelihood estimates for the model are easily obtained in cross-sectional or time-series settings. Conditional on a transformation, maximum likelihood estimates are obtained by ordinary least squares (OLS). Thus a simple grid search procedure yields maximum likelihood estimates.[1]

The BC model has been used in both cross-sectional and time-series settings, but research on BC models for clustered data is limited. One explanation for this is that the increase in the number of parameters that must be estimated in a clustered data model makes maximum likelihood estimates much more difficult to obtain. This problem is mitigated in Bayesian estimation, which in many cases, and in this case in particular, handles high dimensions more readily than classical or frequentist statistical techniques. This dissertation extends the BC

[1]Standard errors from the grid search are *conditional* standard errors and will overstate the reliability of the slope parameters. Unconditional standard errors can be obtained by the inverse of the information matrix. See Spitzer [68] and Judge et al. [40].

transformation to a clustered data setting to show how Bayesian estimates of the model parameters may be obtained and to use the model to evaluate current practices in evaluating hospital performance.

In the next chapter the BC transformation is added to a hierarchical random coefficients model, resulting in a model called the BCRC model. I discuss how the posterior distribution may be simulated by Markov chain Monte Carlo (MCMC) methods and how marginal likelihoods and Bayes factors may be calculated from the simulation output so that Bayesian model selection may be undertaken. Finally, I fit the BCRC model to a set of simulated data to assess the accuracy of the Bayesian estimates.

In third chapter I use the BCRC model to risk-adjust hospital costs for four regional hospitals. As the structure of the healthcare industry has changed over the last decade, there has been an increased interest in cost containment, in holding healthcare providers accountable for the quality of care they provide, and in comparing the performance of healthcare providers. Since healthcare providers tend to treat different populations, meaningful comparisons must use performance measures that take into account the underlying patient characteristics and comorbidities. This has given rise to the practice of risk-adjusting performance by

regressing the performance measure—usually mortality, length of stay, or cost of care—on an index of severity or on patient specific characteristics and comorbidities. Comparisons are then based on the residuals of the regression, which has netted out the effect of risk factors and comorbidities. In this chapter I use the BCRC model to risk-adjust the cost of inpatient hospital care for patients who undergo coronary artery bypass graft (CABG) procedures. In risk adjustment models of cost, the dependent variable is usually transformed by a natural log due to the skewness of cost data. The BCRC model is able to address important issues that have not been explored in the literature, including the appropriateness of the log transformation versus other possible transformations, and whether it is appropriate to apply the same transformation to all hospital clusters.

Results suggest that in these data, the natural log is not the most appropriate transformation, and the risk adjustment is better modeled by allowing different transformations for each hospital. The hospital ranking induced by the BCRC model, however, was not different from the ranking produced by maximum likelihood estimates of the linear and semi-log models. Chapter 4 concludes the dissertation with a discussion of possible directions for future research, including refinements of the model and and additional issues in risk adjustment and hospital ranking.

Chapter 2

The Box-Cox Random Coefficients Model

This chapter presents the BCRC model and develops the methodology that will be used throughout the dissertation. Specifically, I discuss posterior simulation by Markov chain Monte Carlo (MCMC) methods, show how marginal likelihoods can be calculated from the simulation output, and evaluate the performance of the estimates by reporting the results from fitting the BCRC model to simulated data. Before discussing the BC model in a clustered data context, however, it will be useful to review the usual cross-sectional version. This is done in Section 2.1, and the BCRC model is presented in Section 2.2. The posterior distribution of the BCRC model is described in Section 2.3, and Section 2.4 presents an MCMC algorithm to simulate the posterior distribution. In Section 2.5 I discuss how to

compute the marginal likelihoods and Bayes factors for the model. In Section 2.6 the BCRC model is applied to a set of simulated data and the performance of the algorithm and estimates are evaluated. Finally, Section 2.7 concludes the chapter with discussion.

2.1 The BC Model

The BC model in a simple cross-sectional or time-series context as proposed by Box and Cox [5] is

$$y_m^{(\lambda)} = \mathbf{x}_m'\boldsymbol{\beta} + \varepsilon_m \tag{2.1}$$

where m indexes observations, $\mathbf{x}_m(k \times 1) = (1, x_{m2}, \ldots, x_{mk})'$ is a vector of covariates, $\boldsymbol{\beta}(k \times 1) = (\beta_1, \ldots, \beta_k)'$ is a vector of coefficients, $\varepsilon_m \sim \mathcal{N}(0, \sigma^2)$, and the BC transformation is $y_m^{(\lambda)} = \frac{y_m^{\lambda} - 1}{\lambda}$. Two aspects of the model merit special attention. First, the errors are assumed to be normally distributed after an appropriate transformation of the dependent variable.[1] A dependent variable transformation may also stabilize the error variance or make the model more nearly linear. Second, the model nests two of the most common functional forms of regression equations. When $\lambda_m = 1$, then $y_m^{(\lambda)} = y_m - 1$, and the model is linear. When $\lambda_m = 0$ the BC

[1] Although this is usually assumed in the literature, several authors, including Poirier [55], Showalter [63], and Zarembka [74] demonstrate that the likelihood function is misspecified unless the transformation parameter equals zero. Draper and Cox [20] showed that as long as the data are reasonably symmetric, estimates of λ are robust to non-normality.

transformation is $y_m^{(\lambda)} = \ln y_m$ and the model is the semi-log model.[2]

The transformation, λ, is estimated along with the other parameters. This allows the data to determine the most appropriate functional form. Furthermore, estimates of the transformation and model parameters are easily obtained by maximum likelihood. If λ is known, the remaining parameters may be estimated by least squares. Since λ is not known, a grid search, typically over $[-2, 2]$, for the value of λ that minimizes the sum of squared residuals is used to obtain maximum likelihood estimates. While there are several ways to obtain maximum likelihood estimates, iterated OLS is one of the simplest and is widely implemented in statistical packages [40, 68].

Several variations of the BC model have been proposed. Transforming the dependent variables as well as the independent results in a model usually called the *extended* Box-Cox model:

$$y_m^{(\lambda)} = \mathbf{x}_m^{(\lambda)}\beta + \varepsilon_m. \tag{2.2}$$

[2]This is a limiting result:

$$\begin{aligned} \lim_{\lambda \to 0} \frac{x^\lambda - 1}{\lambda} &= \lim_{\lambda \to 0} \frac{d(x^\lambda - 1)/d\lambda}{1} \text{ by L'Hôpital's rule} \\ &= \lim_{\lambda \to 0} x^\lambda \times \ln x \\ &= \ln x \end{aligned}$$

Box and Tidwell [6] apply a different transformation to each of the k independent variables, but not the dependent variable:

$$y_m = \mathbf{x}_m^{(\lambda)'}\beta + \varepsilon_m \tag{2.3}$$

where $\mathbf{x}_m^{(\lambda)'} = (1, x_{m1}^{(\lambda_1)}, x_{m2}^{(\lambda_2)}, \ldots, x_{mk}^{(\lambda_k)})'$. Box and Tidwell's version tends to have problems with overidentification. Since each covariate has both a slope coefficient and a nonlinear transformation, each confounds the other. Box-Cox models have also been proposed in the presence of serial correlation by Savin and White [59], and in the presence of heteroskedasticity by Seaks and Layson [62].

2.2 The BCRC Model

Given the flexibility of the BC model, there are clear benefits from extending the transformation to clustered data settings. This section applies the BC transformation to a random coefficients model and allows a different transformation for each cluster. Applying the BC transformation to a clustered data model is straightforward, but the increase in the number of parameters makes estimation difficult by the usual maximum likelihood techniques.[3] Therefore, Bayesian techniques are

[3] A similar problem arises in estimating some forms of the BC model, for example the extended BC model and the Box-Tidwell variation [68].

used to simulate the posterior distribution of the model. The BCRC model is:

$$\mathbf{y}_i^{(\lambda_i)} = \mathbf{X}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \tag{2.4}$$

where the clusters are indexed by $i = 1, \ldots, n$, $\mathbf{y}_i^{(\lambda_i)}(n_i \times 1) = (y_{i1}^{(\lambda_i)}, \ldots, y_{in_i}^{(\lambda_i)})'$, $\boldsymbol{\varepsilon}_i(n_i \times 1) \sim \mathcal{N}_{n_i}(\mathbf{0}, \sigma^2\mathbf{I}_{n_i})$, $\mathbf{X}_i(n_i \times k) = (\mathbf{1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{ik})$ is a matrix of covariates with $\mathbf{x}_{ij} = (x_{ij_1}, \ldots, x_{ij_{n_i}})'$, and $\mathbf{1}(n_i \times 1)$ is a vector of ones. The collection of transformation parameters is $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$. To give the model random coefficients, assume that any heterogeneity in the constant and slope parameters across clusters is stochastic; specifically $\mathbf{b}_i(k \times 1) \sim \mathcal{N}(\boldsymbol{\beta}, \mathbf{D})$. $\boldsymbol{\beta}$ is the mean of the slope parameters, and diagonal elements of the covariance matrix $\mathbf{D}$ give some indication of the variation of the slope coefficients across clusters. The objects of inference are $\boldsymbol{\beta}$, $\mathbf{D}$, σ^2, and $\boldsymbol{\lambda}$.

There are other models for clustered data that might have been used in this study. The random coefficients model was chosen because it represents a compromise between extreme treatments of the clusters. Zellner's Seemingly Unrelated Regression (SUR) model assumes different slope parameters for each cluster, and pooled linear regression ignores the clustering and assumes the slope coefficients are identical across clusters. The random coefficients model is a compromise because it models the variation across clusters with random coefficients, which are summarized by a mean vector and covariance matrix.

Since the random coefficients model has a natural interpretation in a Bayesian context as a hierarchical model, Bayesian methods are used for inference. Another reason for taking a Bayesian approach in this problem is that maximum likelihood estimates are not easy to obtain for the BCRC model. Allowing the transformation to vary across clusters poses problems for the usual grid search to find maximum likelihood estimates since the dimensionality of the search space rises exponentially with the number of clusters.[4] Bayesian methods, particularly by MCMC methods, have practical advantages in many cases where high dimensionality is a problem and are well-suited to this model.

2.3 Hierarchical Model and Prior Assumptions

This section presents a hierarchical Bayes representation of the BCRC model. In the Bayesian paradigm data are viewed as fixed and parameters as random. Combining prior beliefs about the distributions of the parameters with the likelihood function of the data yields the posterior distribution. Inference is carried out on the posterior. In what they call hierarchical models, Lindley and Smith [45] showed that priors may be assumed not only on the parameters of the model but also on the hyperparameters of the prior distributions. The random coefficients

[4]Searching n clusters over $[-2, 2]$ with a step size of .1 requires 41^n regressions.

model is an example of a two stage hierarchical model, where the assumption that $\mathbf{b}_i \sim \mathcal{N}(\boldsymbol{\beta}, \mathbf{D})$ is viewed as a prior distribution for $\mathbf{b}_i$, and priors are placed on the hyperparameters $\boldsymbol{\beta}$ and $\mathbf{D}$.

To find the posterior distribution of the BCRC model we must first specify the likelihood function and define priors. Normally distributed errors imply the following likelihood function:

$$f(\mathbf{y} \mid \mathbf{b}, \boldsymbol{\beta}, \mathbf{D}, \sigma^2, \boldsymbol{\lambda}) \propto$$
$$\sigma^{(-\sum_i n_i)} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^{n_i} (\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)'(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)\} \prod_{j=1}^{n_i} \prod_{i=1}^{n} y_{ij}^{\lambda_i - 1} \quad (2.5)$$

where the double product term on the right hand side is the Jacobian of the Box-Cox transformation. The Jacobian is required because the normality assumption is made on the transformed data, and the likelihood function is written in terms of the untransformed data.

Hierarchical priors are assigned in two stages: first to the parameters of the model, and second to the hyperparameters of the prior for $\mathbf{b}_i$. Priors for the first stage of the hierarchy are

$$\pi(\mathbf{b}_i) = \mathcal{N}_k(\boldsymbol{\beta}, \mathbf{D}), \qquad \pi(\sigma^2) = \mathcal{IG}(\frac{\nu}{2}, \frac{\delta}{2}), \qquad \pi(\lambda_i) = \mathcal{N}(\lambda_0, \tau^2)$$

where $\mathcal{N}_k$ is the k-variate Normal distribution and $\mathcal{IG}$ represents the Inverse

Gamma distribution. Second stage priors are

$$\pi(\boldsymbol{\beta}) = \mathcal{N}_k(\boldsymbol{\beta}_0, \mathbf{B}_0), \qquad \pi(\mathbf{D}^{-1}) = \mathcal{W}(\eta, \mathbf{R})$$

where $\mathcal{W}$ represents the Wishart distribution. The priors are assumed independent, and the hyperparameters $(\boldsymbol{\beta}_0, \mathbf{B}_0, \nu, \delta, \eta, \mathbf{R}, \lambda_0, \tau^2)$ are assumed known. These priors are standard for hierarchical longitudinal models and are sufficiently flexible that either vague or informative prior information may be included by choosing appropriate values for the hyperparameters [9, 12, 25, 27].

The posterior distribution is obtained by multiplying the likelihood function and the priors and, except for the normalizing constant, may be written as

$$\begin{aligned} f(\mathbf{b}_i, \boldsymbol{\beta}, \mathbf{D}, \sigma^2, \boldsymbol{\lambda} \mid \mathbf{y}) \;\propto\; & \sigma^{(-\sum_i n_i)} \exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n_i}(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)'(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)\} \\ & \times \mid \mathbf{D}^{-\frac{1}{2}} \mid^{\frac{1}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{b}_i - \boldsymbol{\beta})'\mathbf{D}^{-1}(\mathbf{b}_i - \boldsymbol{\beta})\} \\ & \times \mid \mathbf{B}_0^{-1} \mid^{\frac{1}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\mathbf{B}_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\} \\ & \times \sum_{i=1}^{n} \mid \mathbf{D}^{-1} \mid^{\frac{(\eta - n_i - 1)}{2}} \exp\{-\frac{1}{2}\mathrm{tr}(\mathbf{R}^{-1}\mathbf{D}^{-1})\} \\ & \times \sigma^{-(\nu+2)} \exp\{-2(\delta\sigma^2)^{-1}\} \\ & \times \tau^{-n} \exp\{-\frac{1}{2\tau^2}\sum_{i=1}^{n}(\lambda_i - \lambda_0)^2\} \\ & \times \prod_{j=1}^{n_i}\prod_{i=1}^{n} y_{ij}^{\lambda_i - 1}. \end{aligned} \tag{2.6}$$

An interesting special case of this model assumes that $\lambda_i = \lambda$ for all i. This

model will be particularly important when we consider what the data suggest about the appropriateness of alternative model specifications and model comparison.

A traditional Bayesian analysis would proceed by marginalizing the posterior in equation (2.6) over the parameters of interest. The posterior distribution for the BCRC model is from an unknown family of distributions, is high dimensional, and has an unknown normalizing constant; it is not, therefore, amenable to integration by analytical methods. This difficulty is characteristic of most interesting models and has historically been an impediment to applied Bayesian analysis because the integration required to marginalize and summarize the posterior distribution is either impractical or impossible. High speed computing and recent developments in MCMC simulation have helped to overcome this problem. Rather than integrating the posterior analytically to get marginal distributions of the parameters, MCMC allows a (possibly correlated) sample to be drawn directly from the posterior distribution, even if the normalizing constant is unknown. Moments of the distributions can be approximated by the moments of the simulated sample, and the approximation may be made arbitrarily precise by increasing the number of simulated draws from the posterior distribution. The next section describes the MCMC algorithm to draw simulated samples from the posterior distribution and

carry out Bayesian inference.

2.4 A Markov Chain Monte Carlo Algorithm

The intuition underlying MCMC methods is that a Markov chain is constructed whose transition distribution converges to the posterior distribution. Then, starting from an arbitrary point, the chain is allowed to run until it converges to its stationary distribution. The initial draws, generated while the chain is in its transient stage, are discarded, and the remaining draws are a sample from the posterior distribution that may be used for purposes of inference. The Metropolis-Hastings (MH) algorithm and its special case, the Gibbs sampler, are two examples of MCMC algorithms that are becoming more frequently used in Bayesian applications because implementation is straightforward and they are able to estimate models of nearly unlimited complexity. I first discuss the Gibbs sampler and the MH algorithm in general terms and then show how they can be used to simulate the BCRC model in a hybrid approach.

2.4.1 The Gibbs sampler

Assume we are interested in estimating a vector of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$. Let $f(\boldsymbol{\theta} \mid \mathbf{y})$ be the normalized joint posterior density of $\boldsymbol{\theta}$ and define the full conditional distribution as the distribution of θ_i conditional on the remaining θ_js $(j \neq i)$ as $f_i(\theta_i \mid \boldsymbol{\theta}_{-i})$ where $\boldsymbol{\theta}_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_k)$, for $i = 1, \ldots, k$. Note that the conditional distributions are proportional to the posterior.

The Gibbs sampler produces draws from the joint posterior distribution by taking successive draws from the set of full conditional distributions. Because draws are taken directly from the full conditional distributions, they must be from standard families of distributions. Specifically, the normalizing constants must be known and algorithms must be available to generate samples from the full conditional distributions. Beginning at an arbitrary starting point $\boldsymbol{\theta}^0 = (\theta_1^0, \theta_2^0, \ldots, \theta_k^0)$, the Gibbs sampling algorithm proceeds as follows:

$$
\begin{array}{lll}
\theta_1^1 & \text{is drawn from} & f_1(\theta_1 \mid \boldsymbol{\theta}_{-1}^0, \mathbf{y}) \\
\theta_2^1 & \text{is drawn from} & f_2(\theta_2 \mid \theta_1^1, \theta_3^0, \ldots, \theta_k^0, \mathbf{y}) \\
\theta_3^1 & \text{is drawn from} & f_3(\theta_3 \mid \theta_1^1, \theta_2^1, \theta_4^0, \ldots, \theta_k^0, \mathbf{y}) \\
\vdots & \vdots & \vdots \\
\theta_k^1 & \text{is drawn from} & f_k(\theta_k \mid \boldsymbol{\theta}_{-k}^1, \mathbf{y}).
\end{array}
$$

This produces a new vector of parameters θ^1, which is used in place of θ^0 to repeat the sampler and draw θ^2. The vector θ^m on which the procedure conditions is fixed, and the draw of θ^{m+1} is random, which implies that there is a distribution $f^*(\theta^{m+1} \mid \theta^m)$ associated with obtaining θ^{m+1} from θ^m. Since the draw of each θ^{m+1} depends only on θ^m and not on previous draws, the sequence of draws forms a Markov chain with transition density f^*. If f^* is aperiodic and irreducible, this transition density converges to the posterior distribution as the number of draws increases [28, 71, 54].

2.4.2 The Metropolis-Hastings algorithm

If the conditional distributions are not from a standard family of distributions, the parameters may be simulated by the Metropolis-Hastings (MH) algorithm, first proposed by Metropolis et al. [48], and later generalized by Hastings [32]. Since the distribution is not known, a draw is taken from a known density, $q(\theta^{m+1} \mid \theta^m)$, called the proposal density or candidate generating density. Next, define the candidate acceptance probability as

$$\alpha(\theta^{m+1}, \theta^m) = \min\left\{\frac{f(\theta^{m+1})q(\theta^m \mid \theta^{m+1})}{f(\theta^m)q(\theta^{m+1} \mid \theta^m)}, 1\right\}, \tag{2.7}$$

where f is the joint posterior distribution. In the Metropolis–Hastings algorithm, a draw is made from the proposal density. The draw is accepted and retained

with probability $\alpha(\boldsymbol{\theta}^{m+1}, \boldsymbol{\theta}^m)$. If the draw is not retained, then $\boldsymbol{\theta}^{m+1} = \boldsymbol{\theta}^m$ and the current draw equals the previous draw. If the proposal density is symmetric then the probability that the draw will be accepted simplifies to

$$\alpha(\boldsymbol{\theta}^{m+1}, \boldsymbol{\theta}^m) = \min\left\{\frac{f(\boldsymbol{\theta}^{m+1})}{f(\boldsymbol{\theta}^m)}, 1\right\}. \tag{2.8}$$

Note that if the candidate draw comes from a higher point on the joint posterior distribution than the previous draw it is accepted with certainty. But if the candidate draw is from a point lower on the joint posterior it is only accepted with probability $\frac{f(\boldsymbol{\theta}^{m+1})}{f(\boldsymbol{\theta}^m)}$. Also note that the normalizing constant for f cancels in expression (2.8), and therefore it is possible to simulate the posterior even if the normalizing constant is unknown.

