Abstract of the thesis entitled

On a Topic of Generalized Linear Mixed Models and Stochastic Volatility Model

submitted by Yam, Ho Kwan

for the degree of Master of Philosophy

at The University of Hong Kong in October 2002

Generalized Linear Mixed Models (GLMMs) are extensions to the Generalized Linear Models (GLMs). The inclusion of random effects into the models widens the scope of applicability of GLMMs considerably. However, it also increases the computational effort in estimation. There are various methodologies in making inference on the GLMMs nowadays. We attempt to investigate the use of Gibbs output within the Bayesian framework to carry out the Monte Carlo Approximation of the complicated likelihood function involving random effects by a classical likelihood approach. This methodology is a combination of classical likelihood and Bayesian approaches and it serves as a bridge between them. We will demonstrate this methodology using a famous Salamander Mating data reported by McCullagh and Nelder (1989). Moreover, although the normal distribution plays an important role in statistics, it is not suitable for modeling GLMMs with outlying random effects. The use of a general class of random effects models, such as heavy-tailed distributions should be considered. We will further investi-

gate the use of the Student-t distribution on the random effects in the Salamander Mating data. Furthermore, we suggest to adopt a scale mixture of normal form on the Student-t distribution for simplification of calculation and at the same time, locating the outliers directly through the mixing parameters. Nevertheless, since most financial and economic data exhibits a thick tail behavior, we further investigate the use of heavy tail distributions instead of normal distribution on these kinds of data. A two stage hierarchical scale mixture form on the Student-t distribution will be demonstrated on the Stochastic Volatility (SV) model in a Bayesian aspect.

On a Topic of Generalized Linear Mixed Models and Stochastic Volatility Model

BY

YAM, HO KWAN

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Philosophy
at the University of Hong Kong

Hong Kong, October 2002

DECLARATION

I declare that this thesis represents my own work, except where due acknowledge is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualification.

Signed ________________________________

YAM, HO KWAN

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my supervisors, Dr Boris Choy and Dr Jennifer Chan for their encouragement, support and guidance throughout the years of my study. Moreover, I would like to thank all the friends and staff in the Department of Statistics and Actuarial Sciences. Without anyone of them, I could not have such wonderful years. Most of all, my sincere appreciation goes to my family and Mo for their constant understanding and encouragement.

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 Background

The Generalized linear models (GLMs) generalize the classical linear models to the exponential family of sampling distributions and have an immense impact on both theoretical and practical aspects in statistics. Thereafter, the inclusion of random effects terms into the GLMs opens to the class of Generalized linear mixed models (GLMMs) which overcome the problem of over-dispersion and accommodate population heterogeneity. These models become more applicable in many practical situations. However, the inclusion of random effects complicates the calculation considerably. Diversified methodologies arise in the implementation and estimation in the GLMMs. But, there are still plenty of rooms within the GLMMs framework for further investigation and improvements. In this thesis, we aim to make contributions in improving the methodologies in GLMMs using both likelihood and Bayesian approaches. Moreover, a general class of random effects distributions (symmetric, asymmetric and nonparametric) should be adopted to widen the scope of applicability on the GLMMs. We attempt to investigate the use of heavy-tailed distributions on the random effects as well. Furthermore, we will study the use of scale mixture distribution on the GLMMs and in some financial and economic data to investigate the robust behavior of the resulting estimator.

1.2 Structure of the Thesis

This thesis is divided into two parts. The first part (Chapter 2 - Chapter 6) is focused on the inference in Generalized linear mixed models (GLMMs). The second part (Chapter 7) will investigate the use of scale mixture distribution on the financial and economic data. We will introduce a two stage hierarchical scale mixture form on the Student-t distribution on the stochastic volatility (SV) model.

In chapter 2, we have a brief introduction on the Generalized linear mixed models. In fact, GLMMs are extensions to the Generalized linear models (GLMs) by the inclusion of random effects into the model. However, the computation is complicated after the incorporation of random effects. This arises many estimation methods on GLMMs. In chapter 3, we will investigate the existing methodologies from classical and Bayesian aspects in estimating the GLMMs. They include Maximum likelihood method (ML), Restricted maximum likelihood method (REML), Simulated maximum likelihood method (SML), Penalized quasi-likelihood method (PQL), Monte Carlo Newton Raphson method (MCNR), Monte Carlo EM method (MCEM) and a Bayesian method using Gibbs sampler. In chapter 4, we introduce a famous Salamander Mating data reported by McCullagh and Nelder (1989). This dataset is challenging due to its high dimensional random effects and its crossed design structure. Various methodologies have been proposed on this dataset. We will discuss some of the methods in chapter 4. Since each method has its own advantages and disadvantages, in chapter 5, we further suggest a new method called "Monte Carlo approximation through Gibbs output" in fitting the salamander mating data within a GLMM framework. This newly introduced method is a combination of classical and Bayesian approaches. It can serve as a bridge between these two methods. More details can be found in chapter 5. On the other hand, normal distribution has always been assumed on the random effects of the salamander mating data. However, the data may exhibit a thick tailed behaviour. So, chapter 6 will investigate the use of Student-t distribution on the random effects for the salamander mating data. We further express the Student-t distribution as a scale mixtures of normal form. The use of heavy tailed distribution can capture the outliers and downweigh their effects in statistical analysis. Location of outliers is made possible by the scale mixtures of normal representation.

Chapter 7 constitutes the second part of the thesis focusing on the application of scale mixture distribution on the financial and economic data. The stochastic volatility (SV) model will be introduced in this chapter. Student-t distribution is assumed on an exchange data due to its heavy tailed behaviour. We will further adopt a new two stage hierarchical scale mixture form on the Student-t distribution on the stochastic volatility (SV) model. This can simplify the full conditional distributions and at the same time, speed up the calculation.

CHAPTER 2

GENERALIZED LINEAR MIXED MODELS

2.1 Introduction

One important area in statistics is to develop model that can adequately describe the phenomenon of the data. Several decades ago, researchers developed the simple linear model to study the relationship between response and explanatory variables. The simple expression on this linear model is as follows with $\boldsymbol{Y}$ being the response and $\boldsymbol{X}$ being the explanatory variables:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad \text{where } \boldsymbol{e} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$$

where $\boldsymbol{\beta}$ is a vector of parameters and $\boldsymbol{e}$ is the error term which follows a multivariate normal distribution with mean equals to $\mathbf{0}$ and variance equals to $\sigma^2\boldsymbol{I}$ where $\boldsymbol{I}$ is an identity matrix. This simple linear model forms the base of analysis for the continuous data. However, this model does not work well when there is a non-linear relationship between the response and the covariates. It also cannot deal with non-normal distribution. So, Nelder and Wedderburn (1972) introduced the Generalized linear models (GLMs) which allows an exponential family of distributions on the data in order to describe the non-normal responses. Later on, the

GLMs are further extended to the Generalized linear mixed models (GLMMs) by the inclusion of random effects in the GLMs to account for the serial correlation, to overcome the overdispersion problem, or accommodate population heterogeneity. However, the inclusion of random effects complicates the calculation of the likelihood function considerably, especially for high dimensional random effects models. So, this avoked many diversified methodologies on estimation in the GLMMs.

In this chapter, we will first introduce the GLMs and the exponential family of distributions in the model. Then we will discuss the GLMMs and its difficulties in estimation.

2.2 Generalized linear models (GLMs)

When there is a non-normal distribution on the data $\boldsymbol{y}$, we cannot make inference based on a simple linear model. So, Generalized linear models (GLMs) which allow an exponential family of distributions are needed. These models have unified regression methodology for a wide variety of discrete, continuous, and censored responses that can be assumed to be independent. In GLMs, there are three terms which need to specify. First of all, we assign a distribution on the data. Then the linear predictor $\boldsymbol{X\beta}$ should be defined. Finally, we have to specify the link funtion to link together the mean for the distribution of y_i and the linear form of predictors. The definition on a random variable Y which has a distribution within an exponential family is as follows:

$$\begin{aligned} y_i &\overset{indept.}{\sim} f(y_i) \\ f(y_i) &= \exp\left(\frac{A_i[\theta_i y_i - b(\theta_i)]}{\phi} + c(y_i, \frac{\phi}{A_i})\right) \end{aligned}$$

where $b(.)$ is a function of θ_i which is a canonical parameter and $c(.)$ is a function of y_i and ϕ/A_i. A_i is a known prior weight and ϕ is the dispersion parameter.

There are some properties on this exponential family of distributions.

1. Mean of Y is $\mu = E(\mathrm{Y}) = b'(\theta)$.

2. Variance of Y is $Var(\mathrm{Y}) = \phi b''(\theta)/A = \phi v(\mu)/A$ in which $v(\mu) = b''[(b')^{-1}(\mu)]$ is called the variance function.

3. There are various kinds of exponential family of distributions including Binomial distribution, Poisson distribution, Gamma distribution and Normal distribution, etc.

After defining the exponential family of distributions, the next step will be to define a linear function of predictors. The linear predictor where $i = 1, ..., n$ is defined by the explanatory variables:

$$\eta_i = \boldsymbol{x_i}^T \boldsymbol{\beta}$$

where $\boldsymbol{x}_i$ is a vector of the explanatory variables and $\boldsymbol{\beta}$ is a vector of the regression parameters.

Finally, the mean of the response, μ is linked with the explanatory variables via a monotone link function $g(.)$ as $g(\mu_i) = \eta_i$. There are various kinds of link functions. For binary data, the common link functions are logit-link $g(\mu) = \mu/(1-\mu)$ and probit-link $g(\mu) = \Phi^{-1}(\mu)$ for $0 < \mu < 1$. For Poisson count, we have log-link $g(\mu) = ln(\mu)$. For continuous data, we can have an identity-link as $g(\mu) = \mu$. Each member of the exponential family of distributions adopts different canonical link function. Table 2.1 shows some of the relationship between the distributions and their corresponding link functions.

======================================

Table 2.1 about here

======================================

In order to evaluate the parameters in these GLMs, traditionally researchers used the maximum likelihood method by differentiating the log-likelihood function with respect to the parameters $\boldsymbol{\beta}$. More details can be found in McCullagh and Nelder (1989).

Although the GLMs solve the problem of non-normality in the data, it cannot deal with data which are clustered or longitudinal where overdispersion and autocorrelation are often of concern. So, it is further extended to Generalized linear mixed models (GLMMs).

2.3 Generalized linear mixed models (GLMMs)

Over the past twenty years, the GLMs have been improved and modified to a more general class of models, known as the Generalized linear mixed models (GLMMs) by the inclusion of random effects terms into the GLMs. Generalized linear mixed models are extension of the GLMs. The term 'mixed' in the GLMMs means that we mix the random effects with the fixed effects together to get a modified model which can overcome the overdispersion in the data and at the same time, accommodate the population heterogeneity. The main difference in the structure of GLMMs as compared with GLMs is the incorporation of the random effect, term $\boldsymbol{u}$, into the linear predictor. Thus, the model on an independent response y_i where $i = 1, ..., n$ which follows an exponential family of distribution given the random effects $\boldsymbol{u}$ becomes:

$$\begin{aligned} f(y_i|\boldsymbol{u}) &= \exp\left(\frac{A_i[\theta_i y_i - b(\theta_i)]}{\phi} + c(y_i, \frac{\phi}{A_i})\right) \\ E[y_i|\boldsymbol{u}] &= \mu_i \\ g(\mu_i) &= \boldsymbol{x_i}^T\boldsymbol{\beta} + \boldsymbol{z_i}^T\boldsymbol{u} \end{aligned}$$

where $\boldsymbol{x}$ is a vector of covariates, $\boldsymbol{\beta}$ is a vector of fixed effect parameters, $\boldsymbol{z}$ is the design vector for the random effects, $\boldsymbol{u}$ is a vector of the random effect parameters and $g(.)$ is a link function defined in the previous section. The response in the GLMMs depends on the random effects. Furthermore, we assign a distribution to the random effects, ie. $\boldsymbol{u} \sim h(\boldsymbol{u})$. For simplicity, researchers usually assign a multivariate normal distribution with mean $\mathbf{0}$ and variance $\sigma^2\boldsymbol{I}$ for the random effects. For robustness consideration, other heavier-tailed distributions can also be used. This will be discussed in chapter 6.

Concerning the methodologies in estimating the parameters in GLMMs, there are two main approaches: Classical approach and Bayesian approach. In classical approach, the parameters are commonly estimated by the maximum likelihood method. Since the random effects are unobserved, the likelihood function in the GLMMs involves integration. The likelihood function in the GLMMs is defined as follows:

$$
\begin{aligned}
lik(\theta) &= \int \prod_{i=1} f(y_i, \boldsymbol{u}) d\boldsymbol{u} \\
&= \int \prod_{i=1} f(y_i|\boldsymbol{u}) h(\boldsymbol{u}) d\boldsymbol{u}
\end{aligned}
$$

Hence, the inclusion of random effects into the model complicates the computation of likelihood function considerably. If the integral of random effects in the likelihood function is of low dimension or can be factorised as a product of low-dimensional integrals, the numerical integration may not be too difficult. However, if the random effects are of high dimensional, there exists problem in the evaluation of the corresponding high dimensional integral, especially when the random effects are crossed rather than nested so that the factorization of the integral is not possible.

As a result, diversified methodologies in the implementation and estimation on the GLMMs have been proposed and there are still plenty of room within this framework for further investigation and improvement. For example, McCullagh and Nelder (1989) adopted an estimating equations approach in estimating the parameters in the GLMMs. Breslow and Clayton (1993) and Lin and Breslow (1996) adopted an uncorrected and corrected penalized quasi-likelihood approaches respectively. On the other hand, Liu and Pierce (1994) suggested a Gaussian quadrature method for low-dimensional random effects. McCulloch (1997) further studied a simulated ML approach. Various authors have also investigated the use of EM algorithm or Monte Carlo EM (MCEM) method (Wei and Tanner, 1990, McCulloch, 1997, Chan and Kuk, 1997 and Booth and Hobert, 1999). Other authors studied the use of Monte Carlo Newton Raphson (MCNR) approach (Kuk and Cheng, 1997, 1999 and McCulloch, 1997). MCEM and MCNR are extensions to the EM and NR method respectively in approximating the likelihood by a Monte Carlo method. Moreover, there are researchers adopting Laplace expansion method (Shun and McCullagh, 1995 and Shun ,1997) and Laplace importance sampling method (Kuk, 1999).

Apart from these kinds of classical approaches, Zeger and Karim (1991), Karim and Zeger (1992) and Clayton (1996) investigated the estimation method in GLMMs by a Bayesian approach. They studied the use of Markov Chain Monte Carlo (MCMC) methods, in particular, the Gibbs sampler (Gelfand and Smith, 1990). Some of the standard methods will be further discussed in chapter 3.

Distribution	Binomial(m, π)	Poisson(λ)	Gamma$(r, 1/v)$	Normal(μ, σ^2)
link function	logit	log	reciprocal	identity
$g(\mu)$	$log(\mu/(1-\mu))$	$log(\mu)$	$1/\mu$	μ

Table 2.1. Examples of canonical links on various distributions

CHAPTER 3

COMPUTATION ON THE GENERALIZED LINEAR MIXED MODELS

3.1 Introduction

In the previous chapter, we described the Generalized linear mixed models for correlated data. Since the inclusion of random effects complicated the computation of the likelihood function considerably, a lot of researchers have proposed different methodologies to solve the problem. Thus, in this chapter, we study some of the classical methods, namely, maximum likelihood method (ML), restricted maximum likelihood method (REML), simulated maximum likelihood method (SML), penalized quasi-likelihood method (PQL), Monte Carlo EM method (MCEM) and Monte Carlo Newton-Raphson (MCNR) method. Moreover, we will investigate the use of Gibbs sampling method in Bayesian approach on the GLMMs.

First of all, we will discuss the difference between Bayesian and Classical approaches. Afterwards, we will introduce and investigate the use of various Bayesian and Classical approaches. Some comments on the methods will be given too.

3.2 Bayesian and Classical Approaches

Suppose we have a set of observations y_i where $i = 1, 2, ..., n$. We are interested in the parameter estimate $\boldsymbol{\theta}$. We denote the density of y_i as $f(y_i|\boldsymbol{\theta})$. Then the joint density function is expressed as:

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\theta})$$

There are two main approaches, namely Classical and Bayesian approaches for various methodologies in making statistical inference.

In classical inference, we concern about the likelihood function $lik(\boldsymbol{\theta}|\boldsymbol{y}) = f(\boldsymbol{y}|\boldsymbol{\theta})$. The parameter estimate $\boldsymbol{\theta}$ is treated as fixed but unknown. We differentiate the log-likelihood function, $\ell(\boldsymbol{\theta}|\boldsymbol{y})$ to obtain the parameter estimate $\boldsymbol{\theta}$ which maximizes the likelihood function of the observed data.

However, it is difficult to evaluate the marginal likelihood function when this likelihood involves high dimensional intergral. Various methodologies were proposed in the computation of the likelihood function and hence the maximum likelihood estimates.

In Bayesian inference, based on a density function on $\boldsymbol{y}$, namely $f(\boldsymbol{y}|\boldsymbol{\theta})$ and a prior distribution on the parameter $\boldsymbol{\theta}$, namely $\pi(\boldsymbol{\theta})$, we calculate the posterior density, $p(\boldsymbol{\theta}|\boldsymbol{y})$ as

$$p(\boldsymbol{\theta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int f(\boldsymbol{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

That means we have to collect prior information on the parameter $\boldsymbol{\theta}$ and then assign a suitable prior density $\pi(\boldsymbol{\theta})$ to the parameter $\boldsymbol{\theta}$ in order to construct the posterior density. The parameter $\boldsymbol{\theta}$ is treated as random variables in the Bayesian approach. More information on the Bayesian statistics can be found in Lee (1999) and Bernardo and Smith (1994).

However, there exists argument on the specification of prior density $\pi(\boldsymbol{\theta})$. In some cases, conjugate prior is chosen just for convenience. In other occasions, the choice of prior can be rather subjective. Moreover, it may also be difficult to evaluate the posterior density and hence to obtain the posterior mean as the Bayesian estimate.

In the following sections, we will investigate the commonly used methodologies in both Classical and Bayesian approaches.

3.3 Classical approach

3.3.1 Maximum Likelihood method (ML)

Maximum Likelihood method is the traditional methodology in parameter estimation. With the inclusion of random efects in GLMMs, we obtain the marginal likelihood function by integrating over the unobserved random effects.

$$Lik(\boldsymbol{\theta}) = \int f(\boldsymbol{y}, \boldsymbol{u}|\boldsymbol{\theta})d\boldsymbol{u} = \int f(\boldsymbol{y}|\boldsymbol{u}, \boldsymbol{\beta})h(\boldsymbol{u}|\boldsymbol{\tau})d\boldsymbol{u}$$

where $\boldsymbol{y}$ is the responses and $\boldsymbol{u}$ is the random effects with density $h(\boldsymbol{u}|\boldsymbol{\tau})$. $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\tau})$ are our parameters which include parameter for fixed effects, $\boldsymbol{\beta}$, and parameter for random effects, $\boldsymbol{\tau}$.

