A Flexible Approach to Bayesian Multiple Curve Fitting

Carsten H. Botts
Department of Mathematics and Statistics
Williams College
Williamstown, Massachusetts 01267, USA
cbotts@williams.edu

Michael J. Daniels
Department of Epidemiology and Biostatistics, and
Department of Statistics
University of Florida
Gainesville, Florida 32611, USA
mdaniels@stat.ufl.edu

Summary

We model sparse functional data from multiple subjects with a mixed-effects regression spline. In this model, the expected values for any subject (conditioned on the random effects) can be written as the sum of a population curve and a subject-specific deviate from this population curve. The population curve and the subject-specific deviates are both modeled as free-knot b-splines with k and k' knots located at $\mathbf{t}_k$ and $\mathbf{t}_{k'}$, respectively. To identify the number and location of the "free" knots, we sample from the posterior $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ using reversible jump MCMC methods. Sampling from this posterior distribution is complicated, however, by the flexibility we allow for the model's covariance structure. No restrictions (other than positive definiteness) are placed on the covariance parameters ψ and σ^2 and, as a result, no analytical form for the likelihood $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ exists. In this paper, we consider two approximations to $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ and then sample from the corresponding approximations to $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$. We also sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ which has a likelihood that is available in closed form. While sampling from this larger posterior is less efficient, the resulting marginal distribution of knots is exact and allows us to evaluate the accuracy of each approximation. We then consider a real data set and explore the difference between $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ and the more accurate approximation to $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$.

Keywords: B-splines; Laplace approximation; Reversible jump MCMC; Unit-information prior.

1 Introduction

Social and physical scientists are often interested in how certain variables depend on one another. They often assume that a functional relationship exists between these two variables, and they run experiments or collect samples to learn about this relationship. They are aware, however, that the data does not follow a deterministic equation, but follows the general stochastic equation

$$Y = f(x) + \epsilon, \tag{1}$$

where Y is a dependent variable, x is an independent or controlled variable, and ϵ is random error with mean 0. One of the simplest models relating a response, Y, to a single predictor, x, is the $p^{th}-$order polynomial regression model

$$Y(x) = \beta_0 + \sum_{j=1}^{p} \beta_j x^j + \epsilon.$$

While this model is appealing and can be very useful, it may be inadequate since it assumes that only one polynomial describes the average relationship between Y and x. A more complex model would allow several polynomials to do this. An example of such a model is a piecewise polynomial regression spline. A piecewise polynomial regression spline with k knots divides the domain of $f(x)$ into $k+1$ mutually exclusive regions, and to each region corresponds a unique polynomial describing the average relationship between Y and x. With different polynomials describing different parts of the function, $f(x)$ is not restricted to have the same smoothness throughout its domain. A piecewise polynomial regression spline of order p with k knots at $\mathbf{t}_k = (t_1, t_2, \ldots, t_k)$ is written as

$$f(x) = \beta_0 + \beta_1 x + \sum_{i=1}^{k} (x - t_i)_+^p \beta_{1+i} + \sum_{j=2}^{p} \beta_{k+j} x^j, \tag{2}$$

where $\beta = (\beta_0, \beta_1, \ldots, \beta_{p+k})$ is a fixed set of parameters and $(z)_+^p = z^p \mathbb{I}(z \geq 0)$.

A natural polynomial regression spline takes a form nearly identical to the piecewise regression spline seen above. It is different, however, in that it restricts the function to be linear at the boundaries. This sets the last term on the right-hand side of (2) to 0. A natural polynomial regression spline of order p thus takes the form

$$f(x) = \beta_0 + \beta_1 x + \sum_{i=1}^{k} (x - t_i)_+^p \beta_{1+i}.$$

If $f(x)$ were modeled as such, (1) could then be re-written as

$$Y(x) = \mathbf{c}(x, \mathbf{t}_k) \beta + \epsilon,$$

where $\mathbf{c}\left(x, \mathbf{t}_k\right) = \left(1, x, \left(x - t_1\right)_+^p, \ldots, \left(x - t_k\right)_+^p\right)$ and $\beta = \left(\beta_0, \beta_1, \beta_2, \ldots, \beta_{1+k}\right)$, and for m observations collected at $\left(x_1, x_2, \ldots, x_m\right)$, it would be written as

$$Y\left(\mathbf{x}\right) = \begin{pmatrix} Y\left(x_1\right) \\ Y\left(x_2\right) \\ \vdots \\ Y\left(x_m\right) \end{pmatrix} = \begin{bmatrix} \mathbf{c}(x_1, \mathbf{t}_k) \\ \mathbf{c}(x_2, \mathbf{t}_k) \\ \vdots \\ \mathbf{c}(x_m, \mathbf{t}_k) \end{bmatrix} \beta + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix}. \tag{3}$$

The natural polynomial regression spline written in Equation (3) illustrates the concept behind splines, but the design matrix composed of the row vectors $\mathbf{c}\left(x_1, \mathbf{t}_k\right), \ldots, \mathbf{c}\left(x_m, \mathbf{t}_k\right)$ is unstable and thus rarely used. B-splines, which are computationally more stable, are preferred (Hastie and Tibshirani, 1990; Zhou and Shen, 2001). The functional form of a b-spline is more complex than that of the natural polynomial regression spline, but is readily available (deBoor, 1978). We write it as

$$f(x) = \mathbf{b}(x, \mathbf{t}_k)\alpha,$$

where $\mathbf{b}(x, \mathbf{t}_k)$ is the design vector associated with a b-spline evaluated at x with k knots at $\mathbf{t}_k = \left(t_1, t_2, \ldots, t_k\right)$, and α is a fixed set of parameters. For m observations collected at $\left(x_1, \ldots, x_m\right)$, the model would be written as

$$Y\left(\mathbf{x}\right) = \begin{pmatrix} Y\left(x_1\right) \\ Y\left(x_2\right) \\ \vdots \\ Y\left(x_m\right) \end{pmatrix} = \begin{bmatrix} \mathbf{b}(x_1, \mathbf{t}_k) \\ \mathbf{b}(x_2, \mathbf{t}_k) \\ \vdots \\ \mathbf{b}(x_m, \mathbf{t}_k) \end{bmatrix} \alpha + \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{pmatrix}.$$

With this model, estimating $f(x)$ becomes a problem of estimating the number of knots, k, the locations of those knots, $\mathbf{t}_k$, and α.

A variety of methods have been developed in estimating k and $\mathbf{t}_k$. Halpern (1973) approached this problem using Bayesian methods. Allowing knots to only be placed at the design points in the experiment, he considered all of the subsets of the design points. Halpern assigned prior probabilities to all of these subsets, and calculated the corresponding posterior probabilities of these subsets. His estimator of the function was based on these posterior probabilities. Denison, Mallick, and Smith (1998) placed priors on the number of knots, k, and their locations, $\mathbf{t}_k$. With these priors, they calculated a joint posterior distribution which included the variables k and $\mathbf{t}_k$ and then sampled from this posterior distribution using reversible jump MCMC methods. They too restricted the knots to be located only at the design points of the experiment. DiMatteo, Genovese, and Kass (2001) proposed a method similar to that of Denison, et al. They did not restrict the knots to be located only at the design points of the experiment and they penalized models with unnecessarily large values of k by averaging over the fixed effect coefficients, α.

Friedman (1991) and Stone et al. (1997) try to solve the problem of knot selection using backward and/or forward knot selection. This process continues until the "best" model has been identified. Zhou and Shen (2001) used an alternative method in identifying the locations of the knots. They constructed an algorithm which favored adding knots in locations where several knots had already been added. Lindstrom (1999) used similar methods when selecting knot locations.

Splines have also been used to model curves that vary within and between subjects (Shi et al., 1996, Rice and Wu, 2001, Behseta et al., 2005, Crainiceanu et al., 2005). Brumback and Rice (1998) use splines to model nested functions which vary between subjects and between groups of subjects. Functional models which only account for variation between and within subjects take the general form

$$
\begin{aligned}
Y_i\left(x_j\right) &= f_i(x_j) + \epsilon_{ij} \\
&= f\left(x_j\right) + G_i\left(x_j\right) + \epsilon_{ij},
\end{aligned}
\tag{4}
$$

where $Y_i\left(x_j\right)$ is the observation of the ith individual at x_j, $f\left(x\right)$ can be thought of as a population curve, and $G_i\left(x\right)$ is a random curve specific to subject i. While these functions can be modeled in a variety of ways, we model these two functions as free-knot b-splines with k and k' knots located at $\mathbf{t}_k$ and $\mathbf{t}_{k'}$, respectively. Setting $f(x) = \mathbf{b}(x; \mathbf{t}_k)\alpha$ and $G_i(x) = \mathbf{b}(x; \mathbf{t}_{k'})\gamma_i$, where γ_i are independent random vectors such that $\gamma_i \sim N(0, \Sigma_\gamma)$, equation (4) becomes

$$
Y_i(x_j) = \mathbf{b}(x_j; \mathbf{t}_k)\alpha + \mathbf{b}(x_j; \mathbf{t}_{k'})\gamma_i + \epsilon_{ij}.
\tag{5}
$$

Bigelow and Dunson (2007) consider a variation of this model and identify the location of the knots by sampling from the full posterior distribution. They lessen the computational burden of this sampling procedure by setting $\mathbf{t}_{k'} = \mathbf{t}_k$ and restricting Σ_γ to be diagonal. In this paper, we consider a more flexible model and explore the computational issues associated with it. To be more specific, we eliminate the restrictions imposed by Bigelow and Dunson and then try to identify the number and location of the knots by sampling from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$. This posterior can not be sampled from exactly, however, because eliminating the diagonal restriction on Σ_γ makes the likelihood $p\left(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'}\right)$ intractable. We thus sample from two approximations to this posterior; each approximation corresponds to a different approximation to the likelihood $p\left(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'}\right)$. To assess the accuracy of these approximations, we compare the posterior distribution of knots in each case to the marginal distribution of knots when sampling from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$, where ψ is a set of covariance parameters and σ^2 is the within-subject variability. The marginal distribution of knots when sampling from this higher-dimensional posterior is exact as its corresponding likelihood is avail-

able in closed form. Sampling from this larger posterior, however, is less efficient because of the additional parameters that need to be sampled.

In Section 2, we discuss the linear model given in (5) in more detail. In Section 3, the likelihood of interest, $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})$, is discussed as are the two methods we use to approximate this likelihood. Section 4 shows the algorithm used to sample from the resulting approximate posterior distributions, and Section 5 shows the algorithm we use to sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2 | y)$. Section 6 theoretically compares the approximations, and in Section 7, we describe a simulation study performed to explore the differences in these approximations. In Section 8, we apply the preferred approximation to a real data set.

2   The model and priors

Let $Y_i(x_j)$ be the observed value of the ith curve $(i = 1, \ldots, n)$ at x_j. We specify the following mixed model for functional data,

$$Y_i(x_j) = \mathbf{b}(x_j, \mathbf{t}_k)\alpha + \mathbf{b}(x_j, \mathbf{t}_{k'})\gamma_i + \epsilon_{ij}, \tag{6}$$

where $\mathbf{b}(x_j, \mathbf{t}_q)$ is the design vector of a b-spline with q knots at $\mathbf{t}_q$, k is the number of fixed effect knots at locations $\mathbf{t}_k$, k' is the number of random effect knots at $\mathbf{t}_{k'}$, α is a fixed set of parameters with $\dim(\alpha) = k+2$, γ_i are independent and identically distributed random vectors such that $\gamma_i \sim N(0, \Sigma_\gamma)$ and $\dim(\gamma_i) = k'+2$, ϵ_{ij} are independent random variables with $\epsilon_{ij} \sim N(0, \sigma^2)$ for all pairs (i, j), and the random vectors γ_i are independent of all random variables ϵ_{ij}. The appealing feature of this model is its flexibility; recall that no restriction is being placed on Σ_γ (other than $|\Sigma_\gamma| > 0$), and that the random and fixed effect knots are not set equal to one another. We do assume, however, that all individuals have the same random effect knots, and that $k' \leq k$. When the number of observations per subject is small, such assumptions are needed, as they limit the number of complicated models proposed in the Markov chain we consider.

We place prior distributions on k, k', $\mathbf{t}_k$, $\mathbf{t}_{k'}$, α, Σ_γ, and σ^2. Poisson priors are given to k and k'. Additionally, since there is no reason to favor knots at any particular location on the domain of f and G, flat priors are placed on both $\mathbf{t}_k$ and $\mathbf{t}_{k'}$. These priors are written as

$$k \sim \text{Poi}(\mu_k), \qquad\qquad p(\mathbf{t}_k | k) \propto \mathbb{I}\left(a < t_{k(1)} < t_{k(2)} < \cdots < t_{k(k)} < b\right),$$

$$k' | k \sim \text{Poi}(\mu_{k'})\,\mathbb{I}(k' \leq k), \qquad p(\mathbf{t}_{k'} | k') \propto \mathbb{I}\left(a < t_{k'(1)} < t_{k'(2)} < \cdots < t_{k'(k')} < b\right),$$

where $t_{q(j)}$ is the jth smallest knot in the vector $\mathbf{t}_q$, and (a, b) is the domain of f and G.

For Σ_γ and σ^2, we place prior distributions on transformed values of these parameters. The transformations we consider create a set of unconstrained parameters which are preferable to work with since their sampling distributions can be well approximated with the normal density. For Σ_γ, we applied the modified Cholesky parameterization proposed by Pourhamadi (1999). Pourhamadi decomposes the inverse of the covariance matrix Σ_γ as $\Sigma_\gamma^{-1} = TDT'$, where

$$
T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\phi_{2,1} & 1 & 0 & \cdots & 0 \\ -\phi_{3,1} & -\phi_{3,2} & 1 & \cdots & 0 \\ \vdots & & & \ddots & \vdots \\ -\phi_{p,1} & -\phi_{p,2} & \cdots & -\phi_{p,p-1} & 1 \end{bmatrix}, \text{ and } D = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & & \ddots & \cdots \\ 0 & \cdots & 0 & \sigma_p^2 \end{bmatrix}.
$$

These parameters can be calculated directly from the elements of the covariance matrix using the functions $g(\cdot; \cdot)$ and $h(\cdot; \cdot)$, where $\sigma_j^2 = g(\Sigma_\gamma; j) = \Sigma_\gamma[j, j] - \sigma_j' \Sigma_{\gamma,j}^{-1} \sigma_j$ for $2 \leq j \leq p$ ($\sigma_1^2 = \Sigma_\gamma[1, 1]$), $\phi_j = (\phi_{j,1}, \phi_{j,2}, \ldots, \phi_{j,j-1})' = h(\Sigma_\gamma; j) = \Sigma_{\gamma,j}^{-1} \sigma_j$, $\Sigma_{\gamma,j}$ is the minor matrix within Σ_γ composed of its first $j - 1$ columns and rows, and σ_j is the vector within Σ_γ composed of the first $j - 1$ elements of the jth column. To ease our notational burden, we will refer to these modified Cholesky parameters collectively as ψ, where $\psi = \left(\phi_{2,1}, \log(\sigma_1^2), \phi_{3,1}, \phi_{3,2}, \log(\sigma_2^2), \ldots, \log(\sigma_p^2)\right)$, and we will replace the symbol Σ_γ with ψ (or, to illustrate that Σ_γ is a function of ψ, with $\Sigma_\gamma(\psi)$). All of the elements in the vector ψ are unconstrained, and a multivariate normal unit-information prior is assigned to them. The mean of the unit-information prior on ψ is the maximum likelihood estimate of the parameters, $\hat{\psi}$, and the variance is inversely proportional to approximately one unit of information on ψ. The prior can be written as

$$
\psi \,|\, \mathbf{t}_k, \mathbf{t}_{k'} \sim MVN\left(\hat{\psi}, -\left[\frac{1}{n - (k+2)} I_\psi\right]^{-1}\right),
$$

where

$$
I_\psi = \left.\frac{\partial^2 \log\left\{p\left(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \alpha, \psi, \sigma^2\right)\right\}}{\partial \psi' \partial \psi}\right|_{\alpha=\hat{\alpha}, \psi=\hat{\psi}, \sigma^2=\hat{\sigma}^2},
$$

$p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \alpha, \psi, \sigma^2)$ is the distribution of all the observed data in the study, n is the number of subjects in the study, and $\hat{\alpha}, \hat{\psi},$ and $\hat{\sigma}$ are the maximum likelihood estimates of $\alpha, \psi,$ and σ, respectively. The exact formula for I_ψ is given in the Appendix.