2.4.3 A hybrid approach

The BCRC model may be simulated with a hybrid approach that combines the Gibbs sampler and the Metropolis-Hastings algorithm. This results from the fact that conditional on the transformation parameter, the remaining model parameters are from standard distributions with known normalizing constants. These conditional distributions may be sampled directly using the Gibbs sampler. The transformation parameters, however, are not from standard distributions and require sampling by a Metropolis-Hastings step. To find the set of full conditional

distributions, define

$$\mathbf{V}_{b_i} = (\mathbf{D}^{-1} + \sigma^{-2}\mathbf{X}_i'\mathbf{X}_i)^{-1} \qquad \mathbf{V}_\beta = (\mathbf{B}_0 + n\mathbf{D}^{-1})^{-1}$$

$$\hat{\mathbf{b}}_i = \mathbf{V}_{b_i}(\mathbf{D}^{-1}\beta + \sigma^{-2}\mathbf{X}_i'\mathbf{y}_i^{(\lambda_i)}) \qquad \hat{\beta} = \mathbf{V}_\beta(\mathbf{B}_0\beta_0 + \mathbf{D}^{-1}\textstyle\sum_{i=1}^n \mathbf{b}_i).$$

The full conditional distributions for the parameters except for $\boldsymbol{\lambda}$ can be shown to be:

$$[\mathbf{b}_i \mid \beta, \mathbf{D}, \sigma^2, \boldsymbol{\lambda}] = \mathcal{N}(\hat{\mathbf{b}}_i, \mathbf{V}_{b_i}) \tag{2.9}$$

$$[\beta \mid \mathbf{b}_i, \mathbf{D}, \sigma^2, \boldsymbol{\lambda}] = \mathcal{N}(\hat{\beta}, \mathbf{V}_\beta) \tag{2.10}$$

$$[\mathbf{D}^{-1} \mid \mathbf{b}, \beta, \sigma^2, \lambda] = \mathcal{W}(\eta + n, (\mathbf{R}^{-1} + \sum_{i=1}^{n}[(\mathbf{b}_i - \beta)(\mathbf{b}_i - \beta)']^{-1}) \tag{2.11}$$

$$[\sigma^2 \mid \mathbf{b}, \beta, \mathbf{D}, \boldsymbol{\lambda}] =$$
$$\mathcal{IG}(\frac{\nu + \sum_{i=1}^n n_i}{2}, \frac{\delta + \sum_{i=1}^n (\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)'(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)}{2}) \,. \tag{2.12}$$

To find the conditional distribution of $\boldsymbol{\lambda}$, recall that it is proportional to the posterior distribution and suppress all elements of equation (2.6) that do not involve $\boldsymbol{\lambda}$ into the proportionality constant:

$$\begin{aligned}[\boldsymbol{\lambda} \mid \mathbf{b}, \beta, \mathbf{D}, \sigma^2] \;\propto\; & \exp\{-\frac{1}{2}\sum_{i=1}^{n}\frac{1}{\sigma^2}(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)'(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)\} \\ & \times \exp\{-\frac{1}{2\tau^2}\sum_{i=1}^{n}(\lambda_i - \lambda_0)^2\}\prod_{j=1}^{n_i}\prod_{i=1}^{n} y_{ij}^{\lambda_i - 1}.\end{aligned} \tag{2.13}$$

Since this is not a standard distribution in $\boldsymbol{\lambda}$, simulation will require the MH algorithm. The task of simulating the transformation parameter would be simplified if candidate draws could be made for each λ_i individually instead of the vector $\boldsymbol{\lambda}$. To see that this is possible, suppress all terms in equation (2.14) that do not involve λ_i into the constant of proportionality. This yields the kernel of λ_i conditional on $\boldsymbol{\lambda}_{-i}$

$$[\lambda_i \mid \mathbf{b}, \boldsymbol{\beta}, \mathbf{D}, \sigma^2, \boldsymbol{\lambda}_{-i}] \propto \exp\{-\frac{1}{2}[\frac{1}{\sigma^2}(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i)'(\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i\mathbf{b}_i) + \frac{1}{\tau^2}(\lambda_i - \lambda_0)^2]\} \prod_{j=1}^{n_i} y_{ij}^{\lambda_i - 1}. \quad (2.14)$$

Since this is a function only of λ_i each λ_i may be sampled individually.

There are several possibilities for candidate generating densities. One natural choice is a random walk candidate generating density, $\lambda^{m+1} = \lambda^m + u$, where $u \sim \mathcal{N}(0, \xi^2)$. The advantage of this density is that it is symmetric and easy to implement. Chib, Greenberg and Winkleman [16] argue, however, that very often the random walk does not lead to convergence. An alternative is to use a candidate generating density that is tailored to the conditional density, for example a normal or noncentral t distribution centered at the mode of the full conditional. Since the mode of the full conditional is not known, an optimization algorithm may be necessary to find the mode. The Newton-Raphson algorithm is

especially appropriate in this problem since it requires only the first and second derivatives of the log of the density function. The normalizing constant does not need to be known because it is additive in the log density, and therefore disappears following differentiation. To optimize the full conditional distribution of λ_i with the Newton-Raphson algorithm, a starting point, $\lambda_i^{(0)}$, is first chosen. Iterations then evolve according to the following function:

$$\lambda_i^{(t+1)} = \lambda_i^{(t)} - \frac{f'(\lambda_i \mid \mathbf{b}, \boldsymbol{\beta}, \mathbf{D}, \sigma^2, \boldsymbol{\lambda}_{-i})}{f''(\lambda_i \mid \mathbf{b}, \boldsymbol{\beta}, \mathbf{D}, \sigma^2, \boldsymbol{\lambda}_{-i})} \tag{2.15}$$

where f is the natural log of the full conditional distribution, and the first and second derivatives are

$$\begin{aligned} \frac{\partial f}{\partial \lambda_i} &= -\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\{(\frac{y_{ij}^{\lambda_i}-1}{\lambda_i} - \sum_{k=1}^{K}[x_{ijk}b_{ik}])(\frac{y_{ij}^{\lambda_i}\log y_{ij}}{\lambda_i} - \frac{y_{ij}^{\lambda_i}-1}{\lambda_i^2})\} \\ &\quad -\frac{1}{\tau^2}(\lambda_i - \lambda_0) + \sum_{j=1}^{n_i}\log y_{ij} \end{aligned} \tag{2.16}$$

and

$$\begin{aligned} \frac{\partial^2 f}{\partial \lambda_i^2} &= -\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\{(\frac{y_{ij}^{\lambda_i}\log y_{ij}}{\lambda_i} - \frac{y_{ij}-1}{\lambda_i})^2 + (\frac{y_{ij}-1}{\lambda_i} - \sum_{k=1}^{K}[x_{ijk}b_{ik}]) \\ &\quad \times(\frac{y_{ij}^{\lambda_i}\log y_{ij}}{\lambda_i} - \frac{2y_{ij}^{\lambda_i}\log y_{ij}}{\lambda_i^2} + \frac{2(y_{ij}^{\lambda_i}-1)}{\lambda_i^3})\} - \frac{1}{\tau^2}. \end{aligned} \tag{2.17}$$

The derivatives are undefined at zero, but L'Hôpital's Rule shows that their limits as λ_i approaches zero are:

$$\lim_{\lambda_i \to 0}\frac{\partial f}{\partial \lambda_i} = -\frac{1}{\sigma^2}\sum_{j=1}^{n_i}\{\frac{(\log y_{ij})^3}{2} - \frac{(\log y_{ij})^2\sum_{k=1}^{K}[x_{ijk}b_{ik}]}{2}\} - \frac{\lambda_0}{\tau^2} + \sum_{j=1}^{n_i}\log y_{ij} \tag{2.18}$$

and

$$\lim_{\lambda_i \to 0} \frac{\partial^2 f}{\partial \lambda_i^2} = -\frac{1}{\sigma^2} \sum_{j=1}^{n_i} \{ \frac{(\log y_{ij})^4}{4} + (\log y_{ij} - \sum_{k=1}^{K} [x_{ijk} b_{ik}]) (\frac{(\log y_{ij})^3}{3}) \} - \frac{1}{\tau^2}. \quad (2.19)$$

Once the mode of the full conditional of λ_i is found, a draw is taken from a normal or t distribution centered at the mode, and with variance equal to the curvature of the conditional density at the mode, where curvature is defined as $-\left[\frac{\partial^2 f}{\partial \lambda_i^2}\right]^{-1}$.

The posterior of the BCRC model is simulated by taking successive draws from the full conditional distributions just defined conditional on the current draw. Beginning at an arbitrary starting point, $(\mathbf{b}_i^{(0)}, \boldsymbol{\beta}^{(0)}, \mathbf{D}^{-1(0)}, \sigma^{2(0)}, \boldsymbol{\lambda}^{(0)})$, iterate through the following sequence a large number of times:

1. Draw $\mathbf{b}_i$ from $\mathcal{N}(\hat{\mathbf{b}}_i, \mathbf{V}_{b_i})$ for $i = 1 \ldots n$.

2. Draw $\boldsymbol{\beta}$ from $\mathcal{N}(\hat{\boldsymbol{\beta}}, \mathbf{V}_{\beta})$.

3. Draw $\mathbf{D}^{-1}$ from $\mathcal{W}(\nu + n, \mathbf{R}^{-1} + \sum_{i=1}^{n} [(\mathbf{b}_i - \boldsymbol{\beta})(\mathbf{b}_i - \boldsymbol{\beta})']^{-1})$

4. Draw σ^2 from $\mathcal{IG}(\frac{\nu + \sum_{i=1}^{n} n_i}{2}, \frac{\delta + \sum_{i=1}^{n} (\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i \mathbf{b}_i)' (\mathbf{y}_i^{(\lambda_i)} - \mathbf{X}_i \mathbf{b}_i)}{2})$.

5. Draw λ_i from either the random walk chain $\lambda_i^{j+1} = \lambda_i^j + u$ where $u \sim \mathcal{N}(0, \xi^2)$ or use the Newton-Raphson algorithm to find the mode of equation (2.15)

and draw λ_i from a Normal or t distribution centered at the mode. Then accept the candidate draw with probability $\alpha(\lambda_i^{(j+1)}, \lambda_i^{(j)}) = \min\{\frac{f(\lambda_i^{(j+1)})}{f(\lambda_i^{(j)})}, 1\}$ for $i = 1, \ldots, n$, where $f(\lambda_i)$ is equation (2.15).

The first draws, obtained while the Markov chain is in its transition state, are discarded, and the remaining draws are a sample from the posterior distribution.

Chib and Carlin [12] suggest an improved algorithm for hierarchical longitudinal models that could be applied in this model. In the context of a Gaussian linear mixed model the conditional distribution of β can be written in terms of σ^2 and $\mathbf{D}$ only. Therefore, Chib and Carlin suggest sampling β and $\mathbf{b}$ in one block by first sampling from $[\beta|\mathbf{y}, \sigma^2, \mathbf{D}]$ and then from $[\mathbf{b}|\mathbf{y}, \beta, \sigma^2, \mathbf{D}]$. In the BCRC model it would be possible to sample β, $\mathbf{b}$, and $\boldsymbol{\lambda}$ from $[\beta, \mathbf{b}, \boldsymbol{\lambda}|\mathbf{y}, \sigma^2, \mathbf{D}]$ using the same blocking. This modification speeds convergence of the sampler and improves mixing in the sampling space. The algorithms used in this dissertation do not take advantage of this blocking strategy. As I will show, the algorithms I used converged rapidly and did not have problems with mixing. Future applications, however, should take advantage of Chib and Carlin's improved algorithm.

2.5 The Marginal Likelihood and Model Selection

Since Bayesian analysis proceeds on the assumption that parameters are random and that data are fixed, hypothesis testing in the classical sense is not possible. Instead of testing hypotheses, a Bayesian analysis compares models. The goal of model comparison, or model selection, is to determine the posterior odds that a particular model generated the data, given the observed data. This is important for this study since the primary use of the BCRC model will be to discriminate between alternative risk adjustment model specifications (e.g. linear and semi-log). Model selection provides the basis for preferring one model over another. This section discusses model selection in general, especially how marginal likelihoods and Bayes factors can be computed from the output of an MCMC sampler, and then shows how marginal likelihoods and Bayes factors can be computed for the BCRC model.

Assume we want to compare a finite collection of models $M = \{M_1, \ldots, M_I\}$, where a model consists of a likelihood function and a prior: $M_i = \{f(y \mid M_i, \theta_i), \pi(\theta_i \mid M_i)\}$. Define the prior probability of the truth of model M_i as $\Pr(M_i) = p_i$.

Then the goal is to compute and compare $\Pr(M_i \mid y)$ for all i. By Bayes theorem,

$$\Pr(M_i \mid y) \propto p_i m(y \mid M_i), \tag{2.20}$$

where $m(y \mid M_i)$ is the marginal likelihood of y, given by

$$m(y \mid M_i) = \int f(y \mid M_i, \theta_i)\pi(\theta_i \mid M_i)d\theta_i. \tag{2.21}$$

The marginal likelihood is the likelihood function averaged over the prior. It is also the normalizing constant of the posterior distribution.

The posterior odds ratio of model i relative to model j is the ratio of prior probabilities of the truth of the two models times the ratio of marginal likelihoods, also called the Bayes factor:

$$\frac{\Pr(M_i \mid y)}{\Pr(M_j \mid y)} = \frac{p_i}{p_j} \times \frac{m(y \mid M_i)}{m(y \mid M_j)}. \tag{2.22}$$

If all models are assumed equally likely ($p_i = p$ for all i) then the posterior odds equals the Bayes factor.

Model selection provides three important advantages over hypothesis testing. First, it is not confined to pairwise comparisons since the odds ratios of any number of models may be calculated. Second, the weight of evidence is symmetric in model selection. Whereas, in classical hypothesis testing, rejecting the null hypothesis does not necessarily provide support for the alternative, Bayes factors provide

evidence both for and against alternative models.[5] The third, and most important, advantage of model selection is that it is not restricted to nested models. Any possible model that can be proposed to have generated the data can be compared [56].

Strategies for computing marginal likelihoods from the output of MCMC simulation have developed on two fronts. Some authors have constructed Markov chains that jump between different model specifications during sampling iterations. In the 'reversible jump' method proposed by Green [31], the first step of the MCMC sampler chooses a model and accepts or rejects the move of the Markov chain to that model in a Metropolis-Hastings step. Carlin and Chib [8] take a similar approach and introduce a model indicator into the Gibbs sampling scheme. This makes the priors a function of the model indicator and requires the use of pseudopriors, or linking densities, $\pi(\theta_i \mid m \neq i)$, that specify the prior of the parameters for model i given that the model at that iteration is not i. The drawback to this approach is the difficulty of interpreting and specifying the linking densities.

Other authors have attempted to use samples generated by the MCMC algo-

[5] $\frac{m(y|M_i)}{m(y|M_j)}$ is evidence against model j in favor of model i, and $\frac{m(y|M_j)}{m(y|M_i)}$ is evidence in favor of model j and against model i.

rithm to compute marginal likelihoods directly. Newton and Raftery [53] showed that the marginal likelihood can be computed as the harmonic mean of likelihood values

$$\hat{m}_{NR}(\mathbf{y} \mid M_i) = \{\frac{1}{G}\sum_{g=1}^{G}(\frac{1}{f(\mathbf{y} \mid \theta_i^{(g)}, M_i)})\}^{-1} \tag{2.23}$$

where g is a draw from the MCMC sampler. While this estimator is theoretically sound, it has been criticized because in practice it has problems with stability [11]. The source of the instability is that draws are made from the *inverse* likelihood. Gelfand and Dey [24] modify the method of Newton and Raftery by introducing a tuning function with thin tails into the estimation. Their estimate is

$$\hat{m}_{GD}(\mathbf{y} \mid M_i) = \{\frac{1}{G}\sum_{g=1}^{G}(\frac{p(\theta_i^{(g)})}{f(\mathbf{y} \mid \theta_i^{(g)}, M_i)\pi(\theta_i \mid M_i)})\}^{-1}. \tag{2.24}$$

This helps overcome the instability problem, but choosing a tuning function with sufficiently thin tails has proven difficult, especially in high dimensional problems.

The most promising attempt to compute marginal likelihoods directly from MCMC output is due to Chib [11], whose approach is based on the basic marginal likelihood identity (BMI). By simply rearranging the posterior distribution, the marginal likelihood may be written as

$$m(y \mid M_i) = \frac{f(y \mid M_i, \theta_i)\pi(\theta_i \mid M_i)}{\pi(\theta_i \mid M_i, y)}. \tag{2.25}$$

Since m is not a function of θ_i, equation (2.25) is an identity for any value θ_i^*. If

an estimate of the posterior density at the point θ_i^* ($\hat{\pi}(\theta_i^* \mid M_i, y)$,) is available, then Chib's estimate of the marginal likelihood can be written in log scale as

$$\ln\left(\widehat{(y \mid M_i)}\right) = \ln f(y \mid M_i, \theta_i^*) + \ln \pi(\theta_i^* \mid M_i) - \ln \hat{\pi}(\theta_i^* \mid M_i, y). \qquad (2.26)$$

Computing the marginal likelihood requires only an estimate of the ordinate of the likelihood function, the priors and the posterior, all evaluated at θ_i^*. The functional form of the likelihood and priors are typically known, and the challenge is obtaining an estimate of the posterior ordinate. Chib [11] shows how the draws from the MCMC sampler may be used to estimate the posterior at θ^*. Partition θ into J blocks. By the law of conditional probability, the posterior can be written as

$$\pi(\theta^* \mid y) = \pi(\theta_1^* \mid y) \times \pi(\theta_2^* \mid y, \theta_1^*) \times \cdots \times \pi(\theta_J^* \mid y, \theta_1^*, \theta_2^*, \ldots, \theta_{J-1}^*) \qquad (2.27)$$

where the subscripts indicating model i have been suppressed. The first term on the right hand side can be written

$$\pi(\theta_1^* \mid y) = \int \cdots \int \pi(\theta_1^* \mid y, \theta_2, \ldots, \theta_J)\pi(\theta_2 \mid y) \cdots \pi(\theta_I \mid y) d\theta_2 \ldots d\theta_I \qquad (2.28)$$

and can be approximated by averaging the ordinates of the conditional density of θ_1^* over the draws of the MCMC sampler with

$$\hat{\pi}(\theta_1^* \mid \mathbf{y}) = \frac{1}{G} \sum_{g=1}^{G} \pi(\theta_1^* \mid y, \theta_2^{(g)}, \ldots, \theta_J^{(g)}). \qquad (2.29)$$

The remaining terms on the right hand side of equation (2.27) can be estimated in a similar fashion, but the initial draws from the MCMC sampler may not be used. Consider the estimate of a typical term, $\pi(\theta_k^* \mid y, \theta_1^*, \ldots, \theta_{k-1}^*)$. Because the estimate is conditioned on $(\theta_1^*, \ldots, \theta_{k-1}^*)$, the draws over which the function is averaged must reflect this conditioning. An appropriate sample may be obtained by running the MCMC sampler for an additional G iterations with $(\theta_1 \ldots \theta_{k-1})$ restricted to $(\theta_1^* \ldots \theta_{k-1}^*)$. Chib calls this a reduced run of the sampler. The remaining terms on the right hand side of equation (2.27) can be estimated by

$$\hat{\pi}(\theta_k^* \mid \mathbf{y}) = \frac{1}{G} \sum_{g=1}^{G} \pi(\theta_k^* \mid y, \theta_1^*, \ldots, \theta_{k-1}^*, \theta_{k+1}^{(g)}, \ldots, \theta_J^{(g)}). \tag{2.30}$$

where g represents draws from the reduced run of the Gibbs sampler. After the posterior ordinate is evaluated at θ^*, all of the ordinates are substituted into equation (2.26) to obtain the marginal likelihood.

Chib's BMI approach does not suffer from instability problems, because draws are taken from the likelihood and priors directly, and not from the inverse likelihood. In cases where the normalizing constants of the full conditional distributions are not known, Chib and Greenberg [15] show that posterior ordinates may be obtained by kernel smoothing over the distributions with unknown normalizing constants. Alternatively, Chib and Jeliazkov [17] have proposed a method for computing marginal likelihoods directly from Metropolis-Hastings output. This

method is more general and more stable with higher dimensional models than kernel smoothing.

2.5.1 Marginal likelihood of the BCRC model

Marginal likelihoods for the BCRC model can be computed using the BMI approach. From the previous section, we may estimate the marginal likelihood by

$$\begin{aligned} \ln \widehat{m(\mathbf{y})} &= \ln f(\mathbf{y} \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^*) + \ln \pi(\boldsymbol{\beta}^*) \\ &\quad + \ln \pi(\mathbf{D}^*) + \ln \pi(\sigma^{2*}) + \sum_{i=1}^{n} \ln \pi(\lambda_i) \\ &\quad - \ln f(\boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^* \mid \mathbf{y}) \end{aligned} \tag{2.31}$$

where $(\boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^*)$ are chosen from high density regions of the posterior distribution, such as the mean or the mode.[6]

Chib [11] argues that this estimate of the marginal likelihood is unstable for models in which latent variables appear in both the numerator and the denominator. Therefore, $\mathbf{b}$ does not appear in equation (2.31) since the they are latent variables. We can find the likelihood function of $\mathbf{y}$ conditional on $\boldsymbol{\beta}$ (and not $\mathbf{b}$)

[6]Because the BMI is an identity the values may be chosen anywhere in the support of the posterior. Chib [11] argues, however, that in the simulation context it is important that a sufficient number of draws are available near θ^*, therefore, values should be chosen from a high density region of the support such as the mean or the mode.

by using the fact that the random coefficient can be written as

$$\mathbf{b}_i = \boldsymbol{\beta} + \mathbf{u}_i, \tag{2.32}$$

where $\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$. This implies that the model can be written as

$$\begin{aligned} \mathbf{y}_i^{(\lambda_i)} &= \mathbf{X}_i(\boldsymbol{\beta} + \mathbf{u}_i) + \boldsymbol{\varepsilon}_i \\ &= \mathbf{X}_i\boldsymbol{\beta} + \mathbf{w}_i \end{aligned} \tag{2.33}$$

where $\mathbf{w}_i \sim \mathcal{N}\left(\mathbf{0}, \sigma^2\mathbf{I} + \mathbf{X}_i\mathbf{D}\mathbf{X}_i'\right)$. Therefore, the first term in equation (2.31) is the likelihood function for a normal distribution with mean vector 0 and covariance matrix $\sigma^2\mathbf{I} + \mathbf{X}_i\mathbf{D}\mathbf{X}_i'$.