The algorithm for evaluating the ML estimates are:

1. Take logarithm of the likelihood function $Lik(\boldsymbol{\theta})$ given above

$$\ell(\boldsymbol{\theta}|\boldsymbol{y}) = ln \int \prod_{i,j} f(y_{ij}|u_i; \boldsymbol{\beta})h(u_i; \boldsymbol{\tau})du_i$$

2. Differentiate the log-likelihood function $\ell(\boldsymbol{\theta}|\boldsymbol{y})$ with respect to the parameters $\boldsymbol{\theta}$ and set it to be zero to solve for the ML estimate $\hat{\boldsymbol{\theta}}$.

$$\frac{\delta \ell}{\delta \boldsymbol{\theta}}\Big|_{(\boldsymbol{\theta}=\hat{\boldsymbol{\theta}})} = 0$$

Then $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimate which maximizes the log likelihood function $\ell(\boldsymbol{\theta}|\boldsymbol{y})$. Many researchers further improve this method to form various kinds of modified estimation method. Schall (1991) studied the use of maximum likelihood approach in Generalized linear models with random effects. However, the numerical integration method which he used is only appropriate for simple cases in which the likelihood function involves only integral of low dimension or such integral can be factorized into a product of low dimensional integrals.

3.3.2 Restricted Maximum Likelihood method (REML)

Restricted Maximum Likelihood method is an extension to the Maximum likelihood method. It is mainly for estimating the variance components. It maximizes the likelihood of linear combinations of elements $\boldsymbol{y}$, that is $\boldsymbol{K}^T\boldsymbol{y}$ where $\boldsymbol{K}$ is any chosen vector such that $\boldsymbol{K}^T\boldsymbol{y}$ contains none of the fixed effects. This implies that $\boldsymbol{K}$ is chosen to make $\boldsymbol{K}^T\boldsymbol{X} = 0$. After making the adjustment on the data $\boldsymbol{y}$, we estimate from

$$\boldsymbol{K}^T\boldsymbol{Y} \sim N(\boldsymbol{0}, \boldsymbol{K}^T\boldsymbol{V}\boldsymbol{K}) \qquad \text{instead of} \qquad \boldsymbol{Y} \sim N(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V})$$

where $\boldsymbol{V}$ is the variance matrix for data $\boldsymbol{Y}$ and $\boldsymbol{X}$ is the explanatory variables. Following similar procedures as in Maximum likelihood method by differentiating the likelihood function of $\boldsymbol{K}^T\boldsymbol{y}$ with respect to the variance components, we can get the estimates on the variance components easily which are invariant to the values of fixed effects.

However, in REML, it cannot provide estimates for $\boldsymbol{\beta}$. More information are given in Schall (1991) and McCulloch and Searle (2001).

3.3.3 Penalized Quasi-Likelihood method (PQL)

Wedderburn (1974) first developed the quasi-likelihood method. In this method, a quasi-likelihood function is constructed with fewer assumptions than the likelihood function. In constructing the likelihood function, we need to specify a distribution for the data $\boldsymbol{y}$. However, in constructing the quasi-likelihood function, we only need to specify a relationship between the mean and variance of the data, without knowing its underlying distribution. A conditional logarithm of quasi-likelihood function Q_i of parameters $\boldsymbol{\beta}$ given the random effects $\boldsymbol{u}$ is

$$lnQ_i(\boldsymbol{\theta}|\boldsymbol{u}) = \int_{y_i}^{\mu_i} \frac{A_i(y_i - t)}{\phi v(t)} dt$$

where A_i is a known prior weight, ϕ is a dispersion parameter and $v(\mu)$ is called the variance function specified in the previous chapter on the description of the exponential family.

Some researchers further included a term into the quasi-likelihood function to form the penalized quasi-likelihood (PQL) method. Suppose the random effects follow a normal distribution with mean $\mathbf{0}$ and a variance-covariance matrix $\boldsymbol{D}$.

The penalized quasi likelihood function for $\boldsymbol{\theta}$ is

$$
\begin{aligned}
PQL(\boldsymbol{\theta}) &\approx \int \exp\left(\sum_{i=1}^{n} f(y_i|u_i) - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{D}^{-1}\boldsymbol{u}\right) d\boldsymbol{u} \\
&\approx \int \exp\left(\sum_{i=1}^{n} Q_i(\boldsymbol{\theta}|\boldsymbol{u}) - \frac{1}{2}\boldsymbol{u}^T\boldsymbol{D}^{-1}\boldsymbol{u}\right) d\boldsymbol{u}
\end{aligned}
$$

where $\frac{1}{2}\boldsymbol{u}^T\boldsymbol{D}^{-1}\boldsymbol{u}$ is the 'penalized' term added into the quasi-likelihood function. This term is used to prevent arbitary values of $\boldsymbol{u}$ from being selected. More details can be found in Green (1990) and Wolfinger (1993).

Breslow and Clayton (1993) studied the use of PQL in GLMMs. However, the estimates are biased towards zero for some variance components. Thus, Lin and Breslow (1996) suggested a bias-corrected PQL. This improves the asymptotic performance of PQL estimates, but inflates the variance. So, we need to be careful in dealing with the problem of bias in PQL.

3.3.4 Simulated Maximum Likelihood method (SML)

Geyer and Thompson (1992) and Gelfand and Carlin (1993) suggested the use of Simulated Maximum Likelihood method. McCulloch (1997) studied the use of this method on the GLMMs. In SML method, the likelihood is estimated directly without considering the log-likelihood function by simulation.

$$
Lik(\boldsymbol{\theta}|\boldsymbol{y}) = \int f(\boldsymbol{y}, \boldsymbol{u}|\boldsymbol{\theta}) d\boldsymbol{u}
$$

$$
\begin{aligned}
&= \int f(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\beta})g(\boldsymbol{u};\boldsymbol{\tau})d\boldsymbol{u} \\
&= \int \frac{f(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\beta})g(\boldsymbol{u};\boldsymbol{\tau})}{h(\boldsymbol{u})}h(\boldsymbol{u})d\boldsymbol{u} \\
&\approx \frac{1}{M}\sum_{k=1}^{M}\frac{f(\boldsymbol{y}|\boldsymbol{u}_k;\boldsymbol{\beta})g(\boldsymbol{u}_k;\boldsymbol{\tau})}{h(\boldsymbol{u}_k)}
\end{aligned}
$$

where M is the total number of simulated values, $h(\boldsymbol{u})$ is the importance sampling function and $\boldsymbol{u}$ is a vector of random effects simulated from this distribution by any sampling technique. Theoretically, the estimates are independent of the choice of importance samplng function, $h(\boldsymbol{u})$. They are calculated numercially based on the likelihood function approximated using simulations.

However, the efficiency of estimates depends on the choice of importance sampling function. If the importance function $h(\boldsymbol{u})$ in SML is far away from the density of the random effects, $g(\boldsymbol{u}|\boldsymbol{\tau})$, the resulting estimator may be inefficient.

3.3.5 Monte Carlo Netwon Raphson method (MCNR)

Newton Raphson method is a popular iterative method to find the maximum likelihood estimates. We denote $l(\boldsymbol{\theta};\boldsymbol{y}) = ln[f(\boldsymbol{y};\boldsymbol{\theta})]$ be the log-likelihood function on the data $\boldsymbol{y}$ and the parameter space $\boldsymbol{\theta}$. Then $l'(\boldsymbol{\theta};\boldsymbol{y})$ and $l''(\boldsymbol{\theta};\boldsymbol{y})$ are the first and second order derivatives of $l(\boldsymbol{\theta};\boldsymbol{y})$. In each Newton Raphson iteration, current parameter estimates in the k-th iteration $\boldsymbol{\theta}^{(k)}$ can be updated to the $(k+1)$-st iteration $\boldsymbol{\theta}^{(k+1)}$ by

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - [l''(\boldsymbol{\theta}^{(k)}; \boldsymbol{y})]^{-1} l'(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}) \tag{3.1}$$

and the procedure continues until convergence is achieved. Since for GLMMs, the likelihood function and hence its derivatives may be difficult to evaluate in the Newton Raphson procedures, Kuk and Cheng (1997) proposed the use of Monte Carlo Newton Raphson method in calculating the estimates in the GLMMs. The Monte Carlo algorithm requires the random effects $\boldsymbol{u}$ being simulated from a conditional function on $\boldsymbol{u}$ given the observed $\boldsymbol{y}$ and the current estimate $\boldsymbol{\theta}^{(k)}$. The algorithm of using the MCNR is as follows:

Algorithm:

1. Choose a starting value for $\boldsymbol{\theta}^{(0)}$.

2. Simulate $\boldsymbol{u}_i$ where $i = 1, ..., M$ with M being the total number of simulations from a conditional distribution of $f(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)})$ based on $\boldsymbol{\theta}^{(k)}$ which is the current parameter estimates.

3. Use the simulated $\boldsymbol{u}_i$ to approximate the likelihood by Monte Carlo approximation, we have

$$l'_M(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M} l'(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}, \boldsymbol{u}_i) \tag{3.2}$$

$$l''_M(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}) = \frac{1}{M}\sum_{i=1}^{M} l''(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}, \boldsymbol{u}_i) + \tag{3.3}$$

$$\left(\frac{1}{M}\sum_{i=1}^{M} l'(\boldsymbol{\theta}^{(k)};\boldsymbol{y},\boldsymbol{u}_i)l'^T(\boldsymbol{\theta}^{(k)};\boldsymbol{y},\boldsymbol{u}_i) - l'_M(\boldsymbol{\theta}^{(k)};\boldsymbol{y}){l'_M}^T(\boldsymbol{\theta}^{(k)};\boldsymbol{y})\right)$$

Replacing the terms $l'(\boldsymbol{\theta}^{(k)};\boldsymbol{y})$ and $l''(\boldsymbol{\theta}^{(k)};\boldsymbol{y})$ in Newton Raphson iteration given in equation (3.1) by $l'_M(\boldsymbol{\theta}^{(k)};\boldsymbol{y})$ and $l''_M(\boldsymbol{\theta}^{(k)};\boldsymbol{y})$ respectively given by equations (3.2) and (3.3) to get the Monte Carlo Newton Raphson iteration. We will get a new estimate of $\boldsymbol{\theta}^{(k+1)}$.

4. Repeat steps 2-3 until convergence is achieved.

More informatin on this MCNR is given in Kuk and Cheng (1997) and Kuk and Cheng (1999). They showed that the convergent rate for this MCNR was faster than that of Monte Carlo EM. So, it is computationally more efficient.

3.3.6 Monte Carlo EM method (MCEM)

The EM algorithm (Dempster et al., 1977) is an iterative method for the computation of maximizer on the posterior density. It includes an E-step in expectation and then follows an M-step in maximization. The basic idea for an EM algorithm is that given an observed data $\boldsymbol{y}$, we assume there are some missing data $\boldsymbol{u}$. In the E-step, we compute the expectation on the $log[f(\boldsymbol{y},\boldsymbol{u}|\boldsymbol{\theta})]$ with respect to the conditional distribution, $g(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(i)})$ where $\boldsymbol{\theta}^{(i)}$ is the current approximated estimates. Afterwards, we get a maximizer of the conditional expectation in the M-step. The conditional distribution of $\boldsymbol{u}$, $g(\boldsymbol{u}|\boldsymbol{y},\boldsymbol{\theta}^{(i+1)})$ is updated using the new maximizer and the algorithm is iterated until convergence is obtained.

In evaluating the estimates in GLMMs, we can treat the random effects $\boldsymbol{u}$ as the missing data and apply the EM algorithm. However, the expectation is too difficult when the density of data cannot be written in a closed form. Wei and Tanner (1990) and McCulloch (1994) suggested the use of Monte Carlo EM algorithm where the E-step is implementated by a Monte Carlo method.

Algorithm:

1. Choose a starting value for $\boldsymbol{\theta}^{(0)}$.

2. Simulate $\boldsymbol{u}_i$ where $i = 1, ..., M$ with M being the total number of simulations from the conditional distribution of $g(\boldsymbol{u}|\boldsymbol{y}, \boldsymbol{\theta}^{(k)})$ using Metropolis Hastings or Rejection sampling method based on $\boldsymbol{\theta}^{(k)}$ which is the current parameter estimates.

3. Approximate the expectation, $E[log\{f(\boldsymbol{y}, \boldsymbol{u}|\boldsymbol{\theta})\}]$ by a Monte Carlo method

$$\frac{1}{M}\sum_{i=1}^{M} log\{f(\boldsymbol{y}, \boldsymbol{u}^{(i)}|\boldsymbol{\theta})\}$$

4. Maximize the approximated expectation to obtain an updated MLE $\boldsymbol{\theta}^{(k+1)}$.

5. Repeat steps 2-4 until convergence is achieved.

More information on the MCEM algorithm can be found in Chan and Kuk (1997), McCulloch (1997) and Booth and Hobert (1999).

3.4 Bayesian approach

3.4.1 Gibbs sampler

Bayesian approach is an alternative method to Classical approach. In order to get the Bayesian estimates, we have to specify a prior distribution on each parameter and then get the posterior mean of each parameter from its conditional distribution by Markov Chain Monte Carlo (MCMC) method, eg. Gibbs sampler or Metropolis Hasting algorithm. Clayton (1996) and Zeger and Karim (1991) had studied the use of Gibbs sampling approach on the GLMMs.

Suppose we have three variables X, Y, and Z. We define the joint distribution as $[X, Y, Z]$ which is a complicated distribution with the conditional distribution of X, Y and Z as $[X|Y, Z]$, $[Y|X, Z]$ and $[Z|X, Y]$ respectively which are comparatively simpler.

The algorithm of Gibbs sampler is as follows:

1. Started with an initial value of $X^{(0)}$, $Y^{(0)}$ and $Z^{(0)}$.

2. Draw $X^{(1)}$ from the conditional distribution of $[X|Y^{(0)}, Z^{(0)}]$.

3. Draw $Y^{(1)}$ from the conditional distribution $[Y|X^{(1)}, Z^{(0)}]$ based on $Z^{(0)}$ and the newly simulated $X^{(1)}$.

4. Draw $Z^{(1)}$ from $[Z|X^{(1)}, Y^{(1)}]$ based on the newly simulated $X^{(1)}$ and $Y^{(1)}$.

5. Complete the first iteration. We will repeat this algorithm until N iterations have completed and the simulated values converged to the joint density function.

Suppose totally we have N iterations, the first K iterations in the burn-in period before convergence will be discarded. The remaining $M = N - K$ iterations are used to form posterior samples. The parameter estimates are the posterior sample means. Then, we need to check for convergence and auto-correlation by plotting the series of the simulated values and examining their auto-correlation function respectively for each variable. Geman and Geman (1984) and Gelfand and Smith (1990) had discussed the Gibbs sampler in details.

In applying the Gibbs sampler on GLMMs, we have to specify the joint distribution and the corresponding full conditional distribution for each variables. For example, let $\boldsymbol{y}$ be the observed data having joint density function $f(.)$, $\boldsymbol{\beta}$ be a vector of fixed effect parameters following a b-dimensional multivariate normal distribution with mean $\boldsymbol{\mu}_\beta$ and a variance-covariance matrix of $\boldsymbol{\Sigma}_\beta$ while $\boldsymbol{u}$ be a vector of random effect parameters following a p-dimensional multivariate normal distribution with mean $\mathbf{0}$ and a variance-covariance matrix of $\boldsymbol{D}$. We have the following hierarchical model:

$$\begin{aligned} \boldsymbol{y} &\sim f(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{u}) \\ \boldsymbol{u} &\sim MVN_p(\mathbf{0}, \boldsymbol{D}) \end{aligned}$$

$$
\begin{aligned}
\boldsymbol{\beta} &\sim MVN_b(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta) \\
\boldsymbol{D} &\sim Wishart(\boldsymbol{R}_p, p) \\
\boldsymbol{\Sigma}_\beta &\sim Wishart(\boldsymbol{R}_b, b)
\end{aligned}
$$

where $\boldsymbol{\mu}_\beta$ is a fixed vector and $\boldsymbol{R}_p$ and $\boldsymbol{R}_b$ are p and b dimensional fixed matrices for $\boldsymbol{D}$ and $\boldsymbol{\Sigma}_\beta$ respectively.

The joint density is $[\boldsymbol{y}, \boldsymbol{u}, \boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{\Sigma}_\beta]$. The Bayesian estimates are drawn from its conditional distribution by Gibbs sampler. For example, we sample

$\boldsymbol{\beta}$ from $[\boldsymbol{\beta}|\boldsymbol{u}, \boldsymbol{\Sigma}_\beta, \boldsymbol{D}, \boldsymbol{y}]$

$\boldsymbol{u}$ from $[\boldsymbol{u}|\boldsymbol{\beta}, \boldsymbol{\Sigma}_\beta, \boldsymbol{D}, \boldsymbol{y}]$

$\boldsymbol{D}$ from $[\boldsymbol{D}|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{\Sigma}_\beta, \boldsymbol{y}]$

$\boldsymbol{\Sigma}_\beta$ from $[\boldsymbol{\Sigma}_\beta|\boldsymbol{\beta}, \boldsymbol{u}, \boldsymbol{D}, \boldsymbol{y}]$

Sometimes, the conditional distributions are not in standard form. So, we need to use some non-standard random variates sampling approaches such as Metropolis Hastings (Hastings, 1970; Metropolis et al., 1953) or adaptive rejection sampling (Ripley, 1987).

CHAPTER 4

SALAMANDER MATING DATA, GLMM

4.1 Introduction

There are many datasets which are modelled by GLMMs. In this chapter, we introduce a famous salamander mating data reported by McCullagh and Nelder (1989). Many researchers had adopted various methodologies within the GLMM framework to analyse this salamander mating data. We will give a brief overview of these methodologies in this chapter. Then we propose a new methodology for analysing this famous salamander data and then compare the results with others in the next chapter.

4.2 Salamander Mating Data

Salamander mating experiment was conducted by S. Arnold and P. Verrell of the Department of Ecology and Evolution at the University of Chicago. The salamanders came from two populations: Rough Butt (RB) and Whiteside (WS). The objective of this experiment was to investigate whether there were barriers to interbreeding in the salamanders from these two geographically isolated populations.

There were totally three experiments. Each experiment involved 20 female and 20 male salamanders. Each female salamander in the experiment was paired with 6 male salamanders, at which 3 male salamanders came from Rough Butt (RB) and another 3 from Whiteside (WS). Totally, there were 120 observations of mating results per experiment. The mating design was presented in Table 4.1 (with salamanders from RB are numbered from 1-10 and salamanders from WS are numbered from 11-20). For example, the first female salamander was paired with the 1st, 4th and 5th male salamanders from RB and the 11th, 14th and the 15th male salamanders from WS. There exists a crossed structure in the design of experiment. The first experiment was done in summer of 1986. The second experiment was carried out in fall in the same year with the same set of animals. The third experiment was carried out at the same time as the second experiment but with a new set of salamanders.