For σ^2, a unit-information prior was placed on $\log(\sigma^2)$. The prior can be written as

$$
\log(\sigma^2) \,\big|\, \mathbf{t}_k, \mathbf{t}_{k'} \sim N\left(\log(\hat{\sigma}^2), -\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i - \frac{k+2}{n}} I_{\log(\sigma^2),i}\right)^{-1}\right),
$$

where

$$
I_{\log(\sigma^2),i} = \left.\frac{\partial^2 \log\left\{p(y_i|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \alpha, \psi, \sigma^2)\right\}}{\partial\left(\log(\sigma^2)\right)^2}\right|_{\alpha=\hat{\alpha}, \psi=\hat{\psi}, \sigma=\hat{\sigma}},
$$

6

and m_i is the number of observations for subject i. The exact formula for $I_{\log(\sigma^2),i}$ is given in the Appendix.

The prior placed on α is another unit-information prior and can be written as

$$\alpha|\mathbf{t}_k, \mathbf{t}_{k'} \sim MVN\left(\hat{\alpha}, -\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m_i}I_{\alpha,i}\right)^{-1}\right),$$

where

$$I_{\alpha,i} = \left.\frac{\partial^2 \log\left\{p(y_i|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \alpha, \psi, \sigma^2)\right\}}{\partial\alpha'\partial\alpha}\right|_{\alpha=\hat{\alpha},\psi=\hat{\psi},\sigma^2=\hat{\sigma}^2}.$$

The exact formula for $I_{\alpha,i}$ is given in the Appendix.

Kass and Wasserman (1994) and Pauler (1998) defend the use of unit-information priors, especially in the context of model selection. They show that when using such priors, the resulting Bayes factors can be well approximated with the Bayesian information criterion. Although centering unit-information priors at the maximum likelihood estimates has been suggested in Bayesian adaptive regression splines (Paciorek, 2006), the final results presented in this paper were not sensitive to the means selected for these three prior distributions.

Estimating the fixed and random effect functions now becomes a problem in estimating k, $\mathbf{t}_k$, k', and $\mathbf{t}_{k'}$. We do this by sampling from the posterior distribution $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$. While this procedure is intuitively appealing, a problem arises since the likelihood $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ can not be calculated analytically. We consider two approximations to this likelihood in the next section.

3 Approximating $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})$

The likelihood of interest can be expressed as

$$
\begin{aligned}
p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'}) &= \int_{\sigma^2}\int_{\psi}\int_{\alpha}\prod_{i=1}^{n}\left[\int_{\gamma_i} p(y_i|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \alpha, \gamma_i, \sigma^2)\,p(\gamma_i|\psi)d\gamma_i\right]\\
&\quad \times p(\alpha|\mathbf{t}_k, \mathbf{t}_{k'})p(\psi|\mathbf{t}_k, \mathbf{t}_{k'})\,p(\sigma^2|\mathbf{t}_k, \mathbf{t}_{k'})\,d\alpha d\psi d\sigma^2\\
&= \int_{\sigma^2}\int_{\psi} p(y_i|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2)\,p(\psi|k', \mathbf{t}_{k'})\,p(\sigma^2|\mathbf{t}_k, \mathbf{t}_{k'})\,d\psi d\sigma^2, \quad\quad (7)
\end{aligned}
$$

where

$$p\left(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'},\psi,\sigma^2\right) = \left|\mathbf{A}^{-1}\right|^{\frac{1}{2}}\prod_i (2\pi)^{-\frac{m_i}{2}}\left|\mathbf{B}_{R_i}\Sigma_\gamma(\psi)\mathbf{B}'_{R_i}+\sigma^2\mathbf{I}_{m_i}\right|^{-\frac{1}{2}}\left|\Sigma_\alpha\right|^{-\frac{1}{2}}\times$$

$$\exp\left(-\frac{1}{2}\left(\sum_{i=1}^n y'_i\left(\mathbf{B}_{R_i}\Sigma_\gamma(\psi)\mathbf{B}'_{R_i}+\sigma^2\mathbf{I}_{m_i}\right)^{-1}y_i+\hat{\alpha}\Sigma_\alpha^{-1}\hat{\alpha}-\mathbf{d}'\mathbf{Ad}\right)\right),\quad (8)$$

$$\mathbf{d} = \mathbf{A}^{-1}\sum_{i=1}^n\mathbf{B}'_{F_i}\left(\mathbf{B}_{R_i}\Sigma_\gamma(\psi)\mathbf{B}'_{R_i}+\sigma^2\mathbf{I}_{m_i}\right)^{-1}y_i+\Sigma_\alpha^{-1}\hat{\alpha}$$

$$\mathbf{A} = \sum_{i=1}^n\mathbf{B}'_{F_i}\left(\mathbf{B}_{R_i}\Sigma_\gamma(\psi)\mathbf{B}'_{R_i}+\sigma^2\mathbf{I}_{m_i}\right)^{-1}\mathbf{B}_{F_i}+\Sigma_\alpha^{-1},$$

$$\mathbf{B}_{F_i} = \left(\mathbf{b}\left(x_{1_i};\mathbf{t}_k\right)',\mathbf{b}\left(x_{2_i};\mathbf{t}_k\right)',\ldots,\mathbf{b}\left(x_{m_i};\mathbf{t}_k\right)'\right)',$$

$$\mathbf{B}_{R_i} = \left(\mathbf{b}\left(x_{1_i};\mathbf{t}_{k'}\right)',\mathbf{b}\left(x_{2_i};\mathbf{t}_{k'}\right)',\ldots,\mathbf{b}\left(x_{m_i};\mathbf{t}_{k'}\right)'\right)',$$

$\mathbf{x}_i=(x_{1_i},x_{2_i},\ldots,x_{m_i})$ is the vector containing the $x-$values corresponding to subject i, n is the total number of subjects (curves) in the study, and m_i is the total number of observations for subject i.

The integral in (7) can not be calculated analytically. We thus explore two different approximations to $p\left(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'}\right)$. The first simply plugs in the maximum likelihood estimate of ψ and σ^2. The second approximates (7) using a Laplace approximation.

The first approximation that we consider estimates $p\left(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'}\right)$ with

$$p\left(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'},\hat{\psi}\left(\mathbf{t}_k,\mathbf{t}_{k'}\right),\hat{\sigma}^2\left(\mathbf{t}_k,\mathbf{t}_{k'}\right)\right),$$

where $\hat{\psi}\left(\mathbf{t}_k,\mathbf{t}_{k'}\right)$ and $\hat{\sigma}^2\left(\mathbf{t}_k,\mathbf{t}_{k'}\right)$ are the maximum likelihood estimators of ψ and σ^2 corresponding to the model with fixed effect knots at $\mathbf{t}_k$ and random effect knots at $\mathbf{t}_{k'}$. The dependence of these maximum likelihood estimators on $(\mathbf{t}_k,\mathbf{t}_{k'})$ will be suppressed, and the estimators will now simply be denoted as $\hat{\psi}$ and $\hat{\sigma}^2$. We refer to this likelihood approximation as the "plugged-in" approximation and denote it as $\hat{p}_{\text{Plugged In}}(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'})$. Although this approximation ignores the penalty from increasing the dimension of ψ, it is computationally inexpensive and is similar to the methods employed by Taplin (1993), Draper (1995), Raftery et al. (1996), and Dominici et al. (2002). They approximate marginal likelihoods by plugging in the maximum likelihood estimates of the nuisance parameters that can not be integrated out.

An alternative, but computationally more complex approach, is to integrate out ψ and σ^2 using an approximation method. We use a Laplace approximation to estimate $p\left(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'}\right)$. The approximate likelihood is given by

$$\hat{p}_{\text{Laplace}}\left(y|k,\mathbf{t}_k,k',\mathbf{t}_{k'}\right)=p(y|k,\mathbf{t}_k,k',\mathbf{t}'_k,\hat{\psi},\hat{\sigma}^2)(2\pi)^{\frac{\dim(\hat{\psi})+1}{2}}\left|\widehat{\text{Cov}}\left(\hat{\psi},\log\left(\hat{\sigma}^2\right)\right)\right|^{\frac{1}{2}},$$

where

$$\widehat{\text{Cov}}\left(\hat{\psi}, \log\left(\hat{\sigma}^2\right)\right) = - \left(\frac{\partial^2 \log\left\{p(y|k, \mathbf{t}_k, k', \mathbf{t}_k, \alpha, \psi, \sigma^2)p(\alpha|\mathbf{t}_k)p(\psi|\mathbf{t}_k, \mathbf{t}_{k'})p(\sigma^2|\mathbf{t}_k, \mathbf{t}_{k'})\right\}}{\partial\left(\psi, \log(\sigma^2)\right)' \partial\left(\psi, \log(\sigma^2)\right)}\right)^{-1}\Bigg|_{\alpha=\hat{\alpha}, \psi=\hat{\psi}, \ \sigma^2=\hat{\sigma}^2},$$

and $\dim(\psi) = (k'+3)(k'+2)/2$. We chose the unconstrained parameterization for Σ_γ, ψ, and σ^2, $\log(\sigma^2)$, to make this Laplace approximation more accurate. The optimal parameterization is an area of future research. Calculating $\widehat{\text{Cov}}\left(\hat{\psi}, \log\left(\hat{\sigma}^2\right)\right)$ makes the computational expense of the Laplace approximation greater than that of the "plugged-in" approximation; the matrix of second derivatives takes time to compute and can be numerically unstable when $\dim(\psi)$ is large and the total number of subjects, n, is small. We still expect, however, that the Laplace approximation will be more accurate than the "plugged-in" approximation. A major question of interest addressed in this paper is whether the Laplace's gain in accuracy is worth its computational cost.

4 Sampling from $\hat{p}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$

The two likelihood approximations induce the following approximations to the posterior $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$. Let

$$\hat{p}_{\text{Plugged In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y) \propto \hat{p}_{\text{Plugged In}}(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})p(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$$

and

$$\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y) \propto \hat{p}_{\text{Laplace}}(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}).$$

To sample from these posterior approximations to $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ (which we denote as $\hat{p}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$), we use reversible jump MCMC methods. Reversible jump MCMC methods (Green, 1995; DiMatteo et al., 2001) can be used when sampling from a distribution of a random variable θ and $\dim(\theta)$. In this particular case, the dimensions of the vectors $\mathbf{t}_k$ and $\mathbf{t}_{k'}$ vary with the values of k and k'.

Rather than sampling the fixed and random effect knots together in one MCMC iteration (which results in very low acceptance rates), we sample from the posterior $\hat{p}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ using a Gibbs Sampling type algorithm. In the algorithm, one value of the pair $(k, \mathbf{t}_k)$ is sampled from $\hat{p}(k, \mathbf{t}_k|k', \mathbf{t}_{k'}, y)$ using the Metropolis-Hastings algorithm. Conditioned on this sampled pair of $(k, \mathbf{t}_k)$, a pair $(k', \mathbf{t}_{k'})$ is sampled from $\hat{p}(k', \mathbf{t}_{k'}|k, \mathbf{t}_k, y)$ using identical Metropolis-Hastings methods. This algorithm is given below, and will be referred to as Algorithm I.

Algorithm I.

Repeat the Following Steps:

1. Call the current set of knots $\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)$

2. Sample from $\hat{p}\left(k, \mathbf{t}_k | k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, y\right)$

   (a) Propose a move for the fixed effect knots. Accept with probability
   $$p_{\text{Fixed}} =$$
   $$\min\left(1, \frac{P_{\text{Jump}}\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} \longrightarrow k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} | k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)}{P_{\text{Jump}}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \longrightarrow k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} | k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)} \frac{\hat{p}\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} | k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, y\right)}{\hat{p}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} | k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, y\right)}\right),$$
   where $P_{\text{Jump}}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \longrightarrow k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} | k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)$ is the probability of jumping
   from $(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}})$ to $(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}})$ conditioned on the current value of $(k', \mathbf{t}_{k'})$.

   (b) If not accepted, change $(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}})$ to the older values $(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}})$.

3. Sample from $\hat{p}\left(k', \mathbf{t}_{k'} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)$

   (a) Propose a move for the random effect knots. Accept with probability
   $$p_{\text{Random}} =$$
   $$\min\left(1, \frac{P_{\text{Jump}}\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \longrightarrow k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \,\middle|\, k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}\right)}{P_{\text{Jump}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \longrightarrow k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \,\middle|\, k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}\right)} \frac{\hat{p}\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \,\middle|\, k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}{\hat{p}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \,\middle|\, k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}\right),$$
   where $P_{\text{Jump}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \longrightarrow k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \,\middle|\, k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}\right)$ is the probability of jumping
   from $\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)$ to $\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right)$ conditioned on the current value of $(k, \mathbf{t}_k)$.

   (b) If not accepted, change the value of $\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right)$ to the older pair
   $\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)$.

4. Identify the new set of knots $\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right)$ as $\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)$.

The jump probabilities and details of this algorithm (including derivation of the acceptance weights) are given in the Appendix.

5 Sampling from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$

An alternative way of sampling from the posterior distribution of $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ is to sample from the joint posterior distribution which contains both the covariance parameters, (ψ and σ^2), and $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$. This posterior is written as $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. Although sampling from this posterior is expected to be less efficient than sampling from the approximations to $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ (ψ and σ^2 have to be sampled in addition to $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$), the posterior sample will be exact since $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2)$ can be calculated (see (8)). The resulting posterior sample of knots will also allow us to evaluate the accuracy of the two approximations proposed in Section 4. The accuracy of the approximations will be judged by comparing the distribution of $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ when sampling from $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $\hat{p}_{\text{Plugged-In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ to the marginal distribution of $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ when sampling from the "true" posterior $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. The algorithm used to sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ is given below, and will be referred to as Algorithm II.

Algorithm II.

Repeat the Following Steps:

1. Call the current set of knots and the current value of ψ and σ^2 $\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}},\right)$

2. Sample from $p\left(k, \mathbf{t}_k, |k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}, y\right)$

   (a) Propose a move for the fixed effect knots. Accept with probability

   $p_{\text{Fixed}} =$

   $$\min\left(1, \frac{P_{\text{Jump}}\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} \longrightarrow k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \,|\, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}\right)}{P_{\text{Jump}}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \longrightarrow k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} \,|\, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}\right)} \frac{p\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} \,|\, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}, y\right)}{p\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \,|\, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}, y\right)}\right),$$

   where $P_{\text{Jump}}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \longrightarrow k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} \,|\, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}\right)$ is the probability of jumping from $(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}})$ to $(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}})$ conditioned on the current value of $\left(k', \mathbf{t}_{k'}, \psi, \sigma^2\right)$.

   (b) If not accepted, change $(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}})$ to the older values $(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}})$.

3. Sample from $p\left(k', \mathbf{t}_{k'}, \psi, \sigma^2|k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)$

   (a) Propose a move for the random effect knots, ψ, and σ^2. Accept with probability

   p_{Random}, where $p_{\text{Random}} =$

11

$$\min\left(1, \frac{P_{\text{Jump}}\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}} \longrightarrow k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}} \Big| k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}\right)}{P_{\text{Jump}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}} \longrightarrow k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}} \Big| k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}\right)} \frac{p\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}} \Big| k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}{p\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}} \Big| k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}\right),$$

and $P_{\text{Jump}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}} \longrightarrow k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}} \Big| k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}\right)$ is the probability of jumping from $\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}\right)$ to $\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}}\right)$ conditioned on the current value of $(k, \mathbf{t}_k)$.

(b) If not accepted, change the value of $\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}}\right)$ to the older pair $\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}\right)$.

4. Identify the new set $\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}}\right)$ as $\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}}\right)$.

The jump probabilities and details of this algorithm are given in the Appendix.