The functional form of the first five terms on the right hand side of equation (2.31) are known along with their normalizing constants, and ordinates are obtained by evaluating the likelihood function or prior density function at $(\boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^*)$. This is not the case for sixth term, the posterior density ordinate. Rewrite the posterior as

$$\begin{aligned} f(\boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^* \mid \mathbf{y}) &= f(\boldsymbol{\beta}^* \mid \mathbf{y}) \times f(\mathbf{D}^* \mid \boldsymbol{\beta}^*, \mathbf{y}) \\ &\quad \times f(\sigma^{2*} \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \mathbf{y}) \\ &\quad \times f(\boldsymbol{\lambda}^* \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \mathbf{y}). \end{aligned} \tag{2.34}$$

This can be estimated by estimating separately each of the terms on the right

hand side of equation (2.34). The first term can be approximated by

$$\hat{f}(\boldsymbol{\beta}^* \mid \mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G}(\boldsymbol{\beta}^* \mid \mathbf{D}^{(g)}, \sigma^{2(g)}, \boldsymbol{\lambda}^{(g)}, \mathbf{b}^{(g)}, \mathbf{y}) \tag{2.35}$$

where f is the full conditional distribution of $\boldsymbol{\beta}^*$ and g is a draw from the initial run of the MCMC sampler.

Similarly, the second term in equation (2.34) can be approximated by

$$\hat{f}(\mathbf{D}^* \mid \boldsymbol{\beta}^*, \mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G} f(\mathbf{D}^* \mid \boldsymbol{\beta}^*, \sigma^{2(g)}, \boldsymbol{\lambda}^{(g)}, \mathbf{b}^{(g)}, \mathbf{y}) \tag{2.36}$$

where f is the full conditional distribution of $\mathbf{D}$, and g now represents draws from a reduced run of the MCMC sampler where $\boldsymbol{\beta}$ is set to $\boldsymbol{\beta}^*$.

The third term in (2.34) can be approximated by

$$\hat{f}(\sigma^{2*} \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G} f(\sigma^{2*} \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \boldsymbol{\lambda}^{(g)}, \mathbf{b}^{(g)}, \mathbf{y}) \tag{2.37}$$

where f is the full conditional distribution of σ^2 and g represents draws from a reduced run of the MCMC sampler with $(\boldsymbol{\beta}, \mathbf{D})$ set to $(\boldsymbol{\beta}^*, \mathbf{D}^*)$.

Finally, consider how to estimate the last term in equation (2.34). Since the normalizing constant of the full conditional distribution of $\boldsymbol{\lambda}$ is unknown, kernel smoothing can be used to estimate the ordinate. Because the λ_is are conditionally independent, we can write

$$f(\boldsymbol{\lambda}^* \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \mathbf{y}) = f(\lambda_1^* \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \mathbf{y}) \times \ldots \times f(\lambda_n^* \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \mathbf{y}). \tag{2.38}$$

The ordinate of the full conditional distribution of each λ_i can be approximated by

$$\hat{f}(\lambda_i \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \mathbf{y}) = \frac{1}{G}\sum_{g=1}^{G}\frac{1}{h_j}\mathcal{N}\left(\frac{\lambda_i^* - \lambda_i^{(g)}}{h_j}\right) \tag{2.39}$$

where $\mathcal{N}$ is the Gaussian kernel, h_j is a bandwidth parameter, and g is a draw from a reduced run of the MCMC sampler with parameters set to $(\boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^*_{-i})$ [4, 64, 72].

Combining equations (2.28) through (2.33), the posterior density estimate is

$$\begin{aligned}\hat{f}(\boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^* \mid \mathbf{y}) &= \hat{f}(\boldsymbol{\beta}^* \mid \mathbf{y}) \times \hat{f}(\mathbf{D}^* \mid \boldsymbol{\beta}^*, \mathbf{y}) \\ &\quad \times \hat{f}(\sigma^{2*} \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \mathbf{y}) \\ &\quad \times \prod_{i=1}^{n} \hat{f}(\lambda_i^* \mid \boldsymbol{\beta}^*, \mathbf{D}^*, \sigma^{2*}, \boldsymbol{\lambda}^*_{-i}, \mathbf{y}).\end{aligned} \tag{2.40}$$

Substituting this estimate and ordinates of the likelihood function and priors into equation (2.31) yields the log of the marginal likelihood. Finally, the Bayes factor for model i compared to model j is $\exp\{\ln m(\mathbf{y}|M_i) - \ln m(\mathbf{y}|M_j)\}$.

2.6 Performance of the Sampler

In the previous section I showed how Chib's BMI approach [11] may be used to estimate the marginal likelihood for the BCRC model using output from the

MCMC sampler. The algorithms to simulate the BCRC model and to estimate the marginal likelihood were programmed in the GAUSS programming language. The program is contained in Appendix A. In this section I show how well the MCMC approach estimates the parameters of the BCRC model by applying the methods to a set of simulated data. Fitting the model to simulated data assures that the algorithms are coded properly, that the means of the marginal posterior are close to the true values, and that the true model that generated the data is supported by model selection.

The simulated data contained four balanced clusters of 100 observations. Slope parameters were four draws from a bivariate normal distribution with mean $\beta = (.5, .5)$ and covariance $\mathbf{D} = (.25, .1; .1, .25)$. The covariates were generated from $\mathcal{U}(0, 3)$. The error term was $\varepsilon_{ij} \sim \mathcal{N}(0, .10)$, and $\lambda_i = 0$ for all i.

I fit several different versions of the BCRC model to the simulated data. The models I report are:

- The full BCRC model with no restrictions on λ_i
- A restricted version of the BCRC model with $\lambda_i = \lambda$ for all i
- A semi-log model, with $\lambda_i = 0$ for all i

- A square root transformed model, with $\lambda_i = .5$ for all i
- A linear model, where $\lambda_i = 1$ for all i

Priors were chosen to represent relatively vague prior information. The hyper-prior for β was a normal distribution with mean vector (.25, .25) and a covariance matrix $.1 \times I_2$. The prior for σ^2 was an inverse gamma distribution with .1 and .1 degrees of freedom. The prior for D^{-1} was a Wishart distribution with location parameter 4 (twice the number of covariates) and scale matrix $2 \times I$. Finally, λ_i had a normal prior with mean 1 and variance of 3 for all i.

I tested several starting values for the iterations and found that convergence was not sensitive to starting values. I programmed and tested both of the candidate generating densities discussed in Section 2.4.3, the random walk and a tailored candidate generating density. Both approaches produced similar posterior estimates. The random walk candidate generating density was much faster, but suffered from higher serial correlation. The tailored candidate generating density was therefore used in the Metropolis-Hastings step for each simulation.

A full posterior analysis is presented only for Model 1. As can be seen in Figure 2.1, the MCMC sampler converged quickly. It appears that no more than 250 iterations were required for all the parameter distributions to converge. After

determining that the sample achieved sufficient convergence I generated a sample of 5000 draws after a burn-in period of 1000 iterations. The marginal posterior distributions are plotted in Figure 2.2, and the means and standard deviations of the simulated samples are reported in Table 2.1. Note that the means of the marginal posteriors except for λ_3 are all within one standard deviation of the true values, which suggests that the MCMC algorithm produces accurate estimates of the model parameters.

Autocorrelations for the sequence of draws for each of the parameters are plotted in Figure 2.3. The series of draws for β and D shows no evidence of serial correlation. There is diminishing serial correlation for σ^2 and the transformation parameters. The autocorrelations of the simulated sample are used determine whether there has been an adequate run length of the Markov chain. If the series of draws is independent and identically distributed, then the standard deviation of the sample mean is simply $\frac{\sigma}{\sqrt{n}}$. But if the series of draws is serially correlated, then the sample standard deviation is a function of the correlation. For example, the sample standard deviation of an AR(1) process can be shown to be $\frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho}{1-\rho}}$, where ρ is the autocorrelation of the series. The larger and more persistent is the autocorrelation of a series, the larger the sample that is required to have reliable estimates [71]. Bayesian studies usually report the autocorrelations of the

parameters of the model up to approximately 30 lags to indicate the accuracy of the simulated sample means [10, 15, 16].

If persistent serial correlation is present, there are several simulation strategies that may be used to reduce the correlation. For example, instead of keeping every draw after the burn-in period, every n^{th} draw may be retained. Alternatively, only the final iteration may be retained after the the sampler is run a large number of times. While this strategy is computationally less efficient, it does produce an independent sample. The fact that the draws from the posterior showed little or no serial correlation is evidence that accurate estimates have been produced.

Marginal likelihoods and Bayes factors for each of the five models are presented in Table 2.2. To find the posterior odds that Model 3 generated the data versus Model 1, find the element in row 3 and column 1. Model 3 is 4,915 times more likely than Model 1. The evidence most strongly favors Models 2 and 3, the restricted versions with $\lambda_i = \lambda$ and $\lambda_i = 0$ for all i. This was expected since these models generated the data, although Model 1 generated the data as well. The reason Model 1 is not favored is that it has more parameters than the other models. The Bayes factor is similar to Schwartz's Information criterion in penalizing more complex, higher dimensional models as long as simpler models fit the data well.

Model 2 and Model 3 are virtually indistinguishable, and the linear and square root models were overwhelmingly rejected.

2.7 Discussion

In this chapter I presented the BCRC model and developed a MCMC simulation algorithm to obtain Bayesian estimates of the model parameters. I also showed how Chib's BMI approach could be applied so that model selection could be undertaken. Finally, the model was fit to a set of simulated data. Results showed that the means of the marginal posterior distributions of the data were very close to the true values and that there was very little serial correlation in the simulated samples. This implies that the algorithm produced reliable estimates of posterior means. Finally, model selection favored the lowest dimensional model that generated the data because it had fewer parameters than the other models and still fit the data well. In the next chapter the BCRC model is fit to a set of real data to risk adjust the cost of inpatient hospital care.

Parameter	True Value	Posterior Mean	Standard Deviation
β_1	.500	.478	.318
β_2	.500	.612	.370
σ_2	.100	.118	.029
d_{11}	.25	.349	.801
d_{21}	.10	.101	.299
d_{22}	.25	.515	.524
λ_1	.000	-.135	.314
λ_2	.000	-.121	.156
λ_3	.000	.081	.062
λ_4	.000	.026	.099

Table 2.1: Summary of marginal posterior distribution of BCRC model fit to simulated data

	BCRC (-504.99)*	$\lambda_i = \lambda$ (-487.53)	$\lambda_i = 0$ (-504.95)	$\lambda_i = .5$ (-596.50)	$\lambda_i = 1$ (-3260.43)
BCRC	1	2.09×10^{-4}	2.03×10^{-4}	4.41×10^{27}	2.87×10^{142}
$\lambda_i = \lambda$	4786.72	1	0.9738	2.11×10^{31}	1.37×10^{146}
$\lambda_i = 0$	4915.26	1.03	1	2.17×10^{31}	1.41×10^{146}
$\lambda_i = .5$	2.27×10^{-28}	4.73×10^{-32}	4.61×10^{-32}	1	6.51×10^{114}
$\lambda_i = 1$	3.48×10^{-143}	7.27×10^{-147}	7.08×10^{-147}	1.54×10^{-115}	1

*Log of marginal likelihood is in parentheses

Table 2.2: Bayes factors for alternative models fit to simulated data

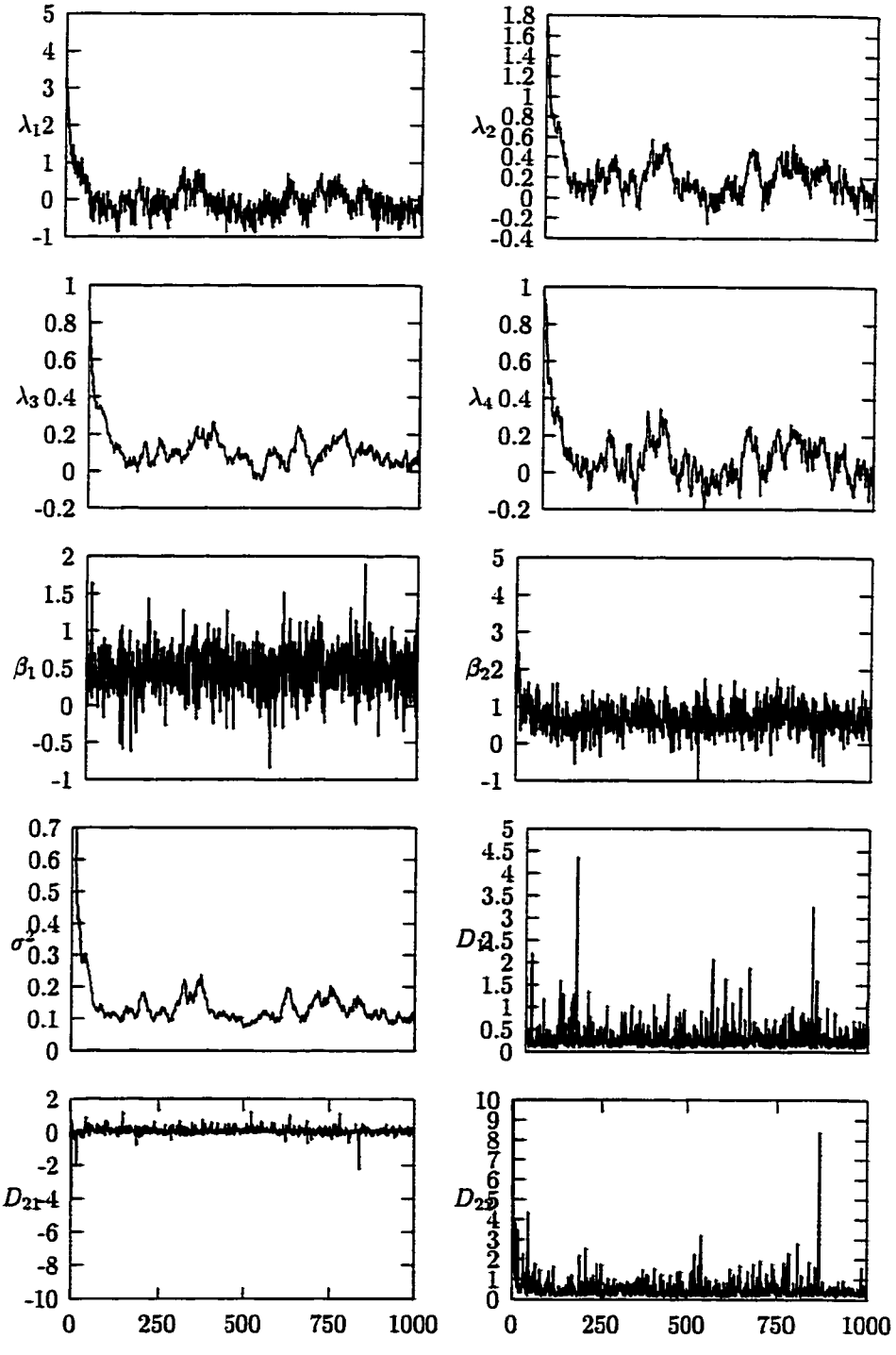

Figure 2.1: Transient stage draws of the BCRC model fit to simulated data

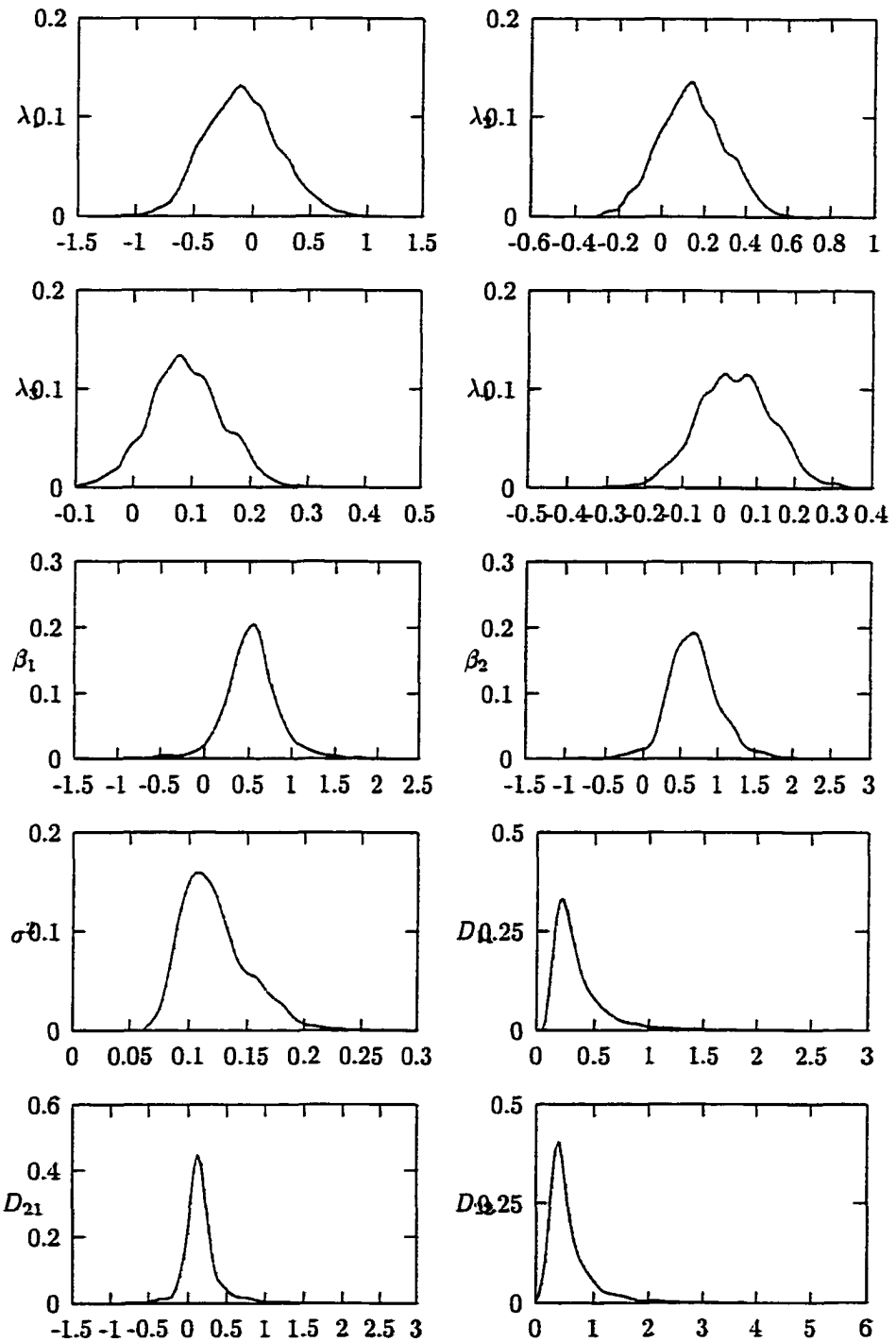

Figure 2.2: Plots of marginal posteriors of model parameters

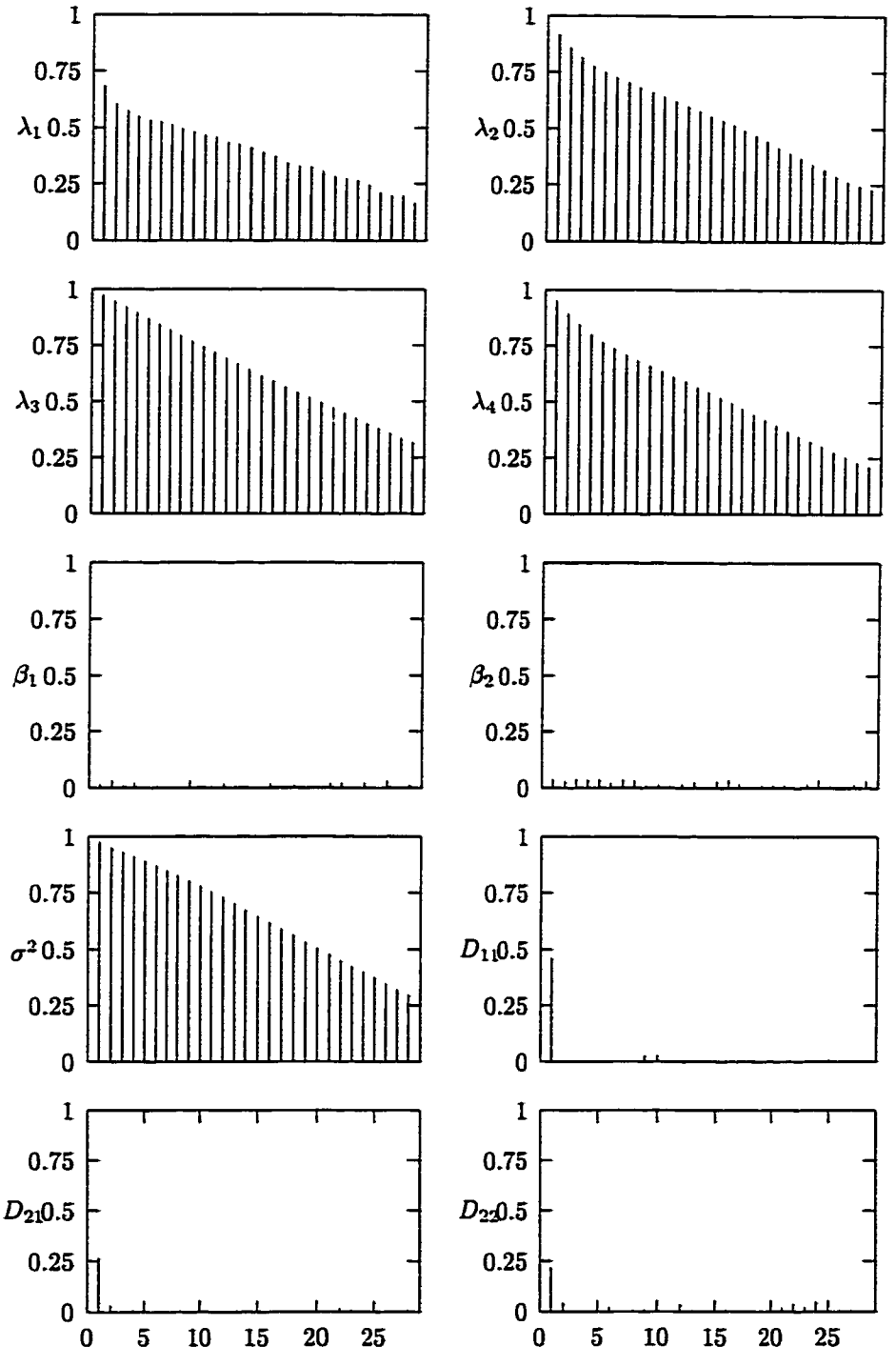

Figure 2.3: Autocorrelations of MCMC draws fit to simulated data

Chapter 3

Risk Adjusting the Cost of Coronary Artery Bypass Graft Surgery

3.1 Background

In recent years the healthcare industry has undergone considerable change. Most notably, there has been an increase in the prevalence of managed care, which has increased competition and introduced capitation payment schemes. These changes have increased interest in improving accountability and in evaluating and comparing the performance of healthcare providers. One method of comparing health care providers that is becoming more prevalent is to issue "report cards" that rank hospital performance based on patient outcomes such as length of stay [58], survival [38], and cost of treatment [70] compared to other hospitals. Simi-

larly, individual physicians are also evaluated and compared in physician profiles [35, 60].

