ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Bayesian Analysis of Structural Change in Trend

Pingping Zheng

A thesis submitted in partial fulfillment of the requirements of The Nottingham Trent University for the degree of Doctor of Philosophy

February 2002

Acknowledgements

I would like to thank Dr John Marriott for his supervision of this PhD, for his support and encouragement given to me during my PhD studies. I also thank Professor Paul Newbold and Professor Neville Davies for their guidance, especially the directional advice from Professor Paul Newbold for my research.

I also thank Professor Charles Nelson of the University of Washington, who kindly provided me the US historical time series which I used in this thesis.

Special thanks are given to my wife, Hui Yang who constantly supported me at all stages of my PhD studies. I shall be grateful for this all my life.

Abstract

In recent years there has been a lot interest in studying the structural changes in time series, especially in economic time series. Statistical methods have been developed to deal with this kind of problem. Bayesian methods are becoming more and more popular in this field.

This thesis focuses on two kinds of structural change, abrupt structural change which happens at some point and the gradual structural change which happens over a period of time. For the former structural change, we discuss the two-phase model and the structural break model; for the latter structural change, we discuss the smooth transition model. We address the problem of parameter estimation for these models using a Bayesian approach. We derive expressions for the posterior densities for parameters which are used to make inference of the parameters and posterior model probabilities which are used to compare models. We also discuss the double smooth transition model which has two smooth transition components. Markov chain Monte Carlo methods are used to estimate these models, including parameter estimation and model selection. Models are fitted with their predictive means. We illustrate our approaches with empirical examples such as the British industrial production index, the US economic time series from Nelson and Plosser (1982) and the global average temperature series.

Finally we apply the reversible jump Markov chain Monte Carlo method to a structural break model which has unknown number of structural break points. The posterior model probabilities are obtained for models with different structural break points and posterior densities for parameters in the preferred model (with the biggest estimated posterior model probability) are also obtained. We fit the model to two US economic times series, the US real GNP series and the US consumer price index, with the number of structural break points selected by our algorithm automatically.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

There has been recent interest in studying and developing statistical models that account for a changing structure in time series in general and in economics time series in particular. Structural change has been a major concern in economics for a long time. Statistical tests and regression analysis are the principal tools in the economic analysis of these models. For example, Bacon and Watts (1971), Broemeling and Tsurum (1987), Wang and Zivot (2000) and Busetti and Harvey (2001) tested and fitted economic data with structural break models. Their models consist of a structural break component and a stationary term or a random walk.

Harrison and Stephens (1976) is one of the earliest to model changes in level and slope using Bayesian approach. Ferelli and Tunnicliffe Wilson (1990) considered the robust Bayesian estimation of level and trend, by making the mass probability large, the model sensibly represents series with a small number of changes in level and trend. The Bayesian approach has become increasingly popular in recent years. See, for example, Marriott and Newbold (1998), Marriott and Newbold (2000), and Wang and Zivot (2000). We present statistical inference on structural change from the Bayesian point of view within the framework of the linear structural break model (with abrupt structural breaks happening at unknown points) and the smooth transition model (with smooth changes of structure within a period of time.) We put emphasis on the analysis of posterior distributions and use them to conduct hypothesis tests as well as to obtain parameter estimates.

Many econometric testing problems involve nuisance parameters which are not identified under the null hypothesis. Problems of this type are often referred to as

Davies type problems (Davies, 1977; 1987) and are characterised by the fact that parameters that are needed to specify the model under the alternative hypothesis are not identified under the null hypothesis so conventional statistical theory cannot be used. The hypothesis testing approach to inference for the linear model with structural change has nuisance parameters because the structure of the model changes and the parameters under the null hypothesis may not appear under the alternative hypothesis and vice versa. While most practical work on testing problems involving nuisance parameters has utilised non-Bayesian procedures, Bayesian methods offer certain advantages that are useful in empirical research. The first part of this work concerns the Bayesian procedures for the hypothesis testing problem involving nuisance parameters in the linear model with a structural change at some unknown point.

Another problem in econometrics concerns modelling econometric growth and inference associated with this. Various theories on stages of economic development and growth assume that an economic relationship changes over time. A considerable proportion of time series econometrics research has been concerned with the debate as to whether economic series are best characterised as non-stationary processes which contain an autoregressive unit root or as stationary process possibly including a deterministic component. Recently the debate has been concerned with stationary processes with a change of structure in the deterministic component. For example, Crafts, Leybourne and Mills (1989), Terasvirta and Anderson (1992), Terasvirta (1994), Greenaway, Leybourne and Sapsford (1997) and Leybourne, Newbold and Vougas (1998). Econometricians think of these models as providing feasible alternative data descriptions to non-stationary models because their behaviour can appear superficially quite similar. The second part of this work concerns the inference for stationary processes with deterministic smooth transition components called smooth transition models.

The development of Markov chain Monte Carlo (MCMC) techniques has provided a new tool for Bayesian statistical analyses. Many papers and books on the use of MCMC have been published recently, for example, Metropolis *etc.* (1953), Hastings (1970), Tierney (1994), Gilks, Richardson and Spiegelhalter (1996), P. J. Green (1995) and Robert and Casella (1999). Many technical reports, which provide

the latest results of this field, can be found on the MCMC preprint at the web site: *http://www.statslab.cam.ac.uk/~mcmc/* of the MRC statistical laboratory, University of Cambridge. We will apply the MCMC techniques to our Bayesian analysis of the smooth transition models. We will also apply the reversible jump Markov chain Monte Carlo (RJMCMC) method to the linear structural break models with unknown number of structural break points.

The contents of this work consist of the following chapters.

In Chapter 2, we introduce the two-phase linear models described by Davies (1977, 1987) and follow the notations by Hinkley (1969). This is a model which changes its slope only. The change of slope is represented by a parameter γ, which is independent of the previous slope. By representing the change of slope as an independent parameter, we are inducing a relationship between the slopes before and after the break. We suggest different priors for the parameters in the model and obtained posterior distribution of the parameters in the model, especially the posterior density for τ, the structural break point of the two-phase model. We compare the two-phase model with the linear model using posterior model probabilities. Simulations were carried out to evaluate the performance of our algorithms.

In Chapter 3, we extend the two-phase model discussed in Chapter 2. We discuss the case the disturbance term in the two-phase model is an AR(1) process, instead of white noise. The two-phase model is compared with the linear model and the posterior for both parameters and models are obtained.

In Chapter 4, we introduce several kinds of smooth transition models. The general description of smooth transition model was given by Bacon and Watts (1971). From then on, a lot of papers introduced different specific smooth transition models and discussed the parameter estimation and hypothesis testing. We concentrate on the smooth transition models introduced by Leybourne, Newbold and Vougas (1998). In this chapter, we only discuss the simplest of the three smooth transition models, which describes the changes of level by the parameter α_2, γ and τ. We obtain the posterior densities for the parameters in the model and the posterior model probabilities for the smooth transition model. We compare the smooth transition model with a linear model, using the posterior model probabilities of them.

In Chapter 5, we consider more complex smooth transition model, representing

the changes of level and slope by parameter α_2, β_2, γ and τ, with an AR(1) disturbance term in stead of white noise. We again compare the smooth transition model with the linear regression model using the posterior model probabilities.

In Chapter 6, we introduce the slice sampler and apply the slice sampler to the more complex smooth transition model discussed in Chapter 5. We illustrate the performance of our algorithm, using the British industrial production index and the US GDP series.

In Chapter 7, we discuss the double smooth transition model introduced by Harvey and Mills (2000), which has two smooth transition components instead of one. Both smooth transition components represent the changes of level and slope together. The posterior densities for the parameters in the double smooth transition model are obtained. In order to sample directly from the posterior densities of these parameters, we use a hybrid Monte Carlo Markov chain sampler consisting of a Gibbs sampler for the parameters for which the full conditionals are standard distributions and the adaptive rejection Metropolis sampler (ARMS) for the parameters for which the full conditionals are complex and not easy to sample from directly. We then apply our approach to the global and hemisphere temperature data sets. Predictive distributions for the future observations are obtained.

Finally, in Chapter 8, we describe a new techniques which was introduced to statistics by P. J. Green (1995), the reversible jump Markov chain (RJMCMC). The RJMCMC is used to sample from the densities for which the dimensions change in the sampling process. Traditional methods cannot be applied in this case. We apply the RJMCMC to the structural break model which has unknown number of components. In the structural break model, the changes of level and slope are independent between the structural break points, which is much more appreciate from a time series point of view. The posterior probabilities of the different models (different number of the structure break components) and the posterior densities of parameters are obtained. Our approach is applied to two historical US economical time series.

Chapter 2

Two-phase Linear Model

We wish to use a Bayesian approach to test hypotheses in the presence of parameters which enter the model only under the alternative. These parameters are called nuisance parameters. To illustrate the approach, we will consider the two-phase linear regression model, which is widely used, and compare the hypotheses using posterior model probabilities. We present two cases with two different priors for the parameters of the model. Simulation will be carried out to investigate the approaches introduced in this chapter.

2.1 Two-phase Linear Regression Model

Davies (1987) used simulation to investigate his approach for two models, one of which is the two-phase model. Davies (1987) described the two-phase model as

> *Observe $X_1, \ldots, X_n$, a sequence of independent normal random variables with unit variance and expectations given by*
>
> $$E(X_i) = \begin{cases} a + bt_i, & (t_i < \theta), \\ a + bt_i + \xi(t_i - \theta), & (t_i \geq \theta), \end{cases}$$
>
> *where t_i denotes the time at which the ith measurement was made and θ the unknown time at which the change in slope occurred.*

Hinkley (1971) consider the cases where there are no discontinuities in the regression function. We follow Hinkley (1971) to write the two-phase model in regression form and adopt his notations.

Suppose $y_1, \ldots, y_n$ are independent normal random variables. Suppose further that under the null hypothesis there is a constant linear trend. Under the alternative, the linear trend changes at some unknown point, τ, but remains continuous. We will drop the strict condition in Davies, which assumes that the variance is a known constant, as this does not seem reasonable in practise. This two-phase linear regression model may be written as follows

$$y_i = \begin{cases} \alpha_0 + \alpha_1 t_i + \varepsilon_i, & t_i \leq \tau; \\ \alpha_0 + d_0 + (\alpha_1 + d_1) t_i + \varepsilon_i, & t_i > \tau, \end{cases}$$

where ε_i is *iid* $\mathrm{N}(0,\ \sigma^2)$, $i = 1, 2, \ldots, n$. t_i denotes the time at which the ith measurement was made and τ denotes the unknown time at which the change in slope occurred. We follow Davies (1987) to suppose the time of measurement to be centred so that $\sum t_i = 0$. α_0, α_1, d_0 and d_1 are unknown parameters. Because the linear trend function is continuous at $t = \tau$, the parameters should satisfy $d_0 + d_1\tau = 0$, and the model may be rewritten as

$$y_i = \alpha_0 + \alpha_1 t_i + \gamma(t_i - \tau)^+ + \varepsilon_i,$$

where $z^+ = \max(0, z)$, $\gamma = d_1$. In this continuous two-phase model, there is only a change of slope which is represented by the parameter γ. In the following, we assume that the prior for γ is independent of the previous slope α_1, which induces a relationship between the slopes before and after the break point τ.

Davies (1987) tests the hypothesis that $\gamma = 0$ against the alternative that $\gamma \neq 0$. We write the null hypothesis ($\gamma = 0$), the alternative ($\gamma \neq 0$) and their respective models separately.

We denote the model under the null hypothesis as M_0. The null hypothesis (model M_0) can be written as follows

$$M_0 : \qquad y_i = \alpha_0 + \alpha_1 t_i + \varepsilon_i, \tag{2.1}$$

where ε_i is *iid* $\mathrm{N}(0,\ \sigma^2)$, $i = 1, 2, \ldots, n$. $\sigma > 0$. $\sum t_i = 0$.

Under the alternative hypothesis the slope of the linear trend changes at some unknown point, τ. We denote the model under the alternative hypothesis as M_1. The alternative hypothesis (model M_1) can be written as follows

$$M_1 : \qquad y_i = \alpha_0 + \alpha_1 t_i + \gamma(t_i - \tau)^+ + \varepsilon_i, \tag{2.2}$$

where $\tau \in (t_1, t_n)$. If we denote $\tau_i = (t_i - \tau)^+$, we can rewrite the alternative hypothesis (model M_1) as follows

$$M_1 : \qquad y_i = \alpha_0 + \alpha_1 t_i + \gamma \tau_i + \varepsilon_i, \tag{2.3}$$

where $\tau \in (t_1, t_n)$ and τ_i is not observable or known because τ is unknown.

We notice that when $\tau = t_n$, the model M_1 becomes the same as model M_0. When $\tau = t_1$ the model M_1 becomes the same as model M_0 except at the left point $t = t_1$. Furthermore, when $\tau \to t_1$ or $\tau \to t_n$, model M_1 goes to model M_0, any approach attempting to distinguish between these two models becomes impossible unless we obtain more data, for example, at time $t_0 < t_1$ or $t_{n+1} > t_n$ which are outside the interval (t_1, t_n). In order to overcome this difficulty, we suppose $\tau \in [t_2, t_{n-1}]$. In this case, the sample data should come from one of the two models.

We now have a sample, $y_1, \ldots, y_n$, denoted as a vector $\mathbf{y} = (y_1, \ldots, y_n)'$ from one of the two models M_0 and M_1, we shall test which model the sample comes from by comparing the two models using a Bayesian approach.

In order to simplify the calculations, we use the following matrix notations for the model M_0

$$M_0 : \quad \mathbf{y} = X_0 \boldsymbol{\alpha} + \mathbf{u}, \tag{2.4}$$

where $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$, $\mathbf{u} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ and

$$X_0 = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}.$$

The model M_1 in matrix notations is then

$$M_1 : \quad \mathbf{y} = X_1 \boldsymbol{\beta}_1 + \mathbf{u}, \tag{2.5}$$

where $\mathbf{y}$ and $\mathbf{u}$ are the same as above, $\boldsymbol{\beta}_1 = (\alpha_0, \alpha_1, \gamma)'$ and

$$X_1 = \begin{pmatrix} 1 & t_1 & \tau_1 \\ 1 & t_2 & \tau_2 \\ \vdots & \vdots & \vdots \\ 1 & t_n & \tau_n \end{pmatrix},$$

where $\tau_i = (t_i - \tau)^+$, $\tau \in [t_2, t_{n-1}]$.

2.2 The Likelihood Functions

An essential element of the Bayesian approach is Bayes' theorem. In order to use a Bayesian approach to test hypotheses, we need to use Bayes' theorem to obtain the posterior probabilities for the models, M_0 and M_1. Bayes' theorem can be written as follows

$$\textit{posterior density} \propto \textit{prior density} \times \textit{likelihood function},$$

We can apply Bayes' theorem for inference about the parameters in the models M_0 and M_1 and for inference about the models M_0 and M_1 themselves. The posterior distributions for the models M_0 and M_1 can be used to test the hypothesis, that is, to test which model the sample data most probably comes from (according to the posterior model probabilities) and the posterior distributions for the parameters in both models M_0 and M_1 can be used to make inferences about the parameters.

In order to use Bayes' theorem, we need the likelihood functions with respect to the two models, M_0 and M_1. The likelihood function with respect to the model M_0 can be written as follows

$$p(\mathbf{y}|\theta_0, M_0) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 t_i)^2\right\}, \tag{2.6}$$

where $\sigma > 0$ and $\theta_0 = (\alpha_0, \alpha_1, \sigma)'$ is the parameter vector. Using matrix notation, we can rewrite the likelihood function above as

$$p(\mathbf{y}|\theta_0, M_0) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X_0\boldsymbol{\alpha})'(\mathbf{y} - X_0\boldsymbol{\alpha})\right\}. \tag{2.7}$$

Following Zellner(1987, p.66), we rewrite this as

$$p(\mathbf{y}|\theta_0, M_0) = \frac{\sigma^{-n}}{(2\pi)^{n/2}} \exp\left\{-\frac{\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})}{2\sigma^2}\right\}, \tag{2.8}$$

where $\nu_0 = n - 2$, $\hat{\boldsymbol{\alpha}} = (X_0'X_0)^{-1}X_0'\mathbf{y}$ and $s_0^2 = \dfrac{(\mathbf{y} - X_0\hat{\boldsymbol{\alpha}})'(\mathbf{y} - X_0\hat{\boldsymbol{\alpha}})}{\nu_0}$.

The likelihood function with respect to the model M_1 can be written as follows

$$p(\mathbf{y}|\theta_1, M_1) = \frac{\sigma^{-n}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - \alpha_0 - \alpha_1 t_i - \gamma(t_i - \tau)^+\right)^2\right\}, \tag{2.9}$$

where θ_1 is the parameter vector $\theta_1 = (\alpha_0, \alpha_1, \gamma, \tau, \sigma)'$. In order to use matrix notation to represent this likelihood function, we shall use τ_i instead of $(t_i - \tau)^+$,

the likelihood function (2.9) can then be written as follows

$$p(\mathbf{y}|\theta_1, M_1) = \frac{\sigma^{-n}}{(2\pi)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \alpha_0 - \alpha_1 t_i - \gamma\tau_i)^2\right\}, \tag{2.10}$$

which can be written in matrix notation as follows

$$p(\mathbf{y}|\theta_1, M_1) = \frac{1}{(2\pi)^{n/2}\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X_1\beta_1)'(\mathbf{y} - X_1\beta_1)\right\}. \tag{2.11}$$

As before, we can rewrite this as follows

$$p(\mathbf{y}|\theta_1, M_1) = \frac{\sigma^{-n}}{(2\pi)^{n/2}} \exp\left\{-\frac{\nu_1 s_1^2 + (\beta_1 - \hat{\beta}_1)'X_1'X_1(\beta_1 - \hat{\beta}_1)}{2\sigma^2}\right\}, \tag{2.12}$$

where $\nu_1 = n - 3$, $\hat{\beta}_1 = (X_1'X_1)^{-1}X_1'\mathbf{y}$ and $s_1^2 = \dfrac{(\mathbf{y} - X_1\hat{\beta}_1)'(\mathbf{y} - X_1\hat{\beta}_1)}{\nu_1}$.

2.3 Bayesian Model Selection

Here we follow Marriott and Newbold (1998), Marriott and Newbold (2000) and Wang and Zivot (2000) and use posterior model probabilities to compare hypothesis or select models. The conditional probability of model M_i, given the sample $\mathbf{y}$, (the posterior probability of M_i, See Bernardo and Smith, 1994, pp.387) is

$$P(M_i|\mathbf{y}) = \frac{P(M_i)p(\mathbf{y}|M_i)}{P(M_0)p(\mathbf{y}|M_0) + P(M_1)p(\mathbf{y}|M_1)}, \tag{2.13}$$

where $i = 0, 1$, $P(M_i)$ is the prior probability assumed for the model M_i and $p(\mathbf{y}|M_i)$ is the integrated likelihood function with respect to model M_i. The integrated likelihood is obtained from

$$p(\mathbf{y}|M_i) = \int p(\mathbf{y}|\theta_i, M_i)p(\theta_i|M_i)\, d\theta_i, \tag{2.14}$$

where θ_i is the parameter vector in model M_i, $p(\mathbf{y}|\theta_i, M_i)$ is the likelihood function with respect to model M_i and $p(\theta_i|M_i)$ is the prior for the parameter vector θ_i in model M_i. Integrals of this type will occur frequently later, so for convenience, we denote

$$L_i = \int p(\mathbf{y}|\theta_i, M_i)p(\theta_i|M_i)\, d\theta_i. \tag{2.15}$$

The simple approach (used in Marriott and Newbold (1998), Marriott and Newbold (2000) and Wang and Zivot (2000)) to testing H_0 against H_1 decides in favour of

the hypothesis with the largest posterior model probability, that is, we would accept model M_i if $P(M_i|\mathbf{y}) > P(M_j|\mathbf{y})$, $\forall j \neq i$.

For the approach we are adopting here we take uniform priors for the models, so that $P(M_0) = P(M_1) = 0.5$. This is equivalent to using a 50% prior weight on $\gamma = 0$ and a 50% prior weight on $\gamma \neq 0$. Given these prior probabilities for the models, the posterior model probabilities are

$$P(M_i|\mathbf{y}) = \frac{L_i}{L_1 + L_2}, \tag{2.16}$$

where $i = 1, 2$.

2.4 Conjugate Prior

We adopt conjugate priors for parameters α_0, α_1 and σ in model M_0. In model M_1, there are two more parameters γ and τ. We will use priors for parameter α_0, α_1, σ and γ that are conjugate priors conditional on τ being known, and since the parameter $\tau \in [t_2, t_{n-1}]$, we use a uniform distribution as a prior for τ.

2.4.1 Posteriors for Model M_0

First we derive a conjugate prior for the parameter vector θ_0 in model M_0. The likelihood for model M_0 is given in (2.8), the conjugate prior for $\boldsymbol{\alpha}$ given σ in this model is

$$p(\boldsymbol{\alpha}|\sigma, M_0) = \frac{|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}})' V_{\boldsymbol{\alpha}}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}})\right\}, \tag{2.17}$$

where $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ is a two-dimensional constant vector and $V_{\boldsymbol{\alpha}}$ is an 2×2 positive definite symmetric constant matrix. We adopt the usual conjugate prior for σ in model M_0, which is the inverted gamma distribution with density function

$$p(\sigma|M_0) = \frac{2}{\Gamma(b) a^b \sigma^{2b+1}} \exp\left\{-\frac{1}{a\sigma^2}\right\}, \quad 0 < \sigma < +\infty, \tag{2.18}$$

where $a, b > 0$ are constants.

The joint conjugate prior distribution for θ_0 in model M_0 is then

$$p(\theta_0|M_0) = \frac{|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{\pi\Gamma(b) a^b \sigma^{2b+3}} \exp\left\{\frac{2/a + (\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}})' V_{\boldsymbol{\alpha}}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}})}{2\sigma^2}\right\}. \tag{2.19}$$

Combining (2.19) with the expression on the right side of (2.8), we obtain the joint density for $\mathbf{y}$ and θ_0 as

$$p(\mathbf{y}, \theta_0|M_0) = \frac{2|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}+1}\Gamma(b)a^b\sigma^{2b+n+3}} \times \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_0\tilde{s}_0^2 + (\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})'\tilde{V}_0(\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})\right]\right\}, \tag{2.20}$$

where $\tilde{V}_0 = X_0'X_0 + V_{\boldsymbol{\alpha}}^{-1}$, $\tilde{\boldsymbol{\alpha}} = \tilde{V}_0^{-1}(X_0'X_0\hat{\boldsymbol{\alpha}} + V_{\boldsymbol{\alpha}}^{-1}\mu_{\boldsymbol{\alpha}})$ and $\nu_0\tilde{s}_0^2 = \nu_0 s_0^2 + 2/a + (\tilde{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}})'X_0'X_0(\mu_{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}})$.

Integrating this with respect to σ gives

$$p(\mathbf{y}, \boldsymbol{\alpha}|M_0) = \frac{2^{b+1+\frac{n}{2}}\Gamma(b+1+\frac{n}{2})|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{n}{2}+1}\Gamma(b)a^b} \times \left[\nu_0\tilde{s}_0^2 + (\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})\tilde{V}_0(\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})\right]^{-(b+1+\frac{n}{2})}. \tag{2.21}$$

We then integrate (2.21) with respect to $\boldsymbol{\alpha}$ and obtain

$$p(\mathbf{y}|M_0) = \frac{2^b\Gamma(b+\frac{n}{2})|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{\pi^{\frac{n}{2}}\Gamma(b)a^b|\tilde{V}_0|^{\frac{1}{2}}(\nu_0\tilde{s}_0^2)^{b+\frac{n}{2}}}. \tag{2.22}$$

If we apply Bayes' theorem to model M_0 to obtain the posterior distribution for the parameter vector θ_0, we obtain

$$p(\theta_0|\mathbf{y}, M_0) \propto \frac{1}{\sigma^{2b+n+3}} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_0\tilde{s}_0^2 + (\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})'\tilde{V}_0(\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})\right]\right\}. \tag{2.23}$$

This posterior distribution can be used to make inferences concerning the parameter vector $\theta_0 = (\alpha_0, \alpha_1, \sigma)$.

If we integrate (2.23) with respect to $\boldsymbol{\alpha}$, we can obtain the marginal posterior distribution for σ conditional on model M_0 as

$$p(\sigma|\mathbf{y}, M_0) \propto \frac{1}{\sigma^{2b+n+1}} \exp\left\{-\frac{\nu_0\tilde{s}_0^2}{2\sigma^2}\right\}, 0 < \sigma < \infty, \tag{2.24}$$

which is an inverted gamma distribution.

We can also integrate (2.23) with respect to σ, we obtain the marginal posterior distribution for the parameter $\boldsymbol{\alpha}$ conditional on M_0 as

$$p(\boldsymbol{\alpha}|\mathbf{y}, M_0) \propto \left[\nu_0\tilde{s}_0^2 + (\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})'X_0'X_0(\boldsymbol{\alpha}-\tilde{\boldsymbol{\alpha}})\right]^{-\frac{2b+n+2}{2}}, \tag{2.25}$$

which is a multivariate student t distribution.

2.4.2 Posteriors for Model M_1

In order to simplify the choice of the parameters of the prior distributions in the two models, we use the same prior distributions for the parameters α_0, α_1 and σ in model M_1 as were used in model M_0. We would also like to obtain conjugate priors for all of the parameters (including the parameters α_0, α_1 and σ) in model M_1. Given the complexity of the structure of the likelihood function with respect to the parameter τ, (see equations (2.9) to (2.12),) no simple expression for a conjugate prior for τ exists. For this reason, we use a uniform distribution as a prior for τ and adopt conditional conjugate priors for parameters α_0, α_1, γ and σ, given τ.

The form of the likelihood suggests that, conditional on σ and τ, the conjugate prior for $\boldsymbol{\beta}_1$ should be the multivariate normal with density

$$p(\boldsymbol{\beta}_1|\sigma, M_1) = \frac{|V_1|^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}\sigma^3} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^0)'V_1(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^0)\right\}, \tag{2.26}$$

where V_1 is a constant 3×3 positive definite symmetric matrix, $\boldsymbol{\beta}_1^0$ a constant vector. Conditional on τ, the conjugate prior for σ should be the inverted gamma with density

$$p(\sigma|M_1) = \frac{2}{\Gamma(b)a^b\sigma^{2b+1}} \exp\left\{-\frac{1}{a\sigma^2}\right\}, \quad 0 < \sigma < +\infty, \tag{2.27}$$

where $a, b > 0$ are constants. We choose the same values for a and b as in model M_0.

If we suppose that $\boldsymbol{\beta}_1$ and σ are independent of τ, then the joint prior for θ_1 is then taken to be

$$p(\theta_1|M_1) = p(\boldsymbol{\beta}_1|\sigma, M_1)p(\sigma|M_1)p(\tau|M_1).$$

As stated earlier, we choose the same conjugate prior distribution for $\boldsymbol{\alpha}$ as in model M_0 by selecting the precision matrix V_1 for parameter vector $\boldsymbol{\beta}_1 = (\boldsymbol{\alpha}, \gamma)$ as

$$V_1 = \begin{pmatrix} V_{\boldsymbol{\alpha}}^{-1} & V_{12} \\ V_{21} & v_{22} \end{pmatrix},$$

where $V_{\boldsymbol{\alpha}}$ is a 2×2 positive definite symmetric matrix defined in model M_0, $V_{12}' = V_{21}$ are 3×1 and 1×3 matrixes respectively, and v_{22} is a positive constant. The marginal conjugate prior for $\boldsymbol{\alpha}$ conditional on σ is then the multivariate normal distribution

with precision matrix $V_{\boldsymbol{\alpha}}$ and density as

$$p(\boldsymbol{\alpha}|\sigma, M_1) = \frac{|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}})' V_{\boldsymbol{\alpha}}^{-1}(\boldsymbol{\alpha} - \boldsymbol{\mu}_{\boldsymbol{\alpha}})\right\}, \tag{2.28}$$

where $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ is the same two-dimensional constant vector used in model M_0. The marginal conjugate prior we adopt for γ, given σ, is the (non-zero) normal distribution $\mathrm{N}(\mu_\gamma, (\sigma K_\gamma)^2)$, where $\gamma \neq 0$. As we do not know whether $\gamma > 0$ or $\gamma < 0$, we place the same weight on $\gamma > 0$ and $\gamma < 0$, this leads to $\mu_\gamma = 0$, $\boldsymbol{\beta}_1^0 = (\boldsymbol{\mu}_{\boldsymbol{\alpha}}', 0)'$. We then have the density function of the marginal conjugate prior for γ, given σ in model M_1 as

$$p(\gamma|\sigma, M_1) = \frac{1}{\sqrt{2\pi}K_\gamma\sigma} \exp\left\{-\frac{\gamma^2}{2K_\gamma^2\sigma^2}\right\}, \tag{2.29}$$

where $\gamma \neq 0$, $K_\gamma = 1/\sqrt{v_{22}}$ is a positive constant.

We know nothing about the correlation relationship between $\boldsymbol{\alpha}$ and γ in model M_1. This relationship is represented by the correlation coefficient matrix V_{12}. For simplicity and in the absence of any other prior information, we suppose $V_{12} = \mathbf{0}$, we then have

$$p(\boldsymbol{\beta}_1|\sigma, M_1) = \frac{|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}K_\gamma\sigma^3} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^0)' V_1 (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^0)\right\}, \tag{2.30}$$

where

$$V_1 = \begin{pmatrix} V_{\boldsymbol{\alpha}}^{-1} & 0 \\ 0 & K_\gamma^{-2} \end{pmatrix},$$

Clearly we can also choose $V_{12} \neq \mathbf{0}$, if we have some prior information about the correlation between $\boldsymbol{\alpha}$ and γ.

We know $\tau \in [t_2, t_{n-1}]$, so the uniform prior for τ is

$$p(\tau|M_1) = \begin{cases} \dfrac{1}{t_{n-1} - t_2}, & \text{when } \tau \in [t_2, t \\ & \text{when } \tau \in [t_2, t_{n-1}]; \\ 0, & \text{otherwise.} \end{cases} \tag{2.31}$$

The joint prior density for θ_1 can therefore be written as

$$\begin{aligned} p(\theta_1|M_1) = {} & \frac{2|V_{\boldsymbol{\alpha}}^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{3}{2}}(t_{n-1} - t_2)K_\gamma\Gamma(b)a^b\sigma^{2b+4}} \times \\ & \exp\left\{-\frac{2/a + (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^0)' V_1 (\boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^0)}{2\sigma^2}\right\}, \end{aligned} \tag{2.32}$$

where $\sigma > 0$, $\tau \in [t_2, t_{n-1}]$.

The joint density for $\mathbf{y}$ and θ_1 is then

$$p(\mathbf{y}, \theta_1|M_1) = p(\theta_1|M_1)p(\mathbf{y}|\theta_1, M_1),$$

and using (2.12) with (2.32), we obtain

$$p(\mathbf{y}, \theta_1|M_1) = \frac{2|V_\alpha^{-1}|^{\frac{1}{2}}}{(2\pi)^{\frac{n+3}{2}}(t_{n-1}-t_2)K_\gamma\Gamma(b)a^b\sigma^{2b+n+4}} \times \exp\left\{-\frac{\nu_1\tilde{s}_1^2 + (\beta_1-\tilde{\beta}_1)'\tilde{V}_1(\beta_1-\tilde{\beta}_1)}{2\sigma^2}\right\}, \tag{2.33}$$

where $\tilde{V}_1 = X_1'X_1 + V_1$, $\tilde{\beta}_1 = \tilde{V}_1^{-1}(X_1'X_1\hat{\beta}_1 + V_1\beta_1^0)$, $\nu_1\tilde{s}_1^2 = \nu_1 s_1^2 + 2/a + (\tilde{\beta}_1 - \hat{\beta}_1)'\tilde{V}_1(\beta_1^0 - \hat{\beta}_1)$ and $\tau \in [t_2, t_{n-1}]$.

Integrating (2.33) with respect to σ gives

$$p(\mathbf{y}, \beta_1, \tau|M_1) = \frac{2^b|V_\alpha^{-1}|^{\frac{1}{2}}\Gamma(\frac{2b+n+3}{2})}{\pi^{\frac{n+3}{2}}(t_{n-1}-t_2)K_\gamma\Gamma(b)a^b} \times \left[\nu_1\tilde{s}_1^2 + (\beta_1-\tilde{\beta}_1)'\tilde{V}_1(\beta_1-\tilde{\beta}_1)\right]^{-\frac{2b+n+3}{2}}, \tag{2.34}$$

where $\tau \in [t_2, t_{n-1}]$.

Integrating this with respect to β_1 gives

$$p(\mathbf{y}, \tau|M_1) = \frac{2^b|V_\alpha^{-1}|^{\frac{1}{2}}\Gamma(\frac{2b+n}{2})}{\pi^{\frac{n}{2}}(t_{n-1}-t_2)K_\gamma\Gamma(b)a^b|\tilde{V}_1|^{\frac{1}{2}}(\nu_1\tilde{s}_1^2)^{\frac{2b+n}{2}}}, \tag{2.35}$$

where $\tau \in [t_2, t_{n-1}]$.

If we integrate (2.35) with respect to τ, we obtain

$$p(\mathbf{y}|M_1) = \frac{2^b|V_\alpha^{-1}|^{\frac{1}{2}}\Gamma(\frac{2b+n}{2})}{\pi^{\frac{n}{2}}(t_{n-1}-t_2)K_\gamma\Gamma(b)a^b}\int_{t_2}^{t_{n-1}} |\tilde{V}_1|^{-\frac{1}{2}}(\nu_1\tilde{s}_1^2)^{-\frac{2b+n}{2}}\,d\tau. \tag{2.36}$$

Applying Bayes' theorem to model M_1, we obtain the joint posterior density for the parameter vector θ_1

$$p(\theta_1|\mathbf{y}, M_1) \propto \frac{1}{\sigma^{2b+n+4}}\exp\left\{-\frac{\nu_1\tilde{s}_1^2 + (\beta_1-\tilde{\beta}_1)'\tilde{V}_1(\beta_1-\tilde{\beta}_1)}{2\sigma^2}\right\}. \tag{2.37}$$

If we integrate (2.37) with respect to β_1, we obtain the joint posterior density for (τ, σ)

$$p(\tau, \sigma|\mathbf{y}, M_1) \propto \frac{1}{|\tilde{V}_1|^{\frac{1}{2}}\sigma^{2b+n+1}}\exp\left\{-\frac{\nu_1\tilde{s}_1^2}{2\sigma^2}\right\}. \tag{2.38}$$

If we know τ, the posterior distribution for σ would be the inverted gamma distribution with the following density

$$p(\sigma|\tau, \mathbf{y}, M_1) \propto \frac{1}{\sigma^{2b+n+1}}\exp\left\{-\frac{\nu_1\tilde{s}_1^2}{2\sigma^2}\right\}. \tag{2.39}$$

Integrating (2.38) with respect to σ, we obtain the posterior density for τ as follows

$$p(\tau|\mathbf{y}, M_1) \propto \frac{1}{|\tilde{V}_1|^{\frac{1}{2}} [\nu_1 \tilde{s}_1^2]^{\frac{2b+n}{2}}}. \tag{2.40}$$

If we integrate (2.37) with respect to σ, we obtain the joint posterior density for (β_1, τ)

$$p(\beta_1, \tau|\mathbf{y}, M_1) \propto \left[\nu_1 \tilde{s}_1^2 + (\beta_1 - \tilde{\beta}_1)' \tilde{V}_1 (\beta_1 - \tilde{\beta}_1)\right]^{-\frac{2b+n+2}{2}}, \tag{2.41}$$

which also means that, conditional on τ, β_1 has a multivariate student t distribution.

2.4.3 Posterior Model Probabilities

Substituting (2.22) and (2.36) in (2.16), we obtain $p(X|M_0)$ and $p(X|M_1)$ as

$$P(M_0|\mathbf{y}) = \frac{1}{1+K}, \tag{2.42}$$

and

$$P(M_1|\mathbf{y}) = \frac{K}{1+K}, \tag{2.43}$$

where

$$K = \frac{|\tilde{V}_0|^{\frac{1}{2}} (\nu_0 \tilde{s}_0^2)^{\frac{2b+n}{2}}}{(t_{n-1} - t_2) K_\gamma} \int_{t_2}^{t_{n-1}} \frac{1}{|\tilde{V}_1|^{\frac{1}{2}} (\nu_1 \tilde{s}_1^2)^{\frac{2b+n}{2}}} \, d\tau. \tag{2.44}$$

The above expression can now be used to discriminate between the two models.

2.5 Noninformative Prior

In this section we consider using noninformative priors for the parameters in model M_0 and model M_1. For parameters appear in both models, we choose noninformative priors. However, any improper priors (including noninformative priors) are only defined up to a constant, see Jeffreys (1961). These constants will cancel in the expression for the posterior model probabilities in (2.16) because these constants appears in both the numerator and denominator in (2.16). In model M_1, there are two additional parameters γ and τ. We need to use proper prior distributions for γ and τ because they are only present in model M_1. See Marriott and Newbold (1998), Marriott and Newbold (2000) and Wasserman (2000). For the parameter γ, we adopt a conjugate prior given the other parameters. We know the parameter $\tau \in [t_2, t_{n-1}]$, so we use a uniform distribution as a prior for the parameter τ.

2.5.1 Posteriors for Model M_0

The noninformative prior for $\boldsymbol{\alpha}$ in model M_0 is

$$p(\boldsymbol{\alpha}|M_0) = K_{\boldsymbol{\alpha}}, \tag{2.45}$$

where $K_{\boldsymbol{\alpha}}$ is a constant. We suppose the prior distribution for σ in model M_0 is $\ln\sigma \sim K_\sigma$, which leads to the following results for σ

$$p(\sigma|M_0) = \frac{K_\sigma}{\sigma}, \tag{2.46}$$

where K_σ is a constant.

If we suppose $\boldsymbol{\alpha}$ and σ are independent of each other, we then have the joint prior distribution for θ_0 ($\boldsymbol{\alpha}$ and σ) as

$$p(\theta_0|M_0) = p(\boldsymbol{\alpha}|M_0)p(\sigma|M_0),$$

that is

$$p(\theta_0|M_0) = \frac{K_{\boldsymbol{\alpha}}K_\sigma}{\sigma}. \tag{2.47}$$

The joint density of $(\mathbf{y}, \theta_0)$ is then given by

$$p(\mathbf{y}, \theta_0|M_0) = \frac{K_{\boldsymbol{\alpha}}K_\sigma}{(2\pi)^{\frac{n}{2}}\sigma^{n+1}} \exp\left\{-\frac{\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)}{2\sigma^2}\right\}. \tag{2.48}$$

Integrating this with respect to σ gives

$$p(\mathbf{y}, \boldsymbol{\alpha}|M_0) = \frac{\Gamma(\frac{n}{2})K_{\boldsymbol{\alpha}}K_\sigma}{2\pi^{\frac{n}{2}}} \left[\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)\right]^{-\frac{n}{2}}. \tag{2.49}$$

We then integrate (2.49) with respect with $\boldsymbol{\alpha}$ to obtain

$$p(\mathbf{y}|M_0) = \frac{\Gamma(\frac{\nu_0}{2})K_{\boldsymbol{\alpha}}K_\sigma}{2(\pi\nu_0 s_0^2)^{\frac{\nu_0}{2}} |X_0' X_0|^{\frac{1}{2}}}. \tag{2.50}$$

Applying Bayes' theorem, we obtain the joint posterior density for the parameter vector θ_0 as

$$p(\theta_0|\mathbf{y}, M_0) \propto \frac{1}{\sigma^{n+1}} \exp\left\{-\frac{\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)}{2\sigma^2}\right\}. \tag{2.51}$$

If we integrate (2.51) with respect to $\boldsymbol{\alpha}$, we obtain the posterior distribution for σ with density

$$p(\sigma|\mathbf{y}, M_0) \propto \frac{1}{\sigma^{n-1}} \exp\left\{-\frac{\nu_0 s_0^2}{2\sigma^2}\right\}, \tag{2.52}$$

which is an inverted gamma distribution.

If we integrate (2.51) with respect to σ, we obtain the posterior distribution for $\boldsymbol{\alpha}$ with density

$$p(\boldsymbol{\alpha}|\mathbf{y}, M_0) \propto \left[\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}}_0)\right]^{-\frac{n}{2}}, \tag{2.53}$$

which is a multivariate student t distribution.

2.5.2 Posteriors for Model M_1

We adopt the same prior distributions for parameters α_0, α_1 and σ in model M_1 as in model M_0. That is,

$$p(\boldsymbol{\alpha}, \sigma|M_1) = \frac{K_{\boldsymbol{\alpha}} K_\sigma}{\sigma}, \tag{2.54}$$

where $K_{\boldsymbol{\alpha}}$, K_σ are the same as in model M_0.

In order to calculate the posterior model probabilities $P(M_i|\mathbf{y})$, $i = 0, 1$, we shall adopt proper priors for the parameters γ and τ. We use the natural conjugate prior for γ given σ in model M_1. For the same reason as in the section 2.4.2, we select the prior for γ given in (2.29).

We know $\tau \in [t_2, t_{n-1}]$, so we adopt the uniform prior distribution given in (2.31).

We suppose that $\boldsymbol{\alpha}$, (γ, σ) and τ are *a priori* independent of each other. Then we have the joint prior distribution for all parameters in model M_1 as

$$p(\theta_1|M_1) = p(\boldsymbol{\alpha}|M_1) p(\gamma|\sigma, M_1) p(\tau|M_1) p(\sigma|M_1),$$

that is

$$p(\theta_1|M_1) = \frac{K_{\boldsymbol{\alpha}} K_\sigma}{\sqrt{2\pi}(t_{n-1} - t_2) K_\gamma \sigma^2} \exp\left\{-\frac{\gamma^2}{2K_\gamma^2 \sigma^2}\right\}, \tag{2.55}$$

where $\tau \in [t_2, t_{n-1}]$, $\gamma > 0$.

Because

$$p(\mathbf{y}, \theta_1|M_1) = p(\theta_1|M_1) p(\mathbf{y}|\theta_1, M_1),$$

straightforward calculation gives the joint density for $\mathbf{y}$ and θ_1 as

$$\begin{aligned} p(\mathbf{y}, \theta_1|M_1) = {} & \frac{K_{\boldsymbol{\alpha}} K_\sigma}{(2\pi)^{\frac{n+1}{2}} (t_{n-1} - t_2) K_\gamma \sigma^{n+2}} \times \\ & \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_1 \tilde{s}_1^2 + (\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)' V_1 (\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)\right]\right\}, \end{aligned} \tag{2.56}$$

where $V_1 = X_1'X_1 + U$, $\tilde{\beta}_1 = V_1^{-1}X_1'X_1\hat{\beta}_1$, $\nu_1\tilde{s}_1^2 = \nu_1 s_1^2 + (\tilde{\beta}_1 - \hat{\beta}_1)'X_1'X_1(\tilde{\beta}_1 - \hat{\beta}_1) + \tilde{\beta}_1'U\tilde{\beta}_1$ and U is a matrix defined by

$$U = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & K_\gamma^{-2} \end{pmatrix}.$$

Integrating (2.56) above with respect to σ gives

$$p(\mathbf{y}, \boldsymbol{\beta}_1, \tau|M_1) = \frac{K_\alpha K_\sigma \Gamma(\frac{n+1}{2})}{2\pi^{\frac{n+1}{2}}(t_{n-1} - t_2)K_\gamma} \times \left[\nu_1\tilde{s}_1^2 + (\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)'V_1(\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)\right]^{-\frac{n+1}{2}}. \tag{2.57}$$

where $\tau \in [t_2, t_{n-1}]$.

Integrating this with respect to $\boldsymbol{\beta}_1$ gives

$$p(\mathbf{y}, \tau|M_1) = \frac{K_\alpha K_\sigma \Gamma(\frac{\nu_0}{2})}{2\pi^{\frac{\nu_0}{2}}(t_{n-1} - t_2)K_\gamma(\nu_1\tilde{s}_1^2)^{\frac{\nu_0}{2}}|V_1|^{\frac{1}{2}}}. \tag{2.58}$$

where $\tau \in [t_2, t_{n-1}]$.