Note that Algorithm II is nearly identical to Algorithm I with the exception of Step 3. In Step 3 of Algorithm II, new values of ψ and σ^2 are proposed in addition to new values of k' and $\mathbf{t}_{k'}$. To be more specific, the candidates of ψ and σ^2 (conditioned on the candidates of k' and $\mathbf{t}_{k'}$) are proposed from a multivariate normal distribution with covariance equaling $\widehat{\text{Cov}}\left(\hat{\psi}, \log(\hat{\sigma}^2)\right)$. Calculating this estimated covariance matrix in Step 3 makes this algorithm's computation time per iteration comparable to that of Algorithm I when $\hat{p}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y) = \hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$. Although the computational cost of these two algorithms are similar, we expect Algorithm II to mix slower and have a lower acceptance rate, as it proposes higher dimensional moves in Step 3.

6 A Theoretical Comparison of the Approximations and Algorithms

To sample from the posterior $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$, three options have been proposed: (1) sample from $\hat{p}_{\text{Plg}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ using Algorithm I, (2) sample from $\hat{p}_{\text{Lp}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ using Algorithm I, and (3) exactly sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ using Algorithm II. We refer to options (1), (2), and (3) as PLG, LP, and EX, respectively. In this section, we theoretically justify why LP is preferred to PLG and EX.

While LP will take longer to implement than PLG (as calculating $\hat{p}_{\text{Lp}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ will take longer than calculating $\hat{p}_{\text{Plg}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$), it is expected that PLG will result in unnecessarily large values of k'. We anticipate this positive bias in the number of random effect knots since the plugged-in approximation to $p(y|k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ does not average over the random effect covariance parameters, ψ; it simply plugs in the

maximum likelihood estimate of ψ. The tendency of PLG to estimate more random effect knots than LP becomes clear after close examination of the ratio

$$\frac{\hat{p}(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}{\hat{p}(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}, \tag{9}$$

calculated in step 3(a) of Algorithm I. In PLG, this ratio is calculated as $\frac{\hat{p}_{\text{Plg}}(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}{\hat{p}_{\text{Plg}}(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}$, and in LP, this ratio is calculated as $\frac{\hat{p}_{\text{Lp}}(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}{\hat{p}_{\text{Lp}}(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}$. In general, we expect

$$\frac{\hat{p}_{\text{Lp}}(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}{\hat{p}_{\text{Lp}}(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)} \leq \frac{\hat{p}_{\text{Plg}}(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)}{\hat{p}_{\text{Plg}}(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y)} \tag{10}$$

when $k'_{\text{new}} > k'_{\text{old}}$. If this inequality holds, then it does follow that PLG will over estimate the number of random effect knots. In Theorem 1, it is proven that when $\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \right)$ is the true set of knots, the inequality given in (10) asymptotically holds.

Theorem 1. *Let the true values of $k, \mathbf{t}_k, k'$ and $\mathbf{t}_{k'}$ be k_{new}, $\mathbf{t}_{k_{\text{new}}}$, k'_{old}, and $\mathbf{t}_{k'_{\text{old}}}$, respectively. If the four conditions below are met, then as n (the total number of subjects) goes to ∞,*

$$\frac{\hat{p}_{\text{Lp}} \left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y \right)}{\hat{p}_{\text{Lp}} \left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y \right)} \leq \frac{\hat{p}_{\text{Plg}} \left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y \right)}{\hat{p}_{\text{Plg}} \left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y \right)}.$$

1. $\mathbf{t}_{k'_{\text{old}}} \subset \mathbf{t}_{k'_{\text{new}}}$ with $k'_{\text{new}} > k'_{\text{old}}$.

2. m_i, the number of observations recorded for individual i, is random and the random variables $m_1, m_2, \ldots, m_n$ are i.i.d. for all n.

3. $\sum_{i=1}^{n} \left(m_i - (k'_{\text{new}} + 2) \right) > k_{\text{new}} + 2$

4. $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ are i.i.d. vectors, where $\mathbf{x}_j = \left(\mathbf{x}_{j_1}, \mathbf{x}_{j_2}, \ldots, \mathbf{x}_{j_{m_j}} \right)$ are the m_j points in time at which subject j is observed.

The proof of this theorem is given in the Appendix.

LP is not only preferred to PLG, but also to EX. The Markov chain resulting from LP converges to the stationary distribution more quickly than that corresponding to EX. This is the case because, in LP, transitions are only made in the space of the fixed and random effect knots. In EX, transitions are made in the space of the fixed knots, the random effect knots, and the covariance parameters. The acceptance probability of the chain corresponding to LP is thus expected to be larger than that corresponding to EX.

While the arguments given in this section suggest that LP is preferred to PLG and EX, it is still of interest to see how these options vary in practice. Of particular interest is to see the extent of the inequality given in (10). If the bias in the number of random effect knots fitted by PLG is only marginal, it might be the preferred option (since it is the easiest approximation to compute). To compare these three options, we conducted a simulation study. This is the subject of Section 7

7   Simulation studies

The accuracies of the posterior distributions constructed using the two likelihood approximations were determined using simulated data sets. Five data sets were simulated from two models at two different sample sizes. Each model had the general form given in (5), but different values of k, $\mathbf{t}_k$, k', $\mathbf{t}_{k'}$ and α. The specific values of these parameters are given in Table (1).

Table 1: True Models Simulated From

	Model 1	Model 2
$\mathbf{t}_k$	$(.35, 2.3, 2.4, 3)$	$(.2, .58, 1, 1.7, 2.8)$
$\mathbf{t}_{k'}$	(2.4)	$(.8, 2.58)$
α	$(5, \text{-}5.5, 1.4, \text{-}6.2, 1, 8)$	$(5, 10, 4, 6, 14.2, 10, 7)$
Σ_γ	$5 \times \mathbf{I}_{3\times 3}$	$5 \times \mathbf{I}_{4\times 4}$
σ^2	$.5$	$.5$

The population curve and five randomly sampled subject specific curves corresponding to each model are given in Figure 1. The simulation experiments were performed at two different values of n ($n = 25$ and $n = 40$, where n is the number of subjects). The subjects in all simulations each had 20 observations, 5 of which were randomly selected ($m_i = 5$ for all i). With such a small sample size per subject, it was not feasible to smooth observations separately for each individual as has been done in previous work (Behseta et al., 2005).

For each simulated data set, we sampled from $\hat{p}_{\text{Plugged In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ using Algorithm I, and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ using Algorithm II. We refer to the posterior sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ as the "truth" or the "true" posterior sample. The number of iterations for each chain of $\hat{p}_{\text{Plugged In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ was at least 10,000, and for $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ the chains were twice as long. All chains were run using `R 2.5.0`, and the average time it took to complete 1000 iterations at $n = 25$ was 18 minutes for $\hat{p}_{\text{Plugged In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$, and 54 minutes for $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. At $n = 40$ the average time it took to complete 1000 iterations was 30 minutes

for $\hat{p}_{\text{Plugged In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$, and 78 minutes for $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. Convergence to a stationary distribution did not seem to be an issue with any of the chains run. The distributions of k and k' corresponding to each approximation and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ are shown in Figures (2) - (5). The captions of each figure also give the average acceptance rates for the plugged-in, Laplace, and "true" chains. The acceptance rate of the Laplace chains are, in each case, higher than those for the "true" or plugged-in chains. Figure 6 gives the posterior distributions of $\mathbf{t}_k$ and $\mathbf{t}_{k'}$ for data simulated from model 1 at $n = 25$. The accuracy of the Laplace and plugged-in as shown in Figure 6 is similar to the accuracy seen in data simulated from model 1 at $n = 25$ and from data simulated from model 2 at $n = 25$ and $n = 40$ (not shown).

The marginal distributions of k' show that the Laplace method is more accurate than the plugged-in method. In each case, the marginal distributions of k' for the Laplace method essentially matches the "true" marginal distributions of k'. The plugged-in estimator greatly overestimates the true number of random effect knots (see Figures 2 - 5). This is to be expected since both ψ and σ^2 were not averaged over (averaging over the parameters would penalize vectors of ψ with large dimensions). The marginal distributions of $\mathbf{t}_{k'}$ in Figure 6 also show that the Laplace method is more accurate than the plugged-in method. The posterior distributions of $\mathbf{t}_{k'}$ corresponding to $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ are essentially the same. The posterior distribution of $\mathbf{t}_{k'}$ corresponding to $\hat{p}_{\text{Plugged-In}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ is far from the true posterior distribution of $\mathbf{t}_{k'}$; the plugged-in method fits far too many random effect knots and thus can not correctly identify their true locations. Thus, in terms of accuracy, the Laplace approximation is definitely preferred to the plugged-in approximation.

The simulation results also suggest that sampling from $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ is preferred to sampling from the exact posterior $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. The distribution of knots corresponding to both posteriors are very similar and the time per iteration is essentially the same. However, sampling from the exact posterior requires that the nuisance parameters ψ and σ^2 be sampled jointly with k' and $\mathbf{t}_{k'}$ in Step 3 of Algorithm II. The acceptance rate of Algorithm II is thus less than that of Algorithm I when sampling from $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$. The lower acceptance rate of Algorithm II also affected the mixing of the chain. Figure 7 gives the trace plots of $\mathbf{t}_{k'}$ for the Laplace and true chains (for iterations 2000-3000) on one (representative) simulated data set. From the plots in Figure 7, it is clear that Algorithm I (applied to the Laplace approximation) explores the posterior distribution of $\mathbf{t}_{k'}$ more efficiently than Algorithm II does.

8 Application

To gain further insight into how the Laplace approximation compares to the "true" posterior, we considered a real data set and sampled from $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. The data set measures the protein content of milk collected from cows who were on a barley diet (Diggle, 1990). Twenty-five cows were in the experiment (n = 25), and milk was collected weekly from each cow for 19 weeks. Five observations from each cow were randomly selected and the reversible jump MCMC algorithms were run on the resulting data set. A plot of the data for all cows is given in Figure 8. Just as with the simulated data sets, 10,000 iterations of Algorithm I were used to sample from $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and 20,000 iterations of Algorithm II were used to sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. The posterior distributions of k and k' corresponding to the $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ and $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$ are given in Figure 9. The posterior distributions of $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ corresponding to $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$ once again matches the marginal distribution of $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$ when exactly sampling from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$.

In addition to examining the two posterior distributions of knots, we also examine the two corresponding posterior distributions of one randomly selected cow's curve. Recall that the curve of one cow is the sum of the fixed curve and that cow's random effect curve. To sample one cow's curve from the chain corresponding to $\hat{p}_{\text{Laplace}}(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$, values of ψ and σ^2 were sampled using an independence sampler. Conditioned on these sampled values, a fixed effect curve was sampled, and conditioned on this sampled fixed effect curve, a random effect curve was sampled. The algorithm used to do this requires a slight addition to Algorithm I. This addition is noted below:

For each sampled set of knots, $(k, \mathbf{t}_k, k', \mathbf{t}_{k'})$, do the following:

1. For $t = 1 : L$ (we set $L = 100$)

    (a) Sample the modified Cholesky parameters ψ and $\log\left(\sigma^2\right)$ from

$$
\left(\begin{array}{c} \psi^t \\ \log\left(\sigma^{2,t}\right) \end{array} \right) \sim MVN \left(\left(\begin{array}{c} \hat{\psi} \\ \log\left(\hat{\sigma}^2\right) \end{array} \right), -I\left(\hat{\psi}, \log\left(\hat{\sigma}^2\right)\right)^{-1} \right),
$$

    where $\left(\hat{\psi}, \hat{\sigma}^2\right)$ are the maximum likelihood estimates of the Pourhamadi parameters and σ^2 corresponding to the current set of knots, and $I\left(\hat{\psi}, \log\left(\hat{\sigma}^2\right)\right)$ is their information matrix evaluated at these estimates.

(b) Accept this draw with probability

$$\min\left(1, \frac{\frac{p\left(y|k,\mathbf{t}_k,k',\mathbf{t}_k,\psi^t,\sigma^{2,t}\right)p\left(\psi^t|k',\mathbf{t}_{k'}\right)p\left(\sigma^{2,t}\right)}{\xi(\psi^t,\log(\sigma^{2,t}))}}{\frac{p(y|k,\mathbf{t}_k,k',\mathbf{t}_k,\psi^{t-1},\sigma^{2,t-1})p(\psi^{t-1}|k',\mathbf{t}_{k'})p(\sigma^{2,t-1})}{\xi(\psi^{t-1},\log(\sigma^{2,t-1}))}}\right)$$

where $p\left(\psi^t|k',\mathbf{t}_{k'}\right)$ is the prior distribution of ψ evaluated at ψ^t, $p\left(\sigma^{2,t}\right)$ is the prior distribution of σ^2 evaluated at $\sigma^{2,t}$, and $\xi\left(\psi^t,\log\left(\sigma^{2,t}\right)\right)$ is the multivariate normal density given in step (a) evaluated at $\left(\psi^t,\log\left(\sigma^{2,t}\right)\right)$.

(c) If the move is not accepted, let $\psi^t = \psi^{t-1},\ \log(\sigma^{2,t}) = \log(\sigma^{2,t-1})$

2. Draw a fixed effect, α^*, from the conditional posterior distribution $\alpha|k,\mathbf{t}_k,k',\mathbf{t}_{k'},\Sigma_\gamma^t,\sigma^{2,t}$ where $\alpha|k,\mathbf{t}_k,k',\mathbf{t}_{k'},\Sigma_\gamma^t,\sigma^{2,t} \sim MVN\left(\left(\mathbf{F}'\mathbf{C}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}'\mathbf{C}^{-1}\left(Y,\hat{\alpha}\right)',\left(\mathbf{F}'\mathbf{C}^{-1}\mathbf{F}\right)^{-1}\right)$,

   $\mathbf{F} = \left(\mathbf{B}'_{F_1},\mathbf{B}'_{F_2},\ldots,\mathbf{B}'_{F_n},\mathbf{I}_{k+2}\right)'$, $\mathbf{C}$ is a block-diagonal matrix composed of the n matrices $\left\{\left(\mathbf{B}_{R_i}\Sigma_\gamma^t\mathbf{B}'_{R_i}+\sigma^{2,t}\mathbf{I}_{m_i}\right)\right\}_{i=1}^n$ and the matrix $-\left(\frac{1}{n}\sum_{i=1}^n\frac{1}{m_i}I_{\alpha,i}\left(\hat{\alpha}\right)\right)^{-1}$, $\mathbf{B}_{F_i}$ and $\mathbf{B}_{R_i}$ are defined in Section 3, Y is all of the observed data, and $\hat{\alpha}$ is the maximum likelihood estimate of α.

3. Draw a random effect, γ^*, from the conditional posterior distribution $\gamma|k,\mathbf{t}_k,k',\mathbf{t}_{k'},\alpha^*,\Sigma_\gamma^t,\sigma^{2,t}$ where $\gamma|k,\mathbf{t}_k,k',\mathbf{t}_{k'},\alpha^*,\Sigma_\gamma^t,\sigma^{2,t} \sim MVN(\lambda,\Lambda)$, $\lambda = \Sigma_\gamma^t\mathbf{B}'_{R_i}\left(\sigma^{2,t}\mathbf{I}_{m_i}\right)^{-1}\left(Y_i-\mathbf{B}_{F_i}\alpha^*\right)$, and

$$\Lambda = \Sigma_\gamma^t\mathbf{B}'_{R_i}\left\{\left(\sigma^{2,t}\mathbf{I}_{m_i}\right)^{-1}-\left(\sigma^{2,t}\mathbf{I}_{m_i}\right)^{-1}\mathbf{B}_{F_i}\left(\frac{1}{\sigma^{2,t}}\sum_{i=1}^n\mathbf{B}'_{F_i}\mathbf{B}_{F_i}\right)^{-1}\mathbf{B}_{F_i}\left(\sigma^{2,t}\mathbf{I}_{m_i}\right)^{-1}\right\}\mathbf{B}_{R_i}\Sigma_\gamma^t$$

This addition to Algorithm I slows it down only marginally.

To sample random effects from the chain corresponding to $p(k,\mathbf{t}_k,k',\mathbf{t}_{k'},\psi,\sigma^2|y)$, steps (2) and (3) given above were just added to Algorithm II. In other words, a fixed and random effect curve were drawn for every random effects covariance matrix drawn. Figure 10 shows the pointwise median curve for the selected cow drawn for each chain, along with pointwise 95% credible intervals. The estimated cow-specific curves and credible intervals for both methods are similar.