Because healthcare providers treat heterogeneous populations, these performance comparisons must take into account the severity of the patients' underlying conditions, the comorbidities present, or the patient mix being treated. Doctors and hospitals that treat older, indigent, or other high risk patients may have higher mortality rates and costs not because they provide poorer quality of care, but rather because of the severity of the illness and the nature of the patients they treat. In order to make meaningful comparisons between healthcare providers, patient outcomes must be adjusted to reflect severity.

The process of accounting for underlying severity is called "risk adjustment."[1] Outcomes are risk-adjusted by first regressing the performance measure on an index of severity, or on patient-specific variables, including risk factors and comorbidities. Then the residuals from this regression, which have netted out any severity effects, are used to rank the doctors or hospitals.

Choosing an appropriate model is an important first step in risk adjustment. In the risk adjustment literature, the term "model" may refer to the choice of

[1]Some authors use the term risk adjustment for binary outcomes and severity adjustment for continuous outcomes [49]. I do not make this distinction.

independent variables or to the proposed functional relationship between performance and the independent variables. In this chapter the term is used with the latter definition in mind. When the performance measure is a binary outcome, such as survival, there is little question that a probit or logit is the appropriate statistical model for the risk adjustment regression. For continuous performance measures such as cost, the most appropriate model is less obvious. Since hospital costs tend to be skewed, the risk adjustment model may involve a transformation of the dependent variable so that the residuals are more nearly normally distributed. The linear model and the semi-log model are used most often in risk adjustment [35]. There has been little research into how appropriate these models are, whether another transformation is more appropriate, or whether the transformation is dependent upon context, for example the particular medical condition or type of healthcare provider under investigation.

The importance of using the most appropriate model was demonstrated by Schnitzler et al. [60], who used three alternative models to risk adjust the cost of treating community acquired pneumonia at 6 hospitals and produced a rank ordering of physicians by cost. The models included the linear model, the semi-log model, and robust estimation. They found that the rankings were model-dependent. Physicians who ranked in the highest (or lowest) 10 percent when cost

was regressed on severity and patient characteristics often ranked much lower (or higher) when the log of cost was used as the dependent variable or when a robust estimation procedure was used.

Using the most appropriate functional form for the risk-adjustment process is important because patients and healthcare providers use this information. For example, managed care organizations use risk-adjusted performance to select physicians for inclusion in their referral networks [49], and consumers may use this information to choose healthcare providers.

In 1990, the *New York Times* began to publish risk-adjusted mortality rates for all physicians and hospitals performing coronary artery bypass graft (CABG) surgery in New York. Mukamel and Mushlin [52] linked these data to physician claims submitted to Medicare and tested whether the published risk-adjusted data affected hospital and physician market shares and prices. They found that hospitals and physicians with better outcomes experienced higher rates of growth in market share and that physicians with better outcomes had higher rates of growth of charges for the procedure. In a similar study, Mennemeyer et al. [47] studied the impact of risk-adjusted outcomes reported by the Health Care Finance Administration (HCFA) between 1984 and 1992 on hospital admission rates for all

community hospitals in the Unites States. They found that hospitals with twice the expected mortality rate experienced fewer admissions; the effect was small but statistically significant. Although not concerned with costs, these studies show that consumers are sensitive to published risk-adjustments. To my knowledge, there have been no studies of the impact of published, risk-adjusted cost data on consumer behavior.

The accuracy of reported risk-adjusted performance, and therefore the model used in the risk adjustment process, is important if consumers and healthcare providers make decisions that are motivated by the information. Incorrect information may lead to suboptimal allocation of healthcare resources. This chapter uses the BCRC model to risk-adjust the cost of treating patients who undergo coronary artery bypass graft (CABG) procedures. Several questions are addressed using the model: 1) What is the most appropriate transformation of costs in a risk adjustment model? 2) What is the probability that λ_i, the BC transformation parameter, is in the neighborhood of 0 or 1? 3) Can it be assumed that the transformation is equal across hospitals? 4) How does the ranking induced by risk adjustment using the BCRC model compare to the rankings produced by maximum likelihood estimation of the default models? and 5) How does accounting for institutional effects affect the ranking?

The first and second questions were addressed by fitting the BCRC model to the hospital data and summarizing the marginal posterior distributions of the transformation parameters. The third question was addressed by model comparison of the BCRC model and other alternatives, including the restricted version of the BCRC model ($\lambda_i = \lambda$) and the semi-log model. Finally, the fourth and fifth questions were addressed by using the draws from the MCMC simulation to rank the hospitals and then making comparisons to rankings based on other models and unadjusted average costs.

The chapter proceeds as follows. Section 3.2 describes the data used in the risk adjustment. Section 3.3 describes the model fitting, including values chosen for hyperparameters and the performance of the MCMC sampler. The marginal posterior distributions are summarized in Section 3.4 and the results from the model selection are presented in Section 3.5. The hospitals are ranked in Section 3.6, and Section 3.7 concludes the chapter with discussion.

3.2 Hospital Data

Hospital cost data for this section were provided by the BJC Health System and were originally collected by the Greater St. Louis Healthcare Alliance (the "Al-

liance"), which was formed in 1992 as a voluntary coalition of regional hospitals, physicians, and managed care organizations with a mission of improving the quality of healthcare delivery and reducing costs. This section uses the subset of 496 CABG patients from four hospitals in the BJC Health System that was used in the risk adjustment analysis of the 1996 Comprehensive Hospital Performance Report [30].

3.2.1 Patients

The 496 patients were selected from all cardiovascular surgery patients at four regional hospitals between January 1, 1995, and December 31, 1995. The Alliance obtained administrative discharge data from Missouri's Hospital Industry Data Institute (HIDI) and selected a sample of patients using principal diagnosis codes based on the International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM) standard and Medicare's Diagnosis Related Groups (DRG), which indicate primary reasons for hospitalization. Appropriate diagnosis codes for patient selection were determined by a panel of cardiothoracic surgeons. Patients admitted for CABG were included in the study if one or more of the following ICD-9-CM principal diagnoses were present: 36.1X Bypass anastomosis for heart revascularization, 402.X1 Hypertensive heart disease with

congestive heart failure (CHF), 404.X1 Hypertensive heart and renal disease with CHF, 404.X3 Hypertensive heart and renal disease with CHF and renal failure, 410.XX Acute myocardial infarction, 411.XX Other acute and subacute forms of ischemic heart disease, 412.XX Old myocardial infarction, 413.XX Angina pectoris, 414.XX Other forms of chronic ischemic heart disease, 428.XX Heart failure, 785.51 Cardiogenic shock, 786.50 Chest pain not otherwise specified, 996.03 Mechanical complication of cardiac device, 996.72 Other complications of internal prosthetic device, E870.6 Accidental cut during catheterization, E878.2 Surgical procedure as cause of later complication, and E879.0 Cardiac catheterization as cause of later complication.

Patients were excluded from the study if one or more of the following principal diagnoses were present: 35.XX Operations on valves and septa of heart, 37.32 Excision of aneurysm, 38.34 Resection of aorta with anastamosis, 38.44 Resection of abdominal aorta with replacement, 38.45 Resection of thoracic aorta with replacement, 38.64 Other excision of aorta, 39.54 Other repair on vessels and reentry operation, 4211 Cervical esophagostomy, and 321 Other bronchial excision. Patients were also excluded if they were categorized under the following DRGs: 104 Cardiac valve procedure with catheterization or 105 Cardiac valve procedure without catheterization.

A random subset of qualifying patients was selected and additional clinical variables, determined by a panel of cardiovascular surgeons, were abstracted from their patient charts by an independent medical record abstraction service.

The cost of each patient's CABG surgery admission was obtained from the hospital's internal cost accounting database (McKesson-HBO, Inc., Atlanta, GA). These costs were generated by a departmental ratio of cost to charges (RCC) methodology. Patient charges in each hospital department were multiplied by the RCC reported in the hospital's annual Medicare cost report. The patient's total cost was the sum of department costs. Costs computed in this manner are not without problems. The most serious problem is that costs generated by the RCC method are more accurate at higher levels of aggregation. For example, the departmental cost is more accurate than the costs attributed to any particular service provided by the department. Furthermore, the literature suggests that in departments with high fixed and low variable costs, or where patient charges are based on a percentage mark-up, RCC estimates tend to be reliable. In such departments as radiology and surgery, where labor and supply inputs vary substantially, RCC estimates are less accurate [42, 43]. There are also problems in making cross-system comparisons with cost estimates derived from RCC methods since hospitals differ in their department structures and departmental cost ratios.

In spite of these defects, costs estimated by the RCC method do reflect resource utilization, and their use is standard practice in studies of the cost of providing health care [70].

3.2.2 Variables

Clinical variables used to risk-adjust the cost of CABG include blood albumin level (ALBUMIN), worsening congestive heart failure (CHF), prior open heart surgery (PRIOR), associated valve surgery (VALVE), hematocrit level (HEMATOCRIT), alveolar arterial oxygen gradient (AAO2), percutaneus transluminal coronary angioplasty (PTCA), and cardiac catheter (CATH), Table 3.1.

Serum albumin is the primary protein in the blood. Lower levels of albumin are associated with diabetes, renal insufficiency, and malnutrition [22]. Therefore, the coefficient for ALBUMIN is expected to be negative.

CHF, or the inability for the heart to provide adequate blood flow to vital organs, is a serious comorbidity for patients who undergo CABG surgery. The variable used in this analysis is a dummy variable that equals 1 if the patient's CHF had grown worse within 30 days prior to surgery. Patients with worsening CHF are expected to accrue higher costs.

Repeat procedures and associated valve surgery have been shown to increase risk for surgical site infections, perioperative death, and post-operative adverse outcomes. Since these lead to higher costs, VALVE and PRIOR are expected to have positive coefficients.

Hematocrit is the proportion of blood volume represented by red blood cells. Low hematocrit has implications for oxygenation of tissues and blood viscosity. The normal range for adult males is 42-54%, and for females is 38-46%. The variable used in this analysis is a dummy variable that equals 1 if hematocrit was below 33%. A gender-specific measure of hematocrit would have been preferred but was not available. The coefficient for HEMATOCRIT is expected to be positive.

AAO2 is the difference between the partial pressure of oxygen in the alveoli and the partial pressure of oxygen in the arterial blood. It is a measure of the efficiency of the lungs in moving oxygen from the air into the blood. Normal AAO2 ranges between 5 and 25 mmHg. The variable used in this analysis is a dummy variable equal to 1 for AAO2 levels greater than 30 mmHg. Higher levels indicate less efficient regulation, therefore a positive coefficient is expected for this variable.

PTCA, or balloon angioplasty, opens clogged coronary arteries via an inflated balloon in the artery. It is typically used as an alternative to CABG surgery in less severe cases since it is much less invasive. If a patient undergoes a PTCA and subsequently requires CABG surgery, then the PTCA was not adequate. The PTCA may indicate higher risk and is expected to have a positive sign.

Finally, cardiac catheterization is a procedure that locates coronary arterial blockages. It is a costly procedure, and patients who receive catheterization should incur higher costs. The expectation for CATH is positive.

Costs and patient characteristics for each of the four hospitals are summarized in Table 3.2. There was substantial variation across the four hospitals. The lowest cost hospital incurred on average $15,616 in costs to treat a CABG patient, while the highest cost hospital incurred on average $28,540, nearly twice as much. The highest cost hospital also appears to have treated more severe cases and had the highest proportion of patients with worsening CHF, low hematocrit, and high AAO2.

A histogram of costs, plotted in Figure 3.1, shows the typical problem with using levels of hospital costs in regression analyses. Costs are highly skewed. The mode of costs is approximately $20,000, but there are several outliers with costs

over $100,000.

3.3 Model Fitting

The BCRC model was fit to the clustered hospital cost data. The sampler was run for an initial burn-in period of 1000 iterations, and an additional 5,000 draws were taken from the posterior and used for inference. Several different starting values were tested, and convergence of the Markov chain did not depend on the starting point. The model that was fit to the hospital data was:

$$\begin{aligned} \text{COST}_{ij}^{(\lambda_i)} &= b_{i0} + b_{i1}\text{ALBUMIN}_{ij} + b_{i2}\text{CHF}_{ij} + b_{i3}\text{PRIOR}_{ij} + b_{i4}\text{VALVE}_{ij} \\ &\quad + b_{i5}\text{HEMATOCRIT}_{ij} + b_{i6}\text{AAO2}_{ij} + b_{i7}\text{PTCA}_{ij} \\ &\quad + b_{i8}\text{CATH}_{ij} + \varepsilon_{ij}, \end{aligned} \tag{3.1}$$

where $i = 1, \ldots, 4$ indexes the hospitals, and $j = 1, \ldots, n_i$ indexes within-cluster observations.

3.3.1 Priors

Priors for the model were based on maximum likelihood estimates that ignored the hospital clusters. A grid search was used to estimate the following risk-adjustment

model:

$$\begin{aligned} \text{COST}_i^{(\lambda)} &= \beta_0 + \beta_1 \text{ALBUMIN}_i + \beta_2 \text{CHF}_i + \beta_3 \text{PRIOR}_i + \beta_4 \text{VALVE}_i \\ &\quad + \beta_5 \text{HEMATOCRIT}_i + \beta_6 \text{AAO2}_i + \beta_7 \text{PTCA}_i \\ &\quad + \beta_8 \text{CATH}_i + \varepsilon_i. \end{aligned} \tag{3.2}$$

As can be seen in Figure 3.2, the likelihood function was maximized when the transformation was -.875, suggesting that a transformation close to the inverse of costs is the most appropriate transformation, and not a log transformation that might be used by default. Maximum likelihood estimates for the parameters in equation (3.2) are reported in Table 3.3.

Priors of $\mathbf{b}_i$ are normal distributions with mean vector equal to the maximum likelihood estimates (.657, .085, .112, .118, .112, .083, .104, .219, -.804) and covariance matrix with .5 on the diagonal and .25 on all off-diagonal elements. The prior of each λ_i is a normal distribution with mean -.875 and a variance of 3. The prior for σ^2 is an inverse gamma distribution with hyperparameters of .1 and .1, which imply a mean of 1 and variance of 10. The prior for $\mathbf{D}^{-1}$ is a Wishart distribution with 18 degrees of freedom (twice the number of independent variables) and a scale matrix with diagonal elements of 9 and off-diagonal elements of zero. All of the priors were centered at maximum likelihood estimates, yet had large

enough variance to represent relatively vague information. Priors were not chosen to reflect complete ignorance (i.e. they are not uniform distributions), because model selection is important to this study. Model comparison is not possible when ignorance priors are used, because marginal likelihoods are derived by integrating the likelihood over the priors, which requires finite mass in the priors.

3.3.2 Performance of the MCMC Sampler

The MCMC output from the model fitting was very well behaved. The first 1000 draws of the sampler are plotted in Figures 3.3 and 3.4, which show that convergence occurred rapidly. In fact, no more than 200 iterations appear to be required for convergence. Autocorrelations up to 30 lags for the 500 retained draws are plotted for transformation parameters in Figure 3.5 and for the slope coefficients in Figure 3.6. While the transformation parameters are serially correlated, the correlation is only approximately .3 by lag 30. The slope coefficients are virtually free of serial correlation.

3.4 Results

The posterior distributions for the model coefficients are plotted in Figure 3.7 and are summarized in Table 3.4. Because a Bayesian analysis treats parameters as random and produces a probability distribution rather than a point estimate, I report the mean of the marginal posterior distribution and the probability that the coefficient is greater than zero.

The means of the marginal posterior distributions had the expected signs. Congestive heart failure ($Pr(CHF > 0) = 0.74$), prior procedures ($Pr(PRIOR > 0) = 0.77$) and associated valve procedures ($Pr(VALVE > 0) = 0.75$) all resulted in higher costs with high probability. Low hematocrit ($Pr(HEMATOCRIT > 0) = 0.77$) and high AAO2 ($Pr(AAO2 > 0) = 0.57$) were also associated with higher costs. Cardiac catheterization ($Pr(CATH > 0) = 0.72$) and PTCA ($Pr(PTCA > 0) = 0.84$) contributed to higher costs, and higher serum albumin ($Pr(ALBUMIN > 0) = 0.30$) resulted in lower costs. Note by comparing Table 3.4 to Table 3.3 that the posterior means from the BCRC model were very similar to maximum likelihood estimates.

The means and standard deviations of the unique elements of the covariance matrix $\mathbf{D}$ are reported in Table 3.5. The diagonal elements indicate how the

effect of the risk factors varies across hospitals. The fact that the variance is large relative to the posterior mean, approximately .09 on average, suggests that there is substantial variation across hospitals.

The posterior distributions of the transformation parameters indicate that the the most appropriate transformation for costs is not natural log, but inverse. The posterior means of the transformation parameters ranged from a low of -1.13 to a high of -0.65. As seen in Table 3.4, there is a very small probability that any of the transformation parameters is greater than zero.

The posterior distributions of the transformation parameters, plotted in Figure 3.8, also indicated that the transformations were not equal. While the distributions for the transformation for Hospitals B and D are very similar, the distributions for the other hospitals are not. For example, the probability that λ_D is greater than the posterior mean of λ_B is .48 and the probability that λ_B is less than the mean of λ_D is .698. But the probability that λ_C is greater than the posterior mean of λ_B is .98 and the probability that λ_B is less than the mean of λ_C is .97.

3.5 Model Selection

Additional evidence that the transformations are not equal across hospitals was found by model selection. I computed Bayes factors to compare a range of models, Table 3.6. Equal prior probability was assigned to each of the alternative models, therefore the Bayes factor equals the posterior odds that one model generated the data compared to another model. Reading across row 1 of Table 3.6 reveals that the data support the flexible BCRC model over all of the other models. The BCRC model was 2,099 times more likely to have generated the data than restricted BCRC model and 73,200 times more likely to have generated the data than the inverse transform model. Not surprisingly, reading in row 2 and column 5 of Table 3.6 we see that the inverse transform and the restricted BCRC model are virtually indistinguishable from each other. Still, the evidence is strongest for the flexible BCRC model. Recall that the Bayes factor is a function of the posterior distribution ordinates and that higher dimensional models are penalized in manner similar to the Schwartz criterion. The fact that the Bayes factor supports the BCRC model in spite of its additional parameters is strong evidence that the transformation differs across hospitals.

3.6 Hospital Ranking

3.6.1 Ranking from Hierarchical Parameters

Since risk adjustment is often used to compare performance, after fitting the risk adjustment model I investigated how using the BCRC model for risk adjustment affects the hospital ranking. Most studies that use regression to estimate the parameters of a risk adjustment model rank the hospitals by the residuals [29, 50]. Predicted costs are subtracted from observed costs, and the difference is averaged across hospitals to represent the hospital's overuse or underuse of resources. The hospital with the largest average residual is the highest cost hospital, and the hospital with the smallest average residual is the lowest cost hospital.

This type of ranking can be produced by computing residuals using the hierarchical parameters (β) of the BCRC model for each draw of the MCMC sample. Although a very small literature exists on ranking, Goldstein and Spiegelhalter suggest using the output from the MCMC simulation by computing the each hospital's rank at each iteration of the MCMC sampler, and then computing the mean or median rank for each hospital [2, 23, 29]. The advantage of this approach is that it is then possible to compute both the probability distribution of ranks for each hospital and the relative odds that one hospital occupies the rank versus

another hospital. A residual and rank was computed for each hospital for each draw of the MCMC sampler, and averages were used as the basis for ranking.

The ranking produced by the hierarchical parameters of the BCRC model is reported in Table 3.7, along with the unadjusted ranking, and the rankings produced by maximum likelihood estimation of the linear and semi-log models. The risk adjustment clearly affected the hospital ranking. Risk adjustment moved Hospital B from the position of second highest cost to lowest cost. Hospital A remained in the position of highest cost, and Hospitals C and D retained their positions relative to each other. The BCRC model produced exactly the same ranking as that produced by maximum likelihood estimation of the linear and semi-log models. This is not entirely surprising since there are only four hospitals in the data and the means of the marginal posterior distributions were so similar to the maximum likelihood estimates. There are also only four clusters in the data and they were fairly evenly distributed across the costs. We would expect differences in ranking to happen where there are small margins of difference between hospitals.