If we now integrate (2.58) with respect to τ, we have

$$p(\mathbf{y}|M_1) = \frac{K_\alpha K_\sigma \Gamma(\frac{\nu_0}{2})}{2\pi^{\frac{\nu_0}{2}}(t_{n-1} - t_2)K_\gamma} \int_{t_2}^{t_{n-1}} (\nu_1\tilde{s}_1^2)^{-\frac{\nu_0}{2}}|V_1|^{-\frac{1}{2}}\, d\tau. \tag{2.59}$$

Applying Bayes' theorem, we obtain the joint posterior density for the parameter vector θ_1 as

$$p(\theta_1|\mathbf{y}, M_1) \propto \frac{1}{\sigma^{n+2}} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_1\tilde{s}_1^2 + (\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)'V_1(\boldsymbol{\beta}_1 - \tilde{\boldsymbol{\beta}}_1)\right]\right\}, \tag{2.60}$$

If we integrate (2.60) with respect to $\boldsymbol{\beta}_1$, we obtain the joint posterior density for (τ, σ)

$$p(\tau, \sigma|\mathbf{y}, M_1) \propto \frac{1}{|V_1|^{\frac{1}{2}}\sigma^{n-1}} \exp\left\{-\frac{\nu_1\tilde{s}_1^2}{2\sigma^2}\right\}. \tag{2.61}$$

If we know τ, then σ will have an inverted gamma density

$$p(\sigma|\tau, \mathbf{y}, M_1) \propto \frac{1}{\sigma^{n-1}} \exp\left\{-\frac{\nu_1\tilde{s}_1^2}{2\sigma^2}\right\}. \tag{2.62}$$

If we integrate the joint posterior for (τ, σ), (2.61), with respect to σ, we obtain the posterior density for τ as

$$p(\tau|\mathbf{y}, M_1) \propto \frac{1}{|V_1|^{\frac{1}{2}}(\nu_1\tilde{s}_1^2)^{\frac{n-2}{2}}}. \tag{2.63}$$

Integrating (2.60) with respect to σ, we obtain the joint posterior density for (β_1, τ) as follows

$$p(\beta_1, \tau|\mathbf{y}, M_1) \propto \left[\nu_1 \tilde{s}_1^2 + (\beta_1 - \tilde{\beta}_1)' V_1 (\beta_1 - \tilde{\beta}_1)\right]^{-\frac{n+1}{2}}. \tag{2.64}$$

If we know τ, β_1 will have a multivariate student t distribution with density

$$p(\beta_1|\tau, \mathbf{y}, M_1) \propto \left[\nu_1 \tilde{s}_1^2 + (\beta_1 - \tilde{\beta}_1)' V_1 (\beta_1 - \tilde{\beta}_1)\right]^{-\frac{n+1}{2}}. \tag{2.65}$$

2.5.3 Posterior Model Probabilities

Using (2.50), (2.59) together with (2.16), we obtain $p(M_0|\mathbf{y})$ and $p(M_1|\mathbf{y})$ as in (2.42) and (2.43) where

$$K = \frac{(\nu_0 s_0^2)^{\frac{\nu_0}{2}} |X_0' X_0|^{\frac{1}{2}}}{(t_{n-1} - t_2) K_\gamma} \int_{t_2}^{t_{n-1}} \frac{1}{(\nu_1 \tilde{s}_1^2)^{\frac{\nu_0}{2}} |V_1|^{\frac{1}{2}}} d\tau. \tag{2.66}$$

These posterior model probabilities will be used in our Bayesian test of models.

2.6 Simulations

Now we are going to illustrate the approach with the noninformative prior we have discussed above using simulated data. First we will illustrate the posterior inference for the "break" parameters γ and τ and then we will illustrate the model choice.

For each simulation below (unless otherwise stated), we generate sample data from the model

$$y_i = 0.5 t_i + \gamma (t_i - \tau)^+ + \varepsilon_i, \tag{2.67}$$

so that $\alpha_0 = 0$, $\alpha_1 = 0.5$, without loss of generality. We choose $\sigma = 1$ and use different values for γ and τ. The measuring time begins at $t_1 = -9.5$ and ends at $t_n = 9.5$, the intervals between t_{i-1} and t_i are equal and n=100. See Davies (1987), example 1 in the section "Simulations".

Simulations of the Posterior Densities I

We have an unknown parameter K_γ in the prior for γ. The specific choices of prior densities for σ and γ should be considered together because the prior for γ depends

on σ. This choice is discussed in more detail in the second part of this simulation and in later chapters. First, we choose $K_\gamma = 1$.

We generate 100 observations from each of 6 versions of (2.67) where (γ, τ) takes the values shown in Table 2.1. Notice that when $\gamma = 0$, τ is nonsense for the model so we put a $*$ in Table 2.1 for this case.

γ	0	5	1	1	1	5
τ	*	-8.55	-4.75	0	4.75	8.55

Table 2.1: Pair of values for (γ, τ) used in the simulation. *: Note τ is meaningless when $\gamma = 0$.

The posterior density of τ is then plotted for each pair (γ, τ) and presented in Figure 2.1 (a) through to (f). The dotted vertical lines present the true values of τ.

For Figure 2.1 (a), we generated sample data with (γ, τ) from the first column of Table 2.1. Because $\gamma = 0$, the sample data are from model M_0 where there is no change in slope, in this case τ can clearly be any value represented by $*$ in Table 2.1. The graph in Figure 2.1 (a) shows that the value of τ can be anywhere in the interval but is most likely to be at an extreme.

For Figure 2.1 (b), $\gamma = 5$, $\tau = -8.55$ from the second column of Table 2.1. The change of slope, γ, is large, but the change happens at the place which is very near the left edge of the interval of the measuring time. We can see in Figure 2.1 (b) that the posterior density of τ concentrates near the true value of τ.

For Figure 2.1 (c), we set $\gamma = 1$, $\tau = -4.75$ according to the third column of Table 2.1. This time τ is in the middle part of the interval but the change in the slope, γ, is relatively small. The posterior density of τ again concentrates close to the true vale of τ.

For Figure 2.1 (d), $\gamma = 1$, $\tau = 0$ from the fourth column of Table 2.1. Here τ is at the centre point of the interval.

For Figure 2.1 (e), $\gamma = 1$, $\tau = 4.75$ from the fifth column of Table 2.1. Here τ is of the symmetric point to $\tau = -4.75$ in the third column of Table Table 2.1. $\gamma = 1$, the same value for γ as in the third column of Table 2.1. The posterior density concentrates around the true value of τ with similar bias compared with Figure 2.1

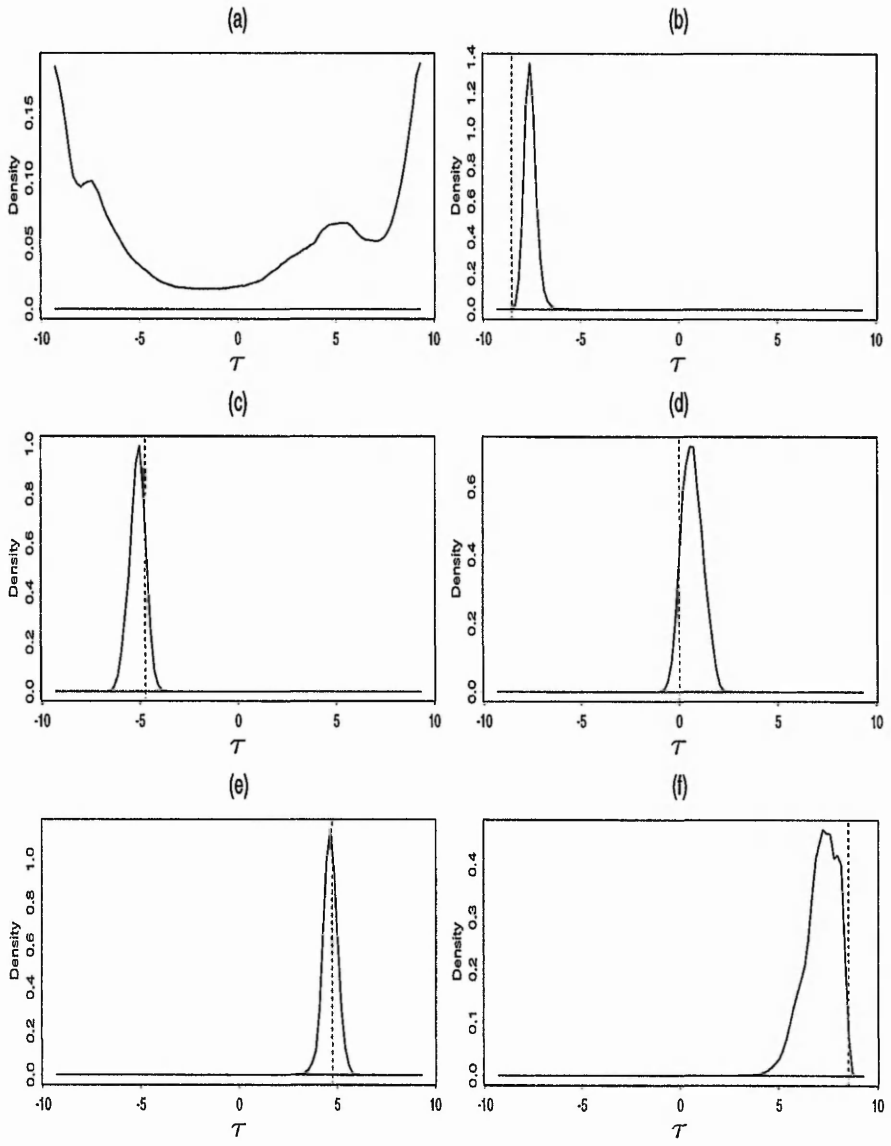


Figure 2.1: Posterior densities of τ in two-phase model. $K_\gamma = 1$.

(c).

In the above three simulations, for Figure 2.1 (c), (d) and (e), we generated data using the same value of γ, different values of τ. We would expect that when τ is in the middle of the interval of the measuring time, we should obtain the best results among these three simulations. However we could not find this phenomenon in these three simulations. The problem is, we just have one simulation for each case. If we would like to find subtle difference between them, we need to repeat each simulation, for example, 1000 times, and then compare these repeated simulations.

For Figure 2.1 (f), $\gamma = 5$, $\tau = 8.55$ from the last column of Table 2.1. Similar situations for γ and τ compared with the second column of Table Table 2.1. $\gamma = 5$, $\tau = 8.55$ gives a much large slope change near the right edge of the interval.

Simulations of Posterior Density II

We are now going to investigate the effect of different values of K_γ on the inference of the posterior density of τ. We compare the posterior densities of τ corresponding to different values of K_γ. We change $K_\gamma = 1$ in the above simulation to $K_\gamma = 2, 10, 100, 1000$ respectively and also generate 100 observations from (2.67), for each value of K_γ and values of γ, τ from Table 2.1.

First, we plot the posterior densities of τ corresponding to $K_\gamma = 2, 10, 100, 1000$ respectively in Figure 2.2 to Figure 2.5. In each figure (a) to (f) refers to different values of (γ, τ) accordingly to Table 2.1. (Notice that we have plotted the posterior densities of τ for $K_\gamma = 1$ in Figure 2.1.

We see from Figure 2.1 to Figure 2.5, where $K_\gamma = 1, 2, 10, 100, 1000$ respectively, that the graphs of the posterior density of τ are similar except (a), the posterior densities for τ when the true value of $\gamma = 0$ and τ is meaningless in the model. When K_γ increases, the posterior densities for τ become more "diffuse" especially when the value of K_γ increases from 2 to 10. The posterior density for τ looks like a uniform distribution over the measuring time interval when $K_\gamma \geq 10$. We know that a uniform distribution provides the lest information than other distributions. So we think when K_γ becomes larger, the estimation of the posterior density for τ becomes better, when the true value of $\gamma = 0$ and τ becomes meaningless in the model. Our simulations also suggest that there is little effect to the estimation of

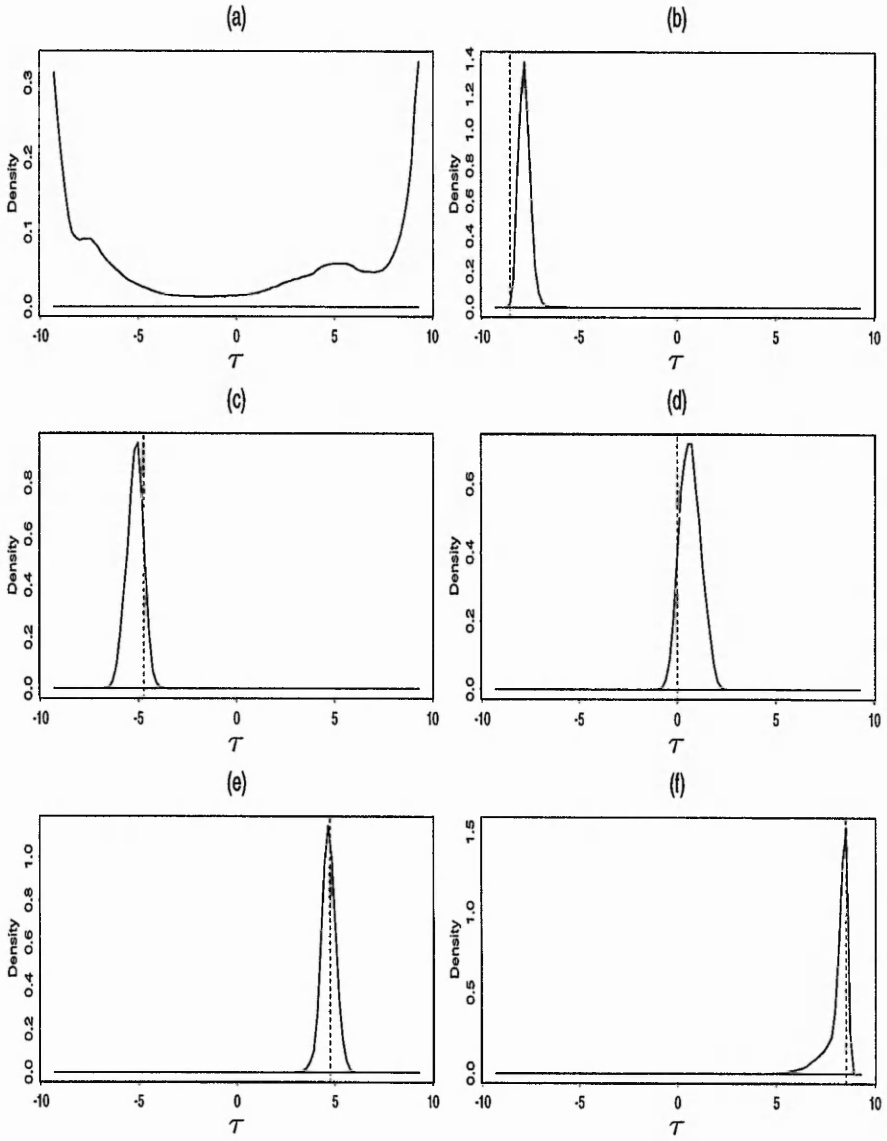


Figure 2.2: Posterior densities of τ in two-phase model. $K_\gamma = 2$.

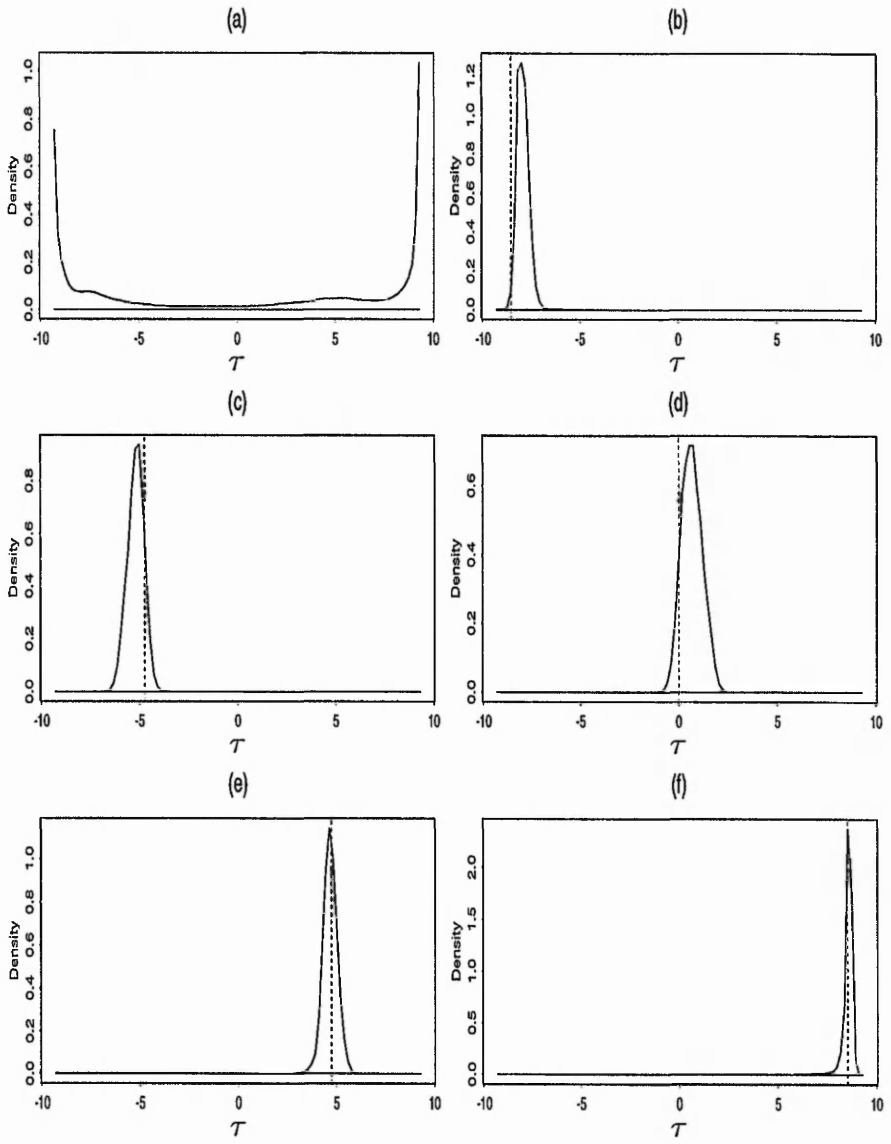


Figure 2.3: Posterior densities of τ in two-phase model. $K_\gamma = 10$.

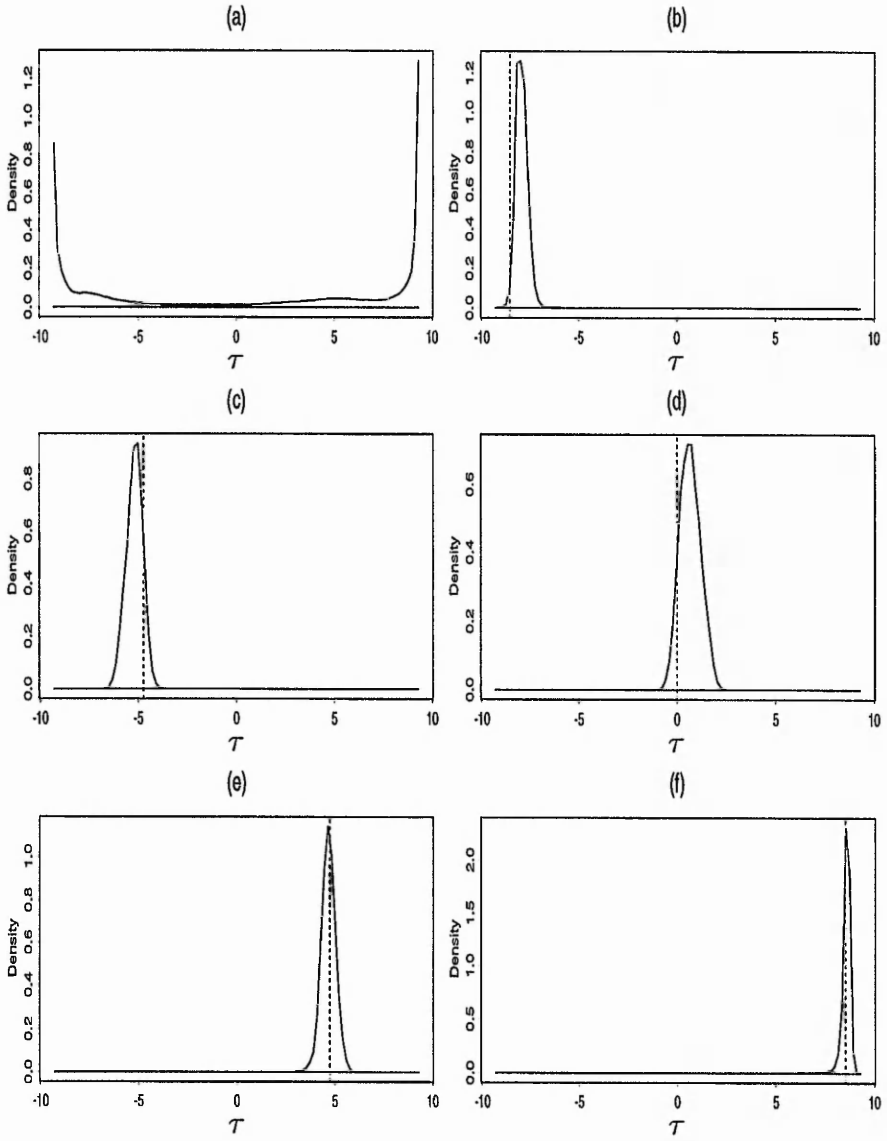


Figure 2.4: Posterior densities of τ in two-phase model. $K_\gamma = 100$.

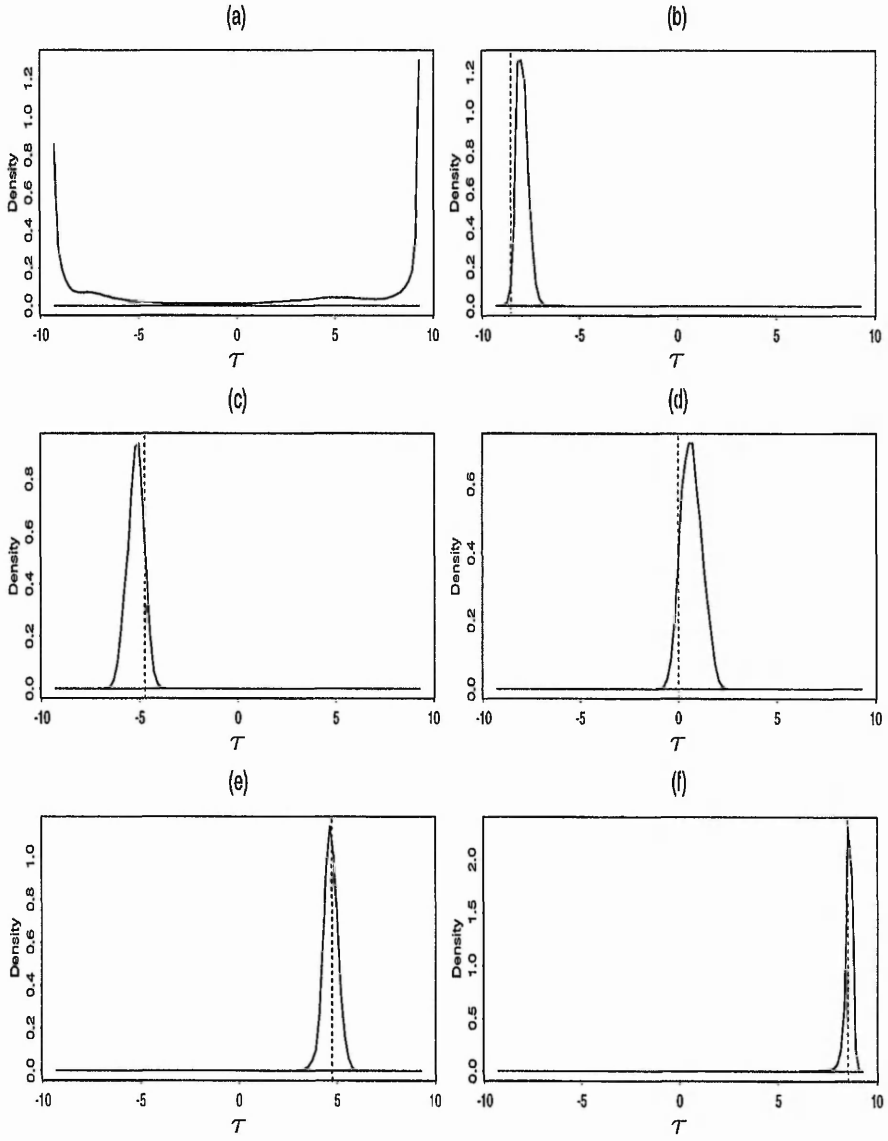


Figure 2.5: Posterior densities of τ in two-phase model. $K_\gamma = 1000$.

the posterior density of τ when K_γ is larger than 10.

Second, we list the posterior mean, standard deviation and 95% probability interval for τ in Table 2.2 in three parts separately, with different values of $K_\gamma = 1, 2, 10, 100, 1000$ and different true values of (γ, τ) in Table 2.1, labelled as (a) to (f).

We see the column (a), the true value of $\gamma = 0$, the sample data comes from model M_0, where τ has no meaning. The posterior standard deviation of τ increases when K_γ increases, reflecting the increasing uncertainty in the prior of τ; the length of the 95% probability interval for τ increases when K_γ increases, which also shows the uncertainty of τ. So from the results at column (a), we prefer to choose bigger value of K_γ. We note that when K_γ is greater than 10, there is little additional effect.

For columns (b) to (f) in Table 2.2, the sample data comes from model M_1. For column (b), the true value of $\tau = -8.55$, which is outside of all the 95% probability intervals; For columns (c) to (f), all the 95% probability intervals contains the corresponding true value of τ except when $K_\gamma = 1$ in column (f). At column (f), we found improvement in inference for τ when K_γ becomes bigger, the standard deviations decrease, the length of the 95% probability intervals decreases too. However, when K_γ is great than 10, no obvious improvement of the inference of τ.

From above discussions, we know that when the sample data come from model M_0, that is, $\gamma = 0$, the change of slope term $\gamma(t_i - \tau)^+$ disappears from the model and τ becomes meaningless. Larger value of K_γ means larger standard deviation of the prior for γ, more "diffuse" prior for γ. Our simulations suggest that when $\gamma = 0$, more "diffuse" prior for γ is better than a normal distribution with a relatively small variance. However too large value for the variance of the prior normal distribution is not necessary because it provides little effect.

Simulations of Model Comparison

Finally, we would like to see how our approach to model comparison behaves for different simulations. We generate data from model (2.67) with the same values for the parameters except γ, τ and the sample size n. Every simulation we use consists of 1000 replications. In the following tables, when $\gamma = 0$, it means that data were

K_γ	(a)	(b)	(c)	(d)	(e)	(f)
	mean					
1	0.068	-7.59	-5.14	0.667	4.64	7.16
2	0.0488	-7.81	-5.19	0.672	4.67	8.14
10	0.324	-7.94	-5.2	0.674	4.68	8.56
100	0.459	-7.95	-5.2	0.674	4.68	8.57
1000	0.461	-7.95	-5.2	0.674	4.68	8.57
	standard deviation					
1	6.79	0.324	0.426	0.537	0.384	0.858
2	7.1	0.312	0.434	0.536	0.382	0.62
10	7.53	0.314	0.437	0.536	0.382	0.302
100	7.61	0.315	0.437	0.536	0.382	0.289
1000	7.61	0.315	0.437	0.536	0.382	0.289
	95% probability interval					
1	(-9.17, 9.17)	(-8.17, -6.9)	(-6.02, -4.34)	(-0.32, 1.76)	(3.86, 5.39)	(5.24, 8.45)
2	(-9.22, 9.22)	(-8.36, -7.14)	(-6.08, -4.37)	(-0.314, 1.76)	(3.91, 5.42)	(6.34, 8.75)
10	(-9.26, 9.27)	(-8.52, -7.28)	(-6.1, -4.38)	(-0.313, 1.76)	(3.92, 5.43)	(7.88, 8.92)
100	(-9.27, 9.28)	(-8.52, -7.28)	(-6.1, -4.39)	(-0.313, 1.76)	(3.92, 5.44)	(7.95, 8.93)
1000	(-9.27, 9.28)	(-8.52, -7.28)	(-6.1, -4.39)	(-0.313, 1.76)	(3.92, 5.44)	(7.95, 8.93)

Table 2.2: Effects of different values of K_γ for inference of τ.

generated from model M_0; otherwise data were generated from model M_1. The frequency of the event $\{P(M_0|\mathbf{y}) > 0.5\}$ in the 1000 replications is denoted as p, so, when $\gamma = 0$, a large value of p corresponds to a "success" performance of our approach; when $\gamma \neq 0$, the smaller value of p is, the better our approach performs. The proportions of times $P(M_0|\mathbf{y}) > 0.5$ for each simulation are listed in Table 2.3.

			γ										
No.	n	τ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
1	20	0	93.9	88.7	75	51.7	29.1	10.7	3.6	0.4	0.1	0	
2	64	0	96.9	87.3	50.2	16	1.8	0					
3	64	-5	96.9	89	72	45.2	22.4	7.7	2.3	0.2	0		
4	64	-7	96.9	89	82.2	73.5	58.9	41.8	29.6	19	11.2	6.3	2.9
5	128	-7	93	89.2	78.9	63	42.7	24.3	13.2	5.1	1.3	0.5	0.1
6	AR(1)		70.8	69.7	65.4	59.2	49.8	39.5	31.4	21.9	16.4	11.2	6.4

Table 2.3: Proportion of times $P(M_0|\mathbf{y}) > 0.5$ in 1000 replications.

We divide the simulations into 6 groups according to the value of (n, τ). Within each group, we have 11 simulations according to different value of γ. So totally we have 66 simulations. When $\gamma = 0$, the data come from model M_0, so the higher proportions of times $P(M_0|\mathbf{y}) > 0.5$ the better; when $\gamma \neq 0$, the data come from model M_1, so the lower proportions of times $P(M_0|\mathbf{y}) > 0.5$ the better. We can observe two trends in the table: when $\gamma = 0$, the larger the sample size the better results; when $\gamma \neq 0$, the larger the value of γ the better results. We explain the details of the simulation results as follows.

For group 1 simulations, we take $\tau = 0$ and $n = 20$. For group 2 simulation, we increase the sample size to $n = 64$. As the sample size increases, our test performs better in group 2 simulations than in group 1 simulations. For group 3 simulations, we change $\tau = 0$ to $\tau = -5$. This means that in model M_1 the regression line changes its slope at time $t = -5$ of the way through the observed sample, from $t = -9.5$ to $t = 9.5$. We still choose $n = 64$. It should be more difficult to test which model the data come from than in group 2 simulations because τ is nearer to the beginning of sampling time. For the extreme case, when τ goes to the start (or end) of sampling time, $(\tau \to t_1)$, M_1 becomes the same as model M_0, and the two models, M_0 and

M_1, are indistinguishable. We notice that when the data are generated from model M_1, our testing results become more favourable to M_0 compared with the results in group 2 simulations, although the difference is not very big. For group 3 simulations, we choose $\tau = -7$, even closer to the beginning of sampling time. The results are worse than those in group 3 simulations. For group 5 simulations, we increase sample size to $n = 128$, still with $\tau = -7$. The results are better than those from group 4 simulations. Finally, for group 6 in simulations, in order to test the robustness of our approach, we replace the *iid* N(0, 1) process ε_t with an AR(1) process

$$u_t = \rho u_{t-1} + \varepsilon_t, \tag{2.68}$$

where $\rho = 0.5$, ε_t is still *iid* N(0, 1) and keep n and τ the same values as for group 5 in simulations. The results for group 6 simulations show the worse performance of our approach for AR disturbance term than the results for the model with white noise.

Chapter 3

Two-phase Model II

In this chapter, we focus on the topic of expanding the two-phase model discussed in Chapter 2 with the disturbance term changed to an AR(1), a first order autoregressive process, because AR(1) disturbances are sometimes encountered in practice. The expanded two-phase model is

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 t_i + \gamma(t_i - \tau)^+ + u_i, \\ u_i &= \rho u_{i-1} + \varepsilon_i, \end{aligned} \tag{3.1}$$

where $x^+ = \max\{0, x\}$, $\tau \in [t_2, t_{n-1}]$, $\{\varepsilon_i\}$ is an iid $\mathrm{N}(0, \sigma^2)$ process, $i = 1, 2, \ldots, n$. $\rho \in (-1, 1)$ because we are assuming that u_i, $i = 1, 2, \ldots, n$ is stationary. If we denote $\tau_i = (t_i - \tau)^+$, we can rewrite the model as follows

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 t_i + \gamma \tau_i + u_i, \\ u_i &= \rho u_{i-1} + \varepsilon_i. \end{aligned} \tag{3.2}$$

In the above model, y_i is the ith observation measured at time t_i, u_i is the first-order autoregressive process. Note that if $\rho = 0$, this model would reduce to the simple model discussed in Chapter 2.

We wish to use a Bayesian approach to choose between a two-phase model and a linear model, when the disturbance terms for both models are AR(1). The null model M_0, is represented as

$$\begin{aligned} M_0 : \quad & y_i = \alpha_0 + \alpha_1 t_i + u_i, \\ & u_i = \rho u_{i-1} + \varepsilon_i, \end{aligned} \tag{3.3}$$

where $\rho \in (-1, 1)$. The alternative two-phase model, M_1, is represented as

$$\begin{aligned} M_1 : \quad & y_i = \alpha_0 + \alpha_1 t_i + \gamma \tau_i + u_i, \\ & u_i = \rho u_{i-1} + \varepsilon_i. \end{aligned} \tag{3.4}$$

3.1 The likelihood functions

In order to test hypothesis in a Bayesian frame work, we need to obtain the likelihood functions corresponding to the null and the alternative models, that is the likelihood functions for model M_0 and model M_1. We begin with the analysis of model M_1, because the analysis of model M_0 is similar and simpler.

By eliminating u_i, u_{i-1} from the above model M_1, we obtain

$$M_1 : \quad y_i - \rho y_{i-1} = (1-\rho)\alpha_0 + (t_i - \rho t_{i-1})\alpha_1 + (\tau_i - \rho \tau_{i-1})\gamma + \varepsilon_i. \tag{3.5}$$

Note that the new symbol y_0 appears when $i = 1$. y_0 represents the initial condition of the model. The parameter vector of model M_1 is $\theta_1 = (\alpha_0, \alpha_1, \gamma, \tau, \sigma, \rho)'$.

If we assume that the process $\{y_i\}$ was operative for $i = 0$, that is, operative for time t_0, then the initial value, y_0, for the model is given by

$$y_0 = \alpha_0 + \alpha_1 t_0 + u_0. \tag{3.6}$$

We discuss the distribution of the initial value y_0 through the properties of the disturbance term $\{u_i\}$. The disturbance term $\{u_i\}$ is an AR(1) stationary process which has the properties $E(u_i) = 0$ and $\text{var}(u_i) = \sigma_u$, where σ_u is a constant. It is easy to verify that $u_i \sim \text{N}\left(0, \frac{\sigma^2}{1-\rho^2}\right)$, see, for example, Box and Jenkins (1976), in particular we have $u_0 \sim \text{N}\left(0, \frac{\sigma^2}{1-\rho^2}\right)$ and therefore $y_0 \sim \text{N}\left(\alpha_0 + \alpha_1 t_0, \frac{\sigma^2}{1-\rho^2}\right)$. We denote $\tilde{u}_0 = \sqrt{1-\rho^2}u_0$, then $\tilde{u}_0 \sim \text{N}(0, \sigma^2)$ and $\tilde{u}_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ are iid $\text{N}(0, \sigma^2)$.

We transform data from $(y_0, y_1, \ldots, y_n)$ to $(x_0, x_1, \ldots, x_n)$ as follows. We set $x_0 = \sqrt{1-\rho^2}y_0 = \sqrt{1-\rho^2}\alpha_0 + \sqrt{1-\rho^2}t_0\alpha_1 + \tilde{u}_0$ and $x_i = y_i - \rho y_{i-1}$, $i = 1, 2, \ldots, n$. If $\rho = 0$, this transform will be identical and the disturbance terms, u_i, $i = 1, 2, \ldots, n$, will be white noise. We denote $\mathbf{x} = (x_0, x_1, x_2, \ldots, x_n)'$, $\boldsymbol{\varepsilon} = (\tilde{u}_0, \varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$,

$\boldsymbol{\beta} = (\alpha_0, \alpha_1, \gamma)'$ so that $\theta = (\boldsymbol{\beta}', \tau, \sigma)$ and

$$X = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}t_0 & 0 \\ 1-\rho & t_1 - \rho t_0 & \tau_1 - \rho\tau_0 \\ 1-\rho & t_2 - \rho t_1 & \tau_2 - \rho\tau_1 \\ \vdots & \vdots & \vdots \\ 1-\rho & t_n - \rho t_{n-1} & \tau_n - \rho\tau_{n-1} \end{pmatrix}.$$

Notice that $\mathbf{x}$ is unobservable, x_0 and x_1 are functions of ρ and y_0, and x_i, $i \neq 1$ are functions of ρ. X has unknown ρ and τ as parameters. We can rewrite model M_1 in matrix notation as follows

$$M_1 : \qquad \mathbf{x} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{3.7}$$

The Jacobian of the transformation $\mathbf{y} \longmapsto \mathbf{x}$ is as follows

$$J = \left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| = \begin{vmatrix} \sqrt{1-\rho^2} & 0 & & & \\ -\rho & 1 & & & \\ & -\rho & 1 & & \\ & & \ddots & \ddots & \\ & & & -\rho & 1 \end{vmatrix} = \sqrt{1-\rho^2}.$$

The joint probability density function for $\mathbf{x} = (x_0, x_1, \ldots, x_n)'$, is given by

$$p(\mathbf{x}|\theta, M_1) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}, \tag{3.8}$$

where $\nu = (n+1) - 3 = n - 2$, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{x}$, and $\nu s^2 = (\mathbf{x} - X\hat{\boldsymbol{\beta}})'(\mathbf{x} - X\hat{\boldsymbol{\beta}})$.

The likelihood function for model M_1, or the joint probability density for $\mathbf{y}$, is given as

$$\begin{aligned} p(\mathbf{y}|\theta, M_1) &= p(\mathbf{x}|\theta, M_1)|J| \\ &= \sqrt{1-\rho^2}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n+1} \times \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}. \end{aligned} \tag{3.9}$$

We can obtain likelihood function for model M_0 by noting that it can be setting $\gamma = 0$ in model M_1.

By eliminating u_t and u_{t-1} from model M_0, we obtain

$$M_0 : \quad y_i - \rho y_{i-1} = (1-\rho)\alpha_0 + (t_i - \rho t_{i-1})\alpha_1 + \varepsilon_i. \tag{3.10}$$

Using similar notation to that in model M_1, we have, for model M_0,

$$X_0 = \begin{pmatrix} \sqrt{1-\rho^2} & \sqrt{1-\rho^2}t_0 \\ 1-\rho & t_1 - \rho t_0 \\ \vdots & \vdots \\ 1-\rho & t_n - \rho t_{n-1} \end{pmatrix},$$

$\boldsymbol{\alpha} = (\alpha_0, \alpha_1)'$ and the parameter vector for model M_0 is $\theta_0 = (\alpha_0, \alpha_1, \sigma, \rho)$, model M_0 is represented in matrix notation by

$$M_0 : \qquad \mathbf{x} = X_0\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{3.11}$$

where $\mathbf{x}$ and $\boldsymbol{\varepsilon}$ are same as in model M_1. We then obtain the joint probability density function for $\mathbf{x} = (x_0, x_1, \ldots, x_n)'$ for model M_0 as

$$p(\mathbf{x}|\theta_0, M_0) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right]\right\}, \tag{3.12}$$

where $\nu_0 = (n+1) - 2 = n-1$, $\hat{\boldsymbol{\alpha}} = (X_0'X_0)^{-1}X_0'\mathbf{x}$, and $\nu_0 s_0^2 = (\mathbf{x} - X_0\hat{\boldsymbol{\alpha}})'(\mathbf{x} - X_0\hat{\boldsymbol{\alpha}})$.

The likelihood function for model M_0, is then given by

$$\begin{aligned} p(\mathbf{y}|\theta_0, M_0) &= p(\mathbf{x}|\theta_0, M_0)|J| \\ &= \sqrt{1-\rho^2}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{n+1} \times \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})\right]\right\}. \end{aligned} \tag{3.13}$$

3.2 Priors and Posteriors for model M_1

We discuss priors for model M_1 first. We choose a conjugate prior for $\boldsymbol{\beta}$ conditional on σ in model M_1

$$p(\boldsymbol{\beta}|\sigma, M_1) = \frac{|V|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^3} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^0)' V^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}^0)\right\}, \tag{3.14}$$

where V is a constant 3×3 positive definite symmetric matrix, $\boldsymbol{\beta}^0$ is a constant 3 dimensional vector. The conjugate prior for σ in model M_1 is an inverted gamma distribution with density

$$p(\sigma|M_1) = \frac{2}{\Gamma(b/2)}\left(\frac{a}{2}\right)^{b/2} \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\}, \quad 0 < \sigma < +\infty, \tag{3.15}$$

where $a, b > 0$ are constants.

We know nothing about τ but $\tau \in [t_2, t_{n-1}]$, so we choose an uniform distribution for τ in model M_1 with the density

$$p(\tau|M_1) = \begin{cases} \dfrac{1}{t_{n-1} - t_2}, & \tau \in (t_2, t_{n-1}), \\ 0, & \text{otherwise.} \end{cases} \tag{3.16}$$

We know $\rho \in (-1, 1)$, which corresponds to the stationary case of the disturbance term, so we choose a uniform distribution for ρ in model M_1 with the density

$$p(\rho|M_1) = \begin{cases} 1/2, & \rho \in (-1, 1), \\ 0, & \text{otherwise.} \end{cases} \tag{3.17}$$

We suppose that $(\boldsymbol{\beta}, \sigma)$, τ and ρ are *a priori* independently distributed in model M_1, so that we have the joint prior for the parameter vector of model M_1 as follows

$$\begin{aligned} p(\theta|M_1) = \frac{1}{\sigma^{b+4}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^0)' V^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}^0)\right\} \\ \times \frac{|V|^{-1/2}}{(2\pi)^{3/2}\Gamma(b/2)(t_{n-1} - t_2)} \left(\frac{a}{2}\right)^{\frac{b}{2}} \exp\left\{-\frac{a}{2\sigma^2}\right\}. \end{aligned} \tag{3.18}$$

Then we can obtain the joint distribution for $\mathbf{y}$ and θ in model M_1 as

$$\begin{aligned} p(\mathbf{y}, \theta|M_1) = \frac{\sqrt{1-\rho^2}}{\sigma^{n+b+5}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^1)' V_1 (\boldsymbol{\beta} - \boldsymbol{\beta}^1)\right\} \\ \times \frac{|V|^{-1/2}}{(2\pi)^{\frac{n+4}{2}}\Gamma(b/2)(t_{n-1} - t_2)} \left(\frac{a}{2}\right)^{\frac{b}{2}} \exp\left\{-\frac{\nu_1 s_1^2}{2\sigma^2}\right\}, \end{aligned} \tag{3.19}$$

where $V_1 = X_1' X_1 + V^{-1}$, $\boldsymbol{\beta}^1 = V_1^{-1}(X_1' X_1 \hat{\boldsymbol{\beta}} + V^{-1}\boldsymbol{\beta}^0)$ and $\nu_1 s_1^2 = a + \nu s^2 + (\boldsymbol{\beta}^1 - \boldsymbol{\beta}^0) V^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0)$.

If we integrate (3.19) with respect to $\boldsymbol{\beta}$, we obtain

$$\begin{aligned} p(\mathbf{y}, \tau, \sigma, \rho|M_1) = \frac{|V|^{-1/2}}{(2\pi)^{\frac{n+1}{2}}\Gamma(b/2)(t_{n-1} - t_2)} \left(\frac{a}{2}\right)^{\frac{b}{2}} \\ \times \frac{\sqrt{1-\rho^2}}{\sigma^{n+b+2}|V_1|^{\frac{1}{2}}} \exp\left\{-\frac{\nu_1 s_1^2}{2\sigma^2}\right\}, \end{aligned} \tag{3.20}$$

and integrating this with respect to σ we obtain

$$\begin{aligned} p(\mathbf{y}, \tau, \rho|M_1) = \frac{2^{\frac{n-1}{2}} a^{\frac{b}{2}} |V|^{-\frac{1}{2}} \Gamma\left(\frac{n+b+1}{2}\right)}{(2\pi)^{\frac{n+1}{2}}\Gamma(b/2)(t_{n-1} - t_2)} \\ \times \frac{\sqrt{1-\rho^2}}{|V_1|^{\frac{1}{2}}(\nu_1 s_1^2)^{\frac{n+b+1}{2}}}. \end{aligned} \tag{3.21}$$

From above (3.21), we obtain the joint posterior density for (τ, ρ) as

$$p(\tau, \rho|\mathbf{y}, M_1) \propto \frac{\sqrt{1-\rho^2}}{|V_1|^{\frac{1}{2}}(\nu_1 s_1^2)^{\frac{n+b+1}{2}}}. \tag{3.22}$$

Finally we integrate (3.21) with respect to τ and ρ to obtain

$$p(\mathbf{y}|M_1) = \frac{2^{\frac{n-1}{2}} a^{\frac{b}{2}} |V|^{-\frac{1}{2}} \Gamma\left(\frac{n+b+1}{2}\right)}{(2\pi)^{\frac{n+1}{2}} \Gamma(b/2)(t_{n-1} - t_2)} \times \int_{-1}^{1} \int_{t_2}^{t_{n-1}} \frac{\sqrt{1-\rho^2}}{|V_1|^{\frac{1}{2}} (\nu_1 s_1^2)^{\frac{n+b+1}{2}}} d\tau d\rho, \tag{3.23}$$

which is needed for the model comparison.