9 Discussion

We have proposed a flexible approach to model multiple curves using free-knot splines. To identify the number and location of fixed and random effect knots in this flexible mixed-effects regression spline, sampling from $\hat{p}_{\text{Laplace}}(k,\mathbf{t}_k,k',\mathbf{t}_{k'}|y)$ is highly preferred to sampling from $\hat{p}_{\text{Plugged In}}(k,\mathbf{t}_k,k',\mathbf{t}_{k'}|y)$ and marginally

preferred to sampling from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$. The posterior sample corresponding to the Laplace approximation is more accurate than that of the plugged-in approximation. The disparity between the Laplace approximation and the plugged-in approximation is most pronounced in the different number of random effect knots fit to the simulated data sets. The Laplace method clearly seems to penalize models with unnecessarily large values of k' more appropriately than the plugged-in method. In addition, as measured by our simulations, the posterior based on the Laplace approximation is much more efficient to sample from (and just as accurate) as the exact posterior sample from $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}, \psi, \sigma^2|y)$.

We are currently investigating other approaches to the knot selection problem. Rather than sampling from the posterior $p(k, \mathbf{t}_k, k', \mathbf{t}_{k'}|y)$, we set $\mathbf{t}_k = \mathbf{t}_{k'}$ and then try to reduce the number of principal component curves associated with the random effects by sampling from the posterior $p(k, \mathbf{t}_k, r|y)$ where $r =$ the number of random effect principal components retained (Botts, 2005). This is similar in spirit to the methods employed by Shi et al. (1996), and James et al. (2000).

Behseta et al. (2005) propose an alternative hierarchical approach that, at the first level, fits curves separately for each subject, allowing each subject to have her own set of knots. However, their methods are only appropriate for large within-subject sample sizes.

References

deBoor, C. (1978). *A Practical Guide to Splines.* New York City: Springer-Verlag.

Behseta, S., Kass, R.E., and Wallstrom, G.L. (2005). Hierarchical models for assessing variability among functions. *Biometrika* **92**, 419-434.

Bigelow, J.L. and Dunson, D.B. (2007). Bayesian adaptive regression splines for hierarchical data. *Biometrics*, in press.

Botts, C. (2005). Bayesian Methods in Single and Multiple Curve Fitting. *Ph.D. Thesis*, Iowa State University.

Brumback, B.A. and Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *Journal of the American Statistical Association* **93**, 961-976.

Crainiceanu, C., Ruppert, D., and Wand, M.P. (2005). Bayesian analysis for penalized spline regression using WinBUGS. *Journal of Statistical Software* **14**, Article 14.

Demidenko, E. (2004). *Mixed Models: Theory and Applications.* John Wiley & Sons, Inc. Hoboken, New Jersey.

Dension, D.G.T., Mallick, B.K. and Smith, A.F.M. (1998). Automatic Bayesian curve fitting. *Journal of the Royal Statistical Society, Series B* **60**, 333-350.

Diggle, P.J. (1990). *Time Series: A Biostatistical Introduction.* Oxford University Press, Oxford.

DiMatteo, I., Genovese, C.R., and Kass, R.E. (2001). Bayesian curve-fitting with free-knot splines. *Biometrika* **88**, 1055-1071.

Dominici, F., Daniels, M., Zeger, S.L., Samet, J.M. (2002). Air pollution and mortality: Estimating regional and national dose-response relationships. *Journal of the American Statistical Association* **97**, 100-111.

Friedman, J.H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* **19**, 1-141.

Gelfand, A.E. and Dey, D.K. (1994). Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society, Series B* **56**, 501-514.

Green, P.J. (1995). Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination. *Biometrika* **82**, 711-732.

Halpern, E.F. (1973). Bayesian spline regression when the number of knots is unknown. *Journal of the Royal Statistical Society, Series B* **35**, 347-360.

Hastie, T.J. and Tibshirani, R.J. (1990). *Generalized Additive Models.* Boca Raton, Florida: Chapman & Hall.

Hooper, P.M. (1993). Iterative weighted least squares estimation in heteroscedastic linear models. *Journal of the American Statistical Association* **88**, 179-184.

James, G.M., Hastie, T.J., and Sugar, C.A. (2000). Principal component models for sparse functional data. *Biometrika* **87**, 587-602.

Kass,R.E. and Wasserman, L. (1994). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz Criterion. *Journal of the American Statistical Association* **90**, 928-934.

Lindstrom, M. J. (1999). Penalized estimation of free-knot splines. *Journal of Computational and Graphical Statistics* **8**, 333-352.

Paciorek, C.J. (2006). Misinformation in the conjugate prior for the linear model with implications for free-knot spline modeling. *Bayesian Analysis* **1**, 375-383.

Pauler, D.K. (1998). The Schwarz criterion and related methods for normal linear models. *Biometrika* **85**, 13-27.

Pourhamadi, M. (1999) Joint mean-covariance covariance models with applications to longitudinal data, I: Unconstrained parameterization. *Biometrika* **86**, 677-690.

Raftery, A.E., Madigan, D., and Volinsky, C.T. (1996). Accounting for model uncertainty in survival analysis improves predictive performance (with discussion). *Bayesian Statistics 5* (J. Bernardo, J. Berger, A. Dawid and A. Smith, eds.), 323-349. Oxford University Press.

Rice, J. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253-259.

Shi, M. Weiss, R.E. and Taylor, J.M.G. (1996). An analysis of pediatric CD4 counts for AIDS patients using flexible random curves. *Applied Statistics* **45**, 151-163.

Stone, C.J., Hansen, M., Kooperberg, C., and Truong, Y.K. (1997). Polynomial splines and their tensor products in extended linear modeling. *The Annals of Statistics* **25**, 1371-1470.

Taplin, R.H. (1993). Robust likelihood calculation for time series. *Journal of the Royal Statistics Society, Series B* **55**, 829-836.

Zhou, S. and Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *Journal of the American Statistical Association* **96**, 247-259.

Appendix

Exact formulas for I_ψ, I_α, and $I_{\log(\sigma^2)}$

The $(j,k)^{th}$ element in the matrix I_ψ can be calculated as

$$
\begin{aligned}
\left[I_\psi\right]_{j,k} &= -\frac{1}{2}\sum_{i=1}^{n}\mathrm{tr}\left\{-\mathbf{L}_i^{-1}\mathbf{G}_{j,i}\left(\mathbf{L}_i^{-1}\right)'\mathbf{G}_{k,i}+\mathbf{L}_i^{-1}\mathbf{H}_{(j,k),i}\right\}\\
&\quad +\frac{1}{2}\sum_{i=1}^{n}\left(y_i-\mathbf{G}_{F_i}\hat{\alpha}\right)'\left(-\mathbf{L}_i^{-1}\left(\mathbf{G}_{j,i}\mathbf{L}_i^{-1}\mathbf{G}_{k,i}'-\mathbf{H}_{(j,k),i}+\mathbf{G}_{k,i}\mathbf{L}_i^{-1}\mathbf{G}_{j,i}'\right)\left(\mathbf{L}_i^{-1}\right)'\right)\left(y_i-\mathbf{G}_{F_i}\hat{\alpha}\right),\ \text{where}\\
&\quad \mathbf{L}_i=\mathbf{B}_{R_i}\Sigma_\gamma(\hat{\psi})\mathbf{B}_{R_i}'+\hat{\sigma}^2\mathbf{I},\\
&\quad \mathbf{G}_{j,i}=\mathbf{B}_{R_i}\mathbf{W}_j(\hat{\psi})\mathbf{B}_{R_i}',\\
&\quad \mathbf{W}_j(\hat{\psi})=\left.\frac{\partial(TDT')}{\partial\psi_j}\right|_{\psi=\hat{\psi}},\ \text{and}\\
&\quad \mathbf{H}_{(j,k),i}=\mathbf{B}_{R_i}\left(\left.\frac{\partial^2(TDT')}{\partial\psi_j\partial\psi_k}\right|_{\psi=\hat{\psi}}\right)\mathbf{B}_{R,i}'\\
I_{\log(\sigma^2),i} &= -\frac{1}{2}\mathrm{tr}\left\{-\left(\hat{\sigma}^2\right)^2\mathbf{L}_i^{-1}\left(\mathbf{L}_i^{-1}\right)'+\sigma^2\mathbf{L}_i^{-1}\right\}\\
&\quad +\frac{1}{2}\left(y_i-\mathbf{B}_{F_i}\hat{\alpha}\right)'\left\{-\mathbf{L}_i^{-1}\left(\left(\hat{\sigma}^2\right)^2\mathbf{L}_i^{-1}-\hat{\sigma}^2\mathbf{I}+\left(\hat{\sigma}^2\right)^2\mathbf{L}_i^{-1}\right)\left(\mathbf{L}_i^{-1}\right)'\right\}\left(y_i-\mathbf{B}_{F_i}\hat{\alpha}\right)\\
I_{\alpha,i} &= -\mathbf{B}_{F_i}'\left(\mathbf{B}_{R_i}\Sigma_\gamma(\hat{\psi})\mathbf{B}_{R_i}'+\hat{\sigma}^2\mathbf{I}_{m_i}\right)^{-1}\mathbf{B}_{F_i}
\end{aligned}
$$

RJMCMC details for Algorithm I:

This section gives the details on Algorithm I. The details show how a move is made from a set of old knots to a new set of knots. Note that the constants which appear in this algorithm (.4 and .2) were selected to optimize its mixing and acceptance probability. The old set of knots will be denoted as $\left(k_{\text{old}},\mathbf{t}_{k_{\text{old}}},k_{\text{old}}',\mathbf{t}_{k_{\text{old}}'}\right)$.

The first step of the algorithm involves a step in the space of the fixed effect knots, Ω^k. Note that Ω^k is the sum of spaces $\Omega_1,\Omega_2,\dots$ where $\Omega_j=[a,b]^j$. Therefore Ω^k can be written as $\Omega^k=\bigoplus_{j=1}^{\infty}\Omega_j$. The first decision to be made is whether to give birth, relocate, or kill a fixed knot. These probabilities are labeled as b_F, r_F, and d_F, respectively, and are calculated as

$$
b_F\left(k_{\text{old}}\right)=.4\times\min\left(1,\frac{p_k\left(k_{\text{old}}+1\right)}{p_k\left(k_{\text{old}}\right)}\right),\quad d_F\left(k_{\text{old}}\right)=\begin{cases}0 & k_{\text{old}}=k_{\text{old}}'\\ .4\times\min\left(1,\frac{p_k\left(k_{\text{old}}-1\right)}{p_k\left(k_{\text{old}}\right)}\right) & o.w.\end{cases},
$$

and $r_F\left(k_{\text{old}}\right)=1-b_F\left(k_{\text{old}}\right)-d_F\left(k_{\text{old}}\right)$, where $p_k\left(k_{\text{old}}\right)$ is the prior distribution assigned to k evaluated at k_{old}.

If birth is chosen, a fixed effect knot, t_k^{selected}, is selected at random and a new knot is given birth to. The new knot, t_k^{new}, is sampled from the distribution $t_k^{\text{new}}\sim TN_a^b\left(t_k^{\text{selected}},.2\right)$ where TN_a^b is the normal distribution truncated at a and b with mean at t_k^{selected} and variance .2. If death is selected, then one of the fixed effect

knots is uniformly selected and killed. When relocation is chosen, a fixed effect knot, $t_k^{\text{relocated}}$, is uniformly selected and then relocated to a position $t_k^{\text{new loc}}$ where $t_k^{\text{new loc}} \sim TN_a^b \left(t_k^{\text{relocated}}, .2 \right)$. The jump probabilities corresponding to each move is given in Table 2. Note: $h\left(y|x \right) = \phi\left(y; x, .2 \right) \left(\Phi\left(b; x, .2 \right) - \Phi\left(a; x, .2 \right) \right)^{-1}$ where $\phi\left(\cdot; \mu, \tau^2 \right)$ is the normal density with mean μ and variance τ^2, and $\Phi\left(\cdot; \mu, \tau^2 \right)$ is the cumulative distribution function of a normal density with mean μ and variance τ^2.

Move in Ω^k	$P_{\text{Jump}}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} \longrightarrow k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} \,\middle|\, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \right)$
Birth	$b_F\left(k_{\text{old}} \right) \frac{1}{k_{\text{old}}} \sum_{x \in t_{k_{\text{old}}}} h\left(t_k^{\text{new}} | x \right)$
Relocation	$r_F\left(k_{\text{old}} \right) \frac{1}{k_{\text{old}}} h\left(t_k^{\text{new loc}} | t_k^{\text{relocated}} \right)$
Death	$d_F\left(k_{\text{old}} \right) \frac{1}{k_{\text{old}}}$

Table 2: Jump Probabilities in Ω^k

A move must then be made with the random effect knots. This is the second step in the algorithm. As with the fixed effect knots, one of three moves can be made in $\Omega^{k'}$, the space of random effect knots: birth, death, or relocation. The probabilities associated with each are denoted as b_R, d_R, and r_R, respectively. They are calculated as

$$b_R\left(k'_{\text{old}} \right) = \begin{cases} 0 & k'_{\text{old}} = k_{\text{old}} \\ .4 \times \min\left(1, \frac{p_{k'}\left(k'_{\text{old}} + 1 \right)}{p_{k'}\left(k'_{\text{old}} \right)} \right) & o.w. \end{cases},$$

$$d_R\left(k'_{\text{old}} \right) = \begin{cases} 0 & k'_{\text{old}} = 1 \\ .4 \times \min\left(1, \frac{p_{k'}\left(k'_{\text{old}} - 1 \right)}{p_{k'}\left(k'_{\text{old}} \right)} \right) & o.w. \end{cases},$$

and $r_R\left(k'_{\text{old}} \right) = 1 - b_R\left(k'_{\text{old}} \right) - d_R\left(k'_{\text{old}} \right)$, where $p_{k'}\left(k'_{old} \right)$ is the prior distribution of k' evaluated at k'_{old}. If death is selected, then one of the random effect knots is uniformly selected and then killed. If birth is selected, then one of the random effect knots is uniformly selected (call this $t_{k'}^{\text{selected}}$), and a knot is added to the set of random effect knots (call this knot $t_{k'}^{\text{new}}$). Note that $t_{k'}^{\text{new}} \sim TN_a^b \left(t_{k'}^{\text{selected}}, .2 \right)$. If relocation is selected, then one of the random effect knots is uniformly selected, $t_{k'}^{\text{relocated}}$, and moved to another location (call this $t_{k'}^{\text{new loc}}$). Again, $t_{k'}^{\text{new loc}} \sim TN_a^b \left(t_{k'}^{\text{relocated}}, .2 \right)$. The jump probabilities associated with these moves are given in Table 3.