The percentage of successful mating per experiment is shown in Table 4.2. The success rates for the salamanders in pairs of 'RB female and RB male' and 'WS female and WS male' are over 60 % in these experiements. It seems that there is a higher successful rate when the pair of salamanders involved in the mating live in the same population. For the pair of salamanders 'RB female and

WS male', the success rates of mating are from 46 % to 66 % in the experiments which are comparatively lower than the success rates of mating for salamanders involved in the same populations. The lowest success rate (around 20 %) occurred in the pair of 'WS female and RB male'. On the other hand, the total number of success is the highest in summer while the number of success for two different sets of salamanders in fall are similar.

====================================

Table 4.1 and Table 4.2 about here

====================================

4.3 Modelling the Salamander Mating Data

It is clear that the responses (1 if the mating is successful, 0 if the mating is unsuccessful) are not independent but correlated since each female salamander is paired with six male salamanders. We incorporate the female and male random effects in the data to deal with the problem of overdispersion. Thus, the salamander mating data are modelled within a GLMM framework.

In the Generalized linear mixed model with a logit link function, the likelihood function of the observed $\boldsymbol{y}$ given the vector of parameters is

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = \prod_{t=1}^{120} \int \cdots \int \frac{\exp(\eta_t y_t)}{1+\exp{(\eta_t)}} h(\boldsymbol{z}) d\boldsymbol{z}$$

where η_t is a linear predictor and $\boldsymbol{z}$ is a vector of 40 dimensional random effects for male and female salamanders. Various authors proposed different forms of linear predictors which may include the female and male population effects, their interaction, the seasonal effects as well as female and male random effects. An example of linear predictor which is used in this research is given in the next chapter.

Since the likelihood function involves integral of 40 dimensions, the evaluation of such a likelihood function is a real challenge to researchers. Moreover, the random effects are crossed rather than nested. This makes the factorization of the high dimensional integral impossible. As a result, it arouses the interest of many researchers to propose various methodologies by either a Bayesian or Classical approach for analysing this dataset.

4.4 Various methodologies on Salamander Mating Data

Among the methodologies within GLMMs developed for analysing the salamander mating data, logit link function is mainly used. Other researchers proposed a probit link function (see McCulloch (1994) and Chan and Kuk (1997)). Since the inclusion of high dimensional crossed random effects complicates the estimation considerably, we now focus on the various estimation methods proposed by various researchers on analysing the salamander mating data in a GLMMs framework.

The data were firstly introduced by McCullagh and Nelder (1989) who adopted an estimating equation approach. They used an approximation obtained by Taylor series expansions for the integrands. However, this method is not efficient when there is high-dimensional integration for the random effects. After this result was published, many researchers adopted different estimation methodologies for this challenging dataset. In a Bayesian aspect, Karim and Zeger (1992) adopted the Gibbs sampling approach. Gibbs sampler is effective in simulating the parameters as well as the unobserved random effects from their joint posterior distributions. However, problems exist in choosing suitable prior distributions for the parameters. Moreover, it is difficult to simulate from the full conditional distributions when some of such conditional distributions are not of a standard forms. Other random variates generation techniques, e.g. Metropolis Hastings or Adaptive rejection sampling are needed in simulating from the non-standard conditional distribution. It also imposes a computational burden when we have to simulate a large number of random effects. Apart from Bayesian approach, there are other methodologies using classical approaches. To name a few, Schall (1991) studied the use of a general algorithm on ML and REML to get an approximated maximum likelihood or quasi-maximum likelihood estimates for the fixed and random effects. Wolfinger (1993) suggested a pseudo-likelihood approach which is based on an approximate marginal model for the mean responses. Breslow and Clayton (1993) investigated the penalized quasi likelihood (PQL) approach. However, the estimates are biased towards zero for some variance components. Lin and Breslow (1996) revised this methodology by a bias-corrected PQL. This improves the asymptotic performance of PQL estimates, but inflates the variance. The efficiency also depends on the sample size. Kuk (1995) proposed an asymtotically unbiased estimation methodology for this salamander data. This estimation was based on an adjusted best linear unbiased prediction (BLUP) approach which avoids integration altogether. It can produce estimates which are nearly unbiased, but with a bit larger variance. Shun (1997) studied the use of modified Laplace approximation. This method focuses on re-grouping the terms in the likelihood function according to an asymptotic order and computing a score function. However, the calculation of the likelihood function depends on its structure. It is not efficient when the terms in the likelihood function are complicated. Kuk (1999) further introduced another Laplace importance sampling method. He chose a normal importance sampling function for the random effects with mean as the maximizer of the joint density and variance as the corresponding information matrix. After getting the simulated random effects, he applied a Monte Carlo method to approximate the likelihood function. All these methodologies we mentioned above adopted a logit link function. McCulloch (1994) studied the use of probit link function on the salamander mating data by an MCEM algorithm using ML estimation for all parameters and REML estimation for the variance components. However, he only considered independent random effects. Chan and Kuk (1997) further extended the MCEM algorithm to the correlated random effects. The advantage of using a probit link function is on the ability of EM algorithm to form a nearly identical continuous normal linear model by a probit link in the M-step after taking conditional expectation of the random effects and the underlying unobserved continuous responses in the E-step. Hence, the calculation of ML estimates is greatly simplified by using the basic results for ML estimates when the sampling distribution for the responses is normal.

We have introduced various methodologies proposed by researchers in making inference on this famous and challenging salamander mating data. Table 4.3 shows the estimation results by these methodologies. The results vary across different estimation methods possibly due to the complicated structure of the salamander data. Generally speaking, the variance for female random effects is greater than that of male random effects. This told us that the variability of female effects in mating is larger.

====================================

Table 4.3 about here

=====================================

On the other hand, nearly all researchers suggested the use of normal distribution on the random effects in the salamander mating data. However, the random effects may exhibit a heavy tailed behaviour. In practice, a wider class of random effects distributions may be more appropriate and can widen the scope of application considerably. We will study the use of other heavy-tailed distributions, eg. Student-t distribution to achieve robustness in statistical inference in chapter 6.

i-th Female	j-th Male 1st	j-th Male 2nd	j-th Male 3rd	j-th Male 4th	j-th Male 5th	j-th Male 6th
RB Female						
1	1	*14*	5	*11*	4	*15*
2	5	*15*	3	*13*	1	*12*
3	2	*11*	1	*14*	3	*13*
4	4	*12*	2	*15*	5	*14*
5	3	*13*	4	*12*	2	*11*
6	*19*	9	*20*	7	*16*	8
7	*18*	8	*19*	9	*17*	6
8	*16*	6	*17*	10	*20*	9
9	*20*	7	*18*	6	*19*	10
10	*17*	10	*16*	8	*18*	7
WS Female						
11	9	*19*	7	*20*	10	*18*
12	7	*16*	9	*17*	6	*20*
13	8	*17*	6	*19*	7	*16*
14	10	*20*	8	*18*	9	*19*
15	6	*18*	10	*16*	8	*17*
16	*15*	2	*13*	4	*12*	1
17	*14*	1	*15*	2	*11*	5
18	*11*	4	*12*	5	*15*	3
19	*13*	3	*11*	1	*14*	4
20	*12*	5	*14*	3	*13*	2

Table 4.1: The mating pattern of each female salamander to six male salamanders from two populations: Rough- Butt (RB) and Whiteside (WS) in which the salamanders coming from WS are written in italic.

	$RB_f - RB_m$	$RB_f - WS_m$	$WS_f - RB_m$	$WS_f - WS_m$	Total
Expt. 1 (summer)	22 (73.3%)	20 (66.7%)	7 (23.3%)	21 (70.0%)	70
Expt. 2 (fall 1st r)	18 (60.0%)	14 (46.7%)	7 (23.3%)	20 (66.7%)	59
Expt. 3 (fall 2nd)	20 (66.7%)	16 (53.3%)	5 (16.7%)	19 (63.3%)	60

Table 4.2: The successful rate of mating for different pairs of salamanders in the three experiments and its corresponding percentage in brackets.

Estimates	β_0	β_1	β_2	β_3	σ_1	σ_2
Moment	0.97	-2.12	-0.30	2.26	1.17	0.84
Laplace importance sampling	1.39	-3.05	-0.45	3.29	1.31	0.50
PQL	0.79	-2.29	-0.54	2.82	1.19	0.30
CPQL	0.79	-2.29	-0.54	2.82	1.31	0.63
Gibbs sampling	1.48	-3.25	-0.50	3.62	1.53	0.37
Laplace approximation	1.39	-3.06	-0.45	3.31	1.34	0.50

Table 4.3. Parameter estimates for the first experiment by various researchers.

CHAPTER 5

MC APPROXIMATION THROUGH GIBBS OUTPUT IN GLMM

5.1 Introduction

A wide variety of methodologies can be used for parameter estimation in the GLMMs where each method has its own merits and drawbacks. In this chapter, we introduce a new parameter estimation method called "Monte Carlo approximation through Gibbs output" in the estimation on GLMMs. This method provides a new perspective of using both classical and Bayesian approaches in estimation.

The chapter will be presented as follows. Section 2 introduces our method of Monte Carlo approximation using prior information on parameters. In Section 3, we will further illustrate this method by using the famous salamander mating data discussed in the previous chapter. Section 4 reports the numerical results with comments. Finally a conclusion is given in Section 5.

5.2 Our proposed method: Monte Carlo approximation through Gibbs output

Our proposed method is an extension to the Monte Carlo relative likelihood approach reported by Kuk and Cheng (1999). The advantage of our method is that we do not need to specify a proper reference point in the estimation procedure. Therefore, in our method, we can adopt arbitrary prior distributions to the parameters and then sample random effects as well as parameters from the joint posterior distribution. Based on the Gibbs output, we can then approximate the likelihood function by the Monte Carlo method.

Suppose, we define the marginal likelihood based on the observed data $\boldsymbol{y}$ over the random effects $\boldsymbol{z}$ as

$$lik(\boldsymbol{\theta}; \boldsymbol{y}) = \int f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{z}$$

Geyer and Thompson (1992) proposed to calculate the marginal relative likelihood using a Monte Carlo approximation with i indexes the number of simulations used in the approximation as follows:

$$\begin{aligned} \frac{lik(\boldsymbol{\theta}; \boldsymbol{y})}{lik(\boldsymbol{\theta}_0; \boldsymbol{y})} &= \int \frac{f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta})}{f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta}_0)} f(\boldsymbol{z}|\boldsymbol{y}; \boldsymbol{\theta}_0) d\boldsymbol{z} \\ &\approx \frac{1}{M} \sum_{i=1}^{M} \frac{f(\boldsymbol{y}, \boldsymbol{z}_i|\boldsymbol{\theta})}{f(\boldsymbol{y}, \boldsymbol{z}_i|\boldsymbol{\theta}_0)} \qquad \text{where } \boldsymbol{z}_i \sim f(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{\theta}_0) \end{aligned}$$

The random effects $\boldsymbol{z}_i$ are drawn from the joint density of $f(\boldsymbol{y}, \boldsymbol{z}|\boldsymbol{\theta}_0)$ based on a given reference point $\boldsymbol{\theta}_0$. However, the local approximation may only be good when the reference point $\boldsymbol{\theta}_0$ is close to the true parameter $\boldsymbol{\theta}$. Kuk and Cheng (1999) demonstrated that the resulting maximizer may differ substantially from the true MLE if the reference point is chosen improperly. One remedy to this problem is to update the reference point $\boldsymbol{\theta}_0$ to the current $\hat{\boldsymbol{\theta}}$ each time and then simulate a new set of $\boldsymbol{z}$ based on the new reference point (Geyer and Thompson, 1992). This can solve the problem of choosing a proper reference point and simulating one particular set of random effects to approximate the likelihood function by Monte Carlo method. However, this method of updating the reference point will require iterations within iterations. Moreover, this method requires a large amount of computation in simulating a new set of random effects based on the current estimate $\hat{\boldsymbol{\theta}}$ as a new reference point each time to approximate the entire likelihood function.

Apart from the Monte Carlo relative likelihood approach, McCulloch (1997) also suggested a simulated maximum likelihood (SML) approach. This method required an optimal importance sampling function to draw the random effects in order to carry out the Monte Carlo approximation. However, the SML estimator performs poorly if the choice of importance sampling distribution is far away from the true distribution for the random effects.

Because of the constraints we mentioned above, we suggest another method of using the Gibbs output in the Monte Carlo approximation. In our proposed method, we assign a conveniently chosen prior density of $\boldsymbol{\theta}^*$, say $h(\boldsymbol{\theta}^*)$ and use it to replace the specified reference point $\boldsymbol{\theta}_0$. The likelihood function is calculated as follows:

$$\begin{aligned} lik(\boldsymbol{\theta}) &= \int\int \frac{f(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\theta})}{f(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\theta}^*)} f(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\theta}^*)d\boldsymbol{z}\ h(\boldsymbol{\theta}^*)d\boldsymbol{\theta}^* \\ &\propto \int\int \frac{f(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\theta})}{f(\boldsymbol{y},\boldsymbol{z}|\boldsymbol{\theta}^*)} f(\boldsymbol{z},\boldsymbol{\theta}^*|\boldsymbol{y})d\boldsymbol{z}d\boldsymbol{\theta}^* \\ &\approx \frac{1}{M}\sum_{i=1}^{M} \frac{f(\boldsymbol{y},\boldsymbol{z}_i|\boldsymbol{\theta})}{f(\boldsymbol{y},\boldsymbol{z}_i|\boldsymbol{\theta}_i)} \qquad \text{where } (\boldsymbol{z}_i,\boldsymbol{\theta}_i) \sim f(\boldsymbol{z},\boldsymbol{\theta}|\boldsymbol{y}) \end{aligned} \tag{5.1}$$

Thus we sample the random effects $\boldsymbol{z}_i$ and parameters $\boldsymbol{\theta}_i$ from a joint posterior density and use them to evaluate the likelihood function by a Monte Carlo approximaton. Moreover, it solves the problem of choosing a proper reference point and it does not required to simulate a new set of random effects for $\boldsymbol{\theta}$ in each iteration. If the sample size is large, the posterior density $f(\boldsymbol{\theta}^*|\boldsymbol{y})$ will be concentrated around the maximum likelihood estimates $\hat{\boldsymbol{\theta}}$ and so we automatically get a good approximaiton of $lik(\boldsymbol{\theta})$ around $\hat{\boldsymbol{\theta}}$.

5.3 Model fitting on the Salamander Mating Data

We use the famous salamander mating data as mentioned in the previous chapter to illustrate our proposed method. We consider the experiment carried out in the summer of 1986. The objective of this experiment was to investigate whether there were barriers to interbreeding in the salamanders from these two geographically isolated populations, Rough Butt (RB) and Whiteside (WS).

Since the inclusion of 40 random effects leads to a high dimensional intergrals in the likelihood function in the model and complicates the parameter estimation considerably, many researchers proposed different estimation methods on this famous salamander data. A brief overview on this methodologies can be found in previous chapter. In this section, we will introduce the method of Monte Carlo approximation through Gibbs output to estimate the likelihood function in the salamander mating data. We will also compare the results with other researchers and study the efficiency of our method.

A logit link function is adopted to the appropriate parameters and the generalized linear mixed model of our proposed method is presented as follows. Let Y_t be the binary response of mating (1 = success, 0 = failure) in the experiment, where t = 1,...,120 correspond to the number of matings in the experiment. The fixed effects indicate which population the male and female salamander belongs to. WSF_t equals to 1 if a female salamander involved in the t-th mating came from Whiteside. Otherwise, WSF_t is 0. Similarily, WSM_t equals to 1 if a male salamander involved in the t-th mating came from Whiteside. Otherwise, WSM_t equals to 0. $WSF_t \times WSM_t$ is the interaction between the two fixed effects. We further define a_j and b_j as the random effects of the j-th female salamanders and the j-th male salamanders with j = 1,...,20. These female and male random effects follow normal distributions with means equal to 0 and variances equal to ${\sigma_1}^2$ and ${\sigma_2}^2$, respectively. We denote the vector of random effects by $\boldsymbol{z} = (a_1, ..., a_{20}, b_1, ..., b_{20})$ and the vector of parameters by $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma_1, \sigma_2)$. The marginal density is given by

$$\begin{aligned} f(\boldsymbol{y}|\boldsymbol{\theta}) &= \prod_{t=1}^{120} \int \cdots \int \frac{\exp(\eta_t y_t)}{1 + \exp(\eta_t)} \prod_{j=1}^{20} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{{a_j}^2}{2{\sigma_1}^2}\right) \times \\ & \quad \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{{b_j}^2}{2{\sigma_2}^2}\right) da_1 ... da_{20}\ db_1 ... db_{20} \\ \text{where } \eta_t &= \beta_0 + \beta_1 WSF_t + \beta_2 WSM_t + \beta_3 WSF_t \times WSM_t + a_{t'} + b_{t''} \end{aligned}$$

t' and t'' correspond to the t-th mating for the female and male salamander respectively.

To obtain the ML estimates, we evaluate the marginal likelihood function by firstly sampling $\boldsymbol{z}$ and $\boldsymbol{\theta}$ from the joint posterior density $f(\boldsymbol{z}, \boldsymbol{\theta}|\boldsymbol{y})$ adopting a vague prior $h(\boldsymbol{\theta})$ for the parameters $\boldsymbol{\theta}$ using Gibbs sampler. The vague priors for $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$ are normal and inverse gamma respectively. Let $(\boldsymbol{z}_i, \boldsymbol{\theta}_i)$ be the i-th simulated set of random effects and parameters, we then use the Gibbs output to calculate the approximated marginal likelihood function given by equation (5.1) where the term $f(\boldsymbol{y}, \boldsymbol{z}_i|\boldsymbol{\theta}_i)$, the joint density of $\boldsymbol{y}$ and $\boldsymbol{z}_i$ given $\boldsymbol{\theta}_i$ in the denominator of the equation is calculated by

$$f(\boldsymbol{y}, \boldsymbol{z}_i; \boldsymbol{\theta}_i) = \prod_{t=1}^{120} \frac{\exp(\eta_{ti} y_t)}{1 + \exp(\eta_{ti})} \prod_{j=1}^{20} \frac{1}{\sqrt{2\pi}\sigma_1{}^i} \exp\left(-\frac{a_j{}^{2i}}{2\sigma_1{}^{2i}}\right) \times \frac{1}{\sqrt{2\pi}\sigma_2{}^i} \exp\left(-\frac{b_j{}^{2i}}{2\sigma_2{}^{2i}}\right)$$

$$\text{where } \eta_{ti} = \beta_0{}^i + \beta_1{}^i WSF_t + \beta_2{}^i WSM_t + \beta_3{}^i WSF_t \times WSM_t + a^i_{t'} + b^i_{t''}$$

Newton Raphson Method is then used to obtain the ML estimates. As we assign priors to the parameters and sample them from a joint posterior density, we have adopted a semi-Bayesian approach. This methodology combines both classical approach of maximizing likelihood function and Bayesian approach of adopting priors for the parameters.