One important question is how likely it is that one hospital occupies another rank. In the classical context, this is done by interval estimation around the

residuals using Monte Carlo simulation or bootstrapping. In the Bayesian context we can compute the probability distribution of the hospitals across ranks, and the relative odds that one hospital versus another occupies a particular rank. In Table 3.8 I report the probability distributions of the hospitals across ranks, which can be used to compute the relative odds. For example, Hospital A is ranked as the highest cost hospital, but there is some probability that both Hospital C and Hospital D are the highest. Based on the probabilities in Table 3.8, Hospital A is 14.8 times more likely than Hospital C and 23.7 times more likely than Hospital D to be the highest cost hospital.

3.6.2 Ranking From Cluster Parameters

Goldstein and Spiegelhalter [29] argue that it is important to account for institutional effects in the risk adjustment model, but doing this has implications for the ranking. On an intuitive level, the idea of risk adjustment is to remove information from the residuals that relates to the individual risk factors and to leave any information that relates to the institution. If institution effects are included in the model, for example by including hospital dummy variables, then the average residual for each hospital will equal zero and ranking based on the residuals will not be possible.

To illustrate the issue, consider the hypothetical data in Figure 3.9. At these two hypothetical hospitals, the risk factor has the same incremental effect on cost, and the hospitals differ only in their intercept terms. Fitting a regression model to these data would produce all positive residuals for the hospital with the larger intercept, and all negative residuals for the hospital with the smaller intercept. Ranking these hospitals based on the residuals would place the higher cost hospital above the lower cost hospital, as desired. If a hospital dummy variable were included, the average residuals for both hospitals would be zero.

If risk factors affect cost in different ways at each hospital then it is important to model hospital-specific effects explicitly. Consider another hypothetical example in Figure 3.10. In this example, one hospital appears to treat the higher risk patients at relatively lower cost, but treats the lower risk patients at relatively higher cost, as might be the case at a university affiliated medical center versus a community medical center. In this case, ranking based on residuals will be arbitrary whether or not institution-specific dummy variables are included in the model. The most appropriate ranking for this case is based on the expected cost conditional on a value of the risk factor. For example, in Figure 3.10 the expected cost for the hospital represented by bullets is lower than the hospital represented by diamonds conditional on a value of 5 for the risk factor, but higher conditional

on a value of 10 for the risk factor. It is clear that if the model includes hospital specific factors then the ranking must be based on expected cost conditional on a value of the risk factor.

Using the BCRC model as the risk adjustment model allows for this type of ranking by computing expected costs from the cluster-level slope parameters ($\mathbf{b}_i$). I have already presented evidence that suggests that hospital-specific effects may be important. Recall that the diagonal elements of the covariance matrix were large relative to the values of the random coefficients. This implies large variation in the effects of the risk factors across hospitals, and the ranking produced from the cluster level parameters may differ from the ranking produced by the hierarchical parameters at some levels of risk.

To rank the hospitals by expected cost I estimated the average predicted cost for each hospital as

$$\bar{c}_{ij} = \frac{1}{n_i G} \sum_{g=1}^{G} \sum_{j=1}^{n_i} (\lambda_i^{(g)} \bar{\mathbf{x}}'_{ij} \mathrm{b}_i^{(g)} + 1)^{\frac{1}{\lambda_i^{(g)}}} \tag{3.3}$$

where g indexes draws from the MCMC sampler, j indexes within-hospital observations, and i indexes hospitals, b_i is the vector of slope parameters, and $\bar{\mathbf{x}}_{ij}$ is the vector of covariates which have been fixed at one of five levels of risk: extreme low, low, average, high, and extreme high. For extreme low risk I set all binary

covariates to zero and continuous covariates the sample minimum. For low risk I set all covariates to one standard deviation below the mean of the sample. Average risk is the mean of the sample. For high risk I set covariates to one standard deviation above the mean, and for extreme high risk I set binary covariates equal to 1 and continuous variables to the sample maximum.

I encountered one obstacle computing predicted costs using this method. The formula in equation (3.3) does not restrict costs to be greater than zero for the entire vector space of the risk factors. Because λ_i is negative with high probability, for large enough values of $\bar{\mathbf{x}}_{ij}$, the quantity $(\lambda_i^{(g)}\bar{\mathbf{x}}_{ij}'\mathbf{b}_i^{(g)} + 1)$ is less than one. Therefore the solution requires raising a negative number to a non-integral power. As might be expected, this was only a problem in creating rankings for high and extreme high risk. Only the iterations of the MCMC sample that yielded real predicted costs for equation (3.3) were used to compute predicted costs and rankings.

The predicted cost for each hospital conditional upon risk level is presented in Table 3.9 along with the average ranking. There are two important results. First, notice that the hospitals are ranked exactly as before for extreme low, low and average risk. But the ranking is different for high and extreme high risk.

Second, the average ranking was not always consistent with the predicted cost. Hospital B has the highest predicted cost, but the second-highest average ranking conditional on both high and extreme high risk. This was because Hospital B had a few observations with extremely large predicted costs that skewed the average predicted cost but not the average ranking.

The probability distributions of Hospitals across ranking categories is presented in Table 3.10, and suggests that for extreme high risk, the relative odds that Hospital A is the highest cost versus Hospital B is 1.14, which implies they are equally likely to be the highest cost hospital. For high risk, the relative odds that Hospital A versus B is the highest cost hospital is 2.7, only mild evidence to suggest that Hospital A is the higher cost hospital.

The fact that Hospital B moved from lowest cost to highest cost suggests that while Hospital B provides care for low and medium risk at a lower cost, it costs much more to care for higher risk patients. This might be expected for a small community medical center. In fact, Hospital B is the smallest community medical center in the data set. The ranking also shows that Hospital A is the highest cost hospital for all levels of risk. Hospital A is a large, university affiliated medical center.

3.7 Discussion

This section used the BCRC model as a risk adjustment model for the cost of treating CABG patients. I fit the BCRC model to data from 496 CABG patients at four regional hospitals in order to determine the most appropriate transformation and model. Posterior means for the hospitals ranged from -1.14 to -.65, suggesting that the most appropriate transformation was most nearly inverse, and not natural log. The results also suggested that the transformation should not be considered equal across hospitals. Not only were the posterior means for the transformation parameters varied, but Bayes factors overwhelmingly supported the most flexible BCRC model. I also used the output of the MCMC sampler to rank the hospitals by risk-adjusted cost. The ranking produced by the hierarchical parameters was identical to the ranking produced by maximum likelihood estimates of the BC, linear, and semi-log models. Using the cluster-level parameter estimates allowed for a consideration of hospital-specific effects. The ranking conditional on low and average risk was identical to the ranking from the hierarchical parameters. The ranking assuming high and extreme high risk, however, was different, and moved the lowest cost hospital to the position of second highest cost hospital. The probability distributions of the hospitals across ranks also showed that Hospital

A and Hospital B were almost equally likely to be the highest cost hospital.

Variable	Description	Expected Sign
COST	Total inpatient hospital costs	
ALBUMIN	Blood albumin level (gm/dl)	-
CHF	Worsening CHF within last 30 days dummy	+
PRIOR	Prior open heart surgery dummy	+
VALVE	Associated valve surgery dummy	+
HEMATOCRIT	Hematocrit $< 33\%$ dummy	+
AAO2	Alveolar arterial oxygen gradient > 30 mmHg dummy	+
PTCA	Balloon angioplasty dummy	+
CATH	Cardiac catheter without PTCA dummy	+

Table 3.1: Description of clinical variables used in risk adjustment

	Hospitals				
	All ($n = 496$)	**A** ($n = 107$)	**B** ($n = 167$)	**C** ($n = 58$)	**D** ($n = 164$)
COST	$20,338	$28,540	$20,811	$17,197	$15,616
ALBUMIN	3.92	3.93	3.74	3.99	4.07
CHF	11%	19%	10%	19%	5%
PRIOR	8%	7%	9%	12%	7%
VALVE	9%	11%	9%	22%	3%
HEMATOCRIT	13%	22%	10%	10%	12%
AAO2	11%	32%	3%	16%	4%
PTCA	2%	3%	2%	0%	2%
CATH	61%	64%	77%	45%	50%

Table 3.2: Characteristics of CABG patients by hospital

Variable	Coefficient	Standard Error	p - value
INTERCEPT	0.657	0.074	0.0001
CHF	0.085	0.025	0.0008
PRIOR	0.112	0.027	0.0001
VALVE	0.118	0.027	0.0001
HEMATOCRIT	0.112	0.023	0.0001
AAO2	0.083	0.024	0.0007
CATH	0.104	0.016	0.0001
PTCA	0.219	0.054	0.0001
ALBUMIN	-0.084	0.018	0.0001

$R^2 = .32$

Table 3.3: Results from maximum likelihood estimation using CABG data

Variable	Mean	Standard Deviation	Median	$\Pr(\beta_k > 0)$
INTERCEPT	0.5956	0.1833	0.5937	0.9980
CHF	0.0911	0.151	0.0943	0.7356
PRIOR	0.1031	0.1496	0.1027	0.7600
VALVE	0.0979	0.1534	0.0981	0.7460
HEMATOCRIT	0.1058	0.1521	0.1029	0.7736
AAO2	0.0212	0.1515	0.0184	0.5564
CATH	0.0827	0.1522	0.0825	0.7178
PTCA	0.1707	0.1814	0.1694	0.8434
ALBUMIN	-0.0773	0.1543	-0.0772	0.2958
σ^2	0.0204	0.0079	0.019	1.0000
λ_1	-0.9198	0.183	-0.9189	0.0000
λ_2	-1.1395	0.3236	-1.1316	0.0002
λ_3	-0.6517	0.2011	-0.6514	0.0020
λ_4	-1.0472	0.352	-1.0359	0.0060

Table 3.4: Summary of marginal posterior distributions of BCRC model fit to CABG data

	1	2	3	4	5
D_1	0.0969				
	(0.0458)				
D_2	-0.0004	0.0915			
	(0.0301)	(0.0421)			
D_3	-0.0003	-0.0004	0.0903		
	(0.0294)	(0.0285)	(0.0418)		
D_4	0.0007	-0.0007	0.0006	0.093	
	(0.031)	(0.0291)	(0.0286)	(0.045)	
D_5	-0.0002	0.0002	0.0008	0.0002	0.0926
	(0.0314)	(0.0299)	(0.0292)	(0.0286)	(0.0445)
D_6	-0.0005	0.0004	0	0.0004	0
	(0.0306)	(0.0296)	(0.0282)	(0.029)	(0.0298)
D_7	-0.0002	0.0003	-0.0002	0.0016	0.0015
	(0.0297)	(0.0288)	(0.0292)	(0.0299)	(0.0294)
D_8	0	-0.0019	0.0013	0.0011	-0.0004
	(0.0317)	(0.0306)	(0.0305)	(0.0313)	(0.0308)
D_9	-0.0002	0.0009	-0.0001	-0.0002	-0.0002
	(0.0306)	(0.0291)	(0.0287)	(0.0303)	(0.0301)
	6	7	8	9	
D_6	0.0917				
	(0.0422)				
D_7	-0.0002	0.0921			
	(0.0294)	(0.044)			
D_8	-0.0003	0.0004	0.1009		
	(0.0309)	(0.0309)	(0.0489)		
D_9	-0.0004	-0.0003	0.0007	0.0913	
	(0.0294)	(0.0285)	(0.0306)	(0.0434)	

Table 3.5: Summary of marginal posterior distribution of slope covariance matrix

Model	**BCRC** (-837.88)*	$\lambda_i = \lambda$ (-845.53)	$\lambda_i = 0$ (-916.64)	$\lambda_i = -0.5$ (-860.29)	$\lambda_i = -1$ (-849.08)
BCRC	1	2098.97	1.60×10^{34}	5.38×10^{9}	7.32×10^{4}
$\lambda_i = \lambda$	4.76×10^{-4}	1	7.61×10^{30}	2.56×10^{6}	34.85
$\lambda_i = 0$	6.26×10^{-35}	1.31×10^{-31}	1	3.37×10^{-25}	4.58×10^{-30}
$\lambda_i = -0.5$	1.86×10^{-10}	3.90×10^{-7}	2.97×10^{24}	1	1.36×10^{-5}
$\lambda_i = -1$	1.37×10^{-5}	0.0287	2.18×10^{29}	7.35×10^{4}	1

*Log of marginal likelihood is in parentheses.

Table 3.6: Bayes factors for models fit to CABG data

	Risk Adjustment Model			
Ranking	**No Adjustment***	**BCRC Model†**	**Linear Model‡**	**Semi-log Model‡**
1	A (28,540)	A (1.126)	A (0.5767)	A (0.2254)
2	B (20,811)	C (2.012)	C (0.0339)	C (0.0419)
3	C (17,197)	D (2.901)	D (-0.2908)	D (-0.1348)
4	D (15,616)	B (3.960)	B (-0.3393)	B (-0.1554)

* Average unadjusted cost is in parentheses.

† Average rank is in parentheses.

‡ Average residual is in parentheses.

Table 3.7: Hospital rankings based on hierarchical parameters.

	Hospital A	Hospital B	Hospital C	Hospital D
Pr(Rank=1)	0.901	0.000	0.061	0.038
Pr(Rank=2)	0.073	0.000	0.864	0.062
Pr(Rank=3)	0.026	0.038	0.074	0.862
Pr(Rank=4)	0.000	0.962	0.000	0.038

Table 3.8: Probability distribution of hospitals across ranks

Risk	Hospital A	Hospital B	Hospital C	Hospital D
Extreme Low Risk	$ 17,385 (1.15)	$ 10,506 (3.99)	$ 16,009 (1.84)	$ 12,535 (3.00)
Low Risk	$ 18,621 (1.03)	$ 11,554 (3.99)	$ 16,793 (1.96)	$ 13,331 (3.00)
Average Risk	$ 22,694 (1.00)	$ 15,183 (3.53)	$ 19,256 (2.00)	$ 15,217 (3.46)
High Risk	$ 40,283 (1.26)	$ 40,895 (1.90)	$ 25,814 (2.92)	$ 21,673 (3.91)
Extreme High Risk	$ 157,826 (1.48)	$ 454,087 (1.66)	$ 35,477 (3.39)	$ 43,190 (3.46)

Average rank is in parentheses

Table 3.9: Predicted cost and rank for five levels of risk

	Hospital A	Hospital B	Hospital C	Hospital D
		Extreme Low Risk		
Pr(Rank=1)	0.842	0.000	0.158	0.000
Pr(Rank=2)	0.158	0.000	0.842	0.000
Pr(Rank=3)	0.000	0.005	0.000	0.995
Pr(Rank=4)	0.000	0.995	0.000	0.005
		Low Risk		
Pr(Rank=1)	0.965	0.000	0.035	0.000
Pr(Rank=2)	0.035	0.000	0.965	0.000
Pr(Rank=3)	0.000	0.001	0.000	0.999
Pr(Rank=4)	0.000	0.999	0.000	0.001
		Average Risk		
Pr(Rank=1)	1.000	0.000	0.000	0.000
Pr(Rank=2)	0.000	0.000	1.000	0.000
Pr(Rank=3)	0.000	0.466	0.000	0.534
Pr(Rank=4)	0.000	0.534	0.000	0.466
		High Risk		
Pr(Rank=1)	0.732	0.268	0.000	0.000
Pr(Rank=2)	0.268	0.593	0.131	0.008
Pr(Rank=3)	0.000	0.110	0.817	0.073
Pr(Rank=4)	0.000	0.029	0.052	0.919
		Extreme High Risk		
Pr(Rank=1)	0.528	0.461	0.000	0.011
Pr(Rank=2)	0.463	0.447	0.033	0.057
Pr(Rank=3)	0.009	0.058	0.542	0.392
Pr(Rank=4)	0.000	0.034	0.425	0.541