Derivation of Metropolis-Hastings Acceptance Probabilities to Satisfy Detailed Balance of Algorithm I

We show that the acceptance probabilities given in Section 4 are such that detailed balance is satisfied. In this case, it needs to be shown that

Move in $\Omega^{k'}$	$P_{\text{Jump}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \longrightarrow k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \,\middle|\, k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}\right)$
Birth	$b_R\left(k'_{\text{old}}\right) \frac{1}{k'_{\text{old}}} \sum_{x \in t_{k'_{\text{old}}}} h\left(t_{k'}^{\text{new}}|x\right)$
Relocation	$r_R\left(k'_{\text{old}}\right) \frac{1}{k'_{\text{old}}} h\left(t_{k'}^{\text{new loc}}|t_{k'}^{\text{selected}}\right)$
Death	$d_R\left(k'_{\text{old}}\right) \frac{1}{k'_{\text{old}}}$

Table 3: Jump Probabilities in $\Omega^{k'}$

$$P_{\text{Trans}}\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}} k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \longrightarrow k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right) p\left(k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\,\middle|\,y\right)$$
$$= \ P_{\text{Trans}}\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}} k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \longrightarrow k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right) p\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\,\middle|\,y\right)$$

(Note: throughout the rest of this proof, $\mathbf{t}_{k_{\text{new}}}$ will be denoted as $\mathbf{t}_{k_n}$, k_{old} will be denoted as k_o, and so on). Assume that both a fixed and random knot are being added. In this case,

$$P_{\text{Trans}}\left(k_{\text{o}}, \mathbf{t}_{k_{\text{o}}}, k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}} \longrightarrow k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}, k'_{\text{n}}, \mathbf{t}_{k'_{\text{n}}}\right) p\left(k_{\text{o}}, \mathbf{t}_{k_{\text{o}}}, k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\,\middle|\,y\right)$$

$$= \underbrace{b_F\left(k_{\text{o}}\right) \frac{1}{k_{\text{o}}} \sum_{x \in \mathbf{t}_{k_{\text{o}}}} h\left(t_k^{\text{new}}|x\right) \times \min\left(1, \frac{P_{\text{Jump}}\left(k_{\text{n}}, \mathbf{t}_{k_{\text{n}}} \longrightarrow k_{\text{o}}, \mathbf{t}_{k_{\text{o}}}\,\middle|\,k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\right)}{P_{\text{Jump}}\left(k_{\text{o}}, \mathbf{t}_{k_{\text{o}}} \longrightarrow k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}\,\middle|\,k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\right)} \frac{p\left(k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}\,\middle|\,k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}, y\right)}{p\left(k_{\text{o}}, \mathbf{t}_{k_{\text{o}}}\,\middle|\,k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}, y\right)}\right)}_{P_{\text{Trans}}\left(k_o, \mathbf{t}_{k_o} \longrightarrow k_n, \mathbf{t}_{k_n} | k'_o, \mathbf{t}_{k'_o}\right)}$$

$$\times \underbrace{b_R\left(k'_{\text{o}}\right) \frac{1}{k'_{\text{o}}} \sum_{x \in \mathbf{t}'_{k_{\text{o}}}} h\left(t_{k'}^{\text{new}}|x\right) \times \min\left(1, \frac{P_{\text{Jump}}\left(k'_{\text{n}}, \mathbf{t}_{k'_{\text{n}}} \longrightarrow k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\,\middle|\,k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}\right)}{P_{\text{Jump}}\left(k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}} \longrightarrow k'_{\text{n}}, \mathbf{t}_{k'_{\text{n}}}\,\middle|\,k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}\right)} \frac{p\left(k'_{\text{n}}, \mathbf{t}_{k'_{\text{n}}}\,\middle|\,k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}, y\right)}{p\left(k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\,\middle|\,k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}, y\right)}\right)}_{P_{\text{Trans}}\left(k'_o, \mathbf{t}_{k'_o} \longrightarrow k'_n, \mathbf{t}_{k'_n} \middle| k_n, \mathbf{t}_{k_n}\right)}$$

$$\times \ p\left(k_{\text{o}}, \mathbf{t}_{k_{\text{o}}}, k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\,\middle|\,y\right).$$

Assume that the ratios within each min argument are < 1 (if this is not the case, the reciprocal of the ratios will be less than 1, and the proof can be done in the other direction), and recall that

1. $b_F\left(k_{\text{o}}\right) \frac{1}{k_{\text{o}}} \sum_{x \in \mathbf{t}_{k_{\text{o}}}} h\left(t_k^{\text{new}}|x\right) = P_{\text{Jump}}\left(k_{\text{o}}, \mathbf{t}_{k_{\text{o}}} \longrightarrow k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}\,\middle|\,k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}}\right)$

2. $b_R\left(k'_{\text{o}}\right) \frac{1}{k'_{\text{o}}} \sum_{x \in \mathbf{t}'_{k_{\text{o}}}} h\left(t_{k'}^{\text{new}}|x\right) = P_{\text{Jump}}\left(k'_{\text{o}}, \mathbf{t}_{k'_{\text{o}}} \longrightarrow k'_{\text{n}}, \mathbf{t}_{k'_{\text{n}}}\,\middle|\,k_{\text{n}}, \mathbf{t}_{k_{\text{n}}}\right)$

3. $p\left(k_n, \mathbf{t}_{k_n}\,\middle|\,k'_o, \mathbf{t}_{k'_o}, y\right) \Big/ p\left(k_o, \mathbf{t}_{k_o}\,\middle|\,k'_0, \mathbf{t}_{k'_o}, y\right) = \dfrac{p\left(y|k_n, \mathbf{t}_{k_n}, k'_o, \mathbf{t}_{k'_o}\right) p(k_n) p(\mathbf{t}_{k_n}|k_n)}{p\left(y|k_o, \mathbf{t}_{k_o}, k'_o, \mathbf{t}_{k'_o}\right) p(k_o) p(\mathbf{t}_{k_o}|k_o)}$

4. $p\left(k'_n, \mathbf{t}_{k'_n}\,\middle|\,k_n, \mathbf{t}_{k_n}, y\right) \Big/ p\left(k'_o, \mathbf{t}_{k'_o}\,\middle|\,k_n, \mathbf{t}_{k_n}, y\right) = \dfrac{p\left(y|k_n, \mathbf{t}_{k_n}, k'_n, \mathbf{t}_{k'_n}\right) p(k'_n) p(\mathbf{t}_{k'_n}|k'_n)}{p\left(y|k'_o, \mathbf{t}_{k'_o}, k_n, \mathbf{t}_{k_n}\right) p(k'_o) p(\mathbf{t}_{k'_o}|k'_o)}$

With these four observations it is clear that the above expression reduces to

$$
\begin{aligned}
& P_{\text{Jump}}\left(k_n, \mathbf{t}_{k_n} \longrightarrow k_o, \mathbf{t}_{k_o} \,\middle|\, k'_o, \mathbf{t}_{k'_o}\right) \frac{p\left(y \,\middle|\, k_n, \mathbf{t}_{k_n}, k'_o, \mathbf{t}_{k'_o}\right) p\left(k_n\right)}{p\left(y \,\middle|\, k_o, \mathbf{t}_{k_o}, k'_o, \mathbf{t}_{k_o}\right) p\left(k_o\right)} \frac{p(\mathbf{t}_{k_n}|k_n)}{p(\mathbf{t}_{k_o}|k_o)} P_{\text{Jump}}\left(k'_n, \mathbf{t}_{k'_n} \longrightarrow k'_o, \mathbf{t}_{k'_o} \,\middle|\, k_n, \mathbf{t}_{k_n}\right) \\
& \qquad \times \frac{p\left(y | k_n \mathbf{t}_{k_n}, k_{n'}, \mathbf{t}_{k'_n}\right) p\left(k'_n\right)}{p\left(y | k_n, \mathbf{t}_{k_n}, k'_o, \mathbf{t}_{k'_o}\right) p\left(k'_o\right)} \frac{p(\mathbf{t}_{k'_n}|k'_n)}{p(\mathbf{t}_{k'_o}|k'_o)} \frac{p\left(y | k_0, \mathbf{t}_{k_o}, k'_o, \mathbf{t}_{k_o}\right) p\left(k_o\right) p\left(k'_o\right) p(\mathbf{t}_{k_o}|k_o) p(\mathbf{t}_{k'_o}|k'_o)}{p\left(y\right)} \\
& = \quad P_{\text{Jump}}\left(k_n, \mathbf{t}_{k_n} \longrightarrow k_o, \mathbf{t}_{k_o} \,\middle|\, k'_o, \mathbf{t}_{k'_o}\right) P_{\text{Jump}}\left(k'_n, \mathbf{t}_{k'_n} \longrightarrow k'_o, \mathbf{t}_{k'_o} \,\middle|\, k_n, \mathbf{t}_{k_n}\right) \\
& \qquad \times \frac{p\left(y | k_n, \mathbf{t}_{k_n}, k'_n, \mathbf{t}_{k_n}\right) p\left(k_n\right) p\left(k'_n\right) p(\mathbf{t}_{k_n}|k_n) p(\mathbf{t}_{k'_n}|k'_n)}{p\left(y\right)} \\
& = \quad \underbrace{P_{\text{Jump}}\left(k_n, \mathbf{t}_{k_n} \longrightarrow k_o, \mathbf{t}_{k_o} \,\middle|\, k'_o, \mathbf{t}_{k'_o}\right)}_{= \, d_F(k_n)\frac{1}{k_n+1}} \\
& \qquad \times \underbrace{\min\left(1, \frac{P_{\text{Jump}}\left(k_o, \mathbf{t}_{k_o} \longrightarrow k_n, \mathbf{t}_{k_n} \,\middle|\, k'_o, \mathbf{t}_{k_o}\right)}{P_{\text{Jump}}\left(k_n, \mathbf{t}_{k_n} \longrightarrow k_o, \mathbf{t}_{k_o} \,\middle|\, k'_o, \mathbf{t}_{k'_o}\right)} \frac{p\left(k_0, \mathbf{t}_{k_o} \,\middle|\, k'_o, \mathbf{t}_{k'_o}, y\right)}{p\left(k_n, \mathbf{t}_{k_n} \,\middle|\, k'_0, \mathbf{t}_{k'_o}, y\right)}\right)}_{= \, 1} \\
& \qquad \times \underbrace{P_{\text{Jump}}\left(k'_n, \mathbf{t}_{k'_n} \longrightarrow k'_o, \mathbf{t}_{k'_o} \,\middle|\, k_n, \mathbf{t}_{k_n}\right)}_{= d_R(k'_n)\frac{1}{k'_n+1}} \\
& \qquad \times \underbrace{\min\left(1, \frac{P_{\text{Jump}}\left(k'_o, \mathbf{t}_{k'_o} \longrightarrow k'_n, \mathbf{t}_{k'_n} \,\middle|\, k_n, \mathbf{t}_{k_n}\right)}{P_{\text{Jump}}\left(k'_n, \mathbf{t}_{k'_n} \longrightarrow k'_o, \mathbf{t}_{k'_o} \,\middle|\, k_n, \mathbf{t}_{k_n}\right)} \frac{p\left(k'_0, \mathbf{t}_{k'_o} \,\middle|\, k_n, \mathbf{t}_{k_n}, y\right)}{p\left(k'_n, \mathbf{t}_{k'_n} \,\middle|\, k_n, \mathbf{t}_{k_n}, y\right)}\right)}_{= \, 1} \times p\left(k_n, \mathbf{t}_{k_n}, k'_n, \mathbf{t}_{k'_n} \,\middle|\, y\right) \\
& = \quad P_{\text{Trans}}\left(k_n, \mathbf{t}_{k_n} \longrightarrow k_o, \mathbf{t}_{k_o} \,\middle|\, k'_o, \mathbf{t}_{k'_o}\right) P_{\text{Trans}}\left(k'_n, \mathbf{t}_{k'_n} \longrightarrow k'_o, \mathbf{t}_{k'_o} \,\middle|\, k_n, \mathbf{t}_{k_n}\right) p\left(k_n, \mathbf{t}_{k_n}, k'_n, \mathbf{t}_{k'_n} \,\middle|\, y\right) \\
& = \quad P_{\text{Trans}}\left(k_n, \mathbf{t}_{k_n}, k'_n, \mathbf{t}_{k'_n} \longrightarrow k_o, \mathbf{t}_{k_o}, k'_o, \mathbf{t}_{k'_o}\right) p\left(k_n, \mathbf{t}_{k_n}, k'_n, \mathbf{t}_{k'_n} \,\middle|\, y\right).
\end{aligned}
$$

This argument shows detailed balance holds for the given acceptance ratios when a fixed and random effect knot are added. Similar arguments can be made to prove detailed balance when a fixed and/or random effect knot is deleted and/or relocated.

RJMCMC details for Algorithm II:

As mentioned in the main text of this paper, Algorithm II is only a slight variation of Algorithm I. In Algorithm II, candidate values of $(k, \mathbf{t}_k)$ are sampled just as they are in Algorithm I. Conditioned on each sampled pair of knots, candidate values of k', $\mathbf{t}_{k'}$, ψ and σ^2 are sampled. These candidate values are denoted ψ_{new} and σ^2_{new}, and they are sampled from the multivariate normal distribution

$$
\begin{pmatrix} \psi_{\text{new}} \\ \log(\sigma^2_{\text{new}}) \end{pmatrix} \sim MVN\left(\begin{pmatrix} \hat{\psi} \\ \log(\hat{\sigma}^2) \end{pmatrix}, \widehat{\text{Cov}}\left(\hat{\psi}, \log(\hat{\sigma}^2)\right)\right).
$$

Note that $\hat{\psi}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates of ψ and $\hat{\sigma}^2$ corresponding to the model with fixed effect knots at $\mathbf{t}_{k_{\text{old}}}$ and random effect knots at $\mathbf{t}_{k'_{\text{new}}}$.

The jump probabilities corresponding to step 2 of Algorithm II are the same as those in step 2 of Algorithm I. The jump probabilities corresponding to step 3 of Algorithm II are given in Table 4. Note: $MVN_3(z, w)$ is the multivariate density, with mean $\left(\hat{\psi}, \log(\hat{\sigma}^2)\right)$ and covariance $\widehat{\text{Cov}}\left(\hat{\psi}, \log(\hat{\sigma}^2)\right)$ evaluated at the vector (z, w). In this case, $\hat{\psi}$ and $\hat{\sigma}^2$ are the maximum likelihood estimates calculated under the model with fixed effect knots at $\mathbf{t}_{k_{\text{old}}}$ and random effect knots at $\mathbf{t}_{k'_{\text{new}}}$.

	$P_{\text{Jump}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \psi_{\text{old}}, \sigma^2_{\text{old}} \longrightarrow k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \psi_{\text{new}}, \sigma^2_{\text{new}} \mid k_{\text{old}}, \mathbf{t}_{k_{\text{old}}}\right)$
Birth	$\left(b_F\left(k'_{\text{old}}\right) \frac{1}{k'_{\text{old}}} \sum_{x \in t_{k'_{\text{old}}}} h\left(t_{k'}^{\text{new}}|x\right)\right) \times MVN_3\left(\psi_{\text{new}}, \log(\sigma^2_{\text{new}})\right)$
Relocation	$\left(r_F\left(k_{\text{old}}\right) \frac{1}{k'_{\text{old}}} h\left(t_{k'}^{\text{new loc}}|t_{k'}^{\text{relocated}}\right)\right) \times MVN_3\left(\psi_{\text{new}}, \log(\sigma^2_{\text{new}})\right)$
Death	$d_F\left(k'_{\text{old}}\right) \frac{1}{k'_{\text{old}}}$

Table 4: Jump Probabilities For Step 3 in Algorithm II

It can be shown, using methods similar to those given before, that these jump probabilities (along with the acceptance probabilities of Algorithm II given in Section 5) guarantee detailed balance.

Proof of Theorem 1

We first prove that the maximum likelihood estimators of α, Σ_γ, and σ^2 are consistent with respect to n, the number of subjects in the study. To be more specific, we prove that $\hat{\alpha}$, $\hat{\psi}$, and $\hat{\sigma}^2$ converge to their true values as n goes to infinity. Proving the consistency of these estimators is not straight-forward since the observed data for all n subjects, $y_1, y_2, \ldots, y_n$, are not i.i.d. (recall that they are independent yet not identically distributed since $y_i \sim MVN\left(\mathbf{B}_{F_i}\alpha, \mathbf{B}_{R_i}\Sigma_\gamma(\psi)\mathbf{B}'_{R_i} + \sigma^2\mathbf{I}_{m_i}\right)$).

Hooper (1993) and Demidenko (2004) show, however, that if one considers the design matrices of the ith individual $(\mathbf{B}_{F_i}, \mathbf{B}_{R_i})$ and the number of observations for the ith individual (m_i) as random, independent of ϵ_i and γ_i, and part of the observed data, then the i.i.d. assumption of the data is satisfied. For the problem presented in this paper, such considerations make sense. The design matrices corresponding to individual i, $\mathbf{B}_{F_i}$ and $\mathbf{B}_{R_i}$, can be thought of as random since the points in time at which the m_i observations occur, $\mathbf{x}_i$, vary. If the number of observations recorded for all individuals, $m_1, m_2, \ldots, m_n$, are i.i.d. (as we assume in Condition (2)), and the times at which these observations occur, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$, are i.i.d. (as we assume in Condition (4)), then the sets $\{y_i, \mathbf{B}_{F_i}, \mathbf{B}_{R_i}, m_i\}_{i=1}^n$ are i.i.d. It follows that the maximum likelihood estimators of α, ψ, and σ^2 converge to their true values as n, the number of subjects, approaches ∞.