3.3 Priors and Posteriors for model M_0

Now we discuss priors and posteriors for model M_0. As for $\boldsymbol{\beta}$ in model M_1, we choose a conjugate prior for $\boldsymbol{\alpha}$ conditional on σ in model M_0, namely

$$p(\boldsymbol{\alpha}|\sigma, M_0) = \frac{|V_0|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)' V_0^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)\right\}, \tag{3.24}$$

where V_0 is a constant 2×2 positive definite symmetric matrix. We suppose that V_0 is the upper-left sub-matrix of V, which appears in (3.14), $V = \begin{pmatrix} V_0 & v_{12} \\ v_{12} & v_{22} \end{pmatrix}$.

We choose the same priors for σ and ρ as in model M_1, so we have

$$p(\sigma|M_0) = \frac{2}{\Gamma(b/2)} \left(\frac{a}{2}\right)^{b/2} \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\}, \quad 0 < \sigma < +\infty, \tag{3.25}$$

where a and b are the same constants used in model M_1 and

$$p(\rho|M_0) = \begin{cases} 1/2, & \rho \in (-1, 1), \\ 0, & \text{otherwise.} \end{cases} \tag{3.26}$$

We again suppose that $(\boldsymbol{\alpha}, \sigma)$ and ρ are *a priori* independently distributed so that we can obtain the joint prior for the parameter vector as follows

$$p(\theta_0|M_0) = \frac{1}{\sigma^{b+3}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)' V_0^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^0)\right\} \times \frac{|V_0|^{-1/2}}{2\pi\Gamma(b/2)} \left(\frac{a}{2}\right)^{\frac{b}{2}} \exp\left\{-\frac{a}{2\sigma^2}\right\}. \tag{3.27}$$

We then obtain the joint distribution for $\mathbf{y}$ and θ_0 in model M_0 as

$$p(\mathbf{y}, \theta_0|M_0) = \frac{\sqrt{1-\rho^2}}{\sigma^{n+b+4}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^1)' V_{01} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^1)\right\} \times \frac{|V_0|^{-1/2}}{(2\pi)^{\frac{n+3}{2}} \Gamma(b/2)} \left(\frac{a}{2}\right)^{\frac{b}{2}} \exp\left\{-\frac{\nu_{01} s_{01}^2}{2\sigma^2}\right\}, \tag{3.28}$$

where $V_{01} = X_0'X_0 + V_0^{-1}$, $\boldsymbol{\alpha}^1 = V_0^{-1}(X_0'X_0\hat{\boldsymbol{\alpha}} + V_0^{-1}\boldsymbol{\alpha}^0)$ and $\nu_{01}s_{01}^2 = a + \nu_0 s_0^2 + (\boldsymbol{\alpha}^1 - \boldsymbol{\alpha}^0)V_0^{-1}(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}^0)$.

If we integrate (3.28) with respect to $\boldsymbol{\alpha}$, we obtain

$$p(\mathbf{y}, \sigma, \rho|M_0) = \frac{|V_0|^{-1/2}}{(2\pi)^{\frac{n+1}{2}}\Gamma(b/2)}\left(\frac{a}{2}\right)^{\frac{b}{2}} \times \frac{\sqrt{1-\rho^2}}{\sigma^{n+b+2}|V_{01}|^{\frac{1}{2}}} \exp\left\{-\frac{\nu_{01}s_{01}^2}{2\sigma^2}\right\}, \tag{3.29}$$

then the posterior joint distribution for (σ, ρ) is

$$p(\sigma, \rho|\mathbf{y}, M_0) \propto \frac{\sqrt{1-\rho^2}}{\sigma^{n+b+2}|V_{01}|^{\frac{1}{2}}} \exp\left\{-\frac{\nu_{01}s_{01}^2}{2\sigma^2}\right\}. \tag{3.30}$$

If we integrate (3.29) with respect to σ, we obtain

$$p(\mathbf{y}, \rho|M_0) = \frac{2^{\frac{n-1}{2}}a^{\frac{b}{2}}|V_0|^{-\frac{1}{2}}\Gamma\left(\frac{n+b+1}{2}\right)}{(2\pi)^{\frac{n+1}{2}}\Gamma(b/2)} \times \frac{\sqrt{1-\rho^2}}{|V_{01}|^{\frac{1}{2}}(\nu_{01}s_{01}^2)^{\frac{n+b+1}{2}}}. \tag{3.31}$$

Correspondingly, the posterior density for ρ is

$$p(\rho|\mathbf{y}, M_0) \propto \frac{\sqrt{1-\rho^2}}{|V_{01}|^{\frac{1}{2}}(\nu_{01}s_{01}^2)^{\frac{n+b+1}{2}}}. \tag{3.32}$$

If we integrate (3.31) with respect to ρ, we obtain

$$p(\mathbf{y}|M_0) = \frac{2^{\frac{n-1}{2}}a^{\frac{b}{2}}|V_0|^{-\frac{1}{2}}\Gamma\left(\frac{n+b+1}{2}\right)}{(2\pi)^{\frac{n+1}{2}}\Gamma(b/2)} \times \int_{-1}^{1} \frac{\sqrt{1-\rho^2}}{|V_{01}|^{\frac{1}{2}}(\nu_{01}s_{01}^2)^{\frac{n+b+1}{2}}}\, d\rho, \tag{3.33}$$

which will be used in model comparison between model M_0 and M_1.

3.4 Posterior Model Probabilities

As we have obtained $p(\mathbf{y}|M_0)$ and $p(\mathbf{y}|M_1)$, if we set equal priors for the models, that is, $P(M_0) = P(M_1) = 0.5$, then we can use (2.13) on page 9 to obtain the posterior probabilities as

$$\begin{aligned} P(M_0|\mathbf{y}) &= \frac{1}{1+k}, \\ P(M_1|\mathbf{y}) &= \frac{k}{1+k}, \end{aligned} \tag{3.34}$$

where

$$k = \frac{|V|^{-\frac{1}{2}}}{(t_{n-1}-t_2)|V_0|^{-\frac{1}{2}}} \frac{\int_0^1 \int_{t_2}^{t_{n-1}} \frac{\sqrt{1-\rho^2}}{|V_1|^{\frac{1}{2}}(\nu_1 s_1^2)^{\frac{n+b+1}{2}}} d\tau d\rho}{\int_0^1 \frac{\sqrt{1-\rho^2}}{|V_{01}|^{\frac{1}{2}}(\nu_{01} s_{01}^2)^{\frac{n+b+1}{2}}} d\rho}.$$

We would favour the model which has the bigger posterior model probability.

Chapter 4

Smooth Transition Models

Smooth transition models are widely used to model economic time series, representing deterministic structural change in a time series regression as an alternative to the structural break models discussed in Chapter 2 and Chapter 3. The structural change of smooth transition model happens between different regimes over time while the structural break happens at some point instantaneously, for example, in the two-phase model. Considerable attention has been devoted to the smooth transition models for decades, including estimation of the parameters in the smooth transition models and testing of smooth transition models against other models, such as linear models, unit root models.

Smooth transition models were originally proposed by Bacon and Watts (1971). Bacon and Watts (1971) argued the two-phase model is only appropriate if it is known that an abrupt transition happens. This is the disadvantage of the two-phase model because usually the nature of the structural change of practical data sets is not known *a priori*. Bacon and Watts (1971) introduced a smooth transition function $\text{trn}\{(x - x_0)/\gamma\}$ and suggested that the smooth transition function $\text{trn}\{(x - x_0)/\gamma\}$ satisfies the following conditions:

$$
\begin{aligned}
&(a) \quad \lim_{s \to \infty} \text{trn}(|s|/\gamma) = 1; \\
&(b) \quad \text{trn}(0) = 0; \\
&(c) \quad \lim_{\gamma \to 0} \text{trn}(s/\gamma) = \text{sgn}(s); \\
&(d) \quad \lim_{s \to \infty} s \, \text{trn}(s/\gamma) = s.
\end{aligned}
$$

By changing the value of γ, $\text{trn}\{(x - x_0)/\gamma\}$ describes abrupt transition ($\gamma \approx 0$) or

very gradual transition ($\gamma > 1$). Bacon and Watts (1971) considered the following smooth transition model

$$Y = \alpha_0 + \alpha_1(x - x_0) + \alpha_2(x - x_0)\text{trn}\{(x - x_0)/\gamma\} + Z, \tag{4.1}$$

where Z is a random variable accounting for error. Bacon and Watts (1971) also suggested specific transition functions that can be used: the cumulative distribution function of any symmetric probability density function or the hyperbolic tangent. A Bayesian approach had been used to estimate the parameters in the model (4.1) with a hyperbolic tangent function as the smooth transition function, $\text{trn}(s/\gamma) = \tanh(s/\gamma)$.

Maddala (1977) discussed the two-phase regression (which he called switching regression) and then suggested an alternative to the two-phase regression, a continuously varying parameter regression given by

$$\begin{aligned} y_t &= \beta_t x_t + u_t, \\ \beta_t &= \bar{\beta} + \frac{c}{1 + e^{\alpha - \delta z_t}}, \end{aligned} \tag{4.2}$$

where $\delta < 0$. Maddala (1977) suggested that the parameters of (4.2) can be estimated by nonlinear least-squares methods.

Teräsvirta and Anderson (1992) investigated various time-series representing business cycles, such as production and unemployment and considered these cycles as being nonlinear and can be adequately described by a smooth transition autoregressive (STAR) model given by

$$y_t = \pi_{10} + \pi_1' w_t + (\pi_{20} + \pi_2' w_t) F(y_{t-d}) + u_t, \tag{4.3}$$

where u_t is *i.i.d.* $\text{N}(0, \sigma^2)$, $\pi_j = (\pi_{j1}, \ldots, \pi_{jp})'$, $j = 1, 2$, $w_t = (y_{t-1}, \ldots, y_{t-p})'$ and F is the transition function proposed by either

$$F(y_{t-d}) = (1 + \exp\{-\gamma(y_{t-d} - c)\})^{-1}, \quad \gamma > 0, \tag{4.4}$$

or

$$F(y_{t-d}) = 1 - \exp\{-\gamma(y_{t-d} - c)^2\}, \quad \gamma > 0, \tag{4.5}$$

and the corresponding smooth transition model is called either LSTAR or ESTAR model. The issue of nonlinearity of business important because if the business cycles

are nonlinear, then the linear model theory will be inadequate and produce inferior prediction in practice. Teräsvirta and Anderson (1992) conducted a test of the linear AR model against the STAR model. If the linearity was rejected, a sequence of tests of nested hypotheses would carry out to choose between LSTAR and ESTAR models and the specific models are estimated by the nonlinear least-square methods.

Lin and Teräsvirta (1994) discussed similar smooth transition regression given by

$$y_t = x_t'\pi_1 + x_t'\pi_2 F(z_t) + u_t, \tag{4.6}$$

where $x_t = (1, y_{t-1}, \ldots, y_{t-p}, x_{1t}, \ldots, x_{qt})'$ is an $m \times 1$ vector, $m = p + 1 + q$, $\pi_1 = (\pi_{11}, \ldots, \pi_{1m})'$, $\pi_2 = (\pi_{21}, \ldots, \pi_{2m})'$ are $m \times 1$ parameter vectors, and u_t is an error term with $Eu_t = 0$, $Ex_t u_t = 0$, and $Ez_t u_t = 0$. The smooth transition function $F(t)$ is given by either

$$F(t) = F(t, \gamma) = (1 + \exp\{-\gamma(t^k + \alpha_1 t^{k-1} + \ldots + \alpha_{k-1} t + \alpha_k)\})^{-1}, \quad k = 1, 3, \tag{4.7}$$

or

$$F(t) = F(t, \gamma) = 1 - \exp\{-\gamma(t - \alpha)^2\}). \tag{4.8}$$

They used the first-order Taylor approximation of $F(t)$ at $t = 0$, $A(t)$, to substitute $F(t)$ in (4.7) or (4.8). They then test the null hypothesis $\gamma = 0$ against the alternative, the approximated (4.6), substituting $F(t)$ with$A(t)$.

Granger and Teräsvirta (1993, ch.7) suggested the logistic smooth transition regression (LSTR) trend model as

$$\ln(y_t) = \alpha_1 + \beta_1 t + \alpha_2 S_t(\gamma, \tau) + \beta_2 t S_t(\gamma, \tau) + \varepsilon_t, \tag{4.9}$$

where $t = 1, 2, \ldots, T$, $\tau \in (1, T)$ and $S_t(\gamma, \tau) = \{1 + \exp[-\gamma(t - \tau)]\}^{-1}$ is the curvilinear logistic function. Granger and Teräsvirta (1993) noticed the difficulties of using the nonlinear least-squares (NLS) methods to estimate the parameters in model (4.9). The estimate for γ may converge very slowly, particularly if the true value of γ is large. In this case, the transition occurs quickly. The reason for this phenomenon is when the true value of γ is large, the smooth transition function $S_t(\gamma, \tau) = \{1 + \exp[-\gamma(t - \tau)]\}^{-1}$ produces similar values when γ varies in a neighbourhood of the true value of γ. (See our discussions of the value of the smooth

transition function below.) Therefore, the standard errors of the NLS estimate of γ may appear very large, which is not indicating the insignificance of the estimate.

Greenaway, Leybourne and Sapsford (1997) modelled economic growth and structural change with smooth transition models presented above in (4.9). They tested the null hypothesis of constancy of the intercept and trend against the smooth transition alternative. They followed Lin and Teräsvirta (1994) to use a polynomial approximation of $S_t(\gamma, \tau) = \{1 + \exp[-\gamma(t - \tau)]\}^{-1}$ and assumed a third-order Taylor expansion is adequate. The NLS estimates are used for the parameters in model (4.9).

Leybourne, Newbold and Vougas (1998) considered testing non-stationary processes, autoregressive unit root, against alternatives that were stationary around a smooth transition. They considered the following three smooth transition regression models

$$\begin{aligned} \text{Model A} \quad & y_t = \alpha_1 + \alpha_2 S_t(\gamma, \tau) + \nu_t, \\ \text{Model B} \quad & y_t = \alpha_1 + \beta_1 t + \alpha_2 S_t(\gamma, \tau) + \nu_t, \\ \text{Model C} \quad & y_t = \alpha_1 + \beta_1 t + \alpha_2 S_t(\gamma, \tau) + \beta_2 t S_t(\gamma, \tau) + \nu_t, \end{aligned} \tag{4.10}$$

where ν_t, $t = 1, 2, \ldots, T$, is a stationary process, $\gamma > 0$ and $S_t(\gamma, \tau)$ is the logistic smooth transition (LST) function, based on the sample of size T,

$$S_t(\gamma, \tau) = \frac{1}{1 + \exp\{-\gamma(t - \tau T)\}}, \tag{4.11}$$

where $\tau \in (0, 1)$. They tested various unit root hypotheses as follows

$$\begin{aligned} \text{Null hypothesis} \quad & y_t = \nu_t, \nu_t = \nu_{t-1} + \varepsilon_t, \nu_0 = \psi, \\ \text{Alternative hypothesis} \quad & \text{Model A, Model B or Model C;} \end{aligned}$$

$$\begin{aligned} \text{Null hypothesis} \quad & y_t = \nu_t, \nu_t = \kappa + \nu_{t-1} + \varepsilon_t, \nu_0 = \psi, \\ \text{Alternative hypothesis} \quad & \text{Model B or Model C;} \end{aligned}$$

then used a nonlinear least-square (NLS) algorithm to estimate the parameters in the smooth transition models (4.10). In Model A and B, the eventual changes of level are represented by α_2, γ and τ. In Model C, the eventual changes of level are

represented by α_2, γ and τ; the eventual changes of slope are presented by β_2, γ and τ.

We develop our approach within the framework of a Bayesian perspective to compare the linear model with the logistic smooth transition regression models which were considered by Leybourne, Newbold and Vougas (1998), in above (4.10). In this chapter, we consider the simplest smooth transition model, Model A in (4.10), with the error term ν_t *i.i.d* $\mathrm{N}(0, \sigma^2)$. In the following chapters, we will consider the remaining smooth transition models with more complex error terms.

The LST function $S_t(\gamma, \tau)$ in (4.11) controls the transitions between different regimes. When $\gamma > 0$, the model transition occurs smoothly from the initial regime

$$y_t = \alpha_1 + \nu_t, \quad t \to -\infty,$$

to the final regime

$$y_t = \alpha_1 + \alpha_2 + \nu_t, \quad t \to +\infty,$$

because $\lim_{t\to-\infty} S_t(\gamma, \tau) = 0$ and $\lim_{t\to+\infty} S_t(\gamma, \tau) = 1$ respectively. When $\gamma < 0$, the model transition occurs gradually in the opposite direction from the initial regime

$$y_t = \alpha_1 + \alpha_2 + \nu_t, \quad t \to -\infty,$$

to the final regime

$$y_t = \alpha_1 + \nu_t, \quad t \to +\infty,$$

because in this case $\lim_{t\to-\infty} S_t(\gamma, \tau) = 1$ and $\lim_{t\to+\infty} S_t(\gamma, \tau) = 0$. Because $S_t(\gamma, \tau) = 1 - S_t(-\gamma, \tau)$, we only need to consider the case $\gamma > 0$ without loss of generality.

Leybourne, Newbold and Vougas (1998) gave the following interpretation of the parameters of $S_t(\gamma, \tau)$ when $\gamma > 0$. The parameter τ determines the timing of the transition midpoint since, for $\gamma > 0$, we have $S_{-\infty}(\gamma, \tau) = 0$, $S_{+\infty}(\gamma, \tau) = 1$ and $S_{\tau T}(\gamma, \tau) = 0.5$. The speed of transition is then determined by the parameter γ. If γ is small then $S_t(\gamma, \tau)$ takes a long period of time to traverse the interval (0, 1), and in the limiting case with $\gamma = 0$, $S_t(\gamma, \tau) = 0.5$ for all t. On the other hand, for large values of γ, $S_t(\gamma, \tau)$ traverses the interval (0, 1) very rapidly, and as γ approaches $+\infty$, this function changes value from 0 to 1 instantaneously at time $t = \tau T$. Thus in (4), y_t is a independent series around a mean which changes from initial value α_1 to final value $\alpha_1 + \alpha_2$.

We will illustrate how the LST function $S_t(\gamma, \tau)$ behaves using numerical examples. For symmetry of $S_t(\gamma, \tau)$ at $\tau = 0.5$, $S_0(\gamma, 0.5) = 1 - S_T(\gamma, 0.5)$, we suppose that $t = 0, 1, \ldots, T$. Now the starting time is at $t = 0$ and the sample size is $T + 1$.

The LST function $S_t(\gamma, \tau)$ is determined by γ and $t - \tau T$, that is, γ and how far away t is from the midpoint τT, and is symmetry at $t = \tau T$. In fact we have $S_{\tau T + \Delta t}(\gamma, \tau) = 1 - S_{\tau T - \Delta t}(\gamma, \tau)$. See Figure 4.1. In Figure 4.1, we plot 4 graphs of $S_t(\gamma, \tau)$ with $\gamma = 1$, $\tau = 0.5$ and different $T = 15, 30, 50, 100$ respectively. Even though the sample size is different, $S_t(\gamma, \tau)$ produces same values around the midpoint τT.

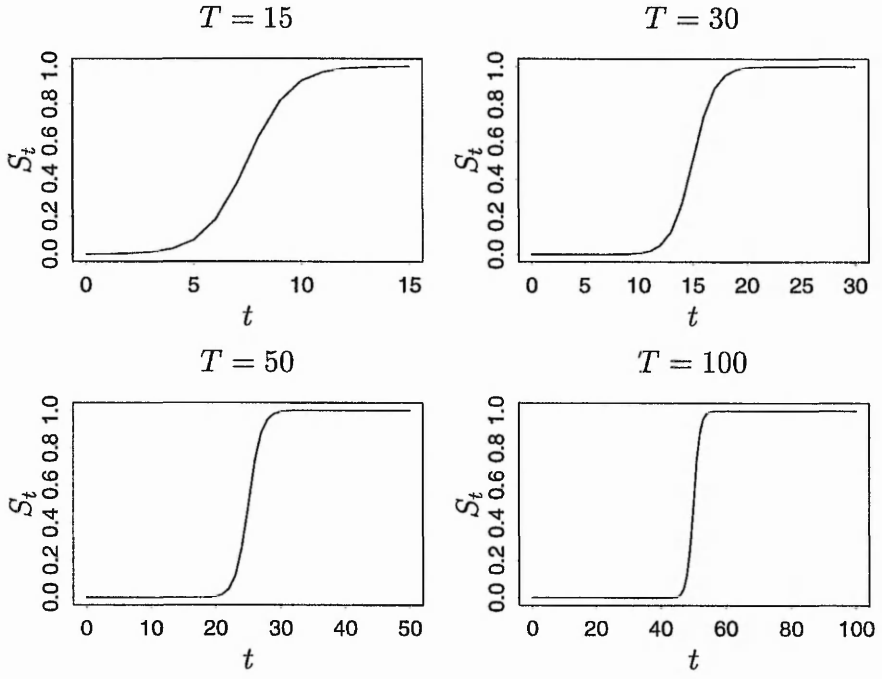


Figure 4.1: Graphs of $S_t(\gamma, \tau)$ with $\gamma = 1$, $\tau = 0.5$ and $T = 15, 30, 50, 100$ respectively.

We calculated values of $S_t(\gamma, \tau)$ around the midpoint τT and the results are presented in Table 4.1, where $\gamma = 0.01, 0.1, 0.5, 1.0, 3.0, 5.0$ respectively. We denote $\Delta t = t - \tau T$. From Table 4.1, we see that, given $\gamma = 0.01$, $S_t(\gamma, \tau)$ is 0.48 when $\Delta t = -8$, increasing to 0.5199 when $\Delta t = 8$. This shows that a small value of γ leads to a slow adjustment (speed of transition) from the old regime of α_1 to the new regime of $\alpha_1 + \alpha_2$. On the other hand, if $\gamma = 5.0$, $S_t(\gamma, \tau)$ is 0.0066 when $\Delta t = -1$

and increases to 0.9933 when $\Delta t = 1$. This shows that an abrupt change from the old regime to the new regime can be expressed using a large value of γ. The data in Table 4.1 show that we can use γ to express the speed of the transition at which $S_t(\gamma, \tau)$ traverses from one regime to the other. The bigger the value of γ, the bigger the speed of the transition. We can use the logistic smooth transition function to express a wide range of structural changes from very smooth transitions to abrupt change.

Besides these two extreme examples, we have $S_{\tau T-8}(1.0, \tau) = 0.0003$ and $S_{\tau T+8}(1.0, \tau) = 0.9996$, so for sample size of 100, the whole range is effectively traversed in 16 observations, a steep step in the bottom picture on the right in Figure 4.1. From Table 4.1, we know that if $\gamma = 3$, the whole range will be traversed in 9 observations symmetrically placed around τT.

From above discussion we know that, for moderate sample size, for example, $T = 100$, values of $\gamma > 1.0$ are extremely unlikely if we know *a priori* that there is a gradual smooth transition occurs with the measuring time interval $t \in [0, T]$. This information will help us to set prior for γ later.

4.1 Model Selection

We wish to use a Bayesian approach to choose between the linear model (constant trend) and the LST model. We will decide from which model the sample data comes on the basis of the posterior model probabilities, $P(M_i|\mathbf{y})$ for the two models. We will favour one model rather than the other if $P(M_i|\mathbf{y})$ is bigger than that of the other. The model M_0 can be stated as follows

$$M_0 : \quad y_t = \alpha_1 + \nu_t, \tag{4.12}$$

where ν_t is *iid* $\mathrm{N}(0, \sigma^2)$, $t = 0, 1, \ldots, T$. The model M_1, is the following LST model

$$M_1 : \quad y_t = \alpha_1 + \alpha_2 S_t(\gamma, \tau) + \nu_t, \tag{4.13}$$

where $\alpha_2 \neq 0$, $\gamma > 0$ and $S_t(\gamma, \tau)$ is the LST function defined at (4.11) above.

We denote all the parameters in model M_0 as the parameter vector $\theta_0 = (\alpha_1, \sigma)'$, all the parameters in model M_1 as the parameter vector $\theta_1 = (\alpha_1, \alpha_2, \gamma, \tau, \sigma)'$, and the sample data as $\mathbf{y} = (y_0, y_1, \ldots, y_T)'$.

Δt	γ							
$(t-\tau T)$	0.01	0.1	0.5	1.0	3.0	5.0	7.0	9.0
-8	0.48	0.31	0.018	0.0003	0	0	0	0
-7	0.4825	0.3318	0.0293	0.0009	0	0	0	0
-6	0.485	0.3543	0.0474	0.0024	0	0	0	0
-5	0.4875	0.3775	0.0758	0.0066	0	0	0	0
-4	0.49	0.4013	0.1192	0.0179	0	0	0	0
-3	0.4925	0.4255	0.1824	0.0474	0.0001	0	0	0
-2	0.495	0.4501	0.2689	0.1192	0.0024	0	0	0
-1	0.4975	0.475	0.3775	0.2689	0.0474	0.0066	0.0009	0.0001
0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5
1	0.5025	0.5249	0.6224	0.731	0.9525	0.9933	0.9991	0.9999
2	0.5049	0.5498	0.731	0.8807	0.9975	0.9999	1	1
3	0.5074	0.5744	0.8175	0.9525	0.9998	0.9999	1	1
4	0.5099	0.5986	0.8807	0.982	0.9999	1	1	1
5	0.5124	0.6224	0.9241	0.9933	0.9999	1	1	1
6	0.5149	0.6456	0.9525	0.9975	1	1	1	1
7	0.5174	0.6681	0.9706	0.999	1	1	1	1
8	0.5199	0.6899	0.982	0.9996	1	1	1	1

Table 4.1: Values of LST function around $t = \tau T$.

Given a sample $\mathbf{y}$, the Bayesian comparison of the two models again proceeds by computing the posterior model probabilities, which are given by Bayes' theorem in (2.13). For the approach we adopt here, we will again assume that the two models are equally likely *a priori*, so that $P(M_i) = 0.5$.

In order to simplify the derivation of our approach, we use matrix notation to represent the model M_1 as follows

$$M_1 : \ \mathbf{y} = X\boldsymbol{\alpha} + \mathbf{u}, \tag{4.14}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ and X is the matrix

$$X = \begin{pmatrix} 1 & S_0 \\ 1 & S_1 \\ \vdots & \vdots \\ 1 & S_T \end{pmatrix},$$

where $S_t = S_t(\gamma, \tau)$ is the LST function, $t = 0, 1, \ldots, T$. The elements in the second column of the matrix are functions of both γ and τ.

4.2 The likelihood Functions

The likelihood function for the model M_0 can be written as

$$p(\mathbf{y}|\theta_0, M_0) = \frac{1}{(2\pi)^{\frac{T+1}{2}}\sigma^{T+1}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=0}^{T}(y_i - \alpha_1)^2\right\}. \tag{4.15}$$

We then rewrite the above as

$$p(\mathbf{y}|\theta_0, M_0) = \frac{\sigma^{-T-1}}{(2\pi)^{\frac{T+1}{2}}} \exp\left\{-\frac{T+1}{2\sigma^2}\left[(\bar{y_t^2} - \bar{y}^2) + (\alpha_1 - \bar{y})^2\right]\right\}, \tag{4.16}$$

where $\bar{y} = \frac{\sum_{t=0}^{T} y_t}{T+1}$, $\bar{y_t^2} = \frac{\sum_{t=0}^{T} y_t^2}{T+1}$.

The likelihood function with respect to model M_1 can be written as follows

$$p(\mathbf{y}|\theta_1, M_1) = \frac{1}{(2\pi)^{\frac{T+1}{2}}\sigma^{T+1}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=0}^{T}(y_t - \alpha_1 - \alpha_2 S_t)^2\right\}. \tag{4.17}$$

Using the matrix formulation in (4.14), we have the likelihood function with respect to model M_1 as

$$p(\mathbf{y}|\theta_1, M_1) = \frac{1}{(2\pi)^{\frac{T+1}{2}}\sigma^{T+1}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\alpha})'(\mathbf{y} - X\boldsymbol{\alpha})\right\}. \tag{4.18}$$

We then can rewrite this as follows

$$p(\mathbf{y}|\theta_1, M_1) = \frac{1}{(2\pi)^{\frac{T+1}{2}}\sigma^{T+1}} \exp\left\{-\frac{\nu s^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' X' X (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})}{2\sigma^2}\right\}, \tag{4.19}$$

where $\nu = T - 1$, $\hat{\boldsymbol{\alpha}} = (X'X)^{-1}X'\mathbf{y}$ and $s^2 = \dfrac{(\mathbf{y} - X\hat{\boldsymbol{\alpha}})'(\mathbf{y} - X\hat{\boldsymbol{\alpha}})}{\nu}$.

4.3 Priors

We adopt conjugate priors for the parameters in both models where possible. The parameters α_1 and σ are present in both models, while α_2, γ and τ are only present in model M_1. Because of the complicated structure for γ and τ in the likelihood function for model M_1, simple conjugate priors do not exist for these parameters. We adopt the logistic distribution for γ and the uniform distribution as a prior for τ in model M_1.

4.4 Posteriors for Model M_0

We adopt a joint conjugate priors for α_1 and σ, first the conjugate prior for α_1 given σ is

$$p(\alpha_1|\sigma, M_0) = \frac{1}{\sqrt{2\pi}k_1\sigma} \exp\left\{-\frac{(\alpha_1 - c)^2}{2(k_1\sigma)^2}\right\}, \tag{4.20}$$

where $k_1 > 0$ is a constant, $E(\alpha_1|\sigma, M_0) = c$ and $\text{var}(\alpha_1|\sigma, M_0) = k_1^2\sigma^2$. As is usual, we adopt the inverted gamma distribution as the prior for σ

$$p(\sigma|M_0) = \frac{2}{\Gamma(b)a^b\sigma^{2b+1}} \exp\left\{-\frac{1}{a\sigma^2}\right\}, \quad 0 < \sigma < +\infty \tag{4.21}$$

where a and b are constants, satisfying $a, b > 0$.

The joint distribution for $(\theta_0, \mathbf{y})$ is then

$$\begin{aligned} p(\theta_0, \mathbf{y}|M_0) = {} & \frac{2\sigma^{-(T+2b+3)}}{(2\pi)^{T/2+1}k_1\Gamma(b)a^b} \exp\left\{-\frac{G}{2\sigma^2}\right\} \\ & \times \exp\left\{-\frac{T+1+1/k_1^2}{2\sigma^2}(\alpha_1 - \bar{u})^2\right\}, \end{aligned} \tag{4.22}$$

where $\bar{u} = \dfrac{c/k_1^2 + (T+1)\bar{y}}{T+1+1/k_1^2}$ and $G = (T+1)\bar{y_t^2} + c^2/k_1^2 + 2/a - (T+1+1/k_1^2)\bar{u}^2$.

If we integrate (4.22) with respect to α_1 we obtain

$$p(\sigma, \mathbf{y}|M_0) = \frac{2\sigma^{-(T+2b+2)}}{(2\pi)^{\frac{T+1}{2}}((T+1)k_1^2+1)^{\frac{1}{2}}\Gamma(b)a^b} \exp\left\{-\frac{G}{2\sigma^2}\right\}. \tag{4.23}$$

Integrating (4.23) with respect to σ then gives

$$p(\mathbf{y}|M_0) = \frac{2^b\Gamma\left(\frac{T+2b+1}{2}\right)}{\pi^{\frac{T+1}{2}}((T+1)k_1^2+1)^{\frac{1}{2}}\Gamma(b)a^b}G^{-\frac{T+2b+1}{2}}. \tag{4.24}$$

Applying Bayes' theorem, we obtain the posterior density for θ_0 as

$$p(\theta_0|\mathbf{y}, M_0) = \frac{1}{\sigma^{T+2b+3}}\exp\left\{-\frac{G+(T+1+1/k_1^2)(\alpha_1-\bar{u})^2}{2\sigma^2}\right\}. \tag{4.25}$$

If we integrate (4.25) with respect to α_1, we obtain the posterior density for σ as

$$p(\sigma|\mathbf{y}, M_0) \propto \frac{1}{\sigma^{T+2b+2}}\exp\left\{-\frac{G}{2\sigma^2}\right\}, \tag{4.26}$$

which is an inverted gamma distribution.

Or we integrate (4.25) with respect to σ, obtain the posterior density

$$p(\alpha_1|\mathbf{y}, M_0) \propto \frac{1}{[G+(T+1+1/k_1^2)(\alpha_1-\bar{u})^2]^{\frac{T+2b+2}{2}}}, \tag{4.27}$$

which is a univariate student t distribution.

4.5 Posteriors for Model M_1

Because of the complexity of the structure of the parameters γ and τ in the likelihood function for model M_1, no simple expression for a conjugate prior for (γ, τ) exists. We suppose that $(\boldsymbol{\alpha}, \sigma)$, γ, τ are *a priori* independent, then we have the joint prior for θ_1 in model M_1 as follows

$$p(\theta_1|M_1) = p(\boldsymbol{\alpha}|\sigma, M_1)p(\gamma|M_1)p(\tau|M_1)p(\sigma|M_1).$$

The conditional conjugate prior distribution for $\boldsymbol{\alpha}$, given σ, in model M_1 is the normal distribution with density function

$$p(\boldsymbol{\alpha}|\sigma, M_1) = \frac{\sqrt{|V_0^{-1}|}}{2\pi\sigma^2}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\boldsymbol{\mu}_0)'V_0^{-1}(\boldsymbol{\alpha}-\boldsymbol{\mu}_0)\right\}, \tag{4.28}$$

where V_0 is a 2×2 positive definite symmetric matrix and $\boldsymbol{\mu}_0$ is a 2-dimension vector.

We adopt the same conjugate prior for σ in model M_1 as in model M_0, which has the density function given in (4.21).

From the discussion of Table 4.1, we know that values of $\gamma > 1.0$ will give rise to very rapid changes that are likely to be difficult to distinguish from structural

breaks. Our prior for γ should reflects the fact that for gradual smooth transitions, we are most likely to see values less than 1. Even in small sample periods, if $\gamma = 1$, the smooth transition function $S_t(\gamma, \tau)$ gives from $S_{\tau T-8} = 0.0003$ to $S_{\tau T+8} = 0.9996$ within 16 samples around the midpoint τT. We choose the positive logistic distribution (See E. B. Fowlkes, 1987) as the prior for γ with density function

$$p(\gamma|M_1) = \frac{2k_2 e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}, \tag{4.29}$$

where $\gamma > 0$, $k_2 > 0$ is a constant. We plot the positive logistic density with $k_2 = 1$ in Figure 4.2. We can see from Figure 4.2 that the probability of $\{\gamma >= 6\}$ is very tiny. So we think the choice of $k_2 = 1$ gives us a suitable prior for γ which spreads out enough to cover the interesting intervals for $\gamma \in (0, 1)$, corresponding to a gradual smooth transition, and $\gamma > 1$, corresponding to an abrupt transition.

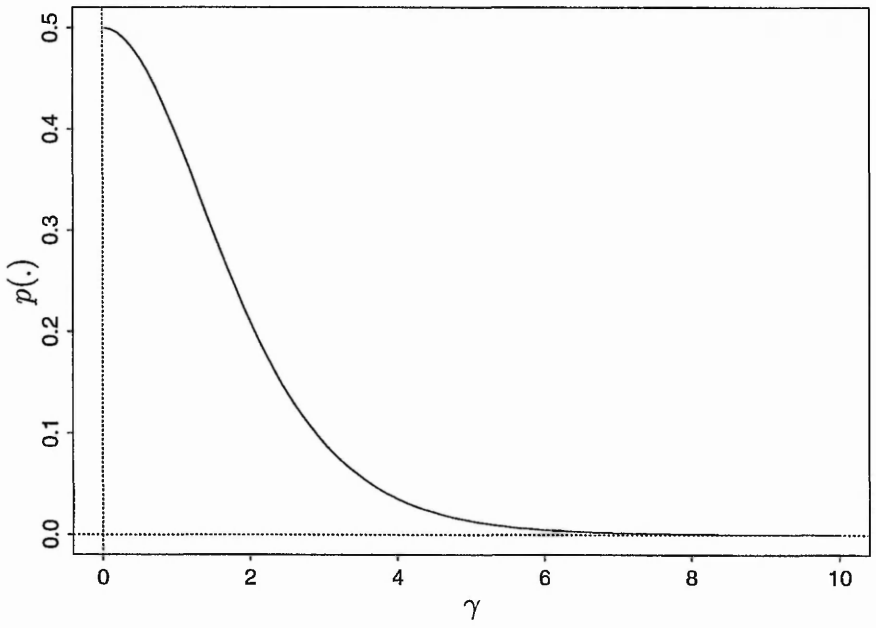


Figure 4.2: Positive logistic density function with $k_2 = 1$.

We know nothing about τ except that $\tau \in (0, 1)$, so we choose a uniform distribution as the prior for τ with density function

$$p(\tau) = 1, \tag{4.30}$$

where $\tau \in (0, 1)$.

We can now write the prior density for the parameter vector θ_1 as follows

$$\begin{aligned} p(\theta_1|M_1) = & \frac{\sqrt{|V_0^{-1}|}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\boldsymbol{\mu}_0)'V_0^{-1}(\boldsymbol{\alpha}-\boldsymbol{\mu}_0)\right\} \\ & \times \frac{2}{\Gamma(b)a^b\sigma^{2b+1}} \exp\left\{-\frac{1}{a\sigma^2}\right\} \frac{2k_2e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}, \end{aligned} \tag{4.31}$$

and so the joint distribution for $(\theta_1, \mathbf{y})$ is

$$\begin{aligned} p(\theta_1, \mathbf{y}|M_1) = & \frac{4k_2\sqrt{|V_0^{-1}|}}{(2\pi)^{\frac{T+3}{2}}\Gamma(b)a^b\sigma^{T+2b+4}} \exp\left\{-\frac{H}{2\sigma^2}\right\} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2} \\ & \times \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\hat{\boldsymbol{\mu}})'\hat{V}^{-1}(\boldsymbol{\alpha}-\hat{\boldsymbol{\mu}})\right\} \end{aligned}, \tag{4.32}$$

where $\hat{V}^{-1} = X'X + V_0^{-1}$, $\hat{\boldsymbol{\mu}} = \hat{V}(X'X\hat{\boldsymbol{\alpha}} + V_0^{-1}\boldsymbol{\mu}_0)$ and $H = (\boldsymbol{\mu}_0 - \hat{\boldsymbol{\alpha}})'X'X(\hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\alpha}}) + \nu s^2 + 2/a$.

Integrating (4.32) with respect to $\boldsymbol{\alpha}$, we obtain

$$\begin{aligned} p(\gamma, \tau, \sigma, \mathbf{y}|M_1) = & \frac{4k_2\sqrt{|V_0^{-1}|}}{(2\pi)^{\frac{T+1}{2}}\Gamma(b)a^b\sigma^{T+2b+2}} \\ & \times \frac{1}{|\hat{V}^{-1}|} \exp\left\{-\frac{H}{2\sigma^2}\right\} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}. \end{aligned} \tag{4.33}$$

Integrating (4.33) with respect to σ, we obtain

$$\begin{aligned} p(\gamma, \tau, \mathbf{y}|M_1) = & \frac{2^{b+1}\Gamma\left(\frac{T+2b+1}{2}\right)k_2\sqrt{|V_0^{-1}|}}{\pi^{\frac{T+1}{2}}\Gamma(b)a^b} \\ & \times \frac{1}{|\hat{V}^{-1}|} H^{-\frac{T+2b+1}{2}} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}. \end{aligned} \tag{4.34}$$

Finally integrating (4.34) with respect to (γ, τ) gives

$$\begin{aligned} p(\mathbf{y}|M_1) = & \frac{2^{b+1}\Gamma\left(\frac{T+2b+1}{2}\right)k_2\sqrt{|V_0^{-1}|}}{\pi^{\frac{T+1}{2}}\Gamma(b)a^b} \\ & \times \iint \frac{1}{|\hat{V}^{-1}|} H^{-\frac{T+2b+1}{2}} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2} d\gamma d\tau. \end{aligned} \tag{4.35}$$

Using Bayes' theorem, we obtain $p(\theta_1|\mathbf{y}, M_1)$, the posterior distribution for θ_1, with respect to model M_1, as

$$\begin{aligned} p(\theta_1|\mathbf{y}, M_1) & \propto p(\mathbf{y}|\theta_1, M_1)p(\theta_1|M_1) \\ & \propto \frac{1}{\sigma^{T+2b+4}} \exp\left\{-\frac{H}{2\sigma^2}\right\} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2} \\ & \quad \times \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\hat{\boldsymbol{\mu}})'\hat{V}^{-1}(\boldsymbol{\alpha}-\hat{\boldsymbol{\mu}})\right\} \end{aligned}. \tag{4.36}$$

Integrating (4.36) with respect to $\boldsymbol{\alpha}$ gives the marginal joint posterior distribution for γ, τ, σ as

$$\begin{aligned} p(\gamma, \tau, \sigma|\mathbf{y}, M_1) &= \iint p(\theta_1|\mathbf{y}, M_1)\, d\boldsymbol{\alpha} \\ &\propto \frac{|\hat{V}|}{\sigma^{T+2b+2}} \exp\left\{-\frac{H}{2\sigma^2}\right\} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}. \end{aligned} \tag{4.37}$$

Integrating again with respect to σ gives the marginal posterior distribution for γ, τ as

$$\begin{aligned} p(\gamma, \tau|\mathbf{y}, M_1) &= \int_0^{+\infty} p(\gamma, \tau, \sigma|\mathbf{y}, M_1)\, d\sigma \\ &\propto |\hat{V}| H^{-\frac{T+2b+1}{2}} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}. \end{aligned} \tag{4.38}$$

We can then integrate (4.38) with respect to either γ or τ to obtain the marginal posterior distribution for τ or γ respectively,

$$p(\gamma|\mathbf{y}, M_1) \propto \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2} \int_0^1 |\hat{V}| H^{-\frac{T+2b+1}{2}}\, d\tau, \tag{4.39}$$

and

$$p(\tau|\mathbf{y}, M_1) \propto \int_{-\infty}^{\infty} |\hat{V}| H^{-\frac{T+2b+1}{2}} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2}\, d\gamma. \tag{4.40}$$

The conditional posterior distribution of $\boldsymbol{\alpha}$ given (γ, τ, σ) from $p(\theta_1|\mathbf{y})$ has density

$$p(\boldsymbol{\alpha}|\gamma, \tau, \sigma, \mathbf{y}, M_1) = \frac{\sqrt{|\hat{V}^{-1}|}}{2\pi\sigma^2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha} - \hat{\boldsymbol{\mu}})'\hat{V}^{-1}(\boldsymbol{\alpha} - \hat{\boldsymbol{\mu}})\right\}, \tag{4.41}$$

which is a multivariate normal distribution with mean $\hat{\boldsymbol{\mu}}$ and precision matrix $\hat{V}^{-1}/\sigma^2$. Here $\hat{\boldsymbol{\mu}}$ and $\hat{V}^{-1}/\sigma^2$ are functions of γ and τ which are unknown.

4.6 Posterior Model Probabilities

We set equal probability as priors for both models, M_0 and M_1, using (2.13), we obtain the posterior probability for model M_0 as

$$P(M_0|\mathbf{y}) = \frac{1}{1+k}, \tag{4.42}$$

and the posterior probability for model M_1 is

$$P(M_1|\mathbf{y}) = \frac{k}{1+k}, \tag{4.43}$$

where

$$k = 2((T+1)k_1^2+1)^{\frac{1}{2}} G^{\frac{T+2b+1}{2}} k_2 \sqrt{|V_0^{-1}|} \times \int_0^1 \int_0^\infty \frac{1}{|\hat{V}^{-1}|} H^{-\frac{T+2b+1}{2}} \frac{e^{-k_2\gamma}}{(1+e^{-k_2\gamma})^2} d\gamma d\tau. \tag{4.44}$$

.

4.7 Simulations

Now we use simulated data to illustrate the above approach to choosing between model M_0 and M_1. We choose the following prior parameters: $k_1 = 1$, $V_0 = \begin{Bmatrix} 1 & 0 \\ 0 & 1 \end{Bmatrix}$, $a = 1$, $b = 2$, $k_2 = 1$.

In order to calculate the posterior probability for model M_0, we need to calculate the value of k in (4.44), which involves integrating a function of two variables over the area $[0,1] \times (0, +\infty)$. We use Gauss-Laguerre rules (See Griffiths, 1993) to calculate the inner integral with respect to γ over the infinite interval $(0, +\infty)$. Then we use Gauss-Legendre rules to calculate the outer integral with respect to τ over the finite interval $[0,1]$.