Table 3.10: Probability distributions of hospitals across ranks by risk

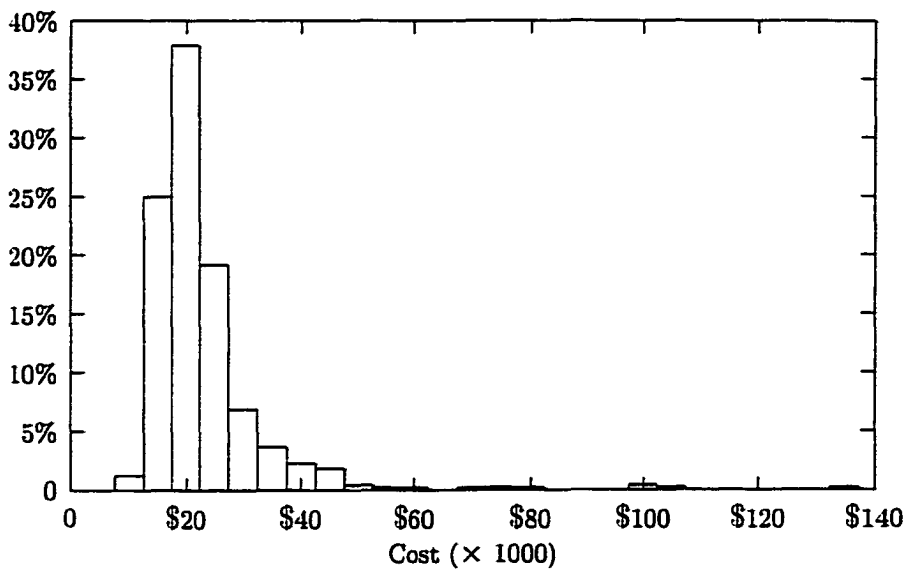

Figure 3.1: Histogram of inpatient hospital costs for CABG patients

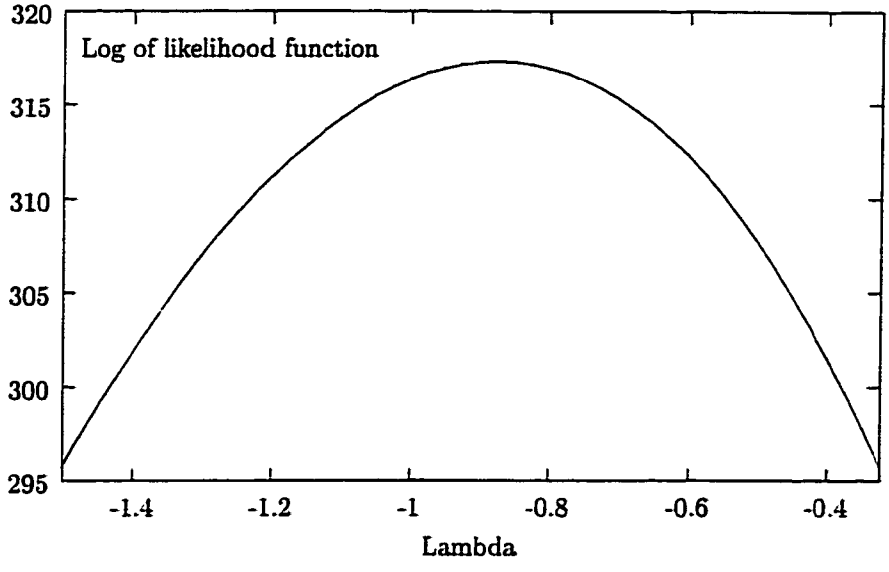

Figure 3.2: Log of likelihood function

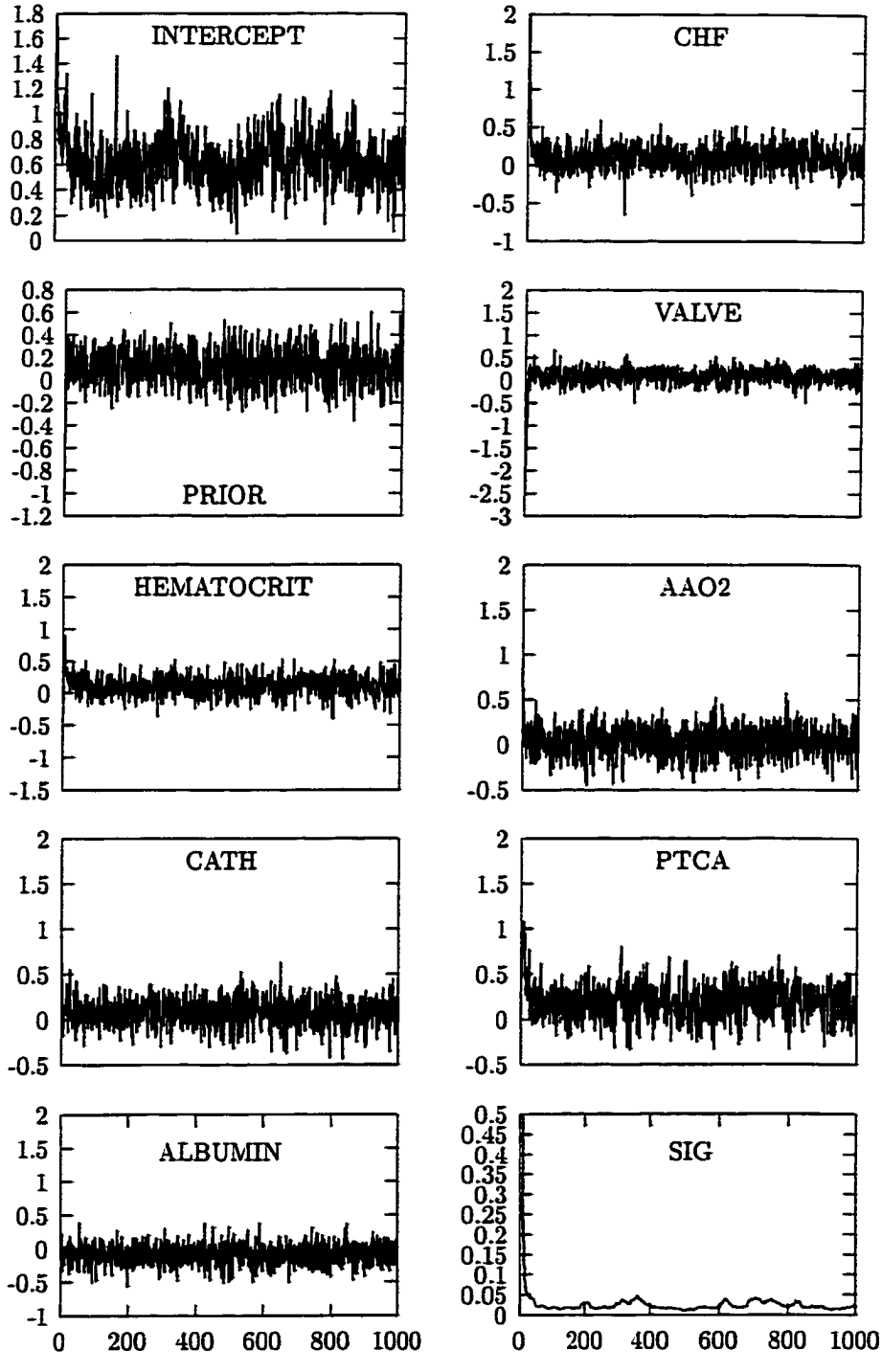

Figure 3.3: Transient draws from slope posteriors of BCRC model

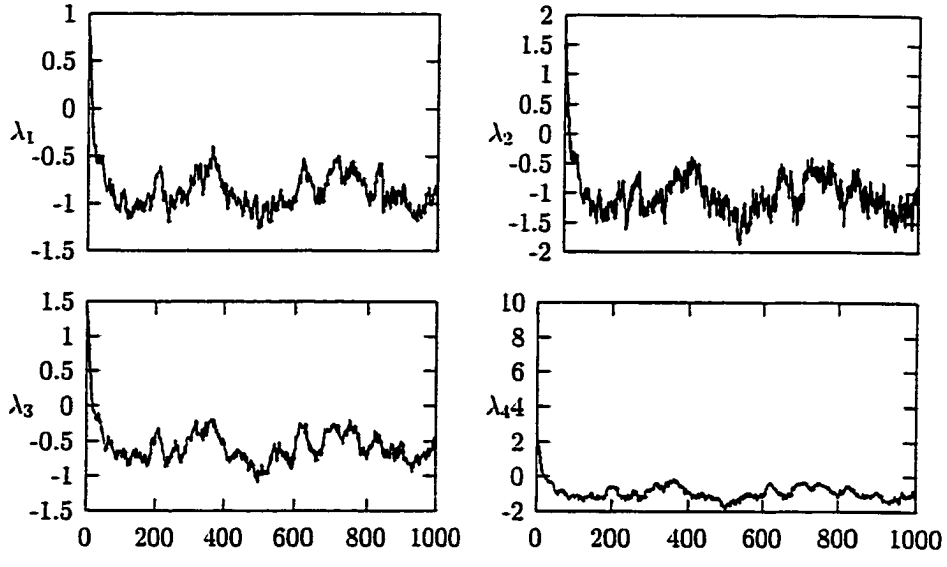

Figure 3.4: Transient draws of transformation posteriors of BCRC model fit to CABG data

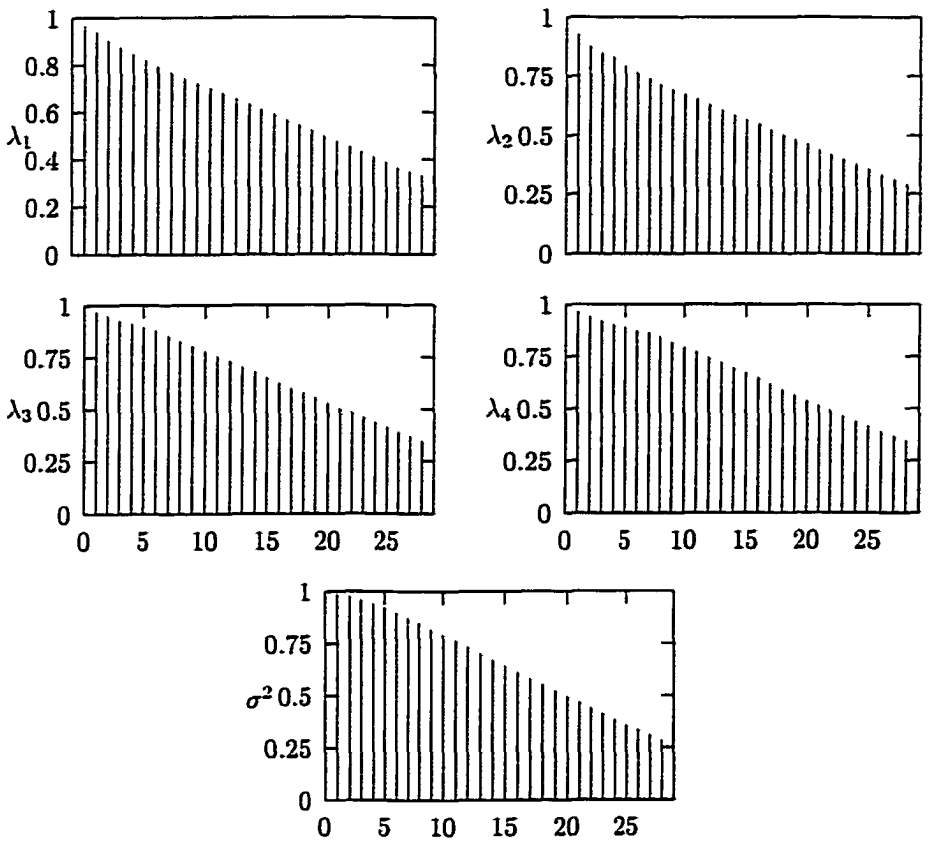

Figure 3.5: Autocorrelations of transformation and variance parameters

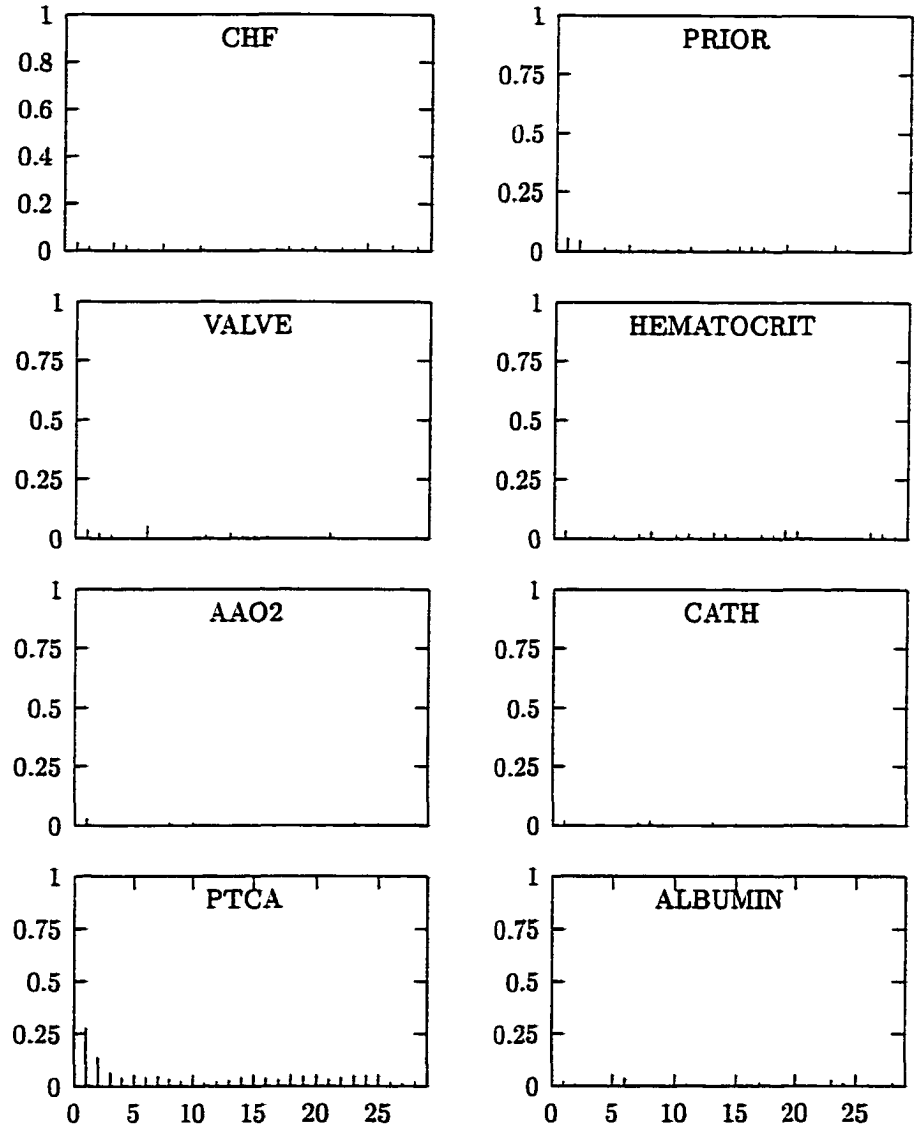

Figure 3.6: Autocorrelations of slope parameters

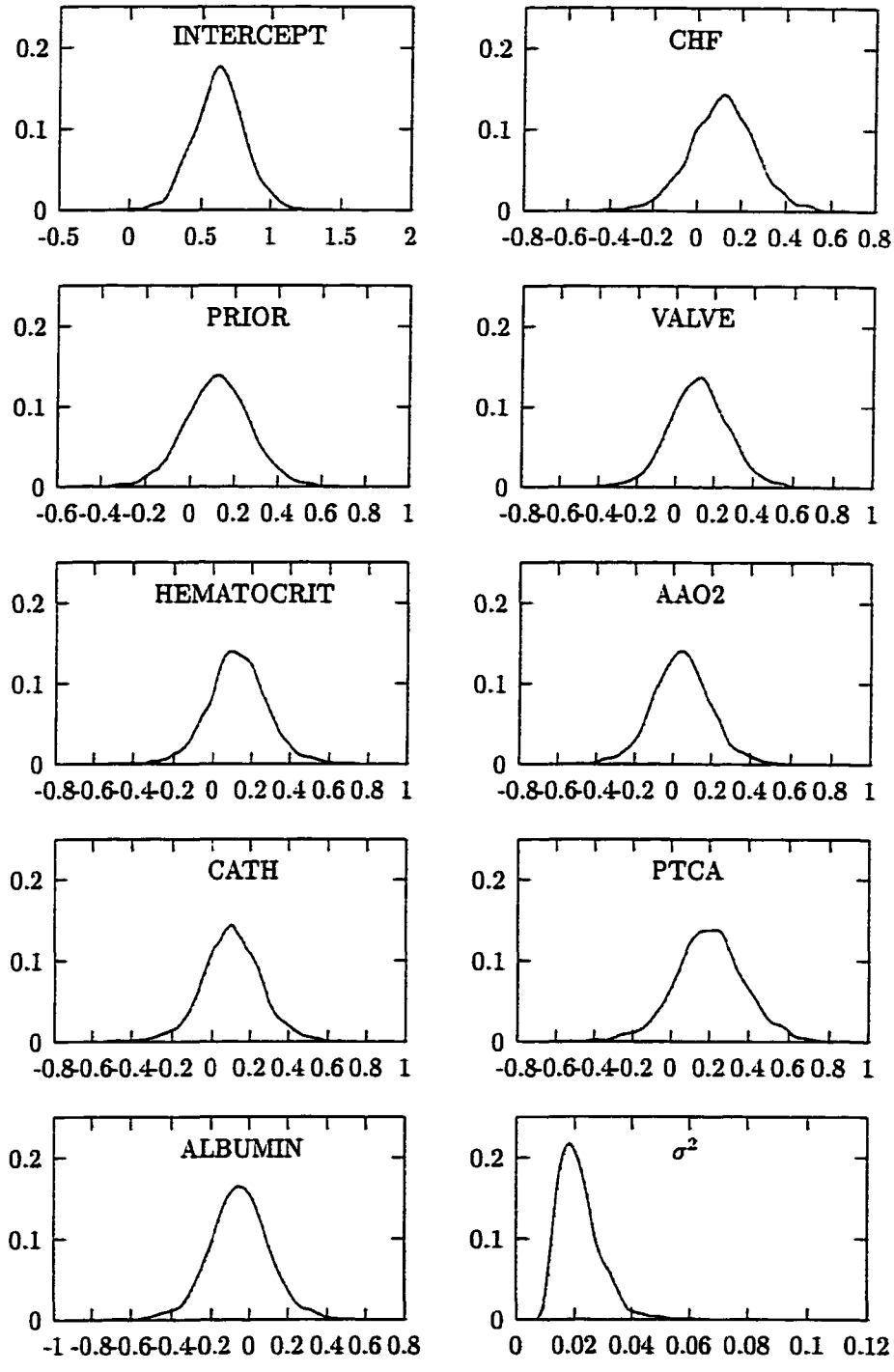

Figure 3.7: Plots of marginal posterior distributions of model coefficients

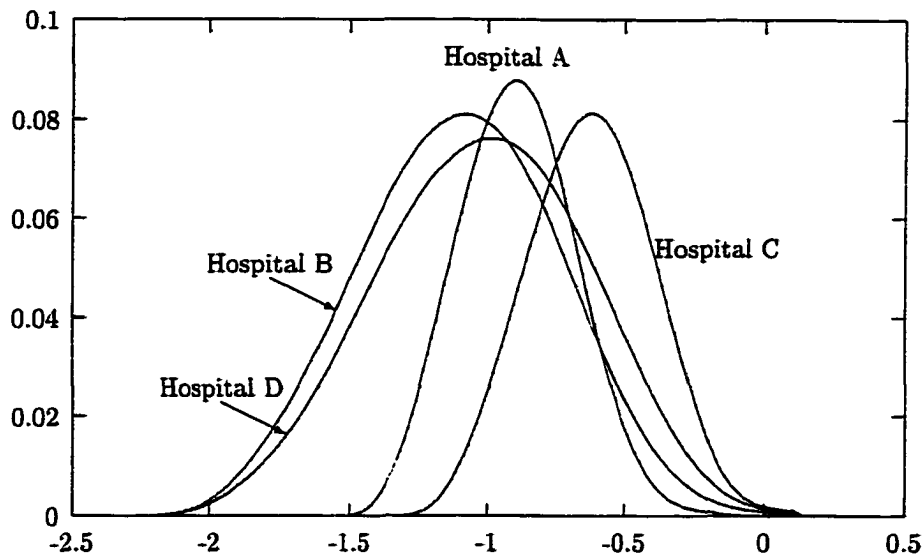

Figure 3.8: Plots of marginal posterior distributions of transformation parameters

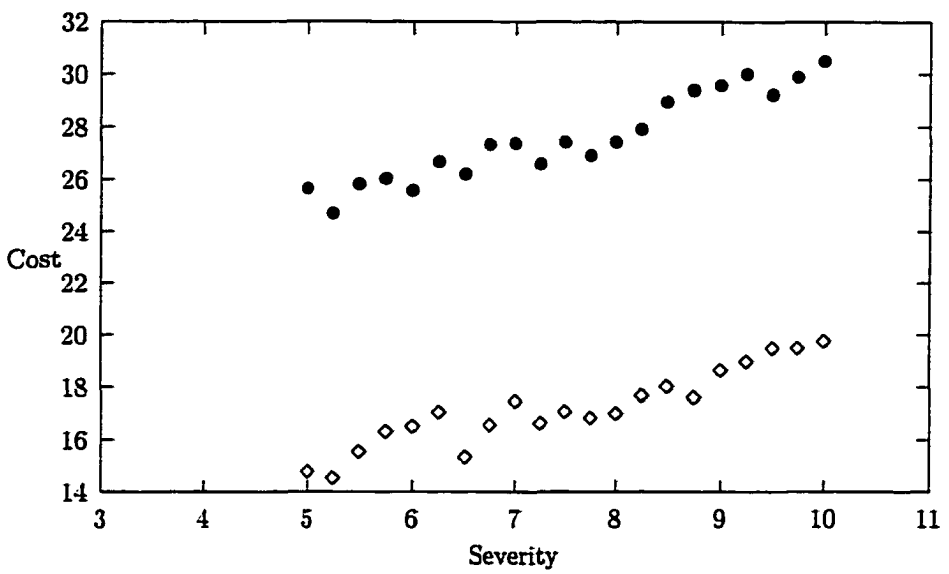

Figure 3.9: Hypothetical cost data, same marginal effect on cost

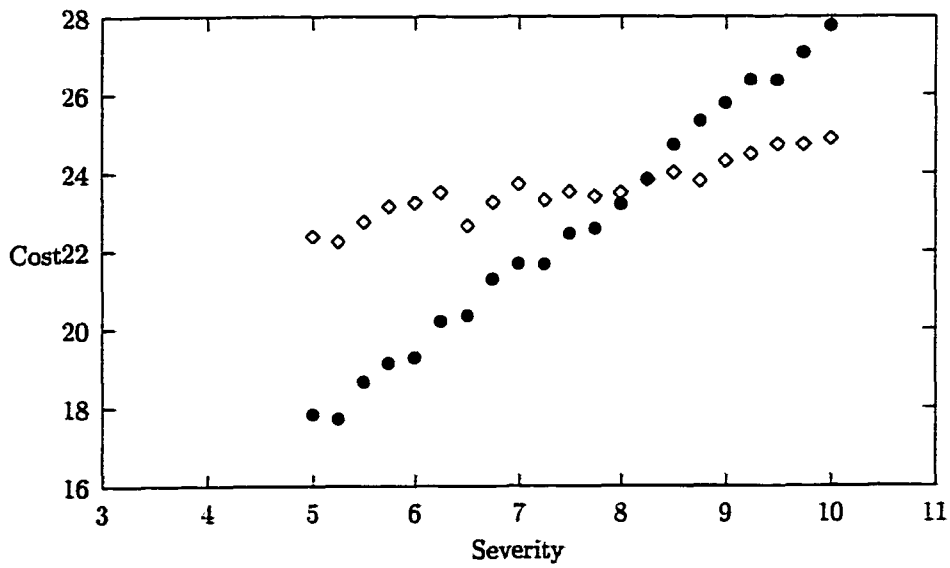

Figure 3.10: Hypothetical cost data, different marginal effect on cost

Chapter 4

Conclusion

The purpose of this dissertation is to present a BC model for clustered data and to use the model to shed light on current practices in risk adjusting continuous health care outcomes. I introduced the BC transformation into a random coefficients model and showed how Bayesian estimates of the model parameters could be obtained by MCMC simulation. Fitting the model to simulated data showed that the MCMC algorithm performed well. Posterior means were very close to the parameters that generated the data, and Bayes factors supported the model that created the data. The BCRC model was then used as a risk adjustment model for the cost of treating patients undergoing CABG surgery. Results from the model fitting suggested that the natural log transformation of costs that is commonly used in risk adjustment models was not appropriate for these data, and that the

transformation varied across hospitals. Finally, the hospitals were ranked based on risk-adjusted costs using both the hierarchical parameters and the cluster-level parameters. The ranking of the hospitals using the hierarchical parameters was identical to the ranking produced by maximum likelihood estimates of the BC, linear and semi-log models. Using the cluster-level parameters, however, the hospital that ranked lowest using the other methods, was ranked second most costly in treating high and extreme high risk patients.

There are two main contributions of this research. First, to my knowledge this is the first use of the BC transformation in a clustered data setting. The model and estimation algorithms presented here are valuable to the extent that functional form is an important consideration in clustered data applications. Second, although one other author has suggested that the BC transformation may be valuable in a risk adjustment context, this is the first study that actually uses the transformation in a risk adjustment model [61].

This research also provides another example of the value of Bayesian methods in applied health economic research. Bayesian methods are not widely used in applied empirical health care research, and the Food and Drug Administration (FDA) does not yet accept Bayesian analyses in New Drug Applications. In spite

of this, Bayesian techniques have found support in such applications as decision analysis [39], modeling disease progression [18], monitoring clinical trials [19], technology assessment [66], and meta-analysis [34], and the use of Bayesian inference is increasing. Perhaps the most compelling reason to use Bayesian methods in a risk adjustment context is the ability to perform model selection and compare alternative, non-nested models. There is no counterpart for comparing non-nested models with classical methods.

The results of the study also leave us with something of a puzzle, and suggest areas for future research. While it appears that the linear and log models were inappropriate, in the sense that the transformation parameters were not close to zero or one, this did not appear to make a difference for the hospital ranking. But the BCRC model also allows for a much richer description of the ranking. Hospitals can be ranked conditional not only upon the risk level of the current sample of patients, but also based on different risk levels.

Refinements of the model may be helpful for future research. For example, the error structure that was assumed for the BCRC model is much more simple than is assumed in most clustered data models. Since this project began, Chib and Greenberg [15] have shown that more general covariance matrices can be simulated by

partitioning the matrix and simulating the parameters in blocks. Future research should include more general assumptions about the error covariance matrix.