We now arrive at the result in Theorem 1 by examining the behavior of $\log\left(\frac{\hat{p}_{\text{Lp}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}|k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}{\hat{p}_{\text{Lp}}\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}|k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}\right)$

as $n \longrightarrow \infty$. Recall that $k'_{\text{old}} < k'_{\text{new}}$ and $\mathbf{t}_{k'_{\text{old}}} \subset \mathbf{t}_{k'_{\text{new}}}$, making the model with knots at $\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right)$ nested within the model with knots at $\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right)$.

$$
\begin{aligned}
&\log\left(\frac{\hat{p}_{\text{Lp}}\left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}{\hat{p}_{\text{Lp}}\left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y\right)}\right) \\
&= \log\left(\frac{\hat{p}_{\text{Lp}}\left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)}{\hat{p}_{\text{Lp}}\left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right)} \frac{p\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)}{p\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right)}\right) \\
&= \log\left(\frac{\hat{p}\left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \hat{\psi}(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}), \hat{\sigma}^2\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right)\right)}{\hat{p}\left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \hat{\psi}(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}), \hat{\sigma}^2\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right)\right)}\right) \\
&\quad + \log\left(\frac{p\left(\hat{\psi} | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right)}{p\left(\hat{\psi} | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right)} \frac{p\left(\hat{\sigma}^2 | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right)}{p\left(\hat{\sigma}^2 | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right)} \frac{\left|\widehat{\text{Cov}}\left(\hat{\psi}\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right), \log\left(\hat{\sigma}^2\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right)\right)\right)\right|^{\frac{1}{2}}}{\left|\widehat{\text{Cov}}\left(\hat{\psi}\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right), \log\left(\hat{\sigma}^2\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right)\right)\right)\right|^{\frac{1}{2}}} (2\pi)^{\frac{d_1 - d_2}{2}}\right) \\
&\quad + \log\left(\frac{p\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}\right)}{p\left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}\right)}\right),
\end{aligned}
$$

where $d_1 = \dim\left(\hat{\psi}\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}\right)\right) = \frac{(k'_{\text{old}}+2)(k'_{\text{new}}+3)}{2}$, $d_2 = \dim\left(\hat{\psi}\left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}\right)\right) = \frac{(k'_{\text{new}}+2)(k'_{\text{new}}+3)}{2}$, $d_2 > d_1$, and $\widehat{\text{Cov}}\left(\hat{\psi}, \log(\hat{\sigma}^2)\right) = -\frac{1}{n} H\left(\hat{\psi}, \log(\hat{\sigma}^2)\right)^{-1}$, with

$$
H\left(\hat{\psi}, \log\left(\hat{\sigma}^2\right)\right) = \left(\frac{\partial^2 \log\left\{p(y|k, \mathbf{t}_k, k', \mathbf{t}_k, \alpha, \psi, \sigma^2) p(\alpha|\mathbf{t}_k) p(\psi|\mathbf{t}_k, \mathbf{t}_{k'}) p(\sigma^2|\mathbf{t}_k, \mathbf{t}_{k'})/n\right\}}{\partial\left(\psi, \log(\sigma^2)\right)' \partial\left(\psi, \log(\sigma^2)\right)}\right)^{-1}\Bigg|_{\alpha=\hat{\alpha}, \psi=\hat{\psi},\ \sigma^2=\hat{\sigma}^2}.
$$

From this, we get that

$$\log \left(\frac{\hat{p}_{\text{Lp}} \left(k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y \right)}{\hat{p}_{\text{Lp}} \left(k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, y \right)} \right)$$

$$= \log \left(\frac{\hat{p} \left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \hat{\psi}(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}), \hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right) \right)}{\hat{p} \left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \hat{\psi}(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}), \hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right) \right)} \right)$$

$$+ \log \left(\frac{p \left(\hat{\psi} | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right)}{p \left(\hat{\psi} | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right)} \frac{p \left(\hat{\sigma}^2 | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right)}{p \left(\hat{\sigma}^2 | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right)} \frac{\left| -H \left(\hat{\psi} \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right), \log \left(\hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right) \right) \right) \right|^{\frac{1}{2}}}{\left| -H \left(\hat{\psi} \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right), \log \left(\hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right) \right) \right) \right|^{\frac{1}{2}}} \left(\frac{n}{2\pi} \right)^{\frac{d_2 - d_1}{2}} \right)$$

$$+ \log \left(\frac{p \left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \right)}{p \left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \right)} \right)$$

$$= \log \left(\frac{\hat{p} \left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}}, \hat{\psi}(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}}), \hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right) \right)}{\hat{p} \left(y | k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}}, \hat{\psi}(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}}), \hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right) \right)} \right) + K_n + \frac{d_2 - d_1}{2} \log(n)$$

$$+ \log \left(\frac{p \left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{old}}, \mathbf{t}_{k'_{\text{old}}} \right)}{p \left(k_{\text{new}}, \mathbf{t}_{k_{\text{new}}}, k'_{\text{new}}, \mathbf{t}_{k'_{\text{new}}} \right)} \right),$$

where

$$K_n = \log \left(\frac{p \left(\hat{\psi} | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right)}{p \left(\hat{\psi} | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right)} \frac{p \left(\hat{\sigma}^2 | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right)}{p \left(\hat{\sigma}^2 | \mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right)} \frac{\left| -H \left(\hat{\psi} \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right), \log \left(\hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{old}}} \right) \right) \right) \right|^{\frac{1}{2}}}{\left| -H \left(\hat{\psi} \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right), \log \left(\hat{\sigma}^2 \left(\mathbf{t}_{k_{\text{new}}}, \mathbf{t}_{k'_{\text{new}}} \right) \right) \right) \right|^{\frac{1}{2}}} \left(2\pi \right)^{\frac{d_1 - d_2}{2}} \right).$$

Gelfand and Dey (1994) show that if the models are nested (which is guaranteed by Condition (1)) and if the maximum likelihood estimators exist, then $K_n = O(1)$. Note that existence of the maximum likelihood estimators follows from Condition (3). Demidenko (2004) shows that if $\sum_{i=1}^{n} \left(m_i - (k'_{\text{new}} + 2) \right) > k_{\text{new}} + 2$, then the maximum likelihood estimators of α, ψ, and σ^2 exist. Since K_n is bounded, for sufficiently large values of n, K_n becomes negligible compared to $\frac{d_2 - d_1}{2} \log(n)$, and we can say that

$$\log\left(\frac{\hat{p}_{\mathrm{Lp}}\left(k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y\right)}{\hat{p}_{\mathrm{Lp}}\left(k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y\right)}\right)$$

$$\approx \log\left(\frac{\hat{p}\left(y|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}},\hat{\psi}(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{old}}}),\hat{\sigma}^2\left(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{old}}}\right)\right)}{\hat{p}\left(y|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}},\hat{\psi}(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{new}}}),\hat{\sigma}^2\left(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{new}}}\right)\right)}\right)+\frac{d_2-d_1}{2}\log(n)$$

$$+\log\left(\frac{p\left(k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}\right)}{p\left(k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}\right)}\right)$$

$$\geq \log\left(\frac{\hat{p}\left(y|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}},\hat{\psi}(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{old}}}),\hat{\sigma}^2\left(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{old}}}\right)\right)}{\hat{p}\left(y|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}},\hat{\psi}(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{new}}}),\hat{\sigma}^2\left(\mathbf{t}_{k_{\mathrm{new}}},\mathbf{t}_{k'_{\mathrm{new}}}\right)\right)}\right)+\log\left(\frac{p\left(k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}\right)}{p\left(k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}\right)}\right)$$

$$= \log\left(\frac{\hat{p}_{\mathrm{Plg}}\left(y|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}\right)}{\hat{p}_{\mathrm{Plg}}\left(y|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}\right)}\frac{p\left(k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}\right)}{p\left(k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}\right)}\right)$$

$$= \log\left(\frac{\hat{p}_{\mathrm{Plg}}\left(k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y\right)}{\hat{p}_{\mathrm{Plg}}\left(k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y\right)}\right),$$

and it follows that

$$\frac{\hat{p}_{\mathrm{Lp}}(k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y)}{\hat{p}_{\mathrm{Lp}}(k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y)} \leq \frac{\hat{p}_{\mathrm{Plg}}(k'_{\mathrm{new}},\mathbf{t}_{k'_{\mathrm{new}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y)}{\hat{p}_{\mathrm{Plg}}(k'_{\mathrm{old}},\mathbf{t}_{k'_{\mathrm{old}}}|k_{\mathrm{new}},\mathbf{t}_{k_{\mathrm{new}}},y)}.$$
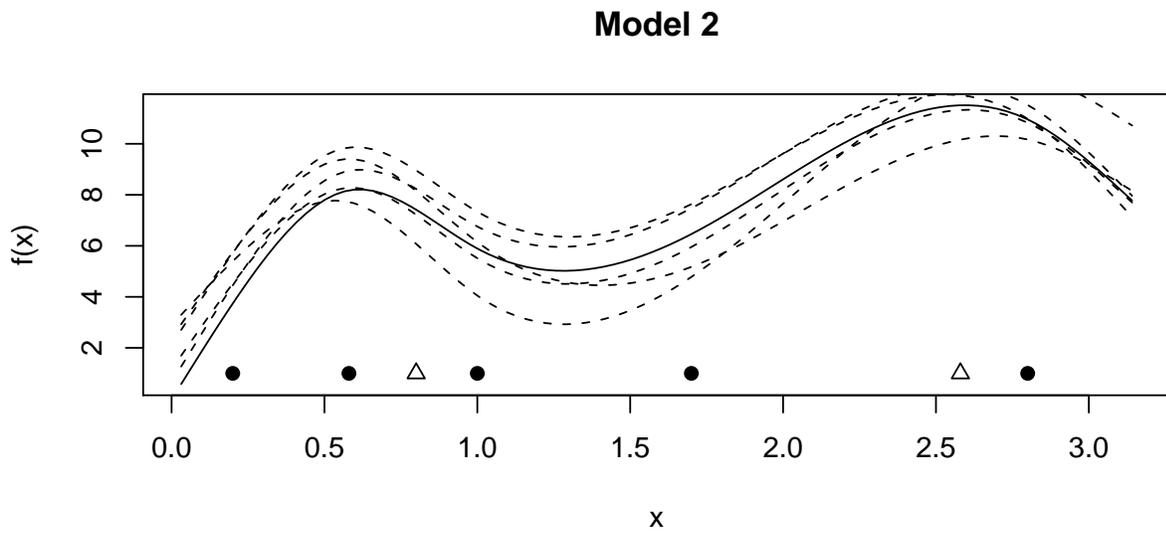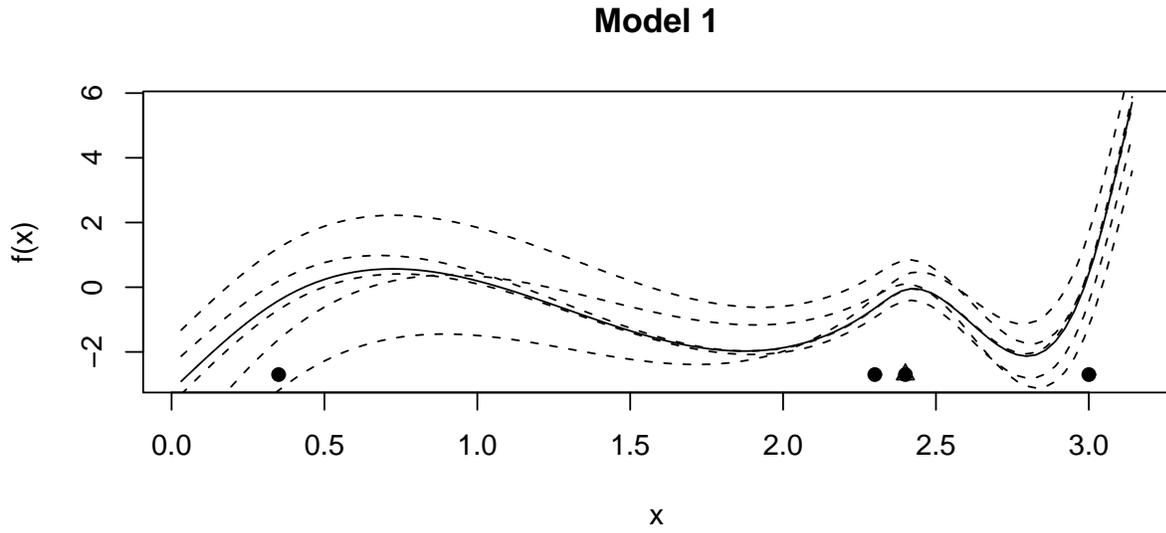
Model 1