To approximate the likelihood function closely using the Monte Carlo approximation, the number of simulations M should be large. In this salamander example, we use $M = 20000$ set of simulations on $(\boldsymbol{z}_i, \boldsymbol{\theta}_i)$.

The full conditional function is not in standard form for $\boldsymbol{z}$. Other non-standard sampling methods, eg. Metropolis Hastings or adaptive rejection sampling can be used to perform the sampling. Our Gibbs outputs are obtained from a statistical software called 'WinBUGS' (Window version of Bayesian inference Using Gibbs sampling). Vague priors are chosen for $\boldsymbol{\theta}$. To obtain the ML estimates using the Newton Raphson method, we have to calculate the first and second derivatives of log-likelihood function with respect to $\boldsymbol{\theta}$ denoted by $l'(\boldsymbol{\theta}; \boldsymbol{y})$ and $l''(\boldsymbol{\theta}; \boldsymbol{y})$ respectively. The equations of the derivatives are given in Appendix

1. The ML estimate is obtained by

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - (l''(\boldsymbol{\theta}^{(k)}; \boldsymbol{y}))^{-1} l'(\boldsymbol{\theta}^{(k)}; \boldsymbol{y})$$

where k is the number of iterations in the Newton Raphson method. More information are given in Kuk and Cheng (1997) on the Monte Carlo Newton-Raphson Algorithm. In our example, the initial values used in the Newton Raphson method are the moment estimates. The implementation of the Newton Raphson iterative procedures and the evaluation of likelihod function and its derivatives are carried out by writting a FORTRAN program.

5.4 Results

Table 5.1 shows the result using our proposed method for the experiment carried out in the summer of 1986. We also include the estimates obtained by other researchers for comparison. Our estimates for $\beta_0, \beta_1, \beta_2, \beta_3$ are close to other $\boldsymbol{\beta}$ estimates obtained by various researchers. Our estimate for σ_1 is close to the estimate obtained by Kuk (1999) by the method of Laplace importance sampling and by Lin and Breslow (1996) on the CPQL approach while the estimate for σ_2 is close to the estimates obtained through Gibbs sampling by Karim and Zeger (1992). It is interesting to see that our estimates lie between the results of Bayesian and Classical approaches. We also include the standard deviation for the estimate of $\boldsymbol{\beta}$ in Table 5.1.

The goodness-of-fit for our model can be assessed by the estimated probabilities of successful mating for different mating types of salamanders from WS and RB. We denote π_{ij} be the probability of successful mating between a female from population i and a male from population j where $i, j = W$ or R corresponding to WS or RB respectively. For example, π_{RW} corresponds to the probability of successful mating between a female living in RB and a male living in WS. The observed and estimated probabilties are shown in Table 5.2 for model checking. It is found that the expected probabilities obtained under our proposed method are close to the observed probabilities. We use a chi-square test to check whether the expected probabilities are equal to the observed probabilities. The chi-square statistics for our proposed method is 0.025 (p-value = 0.987). The result shows that our model is statistically adequate. Moreover, our p-value is the largest among all methods mentioned. It assures that our estimation method is efficient.

====================================

Table 5.1 and Table 5.2 about here

====================================

5.5 Conclusion

We have shown that our method of assigning priors to the parameters and using the Gibbs output for Monte Carlo approximation is useful in obtaining the ML estimates for models that involves high dimensional random effects. The use of prior information solves the problem of choosing a proper reference point and updating such reference point through iterations in the Monte Carlo relative likelihood approach.

On the other hand, our method is innovative in making inference in the Generalized linear mixed models through the Gibbs output for Monte Carlo approximation. Moreover, it eases the problem of evaluating the marginal likelihood. We illustrate our method through the famous salamander mating data. The crossed random effects induces high dimensional integrals in the likelihood function and prohibits the factorization of such integral. The use of Monte Carlo approximation provides a solution to evaluate the high-dimensional integral. In a non-parametric approach, Chib (1995) suggested ways to evaluate the marginal likelihood for a given model through Gibbs output. Using his idea, we can further extend our methodology when the prior density for the parameters will not be fixed but is estimated through the model. This will be another interesting area to be investigated.

Finally, in order to simulate the Gibbs output in the Bayesian step, the use of non-standard sampling method, eg. Metropolis Hastings or Adaptive Rejection sampling or Ratio-of-Uniform are needed. The convergence and auto-correlation of the Gibbs output should be checked to make sure that the estimates obtained by ML method is correct. Moreover, in using the Newton Raphson method in obtaining the ML estimates, we should be careful in choosing a proper starting point since this Newton Raphson method is sensitive to our starting values. Especially in our example of the salamander mating data, it may be difficult to search for a global maximum in this case due to the high-dimensional surfaces in the parameter space. On the other hand, Kuk (2002) suggested that the effect of prior specification can be possibly outweighed by a large sample size. So, if the sample size is not large enough to outweigh any reasonable prior specification, duplication of data can improve the accuracy of the simulated likelihood function. This will be another interesting area for further research.

5.6 Appendix 1

The likelihood function is defined as follows:

$$lik(\boldsymbol{\theta}; \boldsymbol{y}) \approx \frac{1}{M} \sum_{i=1}^{M} \frac{f(\boldsymbol{y}, \boldsymbol{z}_i | \boldsymbol{\theta})}{f(\boldsymbol{y}, \boldsymbol{z}_i | \boldsymbol{\theta}_i)}$$

If we take logarithm on the likelihood $lik(\boldsymbol{\theta};\boldsymbol{y})$, it becomes

$$\ell(\boldsymbol{\theta};\boldsymbol{y}) \approx ln\left(\frac{1}{M}\sum_{i=1}^{M}\frac{f(\boldsymbol{y},\boldsymbol{z}_i|\boldsymbol{\theta})}{f(\boldsymbol{y},\boldsymbol{z}_i|\boldsymbol{\theta}_i)}\right)$$

where $f(\boldsymbol{y},\boldsymbol{z}_i|\boldsymbol{\theta}_i) = \prod_{t=1}^{120}\frac{\exp(\eta_{ti}y_t)}{1+\exp(\eta_{ti})}\prod_{j=1}^{20}\frac{1}{\sqrt{2\pi}{\sigma_1}^i}\exp\left(-\frac{{a_j}^{2i}}{2{\sigma_1}^{2i}}\right)\times\frac{1}{\sqrt{2\pi}{\sigma_2}^i}\exp\left(-\frac{{b_j}^{2i}}{2{\sigma_2}^{2i}}\right)$

We denote $\boldsymbol{\theta} = (\boldsymbol{\beta},\boldsymbol{\sigma})$ as our parameter estimates. We further let

$$f(\boldsymbol{y}|\boldsymbol{z_i},\boldsymbol{\beta}) = \prod_{t=1}^{120}\frac{\exp(\eta_{ti}y_t)}{1+\exp(\eta_{ti})}$$

$$f(\boldsymbol{z}_i|\boldsymbol{\sigma}) = \prod_{j=1}^{20}\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{{a_j}^{2i}}{2{\sigma_1}^2}\right)\times\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{{b_j}^{2i}}{2{\sigma_2}^2}\right)$$

1. Differentiate the log-likelihood function once with respect to β_k where $k = 0,1,2,3$. We have:

$$\frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial\beta_k} = lik(\boldsymbol{\theta};\boldsymbol{y})^{-1}\frac{1}{M}\sum_{i=1}^{M}\frac{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta},\boldsymbol{\sigma})Df_{\beta_k}(\boldsymbol{y}|\boldsymbol{z_i},\boldsymbol{\beta})}{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta_i},\boldsymbol{\sigma_i})}$$

where

$$Df_{\beta_k}(\boldsymbol{y}|\boldsymbol{z_i},\boldsymbol{\beta}) = \sum_{t=1}^{120}\left(X_{tk}[y_t - \frac{\exp(\eta_{ti})}{1+\exp(\eta_{ti})}]\right)$$

2. In order to assure a positive value on the parameter $\boldsymbol{\sigma}^2$, we differentiate the log-likelihood function once with respect to $ln({\sigma_l}^2)$ where $l = 1,2$. We have:

$$\frac{\partial\ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial ln({\sigma_l}^2)} = lik(\boldsymbol{\theta};\boldsymbol{y})^{-1}\frac{1}{M}\sum_{i=1}^{M}\frac{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta},\boldsymbol{\sigma})Df_{\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma})}{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta_i},\boldsymbol{\sigma_i})}$$

where

$$Df_{\sigma_1}(\boldsymbol{z_i}|\boldsymbol{\sigma}) = \sum_{j=1}^{20}\frac{{a_j}^{2i}}{2{\sigma_1}^2} - 10$$

$$Df_{\sigma_2}(\boldsymbol{z_i}|\boldsymbol{\sigma}) = \sum_{j=1}^{20} \frac{{b_j}^{2i}}{2{\sigma_2}^2} - 10$$

3. Differentiate the log-likelihood function twice with respect to β_{k_1} and β_{k_2}. We have the second derivatives as follows:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{y})}{\partial \beta_{k_1} \partial \beta_{k_2}} &= lik(\boldsymbol{\theta}; \boldsymbol{y})^{-2}\{lik(\boldsymbol{\theta}; \boldsymbol{y})\frac{1}{M}\sum_{i=1}^{M} \\ &\quad \frac{[Df_{\beta_{k_1}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta})Df_{\beta_{k_2}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta}) + D2f_{\beta_{k_1}\beta_{k_2}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta})]f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta}, \boldsymbol{\sigma})}{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta_i}, \boldsymbol{\sigma_i})} \\ &\quad -[\frac{1}{M}\sum_{i=1}^{M} \frac{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta}, \boldsymbol{\sigma})Df_{\beta_{k_1}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta})}{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta_i}, \boldsymbol{\sigma_i})}][\frac{1}{M}\sum_{i=1}^{M} \frac{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta}, \boldsymbol{\sigma})Df_{\beta_{k_2}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta})}{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta_i}, \boldsymbol{\sigma_i})}]\}\end{aligned}$$

where

$$D2f_{\beta_{k_1}\beta_{k_2}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta}) = \frac{\partial Df_{\beta_{k_1}}(\boldsymbol{y}|\boldsymbol{z_i}, \boldsymbol{\beta})}{\partial \beta_{k_2}} = -\sum_{t=1}^{120} X_{tk_1} X_{tk_2} \frac{\exp(\eta_{ti})}{[1 + \exp(\eta_{ti})]^2}$$

4. Differentiate the log-likelihood function twice with respect to $ln({\sigma_1}^2)$ and $ln({\sigma_l}^2)$ where $l = 1, 2$. We have the second derivatives as follows:

$$\begin{aligned}\frac{\partial \ell(\boldsymbol{\theta}; \boldsymbol{y})}{\partial ln({\sigma_1}^2) \partial ln({\sigma_l}^2)} &= lik(\boldsymbol{\theta}; \boldsymbol{y})^{-2}\{lik(\boldsymbol{\theta}; \boldsymbol{y})\frac{1}{M}\sum_{i=1}^{M} \\ &\quad \frac{[Df_{\sigma_1}(\boldsymbol{z_i}|\boldsymbol{\sigma})Df_{\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma}) + D2f_{\sigma_1\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma})]f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta}, \boldsymbol{\sigma})}{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta_i}, \boldsymbol{\sigma_i})} \\ &\quad -[\frac{1}{M}\sum_{i=1}^{M} \frac{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta}, \boldsymbol{\sigma})Df_{\sigma_1}(\boldsymbol{z_i}|\boldsymbol{\sigma})}{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta_i}, \boldsymbol{\sigma_i})}][\frac{1}{M}\sum_{i=1}^{M} \frac{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta}, \boldsymbol{\sigma})Df_{\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma})}{f(\boldsymbol{y}, \boldsymbol{z_i}|\boldsymbol{\beta_i}, \boldsymbol{\sigma_i})}]\}\end{aligned}$$

where

$$\begin{aligned}\frac{\partial Df_{\sigma_1}(\boldsymbol{z_i}|\boldsymbol{\sigma})}{\partial ln{\sigma_l}^2} = D2f_{\sigma_1\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma}) &= -\frac{1}{2}\sum_{j=1}^{20} \frac{{a_j}^{2i}}{{\sigma_1}^2} \qquad \text{if } l = 1 \\ D2f_{\sigma_1\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma}) &= 0 \qquad \text{if } l = 2\end{aligned}$$

5. Differentiate the log-likelihood function twice with respect to $ln(\sigma_l{}^2)$ where $l = 1, 2$ and β_k where $k = 0, 1, 2, 3$. We have the second derivatives as follows:

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta};\boldsymbol{y})}{\partial \beta_k \partial ln(\sigma_l{}^2)} &= lik(\boldsymbol{\theta};\boldsymbol{y})^{-2}\{lik(\boldsymbol{\theta};\boldsymbol{y})\frac{1}{M}\sum_{i=1}^{M}\frac{Df_{\beta_k}(\boldsymbol{y}|\boldsymbol{z_i},\boldsymbol{\beta})Df_{\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma})f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta},\boldsymbol{\sigma})}{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta_i},\boldsymbol{\sigma_i})} \\ &\quad -[\frac{1}{M}\sum_{i=1}^{M}\frac{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta},\boldsymbol{\sigma})Df_{\beta_k}(\boldsymbol{y}|\boldsymbol{z_i},\boldsymbol{\beta})}{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta_i},\boldsymbol{\sigma_i})}][\frac{1}{M}\sum_{i=1}^{M}\frac{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta},\boldsymbol{\sigma})Df_{\sigma_l}(\boldsymbol{z_i}|\boldsymbol{\sigma})}{f(\boldsymbol{y},\boldsymbol{z_i}|\boldsymbol{\beta_i},\boldsymbol{\sigma_i})}]\} \end{aligned}$$

Estimates	β_0	β_1	β_2	β_3	σ_1	σ_2
Our proposed method	1.30	-2.83	-0.50	3.22	1.30	0.38
	(0.56)	(0.82)	(0.68)	(0.92)		
Moment	0.97	-2.12	-0.30	2.26	1.17	0.84
Laplace importance sampling	1.39	-3.05	-0.45	3.29	1.31	0.50
PQL	0.79	-2.29	-0.54	2.82	1.19	0.30
CPQL	0.79	-2.29	-0.54	2.82	1.31	0.63
Gibbs sampling	1.48	-3.25	-0.50	3.62	1.53	0.37
Laplace approximation	1.39	-3.06	-0.45	3.31	1.34	0.50

Table 5.1. Parameter estimates by our proposed method (standard errors in parentheses) and other estimates by various methods.

	Observed proportion	Expected proportion
π_{WW}	0.700	0.719
π_{WR}	0.233	0.232
π_{RW}	0.667	0.653
π_{RR}	0.730	0.735

Table 5.2. Observed and expected values for probability of successful mating for salamanders involved in two different populations.

CHAPTER 6

STUDENT-T DISTRIBUTION ON THE SALAMANDER MATING DATA

6.1 Introduction

Most of the current works on the GLMMs assume normal distribution for the random effects. In practice, a wider class of random effects distributions may be more appropriate and can widen the scope of application considerably. For example, Student-t distribution which has thick tail behaviour is an alternative to the normal distribution on modelling the random effects. Choy and Smith (1997a) suggested the scale mixtures of normal (SMN) distributions that included the Student-t, stable, exponential power and Laplace distributions to replace the commonly used normal distribution. SMN distributions are symmetric and can be more platykurtic and leptokurtic than the normal distribution. The property of heavy tails in the SMN distributions motivates the robustness consideration. This robustness analysis can be applied to the GLMMs to locate outliers and downweigh their effects.

The male and female random effects in the salamander mating data have always assumed to be normally distributed. Instead, in this thesis, we propose the use of Student-t distribution for modelling the random effects. The main objective of this chapter is to provide a robust analysis on the salamander data through a fully Bayesian approach. We further assume that the degrees of freedom for the Student-t distribution to be random in the model so that it can be estimated from the data. We will use a logit link function in the model. Then we express the Student-t distribution as scale mixtures of normal (SMN) representation to simplify the Bayesian computation and to provide a means to check for the possible outliers in the model.

The framework of this chapter is presented as follows. Firstly, Section 2 describes how the Student-t distribution can be expressed as the scale mixtures of normal representation. Section 3 presents the model of GLMM on the salamander mating data with Student-t distributed random effects. We will analyse this mixed model in a full Bayesian approach. Section 4 shows our result on the robust analysis of the salamander data. Finally a conclusion on the use of Student-t distribution or other heavier tailed distributions will be given in Section 5.

6.2 Representation of Scale Mixture of Normals for Student-t distribution

Before carrying on to describe our model, we introduce the representation of Student-t distribution as the scale mixture of normals (SMN) first. Student-t is a heavy-tailed distribution which is useful when the data is more dispersed than normal. It is markedly leptokurtic when the degrees of freedom is small. The use of SMN can simplify the simulation and at the same time, give a diagnosis of potential outliers by considering the magnitude of mixing parameters. West(1984, 1987), Choy and Smith (1997b) and Fernandez and Steel (1998) had studied the use of scale mixtures of normal distribution. More details can be found in their papers.

We illustrate how to express Student-t distribution into a SMN form. Suppose X is a random variable which follows a Student-t distribution with mean θ and variance σ^2. The degree of freedom is α. Then

$$X|\theta, \sigma^2 \sim t_\alpha\left(\theta, \sigma^2\right)$$

The representation of Student-t as SMN with λ being the mixing parameter is as follows:

$$X|\theta, \sigma^2, \lambda \quad \sim \quad N\left(\theta, \frac{\sigma^2}{\lambda}\right)$$

$$\lambda \sim Ga\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right)$$

where $N(\cdot)$ is a standard normal distribution, $t_\alpha(\cdot)$ is the Student-t distribution with degrees of freedom α. $Ga(c, d)$ is the gamma distribution with mean c/d.

6.3 Modelling the Salamander Mating Data

The model for analysing the salamander mating data has been defined in Chapter 4 except that in this case, the female and male random effects now follow Student-t distributions with the degrees of freedom being set as parameters and are determined by the data itself.

Let α_f and α_m denote the degrees of freedom for female and male random effects, respectively. Then the vector of parameters in the model is $\boldsymbol{\theta} = (\beta_0, \beta_1, \beta_2, \beta_3, \sigma_f, \sigma_m, \alpha_f, \alpha_m)$. Gamma vague priors are given to α_f and α_m while normal vague priors are given to the $\boldsymbol{\beta}$. For σ_m^2 and σ_f^2, we assign inverse gamma vague priors to them. The full Bayesian model is defined as:

$$logit[Pr(Y_{ij} = 1)] = \beta_0 + \beta_1 WSF_i + \beta_2 WSM_j + \beta_3 WSF_i \times WSM_j + a_i + b_j$$

$$a_i \sim t_{\alpha_f}\left(0, \sigma_f^2\right)$$

$$b_j \sim t_{\alpha_m}\left(0, \sigma_m^2\right)$$

$$\alpha_f \sim Ga\left(c_1, d_1\right)$$

$$\begin{aligned}
\alpha_m &\sim Ga\left(c_2, d_2\right) \\
\beta_k &\sim N\left(\mu_{\beta_k}, \sigma^2_\beta\right) \qquad\qquad \text{where } k = 0, 1, 2, 3 \\
\sigma^2_f &\sim IG\left(c_3, d_3\right) \\
\sigma^2_m &\sim IG\left(c_4, d_4\right)
\end{aligned}$$

where $c_1, c_2, c_3, c_4, d_1, d_2, d_3, d_4 > 0$ are known values.