There are also several issues in risk adjustment that merit further exploration. First, it would be important to know whether variation across hospitals requires different transformations across hospitals, or whether that variation can be adequately modeled with hospital fixed effects. A Bayesian approach is still most consistent with this goal since Bayes factors can be computed to produce the relative odds of these non-nested alternative models.

Second, a larger, more comprehensive database is clearly needed since it is likely that the small number of hospitals accounts for the fact that the BCRC model did not produce a different hospital ranking than did the maximum likelihood results. It is also unclear whether the Bayes factor will favor the BCRC model over restricted versions in the presence of more clusters, because Bayes factors penalize models with higher dimensions. It would be important to study the robustness of this result to the number of clusters. Furthermore, the data used in this study lacked information on nosocomial infections, which have been shown to be a significant determinant of inpatient hospital costs following CABG procedures [36]. I suspect that many of the high cost outliers in the data are pa-

tients with deep chest surgical site infections or antibiotic resistant bloodstream infections. The lack of infection data is a serious omission.

Additional data on other medical conditions besides CABG surgery and in other risk adjustment contexts, such as physician profiling, would also be very useful. The results presented here suggest that the log transformation was not the most appropriate transformation, but it is not clear whether this is true in other settings. It would be important to study whether the most appropriate transformation is dependent entirely upon context, or whether more general statements can be made about the most appropriate risk adjustment model.

Appendix A

GAUSS Program

```
/* ------------------------------------------------
*
    This program estimates the BCRC model with
    a tailored candidate generating density in
    the MH step.
    bi ~ N(.5,.5: .5, .25 | .25, .5)
         model is yit = xit'bi + eit
         bi is N(beta,D)
         beta is N(beta0,B0^(-1))
         D^(-1) is Wish(v00,R00)
         s2 is IG(v0/2,d0/2)
 * ------------------------------------------------ */

new;
library pgraph,kernel;
graphset;
fonts("simplex complex microb simgrma");

format /rd 5,4;
tstart = date;
seed1=3937841;
seed2=2845328;
seed3=1472742;
seed4=2531548;
seed5=1735914;
seed6=3249387;
seed7=1324625;
seed8=2931443;

clear p,n;

/* ----------  Create Simulated Data   ---------- */
one=ones(400,1);
xes=2-2*rndus(400,1,seed1);
X=one~xes;
```

```
eror=sqrt(.1)*rndns(400,1,seed6);
ind=(100|100|100|100);

genbet1=(0.2070 | 0.0038);
genbet2=(0.7656 | 0.1018);
genbet3=(0.7920 | 1.1975);
genbet4=(0.4862 | 0.7364);

geny1=exp(X[1:100,.]*genbet1    + eror[1:100]);
geny2=exp(X[101:200,.]*genbet2  + eror[101:200]);
geny3=exp(X[201:300,.]*genbet3 + eror[201:300]);
geny4=exp(X[301:400,.]*genbet4 + eror[301:400]);

y=(geny1|geny2|geny3|geny4);

nn=sumc(ind);
n=rows(ind);
k=cols(X);
p=k;

_plctrl = -1;
_pstype = 8;
_psymsize = 0;

xy(seqa(1,1,nn),y);

/* === Set Parameters for Priors === */
/*------------ Normal prior for Beta ------------*/
b0_ = .25*ones(k,1);
B0__ = .1*eye(k);
/*------------ IG PRIOR of s2 -------------------*/
v0_ = .1;
d0_ = .1;
/*------------ Wishart PRIOR of Dinv ------------*/
v00_ = 2*p;
R0_ = eye(p);
/*------------ Normal PRIOR of lamda ------------*/
lam_ = 0*ones(n,1);
tau2 = 3;

/* ===  Set Parameters for MCMC Iterations  ====*/
n0 = 1000;
m  = 5000;
capn = n0 + m;

/* ===  Starting values for the Iterations  === */
D0  = inv(v00_*R0_);     /* inverse of the prior mean */
s20 = 2;
beta0 = b0_+ chol(inv(B0__))'rndn(k,1);
lam0 = ones(n,1);

retain=1;
t=1;

/*-------------- Set storage for output ------------*/
pp      = rows(vech(D0));  /* unique elements in D */
dm      = zeros(pp,m);
s2m     = zeros(1,m);
betam   = zeros(k,m);
lambdam = zeros(n,m);
```

```
bm      = zeros(m,k*n);

md = zeros(pp,n0);
madbmal = zeros(n,n0);
mateb   = zeros(k,n0);
m2s     = zeros(1,n0);
mb      = zeros(n0,k*n);

/* ==========  Begin MCMC simulation  ==============*/
do while t le capn;

/*--------------- Generation of b ------------------*/
b1 = zeros(1,k*n);
sse = 0;

i = 1; j = 1; kk=1;

do while i le n;
    ni=ind[i];

    y1 = y[j:j+ni-1,.];
    lami = lam0[i];
    yi = ylam(y1, lami);
    Xi = X[j:j+ni-1,.];
    Vi = inv((inv(D0) + (Xi'Xi)/s20 ));

    bihat = Vi*(inv(D0)*beta0 + (Xi'yi)/s20);
    bi = bihat + chol(Vi)'rndn(p,1);
    b1[kk:kk+k-1] = bi';
    sse = sse + (yi - Xi*bi)'(yi - Xi*bi);
    kk=kk+k;
    j = j + ni;
    i = i + 1;
endo;

/*------------ Generation of s2 -----------------*/
s21 = rninvgam( (v0_+rows(y))/2,(sse+d0_)/2 );

/*------------ Generation of beta ----------------*/
bb1=zeros(p,n);
Dinv = invpd(D0);
V = inv(B0__ + n*Dinv);
i=1; j=1;
    do while i le n;
    bb1[.,i] = b1[1,j:j+k-1]';
    i=i+1; j=j+k;
    endo;
bhat = V*(B0__*b0_ + Dinv*sumc(bb1'));
beta1 =  bhat + chol(V)'rndn(k,1);

/*------------ Generation of D -------------------*/
e = bb1 - beta1;
R1 = invpd(invpd(R0_) + e*e');
D1 = rnwish(v00_+n,R1);
D1 = invpd(D1);

/*------------ Generation of lambda --------------*/
lam1   = zeros(n,1);
nrlams = zeros(n,1);
rwvars  = zeros(n,1);
```

```
i=1; j=1;
do while i le n;

ni=ind[i];
Xi = X[j:j+ni-1,.];
yi = y[j:j+ni-1,.];
bi = bb1[.,i];
lami = lam0[i];

nrlams[i],rwvars[i] = newtraph(ni,Xi,yi,bi,lam_[i],s21,tau2);
i=i+1; j=j+ni;
endo;

/*----------- Begin Metropolis-Hastings step -----------*/
i=1; j=1;
do while i le n;
    ni=ind[i];
    Xi = X[j:j+ni-1,.];
    yi = y[j:j+ni-1,.];
    bi = bb1[.,i];
    lami = lam0[i];
    rwvar = sqrt(rwvars[i]);
    rnx = rnchisq(10);
    candi = nrlams[i] + rwvar*rndns(1,1,seed3);
    lam1[i] = mh(Xi,yi,bi,s21,tau2,lam_[i],lami,candi);
j=j+ni; i = i + 1;
endo;

/*----------- Store Results ----------*/
if t le n0;
    md[.,t]      = vech(D1);
    m2s[.,t]     = s21;
    mateb[.,t]   = beta1;
    madbmal[.,t] = lam1;
    mb[t,.]      = b1;
elseif t gt n0;
    dm[.,t-n0]      = vech(D1);
    s2m[.,t-n0]     = s21;
    betam[.,t-n0]   = beta1;
    lambdam[.,t-n0] = lam1;
    bm[t-n0,.]      = b1;
else;
endif;

beta0  = beta1;
D0 = D1;
s20 = s21;
b0 = bb1;
lam0 = lam1;

t = t + 1;
endo;

save bm, betam, s2m, dm, lambdam;

/* ======= Report Results and Export Draws ======= */
/*--------- Calculate posterior summaries -------- */
bout    = summary(bm');
betaout = summary(betam);
```

```
sout    = summary(s2m);
dout    = summary(dm);
lout    = summary(lambdam);

s1 = "   mean  ";
s2 = "   sd    ";
s3 = "   med   ";
s = s1~s2~s3;

/* ===  Print Results  === */

output file = q:/dissert/simdata/output/nrnl_out.txt reset;

?;

print "              SIMULATION RESULTS";
print "-------------------------------------------------------";
print "Full BCRC Model (lam_i = lam_i), Tailored candidate generating density" ;
?;
format /rd 4,6;  " Simulated sample size = " m;
" Clusters = " n;
?;
print "--------------------------";
format /rd 5,5; " b:";
format /rd 7,6;
print $s;
format /rd 7,4;
print bout;
?;
print "--------------------------";
format /rd 5,5; " beta:";
format /rd 7,6;
print $s;
format /rd 7,4;
print betaout;
?;
print "--------------------------";
format /rd 5,5; " s2:";
format /rd 7,6;
print $s;
format /rd 7,4;
print sout;
?;
print "--------------------------";
format /rd 5,5; " D:";
format /rd 7,6;
print $s;
format /rd 7,4;
print dout;
?;
print "--------------------------";
format /rd 5,5; " Lambda:";
format /rd 7,6;
print $s;
format /rd 7,4;
print lout;
?;
print "--------------------------";
?;
print      "retain rate     = %" ((retain-1)/(capn*n))*100;
```

```
?;

/*=================================================*
 ********** COMPUTE MARGINAL LIKELIHOODS *********
 *=================================================*/

/* ----- Set theta* ----- */

bstar = meanc(bm); bbstar = zeros(k,n); i=1; j=1;
    do while i le n;
    bbstar[.,i] = bstar[j:j+k-1];
    i=i+1; j=j+k;
    endo;
betastar = meanc(betam'); dstar = meanc(dm'); ddstar = xpnd(dstar);
sigstar = meanc(s2m'); lamstar = meanc(lambdam');

/* ---- Compute Prior Ordinates ---- */

lnf = lnlikfnord(y,X,betastar,ddstar,sigstar,lamstar,ind); prordbeta =
lnnormord(betastar,b0_,B0__); prordd    =
lnwishord(ddstar,v00_,inv(R0_)); prordsig  = lnigord(sigstar,v0_,d0_);
lamord = zeros(n,1); i=1; do while i le n;
    lamord[i] = lnnormord(lamstar[i],lam_,tau2);
    i=i+1;
    endo;
prordlam  = sumc(lamord);

/* ---- Compute Posterior of beta ---- */

post1 = zeros(m,1); newb2=zeros(n,k);

t=1; do while t le m;

j=1;
    for i (1,n,1);
      newb1=bm[t,.];
      newb2[i,.] = newb1[j:j+k-1];
      j=j+k;
    endfor;

    sumb  = sumc(newb2);

    D0    = xpnd(dm[.,t]);
    Dinv = invpd(D0);

    V = inv(B0__ + n*Dinv);
    bhat = V*(B0__*b0_ + Dinv*sumb);

    post1[t] = lnnormord(betastar,bhat,V);
    t=t+1;
endo;

postbeta = sumc(post1)/m;

/*=================================================*
 **********  REDUCED RUNS OF THE SAMPLER  **********
 *=================================================*/

/* ===  Run sampler with beta* fixed and compute
        f(D* | beta*, s2(g), lam(g),y)     === */
```

```
/*------------ Set storage for output ----------*/

bm2    = zeros(m,k*n); dm2      = zeros(pp,m); s2m2     = zeros(1,m);
lambdam2 = zeros(n,m);

beta0  = betastar; D0 = ddstar; s20 = sigstar; b0 = bbstar; lam0 =
lamstar;

/* 2222222222  Begin second MCMC simulation  22222222222 */

t=1; do while t le m;

/*-------------- Generation of b ------------------*/ b1 =
zeros(1,k*n); sse = 0;

j = 1; kk=1; for i (1, n, 1);
    ni=ind[i];
    y1 = y[j:j+ni-1,.];
    lami = lam0[i];
    yi = ylam(y1, lami);
    Xi = X[j:j+ni-1,.];
    Vi = inv((inv(D0) + (Xi'Xi)/s20 ));
    bihat = Vi*(inv(D0)*betastar + (Xi'yi)/s20);
    bi = bihat + chol(Vi)'rndn(p,1);
    b1[kk:kk+k-1] = bi';
    sse = sse + (yi - Xi*bi)'(yi - Xi*bi);
    kk=kk+k;
    j = j + ni;
endfor;

bb1=zeros(p,n);
    j=1;
for i (1,n,1);
    bb1[.,i] = b1[1,j:j+k-1]';
    j=j+k;
    endfor;

/*----------- Generation of s2 ------------------*/ sse = 0;

j = 1; for i (1,n,1);
    ni=ind[i];
    y1 = y[j:j+ni-1,.];
    lami = lam0[i];
    yi = ylam(y1, lami);
    Xi = X[j:j+ni-1,.];
    bi = bb1[.,i];
    sse = sse + (yi - Xi*bi)'(yi - Xi*bi);
    j = j + ni;
    endfor;

s21 = rninvgam((v0_+rows(y))/2,(sse+d0_)/2);

/*----------- Generation of D ------------------*/ e = bb1 -
betastar; R1 = invpd(invpd(R0_) + e*e'); D1 = rnwish(v00_+n,R1); D1 =
invpd(D1);

/*----------- Generation of lambda --------------*/
```

```
nrlams2=zeros(n,1);

i=1; j=1; do while i le n;

ni=ind[i]; Xi = X[j:j+ni-1,.]; yi = y[j:j+ni-1,.]; bi = bb1[.,i]; lami
= lam0[i];

nrlams2[i],rwvars[i] = newtraph(ni,Xi,yi,bi,lam_[i],s21,tau2);

i=i+1; j=j+ni;

endo;

/*------------ Begin Metropolis-Hastings step ------------*/ i=1; j=1;

do while i le n;
    ni=ind[i];
    Xi = X[j:j+ni-1,.];
    yi = y[j:j+ni-1,.];
    bi = bb1[.,i];
    lami = lam0[i];
    rwvar = sqrt(rwvars[i]);

candi = nrlams2[i] + rwvar*rndns(1,1,seed4);

lam1[i] = mh(Xi,yi,bi,s21,tau2,lam_[i],lami,candi);

j=j+ni; i = i + 1; endo;

/*------------ Store Results -----------*/ bm2[t,.]      = b1; dm2[.,t]
= vech(D1); s2m2[.,t]     = s21; lambdam2[.,t] = lam1;

beta0  = betastar; D0 = D1; s20 = s21; b0=bb1; lam0 = lam1;

t = t + 1; endo;

/* ---- Compute Posterior of D ---- */

post2 = zeros(m,1); v00n=v00_+n; t=1; do while t le m;

bibeta=zeros(k,k);
    j=1;
    for i (1,n,1);
    newb1=bm2[t,.];
    bibeta=bibeta+(newb1[j:j+k-1]-betastar)*(newb1[j:j+k-1]-betastar)';
    j=j+k;
    endfor;
    R1 = invpd(invpd(R0_) + bibeta);
    post2[t,.] = lnwishord(ddstar,v00n,inv(R1));
    t=t+1;
endo;

postd = sumc(post2)/m;

/* === Run sampler with beta*, D* fixed and compute
     f(s2* | beta*, D*, bi(g), y)      === */

/*------------ Set storage for output ----------*/

bm3  = zeros(m,n*k); s2m3     = zeros(1,m); lambdam3 = zeros(n,m);
```

```
beta0    = betastar; D0       = ddstar; s20      = sigstar; b0       =
bbstar; lam0     = lamstar;

/* 333333333   Begin third MCMC simulation  333333333 */

t=1; do while t le m;

/*--------------- Generation of b ------------------*/ b1 =
zeros(1,k*n); sse = 0;

j = 1; kk=1; for i (1, n, 1);
    ni=ind[i];
    y1 = y[j:j+ni-1,.];
    lami = lam0[i];
    yi = ylam(y1, lami);
    Xi = X[j:j+ni-1,.];
    Vi = inv((inv(D0) + (Xi'Xi)/s20 ));
    bihat = Vi*(inv(D0)*betastar + (Xi'yi)/s20);
    bi = bihat + chol(Vi)'rndn(p,1);
    b1[kk:kk+k-1] = bi';
    sse = sse + (yi - Xi*bi)'(yi - Xi*bi);
    kk=kk+k;
    j = j + ni;
endfor;

bb1=zeros(p,n);
    j=1;
for i (1,n,1);
    bb1[.,i] = b1[1,j:j+k-1]';
    j=j+k;
    endfor;

/*----------- Generation of s2 ------------------*/ sse = 0;

j = 1; for i (1,n,1);
    ni=ind[i];
    y1 = y[j:j+ni-1,.];
    lami = lam0[i];
    yi = ylam(y1, lami);
    Xi = X[j:j+ni-1,.];
    bi = bb1[.,i];
    sse = sse + (yi - Xi*bi)'(yi - Xi*bi);
    j = j + ni;
    endfor;

s21 = rninvgam( (v0_+rows(y))/2,(sse+d0_)/2 );

/*----------- Generation of lambda --------------*/
nrlams3=zeros(n,1);

i=1; j=1; do while i le n;

ni=ind[i]; Xi = X[j:j+ni-1,.]; yi = y[j:j+ni-1,.]; bi = bb1[.,i]; lami
= lam0[i];

nrlams3[i],rwvars[i] = newtraph(ni,Xi,yi,bi,lam_[i],s21,tau2);

i=i+1; j=j+ni;
```

```
endo;

/*----------- Begin Metropolis-Hastings step -----------*/ i=1; j=1;

do while i le n;
    ni=ind[i];
    Xi = X[j:j+ni-1,.];
    yi = y[j:j+ni-1,.];
    bi = bb1[.,i];
    lami = lam0[i];
    rwvar = sqrt(rwvars[i]);

candi = nrlams3[i] + rwvar*rndns(1,1,seed5);

lam1[i] = mh(Xi,yi,bi,s21,tau2,lam_[i],lami,candi);

j=j+ni; i = i + 1; endo;

/*----------- Store Results ----------*/ bm3[t,.]      = b1; s2m3[.,t]
= s21; lambdam3[.,t] = lam1;

s20 = s21; b0 = bb1; lam0 = lam1;

t = t + 1; endo;

/* ---- Compute Posterior Ordinate of s2 ---- */ post3   = zeros(m,1);

t=1; do while t le m;

    newb1=bm3[t,.];
    newb2 = zeros(n,k);

j=1;
    for i (1,n,1);
    newb2[i,.] = newb1[j:j+k-1];
    j=j+k;
    endfor;

ssepost=0;

i=1; j=1; do while i le n;
        bi = newb2[i,.]';
        ni=ind[i];
        y1 = y[j:j+ni-1,.];
        Xi = X[j:j+ni-1,.];
        lami  = lambdam3[t,i];
        yi = ylam(y1, lami);
        ssepost = ssepost + (yi - Xi*bi)'(yi - Xi*bi);
    i=i+1; j=j+ni;
    endo;

    post3[t] = lnigord(sigstar,(v0_+nn)/2,(d0_+ssepost)/2);
    t=t+1;
endo;

postsig = sumc(post3)/m;

/* === Run sampler with beta*, D* fixed and compute
     f(lam* | beta*, D*, s2* bi(g), y)      === */
```

```
/*------------ Set storage for output ----------*/ bm4 = zeros(m,k*n);
lambdam4 = zeros(n,m);

beta0  = betastar; D0 = ddstar; s20 = sigstar; b0 = bbstar; lam0 =
lamstar;

/* 4444444  Begin fourth MCMC simulation  4444444444*/

t=1; do while t le m;

/*-------------- Generation of b ------------------*/ b1 =
zeros(1,k*n); sse = 0;

j = 1; kk=1; for i (1, n, 1);
    ni=ind[i];
    y1 = y[j:j+ni-1,.];
    lami = lam0[i];
    yi = ylam(y1, lami);
    Xi = X[j:j+ni-1,.];
    Vi = inv((inv(D0) + (Xi'Xi)/sigstar ));
    bihat = Vi*(inv(D0)*betastar + (Xi'yi)/sigstar);
    bi = bihat + chol(Vi)'rndn(p,1);
    b1[kk:kk+k-1] = bi';
    sse = sse + (yi - Xi*bi)'(yi - Xi*bi);
    kk=kk+k;
    j = j + ni;
    endfor;

bb1=zeros(p,n);
    j=1;
for i (1,n,1);
    bb1[.,i] = b1[1,j:j+k-1]';
    j=j+k;
    endfor;

/*----------- Generation of lambda --------------*/
nrlams4=zeros(n,1);

j=1; for i (1,n,1);
    ni=ind[i];
    Xi = X[j:j+ni-1,.];
    yi = y[j:j+ni-1,.];
    bi = bb1[.,i];
    lami = lam0[i];
    nrlams4[i],rwvars[i] = newtraph(ni,Xi,yi,bi,lam_[i],sigstar,tau2);
    j=j+ni;
    endfor;

/*----------- Begin Metropolis-Hastings step -----------*/ i=1; j=1;
do while i le n;
    ni=ind[i];
    Xi = X[j:j+ni-1,.];
    yi = y[j:j+ni-1,.];
    bi = bb1[.,i];
    lami = lam0[i];
    rwvar = sqrt(rwvars[i]);

candi = nrlams4[i] + rwvar*rndns(1,1,seed6); lam1[i] =
mh(Xi,yi,bi,sigstar,tau2,lam_[i],lami,candi);
```

```
j=j+ni; i = i + 1; endo;

/*---------- Store Results ----------*/ bm4[t,.] = b1; lambdam4[.,t]
= lam1;

b0   = bb1; lam0 = lam1;

t = t + 1; endo;

/* --- Kernel estimates of Lambdas  --- */

post5a = zeros(1,n); post5b = zeros(1,n); h      = zeros(1,1); i=1; do
while i le n;
    lamrs = lamstar[i];
    lamrr = lambdam4[i,.]';

    post5a[.,i],post5b[.,i],h = ukernel(lamrs,lamrr,0,1,&k_gauss);
    i=i+1;
    endo;

post5c = ln(post5a); postlams = sumc(post5c');

/* --- Compute Marginal Likelihood --- */

priors = prordlam + prordbeta + prordd + prordsig; posteriors =
postbeta + postd + postsig + postlams;

marglik = lnf + priors - posteriors;

tend = date; totsec = ethsec(tstart,tend); str = etstr(totsec);

output file = q:/dissert/second/simdata/output/nl_marglk.txt reset;

print "----------------------------";
print " beta Priors = " prordbeta;
print " Lamda Priors = " prordlam;
print " D Priors = " prordd;
print " s2 Priors = " prordsig;
print "----------------------------";
print " beta Posterior = " postbeta;
print " Lambda Posterior = " postlams;
print " D Posteriors = " postd;
print " s2 Posterior = " postsig;
print "----------------------------";
print " Log Likelihood =  " lnf;
print " Log Priors     =  " priors;
print " Log Posterior  = " posteriors;
?;
print "AND THE MARGINAL LIKELIHOOD IS...  " marglik;
print "_____________________________________________";
print "Total Sample Time for " m "  draws:  "  str;
output off;



/*================================================*
 ****************** Procedures ******************
 *================================================*/
```

```
proc autocor(vector, totlags); local T, autocov, j, mean, autocor,
vec1, vec2, c0, n; T=rows(vector); mean=meanc(vector);
autocov=zeros(totlags,1); j=1; do while j le totlags;
    vec1=vector[1:T-j,.];
    vec2=vector[j:T-1,.];
    autocov[j,.]=(vec1'vec2)/T;
    j=j+1;
    endo;
c0 = autocov[1,.];
autocor = autocov/c0;
retp(autocor); endp;

proc ylam(vector, bcparam);
local trnsfm; if bcparam==0;
    trnsfm=ln(vector);
    else;
    trnsfm=((vector^bcparam - 1)/bcparam);
endif;
retp(trnsfm);
endp;

proc rnchisq(m);  /* m is an integer */
local v,g; v = m/2;
    if round(v)-v == 0;
    g = -2*sumc(ln(rndu(v,1)));
    else;
    v = (m-1)/2;
    g = -2*sumc(ln(rndu(v,1))) + rndn(1,1)^2;
    endif;

retp(g);
endp;

proc rninvgam(n,d);
    local nn,c,x;
    nn = n*2;
    c = rnchisq(nn);
    x = 2*d/c;
    retp(x);
    endp;

proc summary(dat);
    retp(meanc(dat')^stdc(dat')^median(dat'));
    endp;

proc rnwish(n,V); local p,t,i,j,a,y,b ; p = rows(V); T = zeros(p,p); i
= 1; do while i le p;
    j = 1;
        do while j le i;
        if i == j;
        T[i,j] = sqrt(rnchisq(n-i+1));
        else;
        T[i,j] = rndn(1,1);
        endif;
    j = j + 1;
    endo;
i = i + 1;
endo; A = T*T';
Y = chol(v);
B = Y'A*Y;
```

```
retp(B);
endp;

proc normord(x,mu,V);
    local ord,k,detv,const,kern;
    k=rows(x);
    detv=det(V);
    const=((2*pi)^(-k/2))/sqrt(detv);
    kern=exp(-(x-mu)'*invpd(V)*(x-mu)/2);
    ord=const*kern;
    retp(ord);
endp;

proc lnnormord(x,mu,V);
    local ord,k,detv,const,kern;
    k=rows(x);
    detv=det(V);
    const=-k/2*(ln(2*pi)) - ln(sqrt(detv));
    kern=(-(x-mu)'*invpd(V)*(x-mu)/2);
    ord=const + kern;
    retp(ord);
endp;

proc lnigord(x,a,d);
    local ord,const,kern;
    const = (a*ln(d)) - ln(gamma(a));
    kern=(a+1)*ln(1/x) - (d/x);
    ord=const + kern;
    retp(ord);
endp;

proc lnwishord(W,v,Sinv);
    local k,i,prod,const,kern1,kern2,ord,tr;
    k=rows(W);
    i=1; prod=zeros(k,1);
    do while i le k;
        prod[i] = gamma((v+1-i)/2);
        i=i+1;
        endo;
    prod = sumc(ln(prod));
    const= - (((v*k/2)*ln(2)) + (k*(k-1)/4)*ln(pi) + prod);
    tr=sumc(diag(inv(Sinv)*inv(W)));
    kern1 = (v/2)*ln(det(Sinv)) - ((v+k+1)/2)*ln(det(W));
    kern2 = (-tr/2);
    ord=const + kern1 + kern2;
    retp(ord);
endp;


proc lnlikfnord(y,X,beta,D,sig,lam,ind);
local lami,n,ni,nn,const,kern,stor1,
stor2,stor3,prod,ord,yilam,yi,Xi,j,deter,gam;
    nn=rows(y);
    n=rows(ind);
    stor1=zeros(n,1);
    stor2=zeros(n,1);
    stor3=zeros(n,1);
    const=(-nn/2)*ln(2*pi);

j=1;
```

```
    for i (1,n,1);
        ni=ind[i];
        lami=lam[i];
        yi=y[j:j+ni-1];
        Xi=X[j:j+ni-1,.];
        yilam=ylam(yi,lami);
        gam=(sig*eye(ni)+Xi*D*Xi');
        stor1[i]=(-1/2)*(yilam-Xi*beta)'*invpd(gam)*(yilam-Xi*beta);
        stor2[i]=sumc((lami-1).*ln(yi));
        stor3[i]=(-1/2)*ln(det(gam));
        j=j+ni;
    endfor;

kern =sumc(stor1); prod =sumc(stor2); deter=sumc(stor3);

ord=const+deter+kern+prod;

retp(ord); endp;



proc (2) = newtraph(ni,X,y,b,prlam,s21,tau2);
    local l1,nrloop,converge,term1,term2,term3,t
    erm4,p1,l2,l3,yj,xj,ylm,f1,f2,lamit,t1,t2,t20,t3,t4,lt1;

l1=1; nrloop=1; converge=0; do while converge lt 1;

term1=zeros(ni,1); term2=zeros(ni,1); term3=zeros(ni,1);
term4=zeros(ni,1);

/*----------- Compute derivatives   -------------*/ p1=1; do while p1
le ni;
    l2=l1^2;
    l3=l1^3;
    yj=y[p1];
    xj=X[p1,.];
    ylm=ylam(yj,l1);

    /****** First Derivative ******/

    term1[p1]=(ylm-xj*bi)*(((yj^l1)*ln(yj)/l1)+(1-yj^l1)/l2);
    term2[p1]=((ln(yj)^3) - (ln(yj)^2)*(xj*bi));

    /****** Second Derivative ******/

    term3[p1]=((yj^l1)*ln(yj)/l1-(yj^l1)/l2+1/l2)^2 +
    (ylm-xj*bi)*((yj^l1)*(ln(yj)^2)/l1 - 2*(yj^l1)*(ln(yj))/l2 + 2*(yj^l1-1)/l3);
    term4[p1]=(ln(yj^4)/4 + (ln(yj)-xj*bi)*((ln(yj)^3)/3));

    /****** Sums, etc. ************/

    p1=p1+1;
    endo;

    t1 = -sumc(term1)/s21;
    t2 =  sumc(ln(y));
    t20= -sumc(term2)/s21;
    t3 = -sumc(term3)/s21;
    t4 = -sumc(term4)/s21;
```

```
    /****** Define limits at zero *****/

if l1 eq 0;
    f1=(t20 - prlam/tau2 + t2);
    f2=(t4 - 1/tau2);
    else;
    f1 =(t1 - ((l1-prlam)/tau2) + t2);
    f2 =(t3 - (1/tau2));
    endif;

lamit = l1 - (f1/f2);

print "lamit = " lamit;  */

if lamit gt 45;
    lamit = 1/lamit;
    elseif lamit lt -45;
    lamit = .01;
    endif;

if abs(f1) le .05;
    converge=1; l1=lamit;
    else;
    l1=lamit;
    nrloop = nrloop+1;

        if nrloop gt 2000;
            print "NO CONVERGENCE";
            end;
        else;
        endif;

    endif;
endo; lt1 = -1/f2;

retp(lamit,lt1); endp;

proc mh(x,y,bi,s21,tau2,lam_,lam0,cand);
    local yold,ynew,prod,jac1,jac2,sse0,s
    se1,num,den,test,lam1,uni,term1a,term2a,term1b,term2b;

prod = sumc(ln(y)); yold = ylam(y,lam0); ynew = ylam(y,candi);

term1a = (yold-X*bi)'(yold-X*bi)/s21; term1b = (lami - lam_^2)/tau2;
term2a = (ynew-X*bi)'(ynew-X*bi)/s21; term2b = (cand - lam_^2)/tau2;
sse0 = -(term1a + term1b)/2 ; sse1 = -(term2a + term2b)/2 ;

jac1=prodc(yi^(candi-1)); jac2=prodc(yi^(lami-1));

test = exp(sse1 - sse0)*(jac1/jac2);

if test ge 1; lam1 = candi; else;

    uni=rndu(1,1);
    if uni le test;
    lam1 = candi;
    else;
    lam1 = lami;
    endif;
```

```
endif; retp(lam1); endp;
```