Model 2

Figure 1: Plots of Models 1-2. (— —): $f(x) + G_i(x)$, (——): $f(x)$. The location of the fixed effect knots are denoted with a • and the location of the random effect by a △.
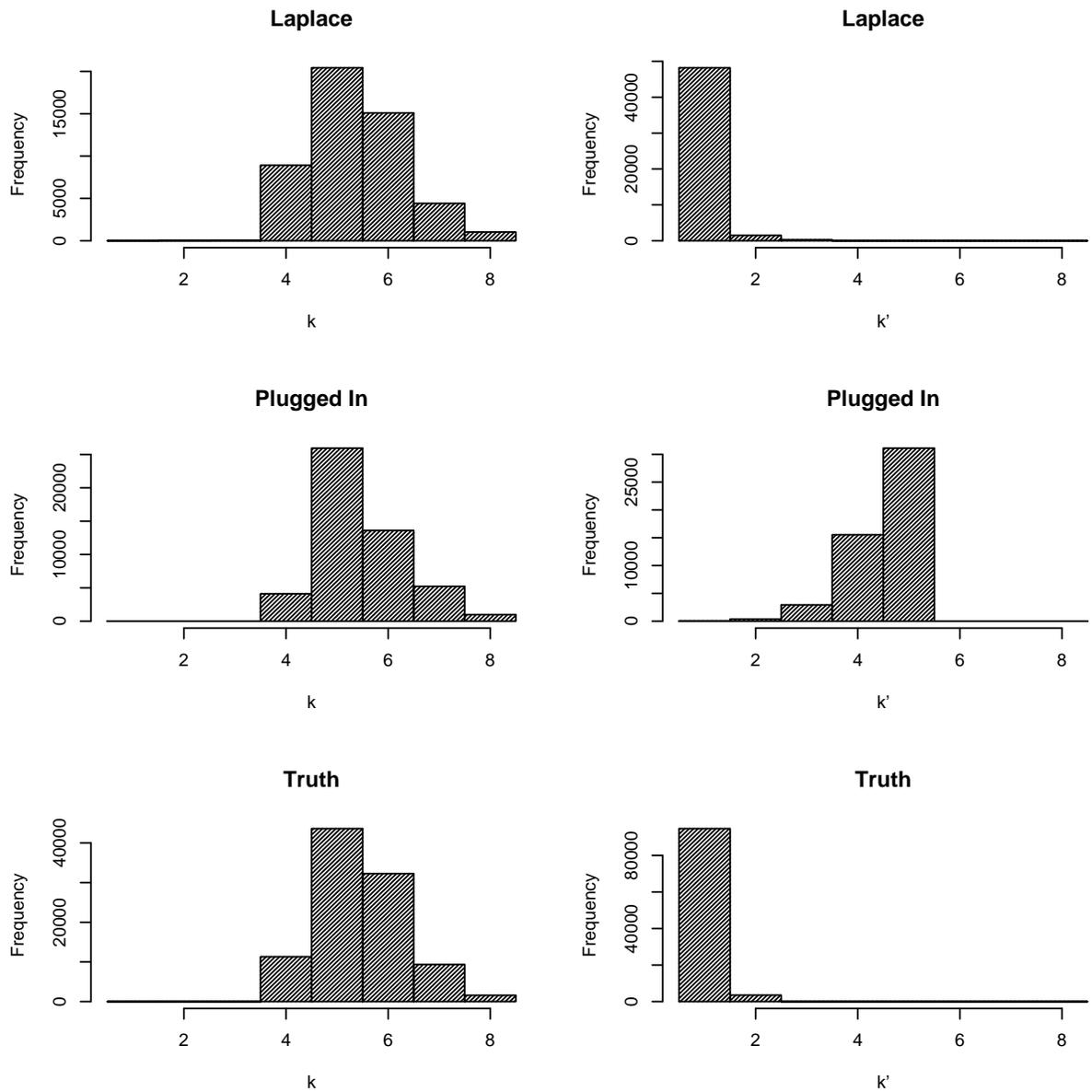
Figure 2: Marginal Distributions of k and k' for Model 1 at $n = 25$. Average acceptance rates: Laplace - 72%, Plugged-In - 44%, Truth - 42%.
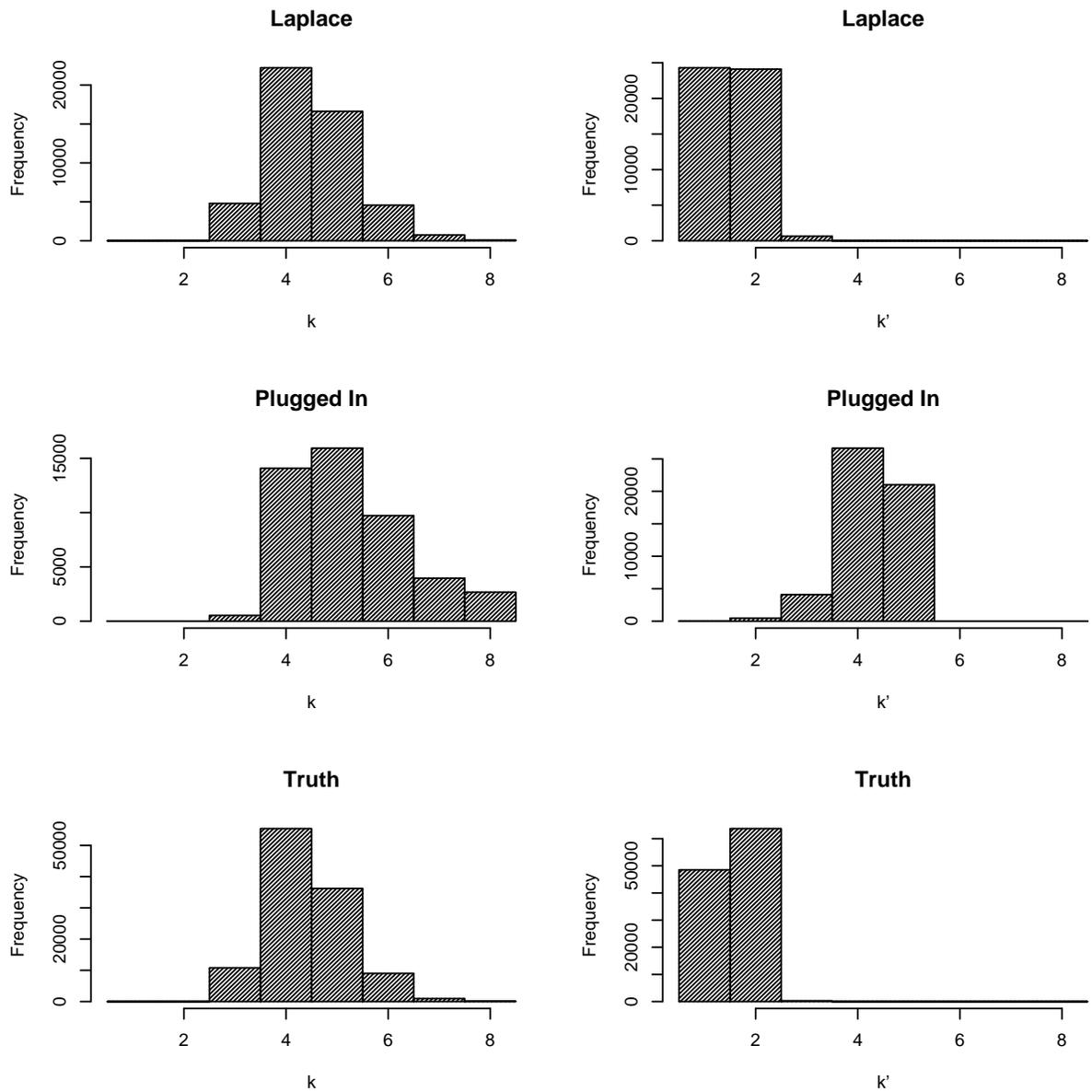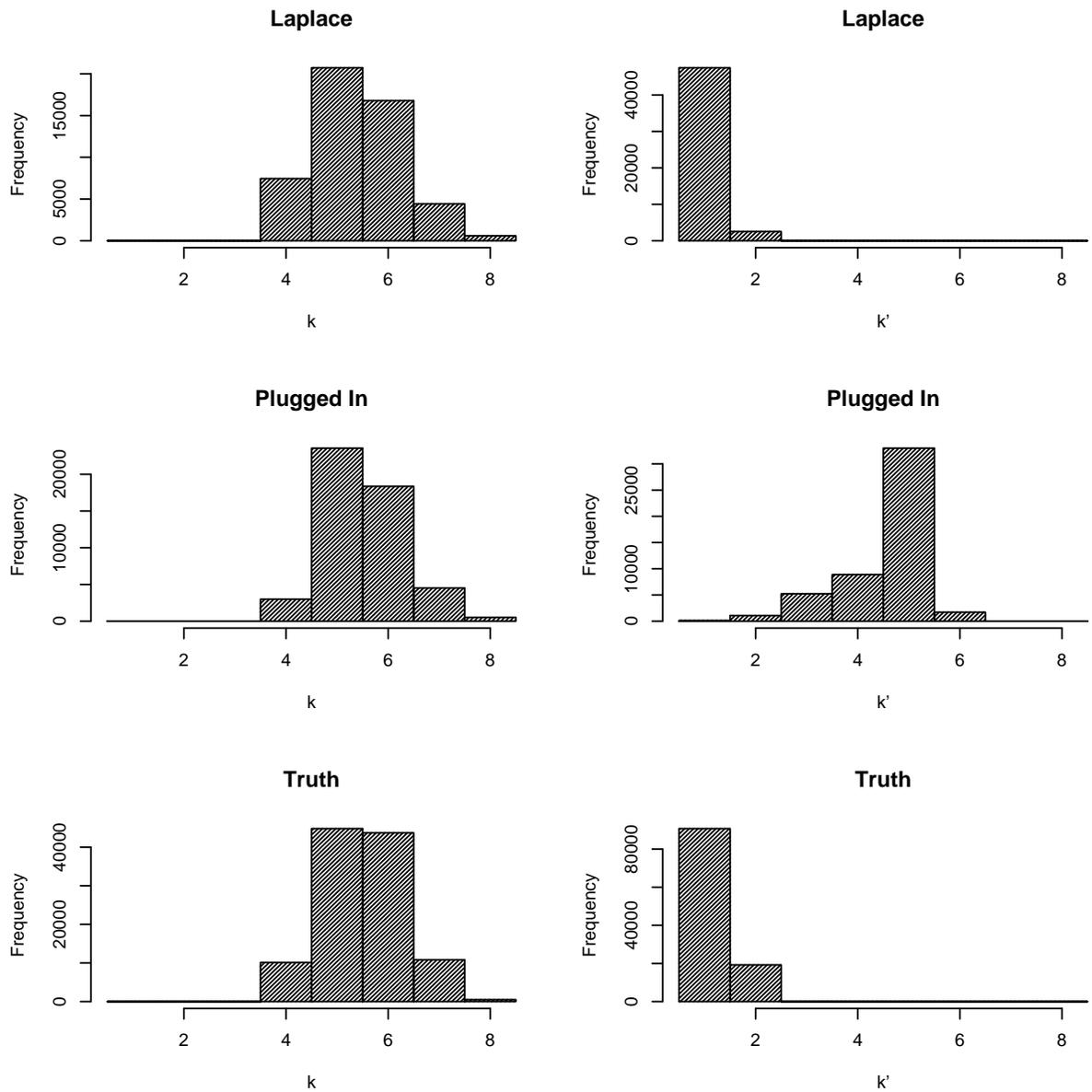
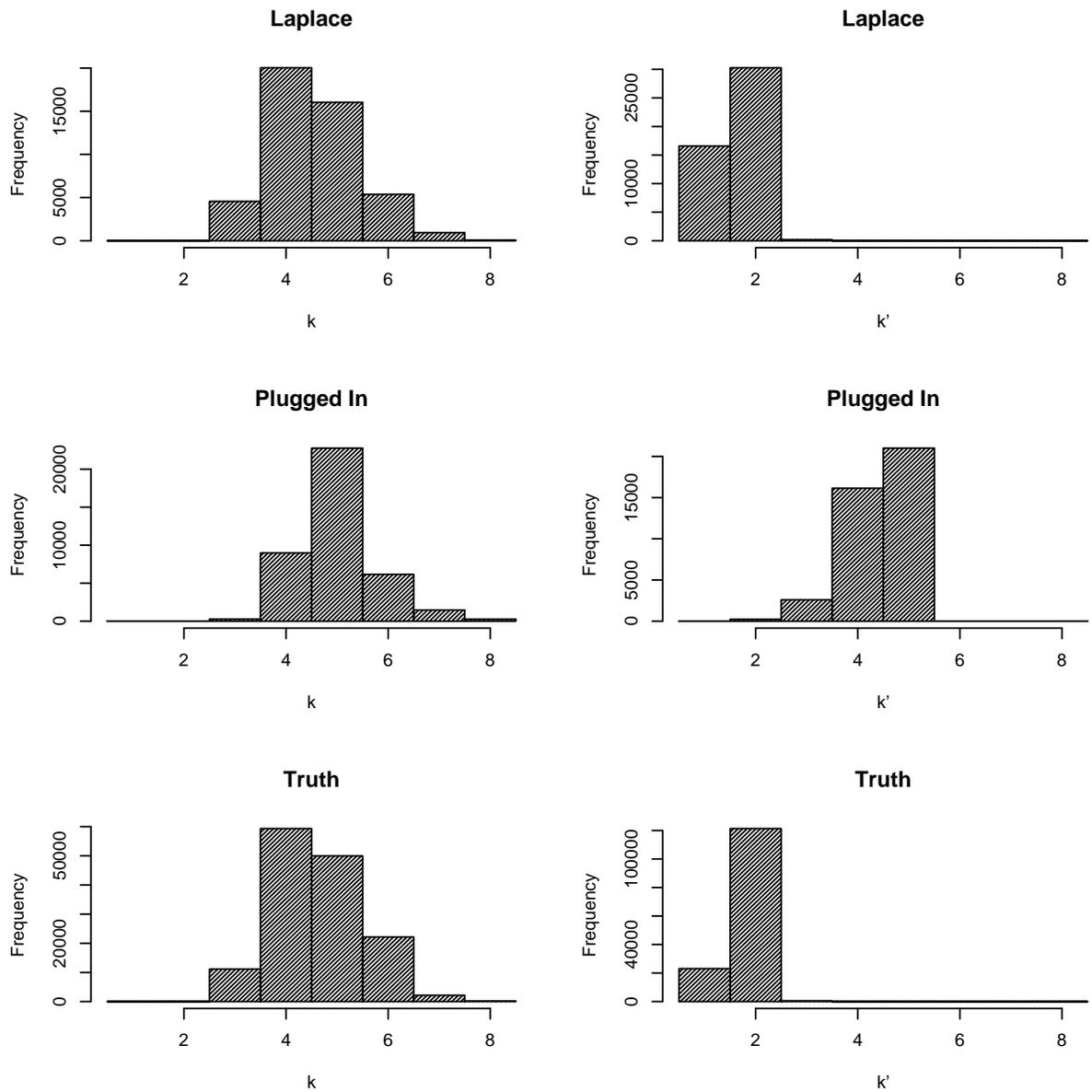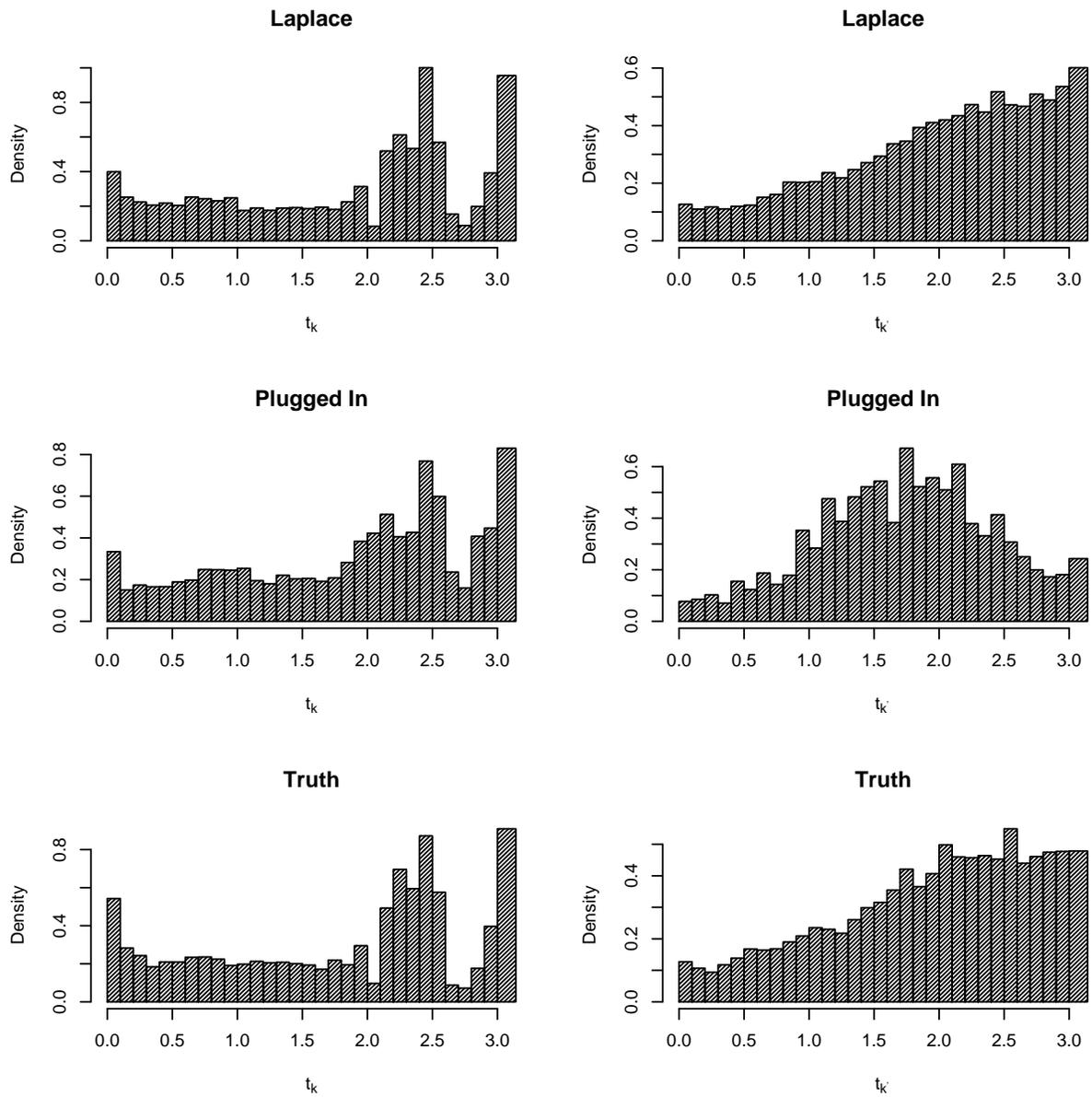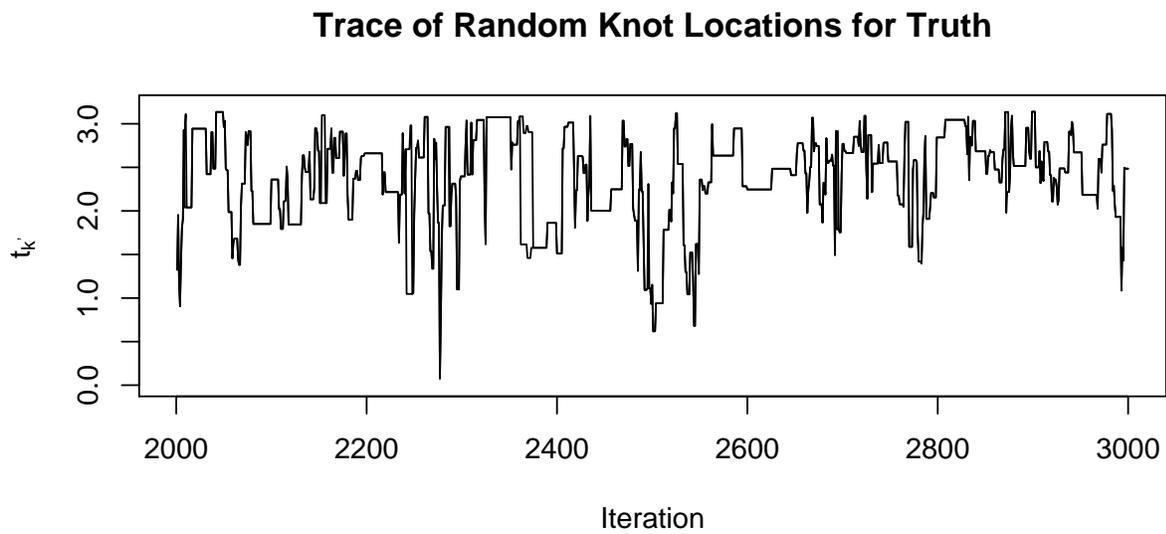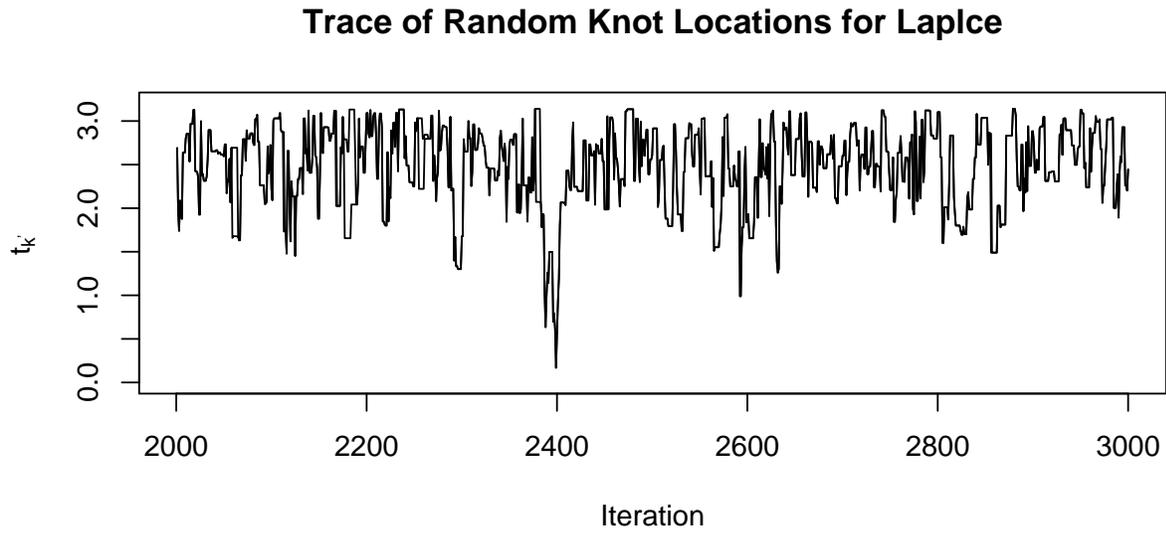Figure 3: Marginal Distributions of k and k' for Model 2 at $n = 25$. Average acceptance rates: Laplace - 65%, Plugged-In - 39%, Truth - 44%.