We further express the Student-t distribution for the random effects into scale mixtures of normal form where λ_{f_i} and λ_{m_j} are the mixing parameters for female and male random effects respectively with $i = 1, ..., 20, j = 1, ..., 20$. The scale mixture of normals on the random effects will be presented as follows:

$$\begin{aligned}
a_i &\sim N\left(0, \frac{\sigma^2_f}{\lambda_{f_i}}\right) \\
b_j &\sim N\left(0, \frac{\sigma^2_m}{\lambda_{m_j}}\right) \\
\lambda_{f_i} &\sim Ga\left(\frac{\alpha_f}{2}, \frac{\alpha_f}{2}\right) \\
\lambda_{m_j} &\sim Ga\left(\frac{\alpha_m}{2}, \frac{\alpha_m}{2}\right)
\end{aligned}$$

In Bayesian approach, let $[\boldsymbol{\beta}, \boldsymbol{a}, \boldsymbol{b}, \alpha_f, \alpha_m, \boldsymbol{\lambda}_f, \boldsymbol{\lambda}_m, \sigma_f, \sigma_m | \boldsymbol{y}]$ denote the joint posterior distribution. Then parameters can be generated from its corresponding conditional distribution through the Gibbs sampler method (Smith and Robert, 1993). However, it is a bit difficult to simulate some of the parameters since their conditional functions are not always in standard form. So, Metropolis Hastings or rejection sampling is needed to simulate them. All the simulatons in this example are carried out using WinBUGS.

6.4 Results

The hierarchical model using Bayesian approach as described in the previous section is fitted to the data of each set of experiments in salamander mating data. To reflect the non-informative prior knowledge on $\alpha_f, \alpha_m, \sigma_f^2, \sigma_m^2$, we choose $c_1 = c_2 = c_3 = c_4 = d_1 = d_2 = d_3 = d_4 = 0.001$. We let μ_{β_k} be the moment estimates of $\boldsymbol{\beta}$ and σ_β^2 be a large value as the normal vague prior for $\boldsymbol{\beta}$. We ran the Gibbs sampler for 35000 iterations while the first 5000 iterations are discarded as the burn-in period. We then pick up simulated values every 30 iterations to reduce the autocorrelation and to mimic a sample of 1000 independent simulated data. The dependency of parameters can be checked by plotting ACF, the Auto-Correlation Function. Figure 6.1 shows the autocorrelation of the parameters $\boldsymbol{\beta}, \sigma_f, \sigma_m, \alpha_f$ and α_m for the first experiment carried out in summer in 1986. The autocorrelation disappeared after first few lags. It seems that the simulated values are not highly correlated. Figure 6.2 presents the time series plots for the parameters $\boldsymbol{\beta}$ which show random patterns.

======================================

Figure 6.1 and Figure 6.2 about here

=====================================

Table 6.1 gives the parameter estimates of the degrees of freedom for Student-t distributed random effects in these three experiments. It is given that α_f and α_m, the degrees of freedom for female and male random effects are 12 and 10 respectively in the first experiment. It shows the presence of outlying male and female random effects. For the second experiment, α_f and α_m are 3 and 6, respectively. This indicates that the male and female random effects for this experiment are more heavily tailed than the others. For the third experiment, 10 and 9 degrees of freedom are obtained for the male and female random effects respectively. These results suggest the dispersion of the male and female random effects are intermediate among the three experiments. Since the degrees of freedom for male and female random effects in all these three experiments are not large, they justify the use of Student-t distributed random effects for robustness consideration.

Table 6.2 shows the parameter estimates and their corresponding standand errors of the salamander data by using normal and Student-t distributions on the random effects for comparison. The estimates of the scale parameters σ_m and σ_f vary across the choice of distribution for the random effects. The estimates of the scale parameters σ_m and σ_f are smaller when we use the Student-t random effects instead of normal random effects in these three experiments.

==

Table 6.1 and Table 6.2 about here

==

By SMN representation of the Student-t random effects, we can detect the possible outliers by checking the magnitude of mixing parameters, λ. As we define the variance of the random effects as $\frac{\sigma^2}{\lambda_i}$ in the SMN representation, a small value of λ for a particular random effect means that it will have a large variance and hence this indicates an outlier. Thus, the possible outliers can be detected by a small value of λ. Figure 6.3 plots the mixing paramaters, λ_f and λ_m for each experiment. In the first experiment, the values for λ are quite similar, but the value of the 7th λ_f is the smallest compared with the others since there was only one successful mating out of 6 occasions for the 7th female salamander. In the second experiment, α_f and α_m are both smaller than the other two experiments. This implies that the distribution of random effects are more heavily tailed in the second experiment. The values of λ are very small for the 15th, 18th and 19th female salamanders since the 15th and the 18th female salamanders failed all the matings while the 19th salamanders succeeded in all the matings. Generally, other salamanders had three to four successful matings out of the six occasions. Similarily, the value of λ for the 2nd male salamander is the smallest because there is no successful mating obtained from that male salamander. Because of these extreme behaviors in mating, they were identified as outliers. This supports the use of Student-t distributions for the female and male random effects with degrees of freedom estimated to be 3 and 6 respectively so that outliers are automatically down-weighed in the model. Moreover, it is interesting to note that even though the same set of salamanders are used in the first two experiments, the female or male salamanders with outlying effects are not the same in these two experiments. In fact, the random effects in the second experiment are more diversified. Since the two experiments are conducted in different seasons, this indicates the presence of significant seasonal effect. See Chan and Kuk (1997) for the result which shows a significant seasonal effects in the data. For the third experiment, the possible female salamanders with outlying random effects are the 2nd, 7th and the 16th salamanders because they all had one successful mating out of six occassions. These outliers can be easily verified by checking the magnitude of λ.

Furthermore, the goodness-of-fit for our model can be assessed by the estimated probabilities of successful mating for different mating types of salamanders from WS and RB. The definition of π_{ij} can be found in the previous chapter. Table 6.3 shows the observed and expected probabilities for each experiment. The expected probabilities are obtained under the assumption of Student-t distribution on the random effects. It is found that the expected probabilites match well to the observed probabilities. This assures that our model is adequate. So, the use of Student-t distribution in the random effects in this dataset can downweigh the effect of outliers and provide better estimates.

=====================================

Figure 6.3 and Table 6.3 about here

=====================================

6.5 Conclusion

For the salamander data, it has been shown that the normal distributions for the male and female random effects cannot accommodate the over-dispersion and capture the outlying effects. In fact, a heavy-tailed distribution is more appropriate and the Student-t distribution is a natural choice. The thicker tail property in this distribution can capture the outliers and down-weight their effects in statistical inference. Furthermore, if we express the Student-t distribution as scale mixtures of normal representation, the possible outliers can be detected easily by the mixing parameters. This is the first attempt among the various researches on salamander data to identify the possible male and female salamanders with outlying effects in mating in each of the three experiments. In the setting, the degrees of freedom for the Student-t distribution is assumed to be unknown and to be estimated given the data. This avoids the possibility of inflating the variance if we choose a wrong degrees of freedom for the Student-t distribution. Moreover, the standard error of the variance of female Student-t distributed random effects is smaller than that of normal random effects in the second experiment. Therefore, our analysis obtained from the Student-t distributed random effects is more robust to outliers.

In fact, other heavy-tailed distributions, such as the exponential power distribution and the symmetric stable distribution worth considering for the analysis of the salamander mating data. In particular, the exponential power density function has a scale mixtures of uniform representation (Choy and Walker, 1999) which enables an efficient Gibbs sampling algorithm for the salamander data. This opens a new area for further research.

	Experiment 1	Experiment 2	Experiment 3
α_f	12.06	3.15	9.81
α_m	9.78	5.97	9.48

Table 6.1. Degrees of freedom of Student-t distribution for random effects.

Estimates	β_0	β_1	β_2	β_3	σ_f	σ_m
Student − t						
Experiment 1	1.502	-3.307	-0.514	3.589	1.384	0.428
	(0.789)	(1.186)	(0.774)	(1.191)	(0.599)	(0.411)
Experiment 2	0.611	-2.811	-0.799	4.311	1.359	0.954
	(0.878)	(1.261)	(0.938)	(1.293)	(0.727)	(0.610)
Experiment 3	1.108	-3.581	-0.915	4.203	0.533	1.512
	(0.846)	(1.029)	(1.095)	(1.194)	(0.471)	(0.609)
Normal						
Experiment 1	1.482	-3.332	-0.452	3.623	1.614	0.464
	(0.803)	(1.166)	(0.727)	(1.100)	(0.580)	(0.431)
Experiment 2	0.650	-2.889	-0.854	4.281	1.902	1.241
	(0.974)	(1.312)	(0.937)	(1.358)	(0.772)	(0.647)
Experiment 3	1.139	-3.651	-0.895	4.293	0.599	1.778
	(0.810)	(1.120)	(1.075)	(1.313)	(0.509)	(0.619)

Table 6.2. Estimates and s.d. (in brackets) for Student-t and normally distributed random effects obtained by WinBUGS.

	Observed proportion			Expected proportion		
	Expt.1	Expt.2	Expt.3	Expt.1	Expt.2	Expt.3
π_{WW}	0.700	0.667	0.633	0.717	0.675	0.636
π_{WR}	0.233	0.233	0.167	0.211	0.227	0.155
π_{RW}	0.667	0.467	0.533	0.673	0.474	0.533
π_{RR}	0.730	0.600	0.667	0.750	0.584	0.681

Table 6.3. Observed and expected probabilities of successful mating for salamanders involved in two different populations.

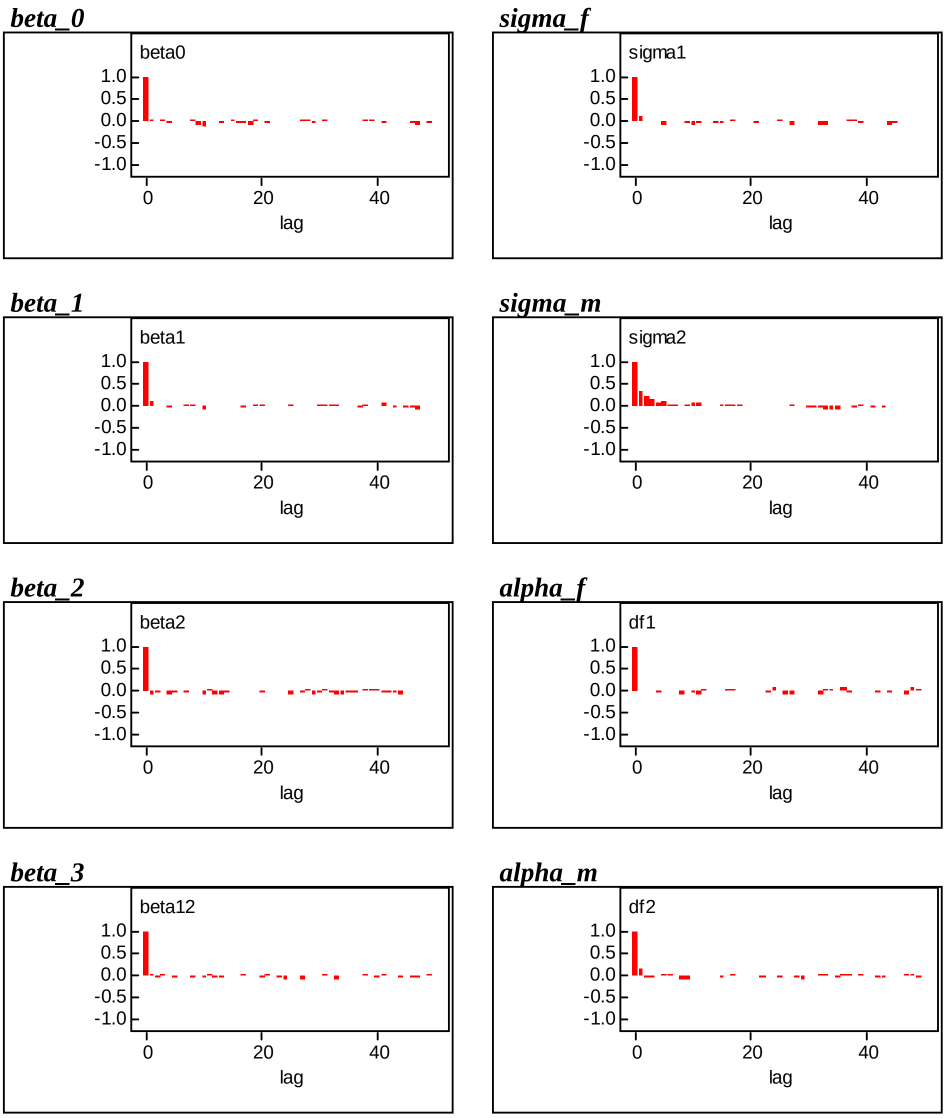

Fig. 6.1. *Autocorrelation functions of 1000 simulated values of* $\boldsymbol{\beta}, \boldsymbol{\sigma}, \boldsymbol{\alpha}$.

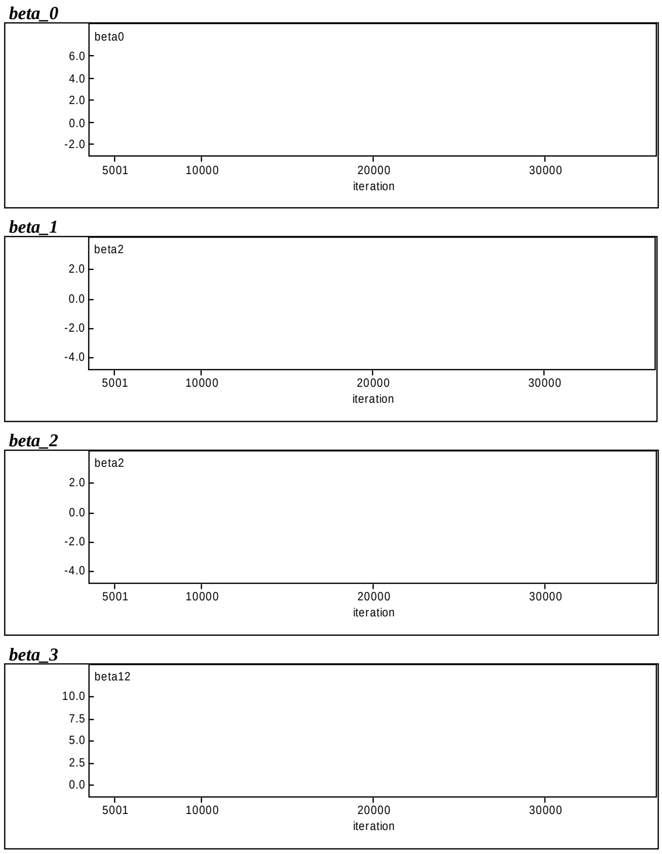

Fig. 6.2. *Time series plot of* $\boldsymbol{\beta}$.

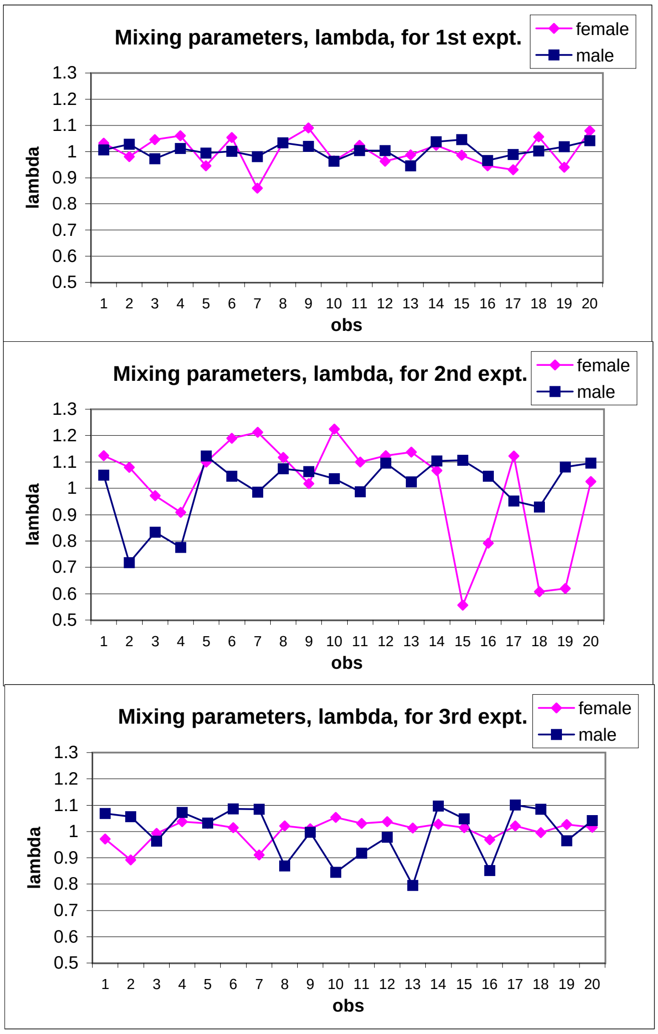

Fig. 6.3. *Bayes estimates of the mixing parameters, λ_f, λ_m for each experiment.*

CHAPTER 7

APPLICATIONS OF SCALE MIXTURE DISTRIBUTIONS FOR STOCHASTIC VOLATILITY MODELS ON FINANCIAL DATA

7.1 Introduction

In the previous chapters, we consider models which accommodate the variability of the data due to clustering, heterogeneity and over-dispersion by inclusion of random effects into the models. These models relate responses to the location of a distribution in terms of parameters. In fact, there are other types of models which measure the variability of the data directly, especially when we expect such variability will change over time. For example, the variability in the returns of an asset is commonly measured by the volatility. There have been many attempts to model the volatility in the financial and economic time series data. Engle (1982) developed an autoregressive conditional heteroscedastic (ARCH) model on modelling returns on the assets. ARCH model assumes an autoregressive feature on the conditional variance. Bollerslev (1986) further improved it to become Generalized ARCH (GARCH) model. In order to deal with non-normality, Nelson

(1991) suggested the Exponential GARCH (EGARCH) model on asset returns. Thereafter, there are a number of variations of ARCH models to deal with a variety of data such as interest rates, exchange rates, equity returns, Treasury Bills, option pricing, etc.. To name a few, Engle *et al.* (1990) investigated the Factor ARCH (FARCH) model on treasury bills. Baillie *et al.* (1996) studied the fractionally integrated GARCH (FIGARCH) model. Hamilton and Susmel (1994) further proposed a switching-regime ARCH (SWARCH) model which is a non-linear extension to ARCH models in studying financial time series with a sudden shift. A comprehensive review of different types of ARCH models can be referred to Engle (1995). It is shown that ARCH models are widely used in financial and economic time series data. However, the conditional variance in these models depend on the previous observations and past variances.