Leybourne, Newbold and Vougas (1998) generated series from the model

$$y_t = 1.0 + 10.0 S_t(\gamma, \tau) + \mu_t \qquad \mu_t = 0.8\mu_{t-1} + \varepsilon_t \qquad \varepsilon_t \sim \text{NID}(0,1),$$

for various values of the speed of transition parameter $\gamma = 0.01, 0.1, 0.5, 1, 5$, with the transition midpoint fraction τ set at 0.2 and 0.5 and considering sample size 100, 200. They obtained results based on 2000 replications.

We follow them to set $\alpha_1 = 1.0$. $\alpha_2 = 10.0$ is too significant for our algorithm to choose between model M_0 and M_1 so we choose $\alpha_2 = 3$ arbitrarily smaller than 10.0. We have only discussed NID$(0, \sigma)$ error term, we cannot apply with an AR(1) error. The standard deviation of μ_t in their AR(1) error is $5/3$, so we choose $\sigma = 5/3$ for our simulations. We select more values for $\gamma = 0.05, 0.5, 1, 2, 3, 4$, $\tau = 0.8$ and sample size 50. So we generate series from the model

$$y_t = 1.0 + 3.0 S_t(\gamma, \tau) + \varepsilon_t \qquad \varepsilon_t \sim \text{NID}(0, 5/3). \tag{4.45}$$

for various values of γ, τ and sample size $T+1$. We obtain the proportions of the

time we would have favoured model M_0, based on 10,000 replications and list them in Table 4.2, Part I.

$T+1$	50			100			200		
τ	0.2	0.5	0.8	0.2	0.5	0.8	0.2	0.5	0.8
γ	Part I								
0.01	0.9955	0.9963	0.9969	0.9707	0.9726	0.9773	0.4747	0.4108	0.5375
0.05	0.8912	0.8833	0.9362	0.2199	0.0407	0.3396	0.0005	0	0.0022
0.1	0.6476	0.4365	0.7955	0.0807	0.0007	0.1658	0.0001	0	0.0002
0.5	0.429	0.0381	0.6381	0.0308	0	0.0769	0	0	0
1	0.425	0.0375	0.6332	0.0304	0	0.0763	0	0	0
2	0.432	0.0395	0.6394	0.0313	0	0.0781	0	0	0.0001
3	0.4359	0.0404	0.6424	0.0319	0	0.0789	0	0	0.0001
4	0.4374	0.0408	0.6448	0.0323	0	0.0796	0	0	0.0001
5	0.4384	0.041	0.6452	0.0323	0	0.08	0	0	0.0001
α_2	Part II								
0	0.9998	0.9998	0.9998	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996
0.5	0.9985	0.9981	0.9988	0.9971	0.9902	0.9973	0.9896	0.9515	0.9908
1	0.993	0.9888	0.9952	0.9704	0.8802	0.9787	0.8635	0.4582	0.8861
1.5	0.976	0.9554	0.9844	0.8539	0.5027	0.8977	0.4618	0.0284	0.529
2	0.9288	0.8621	0.9597	0.5932	0.1261	0.6955	0.0909	0.0001	0.1342
2.5	0.825	0.6695	0.8985	0.2816	0.0104	0.4126	0.0043	0	0.0104
3	0.6476	0.4365	0.7955	0.0807	0.0007	0.1658	0.0001	0	0.0002

Table 4.2: Proportion of $P(M_0|\mathbf{y}) > 0.5$ in 10,000 replications.

In Table 4.2, Part I, data series come from model M_1 so the small values of the proportion of the time we favoured model M_0 in the table indicate that M_1 has been successfully detected. We have three groups of results, corresponding to sample size $T+1 = 50, 100, 200$ respectively. In each group, we have various values for γ and τ. We can see, in each group, that when γ increases, the proportion of favouring model M_0 decreases. That is reasonable because when γ increases, the smooth transition component $\alpha_2 S_t(\gamma, \tau)$ becomes more significant and easier to be "caught" by our algorithm. For the same γ, when $\tau = 0.5$, our algorithm performs better because the smooth transition happens in the middle of the measuring time

interval $[0, T]$ and is easier to be "caught" than it happens in the left edge or right edge of the measuring time interval. Different groups reveal that when the sample size increases, the performance of our algorithm becomes better. We also found in the first group, $T+1 = 50$, when $\tau = 0.2$ or $\tau = 0.8$, no improvement for the results when γ becomes bigger from 1 to 5. This, however, is not unexpected because when γ is bigger than 1, the smooth transition becomes abrupt and sharp, which was carried out in a short period of time. See Table 4.1 and Figure 4.1. No significant change of the smooth transition happens in the data series when γ is great than 1.

α_2 is the magnitude of the smooth transition. We also look at the effect of α_2 by changing α_2 to various values and fixing γ. We choose $\alpha_2 = 0, 0.5, 1, 1.5, 2, 2.5, 3$ and fix γ at an arbitrary value, for example, at $\gamma = 0.1$. We obtain the proportions of the time we would have favoured model M_0 and list them in Table 4.2, Part II. Except $\alpha_2 = 0$, all the data series come from model M_1. When $\alpha_2 = 0$, the data comes from model M_0, we found the proportions of the time we favoured model M_0 are near 1. The difference between different groups should contribute to the generation of different random numbers for different groups. We set the same seed to generate random numbers for each experiment, so the proportions within group are the same while between groups, the sample size changed, the proportions are different in the last digit. From $\gamma = 0.5$ to $\gamma = 3$, all the data series come from model M_1. The performance of our algorithm improved when α_2 becomes bigger and when the sample size becomes bigger. Within each group, when $\tau = 0.5$, we obtained the best results compared with $\tau = 0.2$ and $\tau = 0.8$.

Chapter 5

Smooth Transition Model II

In this chapter, we will extend the smooth transition model discussed in detail in Chapter 4 by introducing additional terms into the smooth transition model. We will discuss the third smooth transition model in Leybourne, Newbold and Vougas (1998), that is, model C, in (4.10), with an AR(1) disturbance term to illustrate how the model can be extend to allow for autocorrelated errors. The smooth transition model we are going to discuss in this chapter is

$$\begin{aligned} y_t &= \alpha_1 + \beta_1 t + \alpha_2 S_t(\gamma, \tau) + \beta_2 t S_t(\gamma, \tau) + u_t, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \end{aligned} \tag{5.1}$$

where $\rho \in (-1, 1)$, ε_t is *iid* N(0, σ^2), σ is unknown constant, $t = 0, 1, \ldots, T$. The sample size is $T + 1$. $S_t(\gamma, \tau)$ is the logistic smooth transition (LST) function in (4.11), based on the sample of size $T + 1$, where $\tau \in (0, 1)$ and $\gamma > 0$. Now we have two trend terms, one is the linear term, $\alpha_1 + \beta_1 t$, which describes the linear change in the trend; the other is the linear logistic smooth transition term, $\alpha_2 S_t(\gamma, \tau) + \beta_2 t S_t(\gamma, \tau)$.

We compare the smooth transition model in (5.1) with linear regression model,

$$\begin{aligned} M_0: \quad y_t &= \alpha_1 + \beta_1 t + u_t, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \end{aligned} \tag{5.2}$$

where $\rho \in (-1, 1)$, ε_t is *iid* N(0, σ^2), correspondingly, we call the smooth transition model in (5.1) model M_1.

5.1 The likelihood functions

In order to use Bayesian methods to compare the two models and obtain the posterior densities for the parameters of model M_1, we first need to obtain the likelihood functions for each model. We start with the likelihood function for model M_1 and will then obtain the likelihood function for model M_0.

By eliminating u_t, u_{t-1} from model M_1, we obtain

$$\begin{aligned} M_1: \quad y_t - \rho y_{t-1} &= (1-\rho)\alpha_1 + ((1-\rho)t + \rho)\beta_1 \\ &\quad +(S_t - \rho S_{t-1})\alpha_2 + (tS_t - \rho(t-1)S_{t-1})\beta_2 + \varepsilon_t, \end{aligned} \tag{5.3}$$

where $S_t = S_t(\gamma, \tau)$ is an abbreviation for the smooth transition function. Notice that when $t = 0$, a new symbol y_{-1} appears in model M_1. y_{-1} is the initial state of the model. As we saw in Chapter 3, $u_t \sim \mathrm{N}(0, \frac{\sigma^2}{1-\rho^2})$, $t = 0, \pm 1, \pm 2, \ldots$, in particular we have $u_{-1} \sim \mathrm{N}(0, \frac{\sigma^2}{1-\rho^2})$. We denote $\tilde{u}_{-1} = \sqrt{1-\rho^2}u_{-1}$, then we have $\tilde{u}_{-1} \sim \mathrm{N}(0, \sigma^2)$, so that $(\tilde{u}_{-1}, \varepsilon_0, \ldots, \varepsilon_T)$ are *iid* $\mathrm{N}(0, \sigma^2)$.

We set $\theta_1 = (\alpha_1, \beta_1, \alpha_2, \beta_2, \gamma, \tau, \rho, \sigma)'$, the parameter vector in model M_1. In order to use matrix notation to describe the model, we denote $x_{-1} = \sqrt{1-\rho^2}y_{-1}$, that is, $x_{-1} = \sqrt{1-\rho^2}\alpha_1 - \sqrt{1-\rho^2}\beta_1 + \sqrt{1-\rho^2}S_{-1}\alpha_2 - \sqrt{1-\rho^2}S_{-1}\beta_2 + \tilde{u}_{-1}$, $x_t = y_t - \rho y_{t-1}$, $t = 0, 1, \ldots, T$, we transform $(y_{-1}, y_0, \ldots, y_T)$ to $(x_{-1}, x_0, \ldots, x_T)$. If we now write $\mathbf{x} = (x_{-1}, x_0, \ldots, x_T)'$, $\boldsymbol{\varepsilon} = (\tilde{u}_{-1}, \varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T)'$, $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)$ and

$$X = \begin{pmatrix} \sqrt{1-\rho^2} & -\sqrt{1-\rho^2} & \sqrt{1-\rho^2}S_{-1} & -\sqrt{1-\rho^2}S_{-1} \\ 1-\rho & \rho & S_0 - \rho S_{-1} & \rho S_{-1} \\ 1-\rho & 1 & S_1 - \rho S_0 & S_1 \\ 1-\rho & 2-\rho & S_2 - \rho S_1 & 2S_2 - \rho S_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1-\rho & (1-\rho)T + \rho & S_T - \rho S_{T-1} & TS_T - \rho(T-1)S_{T-1} \end{pmatrix},$$

we have model M_1 as

$$M_1: \qquad \mathbf{x} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{5.4}$$

Notice that $\mathbf{x}$ is unobservable, x_{-1}, x_0 are functions of y_{-1}, y_0 and ρ, which are all unknown. x_t, $t = 1, 2, \ldots, T$, are functions of ρ. When $\rho = 1$, the transformation from $(y_{-1}, y_0, \ldots, y_T)$ to $(x_{-1}, x_0, \ldots, x_T)$ becomes identical and the disturbance terms in

both models become *iid* $\mathrm{N}(0, \sigma^2)$. The Jacobian of the transformation $\mathbf{y} \longmapsto \mathbf{x}$ is

$$J = \left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| = \begin{vmatrix} \sqrt{1-\rho^2} & 0 & & & \\ -\rho & 1 & & & \\ & -\rho & 1 & & \\ & & \ddots & \ddots & \\ & & & -\rho & 1 \end{vmatrix} = \sqrt{1-\rho^2}.$$

The joint distribution of $\mathbf{x} = (x_{-1}, x_0, \ldots, x_T)$, can be written as

$$p(\mathbf{x}|\theta, M_1) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T+2} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}, \tag{5.5}$$

where $\nu = (T+2) - 4 = T - 2$, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{x}$, and $\nu s^2 = (\mathbf{x} - X\hat{\boldsymbol{\beta}})'(\mathbf{x} - X\hat{\boldsymbol{\beta}})$.

The likelihood function for model M_1 is then given by

$$\begin{aligned} p(\mathbf{y}|\theta, M_1) &= p(\mathbf{x}|\theta, M_1)|J| \\ &= \sqrt{1-\rho^2}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T+2} \times \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}. \end{aligned} \tag{5.6}$$

Similarly, we can obtain the likelihood function for model M_0 because model M_0 is the simple case of model M_1 when $\alpha_2 = \beta_2 = 0$.

By eliminating u_t and u_{t-1} from model M_0, we obtain

$$M_0: \quad y_t - \rho y_{t-1} = (1-\rho)\alpha_1 + ((1-\rho)t + \rho)\beta_1 + \varepsilon_t. \tag{5.7}$$

If we denote the parameter vector in model M_0 as $\theta_0 = (\alpha_1, \alpha_2, \rho, \sigma)'$ and

$$X_0 = \begin{pmatrix} \sqrt{1-\rho} & -\sqrt{1-\rho} \\ 1-\rho & \rho \\ 1-\rho & 1 \\ 1-\rho & 2-\rho \\ \vdots & \vdots \\ 1-\rho & (1-\rho)T + \rho \end{pmatrix},$$

we can rewrite model M_0 as follows

$$M_0: \qquad \mathbf{x} = X_0\boldsymbol{\alpha} + \boldsymbol{\varepsilon}, \tag{5.8}$$

where $\boldsymbol{\alpha} = (\alpha_1, \beta_1)'$.

The joint distribution of $\mathbf{x}$ in model M_0 is given as

$$p(\mathbf{x}|\theta_0, M_0) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T+2} \exp\left\{-\frac{\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})}{2\sigma^2}\right\}, \tag{5.9}$$

where $\nu_0 = (T+2) - 2 = T$, $\hat{\boldsymbol{\alpha}} = (X_0'X_0)^{-1}X_0'\mathbf{x}$, and $\nu_0 s_0^2 = (\mathbf{x} - X_0\hat{\boldsymbol{\alpha}})'(\mathbf{x} - X_0\hat{\boldsymbol{\alpha}})$.

The likelihood function for model M_0 is then given by

$$\begin{aligned} p(\mathbf{y}|\theta_0, M_0) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T+2} & \sqrt{1-\rho^2} \times \\ & \exp\left\{-\frac{\nu_0 s_0^2 + (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})' X_0' X_0 (\boldsymbol{\alpha} - \hat{\boldsymbol{\alpha}})}{2\sigma^2}\right\}, \end{aligned} \tag{5.10}$$

5.2 Priors and Posteriors for model M_1

We select priors for the parameters in model M_1. We choose a conjugate prior for $\boldsymbol{\beta}$ conditional on σ in model M_1 as follows

$$p(\boldsymbol{\beta}|\sigma, M_1) = \frac{|V|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^4} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^0)' V^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}^0)\right\}, \tag{5.11}$$

where V is a 4×4 positive definite symmetric constant matrix, $\boldsymbol{\beta}^0$ is a 4 dimensional constant vector. The conjugate prior for σ in model M_1 is an inverted gamma distribution with density

$$p(\sigma|M_1) = \frac{2}{\Gamma(b/2)} \left(\frac{a}{2}\right)^{b/2} \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\}, \quad 0 < \sigma < +\infty, \tag{5.12}$$

where $a, b > 0$ are constants.

As for γ and τ, we choose the same prior distributions that were used in Chapter 4. These are given by

$$p(\gamma|M_1) = \frac{2ke^{-k\gamma}}{(1+e^{-k\gamma})^2}, \tag{5.13}$$

where $k > 0$, $\gamma > 0$ and

$$p(\tau|M_1) = \begin{cases} 1, & \tau \in (0,1); \\ 0, & \text{otherwise.} \end{cases} \tag{5.14}$$

We know $\rho \in (-1, 1)$, which corresponds to the stationary case, so we choose a uniform distribution over (-1, 1) for ρ which has density

$$p(\rho|M_1) = \begin{cases} 1/2, & \tau \in (-1,1); \\ 0, & \text{otherwise.} \end{cases} \tag{5.15}$$

We suppose that $(\boldsymbol{\beta}, \sigma)$, γ, τ and ρ are *a priori* independently distributed, so that we have the joint prior density for θ_1 as

$$\begin{aligned} p(\theta_1|M_1) = & \frac{2^{1-b/2}|V|^{-1/2}a^{b/2}}{(2\pi)^2\Gamma(b/2)}\times \\ & \frac{1}{\sigma^{b+5}}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)'V^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)\right\}\times \\ & \exp\left\{-\frac{a}{2\sigma^2}\right\}\frac{2kTe^{-kT\gamma}}{(1+e^{-kT\gamma})^2}. \end{aligned} \tag{5.16}$$

Now we can obtain the joint distribution for $(\mathbf{y}, \theta_1)$ in model M_1 as

$$\begin{aligned} p(\mathbf{y}, \theta_1|M_1) = & \frac{2^{1-b/2}|V|^{-1/2}a^{b/2}}{(2\pi)^{\frac{T+6}{2}}\Gamma(b/2)}\times \\ & \frac{\sqrt{1-\rho^2}}{\sigma^{T+b+7}}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}^1)'V_1(\boldsymbol{\beta}-\boldsymbol{\beta}^1)\right\}\times \\ & \exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\}\frac{2kTe^{-kT\gamma}}{(1+e^{-kT\gamma})^2}, \end{aligned} \tag{5.17}$$

where $V_1 = X'X + V^{-1}$, $\boldsymbol{\beta}^1 = V_1^{-1}(X'X\hat{\boldsymbol{\beta}} + V^{-1}\boldsymbol{\beta}^0)$, and $\nu s_1^2 = a + \nu s^2 + (\boldsymbol{\beta}^1 - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})$.

If we integrate (5.17) with respect to $\boldsymbol{\beta}$, we obtain

$$\begin{aligned} p(\mathbf{y}, \gamma, \tau, \rho, \sigma|M_1) = & \frac{2^{1-b/2}|V|^{-1/2}a^{b/2}}{(2\pi)^{\frac{T+2}{2}}\Gamma(b/2)|V_1|^{1/2}}\times \\ & \frac{\sqrt{1-\rho^2}}{\sigma^{T+b+3}}\exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\}\frac{2kTe^{-kT\gamma}}{(1+e^{-kT\gamma})^2}. \end{aligned} \tag{5.18}$$

Integrating this with respect to σ, we obtain

$$\begin{aligned} p(\mathbf{y}, \gamma, \tau, \rho|M_1) = & \frac{2^{\frac{T+2}{2}}|V|^{-1/2}\Gamma\left(\frac{T+b+2}{2}\right)a^{b/2}}{(2\pi)^{\frac{T+2}{2}}\Gamma(b/2)|V_1|^{1/2}}\times \\ & \frac{\sqrt{1-\rho^2}}{(\nu s_1^2)^{\frac{T+b+2}{2}}}\frac{2kTe^{-kT\gamma}}{(1+e^{-kT\gamma})^2}. \end{aligned} \tag{5.19}$$

If we now integrate this with respect to γ, τ, ρ, we obtain

$$\begin{aligned} p(\mathbf{y}|M_1) = & \frac{2^{\frac{T+2}{2}}|V|^{-1/2}\Gamma\left(\frac{T+b+2}{2}\right)a^{b/2}}{(2\pi)^{\frac{T+2}{2}}\Gamma(b/2)}\times \\ & \iiint\frac{\sqrt{1-\rho^2}}{|V_1|^{1/2}(\nu s_1^2)^{\frac{T+b+2}{2}}}\frac{2kTe^{-kT\gamma}}{(1+e^{-kT\gamma})^2}\,d\gamma d\tau d\rho. \end{aligned} \tag{5.20}$$

This will be used in the model comparison later.

5.3 Priors and Posteriors for model M_0

Now we select priors for parameters in model M_0. We select a conjugate prior for $\boldsymbol{\alpha}$ conditional on σ in model M_0 as

$$p(\boldsymbol{\alpha}|\sigma, M_0) = \frac{|V_0|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^0)'V_0^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^0)\right\}, \tag{5.21}$$

where V_0 is a 2×2 positive definite symmetric constant matrix, $\boldsymbol{\alpha}^0$ is a 2 dimensional constant vector. For simplicity, we take V_0 to be sub-matrix of V, where $V = \begin{pmatrix} V_0 & V_{12} \\ V_{12}' & V_{22} \end{pmatrix}$ is the matrix in (5.11).

We choose the same priors for ρ and σ as in model M_1, that is,

$$p(\rho|M_0) = \begin{cases} 1/2, & \rho \in (-1,1); \\ 0, & \text{otherwise}, \end{cases} \tag{5.22}$$

$$p(\tau|M_0) = \begin{cases} 1, & \tau \in (0,1); \\ 0, & \text{otherwise}, \end{cases} \tag{5.23}$$

and

$$p(\sigma|M_0) = \frac{2}{\Gamma(b/2)}\left(\frac{a}{2}\right)^{b/2}\frac{1}{\sigma^{b+1}}\exp\left\{-\frac{a}{2\sigma^2}\right\}, \quad 0<\sigma<+\infty, \tag{5.24}$$

where $a, b > 0$ are the same constants as in model M_1.

We suppose that $(\boldsymbol{\alpha}, \sigma)$ and ρ are *a priori* independent. With these priors for parameters in model M_0 and the independence assumptions, we can obtain the prior for the parameter vector in model M_0 as

$$\begin{aligned} p(\theta_0|M_0) = {} & \frac{2^{1-b/2}a^{b/2}|V_0|^{-1/2}}{2\pi\Gamma(b/2)}\frac{1}{\sigma^{b+3}}\exp\left\{-\frac{a}{2\sigma^2}\right\}\times \\ & \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^0)'V_0^{-1}(\boldsymbol{\alpha}-\boldsymbol{\alpha}^0)\right\}. \end{aligned} \tag{5.25}$$

Then we can obtain the joint distribution for $(\mathbf{y}, \theta_0)$ in model M_0 as follows

$$\begin{aligned} p(\mathbf{y}, \theta_0|M_0) = {} & \frac{2^{1-b/2}a^{b/2}|V_0|^{-1/2}}{(2\pi)^{\frac{T+4}{2}}\Gamma(b/2)}\frac{\sqrt{1-\rho^2}}{\sigma^{T+b+5}}\exp\left\{-\frac{\nu_0 s_{01}^2}{2\sigma^2}\right\}\times \\ & \exp\left\{-\frac{1}{2\sigma^2}\left[(\boldsymbol{\alpha}-\boldsymbol{\alpha}_1)'V_{01}(\boldsymbol{\alpha}-\boldsymbol{\alpha}_1)\right]\right\}, \end{aligned} \tag{5.26}$$

where $V_{01} = X_0'X_0 + V_0^{-1}$, $\boldsymbol{\alpha}_1 = V_{01}^{-1}(X_0'X_0\hat{\boldsymbol{\alpha}} + V_0^{-1}\boldsymbol{\alpha}^0)$, $\nu_0 s_{01}^2 = a + \nu_0 s_0^2 + (\boldsymbol{\alpha}_1 - \hat{\boldsymbol{\alpha}})'X_0'X_0(\boldsymbol{\alpha}^0 - \hat{\boldsymbol{\alpha}})$.

If we integrate this with respect to $\boldsymbol{\alpha}$, we obtain

$$p(\mathbf{y}, \rho, \sigma|M_0) = \frac{2^{1-b/2}a^{b/2}|V_0|^{-1/2}}{(2\pi)^{\frac{T+2}{2}}\Gamma(b/2)|V_{01}|^{1/2}} \times \frac{\sqrt{1-\rho^2}}{\sigma^{T+b+3}} \exp\left\{-\frac{\nu_0 s_{01}^2}{2\sigma^2}\right\}. \tag{5.27}$$

Integrate (5.27) with respect to σ, we obtain

$$p(\mathbf{y}, \rho|M_0) = \frac{2^{\frac{T+2}{2}}a^{b/2}|V_0|^{-1/2}\Gamma\left(\frac{T+b+2}{2}\right)}{(2\pi)^{\frac{T+2}{2}}\Gamma(b/2)|V_{01}|^{1/2}} \frac{\sqrt{1-\rho^2}}{(\nu_0 s_{01}^2)^{\frac{T+b+2}{2}}}. \tag{5.28}$$

If we integrate this with respect to ρ, we obtain

$$p(\mathbf{y}|M_0) = \frac{2^{\frac{T+2}{2}}a^{b/2}|V_0|^{-1/2}\Gamma\left(\frac{T+b+2}{2}\right)}{(2\pi)^{\frac{T+2}{2}}\Gamma(b/2)} \int \frac{\sqrt{1-\rho^2}}{|V_{01}|^{1/2}(\nu_0 s_{01}^2)^{\frac{T+b+2}{2}}} d\rho, \tag{5.29}$$

which can be used in the model comparison.

5.4 Posterior model probabilities

We assume that the investigator does not know which model the sample data is from. So we set equal prior probabilities for the two models, $P(M_0) = P(M_1) = 0.5$. Using (2.13) on page 9, we can obtain the posterior model probabilities as follows

$$\begin{aligned} P(M_0|\mathbf{x}) &= \frac{1}{1+k}, \\ P(M_1|\mathbf{x}) &= \frac{k}{1+k}, \end{aligned} \tag{5.30}$$

where

$$k = \frac{|V|^{-1/2}}{|V_0|^{-1/2}} \frac{\iiint \frac{\sqrt{1-\rho^2}}{|V_1|^{1/2}(\nu s_1^2)^{\frac{T+b+2}{2}}} \frac{2kTe^{-kT\gamma}}{(1+e^{-kT\gamma})^2} d\gamma d\tau d\rho}{\int \frac{\sqrt{1-\rho^2}}{|V_{01}|^{1/2}(\nu_0 s_{01}^2)^{\frac{T+b+2}{2}}} d\rho}.$$

In order to evaluate (5.30) in any particular case, we need to complete the triple-integral which is not straightforward using numerical methods especially when the integrand is typically concentrated in a small region of $\mathbb{R}^3$. For example, in above integral, the range for γ is infinity, $\gamma \in (0, \infty)$. We can expect that the task of this kind of calculation of (5.30) is expensive in time for computer. In the next Chapter, we use a kind of new technology, MCMC, to avoid above analytic integral. So we just leave it and carry on to the next chapter.

Chapter 6

Smooth Transition and MCMC

Research on economic time series has always attracted a lot of interests from economists and statisticians. Economists believe that trends and cycles exist in economic time series. Crafts, Leybourne and Mills (1989) and Newbold and Agiakloglou (1991) discussed trends (evolving growth rates) and cycles in British industrial production, from year 1700 to year 1913. Recently, some economists began to consider the trends in some economic time series to be nonlinear and smooth without abrupt changes. Teräsvirta and Anderson (1992), Teräsvirta (1994), Greenaway, Leybourne and Sapsford (1997) used smooth transition models to describe model growth and structural changes. Leybourne, Newbold and Vougas (1998), Sollis, Leybourne and Newbold (1999) tested the null hypothesis unit roots against the alternative of a smooth transition and used the smooth transition to fit the economic time series.

In this chapter, we use the smooth transition model discussed in Leybourne, Newbold and Vougas (1998) to fit two economic time series. We use a Bayesian approach, that is, we make references through the posterior densities. We use Markov chain Monte Carlo (MCMC) methods to sample from the posterior densities and use the samples to make inferences for the parameters in the model. In this way, we avoid complex analytic integrals which we met in the previous chapter and evaluate by numerical Monte Carlo. In particular, we use the Markov chain Monte Carlo method, called slice sampler, to carry out sampling from the posterior densities.

6.1 Bayesian Models

Here we adopt the smooth transition model discussed in Leybourne, Newbold and Vougas (1998). We set up the priors for the model and obtain the posterior for the parameters of the model.

6.1.1 Smooth transition model

In Leybourne, Newbold and Vougas (1998), their alternative consists of the three smooth transition models with stationary disturbances given in (4.10). For simplicity, we only discuss the third one with an AR(1) process for the disturbance term. Suppose we obtain sample data, $\mathbf{y} = (y_0, y_1, \ldots, y_T)'$, from the third smooth transition model

$$\begin{aligned} y_t &= \alpha_1 + \beta_1 t + \alpha_2 S_t(\gamma, \tau) + \beta_2 t S_t(\gamma, \tau) + u_t, \\ u_t &= \rho u_{t-1} + \varepsilon_t, \end{aligned} \tag{6.1}$$

where $t = 0, 1, \ldots, T$, $\rho \in (-1, 1)$, ε_t is *iid* $\mathrm{N}(0, \sigma^2)$ and σ is an unknown positive constant. The sample size is $T + 1$. $S_t(\gamma, \tau)$ is the logistic smooth transition (LST) function,

$$S_t(\gamma, \tau) = \frac{1}{1 + \exp\{-\gamma(t - \tau T)\}}, \tag{6.2}$$

where $\tau \in (0, 1)$, $\gamma > 0$. This smooth transition model can be used to describe both increasing and decreasing behaviour in observed time series. See discussion in Chapter 4 and Chapter 5. The interpretation of the parameters of the smooth transition function $S_t(\gamma, \tau)$ was discussed in Chapter 4.

6.1.2 Likelihood

In order to use Bayes theorem to make references about the model, we need to obtain the likelihood function for the model. We follow Zellner (1987), pp.87 and suppose the process represented by (6.1) has been operative for $t = 0, -1, -2, \ldots, -T_0$, where T_0 is unknown. We suppose $T_0 > 1$ in order to obtain distribution of u_0 in the following.

From the stationarity of u_t, we have $\mathrm{var}(u_t) = \mathrm{var}(u_{t-1})$. Because u_{t-1} is independent of ε_t, if we calculate the variance of the two sides of the AR(1) equation,

we obtain $\text{var}(\varepsilon_t) = \text{var}(\varepsilon_{t-1}) = \sigma^2/(1-\rho^2)$. Easily, we have $E(u_t) = E(u_{t-1}) = 0$. Because $\{u_t\}$ is a Gaussian process, the distribution of u_t is normal, determined by its mean and variance only, so we have

$$u_t \sim \text{N}\left(0, \frac{\sigma^2}{1-\rho^2}\right), \tag{6.3}$$

in particular, we have $u_0 \sim \text{N}\left(0, \frac{\sigma^2}{1-\rho^2}\right)$. If we denote $\tilde{u}_0 = \sqrt{1-\rho^2}u_0$, then we have $\tilde{u}_0 \sim \text{N}(0, \sigma^2)$, u_0 is independent of ε_t, $t > 0$, so that $(\tilde{u}_0, \varepsilon_0, \ldots, \varepsilon_T)$ are *iid* $\text{N}(0, \sigma^2)$.

By eliminating u_t, u_{t-1} from the model, we obtain

$$\begin{aligned} y_t - \rho y_{t-1} = {} & (1-\rho)\alpha_1 + ((1-\rho)t + \rho)\beta_1 \\ & + (S_t - \rho S_{t-1})\alpha_2 + (tS_t - \rho(t-1)S_{t-1})\beta_2 + \varepsilon_t, \end{aligned} \tag{6.4}$$

where $S_t = S_t(\gamma, \tau)$ is an abbreviation for the smooth transition function, for $t = 1, 2, \ldots, T$. We now construct a new series, denoted $x_t = y_t - \rho y_{t-1}$, for $t = 1, 2, \ldots, T$. If we denote $x_0 = \sqrt{1-\rho^2}y_0$, $\mathbf{x} = (x_0, x_1, \ldots, x_T)'$, then we obtain a transformation from $\mathbf{y}$ to $\mathbf{x}$, which has the Jacobian

$$J = \left|\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right| = \begin{vmatrix} \sqrt{1-\rho^2} & & & \\ -\rho & 1 & & \\ & & \cdots & \\ & & -\rho & 1 \end{vmatrix} = \sqrt{1-\rho^2}. \tag{6.5}$$

The relationship between the two joint distributions for $\mathbf{x}$ and $\mathbf{y}$ is $p(\mathbf{y}) = |J|p(\mathbf{x})$.

For convenience, we denote the parameter vector as $\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2, \gamma, \tau, \rho, \sigma)'$, and we denote $\boldsymbol{\varepsilon} = (\tilde{u}_0, \varepsilon_1, \ldots, \varepsilon_T)'$, which is *i.i.d.* $\text{N}(0, \sigma^2)$, $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2)$ and

$$X = \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \sqrt{1-\rho^2}S_0 & 0 \\ 1-\rho & 1 & S_1 - \rho S_0 & S_1 \\ 1-\rho & 2-\rho & S_2 - \rho S_1 & 2S_2 - \rho S_1 \\ \vdots & \vdots & \vdots & \vdots \\ 1-\rho & (1-\rho)T + \rho & S_T - \rho S_{T-1} & TS_T - \rho(T-1)S_{T-1} \end{pmatrix}.$$

We have $x_0 = \sqrt{1-\rho^2}(\alpha_1 + \alpha_2 S_t) + \tilde{u}_0$, then the model can be presented as

$$\mathbf{x} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{6.6}$$

Using the results for the multivariate normal regression in the appendix, the joint distribution for $\mathbf{x}$ can be written as

$$p(\mathbf{x}|\theta) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})\right]\right\}, \tag{6.7}$$

where $\nu = (T+1) - 4 = T - 3$, $\hat{\beta} = (X'X)^{-1}X'\mathbf{x}$, and $\nu s^2 = (\mathbf{x} - X\hat{\beta})'(\mathbf{x} - X\hat{\beta})$.

The joint distribution for $\mathbf{y}$ then can be written as

$$p(\mathbf{y}|\theta) = \sqrt{1-\rho^2}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^{T+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\beta - \hat{\beta})' X' X (\beta - \hat{\beta})\right]\right\}. \tag{6.8}$$

6.1.3 Priors and Posteriors

We choose a joint conjugate prior for β and σ as follows:

$$p(\beta|\sigma) = \frac{|V|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^4} \exp\left\{-\frac{1}{2\sigma^2}(\beta - \beta^0)' V^{-1} (\beta - \beta^0)\right\}, \tag{6.9}$$

where V is a 4×4 positive definite symmetric constant matrix, β^0 is a 4 dimension constant vector,

$$p(\sigma) = \frac{2}{\Gamma(b/2)}\left(\frac{a}{2}\right)^{b/2} \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\}, \quad 0 < \sigma < +\infty, \tag{6.10}$$

where $a, b > 0$ are constants.

Monahan (1983) and Naylor and Marriott (1996) gave suggestions about how to choose the values for the prior covariances of β, and a, b. We follow them and choose

$$V = k_\beta^2 \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{6.11}$$

where $k_\beta = 16$ and $a = b = 1/128$. Under these selections, the prior distribution for σ is "diffuse" and is shown in Figure 6.1. The plot is truncated at quantile $q = 15.6$, correspondingly to the cumulative probability 0.04, that is, $P(\sigma \leq 15.6) = 0.04$. The prior variances for α_1, β_1, α_2, β_2 are the same $k_\beta^2\sigma^2$.

When we fit the smooth transition model to data, we believe that a smooth transition happens in the period of time of measuring the data, that is, $\tau \in (0, 1)$. We use gamma and uniform distributions for γ and τ as priors respectively,

$$p(\gamma) = G(a_1, b_1), \tag{6.12}$$

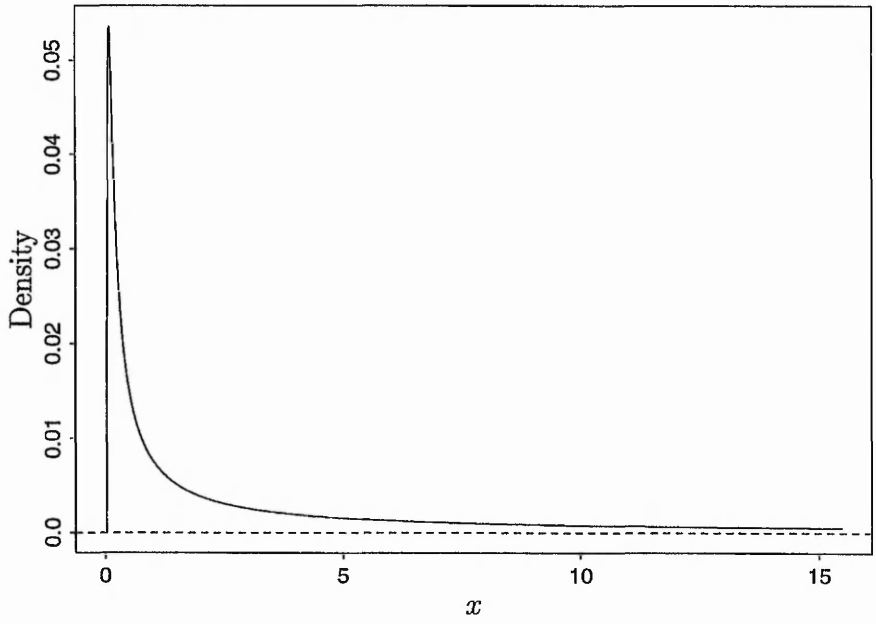


Figure 6.1: Density of inverted gamma, $IG(1/128, 1/128)$, truncated at quantile with cumulative probability 0.04.

and

$$p(\tau) = \begin{cases} 1, & \tau \in (0,1); \\ 0, & \text{otherwise.} \end{cases} \tag{6.13}$$

where $a_1, b_1 > 0$. We choose $a_1 = 4$ so that the prior mean for γ is $4b_1$ and the standard deviation is $2b_1$. The prior for γ covers the interval $(0, 10b_1)$ with probability about 0.99. From the discussion in Chapter 4 we know, when $\gamma = 0.1$, $\tau = 0.5$, about 99% of a smooth transition can be carried out, from 0.00669 to 0.99331, in the time interval $t \in (0, 100)$. In our experience, the value of γ should not be much bigger than 0.4 unless we know *a priori* that there is an abrupt transition during the time of measuring the data.

We know $\rho \in (-1, 1)$, which corresponds to the stationary case, so we choose the uniform distribution over (-1, 1) for ρ

$$p(\rho) = \begin{cases} 1/2, & \tau \in (-1,1); \\ 0, & \text{otherwise.} \end{cases} \tag{6.14}$$

We suppose that $(\boldsymbol{\beta}, \sigma)$, γ, τ and ρ are *a priori* independently distributed, so

that we have the following joint prior distribution for the parameter vector

$$p(\theta) \propto \frac{1}{\sigma^4} \exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta^0)'V^{-1}(\beta-\beta^0)\right\} \times \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\} G(a_1, b_1). \tag{6.15}$$

Now we can obtain the posterior for θ as

$$p(\theta|\mathbf{y}) \propto \frac{1}{\sigma^4} \exp\left\{-\frac{1}{2\sigma^2}(\beta-\beta^1)'V_1(\beta-\beta^1)\right\} \times \frac{1}{\sigma^{T+b+2}} \exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\} G(a_1, b_1)\sqrt{1-\rho^2}, \tag{6.16}$$

where $V_1 = X'X + V^{-1}$, $\beta^1 = V_1^{-1}(X'X\hat{\beta} + V^{-1}\beta^0)$, and $\nu s_1^2 = a + \nu s^2 + (\beta^1 - \hat{\beta})'X'X(\beta^0 - \hat{\beta})$. Notice that V_1, β^1 and νs_1^2 depend on ρ, γ and τ; $G(a_1, b_1)$ depends on γ.

6.2 Introduction of Markov Chain Monte Carlo

Markov chain Monte Carlo (MCMC) is a general method to simulate complex, non-standard multivariate distributions. MCMC methods have a profound influence on statistics especially in Bayesian inference. The theoretical discussion and applications of MCMC can be found in Geman and Geman (1984), Besag (1989), Geyer (1992), Smith and Roberts (1993), Besag and Green (1993), Tierney (1994), Gilks, Richards and Spiegelhater (1996), Robert and Casella (1999) and Chen, Shao and Ibrahim (2000), Besag (2000). Instead of sampling independently from a univariate (or multivariate) distribution, $p(\theta)$, MCMC methods simulate dependent realisations $\theta_1, \ldots, \theta_n$, which form an irreducible Markov chain with $\pi(\theta)$ as its stationary distribution (limiting distribution of θ_n). Samples from the Markov chain then can be used for statistical inference about $p(\theta)$, for example, $E_\pi(g(\theta))$ can be approximated by averaging the function $g(.)$ over the samples from the chain as

$$E_\pi(g(\theta)) \approx \frac{1}{n}\sum_{i=1}^{n} g(\theta_i). \tag{6.17}$$

When $g(\theta) = I_{\{\theta \in B\}}$, the above equation becomes an approximate evaluation of probability of $P(\theta \in B)$, where I is the usual indicator function. See, for example, Geyer (1992), Smith and Roberts (1993), Besag (2000) for more discussion.

The statistical analysis, using MCMC methods, is based on the formula in (6.17). If the distribution function is the posterior density of parameter θ, $p(\theta|\mathbf{y})$, then (6.17) can be used to estimate the posterior mean with $g(\theta) = \theta$, or posterior standard deviation with $g(\theta) = (\theta - E(\theta|\mathbf{y}))^2$, where $E(\theta|\mathbf{y})$ is the posterior mean of θ. The more complex $g(.)$ can provide estimates of other moments and probability quantiles.

It is possible to estimate the predictive density using the sample from the posterior density. If we denote $\tilde{y}$ as a future observation, corresponding to time t, we have

$$p(\tilde{y}, \theta|\mathbf{y}) = p(\tilde{y}|\theta, \mathbf{y})p(\theta|\mathbf{y}) \tag{6.18}$$

as the joint density for $\tilde{y}$ and the parameter vector θ, conditioned on the observed sample data $\mathbf{y}$, where $p(\tilde{y}|\theta, \mathbf{y})$ is the conditional distribution for $\tilde{y}$, given θ and $\mathbf{y}$. $p(\theta|\mathbf{y})$ is the posterior density for θ. In fact the distribution for $\tilde{y}$ is only determined by θ and has nothing to do with $\mathbf{y}$, so we can write $p(\tilde{y}|\theta, \mathbf{y}) = p(\tilde{y}|\theta)$. By integrating the above with respect to θ, we obtain the predictive density for $\tilde{y}$ as

$$p(\tilde{y}|_{p(\tilde{y}|\mathbf{y})} = \int_{\theta} p(\tilde{y}|\theta)p(\theta|\mathbf{y})\, d\theta. \tag{6.19}$$

The above integration may be complicated, however we have obtained the sample data from the posterior density $p(\theta|\mathbf{y})$, so we can avoid this difficulty by using Monte Carlo integration and calculate

$$p(\tilde{y}|\mathbf{y}) \approx \frac{1}{n}\sum_{i=1}^{n} p(\tilde{y}|\theta_i), \tag{6.20}$$

where θ_i, $i = 1, 2, \ldots, n$ is the sample from the posterior density for θ

The best known MCMC sampling algorithms are Gibbs sampler, (Grenander (1983), Geman and Geman (1984)) Metropolis-Hastings' sampler (Metropolis*et al.* (1953), Hastings (1970)) and the latest Reversible Jump Markov chain Monte Carlo (RJMCMC) (Green (1995)). Casella and George (1992) and Chib and Greenberg (1995) give excellent tutorials on Gibbs sampler and Metropolis-Hastings' sampler respectively. In the following, we will briefly introduce the Gibbs sampler and the Metropolis-Hastings' sampler. We will introduce the slice sampler in detail and use the slice sampler in our smooth transition model.

6.2.1 Gibbs sampler

We have the p-dimensional random vector variable $\mathbf{x} = (x^1, x^2, \ldots, x^p) \sim f(\mathbf{x})$ and marginal full conditionals $f_i = p(x^i|x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^p)$. The Gibbs sampler constructs the Markov chain through the following iterative steps as:

Given $\mathbf{x}_n = (x_n^1, x_n^2, \ldots, x_n^p)$, generate

1. $x_{n+1}^1 \sim f_1(x^1|x_n^2, \ldots, x_n^p)$;
2. $x_{n+1}^2 \sim f_2(x^2|x_{n+1}^1, x_n^3, \ldots, x_n^p)$;

   $\vdots$

p. $x_{n+1}^p \sim f_p(x^p|x_{n+1}^1, \ldots, x_{n+1}^{p-1})$.

6.2.2 Metropolis-Hastings sampler

Suppose the target density is f. In order to use Metropolis-Hastings' sampler, we arbitrarily select a proposal distribution $q(\cdot|x)$ which is easy to simulate from, then through the following iterative steps, we obtain the Markov chain with f as its stationary distribution,

Given x_n,

1. generate $y \sim q(y|x_n)$;
2. take

$$x_{n+1} = \begin{cases} y, & \text{with probability } A; \\ x_n, & \text{with probability } 1 - A, \end{cases}$$

where $A = \min\left\{1, \dfrac{f(y)q(x_n|y)}{f(x_n)q(y|x_n)}\right\}$.

There are some varieties of algorithms for Gibbs sampler and Metropolis-Hastings sampler. In the following, we introduce the slice sampler which is a special case of Gibbs sampler and will be used in the inference of the smooth transition model in this chapter.