Figure 4: Marginal Distributions of k and k' for Model 1 at $n = 40$. Average acceptance rates: Laplace - 68%, Plugged-In - 46%, Truth - 39%

Figure 5: Marginal Distributions of k and k' for Model 2 at $n = 40$. Average acceptance rates: Laplace - 59%, Plugged-In - 51%, Truth - 38%.

Figure 6: Marginal Distributions of $\mathbf{t}_k$ and $\mathbf{t}_{k'}$ for Model 1 at $n = 25$.

Trace of Random Knot Locations for Laplce

Trace of Random Knot Locations for Truth

Figure 7: Trace plots of $\mathbf{t}_{k'}$ for the Laplace and true chains for one simulated data set (for iterations 2000 - 3000). Note: The figures trace the positions of **all** elements in $\mathbf{t}_{k'}$, not just the first. Thus, when the trace splits into two lines, a random effect knot has been added.
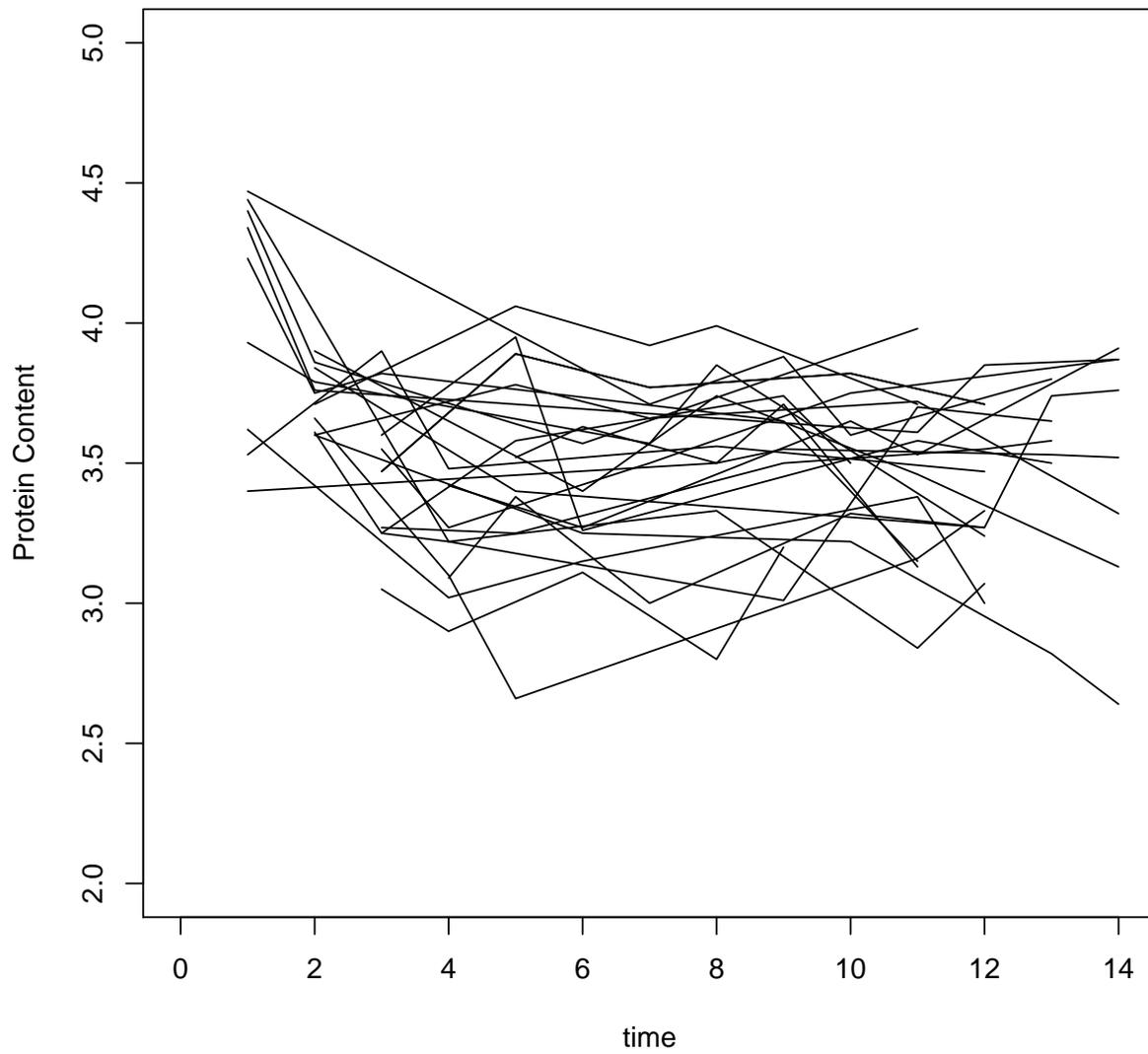
Figure 8: Protein content of milk measured in 25 cows on Barley diet
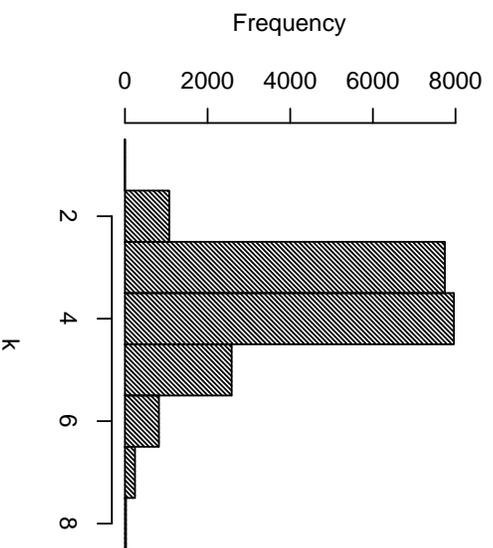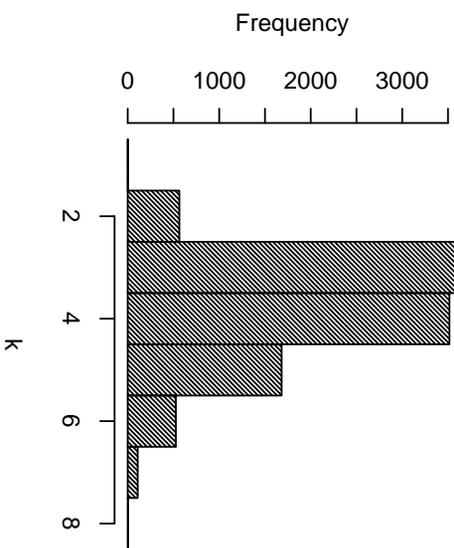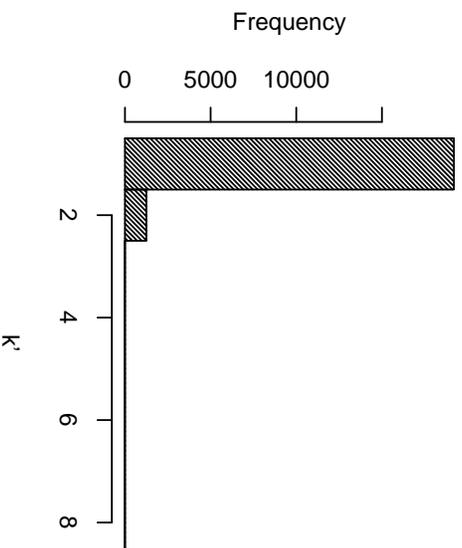
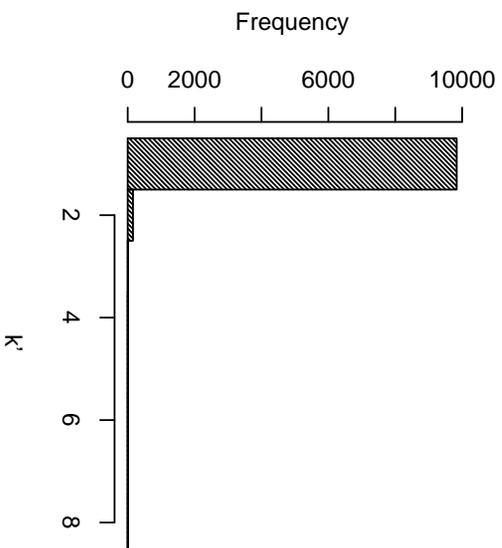Figure 9: Marginal distributions of k and k'

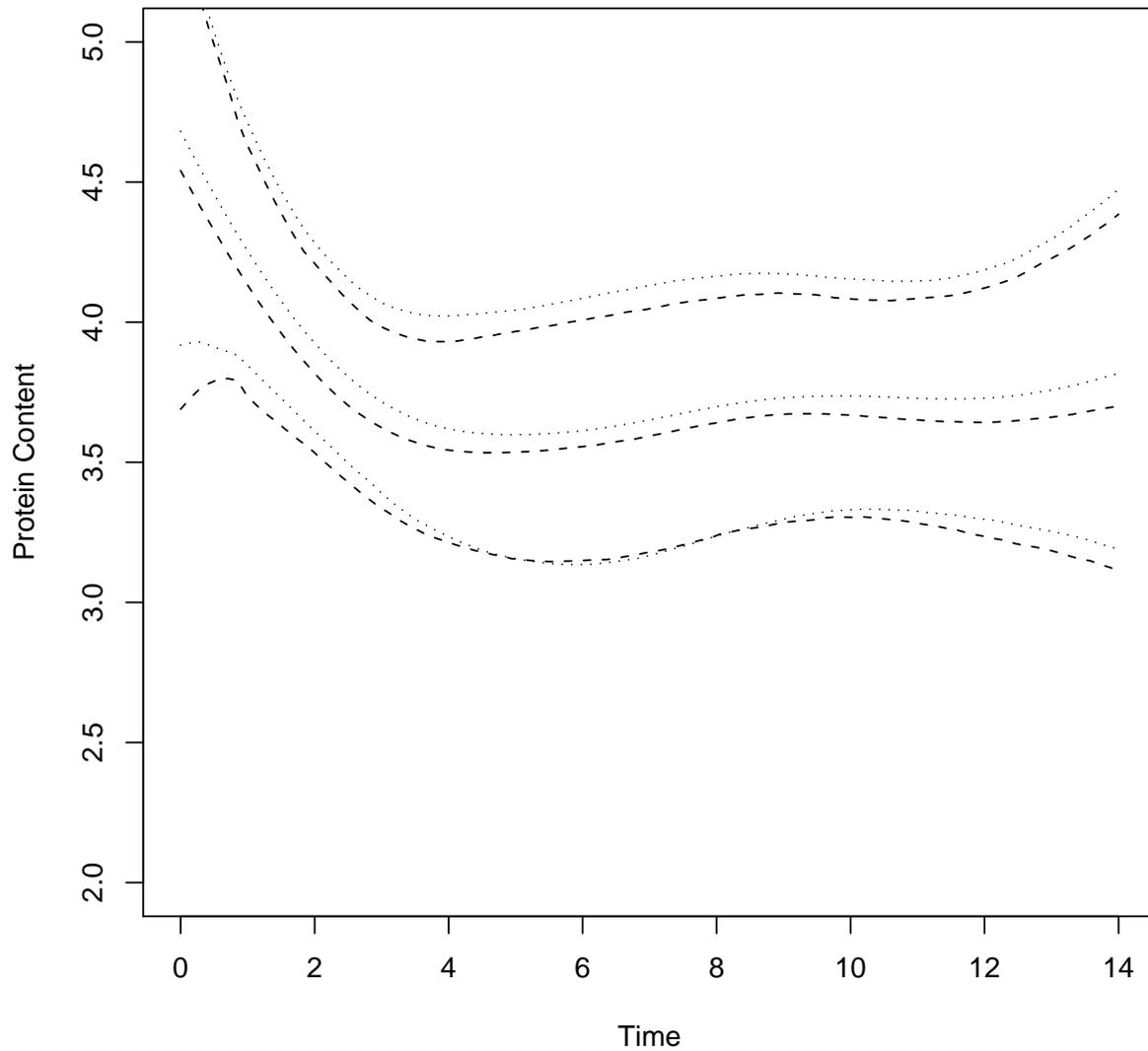Figure 10: Median and 95% pointwise credible intervals for cow-specific curve. (———): Laplace. (··········): Truth.