Bibliography

[1] Albert, James and Siddhartha Chib (1996). Computation in Bayesian econometrics: An introduction to Markov chain Monte Carlo. *Advances in Econometrics*, vol. 11, part A, pp. 3–24.

[2] Berger, James O. and Deely, John (1988). A Bayesian approach to ranking and selection of related means with alternatives to Analysis-of-Variance methodology. *Journal of the American Statistical Association*, vol. 83, no. 402, pp. 364–373.

[3] Bindman, Andrew B. (1990). Can physician profiles be trusted? *Journal of the American Medical Association*, vol. 281, no. 22, pp. 2142–2143.

[4] Bowman, Adrian and Adelchi Azzalini (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Applications* Oxford: Clarendon Press.

[5] Box, G. E. P. and D. R. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society B*, pp. 211–264.

[6] Box, G.E.P. and P. Tidwell (1962). Transformation of the independent variables. *Technometrics*, vol. 4, pp. 531–550.

[7] Buchinsky, Moshe (1995). Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987. *Journal of Econometrics*, vol. 65, pp. 109–154.

[8] Carlin, Bradley and Siddhartha Chib (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society B*, vol. 157, pp. 473–484.

[9] Carlin, Bradley and Thomas Louis (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall.

[10] Chib, Siddhartha and Barton Hamilton, Bayesian Analysis of Cross Section and Clustered Data Treatment Models. *Journal of Econometrics*, in press.

[11] Chib, Siddhartha (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1313–1321.

[12] Chib, Siddhartha, and Bradley Carlin (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing*, vol. 9, pp. 17–26.

[13] Chib, Siddartha and Edward Greenberg (1995). Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models. *Journal of Econometrics*, vol. 68, pp. 339–360.

[14] Chib, Siddhartha and Edward Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, vol. 49, no. 4, pp. 327–335.

[15] Chib, Siddhartha and Edward Greenberg (1998). Bayesian analysis of multivariate probit models. *Biometrika*, vol. 85, pp. 347–361.

[16] Chib, Siddhartha, Edward Greenberg and Rainer Winkelmann (1998). Posterior simulation and Bayes factors in panel count data models. *Journal of Econometrics*, vol. 86, pp. 33-54.

[17] Chib, Siddhartha and Ivan Jeliazkov (1999). Marginal likelihood from the Metropolis-Hastings output. Unpublished manuscript.

[18] Craig, B. A., D. G. Fryback, R. Klein, and B. E. Klein (1999). A Bayesian approach to modeling the natural history of a chronic condition from observations with intervention. *Statistics in Medicine*, vol. 18, no. 11, pp. 1355-1371.

[19] Cronin, K. A., L. S. Freedman, R. Lieberman, H. L. Weiss, S. W. Beenken, and G. J. Kelloff (1999). Bayesian monitoring of phase II trials in cancer chemoprevention. *Journal of Clinical Epidemiology*, vol. 52, no. 8, pp. 705-711.

[20] Draper, Norman R. and David R. Cox (1969). On distributions and their transformation to normality. *Journal of the Royal Statistical Society B*, vol. 31, pp. 472–476.

[21] Feinglass, Joe, Gary J. Martin, and Ashish Sen (1991). The financial effect of physician practice style on hospital resource use. *Health Services Research*, vol. 26, no. 2, pp. 183–205.

[22] Ferraris, Victor A., Suellen P. Ferraris, and Amandeep Singh (1998). Operative outcome and hospital cost. *The Journal of Thoracic and Cardiovascular Surgery*, vol. 115, no. 3, pp. 593–603.

[23] Fong, Duncan K. H. (1992). Ranking and estimation of related means in the presence of a covariate—A Bayesian approach. *Journal of the American Statistical Association*, vo. 87, no. 420, pp. 1128–1136.

[24] Gelfand, A. E. and D. K. Dey (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society B*, vol. 56, pp. 501–514.

[25] Gelfand, Alan, Susan Hills, Amy Racine-Poon, and Adrian Smith (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, vol. 85, no. 412, pp. 972–985.

[26] Gelfand, A. and A. Smith (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, ser. B, vol. 56, pp. 398–409.

[27] Gelman A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. Suffolk: Chapman & Hall.

[28] Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741.

[29] Goldstein, Harvey and David J. Spiegelhalter (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society A*, vol. 159, pp. 385–443.

[30] Greater St. Louis Healthcare Alliance (1996). *Comprehensive Hospital Performance Report*, vol. II.

[31] Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, vol. 82, pp. 711–732.

[32] Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, vol. 57, pp. 97–109.

[33] Heckman, James and Solomon Polachek (1974). Empirical evidence on the functional form of the earnings-schooling relationship. *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 350–354.

[34] Hellmich, M., K.R. Abrams, and A. J. Sutton (1999). Bayesian approaches to meta-analysis of ROC curves. *Medical Decision Making*, vol. 19, no. 3, pp. 252-264.

[35] Hofer, Timpthy P., Rodney A. Hayward, Sheldon Greenfield, Edward H. Wagner, Sherrie H. Kaplan, and Willard G. Manning (1999). The unreliability of individual physician "report cards," for assessing the costs and quality of care of a chronic disease. *Journal of the American Medical Association*, vol. 281, no. 22, pp. 2098–2105.

[36] Hollenbeak, Christopher S., Denise M. Murphy, Stephanie Koenig, Robert S. Woodward, William C. Dunagan, and Victoria J. Fraser (2000). The clinical and economic impact of deep chest surgical site infections following coronary artery bypass graft, Forthcoming in *CHEST*.

[37] Hsiao, Cheng (1986). *Analysis of Panel Data*. Cambridge: University Press.

[38] Iezzoni, Lisa, Michael Schwartz, Arlene Ash, John Hughes, Jennifer Daley and Yevgenia Mackiernan (1996). Severity measurement methods and judging hospital death rates for pneumonia. *Medical Care*, vol. 34, no. 1, pp. 11–28.

[39] Jones, H. (1999). Bayesian analysis: An objective, scientific approach to better decisions. *Clinical Laboratory Management Review*, vol. 13, no. 3, pp. 148-153.

[40] Judge, G. P. Hill, W. Griffiths, H. Lütkepohl, and T. Lee (1988). *Introduction to the Theory and Practice of Econometrics (Second Edition)*, Wiley:New York.

[41] Kassirer, Jerome P. (1994). The use and abuse of practice profiles. *The New England Journal of Medicine*, vol. 330, no. 9, pp. 634–635.

[42] Krueger, Deborah and Thomas Davidson (1987). Alternative approaches to cost accounting. *Topics in Healthcare Financing*, vol. 13, no. 4, pp. 1–9.

[43] Kukla, Steven and Henry Bachofer (1983). Building up reporting systems. *Hospitals*, vol. 57, no. 14, pp. 77–80.

[44] Leamer, Edward (1991). A Bayesian perspective on inference from macroeconomic data, in Hylleberg, Svend and Paldam, Martin, eds., *New approaches to empirical macroeconomics*. Oxford and Cambridge: Blackwell.

[45] Lindley, D. and A. Smith (1972). Bayes estimation for the linear model. *Journal of the Royal Statistical Association*, ser. B, no. 34, pp. 1–41.

[46] Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

[47] Mennemeyer, Stephen, Michael Morrisey, and Leslie Howard (1997). Death and reputation: How consumers acted upon HCFA mortality Information. *Inquiry*, vol. 34, September, pp. 117–128.

[48] Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21, pp. 1087–1092.

[49] Micheletti, Julie A., Thomas J. Shlala, and Colin R. Goodall (1998). Evaluating performance outcomes measurement systems: Concerns and considerations. *Journal of Healthcare Quality*, vol. 20, no. 2, pp. 177–83.

[50] Miller, Michael E., Siu H. Liu, William M. Tierney, and Clement J. McDonald (1993). Estimating physician costliness: An empirical Bayes approach. *Medical Care*, vol. 31, no. 5, pp. YS16–YS28.

[51] Mincer, Jacob (1974). *Schooling, Experience, and Earnings*. New York: NBER.

[52] Mukamel, Dana, and Alvin Mushlin (1998). Quality of care information makes a difference: An analysis of market share and price changes after publication of the New York State Cardiac Surgery Mortality Reports. *Medical Care*, vol. 36, no. 7, pp. 945–954.

[53] Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian Inference by the weighted likelihood bootstrap. *Journal of the Royal Statistical Society B*, vol. 56, pp. 3–48.

[54] O'Hagan, Anthony (1994). *Kendall's Advanced Theory of Statistics*, vol. 2B. Wiley & Sons: New York.

[55] Poirier, Dale (1978). The use of the Box-Cox transformation in limited dependent variable models. *Journal of the American Statistical Association*, vol. 73, no. 362, pp. 284–287.

[56] Poirier, Dale (1995). *Intermediate Statistics and Econometrics: A Comparative Approach.* Cambridge: MIT Press.

[57] Ramsey, J. B., and Paul Zarembka (1971). Specification error tests and alternative functional forms of the aggregate production function. *Journal of the American Statistical Association*, vol. 66, pp. 471–477.

[58] Rosenthal, Gary E., Dwain L. Harper, Linda M. Quinn, and Gregory S. Cooper (1997). Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. *Journal of the American Medical Association*, vol. 13, no. 6, pp. 485–490.

[59] Savin, N. E., and K. J. White (1978). Estimation and testing for functional form and autocorrelation: A simultaneous approach. *Journal of Econometrics*, vol. 8, pp. 1-12.

[60] Schnitzler, Mark, Robert Woodward, Linda Mundy, Robert Sutter, Stuart Boxerman and William Dunagan (1997). Impact of risk adjustment estimators on ranking physican costs for pneumonia. *Best Practices and Benchmarking in Healthcare*, March/April, pp. 82–87.

[61] Schwarts, Michael, and Arlene S. Ash (1994). Evaluating the performance of risk-adjustment methods: Continuous measures. in Iezzoni, Lisa, ed., *Risk Adjustment for Measuring Health Care Outcomes.* Ann Arbor: Health Administration Press.

[62] Seaks, Terry G. and Layson, Stephen K. (1983). Box-Cox estimation with standard econometric problems. *Review of Economics and Statistics*, vol. 65, pp. 160–164.

[63] Showalter, Mark (1994). A Monte Carlo study of the Box-Cox model and a non-linear least squares alternative. *Review of Econometrics and Statistics*, vol. 76, no. 3 pp. 560–570.

[64] Silverman, B.W. (1986). *Density estimation for statistics and data analysis.* New York: Chapman and Hall.

[65] Smith, Peter (1990). The use of performance indicators in the public sector. *Journal of the Royal Statistical Society A*, vol 153, pp. 53–72.

[66] Spiegelhalter, D. J., Myles, J. P., Jones, D. R., and Abrams, K. R., (1999). Methods in health service research. An introduction to Bayesian methods in health technology assessment. *British Medical Journal*, vol. 319, no. 7208, pp. 508-512.

[67] Spitzer, John J. (1982). A fast and efficient method for estimation of parameters in models with the Box-Cox transformation. *Journal of the American Statistical Association*, vol. 77, pp. 760–766.

[68] Spitzer, John J. (1982). A primer on Box-Cox estimation. *Review of Economics and Statistics*, vol. 64, pp. 307–313.

[69] Spoeri, Randall K., and Ralph Ullman (1997). Measuring and reporting managed care performance: Lessons learned and new initiatives. *Annals of Internal Medicine*, vol. 127, pp. 726–732.

[70] Thomas, J. William and Marie Ashcroft (1991). Measuring severity of illness: Six severity systems and their ability to explain cost variations. *Inquiry*, vol. 28 pp. 39–55.

[71] Tierney, Luke (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics*, vol. 22, no. 4, pp. 1701–1762.

[72] Wand, M. P. and M. C. Jones (1995). *Kernel Smoothing.* London: Chapman & Hall.

[73] Welch, H. Gilbert, Merk E. Miller, and W. Pete Welch (1994). Physician profiling: An analysis of inpatient practice patterns in Florida and Oregon. *New England Journal of Medicine*, vol. 330, no. 9, pp. 607–611.

[74] Zarambka, Paul (1974). Transformation of variables in econometrics, in *Frontiers in Econometrics.* Boston: Academic Press.