6.2.3 Slice Sampler

Slice samplers are a form of *auxiliary variable* technique which introduce auxiliary random variables to facilitate the design of an Markov chain Monte Carlo (MCMC) sampling algorithm. Swensden and Wang (1987) first used auxiliary variable technique in MCMC for the inference of the Ising model. The Ising model can be used

in image processing. Suppose we have a black-and-white image consists of a two-dimensional table of pixels, with value 1 for white and 0 representing black, say. Then the image can be described as a binary process, $\eta = (\eta_1, \ldots, \eta_n)$, with joint distribution as

$$p(\eta) \propto \exp\left\{\sum_i \alpha_i(\eta_i) + \sum_{i<j} \beta_{ij} I_{\{\eta_i = \eta_j\}}\right\}, \tag{6.21}$$

where $\beta_{ij} \geq 0$, $i < j$ and I is the indicator function. (6.21) usually be referred to as Ising model. Edwards and Sokal (1988) generalised the Swensden-Wang technique, using the auxiliary variable technique to general models. The form of the slice sampler introduced below is due to Edwards and Sokal (1988). Besag and Green (1993) introduced the slice sampler into statistics literature. Since then the use of this technique has gradually increased. A. Mira and L. Tierney (1998) compared the slice sampler with independence Metropolis-Hastings sampler. Roberts and Rosenthal (1999) discussed the convergence of slice sampler Markov chains. Damien, Wakefield and Walker (1999) demonstrated the use of auxiliary variables for sampling non-standard densities with a variety of examples. More details of slice sampler were discussed in Neal (2000). The strength of the slice sampler is in its generality and ease of implementation. Frequently, the implementation is straightforward, which is a merit from a practical point of view.

Suppose $f(\mathbf{x})$ is a density with respect to d-dimensional Lebesgue measure, which is given, up to a constant, by

$$f(\mathbf{x}) \propto \pi(\mathbf{x}) \Pi_{i=1}^{m} l_i(\mathbf{x}), \tag{6.22}$$

where $\pi(.)$ is a density of known form and the $l_i(.)$ are non-negative functions. We obtain a Gibbs sampler by introducing *auxiliary variables* as in the following steps:

Given $\mathbf{x}_n$,

- Sample m independent uniform random variables $u_{n+1,1}, u_{n+1,2}, \ldots, u_{n+1,m}$, with

  $u_{n+1,i} \sim U(0, l_i(\mathbf{x}_n))$;

- Sample $\mathbf{x}_{n+1} \sim \pi(\mathbf{x})$ conditional on the set $L = \{\mathbf{x} : l_i(\mathbf{x}) > u_{n+1,i}\}$,

where $U(a, b)$ denotes the uniform density on the interval (a, b).

The idea behind the algorithm is straightforward. By introducing auxiliary variables $\mathbf{u} = (u_1, u_2, \ldots, u_m)'$, which are independently distributed, $u_i \sim U(0, l_i(\mathbf{x}))$, the joint density for $(\mathbf{x}, \mathbf{u})$ is

$$f(\mathbf{x}, \mathbf{u}) \propto \pi(\mathbf{x}) \Pi_{i=1}^{m} I\{u_i < l_i(\mathbf{x})\}, \tag{6.23}$$

where $I\{.\}$ denotes the indicator function. (6.23) can be proved easily by integrating the right hand side of (6.23) with respect to $\mathbf{u}$ which will give the density of $\mathbf{x}$. From the joint density for $(\mathbf{x}, \mathbf{u})$, we have the full conditional density for $\mathbf{x}$ as

$$f(\mathbf{x}|\mathbf{u}) \propto \pi(\mathbf{x}) \Pi_{i=1}^{m} I\{l_i(\mathbf{x}) > u_i\}, \tag{6.24}$$

which can be sampled as $\mathbf{x} \sim \pi(\mathbf{x})$ conditional on the $L = \{\mathbf{x} : l_i(\mathbf{x}) > u_i\}$. The full conditional density for $\mathbf{u}$ is

$$f(u_i|\mathbf{x}) \sim U(0, l_i(\mathbf{x})). \tag{6.25}$$

So conditional on $\mathbf{x}$, u_i are independently uniformly distributed over $(0, l_i(\mathbf{x}))$, $i = 1, 2, \ldots, m$ respectively. Since we can obtain the full conditionals for $\mathbf{x}$ and $\mathbf{u}$, implementing the Gibbs sampler for $(\mathbf{x}, \mathbf{u})$ is straightforward using the steps above.

6.3 Diagnosing Convergence

We have introduced several methods of constructing Markov chain which can be used in (6.17), providing an approximation of probabilities or probability quantiles we want. For the implementation of any MCMC method, we need a *stopping rule* to guarantee that the number of iterations is sufficient for the approximation of (6.17). Another problem is about the *burn-in*, that is, how to get rid of the effect of the starting state θ_0? For a Markov chain we generated, $\{\theta_0, \theta_1, \ldots, \theta_n\}$, the chain will gradually "forget" its initial state θ_0 and $P(\theta_n|\theta_0)$ will eventually converge to a unique stationary distribution of the Markov chain, denoted as $\pi(\theta)$, which is independent of n and θ_n. (See the reviews of Brooks and Roberts (1998) and Mengersen *et al.* (1999).) So far we have discussed only running one chain. It is permissible to run parallel chains. It is believed that one long run has the best chance to find new features of the distribution such as new modes while comparison between

parallel chains can provide differences if the chains have not approached stationarity. Gelman and Rubin (1992) recommended parallel chains while Geyer (1992) recommended one long chain. Cowles, Roberts and Rosenthal (2000) discussed the possibility of introducing biases into estimation based on the sampler output by applying convergence diagnostics in some ways. They recommended choosing the number of burn-in by applying convergence diagnostics to one or more *pilot* chains and then making inference on a *separate* long chain. It seems the debates between the one long chain and several parallel chains will continue. In the following, we will introduce two kinds of convergence diagnostics for parallel chains and single chain respectively.

6.3.1 Raftery and Lewis Diagnostic

Raftery and Lewis (1992) considered the convergence diagnostic of a single long run of simulation. Instead of studying the convergence of the Markov chain of $\{\theta_n\}$, Raftery and Lewis studied the derived two-state Markov chain, where an explicit analysis of convergence control is obtained. There are three numbers about the convergence, the minimum sub-sampling step in order the derived chain is a Markov chain, the number of burn-in and the length of the chain for given precision on the empirical average.

Suppose we consider quantiles of U, a function of parameter vector. We denote U_t as the value of U for the t th iteration, then we have the following derived binary chain

$$Z_t = \begin{cases} 1, & \text{if} \quad U_t \leq u; \\ 0, & \text{otherwise.} \end{cases} \tag{6.26}$$

The sequence $\{Z_t\}$ is not a Markov chain. Raftery and Lewis suppose that the dependence in $\{Z_t\}$ falls off rapid with increasing lag k, that is, if we form a subsequence of $\{Z_t\}$ as

$$Z_t^{(k)} = Z_{1+(t-1)k}, \tag{6.27}$$

consisting of every kth iteration from the original chain, then $Z_t^{(k)}$ will be approximately a Markov chain if k is sufficiently large.

Raftery and Lewis determine the lag k by testing if $\{Z_t^{(k)}\}$ is a Markov chain against the alternative that $\{Z_t^{(k)}\}$ is a second order Markov chain, that is, the vector

$(Z_t^{(k)}, Z_t^{(k+1)})$ is a Markov chain. Raftery and Lewis choose the smallest value of k for which the null hypothesis, $\{Z_t^{(k)}\}$ is a Markov chain, is preferred. After achieving the value k for the Markov chain $\{Z_t^{(k)}\}$, Raftery and Lewis determine the number of burn-in iterations, $M = mk$ by considering the two-state Markov chain with the following transition matrix

$$Q = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \tag{6.28}$$

where α is the probability of changing from the first state to the second state and β is the probability of changing from the second state to the first state. The two-state Markov chain is then converges to the stationary distribution $\pi = (\pi_0, \pi_1) = (\beta, \alpha)/(\alpha + \beta)$. Therefore Raftery and Lewis determine the burn-in length by requiring that

$$|P(Z_m^{(k)} = i | Z_0^{(k)} = j) - \pi_i| < \varepsilon \tag{6.29}$$

for $i, j = 0, 1$, where $\varepsilon > 0$ is any given precision with default value $\varepsilon = 0.001$ given by Raftery and Lewis. Raftery and Lewis gives out

$$m = \frac{\log\left(\frac{(\alpha+\beta)\varepsilon}{\max\{\alpha,\beta\}}\right)}{\log\lambda}, \tag{6.30}$$

where $\lambda = 1 - \alpha - \beta$, which is usually positive in practice. Then the number of burn-in $M = mk$ is determined. Here $\bar{Z}_n^{(k)} = \frac{1}{n}\sum_{i=1}^{n} Z_i^{(k)}$ is the estimate of $q = P(U \leq u|\text{Data})$. For large n, $\bar{Z}_n^{(k)}$ will be approximately a normal distribution. Raftery and Lewis require that $P(q - r \leq \bar{Z}_n^{(k)} \leq q + r) = s$, with default values $r = 0.0125$, $s = 0.95$, that is, the estimate of cumulative distribution function locates within ± 0.0125 of the true cumulative distribution function with probability 0.95. Raftery and Lewis then gives out the length of iterations which satisfies this requirement as $N = nk$, where

$$n = \frac{(2-\alpha-\beta)\alpha\beta}{(\alpha+\beta)^3}\left\{\frac{\Phi^{-1}\left(\frac{s+1}{2}\right)}{r}\right\}^2, \tag{6.31}$$

where Φ is the standard normal cumulative distribution function. Moreover Raftery and Lewis suggest that we can use the non-thinning chain to obtain greater accuracy than their criterion. Raftery and Lewis also give out the initial number of iterations

for the pilot run of the chain as

$$N_{\min} = \left\{\Phi^{-1}\left(\frac{s+1}{2}\right)\right\}^2 \frac{q(1-q)}{r^2}. \tag{6.32}$$

For example, when q =0.025 or 0.975, $r = 0.0125$ and $s = 0.95$, then $N_{\min} = 600$. It would be possible to run a final run of the simulation chain with the above M, N and k. Raftery and Lewis (1992b) suggest that a further statistic $I = \frac{M+N}{N_{\min}}$ could be calculated to measure the increase in the number of iterations due to dependence in the sequence. Values of I greater than 1 indicate a high level of dependence. Raftery and Lewis suggest that I greater than 5 often indicates problems of the implementation.

6.3.2 Gelman and Rubin Diagnostic

Gelman and Rubin (1992) diagnostic is applicable to parallel chains, based on detecting when the Markov chains have "forgotten" their starting points, that is, the parallel Markov chains drawn from different starting points become indistinguishable by comparing the between-sequence variance and the within-sequence. Gelman and Rubin (1992) advise that a set of starting values which are over-dispersed with respect to the stationary distribution π should have been obtained for the parallel Markov chains. Gelman and Rubin diagnostic can be applied to all scalar summaries of interest from the target distribution (posterior distribution of parameters), for example, mean, standard deviation, 2.5% and 97.5% order statistics which consist of 95% posterior probability interval, of each parameters. We denote the scalar summary of interest as ψ, assume m parallel simulation chains, each of length n. Gelman and Rubin define the between sequence variance B and the within-sequence variance W as

$$\begin{aligned} B &= \frac{n}{m-1}\sum_{i=1}^{m}(\bar{\psi}_{i.} - \bar{\psi}_{..})^2, \\ W &= \frac{1}{m}\sum_{i=1}^{m} s_i^2, \end{aligned} \tag{6.33}$$

where $\bar{\psi}_{i.} = \frac{1}{n}\sum_{j=1}^{n}\psi_{ij}$, $\psi_{..} = \frac{1}{m}\sum_{i=1}^{m}\bar{\psi}_{i.}$ and $s_i^2 = \frac{1}{n-1}\sum_{j=1}^{n}(\psi_{ij} - \bar{\psi}_{i.})^2$. Then Gelman and Rubin define the "conservative" estimate of the variance of ψ as

$$V(\psi) = \frac{n-1}{n}W + \frac{1}{n}B, \tag{6.34}$$

which is unbiased under stationarity, that is, if the starting points were actually drawn from the target distribution. This estimate is "an overestimate" because the starting points were over-dispersed and may not be from the target distribution. Gelman and Rubin call the within-sequence variance W an "underestimate" variance of ψ because the simulation Markov chains have not explored the whole support of target distribution within limit length of simulation and therefore have less variability. When $n \to \infty$, both $V(\psi)$ and W will approach $\text{var}(\psi)$.

Gelman and Rubin suggest that we can monitor the convergence of the parallel Markov chains by estimating the ratio between the "overestimate" $V(\psi)$ and "underestimate" W

$$\hat{R} = \frac{V(\psi)}{W}, \tag{6.35}$$

which will approach 1 when the parallel Markov chains converge. Gelman and Rubin also suggest that if all the values of $\hat{R}$ are less than 1.1 or 1.2 then we can assess the convergence of the parallel chains.

Brooks and Gelman (1998) suggest a modification of the statistic $\hat{R}$ accounting for sampling variability in the variance estimates as

$$\hat{R}_c = \frac{d+3}{d+2}\hat{R} = \frac{d+3}{d+2}\frac{V(\psi)}{W}, \tag{6.36}$$

where d is the estimated degrees of freedom for a student-t approximation to the posterior inference based on the simulations. The degrees of freedom d can be estimated by the method of moments as $d \approx 2V(\psi)/\widehat{\text{var}}(V(\psi))$, given by Brooks and Gelman.

S-plus programmes for these two convergence diagnostics can be found in Statlib, S archive and I use these S-plus programmes to carry out the convergence diagnostics.

6.4 Application of the Slice Sampler

Here we indicate how we can use Monte Carlo Markov Chain methods to obtain inference for the smooth transition model. In particular, we will use the slice sampler to obtain samples from the posterior density (6.16). We can factorize the expression

in (6.16) as

$$p(\theta|\mathbf{y}) \propto \pi(\theta) \prod_{i=1}^{3} l_i(\theta), \tag{6.37}$$

where $\pi(\theta)$ is a proper joint density function for θ, $l_i(\theta)$, $i = 1, 2, 3$, are positive functions given below

$$\begin{aligned}
\pi(\theta) \propto & \frac{|V_1|^{1/2}}{(\sqrt{2\pi}\sigma)^4} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^1)' V_1 (\boldsymbol{\beta} - \boldsymbol{\beta}^1)\right\} \times \\
& \frac{(\nu s_1^2)^{\frac{T+b+1}{2}}}{\sigma^{T+b+2}} \exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\} G(a_1, b_1), \\
l_1(\theta) = & |V_1|^{-1/2}, \\
l_2(\theta) = & (\nu s_1^2)^{-\frac{T+b+1}{2}}, \\
l_3(\theta) = & \sqrt{1-\rho^2}.
\end{aligned} \tag{6.38}$$

We can sample from $\pi(\theta)$ directly, for example, $\pi(\theta)$ can be factorized as

$$\pi(\theta) = p(\gamma, \tau, \rho) p(\sigma|\gamma, \tau, \rho) p(\boldsymbol{\beta}|\sigma, \tau, \rho, \sigma), \tag{6.39}$$

if we sample γ, τ and ρ independently, then we have the factorization as

$$\pi(\theta) = p(\gamma) p(\tau) p(\rho) p(\sigma|\gamma, \tau, \rho) p(\boldsymbol{\beta}|\gamma, \tau, \rho, \sigma), \tag{6.40}$$

where

$$\begin{aligned}
p(\gamma) &= G(a_1, b_1), \\
p(\tau) &= I_{\{0<\tau<1\}}, \\
p(\rho) &= I_{\{-1<\rho<1\}}, \\
p(\sigma|\gamma, \tau, \rho) &= \frac{(\nu s_1^2)^{\frac{T+b+1}{2}}}{\sigma^{T+b+2}} \exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\}, \\
p(\boldsymbol{\beta}|\gamma, \tau, \rho, \sigma) &= \frac{|V_1|^{1/2}}{(\sqrt{2\pi}\sigma)^4} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^1)' V_1 (\boldsymbol{\beta} - \boldsymbol{\beta}^1)\right\},
\end{aligned} \tag{6.41}$$

where $I_{\{.\}}$ is the indicator function. So the sampling θ from $\pi(\theta)$ can be carried out in the following steps

- First, sample γ, τ and ρ independently from gamma distribution, uniform distribution on $(0, 1)$ and uniform distribution on $(-1, 1)$ respectively;
- then sample σ, conditional on γ, τ and ρ, from inverted gamma distribution;
- finally sample $\boldsymbol{\beta}$, conditional on γ, τ, ρ and σ, from multivariate normal distribution.

The implementation of the slice sampler is therefore straightforward by iterating between

- sampling θ from $\pi(\theta)$ conditional on θ being in the set $L(u_1, u_2, u_3) = \{\theta : l_i(\theta) \geq u_i, i = 1, 2, 3\}$ by rejection sampling and
- independently sampling new values of u_i from $U(0, l_i(\theta))$, $i = 1, 2, 3$.

We notice $\{\theta : l_3(\theta|\mathbf{y}) \geq u_3\} = \{\theta : -\sqrt{1-u_3^2} \leq \rho \leq \sqrt{1-u_3^2}\}$ and $l_i(\theta|\mathbf{y}), i = 1, 2$ are functions of γ, τ and ρ only. So we do not need to sample σ and $\boldsymbol{\beta}$ until we have proper samples for ρ, γ and τ which satisfy $l_i(\theta) \geq u_i$, $i = 1, 2, 3$. We can adjust the steps as follows

- sampling ρ from $U(-\sqrt{1-u_3^2}, \sqrt{1-u_3^2})$;
- sampling independently γ and τ conditional on being in the set $L(u_1, u_2) = \{(\gamma, \tau, \rho) : l_i(\theta|\mathbf{y}) \geq u_i, i = 1, 2\}$ by means of rejection sampling;
- sampling σ from the inverted gamma distribution conditional on γ, τ and ρ;
- sampling $\boldsymbol{\beta}$ from multivariate normal distribution conditional on γ, τ, ρ and σ;
- sampling new values of u_i from $U(0, l_i(\theta|\mathbf{y}))$ independently for $i = 1, 2, 3$.

Before we begin our sampling, we need to discuss how to generate data from some familiar distributions. First we need to generate data from an inverted gamma (IG) distribution or gamma G distribution. Inverted gamma and gamma distributions have the following simple relationship. Given $g \sim G(\alpha, \beta)$ with the two parameters (α, β), we set $\sigma = 1/\sqrt{g}$, then $\sigma \sim IG(\nu, s)$, with the two parameters (ν, s), where the relationship between (α, β) and (ν, s) is given by

$$\begin{cases} \alpha = \dfrac{\nu}{2}, \\ \beta = \dfrac{2}{\nu s^2}. \end{cases}$$

Because we have ν very large (compared with 1), we have $\alpha > 1$. So, we only need the algorithm to generate data from a gamma distribution, $G(\alpha, \beta)$, with the parameters $\alpha > 1$, $\beta > 0$. It is easy to prove (see Appendix A) that $h \sim G(\alpha, 1)$,

then $g = \beta h \sim G(\alpha, \beta)$. The method we adopt for generating data from a $G(\alpha, 1)$ with $\alpha > 1$ was suggested by Best (1978),

1. Define $b = \alpha - 1$, $c = (12\alpha - 3)/4$;

2. Generate u, v *iid* $U(0,1)$ and define

$$w = u(1-u), \quad y = \sqrt{\frac{c}{w}}\left(u - \frac{1}{2}\right), \quad x = b + y;$$

3. If $x > 0$, take $z = 64v^2w^3$ and accept x when

$$z \leq 1 - 2y^2/x$$

or when

$$2(b\log(x/b) - y) \geq \log(z);$$

4. Otherwise, start from 2.

For the multivariate normal distribution, we generate $\mathbf{Z}$ from the standard multivariate normal distribution and then use the Choleski method to factorize the required covariance matrix V so that $V = LL'$, where L is a lower triangular matrix. If we set $\mathbf{Y} = L'\mathbf{Z} + \boldsymbol{\mu}$, we then obtain the multivariate normal vector variable $\mathbf{Y} \sim \mathbf{N}(\boldsymbol{\mu}, V)$. (See A. Jennings and J. J. McKeown (1992), second edition for the discussion of this method and Choleski decomposition of symmetric matrix.)

In the following, we will discuss the application of our approach to two real economic time series, leaving the discussion of the performance of the slice sampler we used in the last section.

6.5 Empirical Applications

In the previous section, we introduced the slice sampler for sampling from the joint posterior density function for the parameters in the smooth transition model. Now we apply our method to two empirical examples to illustrate our procedure. The first series is the annual British industrial production index, from year 1780 to 1913 (134 observations) reproduced in Table A.1 on page 160 in the Appendix. Crafts, Leybourne and Mills (1989) constructed the first series, and decomposed

the series into its trend and cycle components and compared with those obtained from traditionally available indices. Newbold and Agiakloglou (1991) provided an alternative analysis of this series with the standard autoregressive integrated moving average model building approach. Leybourne, Newbold and Vougas (1998) tested the null hypothesis of an autoregressive unit root for this series. One of their tests rejected the null hypothesis for the annual British industrial production series at 1% significance level. Then they fitted the alternative, the smooth transition model with AR(2) error, to this series using a nonlinear least squares (NLS) algorithm. The second is the US quarterly gross domestic production series, from 1961 to the first two quarters of 2000 (158 observations). We follow Leybourne, Newbold and Vougas (1998) to fit the smooth transition model to the first series, as well as the second series, with the logarithm of the scaled data (that is, if the original series is x_t, we use series $y_t = \log(x_t.)$

In the following, we will present the results of the Bayesian analysis of the two series and then conduct the discussion of the performance of the slice sampler Markov chains for the two series, the convergence diagnostic, the burn-in and how long the Markov chains used for Bayesian analysis, *etc.*

6.5.1 Example 1: British Industrial Production Index

We first apply our approach to the British industrial production index series. We obtain the fitted smooth transition model, using posterior means as point estimates for the parameters

$$\begin{aligned} \hat{y}_t &= 1.28 + 0.0128t + 0.693S_t \\ &\quad +0.00692tS_t + u_t \\ u_t &= 0.585u_{t-1} + \varepsilon_t, \end{aligned} \tag{6.42}$$

where $S_t = \dfrac{1}{1+\exp\{-0.0842(t-51.471)\}}$, $\hat{\tau} = 0.387$, $T = 133$, ε_t *iid* $\mathrm{N}(0, \hat{\sigma}^2)$ and $\hat{\sigma} = 0.0403$. The estimated trend in the smooth transition model and the British industrial production series are plotted in Figure 6.2, where the solid line is the actual data and the dotted line is the estimated trend.

The Bayesian analysis provides more information than the posterior mean of parameters (point estimates used in (6.42)). We also obtained the posterior second

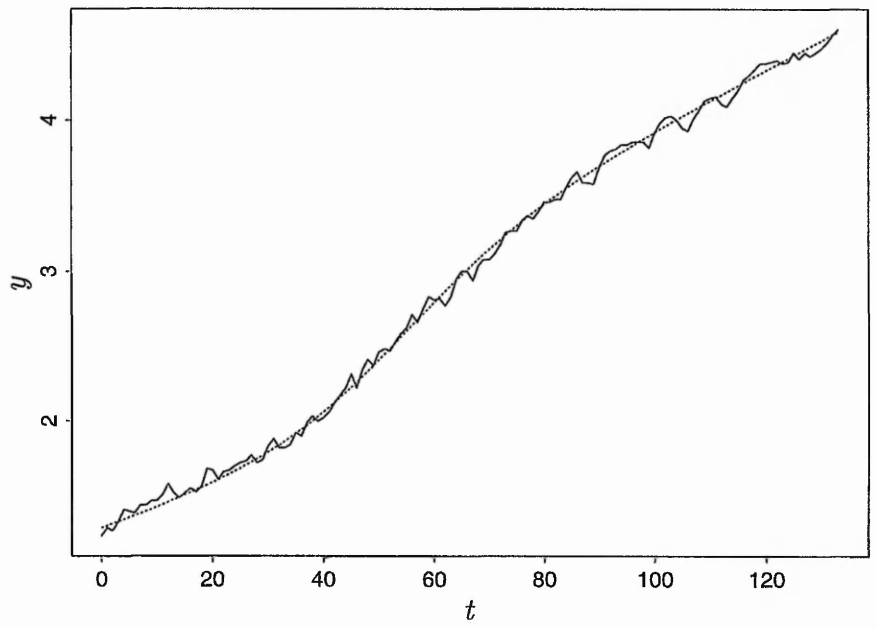


Figure 6.2: UK industrial production and the fitted trend.

moments and an estimate of posterior density for each of the parameters. The posterior mean and standard deviations for each parameter are listed in Table 6.1. The posterior densities for all the parameters in the model are plotted in Figure 6.3. The quantiles with different cumulative probabilities for each parameter are listed in Table 6.2.

	α_1	β_1	α_2	β_2
mean	1.28	0.0128	0.693	0.00692
sd	0.0477	0.00447	0.222	0.00412
	γ	τ	ρ	σ
mean	0.0842	0.387	0.585	0.0403
sd	0.0183	0.0426	0.097	0.00255

Table 6.1: Posterior mean and standard deviation for UK industrial production index series.

Using the posterior quantiles in Table 6.2, we can obtain Bayesian interval estimations for each parameter, for example, a 95% posterior probability interval for γ

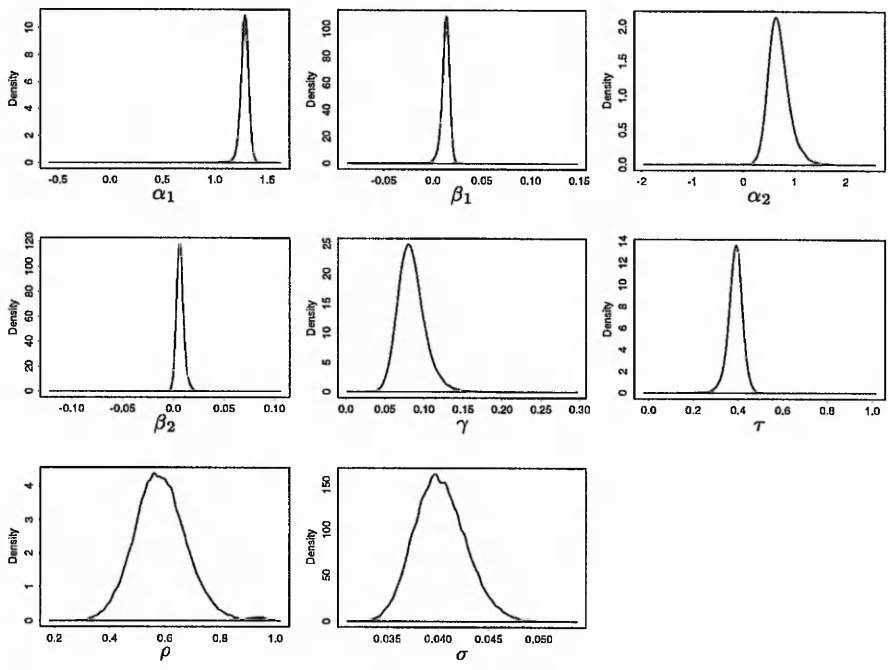


Figure 6.3: Posterior densities of the parameters for UK industrial production index series.

	$p = P(X \leq x)$						
p	0.025	0.05	0.25	0.5	0.75	0.95	0.975
α_1	1.19	1.21	1.26	1.29	1.31	1.34	1.36
β_1	0.00271	0.00531	0.0106	0.0132	0.0156	0.0189	0.02
α_2	0.328	0.386	0.55	0.67	0.811	1.08	1.2
β_2	4.98e-05	0.00115	0.00433	0.00656	0.00897	0.0135	0.0158
γ	0.0538	0.0582	0.072	0.0824	0.0941	0.116	0.125
τ	0.308	0.329	0.369	0.39	0.409	0.44	0.451
ρ	0.407	0.435	0.521	0.58	0.643	0.747	0.788
σ	0.0356	0.0363	0.0385	0.0401	0.0419	0.0447	0.0456

Table 6.2: Quantiles for the posterior distributions of the parameters for UK industrial production index series.

is $(0.0538, 0.125)$; or $(0.0582, 0.116)$ with posterior probability 0.9. A 95% posterior probability interval for τ is $(0.308, 0.451)$; or $(0.329, 0.44)$ with posterior probability 0.9.

We found from the plot of the posterior densities and quantiles for the posterior densities of γ and τ that, our approach favours a gradual smooth transition with the speed γ around 0.0842 and the midpoint of the smooth transition locates at τT around 51.471, corresponding to May 1831.

We can obtain the predictive density for $\tilde{y} = \tilde{y}_{T+1}$, which is presented in Figure 6.4. The predictive mean, standard deviation and quantiles for $\tilde{y}_{T+1}$ are listed in Table 6.3.

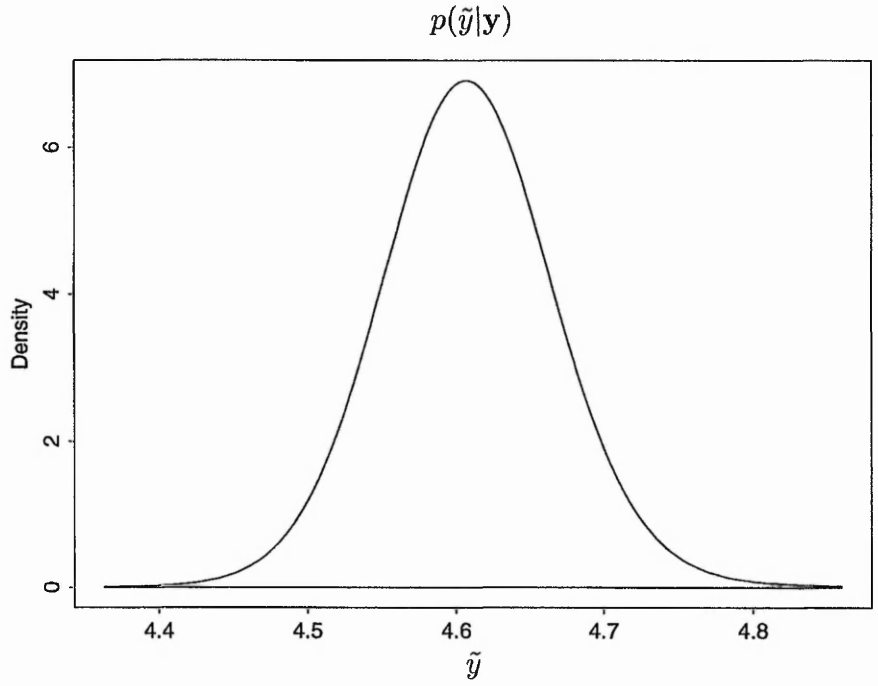


Figure 6.4: First step predictive density for UK industrial production index series.

$p = P(X \leq q)$	0.025	0.05	0.25	0.5	0.75	0.95	0.975	mean	sd
q	4.49	4.51	4.57	4.61	4.65	4.71	4.73	4.61	0.0598

Table 6.3: Statistics for the first step predictive density for UK industrial production index series.

The predictive distribution can also be obtained for $\tilde{y}_t$, $t = 0, 1, \ldots, T$. With the

predictive densities for $\tilde{y}_t$, $t = 0, 1, \ldots, T$, we can obtain mean and quantiles of these predictive distributions. Fitting model with the predictive mean of $\tilde{y}_t$, $E(\tilde{y}_t|\mathbf{y})$, is preferable in Bayesian literature, because this approach is obtained from the Bayesian point of view, making all inferences from the corresponding distributions conditional on the observed sample $\mathbf{y}$.

We obtained two quantiles with cumulative probabilities 0.025 and 0.975 respectively for each $\tilde{y}_t$, $t = 0, 1, \ldots, T$. These can then be plotted with the predictive mean, $E(\tilde{y}_t|\mathbf{y})$, and actual data in Figure 6.5 to give approximate 95% predictive probability intervals for the model at each time $t = 0, 1, \ldots, T$. The solid line is the actual data, the dotted line is the predictive mean, $E(\tilde{y}_t|\mathbf{y})$, and the dashed line is the 95% predictive probability intervals. We can see from the plot that all the actual data fall in the 95% predictive probability intervals in this case.

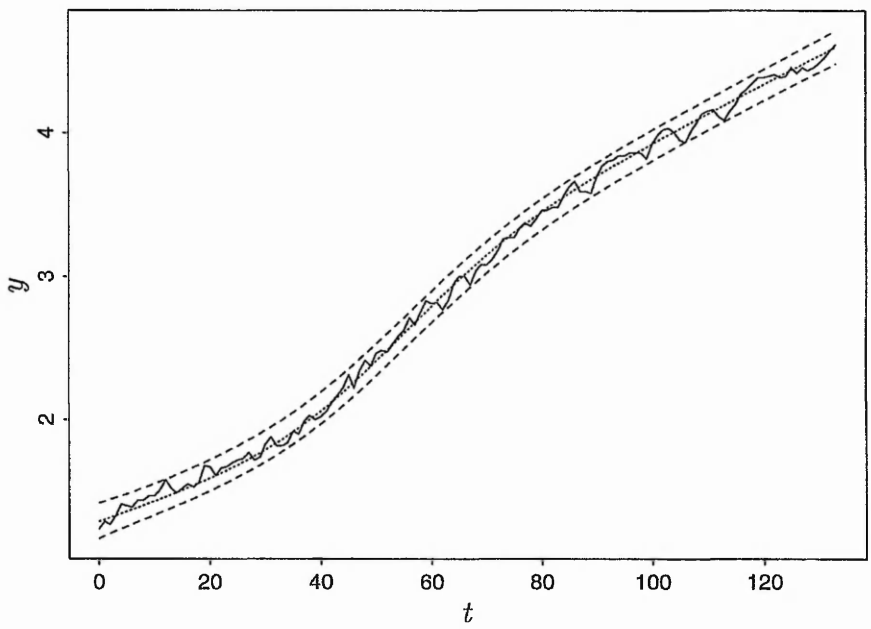


Figure 6.5: 95% predictive probability intervals for UK industrial production index series.

In the smooth transition model, we use the smooth transition function $S_t = S_t(\gamma, \tau)$ to describe the transition. The smooth transition function we selected $S_t = S_t(\gamma, \tau)$ satisfies some properties as $S_t \in (0, 1)$ and $S_{\tau T+\Delta t} = 1 - S_{\tau T-\Delta t}$. We would like to know if these properties preserved for the posterior mean of the smooth

transition function

$$E(S_t|\mathbf{y}) = \int S_t\, p(\theta|\mathbf{y})\, d\theta. \tag{6.43}$$

With the posterior sample from the posterior density of the parameters , we can use the Monte Carlo integrating method to obtain the posterior mean of the smooth transition function as

$$E(S_t|\mathbf{y}) \approx \frac{1}{n}\sum_{i=1}^{n} S_t(\gamma_i, \tau_i), \tag{6.44}$$

where γ_i, τ_i are the marginal sample from the sample θ_i, $i = 1, 2, \ldots, n$, where n is the sample size. We have plotted $E(S_t|\mathbf{y})$ as a function of t in Figure 6.6. We found that $E(S_t|\mathbf{y}) \in (0, 1)$ but $E(S_{\tau T+\Delta t}|\mathbf{y}) \neq 1 - E(S_{\tau T-\Delta t}|\mathbf{y})$, that is, the range of $E(S_t|\mathbf{y})$ is the same of the range of S_t but the shape of $E(S_t|\mathbf{y})$ is different from S_t.

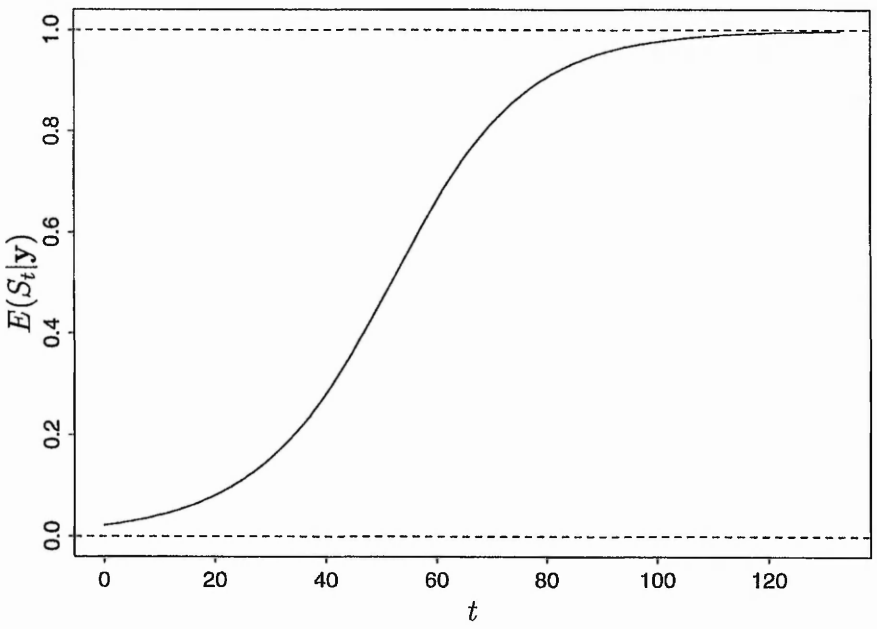


Figure 6.6: Posterior mean of the smooth transition function for UK industrial production index series.

6.5.2 Example 2: US Quarterly GDP series

Now we will apply our approach to the US quarterly gross domestic production series. We found the disturbance term for this series is very weak compared with

the trend, we feel that it is unnecessary to consider the disturbance term as AR(1). Using the posterior means as point estimates for each parameter, we obtained the fitted smooth transition model with white noise as

$$\hat{y}_t = 6.27 + 0.0174t + 0.802S_t - 0.00389tS_t + \varepsilon_t, \tag{6.45}$$

where $S_t = \dfrac{1}{1+\exp\{-0.0854(t-70.65)\}}$, $\hat{\tau} = 0.45$, $T = 157$, ε_t *i.i.d.* $\mathrm{N}(0, 0.0347^2)$. The actual data and the estimated trend are plotted in Figure 6.7 with a solid line and a dotted line respectively. The posterior means and standard deviations for each parameter in the model are listed in Table 6.4 and posterior quantiles are presented in Table 6.5. The posterior densities for the parameters are plotted in Figure 6.8. Our approach favours a gradual transition for this series with the speed of the smooth transition around $\gamma = 0.0854$ and the midpoint of the smooth transition locates around $\tau * T = 70.65$, corresponding to the third quarter of 1978.

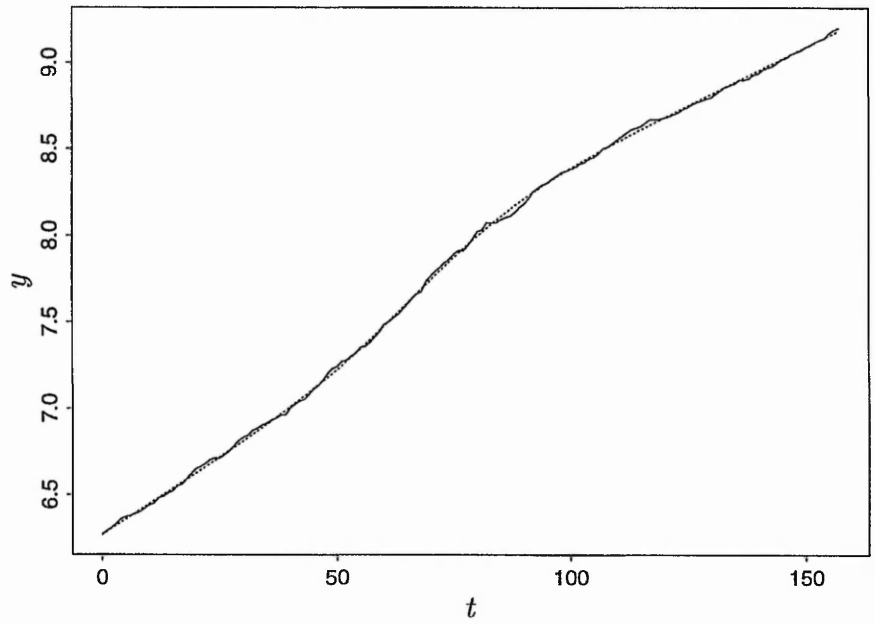


Figure 6.7: US gross domestic production and the fitted smooth transition components.

We also present the predictive density for $\tilde{y}_{T+1}$ in Figure 6.9, the posterior predictive moments and quantiles in Table 6.6.

	α_1	β_1	α_2	β_2
mean	6.27	0.0174	0.802	-0.00389
sd	0.0106	0.000763	0.0547	0.000663
	γ	τ	σ	
mean	0.0854	0.45	0.0347	
sd	0.0105	0.0132	0.00198	

Table 6.4: Posterior mean and standard deviation for model of US GDP series.

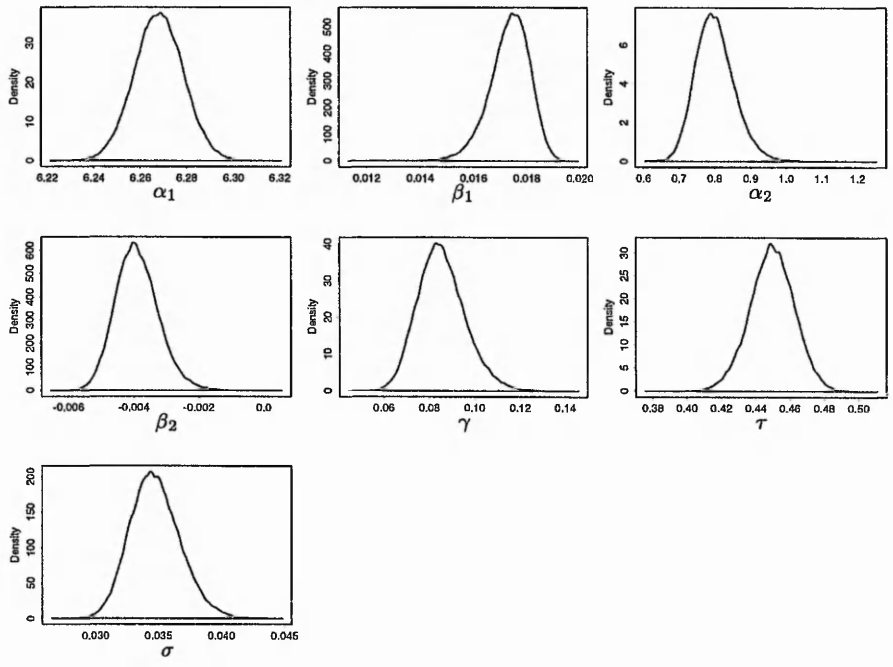


Figure 6.8: Posterior densities for parameters in the model for US GDP series.

	$p = P(X \leq x)$						
p	0.025	0.05	0.25	0.5	0.75	0.95	0.975
α_1	6.247	6.251	6.261	6.268	6.275	6.285	6.289
β_1	0.0157	0.016	0.0169	0.0174	0.0179	0.0185	0.0186
α_2	0.705	0.719	0.764	0.798	0.836	0.898	0.92
β_2	-0.0051	-0.00492	-0.00435	-0.00393	-0.00347	-0.00275	-0.00248
γ	0.0667	0.0693	0.0781	0.0847	0.0919	0.104	0.108
τ	0.423	0.427	0.441	0.45	0.458	0.471	0.475
σ	0.0311	0.0317	0.0334	0.0346	0.036	0.0382	0.0389

Table 6.5: Quantiles for the posterior in model for US GDP series.

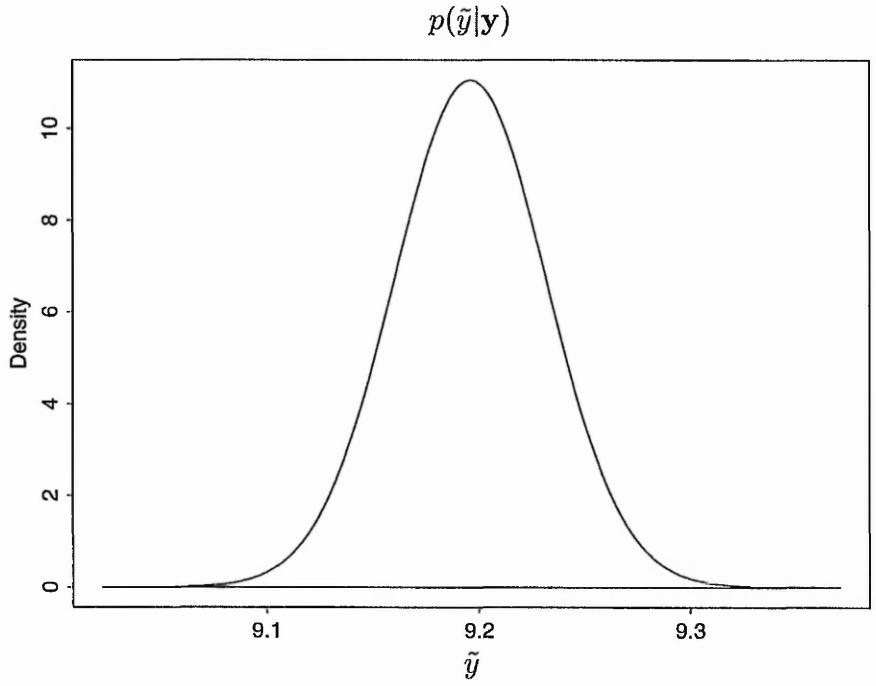


Figure 6.9: First step predictive density for US GDP series.