So, Taylor (1986) developed an alternative class of models called Stochastic Volatility (SV) models for modelling volatility. Unlike the ARCH-type models where the volatility depends on past realizations, the SV models formulate the volatility by an unobservable process that allows the volatilities to vary stochastically. These SV models are also investigated by various researchers to make inference on the financial and economic time series data. For example, Harvey *et al.* (1994) studied the multivariate aspects of SV model. Hull and White (1987) studied the Black Scholes option pricing formula and further applied the stochastic volatility model on the pricing of options or assets. Melino and Turnbull (1990) adopted the SV model on pricing Foreign Currency Options by using a generalized method of moments in estimation. Taylor (1994) investigated the application of SV model in the time series model in an Econometrics field. Moreover, So *et al.* (1998) studied the SV model with Markov Switching to capture sudden shifts in volatility due to some important events.

There are wide variety of aspects in the SV models. However, because of the high dimensional numerical integration required in the evaluation of the likelihood function, it is more difficult to estimate the parameters in the SV models compared with the ARCH models. There have been a large amount of literatures in modifying and improving the existing SV model. Kim *et al.* (1998) had investigated the likelihood-based framework for SV model and compared it with the ARCH model. Some researchers such as Melino and Turnbull (1990) tried to use the Method of Moments (MM) or quasi-maximum likelihood (QML) approach to avoid the integration problems associated with evaluating the likelihood directly. However, Jacquier *et al.* (1994) pointed out that these methods are inefficient. He suggested the use of Markov Chain Monte Carlo (MCMC) algorithm by Gelfand and Smith (1990) and Smith and Roberts (1993) in Bayesian approach on modelling the SV models. He adopted a Gibbs sampling scheme. On the other hand, Shephard and Pitt (1997) employed a Metropolis-Hastings scheme.

Although a number of improvements have been proposed to make statistical inference possible and simple, a deviation from the normality assumption for the time series data does impose an increasing computational burden to the analysis. Moreover, many financial data always exhibits a thick tailed behaviour. Therefore, statisticians and econometricians attempted to model asset returns using heavy-tailed distributions such as Student-t distribution, symmetric stable distribution or exponential-power distribution. The Student-t distribution had been considered by Harvey *et al.* (1996) and Jacquier *et al.* (1994) while exponential-power distribution was adopted by Nelson (1988, 1994). Steel (1998) further introduced SV models with flexible tails. More information on the asymmetric choice of distribution can be found in Harvey and Shephard (1996) and Fernández and Steel (1998). However, the extension from normal to heavy-tailed distributional assumption increases the computation effort dramatically. Therefore, some researchers suggested to express the distribution into scale mixture form to ease the calculation. Andrew and Mallow (1974) had studied scale mixtures of normal distribution. West (1987) further investigated the stable distribution on normal scale mixture form. Walker and Gutiérrez-Peña (1999) discussed the robustifying Bayesian procedure in which a generalized class of densities, eg. the scale mixture of uniform distribution was introduced. On the other hand, Choy and Chan (2000) studied the use of exponential power distribution on the asset returns with a scale mixtures of uniform form. In fact, the use of scale mixture of distributions in the Bayesian computation can be easily performed and at the same time, it can speed up the calculation for the SV models.

This chapter aims at studying the Stochastic Volatility models through the scale mixture distributions in a Bayesian approach. We assume the asset return to follow Student-t distribution instead of normal distribution. A new two-stage scale mixtures representation for the Student-t density will be proposed, ie. we will express the Student-t distribution into a scale mixtures of normal form first. Then the normal form will be further expressed as a scale mixtures of uniform form. This can further simplify the Bayesian computation for the analysis of Stochastic Volatility (SV) models from Bayesian point of view. The structure of the chapter is as follows. In Section 2 we introduce a two-stage scale mixtures representation for the Student-t density function. Section 3 presents a basic SV model in a Bayesian framework for asset return with Student-t distribution and develops the basic formulae for the Gibbs sampling algorithm. For illustrative purpose, an exchange rate data set is analysed in Section 4. Finally, a discussion is given in Section 5.

7.2 Two-stage Scale Mixtures Representation for Student-t Distribution

The Student-t distribution will be represented as a two stage scale mixtures form in this section. We will first use a scale mixture of normals (SMN) to express the Student-t distribution. The SMN form is further expressed as a scale mixtures of uniform form (SMU). So, this is a new SMN-SMU form on the Student-t distribution. Before expressing the Student-t distribution into scale mixtures form, we discuss the application of SMN and SMU first.

The Student-t distribution is one of the examples of the well-known SMN distributions. West (1987) and Choy and Smith (1997b) had studied the robustness properties of these distributions. Wakefield *et al.* (1994), Choy and Smith (1997a) and Fernandez and Steel (2000) had also investigated the applications of these kinds of distributions in Bayesian hierarchical models. On the other hand, Andrews and Mallows (1974) characterized the univariate class of scale mixtures of normal (SMN) distributions using the Laplace transformation approach. The scale mixtures of normal form (SMN) is defined as follows:

Let X be a standardised random variable having this normal scale mixtures representation. It can be expressed in the form of $X = Z \times \lambda$ where Z is the standard normal random variate and λ is a positive random variate known as the mixing variable, having a mixing distribution g, which can be either continuous or discrete.

We further let θ and σ be the location and scale parameters of the scale mixtures of normal random variable X. Then the probability density function of X has the following mixture form

$$f(x) = \int_{R^+} N(x|\theta, \lambda\sigma^2)\, g(\lambda) d\lambda \tag{7.1}$$

where $N(\cdot|\cdot)$ denotes the normal density, $g(\cdot)$ is a probability density function defined on $\Re^+ = (0, \infty)$ and λ is referred to as a mixing parameter which is commonly used as a global diagnostic check for outliers. See Choy and Smith (1997a) for details. In Bayesian framework, the mixture density in (7.1) can be expressed into a two-stage hierarchy of the form

$$\begin{aligned} X|\theta, \sigma^2, \lambda &\sim N(\theta, \lambda\sigma^2) \\ \lambda &\sim g(\lambda). \end{aligned}$$

For suitably chosen mixing density function g, a wide class of symmetric and unimodal SMN distributions can be obtained. In particular, the Student-t distribution with degrees of freedom α corresponds to an inverse gamma mixing distribution, i.e.

$$\lambda \sim IG\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right)$$

where $IG(a, b)$ is the inverse gamma distribution with density

$$g(\lambda) = \frac{b^a}{\Gamma(a)\lambda^{a+1}} \, e^{-b/\lambda} \qquad \lambda > 0, \; a, b > 0.$$

To facilitate an efficient computation for the SV models, we further make use of the class of scale mixtures of uniform (SMU) representation for the normal density. Details about the SMU distributions can be found in Walker and Gutiérrez-Peña (1999), Damien *et al.* (1999) and Choy and Walker (1999). Suppose X is a normal random variable with mean θ and variance σ^2, it can be easily shown that its density function can be rewritten into SMU form as follows:

$$N(x|\theta, \sigma^2) = \int_{\theta-\sigma\sqrt{u}}^{\theta+\sigma\sqrt{u}} \frac{1}{2\sigma\sqrt{u}} \, Ga\left(u \left| \frac{3}{2}, \frac{1}{2}\right.\right) du$$

where $Ga(u|a, b)$ is the gamma density function with parameters a and b. Theoretically, all SMN distributions can be expressed into the form of SMU distributions. Therefore, we can express the Student-t distribution with degrees of freedom α into the following hierarchy

$$\begin{aligned} X|\theta, \sigma^2, \lambda, u &\sim U(\theta - \sigma\lambda^{1/2} u^{1/2}, \theta + \sigma\lambda^{1/2} u^{1/2}) \\ \lambda &\sim IG\left(\frac{\alpha}{2}, \frac{\alpha}{2}\right) \\ u &\sim Ga\left(\frac{3}{2}, \frac{1}{2}\right) \end{aligned}$$

where $U(a, b)$ is a uniform distribution defined on the interval (a, b). Now, we shall show the advantages of using this two-stage scale mixtures representation for the Student-t distribution in the SV models in Section 3.

7.3 Bayesian Student-t SV Models

For modelling financial data, more and more attention has been focused on using heavy-tailed distributions such as Student-t, symmetric stable and exponential power distributions etc. Student-t distribution is probably the most popular among them and is therefore used here.

Let r_t be the asset value of an equity at time $t = 0, 1, 2, \ldots, n$. The mean adjusted asset return y_t at time t is defined as

$$y_t = \ln\left(\frac{r_t}{r_{t-1}}\right) - \frac{1}{n}\sum_{i=1}^{n} \ln\left(\frac{r_i}{r_{i-1}}\right), \qquad t = 1, 2, \ldots, n.$$

Let H_t and h_t be the volatilities and log-volatilities, respectively. The standard SV model for the asset return, y_t is defined as

$$\begin{aligned} y_t &= \beta \exp(h_t/2)\, \epsilon_t, \qquad t = 1, 2, \ldots, n \\ h_t &= \begin{cases} \sigma\eta_1/\sqrt{1-\phi^2} & t = 1 \\ \phi\, h_{t-1} + \sigma\eta_t & t > 1 \end{cases} \end{aligned}$$

where $\{\epsilon_t\}$ and $\{\eta_t\}$ are independent standard Gaussian processes. β is a constant factor that represents the model instantaneous volatility which is usually set to one in many literatures. σ is the variance of the log-volatilities and $\phi \in (-1, 1)$ is the persistence of the volatility.

The mean adjusted returns are then modelled by a Student-t distribution while the log-volatility is assumed to follow a normal distribution. In fact, we can also use heavy-tailed distribution for modelling the log-volatility to achieve a higher degree of robustness. However, our emphasis is on the development of an efficient simulation algorithm for the Gibbs sampler scheme. The extension to robustifying the log-volatility can be done without substantial increase in computational effort.

7.3.1 Advantage of two-stage scale mixtures representation in the Bayesian t-N SV models

The advantage of two-stage scale mixtures representation is on the simplification of calculation on the conditional distribution in the Gibbs sampler. Moreover, it enhances the identification of potential outliers.

In the current setup, we write

$$y_t|h_t \overset{i.i.d.}{\sim} t_\alpha\left(0, \beta^2 \exp(h_t)\right) \qquad t = 1, 2, \ldots, n.$$

The conditional distribution of the log-volatility h_t has a normal distribution of the form

$$h_t|h_{t-1}, \phi, \sigma^2 \overset{i.i.d.}{\sim} N\left(\phi h_{t-1}, \sigma^2\right)$$

while the marginal distribution is

$$h_t|\phi, \sigma^2 \overset{i.i.d.}{\sim} N\left(0, \frac{\sigma^2}{1-\phi^2}\right)$$

where $\phi \in (-1, 1)$ is the instantaneous persistence parameter and σ is a scale parameter of the log-volatility. As in many literatures, we may assume that β is fixed. For simplicity, we shall call this a t-N SV model. To complete a Bayesian framework, we assign independent priors for ϕ and σ^2 which follow shifted beta and inverse gamma distributions, respectively, i.e.

$$\frac{\phi+1}{2} \sim Be(a_\phi, b_\phi) \qquad \text{and} \qquad \sigma^2 \sim IG(a_\sigma, b_\sigma)$$

If we express the student-t distribution into one-stage representation (SMN), the conditional distributions for σ^2 and $\boldsymbol{\lambda}$ are inverse gamma distributions while that for ϕ is a mixture of normal and shifted beta distribution. However, the conditional distribution for h_t is a non-standard distribution which is difficult to simulate. We can solve this problem by the use of two-stage scale mixtures representation since after further expressing the SMN into SMU, the conditional distribution for h_t becomes a truncated normal distribution which can be simulated easily. Moreover, with the mixing parameters in SMN and SMU representation, it enhances the identification of potential outliers. Because of the advantages we describe above, we suggest the use of two-stage scale mixtures representation (SMN-SMU) in this chapter. Now we can rewrite the t-N SV model hierarchically as

$$
\begin{aligned}
y_t|h_t, \lambda_t, u_t &\sim U\left(-\beta \exp(h_t/2)\lambda_t^{1/2}u_t^{1/2}, \beta \exp(h_t/2)\lambda_t^{1/2}u_t^{1/2}\right) \\
\lambda_t &\sim IG\left(\alpha/2, \alpha/2\right) \\
u_t &\sim Ga\left(3/2, 1/2\right) \\
h_t|h_{t-1}, \phi, \sigma^2 &\sim N\left(\phi h_{t-1}, \sigma^2\right) \\
\frac{\phi+1}{2} &\sim Be\left(a_\phi, b_\phi\right) \\
\sigma^2 &\sim IG\left(a_\sigma, b_\sigma\right)
\end{aligned}
$$

for $t = 1, 2, \ldots, n$.

7.3.2 Gibbs Sampler for the t-N SV model

For statistical analysis, we implement the SV model using simulation-based Gibbs sampling approach. Let $\boldsymbol{y} = (y_1, \ldots, y_n)$, $\boldsymbol{h} = (h_1, \ldots, h_n)$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$, $\boldsymbol{u} = (u_1, \ldots, u_n)$ and, for $i = 1, \ldots, n$, $\boldsymbol{h}_{-t} = (h_1, \ldots, h_{t-1}, h_{t+1}, \ldots, h_n)$, $\boldsymbol{\lambda}_{-t} = (\lambda_1, \ldots, \lambda_{t-1}, \lambda_{t+1}, \ldots, \lambda_n)$ and $\boldsymbol{u}_{-t} = (u_1, \ldots, u_{t-1}, u_{t+1}, \ldots, u_n)$. With arbitrarily chosen starting values for $\boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}, \sigma^2$ and ϕ, the Gibbs sampler iteratively samples random variates from a system of full conditional distributions and the resulting simulations are used to mimic a random sample from the target joint posterior distribution. Now the system of full conditionals is given below:-

1. Full conditional distribution for $h_t, t = 1, \ldots, n$:-

$$h_t | \boldsymbol{h}_{-t}, \boldsymbol{\lambda}, \boldsymbol{u}, \sigma^2, \phi, \boldsymbol{y} \sim N(a_t, b_t \sigma^2) \qquad h_t > \ln y_t^2 - 2\ln\beta - \ln\lambda_t - \ln u_t$$

where

$$a_t = \begin{cases} \phi\, h_{t+1} - \sigma^2/2 & t = 1 \\ (1+\phi^2)^{-1}\Big(\phi(h_{t-1} + h_{t+1}) - \sigma^2/2\Big) & 2 \le t \le n-1 \\ \phi\, h_{t-1} - \sigma^2/2 & t = n \end{cases}$$

and

$$b_t = \begin{cases} 1 & t = 1, n \\ (1+\phi^2)^{-1} & 2 \le t \le n-1 \end{cases}$$

2. Full conditional distribution for σ^2:-

$$\sigma^2 | \boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}, \phi, \boldsymbol{y} \sim IG\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}\left[(1-\phi^2)h_1^2 + \sum_{t=2}^{n}(h_t - \phi h_{t-1})^2\right]\right)$$

3. Full conditional distribution for $\lambda_t, t = 1, \ldots, n$:-

$$\lambda_t | \boldsymbol{h}, \boldsymbol{\lambda}_{-t}, \boldsymbol{u}, \phi, \sigma^2, \boldsymbol{y} \sim IG\left(\frac{\alpha+1}{2}, \frac{\alpha}{2}\right) \qquad \lambda_t > \frac{y_t^2}{\beta^2 \exp(h_t) u_t}$$

4. Full conditional distribution for $u_t, t = 1, \ldots, n$:-

$$u_t | \boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}_{-t}, \phi, \sigma^2, \boldsymbol{y} \sim Exp\left(\frac{1}{2}\right) \qquad u_t > \frac{y_t^2}{\beta^2 \exp(h_t)\lambda_t}$$

5. Full conditional distribution for ϕ:-

$$\phi | \boldsymbol{h}, \boldsymbol{\lambda}, \boldsymbol{u}, \sigma^2, \boldsymbol{y} \sim N\left(\phi \,\middle|\, \frac{\sum_{t=2}^{n} h_{t-1}h_t}{\sum_{t=2}^{n-1} h_t^2}, \frac{\sigma^2}{\sum_{t=2}^{n-1} h_t^2}\right) (1+\phi)^{a_\phi - 1/2}(1-\phi)^{b_\phi - 1/2} \qquad |\phi| \le 1.$$

In the above expressions, conditional distributions for h_t, λ_t, u_t and σ^2 are the truncated normal, truncated inverse gamma, truncated exponential and inverse gamma distributions while that for ϕ is a product of a normal and a shifted beta distributions. Simulation of random variates from truncated exponential distribution can be straightforwardly done by the inversion method while simulation from truncated normal and gamma distributions can be done by applying the algorithms proposed by Robert (1995) and Philippe (1997), respectively. However, to simulate the non-standard density of ϕ, we need to use the rejection method, or other simulation methods.

7.4 Illustrative Example

To demonstrate our proposed t-N SV model, we analyse the exchange rates of US dollars to Sterling pounds. Fig. 7.1 presents 1000 mean adjusted daily closing exchange rate returns from January 2, 1981. Obviously, some trading days produce extreme returns. In the following simulation study, we set $\beta = 1$ for simplicity. To reflect the non-informative prior knowledge about the variance of the log-volatility σ^2, we choose $a_\sigma = b_\sigma = 0.001$. We adopt a beta $Be(20, 1)$ informative prior for $(\phi + 1)/2$. In the example, we ran the Gibbs sampler for a single series of 35000 iterations. The first 5000 iterations are discarded as the 'burn-in' period. To avoid high correlation between successive drawings of the financial data, we pick up simulated values at every 30th value to mimic a random sample of 1000 drawings from the target joint posterior distribution. Fig. 7.2 displays the ACF of $h_1, \lambda_1, u_1, \sigma$ and ϕ based on these 1000 values. The sampled values are rather uncorrelated except for σ.

==

Fig. 7.1, Fig. 7.2 and Table 7.1 about here

==

7.4.1 Parameter estimation

For various degrees of freedom α, posterior estimates of σ and ϕ, together with the standard errors and 95% posterior intervals, are presented in Table 7.1 and Table 7.2, respectively and the corresponding boxplots are presented in Fig. 7.3. We notice that σ increases gradually with α. It can be seen that a heavier tailed distribution for the returns results in a smaller variation of the log-volatility and a higher degree of persistence. Fig. 7.4 plots the volatilities, H_t, over the study period for a Cauchy and a normal sampling distributions, respectively and shows that small H_t values (and h_t values) are associated with small α values. This phenomenon can be explained by the robustifying property of the Student-t distribution that it protects inference of the volatility as its tails are heavy enough to capture the outlying observations.