$p = P(X \leq q)$	0.025	0.05	0.25	0.5	0.75	0.95	0.975	mean	sd
q	9.13	9.14	9.18	9.2	9.22	9.26	9.27	9.2	0.0362

Table 6.6: Statistics for the first step predict for US GDP series.

The US government published new data of the US gross domestic production, the data for the third quarter of year 2000, which was 10,052.2 billion dollars. The logarithm of the scaled datum for the third quarter is therefore 9.21555, which is between the two quantiles 9.2 and 9.22, correspondingly to the cumulative probability 0.5 and 0.75 respectively.

We again obtained an approximate a 95% predictive probability interval for each of $\tilde{y}_t$, $t = 0, \ldots, T$ and plot these predictive probability intervals with the actual data and the predictive mean of $\tilde{y}_t$, $E(\tilde{y}_t|\mathbf{y})$, in Figure 6.10. The two dashed lines outside consist of the approximate 95% predictive intervals, the dotted line is the predictive means, $E(\tilde{y}_t|\mathbf{y})$, and the solid line is the true data. The posterior mean of the smooth transition function for the model fits the US gross domestic production series is plotted in Figure 6.11.

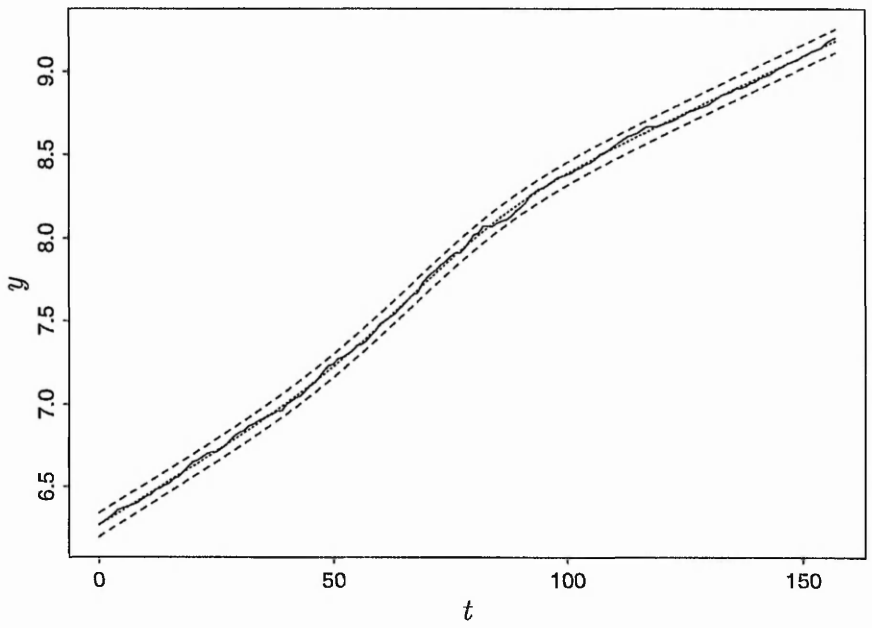


Figure 6.10: 95% predictive probability intervals for US GDP series.

6.6 Performance of the Slice Sampler

We have proposed using the slice sampler to carry out sampling from the posterior density of the parameters of the smooth transition model because it is the most

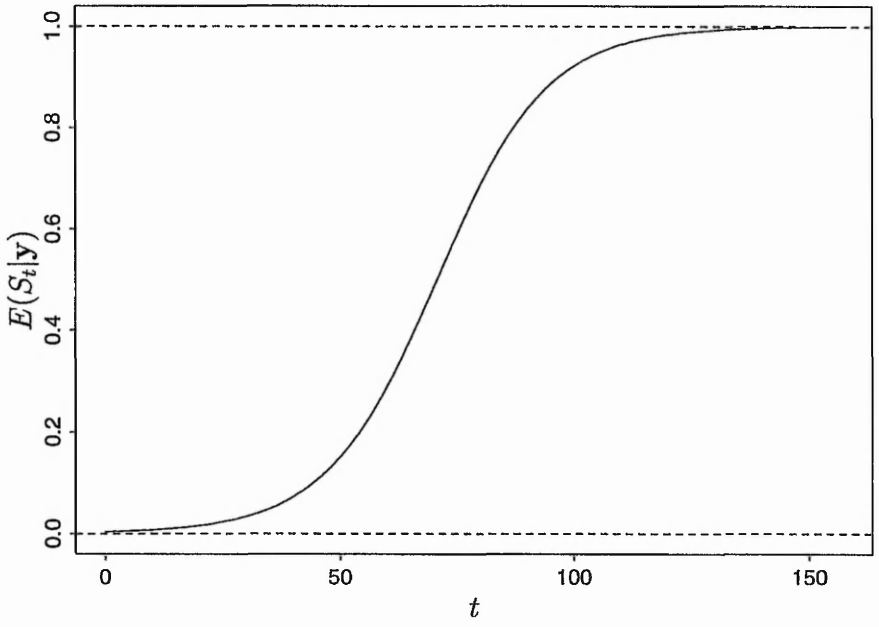


Figure 6.11: Posterior mean of the smooth transition function for US GDP series.

direct way to do this in this case. The expression for the posterior density for the parameters in this model is complex, especially for γ, τ and ρ. If we had some other information for the densities of these parameters, we might be able to use other methods, for example, the Metropolis-Hastings algorithm, using the prior information to determine the proposal distributions, especially for γ, τ and ρ. The slice sampler is direct but in practice, there may be problem, especially when the number of auxiliary variable increases, the determination of the set $L(\mathbf{u}) = \{\theta : l_i(\theta) \geq u_i\}$ may become increasingly difficult; if we still use rejection method for the determination of the set $L(\mathbf{u}) = \{\theta : l_i(\theta) \geq u_i\}$, this may cause low efficiency. Moreover, the increasing of auxiliary variables may cause the slowdown of the convergence of the Markov chain. (See discussion of Robert and Casella (1999), p291.) There are other versions of slice sampler and these methods usually will not sample more efficiently than Gibbs sampler or a well-tuned Metropolis-Hastings sampler. (See Neal (2000).) In the next chapter, when we consider the double smooth transition model, we will adopt other methods to sample from the posterior density, avoiding these problems.

We have discussed how to choose a_1 and b_1 for the prior for γ in the section 6.1.3,

where we suggested $a_1 = 4$ so that the standard deviation for γ was not too small compared with its mean. When $a_1 = 4$, we have the prior mean as $4b_1$ and prior standard deviation as $2b_1$. Economists think there are gradual smooth transitions in the UK industrial production series and US GDP series. So we select a small value for b_1, for example, $b_1 = 0.025$, then the prior mean of γ is 0.1. $\gamma = 0.1$ is suitable for describing a graduate smooth transition which carries out over an medium size of time interval, for example, $[0, 100]$.

We use two different methods to diagnose the convergence of the Markov chain for UK industrial production index series, Raftery and Lewis (1992a) and Gelman and Rubin (1992) diagnostics.

For the Raftery and Lewis diagnostic, we choose initial values for $\gamma = 0.1$, $\tau = 0.5$ and $\rho = 0$. For the Gelman and Rubin diagnostic, we generate six parallel simulation chains, each of which has dispersed values for γ, τ and ρ according to our prior knowledge about them: γ is positive and less than 1 for a graduate smooth transition; $\tau \in (0, 1)$; and $\rho \in (-1, 1)$. We list the dispersed initial values for the six parallel simulation chains in Table 6.7.

Series	γ	τ	ρ
1	0.1	0.5	0
2	0.01	0.5	0.5
3	1	0.1	-0.9
4	0.1	0.9	0.9
5	0.01	0.1	-0.5
6	1	0.9	0

Table 6.7: Dispersed initial values for γ, τ and ρ in six parallel simulations for UK industrial production index series.

Raftery and Lewis' method suggests that a "burn-in" should be under 100 iterations and the length of chain to obtain the default accuracy ($r = 0.0125$, $s = 0.95$) is about 3000. We check both quantiles $q = 0.025$ and $q = 0.975$ which are frequently used in practice. We list the diagnostic of Raftery and Lewis in Table 6.8 for the chain between iteration 1000 and 2000 with left-hand cumulative distribution at

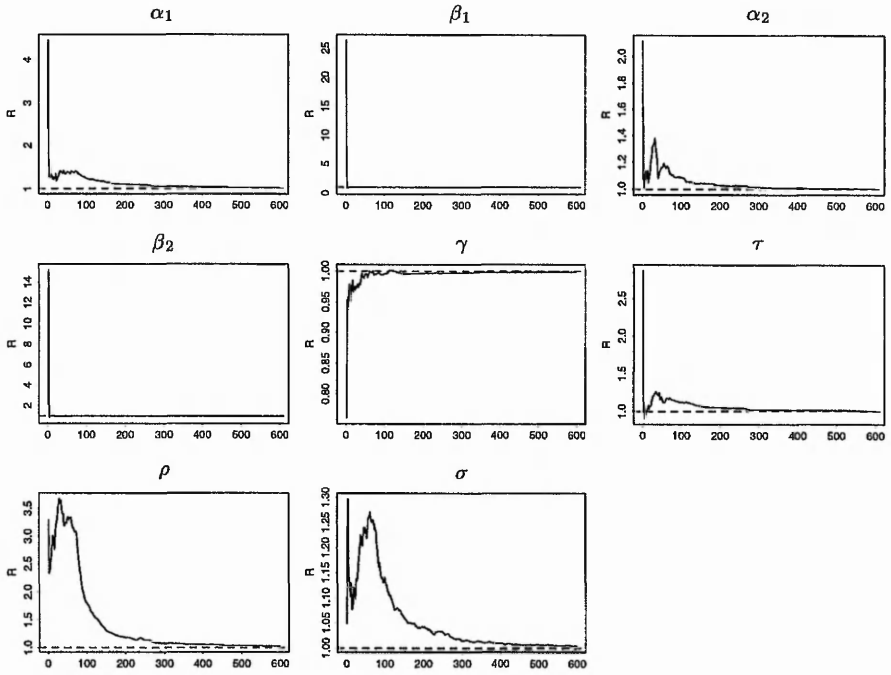


Figure 6.12: Gelman and Rubin diagnostic for Markov chain for UK industrial production index series.

0.025 and 0.975 respectively. In Table 6.8, k_{ind} is the thinning parameter required to make the chain into an independent chain. Raftery and Lewis diagnostic shows that the simulation chains for all the parameters in the smooth transition model, except ρ, converge after 1000 iterations. We do not need to worry about the inference of ρ too much because we are much more interested in other parameters in the smooth transition model than ρ and σ because other parameters have specific interpretations in a practice series. For example, γ and τ are the speed and proportion of midpoint respectively for a smooth transition. Further diagnostics using simulation chain between other iterations (2000–3000, 3000–4000 *etc*) give out similar results for k, M, N, I and k_{ind}.

Gelman and Rubin's method suggests that a "burn-in" of 200 iterations is enough. Values for $\hat{R}_c$ are all smaller than 1, and the smallest is 0.9977. The biggest value of the upper bound of $\hat{R}_c$ is 1.002, the smallest is 0.9977. All the values of $\hat{R}_c$ are under and near 1 indicate that the six parallel simulations are indistinguishable. We list the diagnostic results in Table 6.9. We plot the Gelman

	$P(U \leq u\|\text{Data}) = 0.025$					$P(U \leq u\|\text{Data}) = 0.975$				
	k	M	N	I	k_{ind}	k	M	N	I	k_{ind}
α_1	1	3	749	1.25	2	1	2	570	0.95	1
β_1	1	3	692	1.16	2	1	3	729	1.22	2
α_2	1	2	591	0.99	1	1	2	618	1.03	2
β_2	1	3	749	1.25	2	1	3	671	1.12	2
γ	1	3	692	1.16	2	1	3	671	1.12	2
τ	1	4	811	1.36	2	1	3	729	1.22	2
ρ	1	4	811	1.36	2	1	18	3053	5.12	8
σ	1	2	591	0.99	1	1	2	570	0.95	1

Table 6.8: Raftery and Lewis diagnostic for the UK industrial production series.

and Rubin diagnostic for the posterior means in Figure 6.12, which also suggest convergence after 200 iterations.

	$P(U \leq u\|\text{Data})$						
	2.5%	25%	50%	75%	97.5%	$\hat{R}_c$	$\sup\{\hat{R}_c\}$
α_1	1.217	1.258	1.281	1.305	1.354	0.9979	0.9983
β_1	0.0008016	0.01032	0.01280	0.01523	0.01881	0.9977	0.9977
α_2	0.3360	0.5755	0.6894	0.8084	1.139	0.9980	0.9989
β_2	0.0007584	0.004758	0.007127	0.009181	0.01463	0.9978	0.9979
γ	0.05389	0.07041	0.08092	0.08951	0.1199	0.9987	1.001
τ	0.3023	0.3649	0.3862	0.4061	0.4419	0.9978	0.9980
ρ	0.3889	0.5187	0.5839	0.6497	0.7536	0.9982	0.9992
σ	0.03634	0.03982	0.04026	0.04079	0.04305	0.9990	1.002

Table 6.9: Gelman and Rubin diagnostic for the UK industrial production series.

Additionally, we plot the simulated sample of each parameter in the long single simulation for the first 600 iterations in Figure 6.13. These simulated samples do not show any sign of lack of convergence after 300 iterations. We also plot the autocorrelation function for each parameter in Figure 6.14. These autocorrelation functions show us how fast the autocorrelation functions go to 0 with the increasing

of iterations. This is important for the MCMC methods because the autocorrelation function measure the dependence within a chain. High dependence usually means slow convergence of the chain. We can see from the plot of autocorrelation functions that all the autocorrelation functions decrease rapidly.

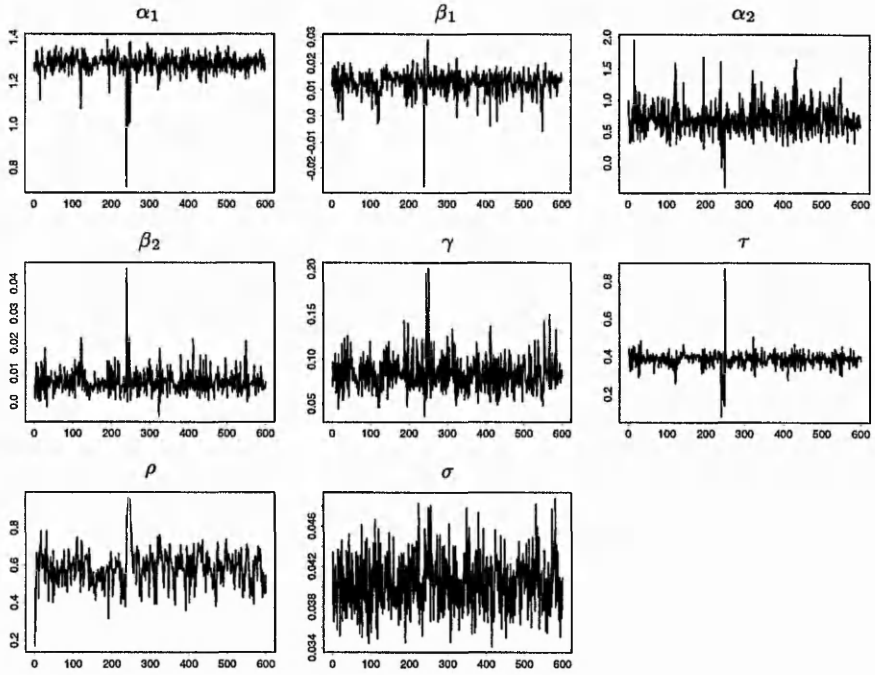


Figure 6.13: Simulated samples for the UK industrial production index series.

After the analysis of the convergence of the Markov chains, both six parallel chains and one long chain, we chose a burn-in of 1000 and ran a long single simulation of 50,000. We used this single simulation chain to obtain the results for the Bayesian analysis of the UK industrial production index series reported in the section 6.5.1.

For the US gross domestic production series, the Raftery and Lewis diagnostic suggests the burn-in should be under 10 and the length of the chain of about 1000 is enough. We also checked both quantiles $q = 0.025$ and $q = 0.975$. We list the results of the Raftery and Lewis diagnostic in Table 6.10. The application of the Raftery and Lewis diagnostic to other range of the samples (for example, samples between iteration 1000–2000, 2000–3000 *etc*) gives out very similar results, which suggests the adequateness of the recommended values of the burn-in M and length of the chain N. Low values of I in Table 6.10 for all parameters suggest the rapid

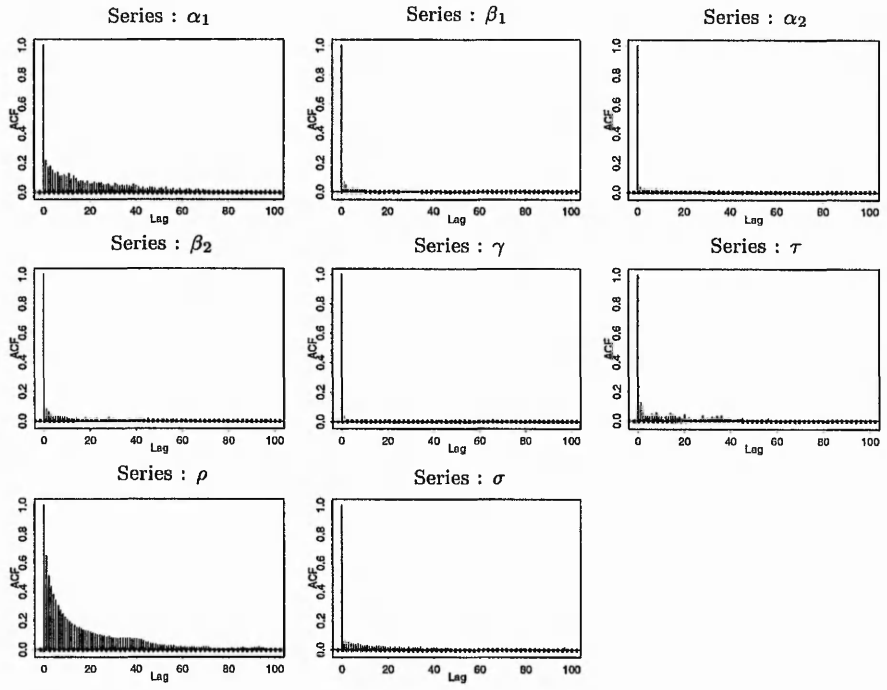


Figure 6.14: Autocorrelation function of the Markov chains for the UK industrial production index series.

convergence of the chains because of the low dependence. The simulated sample of the first 600 iterations for each parameter is plotted in Figure 6.15, which could help us to find lack of convergence if an obvious trend exists. The autocorrelation functions in the lag 100 are plotted in Figure 6.16, which show us the very low dependence of the chains, consistently with the results from the Raftery and Lewis diagnostic. We ran a single simulation of 50000 for the Bayesian analysis of the US GDP series.

	$P(U \leq u|\text{Data}) = 0.025$					$P(U \leq u|\text{Data}) = 0.975$				
	k	M	N	I	k_{ind}	k	M	N	I	k_{ind}
α_1	1	3	672	1.12	2	1	2	570	0.95	1
β_1	1	4	793	1.33	2	1	2	619	1.03	2
α_2	1	2	619	1.03	2	1	5	823	1.38	2
β_2	1	4	759	1.27	2	1	5	863	1.45	2
γ	1	6	1060	1.78	4	1	4	793	1.33	2
τ	1	3	730	1.22	2	1	3	700	1.17	2
σ	1	2	570	0.95	1	1	2	570	0.95	1

Table 6.10: Raftery and Rubin diagnostic for the US GDP series.

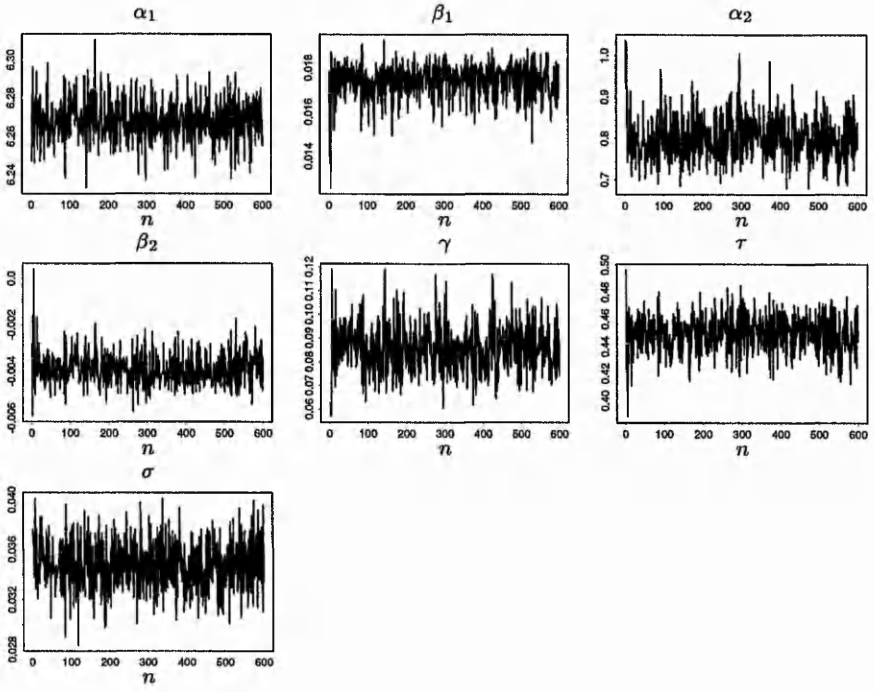


Figure 6.15: Simulated sample of the Markov chain for the US GDP series.

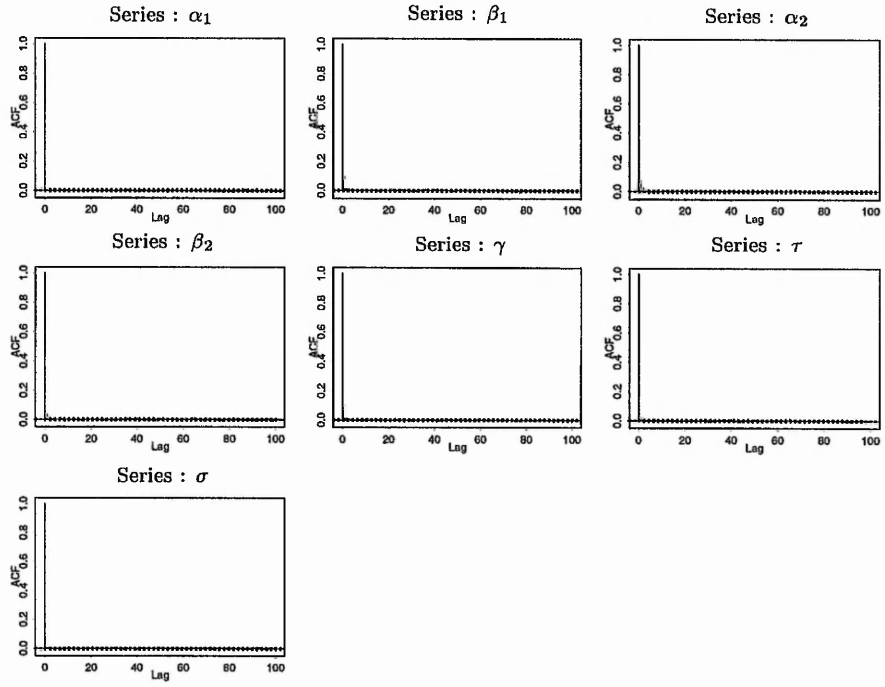


Figure 6.16: Autocorrelation functions of the Markov chain for the US GDP series.

Chapter 7

Double Smooth Transition

In this chapter, we discuss the double smooth transition model, which has been used to describe more complex models than those we discussed in the previous two chapters. This double smooth transition model has been used in the case of the global average temperature series by D. Harvey, 2000. We will discuss how to use Monte Carlo Markov chain methods to carry out a Bayesian analysis of the double smooth transition model.

The double smooth transition model is

$$y_t = \alpha_1 + \beta_1 t + \alpha_2 S_t^{(1)} + \beta_2 t S_t^{(1)} + \alpha_3 S_t^{(2)} + \beta_3 t S_t^{(2)} + \varepsilon_t, \tag{7.1}$$

where $S_t^{(i)}$ is the smooth transition function for the ith smooth transition component

$$S_t^{(i)} = \frac{1}{1 + \exp\{-\gamma_i (t - \tau_i T)\}}, \tag{7.2}$$

$0 < \tau_1 < \tau_2 < 1$. The disturbance terms ε_t, $t = 0, 1, \ldots, T$, are *i.i.d.* $\mathrm{N}(0, \sigma^2)$, where $\sigma > 0$.

7.1 Likelihood Function

We introduce the following matrix notation

$$X = \begin{pmatrix} 1 & 0 & S_0^{(1)} & 0 & S_0^{(2)} & 0 \\ 1 & 1 & S_1^{(1)} & S_1^{(1)} & S_1^{(2)} & S_1^{(2)} \\ & \cdots & \cdots & & & \\ 1 & T & S_T^{(1)} & T S_T^{(1)} & S_T^{(2)} & T S_T^{(2)} \end{pmatrix}, \tag{7.3}$$

which is a $(T+1) \times 6$ matrix, $\mathbf{y} = (y_0, y_1, \ldots, y_T)'$ is the sample data vector, $\boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T)'$ is the disturbance vector and $\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3)'$ is a sub-vector of the parameter vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \gamma_1, \tau_1, \gamma_2, \tau_2, \sigma)'$. For convenience, we denote another sub-vector as $\boldsymbol{\theta}_1 = (\gamma_1, \tau_1, \gamma_2, \tau_2)$ so we have $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\theta}_1, \sigma)'$. Then the double smooth transition model in matrix notation is

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{7.4}$$

It is straightforward to write the likelihood function as

$$p(\mathbf{y}|\boldsymbol{\theta}) \propto \frac{1}{\sigma^{T+1}} \exp\left\{-\frac{\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{2\sigma^2}\right\}, \tag{7.5}$$

where $\nu = T - 5$, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ and $\nu s^2 = (\mathbf{y} - X\hat{\boldsymbol{\beta}})'(\mathbf{y} - X\hat{\boldsymbol{\beta}})$.

7.2 Priors

We choose a conjugate prior for $\boldsymbol{\beta}$, conditional on the remaining parameters, as follows

$$p(\boldsymbol{\beta}|\boldsymbol{\theta}_1, \sigma) = \frac{|V|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^6} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}^0)' V^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta}^0)\right\}, \tag{7.6}$$

where V is a 6×6 positive definite symmetric matrix and $\boldsymbol{\beta}^0$ is a 6 dimensional vector. As we have no *a priori* information about the nature of the correlation between the elements of $\boldsymbol{\beta}$, we suppose V and $\boldsymbol{\beta}^0$ are constant.

For σ, we choose an inverted gamma distribution as prior

$$p(\sigma) = \frac{2}{\Gamma(b/2)} \left(\frac{a}{2}\right)^{b/2} \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\}, \tag{7.7}$$

where $a, b > 0$ are constants. Here we follow Monahan (1983) and Naylor and Marriott (1996) to suggest $a = b = 1/128$, $k_V = 16$, as we did in Chapter 6. We also suggest V is k_V^2 in the diagonal and zero elsewhere.

We know $\gamma_i > 0$, $i = 1, 2$. We again used the "diffuse" inverted gamma distribution as a prior for γ_1 and γ_2,

$$p(\gamma_i) = \frac{2}{\Gamma(b/2)} \left(\frac{a}{2}\right)^{b/2} \frac{1}{\gamma_i^{b+1}} \exp\left\{-\frac{a}{2\gamma_i^2}\right\}, \tag{7.8}$$

$i = 1, 2$, where a, b take the same values as in (7.7).

When we use double smooth transition model to fit data, we believe *a priori* that there are two smooth transitions in the time interval in which the data are observed and the two midpoints of the smooth transition are located within the same time interval, that is, $\tau_1, \tau_2 \in (0,1)$, where $\tau_1 < \tau_2$. We adopt a joint uniform distribution as prior for (τ_1, τ_2) conditional on $\tau_1 < \tau_2$ so that

$$p(\tau_1, \tau_2) = 2I_{\{\tau_1 < \tau_2\}}, \tag{7.9}$$

where $\tau_i \in (0,1)$, $i = 1,2$ and I is the indicator function $I_{\{\tau_1<\tau_2\}} = 1$, if $\tau_1 < \tau_2$; otherwise $I_{\{\tau_1<\tau_2\}} = 0$.

We assume $\boldsymbol{\theta}_1$ is *a priori* independent of $\boldsymbol{\beta}$ and σ and also the components of $\boldsymbol{\theta}_1$ are *a priori* independent of each other. From the above discussion, we have the joint prior for the parameter vector $\boldsymbol{\theta}$ as

$$\begin{aligned} p(\boldsymbol{\theta}) \propto & \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)'V^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}^0)\right\} \\ & \times \frac{1}{\sigma^{b+7}} \exp\left\{-\frac{a}{2\sigma^2}\right\} \prod_{i=1}^{2} \frac{1}{\gamma_i^{b+1}} \exp\left\{-\frac{a}{2\gamma_i^2}\right\}, \end{aligned} \tag{7.10}$$

where $\gamma_i > 0$, $i = 1,2$ and $0 < \tau_1 < \tau_2 < 1$.

7.3 Posteriors

Now that we have the likelihood function and the prior for the parameter vector of the double smooth transition model, we can obtain the posterior for $\boldsymbol{\theta}$ as

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) \propto & \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})'V_1(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}})\right\} \\ & \times \frac{1}{\sigma^{T+b+8}} \exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\} \prod_{i=1}^{2} \frac{1}{\gamma_i^{b+1}} \exp\left\{-\frac{a}{2\gamma_i^2}\right\}, \end{aligned} \tag{7.11}$$

where $V_1 = X'X + V^{-1}$, $\tilde{\boldsymbol{\beta}} = V_1^{-1}(X'X\hat{\boldsymbol{\beta}} + V^{-1}\boldsymbol{\beta}^0)$, $\nu s_1^2 = a + \nu s^2 + (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta}^0 - \hat{\boldsymbol{\beta}})$ and $\gamma_i > 0$, $i = 1,2$ and $0 < \tau_1 < \tau_2 < 1$.

We know from the above expression that the posterior for $\boldsymbol{\beta}$ is a multivariate normal distribution conditional on the other parameters and the posterior for σ is an inverted gamma distribution conditional on the other parameters.

7.4 ARMS method in Gibbs Sampling

We have seen from the above section that sampling from non-typical distributions is important in the practical application of Monte Carlo Markov chain. Gilks and Wild (1992) and Gilks, Best and Tan (1995) discussed two methods for sampling from full conditional distributions. One is the adaptive rejection sampling (ARS), which can be used to sample from log-concave distributions (where the logarithm of the density function is concave). The other is the adaptive rejection Metropolis sampling (ARMS) for sampling from non-log-concave distributions. In order to explain these two methods, we need first to explain what is the accept-rejection algorithm.

7.4.1 Accept-Rejection Algorithm

There are a lot of non-standard distributions from which sampling is difficult. For example, sampling from an inverted gamma distribution which has the form of (7.7) above is difficult if we would like to sample from (7.7) directly. The accept-rejection algorithm provide us a method to solve this kind of problems.

The accept-rejection algorithm can be described as follows. Suppose we would like to sample from the density $f(x)$ (called the target density). Instead of sampling directly from f, we sample from another density $g(x)$ (called the proposal density), which is easy to sample from, provided that there exists a constant $M > 0$ such that

$$f(x) \leq Mg(x) \tag{7.12}$$

on the support of $f(x)$ ($\{x : f(x) > 0\}$). Then the accept-rejection algorithm can be carried out in the following steps

1. Generate $X \sim g(x)$, $U \sim U(0,1)$;

2. Accept $Y = X$ if $U \leq f(X)/(Mg(X))$; Otherwise return to 1,

where $U(0,1)$ denotes a uniform distribution over $(0,1)$, then we have $Y \sim f(x)$.

7.4.2 Adaptive Rejection Sampling

Although it seems that the accept-rejection algorithm can be used to sample from any density, it is difficult to choose M and the proposal density g in practice. Arbitrary choosing of M and g will cause inefficiency in the sampling algorithm and

therefore is impracticable. The design of an accept-rejection algorithm will depend on the specific distribution density function. Gilks and Wild (1992) suggested an algorithm to sample from a log-concave density, called the Adaptive rejection sampling (ARS). ARS is based on the construction of an envelope and the derivation of a corresponding Accept-Rejection algorithm when the density function is log-concave.

Suppose we have a density $f(x)$, which is log-concave, that is, if we set $h(x) = \log(f(x))$, then $h(x)$ is concave ($h''(x) < 0$). Let S_n be a set of points x_i, $i = 0, 1, \ldots, n+1$, in the support of $f(x)$. Suppose we know $h(x_i) = \log(f(x_i))$ up to the same constant for all $i = 0, 1, \ldots, n+1$. Because $h(x)$ is concave, the segment between $(x_i, h(x_i))$ and $(x_{i+1}, h(x_{i+1}))$ is below the graph of $h(x)$ over the interval $[x_i, x_{i+1}]$ and is above the graph of $h(x)$ outside the interval $[x_i, x_{i+1}]$ (see Figure 7.1).

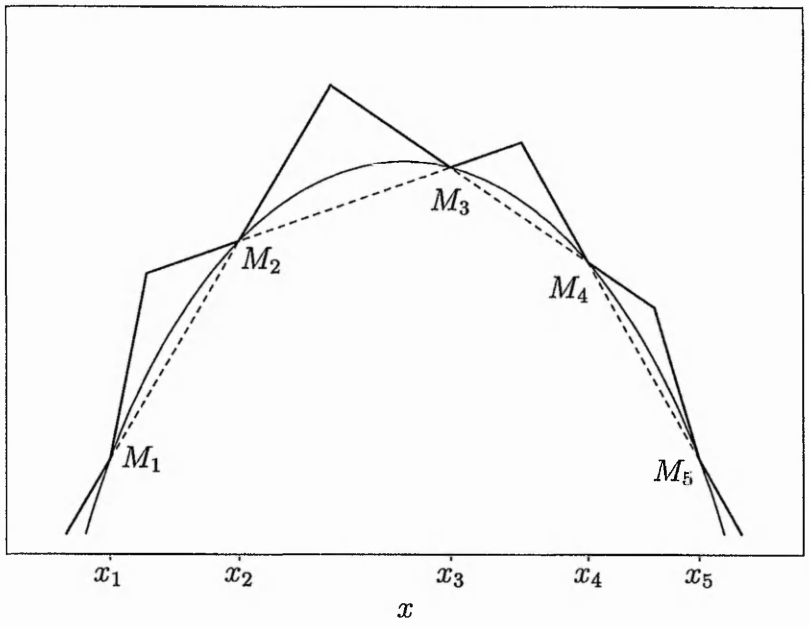


Figure 7.1: Lower and upper envelopes of a log-concave density.

We denote M_i, $i = 0, 1, \ldots, n$ are the points on the curve of $y = h(x)$ and $L_{i,i+1}(x)$ is the line through the two points (M_i, M_{i+1}). For $x \in [x_i, x_{i+1}]$, we define

$$\bar{h}_n(x) = \min\{L_{i-1,i}(x), L_{i+1,i+2}(x)\}, \tag{7.13}$$

$$\underline{h}_n(x) = L_{i,i+1}(x) \tag{7.14}$$

on the interval $[x_0, x_{n+1}]$. For $x \notin [x_0, x_{n+1}]$, we also define

$$\bar{h}_n(x) = \min\{L_{0,1}(x), L_{n,n+1}(x)\}, \tag{7.15}$$

$$\underline{h}_n(x) = -\infty, \tag{7.16}$$

then we have

$$\underline{h}_n(x) \leq h(x) \leq \bar{h}_n(x). \tag{7.17}$$

We call $\underline{h}_n(x)$ and $\bar{h}_n(x)$ the low and the upper envelopes for $h(x)$ respectively. Therefore, if we define $\underline{f}_n(x) = \exp(\underline{h}_n(x))$ and $\bar{f}_n(x) = \exp(\bar{h}_n(x))$, then $\underline{f}_n(x)$ and $\bar{f}_n(x)$ are the lower and the upper envelopes for $f(x)$ respectively,

$$\underline{f}_n(x) \leq f(x) \leq \bar{f}_n(x). \tag{7.18}$$

If we define C_n as the normalisation constant of $\bar{f}_n(x)$ so that $\bar{f}_n(x) = C_n g_n(x)$ and $g_n(x)$ is a proper density, then the ARS algorithm is as follows

1. Initialise n and S_n;
2. Generate $X \sim g_n(x)$, $U \sim U(0,1)$;
3. If $U \leq \underline{f}_n(X)/(C_n g_n(X)$, $Y = X$; otherwise, if $U \leq f_n(X)/(C_n g_n(X)$, $Y = X$; otherwise update S_n to $S_{n+1} = S_n \cup \{X\}$ and go to step 2 until enough sample data has been accepted.

Then we have $Y \sim f(x)$.

7.4.3 Adaptive Rejection Metropolis Sampling

Gilks, Best and Tan (1995) generalised the ARS method to deal with non-log-concave density. Gilks, Best and Tan called the new method the adaptive rejection Metropolis sampling (ARMS) (see Gilks, Best and Tan (1995) for details). Here we outline their algorithm.

We denote by $S_n = \{x_i; i = 0, 1, \ldots, n+1\}$ the current set of abscissae in ascending order, that is, $x_0 < x_1 < \ldots < x_{n+1}$, where x_0 and x_{n+1} are the lower and upper limits of the domain D of the support of $f(x)$. The lower and upper limits of D could be infinite if $f(x)$ is not bounded on the left side and right side of D respectively. Denote $h(x) = \ln(f(x))$.

Let $L_{ij}(x; S_n)$ denote the line between the two points $(x_i, h(x_i))$ and $(x_j, h(x_j))$, $j = i+1$, $i = 0, 1, \ldots, n$. Define a piecewise linear function $h_n(x)$ on (x_0, x_{n+1}) as

$$h_n(x) = \max\{L_{i,i+1}(x, S_n), \min\{L_{i-1,i}(x; S_n), L_{i+1,i+2}(x; S_n), \}\}, \tag{7.19}$$

when $x_i \leq x < x_{i+1}$, where $\min\{a, b\} = a$ if b is not defined. And we define

$$g_n(x) = \frac{1}{m_n} \exp(h_n(x)), \tag{7.20}$$

where

$$m_n = \int \exp(h_n(x))\, dx. \tag{7.21}$$

The ARMS algorithm is then as follows

1. initialise n and S_n independently of X_0, where X_0 is the current sample from $f(x)$;
2. Sample X from $g_n(x)$, $U \sim U(0, 1)$;
3. If $U > f(X)/\exp(h_n(x))$, then
   (a) Set $S_{n+1} = S_n(x) \bigcup \{X\}$;
   (b) Relabel points in S_{n+1} in ascending order;
   (c) n=n+1;
   (d) Go back to step 2;
4. Else accept X and go to next step;
5. S
6. If $U > \min\left\{1, \frac{f(X)\min\{f(X_0), \exp(h_n(X_0))\}}{f(X_0)\min\{f(X), \exp(h_n(X))\}}\right\}$, then
6. If $U > \min\left\{1, \frac{f(X)\min\{f(X_0), \exp(h_n(X_0))\}}{f(X_0)\min\{f(X), \exp(h_n(X))\}}\right\}$, then
   (a) Set $X_m = X_0$;
   (b) Else set $X_m = X$;
7. Return X_m;

Through the above steps we generate a sample X_m from $f(x)$. Repeat the above steps we generate a rejection sampling chain proposed by Tierney (1991). Tierney (1994) discussed the convergence property of this "hybrid" adaptive chain.

C programmes to carry out the ARMS algorithm can be found in the web site of MRC Biostatistics Unit, Cambridge. I use these subroutines within my own C++ programmes to implement our approach for the double smooth transition model.

7.4.4 ARMS for Double Smooth Transition Model

Now we apply the ARMS algorithm to our double smooth transition model. With the discussions of the general ARMS algorithm and the posterior density function for the parameter vector $\boldsymbol{\theta}$ in the double smooth transition model given in the previous sections, we can construct a hybrid of the Gibbs sampling algorithm as follows.

As is usual in Gibbs sampling, we denote $\boldsymbol{\theta}^{(n)}$ as the current sample. Given the current sample $\boldsymbol{\theta}^{(n)}$, we would like to obtain the next sample $\boldsymbol{\theta}^{(n+1)}$ in the following steps:

1. Update νs_1^2, $\tilde{\boldsymbol{\beta}}$ and V_1^{-1} with values of $\boldsymbol{\theta}_1^{(n)}$;
2. Sample $\boldsymbol{\beta}^{(n+1)}$ from $p(\boldsymbol{\beta}|\boldsymbol{\theta}_1^{(n)}, \sigma^{(n)}, \mathbf{y})$ which is a multivariate normal distribution;
3. Sample $\sigma^{(n+1)}$ from $p(\sigma|\boldsymbol{\beta}^{(n+1)}, \boldsymbol{\theta}_1^{(n)}, \mathbf{y})$ which is an inverted gamma distribution;
4. Sample $\gamma_1^{(n+1)}$ from $p(\gamma_1|\boldsymbol{\beta}^{(n+1)}, \tau_1^{(n)}, \gamma_2^{(n)}, \tau_2^{(n)}, \sigma^{(n+1)})$ using the ARMS sampler;
5. Sample $\gamma_2^{(n+1)}$ from $p(\gamma_2|\boldsymbol{\beta}^{(n+1)}, \gamma_1^{(n+1)}, \tau_1^{(n)}, \tau_2^{(n)}, \sigma^{(n+1)})$ using the ARMS sampler
6. Sample $\tau_1^{(n+1)}$ from $p(\tau_1|\boldsymbol{\beta}^{(n+1)}, \gamma_1^{(n+1)}, \gamma_2^{(n+1)}, \tau_2^{(n)}, \sigma^{(n+1)})$, where $\tau_1 \in (0, \tau_2^{(n)})$, using the ARMS sampler;
7. Sample $\tau_2^{(n+1)}$ from $p(\tau_2|\boldsymbol{\beta}^{(n+1)}, \gamma_1^{(n+1)}, \tau_1^{(n+1)}, \gamma_2^{(n+1)}, \sigma^{(n+1)})$, where $\tau_2 \in (\tau_1^{(n+1)}, 1)$, using the ARMS sampler;
8. Update $n = n + 1$;

7.5 Empirical Applications

We now apply our algorithm to three global average temperature data sets (obtained from the Climatic research unit, University of East Anglia, UK). These temperature series are a combination of land air temperature anomalies (Jones (1994)) and sea surface temperature anomalies (Parker, Folland and Jackson (1995)) on a $5°\times5°$ grid-box basis. The combination of the land air temperature and sea surface temperature data sets was discussed in Parker, Jones, Bevan and Folland (1994). These average global temperature data sets are widely used in the research of the international climate change. We use the annual average hemispheres (Northern, Southern) and global temperature series provided by Jones, Osborn and Briffa (1997) from 1856 to 1998, 143 observations.

7.5.1 the temperature data sets

The annual average temperature data sets of the two hemispheres (Southern and Northern) and the global average temperature data sets are studied using different methods. Harvey and Mills (2000) tested the unit roots against the alternative of a stationarity around a double smooth transitions in trend. Harvey and Mills test their hypothesis for two time series, the global average temperature series. When the null unit roots was rejected, double smooth transition model was fitted to these series. We follow them to fit the double smooth transition model to the three temperature series, the global average temperature and two hemisphere temperature series. We look at the results of the Bayesian analysis of the double smooth transition model for these three series first, leaving the discussion of the convergence of the Markov chains later with the discussion of the performance of the ARMS algorithm.

The posterior means and standard deviations for each parameter in the double smooth transition model for these three temperature data sets are listed in Table 7.1.