==================================

Fig. 7.3 and Fig. 7.4 about here

==================================

7.4.2 Outlier diagnosis

We propose the SMN-SMU scale mixtures presentation for the Student-t density in SV models in which computational issues and outlier identification are our primary concern. Though this is not compulsory as many other different simulation methods can be employed, our approach does simplify the system of full conditional distributions and hence provides a very efficient Gibbs sampling scheme. In addition, mixing parameters λ_t and u_t enable us to perform a global diagnosis of possible outliers. The posterior means of λ_t and u_t under a Cauchy-N SV model are plotted in Fig. 7.5. A large values of λ_t and u_t correspond to possible outlying daily returns. However, u_t cannot capture the outliers well. It may be due to the fact that u_t is the mixing parameter of the second stage scale mixtures representation and the outliers has been mainly captured by the first stage mixing parameter λ_t already. Furthermore, the product of the posterior means of λ_t and u_t under a Cauchy-N SV model is plotted in Fig. 7.6. Large values of $\lambda_t\, u_t$ also correspond to possible outlying daily returns. The magnitude of $\lambda_t\, u_t$ (product of λ_t and u_t) in Fig. 7.6 is larger than that of λ_t in Fig. 7.5. So, with the two-stage scale mixtures representation, it can enhance the identification of outliers. With reference to Fig. 7.1, extreme daily returns are clearly identified in Fig. 7.6. The five most influential daily returns are on Days 106, 697, 840, 34 and 858, respectively. Table 7.3 presents the extreme daily returns and their volatilities for Cauchy-N and N-N SV models. The volatilities of these days for the Cauchy-N and N-N models are very different - the former's volatilities are smaller than those of the latter's because the Cauchy-N model can capture outliers much more effectively than the N-N model.

It is well-known that normal conjugate models do not provide a robust analysis, the advantage of using heavy-tailed distribution is obvious when we compare the results for using the Cauchy-N and N-N models. The extreme daily returns are automatically downweighed in the analysis and inference of H_t and hence h_t are protected. We can notice from Table 7.1 that the posterior estimate of σ drops from 0.3796 to 0.3293 when the sampling distribution changes from normal to Cauchy.

==

Table 7.3 and Fig. 7.5 and Fig. 7.6 about here

==

7.4.3 Model selection

To choose the most suitable value of α for the $t - N$ SV model, we adopt the model selection criterion suggested by San Martini and Spezzaferri (1984). Let M_α be the t-N SV model with α degrees of freedom. The posterior expected utility $U(\alpha)$ of this model is defined by

$$U(\alpha) = \frac{1}{n} \sum_{t=1}^{n} \ln p\left(y_t | M_\alpha\right)$$

and is evaluated from the Gibbs sampling outputs. The best model corresponds to the one that gives the largest value of $U(\alpha)$. Fig. 7.7 shows the pattern of the posterior expected utility $U(\alpha)$ against α. Obviously, the Cauchy-N model is the best choice in this comparative study.

================================

Fig. 7.7 about here

================================

Alternatively, we can treat the degrees of freedom α as random and assign, probably, a gamma $G(a_\alpha, b_\alpha)$ prior distribution to it. In that case the full conditional density of α has the form

$$
\begin{aligned}
p(\alpha|\boldsymbol{h},\boldsymbol{\lambda},\boldsymbol{u},\sigma^2,\phi) &\propto p(\boldsymbol{\lambda}|\alpha)\,p(\alpha) \\
&\propto \left(\frac{(\alpha/2)^{\alpha/2}}{\Gamma(\alpha/2)}\right)^n Ga\left(\alpha \,\middle|\, a_\alpha, b_\alpha + \frac{1}{2}\sum_{t=1}^{n}(\lambda_t - \ln\,\lambda_t)\right)
\end{aligned}
$$

where random variates can be simulated using, for example, the Metropolis-Hastings and ratio-of-uniforms (see Wakefield *et al.*, 1991) algorithms or using the adaptive rejection sampling (see Gilks and Wild, 1992) algorithm as the above conditional density can easily be shown to be log-concave.

7.5 Discussion

Student-t distribution has been widely used for modelling heavy-tailed events and for robustness consideration. For Bayesian inference, statisticians always express the distribution into a scale mixtures of normal form so as to simplify the Bayesian calculations, in particular when Gibbs sampling scheme is used. See Wakefield *et al.* (1994) and Choy and Smith (1997b) for details. For some complicated models where data are dependent such as the GARCH and SV models, the SMN representation may not provide a substantial improvement in computational efficiency. The significance here is on the introduction of a two-stage scale mixture representation for the Student-t density. This new representation enables us to simplify the Bayesian computation for the SV models. Similar to the SMN representation, the product of the mixing parameters, $\lambda_i u_i$, provides a mean for outlier diagnosis.

We do not attempt to find the most suitable model for modelling the returns of the exchange rates. In fact, we can assume the degrees of freedom α to be an unknown quantity and assign a suitable prior distribution to it. In addition, for robustification consideration, it is an advantage to adopt heavy-tailed distribution as an alternative to the normal distribution in statistical modelling since the use of a heavy-tailed distributions can protect our inference from model misspecification when the normal assumption is not correct. In addition, other heavy-tailed distributions can be used to model the log-volatilities and we expect that this additional modification can further protect inference from unusual volatilities of the time series.

α	$E[\sigma\|\boldsymbol{y}]$	$SE[\sigma\|\boldsymbol{y}]$	95% posterior interval
1	0.3293	0.0411	(0.2536, 0.4238)
3	0.3304	0.0434	(0.2530, 0.4209)
5	0.3384	0.0456	(0.2563, 0.4404)
10	0.3524	0.0473	(0.2652, 0.4519)
15	0.3611	0.0484	(0.2772, 0.4652)
20	0.3623	0.0478	(0.2791, 0.4638)
∞	0.3796	0.0507	(0.2863, 0.4906)

Table 7.1: Bayes estimates, standard errors and 95% confidence intervals of σ for various α values.

α	$E[\phi\|\boldsymbol{y}]$	$SE[\phi\|\boldsymbol{y}]$	95% posterior interval
1	0.9903	0.0053	(0.9780,0.9988)
3	0.9839	0.0079	(0.9658,0.9967)
5	0.9802	0.0089	(0.9604,0.9960)
10	0.9752	0.0111	(0.9508,0.9953)
15	0.9725	0.0122	(0.9443,0.9930)
20	0.9713	0.0121	(0.9442,0.9923)
∞	0.9656	0.0140	(0.9370, 0.9902)

Table 7.2: Bayes estimates, standard errors and 95% confidence intervals of ϕ for various α values.

Time t	$y(t)$	H_t under N-N model	H_t under Cauchy-N model
34	2.8239	1.4833	0.2250
106	3.7680	1.9490	0.2851
697	-1.3220	0.3100	0.0387
840	1.5361	2.2450	0.0542
858	-1.3656	0.3231	0.0647

Table 7.3: Observed values of the extreme daily returns and their Bayes estimates of H_t under a Normal-Normal and Cauchy-Normal models.

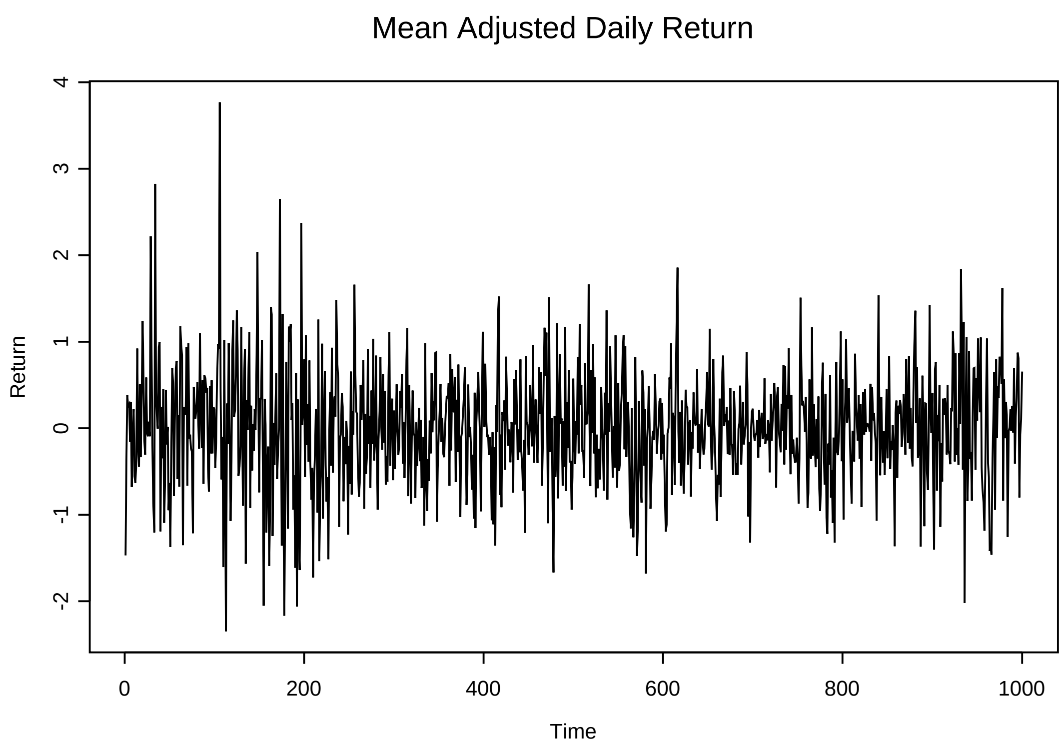

Fig. 7.1. *Time series plot of mean adjusted daily exchange rate returns.*

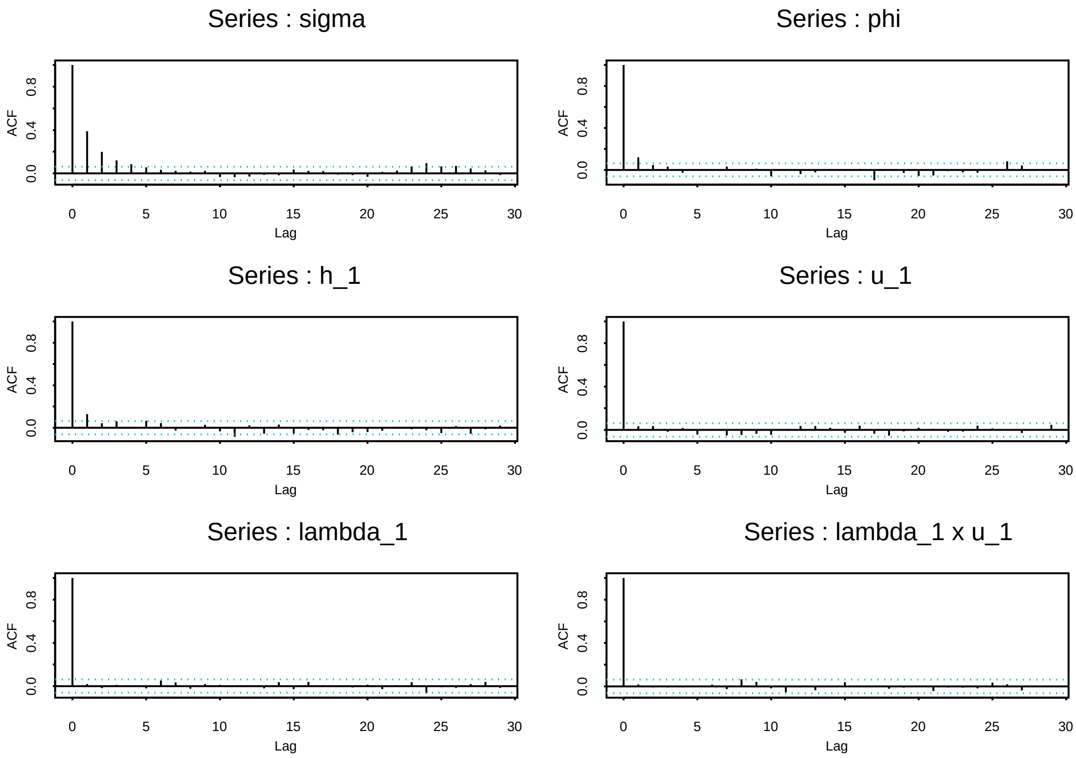

Fig. 7.2. *Autocorrelation functions of 1000 simulated values of* $\sigma, \phi, h_1, \lambda_1, u_1, \lambda_i$ *and* $\lambda_1 u_1$.

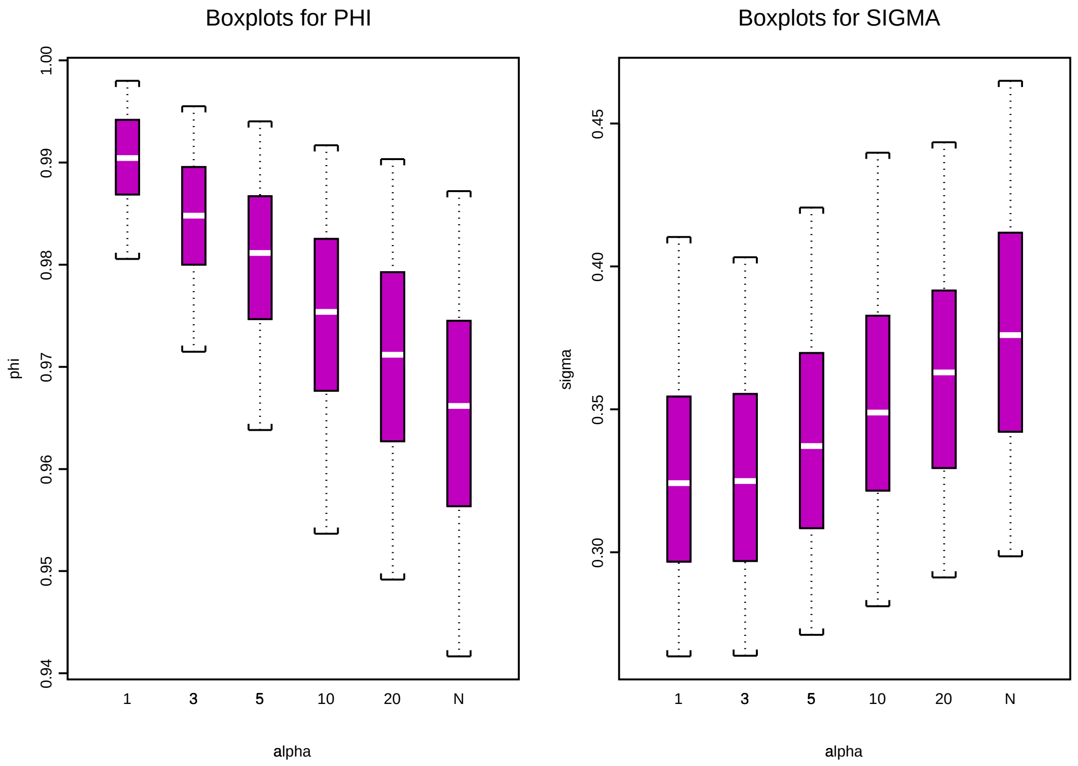

Fig. 7.3. *Boxplots of σ and ϕ for different degrees of freedom α. The whiskers correspond to 5%, 25%, 50%, 75% and 95% of the values.*

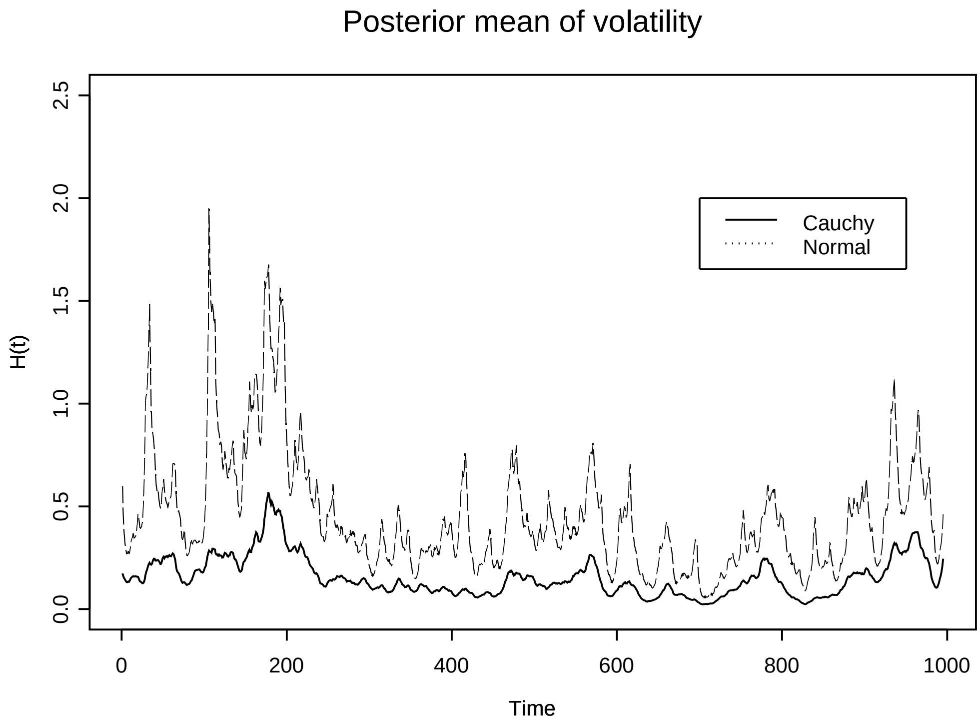

Fig. 7.4. *Bayes estimates of volatility H_t for Cauchy-Normal and Normal-Normal SV models, respectively.*

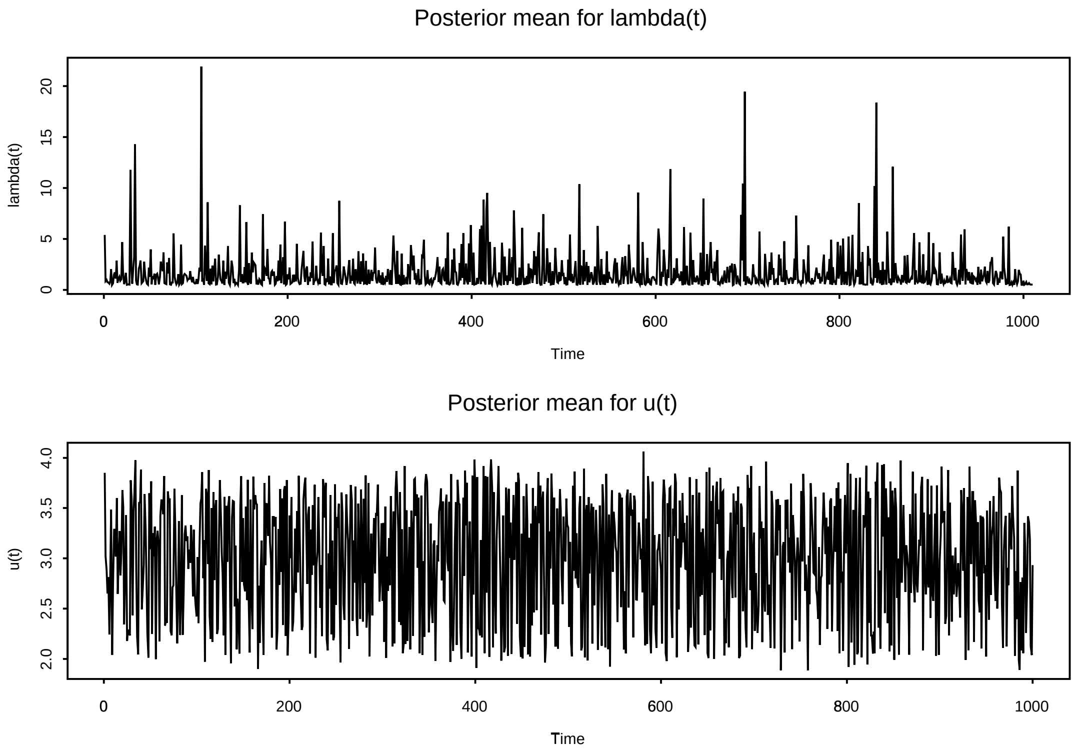

Fig. 7.5. *Bayes estimates of the mixing parameters, λ_t and u_t for the Cauchy-Normal SV model. Large values correspond to extreme daily returns.*