We also plotted the posterior densities of all the parameters for the Global average, the Northern and the Southern Hemisphere models in Figure 7.2, Figure 7.3 and Figure 7.4 respectively. The posterior densities for γ_1 and γ_2 in all the three temperature models show that the algorithm can successfully estimate the different

	Global average		Northern Hemisphere		Southern Hemisphere	
	mean	sd	mean	sd	mean	sd
α_1	-0.367	0.0471	-0.234	0.0917	-0.443	0.0776
β_1	0.00414	0.00477	0.0022	0.00983	0.00975	0.0076
α_2	-1.28	0.666	-1.26	0.982	-1.24	0.698
β_2	0.0182	0.0111	0.0188	0.0171	0.00918	0.00695
α_3	-1.49	0.874	-2.43	1.17	0.349	0.644
β_3	0.00244	0.0107	0.0098	0.0144	-0.00755	0.00732
γ_1	0.523	0.516	0.466	0.504	0.307	0.423
τ_1	0.355	0.0441	0.384	0.0737	0.336	0.0466
γ_2	0.118	0.071	0.157	0.213	1.47	0.967
τ_2	0.717	0.0476	0.756	0.0561	0.636	0.0149
σ	0.103	0.00623	0.131	0.00785	0.097	0.00598

Table 7.1: Posterior means and standard deviations for three temperature data sets.

ranges of support for γ_1 and γ_2. In Figure 7.2, for the Global average temperature, the range of support for γ_2 is much narrower than that of γ_1 and the same is from in Figure 7.3 for the Northern Hemisphere temperature whereas in Figure 7.4 for the Southern Hemisphere the range of support for γ_2 is much larger than that of γ_1. The speeds of the two smooth transitions for each data set are different. For the Global average temperature, the speed of the first smooth transition is much larger than that of the second. The ratio of the two speeds is $\gamma_1/\gamma_2 = 4.43$. For the Northern Hemisphere temperature, the ratio of the two speeds is $\gamma_1/\gamma_2 = 2.97$ and is obviously less than the ratio for the Global average temperature. For the Southern Hemisphere temperature, the ratio of the two speeds is $\gamma_1/\gamma_2 = 0.209$, which is much less than 1. The second speed is much larger than the first speed for the Southern Hemisphere.

We can see from Table 7.1 that the proportions of the middle point of the first smooth transition, τ_1, are 0.355, 0.384 and 0.336 for the Global average, the Northern Hemisphere and the Southern Hemisphere respectively. The corresponding midpoints for the first smooth transition are in the years 1906, 1910 and 1903. The

proportions of the middle point of the second smooth transition, τ_2, are 0.717, 0.756 and 0.636 for the Global average, the Northern Hemisphere and the Southern Hemisphere respectively. The corresponding midpoints for the second smooth transition are in the years 1957, 1963 and 1946. So it would appear that both the first and second transitions for the Global average temperature had their midpoints between the first and second transitions for the Northern Hemisphere and the Southern Hemisphere temperature respectively.

We list the range of years in which 95% of the two transitions (from $S_t(\gamma, \tau) = 0.025$ to $S_t(\gamma, \tau) = 0.975$) would have probably taken place for the three temperature data sets in Table 7.2.

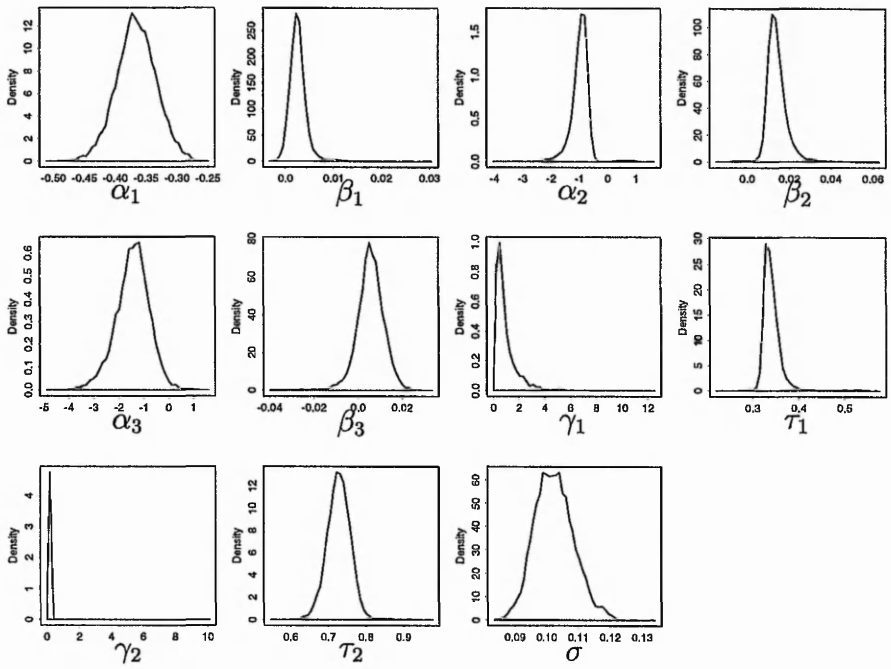


Figure 7.2: Posterior densities for the Global average temperature.

We obtained quantiles for the posterior densities for the parameters of the Global average, the Northern and the Southern Hemisphere which can be used to construct posterior probability intervals for each parameter. We list these quantiles for all the parameters of the Global average, the Northern and the Southern Hemisphere in Table 7.3, Table 7.4 and Table 7.5 respectively.

We have discussed how to obtain the predictive density of $\tilde{y}_t$ in Chapter 6, see

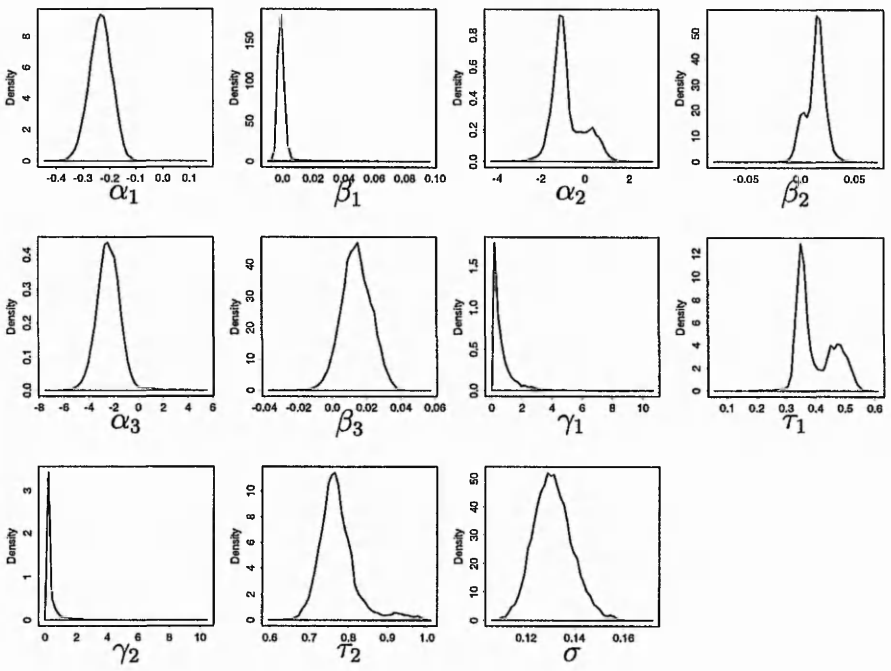


Figure 7.3: Posterior densities for the Northern Hemisphere temperature.

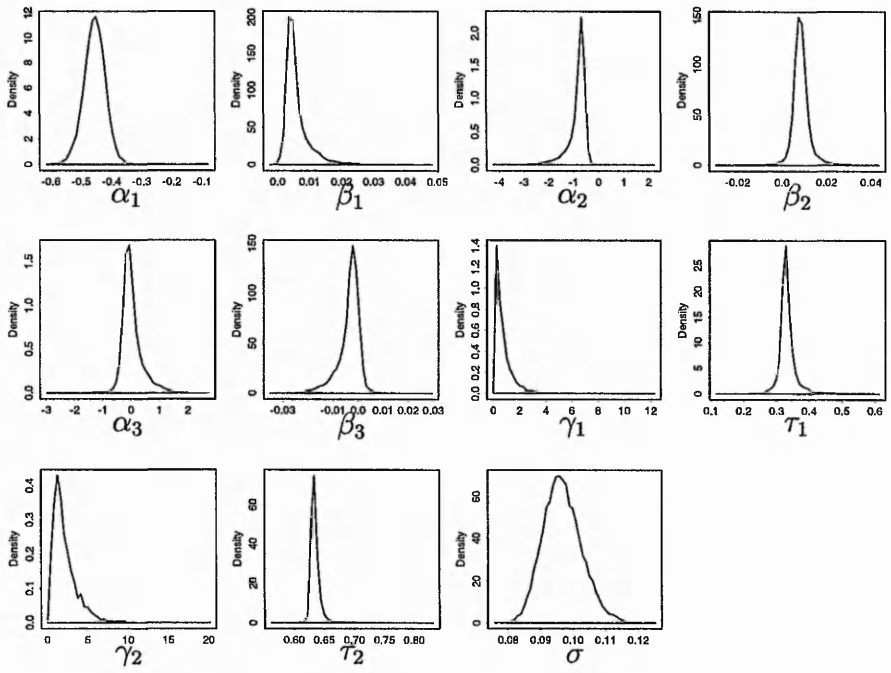


Figure 7.4: Posterior densities for the Southern Hemisphere temperature.

	Transition	From (year)	To (year)	Range (years)
Global	First	1899	1913	14
average	Second	1926	1988	62
Northern	First	1902	1918	16
Hemisphere	Second	1940	1986	47
Southern	First	1891	1915	24
Hemisphere	Second	1943	1948	5

Table 7.2: Range of years 95% of two transition take place.

	0.025	0.05	0.25	0.5	0.75	0.95	0.975
α_1	-0.449	-0.432	-0.395	-0.368	-0.341	-0.301	-0.288
β_1	-0.00119	-0.000583	0.00155	0.00319	0.00521	0.012	0.0156
α_2	-2.96	-2.56	-1.54	-1.14	-0.894	-0.611	-0.458
β_2	0.00119	0.00555	0.0121	0.0163	0.0221	0.0381	0.0465
α_3	-3.35	-3	-2.02	-1.46	-0.946	-0.0987	0.246
β_3	-0.0247	-0.0168	-0.00188	0.00411	0.00893	0.0158	0.0186
γ_1	0.021	0.0421	0.168	0.364	0.682	1.53	1.93
τ_1	0.292	0.309	0.332	0.346	0.366	0.445	0.485
γ_2	0.0103	0.0206	0.0761	0.112	0.155	0.217	0.238
τ_2	0.613	0.642	0.693	0.72	0.744	0.78	0.795
σ	0.0916	0.0933	0.0985	0.103	0.107	0.114	0.116

Table 7.3: Posterior quantiles for parameters of the Global average temperature.

	0.025	0.05	0.25	0.5	0.75	0.95	0.975
α_1	-0.353	-0.339	-0.281	-0.241	-0.2	-0.134	-0.0941
β_1	-0.00635	-0.00518	-0.00225	0.000263	0.00309	0.0162	0.0273
α_2	-3.38	-2.83	-1.73	-1.27	-0.873	0.595	0.958
β_2	-0.0118	-0.00625	0.0103	0.0188	0.0272	0.0431	0.0507
α_3	-4.83	-4.36	-3.14	-2.42	-1.73	-0.579	0.013
β_3	-0.0188	-0.0116	0.00325	0.0108	0.0181	0.0289	0.0328
γ_1	0.0197	0.0395	0.174	0.322	0.616	1.43	1.87
τ_1	0.222	0.278	0.345	0.372	0.429	0.513	0.531
γ_2	0.00633	0.0127	0.0633	0.124	0.18	0.333	0.55
τ_2	0.654	0.679	0.725	0.752	0.779	0.848	0.924
σ	0.116	0.119	0.126	0.131	0.136	0.145	0.148

Table 7.4: Posterior quantiles for parameters of the Northern Hemisphere temperature.

	0.025	0.05	0.25	0.5	0.75	0.95	0.975
α_1	-0.549	-0.531	-0.487	-0.453	-0.418	-0.331	-0.25
β_1	0.00137	0.00195	0.00466	0.00755	0.0126	0.0243	0.0297
α_2	-3.05	-2.67	-1.59	-1.02	-0.73	-0.48	-0.425
β_2	-0.00477	-0.000451	0.00583	0.00887	0.0124	0.0209	0.0245
α_3	-0.632	-0.439	-0.0994	0.216	0.734	1.56	1.83
β_3	-0.0245	-0.0215	-0.0121	-0.00598	-0.00213	0.00166	0.00324
γ_1	0.0172	0.0343	0.172	0.313	0.593	1.36	1.76
τ_1	0.248	0.271	0.315	0.332	0.353	0.417	0.446
γ_2	0.157	0.254	0.811	1.27	1.93	3.38	3.89
τ_2	0.615	0.62	0.627	0.634	0.64	0.655	0.671
σ	0.0861	0.0876	0.0928	0.0967	0.101	0.107	0.11

Table 7.5: Posterior quantiles for parameters of the Southern Hemisphere temperature.

equation (6.20). We obtained two quantiles with cumulative probability 0.025 and 0.975 respectively for $\tilde{y}_t$, $t = 0, 1, \ldots, T$, predictive mean of $\tilde{y}_t$, $t = 0, 1, \ldots, T$ using the sample from the posterior distribution and plotted them with the actual three temperature series in Figure 7.5, Figure 7.6 and Figure 7.7 respectively. The quantiles with cumulative probability 0.025 and 0.975 consist of the predictive 95% probability intervals for the unobserved $\tilde{y}_t$, $t = 0, 1, \ldots, T$.

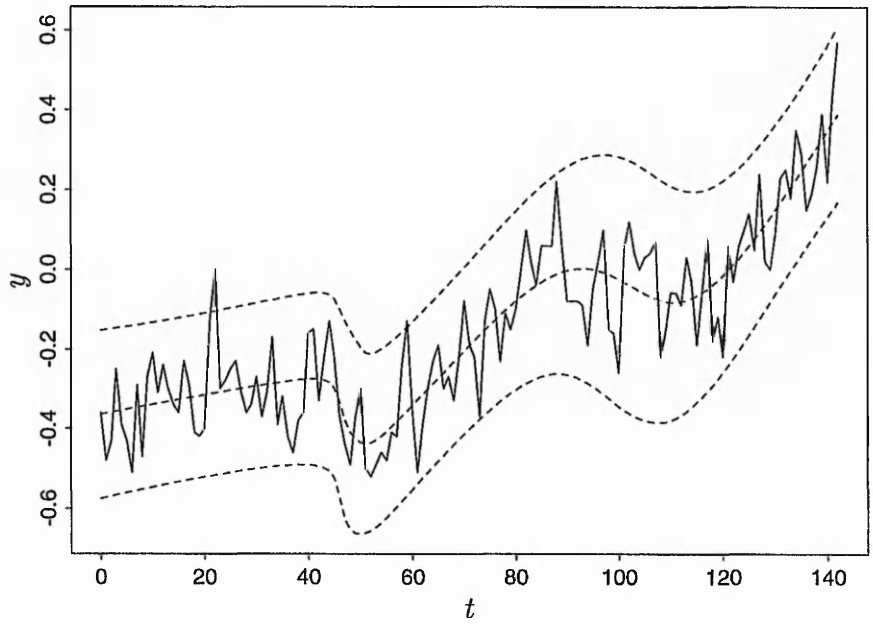


Figure 7.5: Predictive 95% probability intervals, predictive means and the actual Global average temperature.

We can also obtain predictive densities for the future temperatures ($t > T$). We have obtained the first step ($t = T+1$) predictive densities for the three temperature models. The predictive mean, standard deviation and quantiles at the future time $t = T + 1$ for the three temperature models are presented in Table 7.6.

The posterior means of the two smooth transition function $S_t(\gamma_i, \tau_i)$, $i = 1, 2$ at points $t = 0, 1, \ldots, T$ for the Global average, the Northern and the Southern temperature are plotted in Figure 7.8, Figure 7.9 and Figure 7.10 respectively. We can compare these graphs with the calculations in Table 7.2. From the graphs it is obvious that the second transition of the Southern Hemisphere carries out in a short

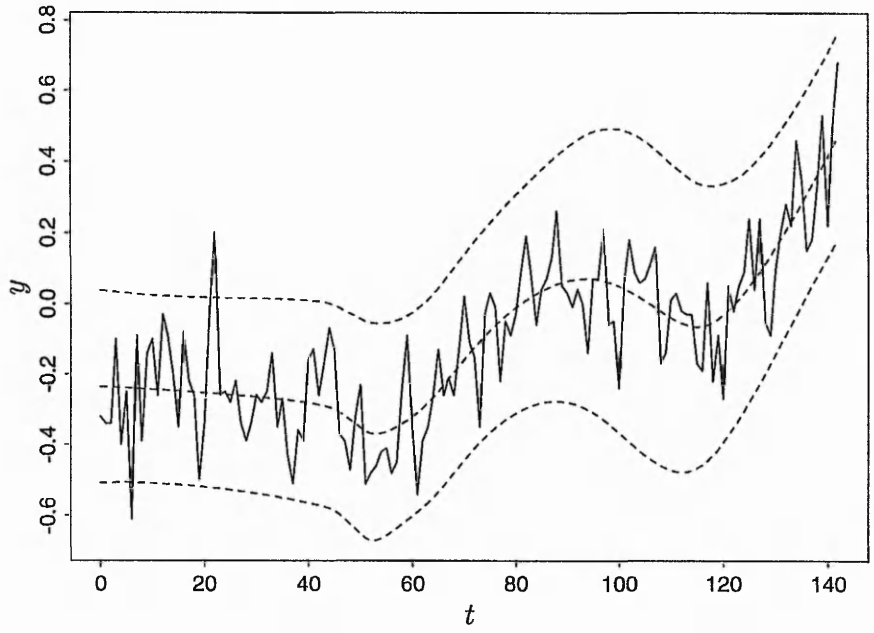


Figure 7.6: Predictive 95% probability intervals, predictive means and the actual Northern Hemisphere temperature.

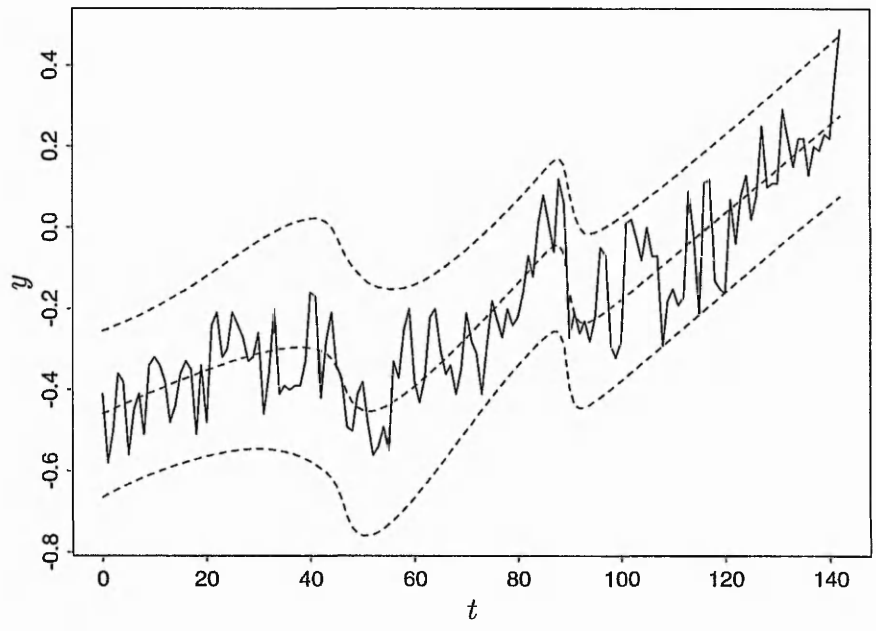


Figure 7.7: Predictive 95% probability intervals, predictive means and the actual Southern Hemisphere temperature.

	Global average	Northern Hemisphere	Southern Hemisphere
mean	0.414	0.495	0.286
sd	0.113	0.15	0.101
cumulative probabilities		quantiles	
0.025	0.192	0.205	0.0875
0.05	0.228	0.252	0.12
0.25	0.338	0.395	0.218
0.5	0.414	0.493	0.286
0.75	0.489	0.592	0.354
0.95	0.599	0.74	0.453
0.975	0.635	0.792	0.486

Table 7.6: First step predictive mean, standard and quantiles of the Global average, the Northern and the Southern Hemisphere temperature.

time (5 years for 95%) while the second transition of the Global average temperature carries out in a comparative longer time (62 years for 95%).

7.6 Reparameterisation

Reparameterisation has been considered in the MCMC literature as a way to speed up convergence in a Monte Carlo Markov chain sampler. It can be used in a Gibbs sampler, a Metropolis-Hastings algorithm or in a hybrid sampler. The general idea of the reparameterisation is to use transformations of the parameters to reduce correlations between parameters of the target distribution so as to produce rapid mixing Markov chain. A simple extreme example can be found in Robert and Casella (1999), Example 7.1.10 on page 294. A more detail review of reparameterisation can be found in Gilks and Roberts (1996). General methods of reparameterisation still need be addressed with more efforts and the performance of the present methods in complex models and high dimensions is still unknown and is a matter for experimentation on a case by case basis.

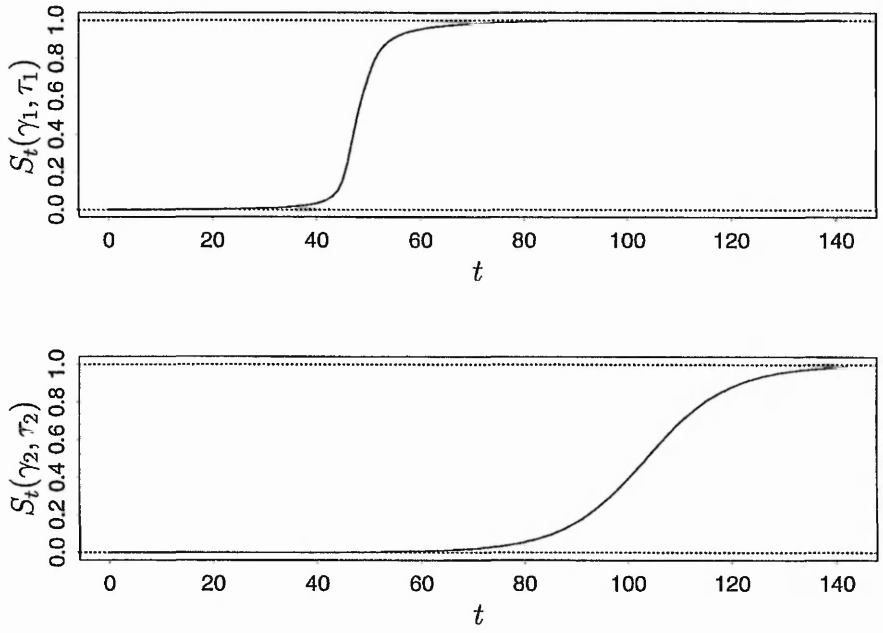


Figure 7.8: Posterior mean of smooth transition function for the Global average temperature.

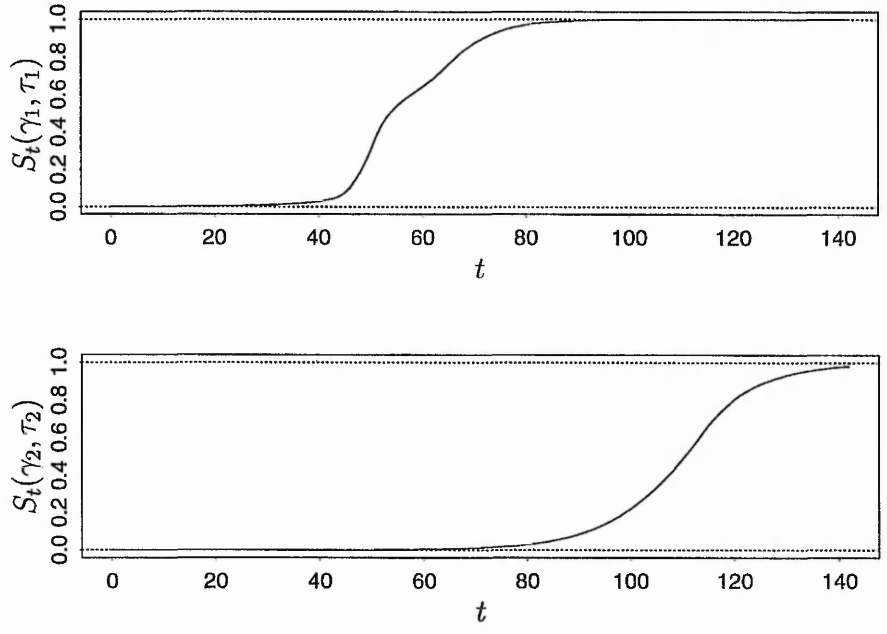


Figure 7.9: Posterior mean of smooth transition function for the Northern Hemisphere temperature.

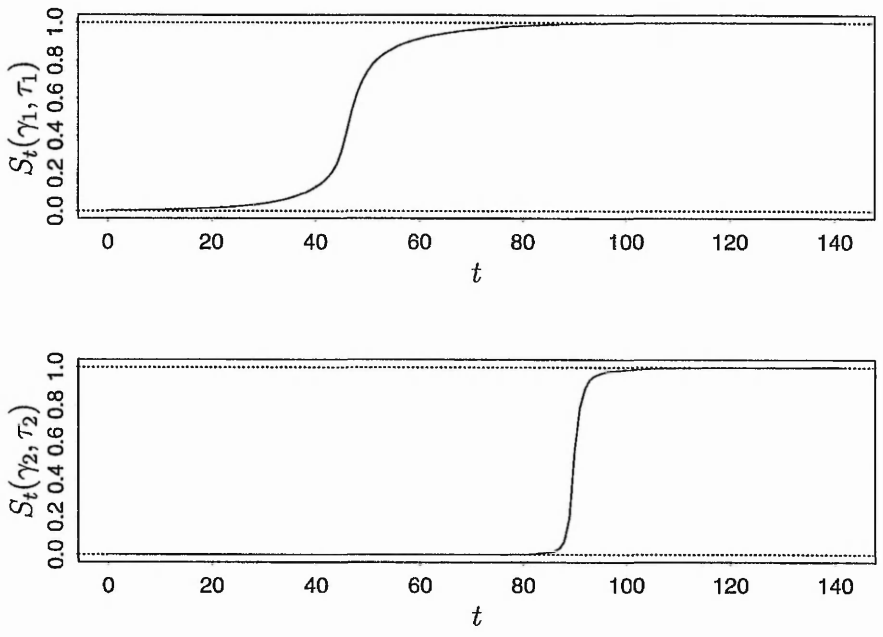


Figure 7.10: Posterior mean of smooth transition function for the Southern Hemisphere temperature.

For our double smooth transition model, we know that all reasonable values for γ are within the interval $(0, 10)$, or more precisely $(0.01, 10)$ (see Table 4.1 in Chapter 4). $\gamma = 0.01$ corresponds to an "approximate" straight line within the time interval $t \in (0, 100)$ and $\gamma = 10$ corresponds to an "abrupt" jump that happens within $t = \tau T \pm 1$. In theory the support for the posterior for γ is $(0, \infty)$. Our problem is how to sample from the posterior distribution for γ with infinite support interval without imposing any unreasonable limits on γ. We also need to achieve good mixing in the Markov chain, that is, the sample path of γ should traverse its support quickly. To address all of these, we consider the transformation, $v = \exp(-\gamma/2)$. With this transformation, we know that $v \in (0, 1)$, which is a limited interval and can be easily sampled from with any MCMC sampler. A more important point is that the range of v corresponding to the reasonable range of γ discussed above, will cover more than 99% of the interval $(0, 1)$ for v, with $\gamma = 0.01$ corresponding to $v = 0.995$ and $\gamma = 10$ corresponding to $v = 0.0067$. By introducing this transformation, we impose no limits upon the range of γ. In our applications above, we found this

reparameterisation gave as good mixing of the chain for γ as for other parameters.

In our sampling algorithm for the double smooth transition model, we have used this reparameterisation for γ_1 and γ_2 so that $v_1 = \exp(-\gamma_1/2)$ and $v_2 = \exp(-\gamma_2/2)$ and used ARMS method to sample for v_1 and v_2 with left bound 0 and right bound 1 which are required. So in fact we imposed no limits on the range of γ_i, $i = 1, 2$. The discussion of the reasonable range of γ_i, $i = 1, 2$ help us to make sure that our transformations of γ_1 and γ_2 are suitable and the Markov chains for γ_1 and γ_2 converge as well as those of other parameters.

7.7 Performance of ARMS

We use the hybrid Monte Carlo Markov chain method, the adaptive rejection Metropolis sampling within Gibbs Sampling, to sample from the posterior density of the parameter vector. The adaptive rejection Metropolis sampling gives us a "proposal" distribution in the Metropolis step automatically while in fact sometimes we have little idea about the true distribution in each Metropolis step and so it is difficult to choose a "proposal" distribution which is "similar" to the true distribution. Gibbs sampling is a better algorithm then Metropolis-Hastings because Gibbs sampling has no rejection step and is therefore more efficient if the conditional distribution can easily be sampled from directly. However in practice applications, sometimes we meet both conditional distributions for some parameters that are easily sampled from and others that are not. In these cases the hybrid Monte Carlo Markov chain method is an obvious choice. A hybrid Monte Carlo Markov chain algorithm is a MCMC method which simultaneously utilises both Gibbs sampling and Metropolis-Hastings steps. The advantages of the hybrid algorithm is that: if the conditional distributions for some of the parameters are standard distributions and we know how to sample from them directly, then we use Gibbs steps; however if the conditional distributions for other parameters are non-standard distributions that cannot be sampled from directly, then we use Metropolis-Hastings steps. Tierney (1994) discussed the sufficient conditions for the uniform ergodicity of the chain and proved that the irreducibility and aperiodicity of the chain generated by a hybrid method follow directly from the irreducibility and aperiodicity of each of the

Gibbs and Metropolis-Hastings steps. Nobile(1998) showed how the introduction of a Metropolis-Hastings step in the Gibbs steps speeds up the exploration of the support of the stationary distribution.

In our double smooth transition model, the conditional posterior distributions for α_1, β_1, α_2, β_2, α_3, β_3 are normal distributions and for σ is an inverted gamma distribution, all of which are easily sampled from. For these parameters we choose the Gibbs steps to sample from their conditional posterior distributions. The conditional posterior distributions for γ_1, τ_1, γ_2 and τ_2 are non-standard distributions and in fact the expressions for the conditional posterior distributions are too complex to easily explore their shapes. It is hard to give "reasonable" proposal distribution for them in order to use standard Metropolis-Hastings steps. In this case, we choose the adaptive rejection Metropolis sampling steps in order to sample from these conditional posterior distributions.

In general cases of diagnostic of the convergence of the chain, we need to determine the burn-in, from where the chain begins to converge, the number of iterations needed to achieve the specified precision. We generate 100,000 samples from the posterior and use the Raftery and Lewis (1992) method to find the burn-in and length of the chain needed to achieve enough accurate estimation of the two primary quantiles $q = 0.025$ and $q = 0.975$. In the Raftery and Lewis diagnostic, we have statistics k, M, N, I and k_{ind} as thinning lag to make the chain first order Markov, the burn-in, the number of length, measurement of dependence and thinning lag to make the chain into an independent chain respectively (see Chapter 6, section 6.3.1). For all the three temperature series, Raftery and Lewis diagnostic suggests the burn-in under 20, the number of length of the chain under or around 2000 for all parameters. However, when we diagnose the convergence of the chain for the Southern Hemisphere series, we find the statistic I of the chain for γ_2 between iteration 1 to 600 is 4.97, which is very near 5. Raftery and Lewis suggest that if I greater than 5 often indicate problems. So we discard the sample between iteration 1 to 600. We present the Raftery and Lewis diagnostic results for the Global average (iteration 1 to 600), the Southern Hemisphere (iteration 1 to 600) and the Southern Hemisphere (iteration 601 to 1200) in Table 7.7, Table 7.8 and Table 7.9 respectively. We apply the Raftery and Lewis diagnostic to the chains of these three

temperature series for the sequential chains and obtain similar results as in Table 7.7, Table 7.8 and Table 7.9. So we chose 600 as the burn-in, 6000 as the number of length of the chain to make a Bayesian analysis of the double smooth transition model for all the three temperature series.

	$P(U \leq u\|\text{Data}) = 0.025$					$P(U \leq u\|\text{Data}) = 0.975$				
	k	M	N	I	k_{ind}	k	M	N	I	k_{ind}
α_1	1	2	571	0.96	1	1	4	751	1.26	2
β_1	1	4	751	1.26	2	1	5	863	1.45	2
α_2	1	3	654	1.10	2	1	11	1908	3.20	4
β_2	1	11	1908	3.20	4	1	9	1595	2.67	5
α_3	1	3	654	1.10	2	1	3	654	1.10	2
β_3	3	15	2598	4.36	5	1	5	996	1.67	4
γ_1	1	4	751	1.26	2	1	5	863	1.45	2
τ_1	1	5	996	1.67	2	1	11	1908	3.20	4
γ_2	1	6	1156	1.94	3	1	4	751	1.26	2
τ_2	1	9	1595	2.67	3	1	9	1595	2.67	3
σ	1	2	536	0.90	1	1	4	751	1.26	2

Table 7.7: Raftery and Lewis diagnostic for the Global average temperature series.

The sample paths of the first 600 iterations are plotted in Figure 7.11, Figure 7.12 and Figure 7.13 for the Global average, the Northern and the Southern Hemisphere temperature series respectively. No obvious evidence of non-convergence of the chains is found in these sample paths. The autocorrelation functions of the first 100 iterations are plotted in Figure 7.14, Figure 7.15 and Figure 7.16 for the Global average, the Northern and the Southern Hemisphere temperature respectively. These autocorrelation functions provide measurements of the dependence within the chain for each parameters. All these autocorrelation functions decrease rapidly especially when the lag is larger than 20. The rapid decreasing of the autocorrelation functions suggest that the convergence of the chains is quick.

	$P(U \leq u\|\text{Data}) = 0.025$					$P(U \leq u\|\text{Data}) = 0.975$				
	k	M	N	I	k_{ind}	k	M	N	I	k_{ind}
α_1	1	2	571	0.96	1	1	5	996	1.67	2
β_1	1	3	654	1.10	2	1	6	1156	1.94	2
α_2	1	2	571	0.96	1	1	5	996	1.67	2
β_2	1	4	751	1.26	2	1	8	1351	2.26	3
α_3	1	3	654	1.10	2	1	3	654	1.10	2
β_3	1	4	751	1.26	2	1	4	751	1.26	2
γ_1	1	5	863	1.45	2	1	2	571	0.96	1
τ_1	1	5	863	1.45	2	1	5	996	1.67	2
γ_2	1	4	751	1.26	2	1	4	751	1.26	2
τ_2	1	9	1595	2.67	3	1	8	1351	2.26	3
σ	1	2	536	0.90	1	1	2	571	0.96	1

Table 7.8: Raftery and Lewis diagnostic for the Northern Hemisphere temperature series.

	$P(U \leq u\|\text{Data}) = 0.025$					$P(U \leq u\|\text{Data}) = 0.975$				
	k	M	N	I	k_{ind}	k	M	N	I	k_{ind}
α_1	1	2	571	0.96	1	1	5	996	1.67	2
β_1	1	2	571	0.96	1	1	5	863	1.45	2
α_2	1	3	654	1.10	2	1	4	751	1.26	2
β_2	1	3	654	1.10	2	1	3	654	1.10	2
α_3	1	5	996	1.67	3	1	3	654	1.10	2
β_3	1	3	654	1.10	2	1	6	1156	1.94	2
γ_1	1	6	1156	1.94	2	1	3	654	1.10	2
τ_1	1	3	654	1.10	2	1	2	571	0.96	1
γ_2	1	4	751	1.26	2	1	3	654	1.10	2
τ_2	1	5	863	1.45	2	1	11	1908	3.20	4
σ	1	3	654	1.10	2	1	2	571	0.96	1

Table 7.9: Raftery and Lewis diagnostic for the Southern Hemisphere temperature series.

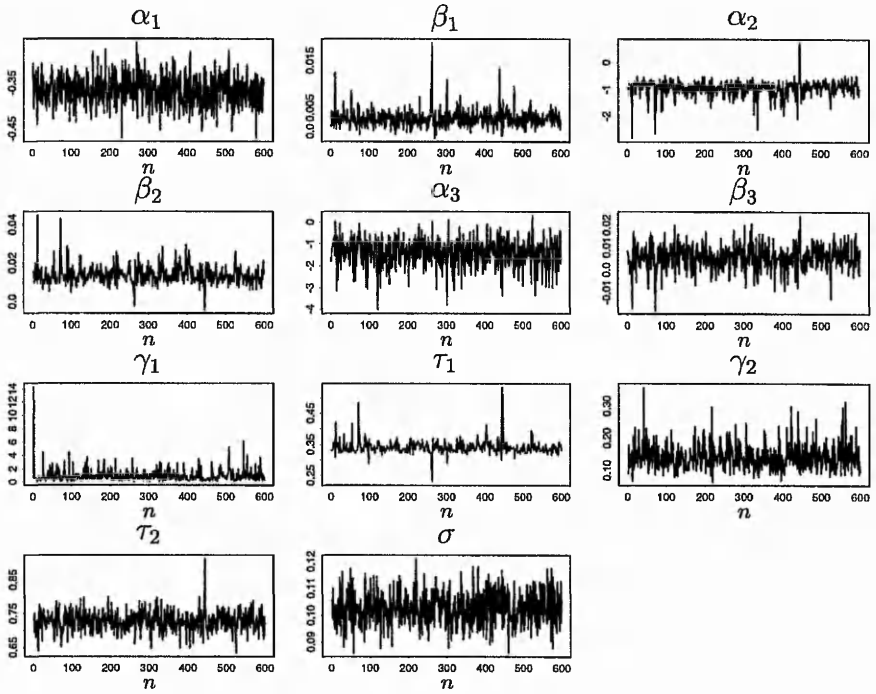


Figure 7.11: Sample path of Markov chain for the Global average temperature.

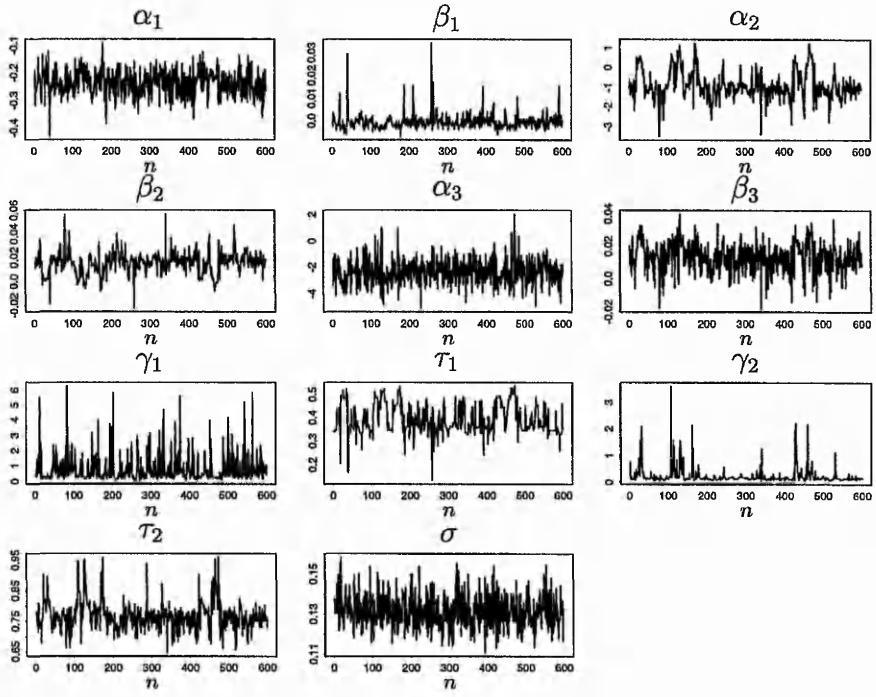


Figure 7.12: Sample path of Markov chain for the Northern Hemisphere temperature.

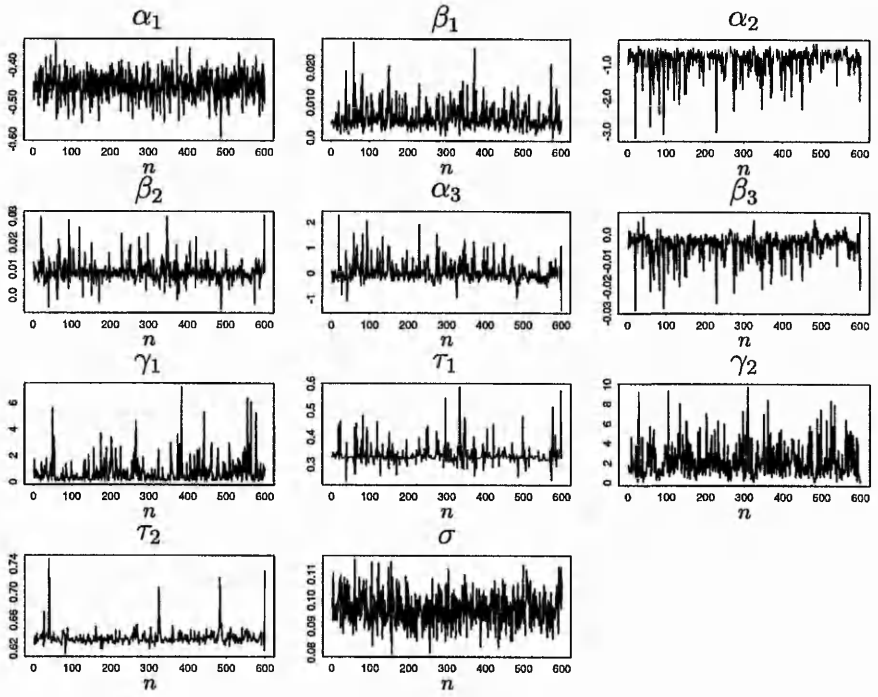


Figure 7.13: Sample path of Markov chain for the Southern Hemisphere temperature.

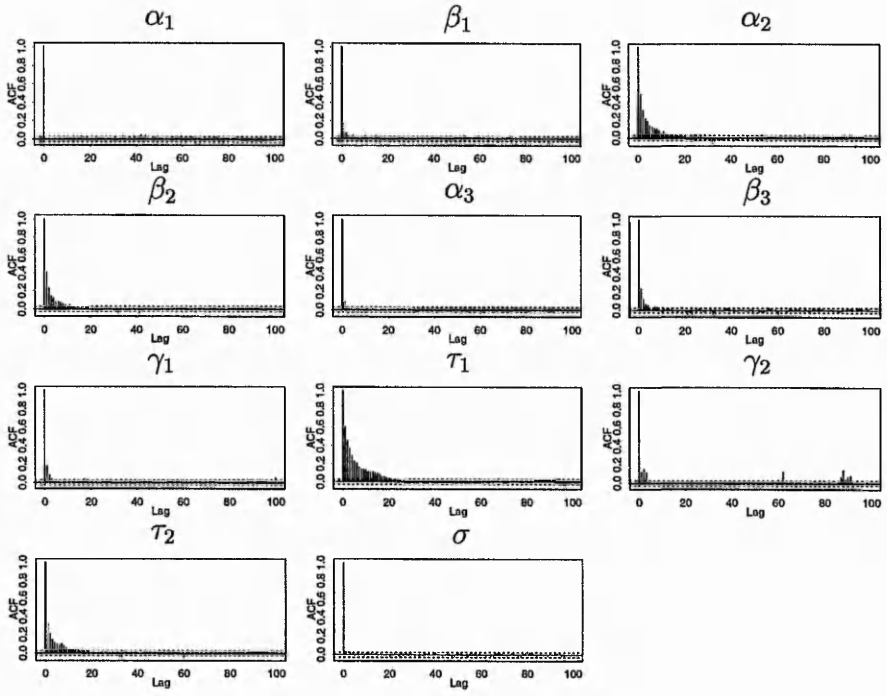


Figure 7.14: Autocorrelation of Markov chain for the Global average temperature.