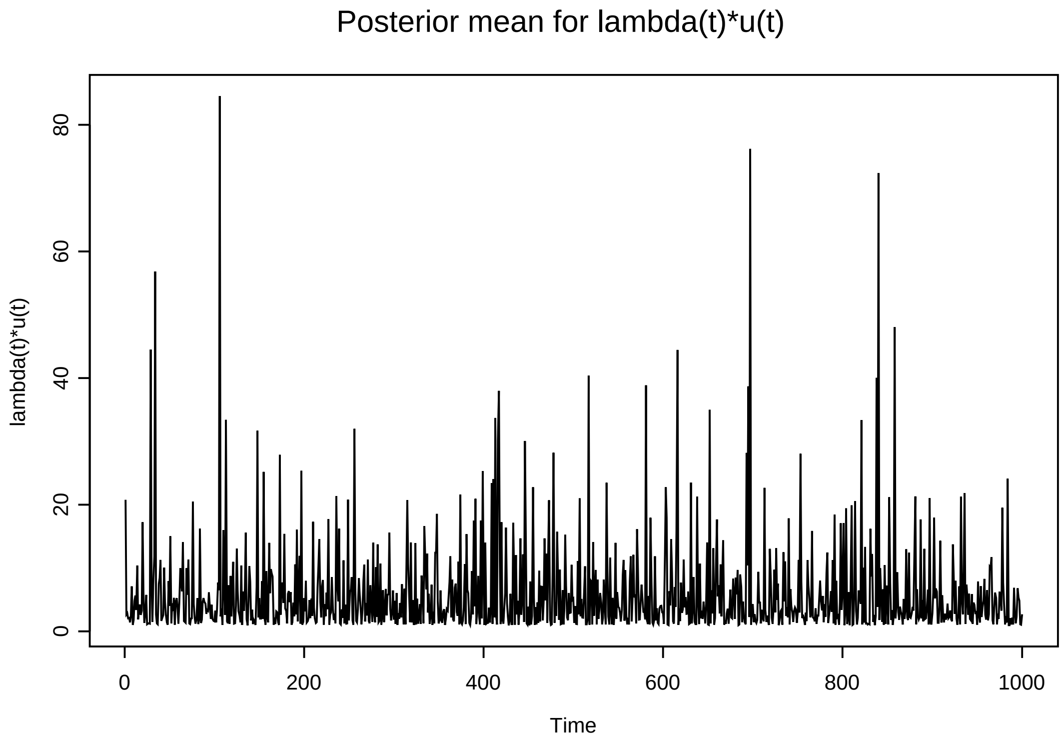

Fig. 7.6. *Bayes estimates of the product of mixing parameters,* $\lambda_t u_t$ *for the Cauchy-Normal SV model. Large values correspond to extreme daily returns.*

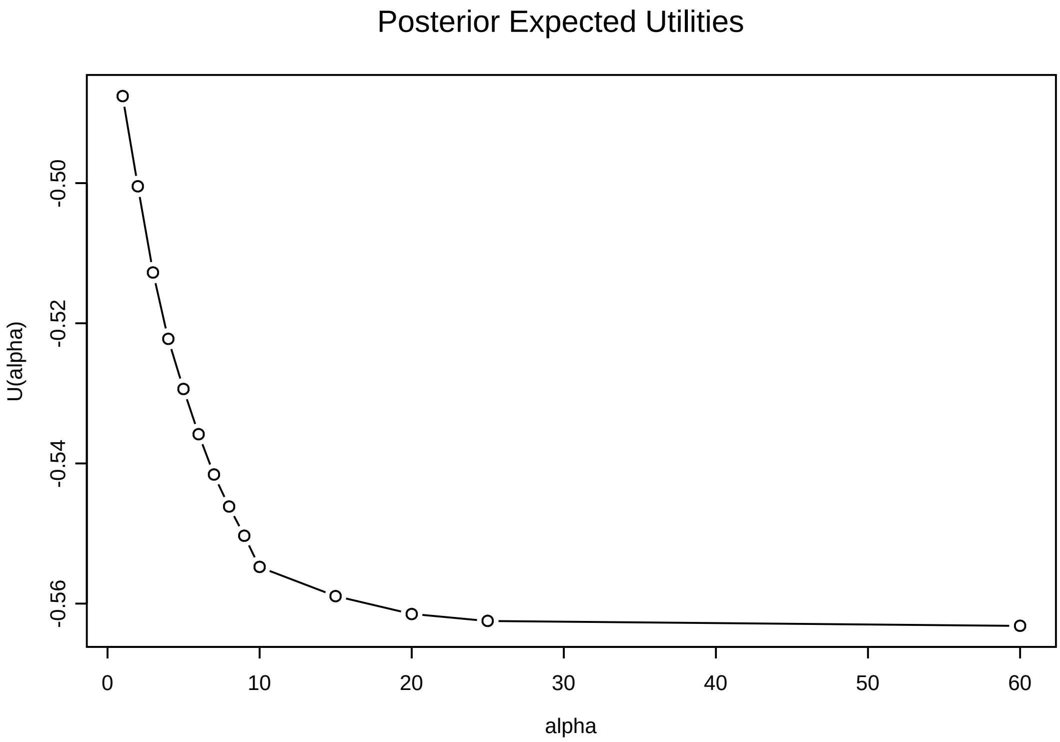

Fig. 7.7. *Posterior expected utility* $U(\alpha)$ *of* $t-N$ *SV models with different degrees of freedom* α*.*

REFERENCES

[1] Andrews, D. F. and Mallows, C. L. (1974), "Scale mixtures of normal distribution," *Journal of the Royal Statistical Society, B*, **36**, 99-102.

[2] Baillie, R. T., Bollerslev, T. and Mikkelson, H. O. (1996), "Fractionally integrated generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, **74**, 3-30.

[3] Bernardo, J. M. and Smith, A. F. M. (1994), *Bayesian Statistics*, Wiley.

[4] Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, **31**, 307-327.

[5] Booth, J. G. and Hobert, J. P. (1999), "Maximizing generalized linear mixed model likelihoods with an automated Monte Carlo EM algorithm," *Journal of the Royal Statistical Society, B*, **61**, 265-285.

[6] Breslow, N. E. and Clayton, D.G. (1993), "Approximate inference in generalized linear mixed models," *Journal of American Statistical Association*, **88**, 9-25.

[7] Carter, C. K. and Kohn, R. (1996), "Markov chain Monte Carlo in conditionally Gaussian state space models," *Biometrika*, **83**, 589-601.

[8] Chan, J. S. K. and Kuk, A. Y. C. (1997), "Maximum likelihood estimation for

probit-linear mixed models with correlated random effects," *Biometrics*, **53**, 86-97.

[9] Chib, S. (1995), "Marginal Likelihood from the Gibbs Output," *Journal of American Statistical Association*, **90**, 1313–1321.

[10] Clayton, D. G. (1996), "Generalized linear mixed models," in *Markov chain Monte Carlo in practice*, eds. W.R. Gilks, S. Richardson and D.J. Spiegelhalter, London: Chapman & Hall, 275-302.

[11] Choy, S. T. B. and Chan, C. M. (2000), "Bayesian estimation of stochastic volatility model via scale mixtures distributions," in *Proceedings of the Hong Kong International Workshop on Statistics and Finance: An interface*, Imperial College Press. 185-204.

[12] Choy, S. T. B. and Smith, A. F. M. (1997a), "On robust analysis of a normal location parameter," *Journal of the Royal Statistical Society, B*, **59**, 463-474.

[13] Choy, S. T. B. and Smith, A. F. M. (1997b), "Hierarchical models with scale mixtures of normal distribution," *TEST*, **6**, 205-211.

[14] Choy, S. T. B. and Walker, S. G. (1999), "The extended exponential power distribution and Bayesian robustness," Technical Report, University of Hong Kong.

[15] Damien, P., Walkefield, J. C. and Walker, S. G. (1999), "Gibbs sampling for Bayesian non-conjugate and hierarchical models using auxiliary variables," *Journal of the Royal Statistical Society, B*, **61**, 331-344.

[16] Dempster, A. P., Larid, N. M., and Rubin, D. B. (1997), "Maximum likelihood from incomplete data via the Em algorithm," *Journal of the Royal Statistical Society, B*, **59**, 1-38.

[17] De Jong, P. and Shephard, N. (1995), "The Simulation Smoother for Time-Series Models," *Biometrika*, **82**, 339-850.

[18] Engle, R. F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation," *Econometrica*, **50**, 987-1006.

[19] Engle, R. F. and Bollerslev, T.(1986), "Modelling the persistence of conditional variance," *Econometric Review*, **5**, 1-50.

[20] Engle, R. F., Ng, V. K. and Rothschild, M. (1990), "Asset pricing with a FACTOR-ARCH covariance structure: Empirical estimates for Treasury Bills," *J. Econometrics*, **45**, 213-237.

[21] Engle, R.F. (1995) *ARCH Selected Readings*, Oxford University Press: New York.

[22] Fernandez, C. and Steel, M. F. J. (1998), "On Bayesian modelling of fat tails

and skewness," *Journal of American Statistical Association*, **93**, 359-371.

[23] Fernandez, C. and Steel, M. F. J. (2000), "Bayesian regression analysis with scale mixtures of normals," *Econometric Theory*, **16**, 80-101.

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

[25] Gelfand, A. and Carlin, B. (1993) "Maximum-likelihood estimation for constrained- or missing data models," *Canadian Journal of Statistics*, **21**, 303-311.

[26] Geman, S., and Geman, D. (1984) "Stochastic relaxation, Gibbs Distributions and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6**, 721-741.

[27] Geyer, C. J. and Thompson, E. A. (1992), "Constrained Monte Carlo maximum likelihood for dependent data," *Journal of the Royal Statistical Society, B*, **54**, 657–699.

[28] Geyer, C. J. (1994), "On the convergence of Monte Carlo Maximum Likelihood Calculations," *Journal of the Royal Statistical Society, B*, **56**, 261–274.

[29] Gilks, W. R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs

sampling," *Applied Statistics*, **41**, 337-348.

[30] Green, P. J. (1990) "On use of EM algorithm for penalized likelihood estimation," *Journal of the Royal Statistical Society, B*, **52**, 443-452.

[31] Hamilton, J. D. and Susmel, R. (1994), "Autoregressive conditional heteroskedasticity and changes in regime," *Journal of Econometrics*, **64**, 307-333.

[32] Harvey, A. C., Ruiz, E. and Shephard, N. (1994), "Multivariate stochastic variance models," *Review of Economic Studies*, **61**, 247-264. Reprinted as 256-276.

[33] Harvey, A.C. and Shephard, N. (1996), "Estimation of an Asymmetric Stochastic Volatility Model for Asset Returns," *Journal of Business and Economic Statistics*, **14**, 429-434.

[34] Hastings, W. K. (1970) "Monte Carlo sampling methods using Markov Chains and their applications," *Biometrika*, **57**, 97-109.

[35] Hull, J. and White, A. (1987), "The pricing of options on assets with stochastic volatilities," *Journal of Finance*, **42**, 281-300.

[36] Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), " Bayesian analysis of stochastic volatility models (with discussion)," *Journal of Business and Economic Statistics*, **12**, 371-417.

[37] Karim, M. R. and Zeger, S. L. (1992), "Generalized linear models with random effects; salamander mating revisited," *Biometrics*, **48**, 631–644.

[38] Kim, S., Shephard, N. and Chib, S. (1998), "Stochastic volatility: likelihood inference and comparison with ARCH models," *Review of Economic Studies*, **65**, 361-393.

[39] Kuk, A. Y. C. (1995), "Asymototically unbiased estimation in generalized linear models with random effects," *Journal of Royal Statistical Society, B*, **57**, 395–407.

[40] Kuk, A. Y. C. (1999), "Laplace importance sampling for generalized linear mixed models," *Journal of Statistical Computing Simulation*, **63**, 143-158.

[41] Kuk, A. Y. C. (2002), "Automatic choice of driving values in Monte Carlo likelihood approximation via posterior simulation," *Statistics and Computing*, To appear.

[42] Kuk, A. Y. C. and Cheng, Y. W. (1997), "The Monte Carlo Newton-Raphson Algorithm," *Journal of Statistical Computing Simulation*, **59**, 233–250.

[43] Kuk, A. Y. C. and Cheng, Y. W. (1999), "Pointwise and functional approximations in Monte Carlo maximum likelihood estimation," *Statistics and Computing*, **9**, 91-99.

[44] Lee, P. M. (1989), *Bayesian Statistics: An Introduction* , Edward Arnold.

[45] Lin, X. and Breslow, N. E. (1996), "Bias correction in generalized linear mixed models with multiple components of dispersion," *Journal of American Statistical Association*, **91**, 1007–1016.

[46] Liu, Q. and Pierce, D. A. (1994), "A note on Gauss-Hermite quadrature," *Biometrika*, **81**, 624-629.

[47] McCullagh, P. and Nelder, J. A. (1989), *Generalized Linear Models* 2nd edn. London: Chapman and Hall.

[48] McCulloch, C. E. (1994), "Maximum likelihood variance components etimation for binary data," *Journal of American Statistical Association*, **89**, 330–335.

[49] McCulloch, C. E. (1997), "Maximum likelihood algorithms for generalized linear mixed models," *Journal of American Statistical Association*, **92**, 162–170.

[50] McCulloch C. E. and Searle S. R. (2001) *Generalized, linear, and mixed models*, New York: John Wiley & Sons, Inc.

[51] Melino, A. and Turnbull, S. M. (1990), "Pricing foregin currency options with stochastic volatility," *Journal of Econometrics*, **45**, 239-265.

[52] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H.

(1953), "Equations of State calculations by fast computing machines," *Journal of Chemical Physics*, **21**, 1087-1091.

[53] Nelder, J. A. and Wedderburn, R. W. M. (1972), "Generalized linear models," *Journal of the Royal Statistical Society, A*, **135**, 370-384.

[54] Nelson, D. B. (1988), "Time series behavior of stock market volatility and returns," Unpublished Ph.D Dissertation, Massachusetts Institute of Technology.

[55] Nelson, D. B.(1991), "Conditional heteroskedasticity in asset returns: A new approach," *Econometrica*, **59**, 347-370.

[56] Nelson, D. B.(1994), "Comment on Bayesian analysis of stochastic volatility models," *Journal of Business and Economic Statistics*, **12**, 403-406.

[57] Philippe, A. (1997), "Simulation of right and left truncated gamma distribution by mixtures," *Statistics and Computing*, **7**, 173-181.

[58] Ripley, B. (1987), *Stochastic Simulation*, New York: John Wiley.

[59] Robert, C.P. (1995), "Simulation of truncated normal variables," *Statistics and Computing*, **5**, 121-125.

[60] San Martini, A. and Spezzaferri, F. (1984), "A predictive model selection criterion," *Journal of the Royal Statistical Society, B*, **46**, 296-303.

[61] Schall, R. (1991), "Estimation in generalized linear models with random

effects," *Biometrika*, **78**, 719-727.

[62] Shephard, N. (1996), "Statistical aspect of ARCH and stochastic volatility," in *Time Series Models: In Econometrics, Finance and other fields*, Chapman & Hall, 1-67.

[63] Shephard, N. and Pitt, M.K. (1997), "Likelihood Analysis of Non-Gaussian measurement time series," *Biometrika*, **84**, 653-667.

[64] Shun, Z. and McCullagh, P. (1995), "Laplace approximations of high dimensional integrals," *Journal of the Royal Statistical Society, B*, **57**, 749–760.

[65] Shun, Z. (1997), "Anther look at the salamander mating data: a modified Laplace approximation approach," *Journal of American Statistical Association*, **92**, 341–349.

[66] Smith, A. F. M. and Roberts, G. O. (1993), "Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society, Series B*, **55**, 3-23.

[67] So, M. K. P., Lam, K. and Li, W. K. (1998), "A stochastic volatility model with markov switching," *Journal of Business and Economic Statistics*, **16**, 244-253.

[68] Steel, M. F. J. (1998), "Bayesian analysis of stochastic volatility models with flexible tails," *Econometric Reviews*, **17**, 109-143.

[69] Taylor, S. J. (1986) *Modelling Financial Time Series*, New York: John Wiley.

[70] Taylor, S. J. (1994), "Modelling stochastic volatility," *Mathematical Finance*, **4**, 183-204.

[71] Tierney, L. (1994), "Markov Chain for exploring posterior distributions (with discussion)," *The Annals of Statistics*, **22**, 1701-1762.

[72] Wakefield, J. C., Gelfand, A. E. and Smith, A. F. M. (1991), "Efficient generation of random variates via the ratio-of-uniforms method," *Statistics and Computing*, **1**, 129-133.

[73] Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. E. and Gelfand, A. E.(1994), "Bayesian analysis of linear and non-linear population models by using the Gibbs sampler," *Applied Statistics*, **43**, 201-221.

[74] Walker, S. G. and Gutiérrez-Peña, E. (1999), "Robustifying Bayesian Procedures (with discussion)," in *Bayesian Statistics 6*, eds. J.M. Bernardo, J.O. Berger, A.P. Dawid and A.F.M. Smith, New York: Oxford University Press, 685-710.

[75] Wedderburn, R. W. M. (1974), "Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method," *Biometrika*, **61**, 439-447.

[76] Wei G. C. G. and Tanner M. A. (1990), "A Monte Carlo implementation of the EM algorithm and the Poor Man's data augmentation algorithms,"

Journal of the American Statistical Association, **85**, 699-704.

[77] West, M. (1984), "Outlier models and posterior distributions in Bayesian linear regression," *Journal of the Royal Statistical Society, Series B*, **46**, 431-439.

[78] West, M. (1987), "On scale mixtures of normal distributions," *Biometrika*, **74**, 646-648.

[79] Wolfinger, R. W. (1993), "Laplace's approximation for nonlinear mixed models," *Biometrika*, **80**, 791-795.

[80] Zeger, S. L. and Karim, M. R. (1991), "Generalized linear models with random effecrs; A Gibbs sampling approach," *Journal of American Statistical Association*, **86**, 79-86.