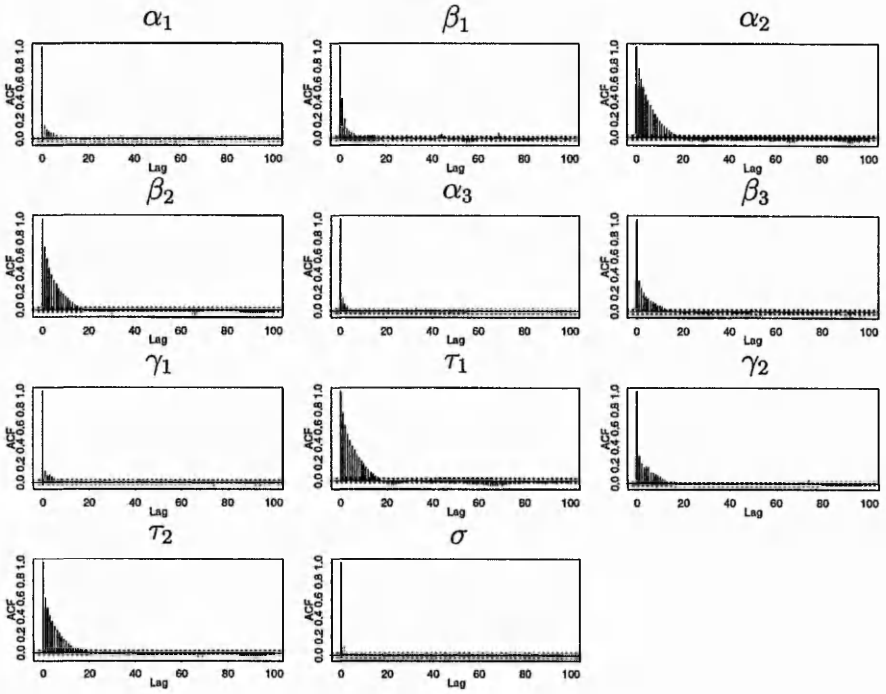


Figure 7.15: Autocorrelation of Markov chain for the Northern Hemisphere temperature.

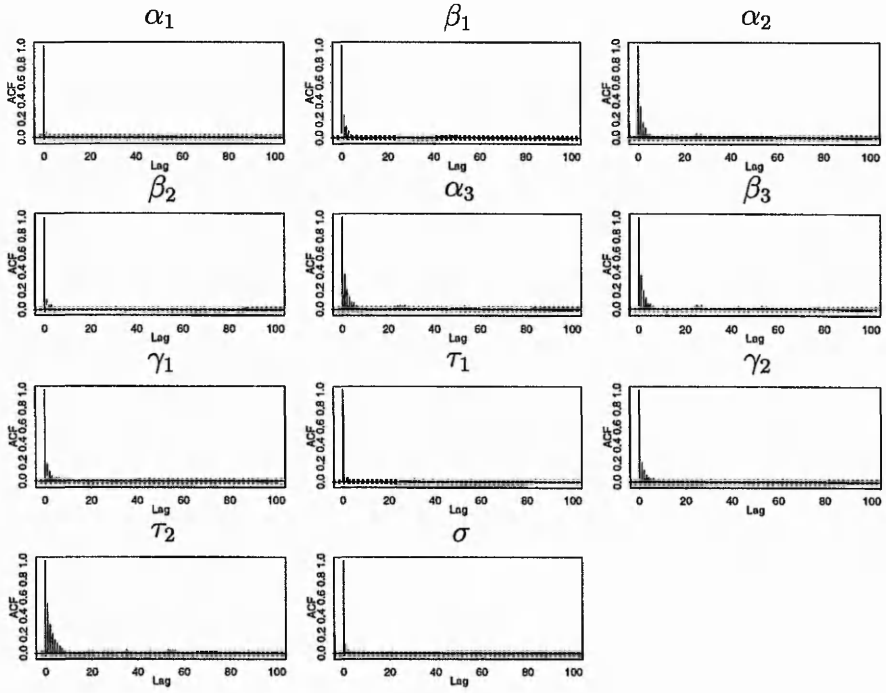


Figure 7.16: Autocorrelation of Markov chain for the Southern Hemisphere temperature.

Chapter 8

Reversible Jump MCMC

We have used Markov chain Monte Carlo methods to sample from different posterior distributions in Chapter 6 and Chapter 7. In this chapter, we would like to use MCMC in model selection. In model selection, different models with different variable dimensions are compared and the number of parameters in the model is itself one of the parameters that need to be estimated. In this case, traditional MCMC methods will not be applicable.

Recently a lot of work has been done to use Markov chain Monte Carlo methods to compare models in order to determine which model the sample comes from. A method of constructing a Markov chain for models with variable dimensions of the parameter space called reversible jump Markov chain Monte Carlo (RJMCMC) has been proposed by Green (1995). Richardson and Green (1997) applied this method to the model of univariate normal mixtures. Encouraged by this success, a lot of work has been done since then. Troughton and Godsill (1997) employed full conditionals to achieve efficient reversible jump Markov chains for autoregressive time series. Robert, Rydén and Titterington (2000) applied the RJMCMC to hidden Markov models, Yan and Brooks (2000) used the RJMCMC in archaeology, applied the RJMCMC to model prehistoric corbelled domes. Waagepetersen and Suzuki (2001) wrote a tutorial paper on RJMCMC, which is worth reading.

8.1 Reversible Jump MCMC

Green (1995) introduced a new framework for the construction of reversible Markov chain samplers that jump between parameter subspaces of different dimensionality. In this section, we describe this reversible jump Markov chain Monte Carlo method.

Suppose we want to choose between a collection of models $\{M_i, i = 0, 1, \ldots, k\}$. The parameter vector of model M_i is denoted by $\boldsymbol{\theta}^{(i)}$. The dimensions of the $\boldsymbol{\theta}^{(i)}$, n_i, $i = 0, 1, \ldots, k$, may be different. Without loss of generality, we suggest that $n_i \leq n_j$, if $i < j$. We observe the sample $\mathbf{y}$, then the joint distribution of M_i, $\boldsymbol{\theta}^{(i)}$ and $\mathbf{y}$ can be written in general as

$$p(M_i, \boldsymbol{\theta}^{(i)}, \mathbf{y}) = p(M_i)p(\boldsymbol{\theta}^{(i)}|M_i)p(\mathbf{y}|M_i, \boldsymbol{\theta}^{(i)}), \tag{8.1}$$

where $i = 0, 1, \ldots, k$. The components of this product are the prior probability of the model, the prior density for the parameter vector conditioned on the model and the likelihood function. The posterior joint distribution of $(M_i, \boldsymbol{\theta}^{(i)})$ is

$$p(M_i, \boldsymbol{\theta}^{(i)}|\mathbf{y}) \propto p(M_i)p(\boldsymbol{\theta}^{(i)}|M_i)p(\mathbf{y}|M_i, \boldsymbol{\theta}^{(i)}). \tag{8.2}$$

For convenience of notation, we shall suppress dependence on $\mathbf{y}$, for example, we shall denote $p(M_i, \boldsymbol{\theta}^{(i)}|\mathbf{y})$ as $p(M_i, \boldsymbol{\theta}^{(i)})$.

Green (1995) constructed a Markov chain which is aperiodic and irreducible, using the "reversible jump" method. Now the moves of the Markov chain are of two types, one consists of moves within a model; the other consists the moves between models. Within model moves are performed using traditional MCMC updates, for example, using the Metropolis Hastings, Gibbs or Slice sampler. Between model moves are performed using the "reversible jump" method. Suppose we are in the state $x_i = (M_i, \boldsymbol{\theta}^{(i)})$, we propose a move to state $x_j = (M_j, \boldsymbol{\theta}^{(j)})$, by drawing x_j from an arbitrary, convenient distribution $p(x_j|x_i)$. With probability $\min\{1, A\}$ we accept the move; otherwise we reject the move, where

$$A = \min\left\{1, \frac{p(M_j, \boldsymbol{\theta}^{(j)})p(x_i|x_j)}{p(M_i, \boldsymbol{\theta}^{(i)})p(x_j|x_i)}\right\}. \tag{8.3}$$

The above expression for the acceptance probability in (8.3) is rather obscure. Green gave another more intuitive expression for (8.3) which also provides a straightforward method for constructing the moves between models. We describe Green's

construction of between model moves as follows. Suppose we choose the move from $(M_i, \boldsymbol{\theta}^{(i)})$ to $(M_j, \boldsymbol{\theta}^{(j)})$ with probability $p(i,j)$. This move is suggested with probability $p(i,j)$, at any particular iteration, starting from anywhere in $(M_i, \boldsymbol{\theta}^{(i)})$. We consider the case $n_i < n_j$ first. We need to generate a vector $\mathbf{u} \sim p(\mathbf{u})$ of length $n_j - n_i$ so that $\boldsymbol{\theta}^{(j)} = f(\boldsymbol{\theta}^{(i)}, \mathbf{u})$ for some invertible function f. Green proved that such a move would be accepted with probability $\min\{1, A\}$, where

$$A = \frac{p(M_j, \boldsymbol{\theta}^{(j)})p(j,i)}{p(M_i, \boldsymbol{\theta}^{(i)})p(i,j)p(\mathbf{u})} \left| \frac{\partial \boldsymbol{\theta}^{(j)}}{\partial(\boldsymbol{\theta}^{(i)}, \mathbf{u})} \right|. \tag{8.4}$$

See Green (1995) for more details. The reverse move from $(M_j, \boldsymbol{\theta}^{(j)})$ to $(M_i, \boldsymbol{\theta}^{(i)})$ would then be accepted with probability $\min\{1, 1/A\}$. With this algorithm, we construct a Markov chain, which includes both the within model moves and between model moves, with the desired stationary distribution.

Green (1995) called the between model moves a "birth-death pair", which increases the dimension n_i to n_j, or reduce it from n_j to n_i, when $n_i < n_j$, $i < j$ in the change-point model. Richardson and Green (1997) called the move "split" when the move is from state $(M_i, \boldsymbol{\theta}^{(i)})$ to state $(M_j, \boldsymbol{\theta}^{(j)})$ with $n_i < n_j$; "combine" if $n_i > n_j$.

The "reversible jump" approach is flexible and the details change depending on the models to be discussed. We have two expressions for A which determine the probability of accepting a move between models. In the following, we try to apply Green's reversible jump sampler to our problem using different expressions for A in (8.3) and (8.4).

8.2 Structural Break Model

We are now going to discuss the problem of choosing between different structural break models (with different number of unknown components) so that Green's approach can be used.

8.2.1 The model

The structural break model can be written as

$$y_t = \begin{cases} \alpha_1 + \beta_1 t + \varepsilon_t, & t < \tau_1 T; \\ \alpha_2 + \beta_2 t + \varepsilon_t, & \tau_1 T \leq t < \tau_2 T; \\ \ldots\ldots & \\ \alpha_{m+1} + \beta_{m+1} t + \varepsilon_t, & \tau_m T \leq t, \end{cases} \tag{8.5}$$

with m structural break points at $\tau_i T$, $i = 1, 2, \ldots, m$ and ε_t is iid $\mathrm{N}(0, \sigma^2)$, $t = 0, 1, \ldots, T$. The sample size is $T + 1$. We suppose that all the m structural break points are located within the measuring time interval $(0, T)$ so that we have

$$0 < \tau_1 < \tau_2 < \cdots < \tau_m < 1. \tag{8.6}$$

We further suppose that the maximum value m can take is k_{max}, that is, $0 \leq m \leq k_{max}$.

For convenience of the following description, we put the parameters (except σ) into groups like $(\alpha_1, \beta_1, \tau_1)$, $(\alpha_2, \beta_2, \tau_2)$, ..., $(\alpha_{m+1}, \beta_{m+1}, \tau_{m+1})$, where $\tau_{m+1} = 1$. So the structural break model with m structural break points has $m+1$ structural break components, for example, the ith structural break component would be $(\alpha_i, \beta_i, \tau_i)$, $i = 1, 2, \ldots, m + 1$.

There are two special cases for the structural break model. First when $\tau_m \to 1$, the number of structural break points decreases from m to $m - 1$. Secondly when $m = 0$, we define y_t as

$$y_t = \alpha_1 + \beta_1 t + \varepsilon_t, \qquad t \in [0, T]. \tag{8.7}$$

In this special case, y_t reduces to a simple linear regression model.

In order to introduce matrix notation for description, we rewrite the structural break model as

$$y_t = \sum_{i=1}^{m+1} (\alpha_i + \beta_i t) I_{\{\tau_{i-1} T \leq t < \tau_i T\}} + \varepsilon_t, \tag{8.8}$$

where $\tau_0 = 0$, $\tau_{m+1} = 1$ and I is the indicator function that equals to 1 when the condition is true; 0 when the condition is false. We should notice that the last term in the above summation, when $i = m + 1$, is $(\alpha_{m+1} + \beta_{m+1} t) I_{\{\tau_m T \leq t \leq \tau_{m+1} T\}}$, which includes the right edge point $\tau_{m+1} T = T$ in the set, rather than the expression in

the above summation, otherwise we will lose the last datum from the sample. For the convenience of notation, we still use the expression (8.8).

The structural break components consist of segments over different intervals such as $[\tau_{i-1}T, \tau_i T)$, $i = 1, 2, \ldots, m+1$. Each line needs two points to determine its position without a disturbance term. At the point of Bayesian prior, we do not need any constraint in the number of data points to estimate one structural break component. In order to avoid too many "short" structural break intervals which are not interesting to economists and may cause inefficiency of our approach, we suppose that at least two points fall in each interval $[\tau_{i-1}T, \tau_i T)$, that is, $\tau_i T - \tau_{i-1}T > 2$. If we denote $\Delta\tau_i = \tau_i - \tau_{i-1}$, $\Delta\tau = \min\{\tau_i - \tau_{i-1}, i = 1, 2, \ldots, m+1\}$, then we have

$$\Delta\tau > \frac{2}{T}. \tag{8.9}$$

In order to use matrix notation to simplify the expressions, we introduce the following symbols:

1. $\boldsymbol{I}_{i1}$ is a column vector of length $T+1$ with jth element $I_{\{\tau_{i-1}T \le j-1 < \tau_i T\}}$, $j = 1, 2, \ldots, T+1$, $i = 1, 2, \ldots, m+1$;
2. $\boldsymbol{I}_{i2}$ is a column vector of length $T+1$ with jth element $(j-1)I_{\{\tau_{i-1}T \le j-1 < \tau_i T\}}$, $j = 1, 2, \ldots, T+1$, $i = 1, 2, \ldots, m+1$;
3. X is a $(T+1) \times (2m+2)$ matrix $X = (\boldsymbol{I}_{11}, \ldots, \boldsymbol{I}_{(m+1)1}, \boldsymbol{I}_{12}, \ldots, \boldsymbol{I}_{(m+1)2})$;
4. $\boldsymbol{\varepsilon} = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_T)'$;
5. $\boldsymbol{\beta} = (\alpha_1, \ldots, \alpha_{m+1}, \beta_1, \ldots, \beta_{m+1})'$;
6. $\mathbf{y} = (y_0, y_1, \ldots, y_T)'$ is the sample data;

then the model can be rewritten as

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{8.10}$$

and the likelihood function can be written as

$$p(\mathbf{y}|m, X, \boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma^{T+1}} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' X' X (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}. \tag{8.11}$$

where $\nu = T - 1 - 2m$, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ and $\nu s^2 = (\mathbf{y} - X\hat{\boldsymbol{\beta}})'(\mathbf{y} - X\hat{\boldsymbol{\beta}})$. It is easy to see that under conditions (8.6) and (8.9), $X'X$ is invertible.

8.2.2 Priors

The choices of prior always play an important role in the Bayesian approach, because the posterior distributions for the parameters in the model and the posterior model probabilities in the model selection problem could be sensitive to the choice of prior distribution. One choice is to seek experts' opinions, to translate experts' beliefs into the prior distribution for the model parameters (see Kadane and Wolfson, 1998 and O'Hagan, 1998).

The alternative is to use distributions which produce mathematically attractive posterior distributions. For example, the conjugate prior, the Jeffreys' prior, the reference prior *etc* (see Bernardo and Smith, 1994, Zellner, 1987 and Ibrahim and Chen, 2000). The "diffuse" prior (vague prior) distributions are frequently used to represent very few prior information about the parameters in the model, for example, the noninformative prior. The Monahan's prior which is used in the previous chapters is a "diffuse" proper prior and can be used in model selection. We continue to use Monahan's prior in this chapter.

Suppose we want to choose from a collection of structural break models denoted as $\{M_m : m = 0, 1, \ldots, k_{max}\}$ where M_m indicates a structural break model with m structural break points. The parameter vector in the model M_m is denoted by $\boldsymbol{\theta}^{(m)}$, so a state, the model indicator M_m and the associated parameter vector $\boldsymbol{\theta}^{(m)}$, is denoted by $(M_m, \boldsymbol{\theta}^{(m)})$.

We choose the prior distribution for model M_m as

$$p(M_m) = \begin{cases} \dfrac{1}{k_{max}+1}, & m \in \{0, 1, \ldots, k_{max}\}; \\ 0, & \text{otherwise}, \end{cases} \tag{8.12}$$

and the prior for $\boldsymbol{\beta}$ in model M_m as

$$p(\boldsymbol{\beta}|M_m, \sigma) = \frac{|V|^{-1/2}}{(2\pi\sigma^2)^{m+1}} \exp\left\{-\frac{\boldsymbol{\beta}' V^{-1} \boldsymbol{\beta}}{2\sigma^2}\right\}, \tag{8.13}$$

where V is a $2m+2$ order matrix with element k_v^2 on the diagonal and 0 elsewhere.

We choose an inverted gamma distribution as prior for σ in model M_m

$$p(\sigma|M_m) = \frac{2}{\Gamma(b/2)} \left(\frac{a}{2}\right)^{b/2} \frac{1}{\sigma^{b+1}} \exp\left\{-\frac{a}{2\sigma^2}\right\}, \tag{8.14}$$

where $a = b = 1/128$ and $k_v = 16$ so that the prior for σ is "diffuse" (see section 7.2 on page 99).

We choose the even-numbered order statistics from $2m+1$ points uniformly distributed on (0, 1) for $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_m)'$ in model M_m

$$p(\boldsymbol{\tau}|M_m) = (2m+1)! \prod_{i=1}^{m+1} (\tau_i - \tau_{i-1}) \times I_{\{\tau_1 < \tau_2 < \cdots < \tau_m\}}. \tag{8.15}$$

This was suggest by Green (1995) in order to avoid too many "short" steps with $\tau_{j+1} - \tau_j$ small which can arise if a uniform prior is used for $\boldsymbol{\tau}$ on $(0, 1)$.

We suppose $(\boldsymbol{\beta}, \sigma)$ and $\boldsymbol{\tau}$ to be *a priori* independent conditional on model M_m so that we have the prior for the state $(M_m, \boldsymbol{\theta}^{(m)})$ as

$$p(M_m, \boldsymbol{\theta}^{(m)}) = p(M_m)p(\boldsymbol{\beta}|M_m, \sigma)p(\sigma|M_m)p(\boldsymbol{\tau}|M_m). \tag{8.16}$$

8.2.3 Posterior Densities

With the prior distributions for the model indicator M_m, parameter vector $\boldsymbol{\theta}^{(m)}$, *a priori* independence and the likelihood above, the posterior density for the state $(M_m, \boldsymbol{\theta}^{(m)})$ is then

$$p(M_m, \boldsymbol{\beta}, \sigma, \boldsymbol{\tau}|\mathbf{y}) \propto p(\mathbf{y}|M_m, \boldsymbol{\theta}^{(m)})p(M_m)p(\boldsymbol{\beta}|M_m, \sigma)p(\sigma|M_m)p(\boldsymbol{\tau}|M_m). \tag{8.17}$$

The posterior densities for parameters under model $(M_m, \boldsymbol{\theta}^{(m)})$ are complicated. We can still make use of the full conditional posterior densities for $\boldsymbol{\beta}$ and σ. The full conditional distribution for $\boldsymbol{\beta}$ is a multivariate normal distribution,

$$p(\boldsymbol{\beta}|\mathbf{y}, M_m, \boldsymbol{\theta}_{\boldsymbol{\beta}}) \propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'V_1(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})\right\}, \tag{8.18}$$

where $\boldsymbol{\theta}_{\boldsymbol{\beta}}$ represents all the parameters except $\boldsymbol{\beta}$, $V_1 = X'X + V^{-1}$ and $\tilde{\boldsymbol{\beta}} = V_1^{-1}X'\mathbf{y}$. The full conditional distribution for σ is an inverted gamma distribution,

$$p(\sigma|\mathbf{y}, M_m, \boldsymbol{\theta}_\sigma) \propto \frac{1}{\sigma^{T+b+2m+3}} \exp\left\{-\frac{\nu s_1^2}{2\sigma^2}\right\}, \tag{8.19}$$

where $\boldsymbol{\theta}_\sigma$ represents all the parameters except σ and

$$\nu s_1^2 = a + \nu s^2 - (\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})'X'X\hat{\boldsymbol{\beta}} + (\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}})'V_1(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}).$$

8.3 Sampling Strategy for Structural Break Model

Now we can construct our own algorithm for a reversible jump MCMC for the structural break model selection problem. The parameters to be sampled are M_m, $\boldsymbol{\beta}$, $\boldsymbol{\tau}$ and σ. The moves of the Markov chain consist of two kinds: between model moves and within model moves. For simplicity of the algorithm, we restrict the between model moves by changing the number of the structural break components by 1 in each step (the number of parameters of the model changes by 3 in each step). Each sweeping cycle consists of the following move types:

1. updating $\boldsymbol{\tau}$;
2. updating σ;
3. updating $\boldsymbol{\beta}$;
4. updating the structural break components from m to $m-1$, or from m to $m+1$;

Moves 1 to 3 are within model moves, which can be carried out with traditional MCMC methods like the Metropolis-Hastings sampler and the Gibbs sampler. Move 4 consists of between model moves that involves changing the model dimension and the corresponding model parameters. We make use of the full conditional distributions of σ and $\boldsymbol{\beta}$, conditional on the other parameters, using a Metropolis-Hastings sampler.

8.3.1 Between model moves

We use the RJMCMC sampler for the between model moves. We write (i, j) to denote the proposal for the move where the present state is the model with i-component structural breaks, $(M_i, \boldsymbol{\theta}^{(i)})$, and we propose a move to the state with j-component structural breaks, $(M_j, \boldsymbol{\theta}^{(j)})$, with probability $p(i, j)$. As indicated above, in each between model move, we change the number of structural break components by 1 (*i.e.* $i - j = \pm 1$), with equal probabilities for an increasing or decreasing proposal, that is, $p(m, m \pm 1) = 0.5$.

If we suggest a move from the present state $(M_m, \boldsymbol{\theta}^{(m)})$ to new state $(M_{m+1}, \boldsymbol{\theta}^{(m+1)})$ with the number of structural break points increasing by one, then according to Green's equation (8.3), we accept this move with acceptance probability $\min\{1, A\}$, where

$$A = \frac{p(M_{m+1}, \boldsymbol{\theta}^{(m+1)})p(m|m+1)}{p(M_m, \boldsymbol{\theta}^{(m)})p(m+1|m)}, \tag{8.20}$$

where $p(m+1|m)$ is an arbitrary distribution density, proposing the new state $(M_{m+1},$
$\boldsymbol{\theta}^{(m+1)})$ conditional on the present state $(M_m, \boldsymbol{\theta}^{(m)})$. $p(m|m+1)$ is the distribution density proposing the reverse move. Alternatively

$$A = \frac{p(M_{m+1}, \boldsymbol{\theta}^{(m+1)})}{p(M_m, \boldsymbol{\theta}^{(m)})p(\mathbf{u})} \left| \frac{\partial \boldsymbol{\theta}^{(m+1)}}{\partial(\boldsymbol{\theta}^{(m)}, \mathbf{u})} \right|, \tag{8.21}$$

if we generate $\mathbf{u}$ from a specific proposal density $p(\mathbf{u}) = p(\mathbf{u}|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)})$, which depends on the present state $(M_m, \boldsymbol{\theta}^{(m)})$ and the new model M_{m+1}.

The reverse proposal, $m+1 \to m$, will be accepted with probability $\min\{1, 1/A\}$.

We suggest two methods for proposing the between model moves corresponding to the different expressions for A in (8.21) and (8.20), and call these method I and method II respectively.

Method I

The first method is to use the expression of A in (8.21) with the corresponding method for construction the between model moves. This method is widely used, for example, Green (1995) applies this method to the change-point model, Richardson and Green (1997) apply this method to the mixture model of univariate normal distributions, Robert, Rydén and Titterington (2000) apply this method to their hidden Markov model and Yan and Brooks (2000) applied this method to model the prehistoric corbelled domes. We will apply this method to our structural break model and follow Brooks, Giudici and Roberts (2000) suggestions for efficient construction of reversible jump MCMC proposal distributions.

Suppose we are currently in the state $(M_m, \boldsymbol{\theta}^{(m)})$ with m structural break points. We update the number of structural break components by increasing it by one or decreasing it by one randomly, with probability 0.5, in one step.

When we suggest a move which includes an additional structural break component, we select a component and split it into two. We need to generate new parameters for the new component. Green (1995) suggested that the parameters for the present state $(\alpha_j, \beta_j, \tau_j)$ should not be completely discarded because the present state is typically well-supported in the posterior distribution. See Green (1995), section 4; Richardson and Green (1997), section 3.2. We adopt this strategy in what follows.

If we have chosen a decreasing proposal, that is to reduce the number of structural break components by 1, then we randomly choose a label j, $j = 2, 3, \ldots, m+1$, and combine the two structural break components $(\alpha_{j-1}, \beta_{j-1}, \tau_{j-1})$ and $(\alpha_j, \beta_j, \tau_j)$ to give a new component labelled j^* with component parameters

$$\begin{aligned} \tau_{j^*} &= \tau_j, \\ \alpha_{j^*} &= \frac{(\tau_{j-1} - \tau_{j-2})\alpha_{j-1} + (\tau_j - \tau_{j-1})\alpha_j}{\tau_j - \tau_{j-2}}, \\ \beta_{j^*} &= \frac{(\tau_{j-1} - \tau_{j-2})\beta_{j-1} + (\tau_j - \tau_{j-1})\beta_j}{\tau_j - \tau_{j-2}}. \end{aligned} \tag{8.22}$$

We accept this proposal with probability $\min\{1, 1/A\}$ where A will be defined below after we consider the proposal for increasing the number of structural break components. If the decreasing proposal is accepted, we need to relocate the structural break components to satisfy the order in (8.6).

The increasing proposal for the structural break components starts with the random choice of a label j from $\{1, 2, \ldots, m+1\}$. We then try to increase the dimension by splitting the component $(\alpha_j, \beta_j, \tau_j)$ into two new components, denoted as $(\alpha_1^*, \beta_1^*, \tau_1^*)$ and $(\alpha_2^*, \beta_2^*, \tau_2^*)$. In this move, the degrees of freedom increase by 3, so we need to generate three independent random variables $\mathbf{u} = (u_1, u_2, u_3)$ with distribution densities

$$\begin{aligned} u_1 &\sim \mathrm{U}(0,1); \\ u_2 &\sim \mathrm{N}(0, \sigma_u^2); \\ u_3 &\sim \mathrm{N}(0, \sigma_u^2). \end{aligned} \tag{8.23}$$

The two new components are defined by choosing the parameters

$$\begin{aligned}
\tau_1^* &= \tau_{j-1} + (\tau_j - \tau_{j-1})u_1, \\
\tau_2^* &= \tau_j, \\
\alpha_1^* &= \alpha_j - u_2\sqrt{\frac{\tau_j - \tau_1^*}{\tau_1^* - \tau_{j-1}}}\beta_j(\tau_j - \tau_{j-1})T, \\
\alpha_2^* &= \alpha_j + u_2\sqrt{\frac{\tau_1^* - \tau_{j-1}}{\tau_j - \tau_1^*}}\beta_j(\tau_j - \tau_{j-1})T, \\
\beta_1^* &= \beta_j\left(1 - u_3\sqrt{\frac{\tau_j - \tau_1^*}{\tau_1^* - \tau_{j-1}}}\right), \\
\beta_2^* &= \beta_j\left(1 + u_3\sqrt{\frac{\tau_1^* - \tau_{j-1}}{\tau_j - \tau_1^*}}\right),
\end{aligned} \tag{8.24}$$

where $\beta_j(\tau_j - \tau_{j-1})T$ is the range of change for $\alpha_j + \beta_j t$ over $t \in [\tau_{j-1}T, \tau_j T)$.

If we select $j = 1$, then $\tau_{j-1} = \tau_0 = 0$; if we select $j = m+1$, then $\tau_j = \tau_{m+1} = 1$. It is easy to see that $\tau_1^* \in (\tau_{j-1}, \tau_j)$. We need to reallocate τ_i, $i = 1, 2, \ldots, m, i \neq j-1, j$ and τ_1^*, τ_2^* so as to label them in ascending order as in equation (8.6). Also the parameters α_i, β_i, $i = 1, 2, \ldots, m+1, i \neq j-1, j$ and $\alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*$ need be reallocated correspondingly for the next move. We will continue to use non-reallocated variable symbols in the following for convenience.

In the move that increases the number of structural break components, all the parameters except $(\alpha_j, \beta_j, \tau_j)$ are fixed with the same values as before this move. The Jacobian of the transformation from $(\boldsymbol{\theta}^{(m)}, \mathbf{u})$ to $\boldsymbol{\theta}^{(m+1)}$ is

$$|J| = (\tau_j - \tau_{j-1})^2 T^2 \beta_j^2 (r + 1/r)^2,$$

where $r = \sqrt{\frac{\tau_j - \tau_1^*}{\tau_1^* - \tau_{j-1}}}$ and the vector $\mathbf{u} = (u_1, u_2, u_3)'$ has the density

$$p(\mathbf{u}) = \frac{1}{2\pi\sigma_u^2} \exp\left\{-\frac{u_2^2 + u_3^2}{2\sigma_u^2}\right\}. \tag{8.25}$$

The acceptance probability for the increasing proposal for the structural break components is $\min\{1, A\}$ where

$$\begin{aligned}
A \;=\; & \text{likelihood ratio} \\
& \times \frac{1}{2\pi k_v^2 \sigma^2} \exp\left\{-\frac{(\alpha_1^*)^2 + (\beta_1^*)^2 + (\alpha_2^*)^2 + (\beta_2^*)^2 - \alpha_j^2 - \beta_j^2}{2k_v^2\sigma^2}\right\} \\
& \times (2m+2)(2m+3)\frac{(\tau_1^* - \tau_{j-1})(\tau_j - \tau_1^*)}{\tau_j - \tau_{j-1}} \\
& \times 2\pi\sigma_u^2 \exp\left\{\frac{u_2^2 + u_3^2}{2\sigma_u^2}\right\} (\tau_j - \tau_{j-1})^2 T^2 \beta_j^2 (r + 1/r)^2.
\end{aligned} \tag{8.26}$$

In (8.26), the first line is the ratio of (8.11) (details will be given below), the second line is the ratio of the priors for $\boldsymbol{\beta}$ conditional on the model indicator and σ, and the third line is the ratio of priors for $\boldsymbol{\tau}$. All these ratios are for the new parameters $\boldsymbol{\theta}^{(m+1)}$ to that for the old parameters $\boldsymbol{\theta}^{(m)}$. The last line is $|J|/p(\mathbf{u})$.

If we denote $I_j = \{t : \tau_{j-1} \leq t < \tau_j\}$, if $j \neq m+1$, or $I_{m+1} = \{t : \tau_m \leq t \leq \tau_{m+1}\}$, $I_j^{(1)} = \{t : \tau_{j-1} \leq t < \tau_1^*\}$, $I_j^{(2)} = \{t \in I_j : t \notin I_j^{(1)}\}$, then we can write the likelihood ratio in (8.26) as

$$\text{likelihood ratio} = \exp\left\{-\frac{1}{2\sigma^2}\left[\sum_{I_j^{(1)}}(y_t - (\alpha_1^* + \beta_1^* t))^2 + \sum_{I_j^{(2)}}(y_t - (\alpha_2^* + \beta_2^* t))^2 - \sum_{I_j}(y_t - (\alpha_j + \beta_j t))^2\right]\right\}.$$

For a decreasing move in (8.22), that is, the removal of a break component, the acceptance probability is $\min\{1, 1/B\}$, the expression for B will be given in the following.

Suppose the current state is $(M_m, \boldsymbol{\theta}^{(m)})$ and we propose a decreasing move. We randomly select a label j from $2, 3, \ldots, m+1$ and are to combine the two structural break components $(\alpha_{j-1}, \beta_{j-1}, \tau_{j-1})$ and $(\alpha_j, \beta_j, \tau_j)$ to give a new one with parameters $(\alpha_{j*}, \beta_{j*}, \tau_{j*})$ generated by (8.22) and the new state is $(M_{m-1}, \boldsymbol{\theta}^{(m-1)})$. Let us consider the corresponding reverse move, from state $(M_{m-1}, \boldsymbol{\theta}^{(m-1)})$ to state $(M_m, \boldsymbol{\theta}^{(m)})$, which splits $(\alpha_{j*}, \beta_{j*}, \tau_{j*})$ into two $(\alpha_{j-1}, \beta_{j-1}, \tau_{j-1})$ and $(\alpha_j, \beta_j, \tau_j)$. In this reverse move, we generate three random variables u_1, u_2 and u_3 in order to generate $(\alpha_{j-1}, \beta_{j-1}, \tau_{j-1})$ and $(\alpha_j, \beta_j, \tau_j)$ so that we have

$$\begin{aligned} u_1 &= \frac{\tau_{j-1} - \tau_{j-2}}{\tau_j - \tau_{j-2}}, \\ u_2 &= \frac{\alpha_j - \alpha_{j*}}{s\beta^*(\tau_j - \tau_{j-2})T}, \\ u_3 &= \frac{\beta_j - \beta_{j*}}{s\beta_{j*}}, \end{aligned} \tag{8.27}$$

where $s = \sqrt{\dfrac{\tau_{j*} - \tau_{j-1}}{\tau_{j-1} - \tau_{j-2}}}$. Then we have the acceptance probability for the corresponding increasing move from $(\alpha_{j*}, \beta_{j*}, \tau_{j*})$ to $(\alpha_{j-1}, \beta_{j-1}, \tau_{j-1})$ and $(\alpha_j, \beta_j, \tau_j)$ as

$\min\{1, B\}$ where

$$\begin{aligned} B &= \text{likelihood ratio} \\ &\times \frac{1}{2\pi k_v^2 \sigma^2} \exp\left\{ -\frac{(\alpha_{j-1})^2 + (\beta_{j-1})^2 + (\alpha_j)^2 + (\beta_j)^2 - \alpha_{j*}^2 - \beta_{j*}^2}{2k_v^2\sigma^2} \right\} \\ &\times 2m(2m+1) \frac{(\tau_{j-1} - \tau_{j-2})(\tau_{j*} - \tau_{j-1})}{\tau_{j*} - \tau_{j-2}} \\ &\times 2\pi\sigma_u^2 \exp\left\{ \frac{u_2^2 + u_3^2}{2\sigma_u^2} \right\} (\tau_{j*} - \tau_{j-2})^2 T^2 \beta_{j*}^2 (s + 1/s)^2. \end{aligned} \tag{8.28}$$

In (8.28), the first line is the ratio of (8.11) for parameters $\boldsymbol{\theta}^{(m-1)}$ to that for parameters $\boldsymbol{\theta}^{(m)}$,

$$\begin{aligned} \text{likelihood ratio} &= \exp\left\{ -\frac{1}{2\sigma^2} \left[\sum_{I_j^{(1)}} (y_t - (\alpha_{j-1} + \beta_{j-1}t))^2 + \sum_{I_j^{(2)}} (y_t - (\alpha_j + \beta_j t))^2 \right.\right. \\ &\qquad \left.\left. - \sum_{I_j} (y_t - (\alpha_{j^*} + \beta_{j^*} t))^2 \right] \right\}, \end{aligned}$$

where $I_j = \{t : \tau_{j-2} \le t < \tau_j\}$, if $j \neq m+1$, or $I_{m+1} = \{t : \tau_m \le t \le \tau_{m+1}\}$, $I_j^{(1)} = \{t : \tau_{j-2} \le t < \tau_{j^*}\}$, $I_j^{(2)} = \{t \in I_j : t \notin I_j^{(1)}\}$. Other lines have the similar interpretations as in (8.26) with ratios for parameters $\boldsymbol{\theta}^{(m-1)}$ to that for parameters $\boldsymbol{\theta}^{(m)}$.

So the decreasing move, from state $(M_m, \boldsymbol{\theta}^{(m)})$ to state $(M_{m-1}, \boldsymbol{\theta}^{(m-1)})$, will be accepted with probability $\{1, 1/B\}$.

"Birth" or "death" of structural break components

When the current state is $(M_0, \boldsymbol{\theta}^{(0)})$, which means no structural break point, the structural break component is a single unbroken line over $[0, T]$. If we choose to increase the number of breaks, we need to generate the new structural break component from a simple line. We call this kind of move a "birth". On the other hand, when the current state is $(M_1, \boldsymbol{\theta}^{(1)})$ and we choose to decrease the number of structural break components, we will delete the only structural break point in the model, leaving a model with no structural break point. We call this kind of move a "death". So "birth" and "death" moves are special cases of the between model moves for updating the number of structural break components and can be done in a similar way to the updating moves discussed above.

The "birth" moves happen when the current state is $(M_0, \boldsymbol{\theta}^{(0)})$ and we choose to increase the number of structural break components. In the state $(M_0, \boldsymbol{\theta}^{(0)})$, we have $y_t = \alpha_1 + \beta_1 t + \varepsilon_t$, $t \in [0, T]$. The "birth" move can be carried out as in equations (8.25) with the special values $\tau_{j-1} = 0$, $\tau_j = 1$, $j = 1$ and $m = 0$. We accept the "birth" move with probabilities $\min\{1, A_{birth}\}$ where $A_{birth} = A$ with $\tau_{j-1} = 0$, $\tau_j = 1$, $j = 1$ and $m = 0$ substituted in the expression for A in (8.26).

"Death" moves can occur when the current state is $(M_1, \boldsymbol{\theta}^{(1)})$ and can be easily undertaken by deleting the only structural break point with probability $\min\{1, 1/B_{birth}\}$, where B_{birth} is the probability of acceptance for a "birth" move in (8.28) with $j = 1$ and $m = 1$ substituted in (8.28).

Choosing σ_u

In order to obtain the acceptance probabilities that are as high as possible for both increasing and decreasing moves, we need to choose the value for σ_u carefully, see the discussion in Brooks, Giudici and Roberts (2000). Brooks, Giudici and Roberts gave some advice on choosing the variances of the random variables which are used to generate new parameters in a split move, such as u_1, u_2 and u_3. For convenience, we choose the uniform distribution over $(0, 1)$ for u_1, normal distributions for u_2 and u_3 and set equal variances for u_2 and u_3. However, in theory, we can choose any suitable distributions for u_1, u_2 and u_3. In the following, we consider choosing σ_u according to the suggestion from Brooks, Giudici and Roberts (2000).

From the above discussion we know that the acceptance probability for between model moves is determined by A. The simplest idea from Brooks, Giudici and Roberts (2000) is that we should choose values for the parameters in the proposal distributions which satisfy $A = 1$ for some "representative" point, in order to increase the acceptance probabilities for both the increasing and decreasing moves. Brooks, Giudici and Roberts (2000) considered choosing the parameters for the proposal distributions by imposing some constraints upon A and suggested that higher order derivatives of A are worth consideration. A traditional method of increasing the acceptance probabilities is to tune these parameters in the proposal distributions in a "pilot run" of the reversible jump Markov chain, in order to enhance the acceptance probabilities of between model moves. Brooks, Giudici and Roberts

method is straightforward and sometimes easy to apply while the "tuning" method is somewhat expensive in time and lacking in direction. Brooks, Giudici and Roberts showed that sometimes simple application of their method gives parameters for the proposal distributions which achieve at least as good results as that of the more computational intensive "tuning", see the examples given by Brooks, Giudici and Roberts (2000). A good suggestion for optimal scaling of random walk metropolis algorithms can be found in Roberts, Gelman and Gilks (1997) and the references therein. Roberts, Gelman and Gilks (1997) suggested that we should "tune the proposal variance so that the average acceptance rate is roughly 1/4". But their suggestion does not apply for RJMCMC, and in our experience, acceptance probabilities for between model moves in a RJMCMC sampler are usually much less than 25%.

In our structural break model, the expression for the acceptance probability is complex and it is not easy to impose constraints on higher order derivatives of A. We only impose the constraint that $A = 0$ on some "representative" point. We denote the "representative" point for the three random variables $\mathbf{u}$ as $\mathbf{u}_0$. We set $\mathbf{u}_0 = E(\mathbf{u}) = (0.5, 0, 0)'$. We then set $A|_{\mathbf{u}_0} = 1$ at the "representative" point to choose the "optimal" values for σ_u. Given these values, we obtain

$$\sigma_u = \frac{k_v \sigma}{\sqrt{(2m+3)(2m+2)\beta_j(\tau_j - \tau_{j-1})T}} \exp\left\{\frac{\alpha_j^2 + \beta_j^2}{4k_v^2\sigma^2}\right\}. \tag{8.29}$$

When we run the reversible jump Markov chain constructed as above, using Method I, we experienced difficulty. The Markov chain has a very poor acceptance rate for the between model moves, which is less than 0.1%. This causes trouble in the convergence of the chain. So we carry on to discuss another method, Method II and present the results for Method II in the following.

Method II

As a possible alternative to the between model moves suggested in method I, we propose method II which uses the full conditional posterior for $\boldsymbol{\beta}$ in (8.18) and A in (8.20) for the between model moves.

We first discuss the move which increases the dimensions of the parameter space. Suppose that we suggest a move from the present state $(M_m, \boldsymbol{\theta}^{(m)})$ to a new state

$(M_{m+1}, \boldsymbol{\theta}^{(m+1)})$. We will generate the new state $(M_{m+1}, \boldsymbol{\theta}^{(m+1)})$ from an arbitrary proposal distribution $p(m+1|m)$, conditional on the present state $(M_m, \boldsymbol{\theta}^{(m)})$, that is,

$$\begin{aligned} p(m+1|m) &= p(M_{m+1}, \boldsymbol{\theta}^{(m+1)}|M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y}) \\ &= q(\boldsymbol{\theta}^{(m+1)}|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y})p(m, m+1), \end{aligned} \tag{8.30}$$

where $p(m, m+1) = 0.5$, the probability for the increasing proposal. The choice of an appropriate proposal distribution, q, is important for the between model moves. For the move $(M_m, \boldsymbol{\theta}^{(m)})$ to $(M_{m+1}, \boldsymbol{\theta}^{(m+1)})$, we fix σ and randomly select a j, $j \in \{1, 2, \ldots, m\}$, then we split the jth component of the structural break $(\alpha_j, \beta_j, \tau_j)$ into two structural break components, denoted as $(\alpha_1^*, \beta_1^*, \tau_1^*)$ and $(\alpha_2^*, \beta_2^*, \tau_2^*)$ where $\tau_2^* = \tau_j$. The parameters for all the other structural break components in the current model are fixed. In this case, the proposal distribution for the increasing move will be $q = q(\alpha_1^*, \beta_1^*, \tau_1^*, \alpha_2^*, \beta_2^*|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y})$, which can be written as

$$q = q(\alpha_1^*, \beta_1^*, \alpha_2^*, \beta_2^*|\tau_1^*, M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y})q(\tau_1^*|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y}). \tag{8.31}$$

For $q(q = q(\alpha_1^*, \beta_1^*, \alpha_2^*, \beta_2^*|\tau_1^*, M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y})q(\tau_1^*|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y}).$ (8.31)

use so instead we use the uniform distribution over (τ_{j-1}, τ_j), that is,

$$q(\tau_1^*|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y}) = \mathrm{U}(\tau_{j-1}, \tau_j).$$

As for $q(\alpha_1^*, \beta_1^*, \alpha_2^*, \beta_2^*|\tau_1^*, M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y})$, we can use the posterior density for $\boldsymbol{\beta}$ in (8.18) provided that $t \in [\tau_{j-1}T, \tau_j T)$, which corresponds to $\boldsymbol{\beta} = (\alpha_1^*, \beta_1^*, \alpha_2^*, \beta_2^*)$. If we denote $t \in [\tau_{j-1}T, \tau_1^* T) = \{t_1^1, t_2^1, \ldots, t_{n_1}^1\}$, $t \in [\tau_1^* T, \tau_j T) = \{t_1^2, t_2^2, \ldots, t_{n_2}^2\}$, and define

$$X = \begin{pmatrix} 1 & 0 & t_1^1 & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & t_{n_1}^1 & 0 \\ 0 & 1 & 0 & t_1^2 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 1 & 0 & t_{n_2}^2 \end{pmatrix}, V = \begin{pmatrix} k_v^2 & 0 & 0 & 0 \\ 0 & k_v^2 & 0 & 0 \\ 0 & 0 & k_v^2 & 0 \\ 0 & 0 & 0 & k_v^2 \end{pmatrix},$$

$\boldsymbol{\beta} = (\alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*)'$, and $\mathbf{y} = (y_{t_1^1}, \ldots, y_{t_{n_1}^1}, y_{t_1^2}, \ldots, y_{t_{n_2}^2})'$. The proposal distribution for $(\alpha_1^*, \alpha_2^*, \beta_1^*, \beta_2^*)'$ is the multivariate normal distribution in (8.18) with X, V and $\mathbf{y}$ defined above.

The acceptance probability for the move from state $(M_m, \boldsymbol{\theta}^{(m)})$ to state $(M_{m+1}, \boldsymbol{\theta}^{(m+1)})$ will also depend on the proposal $p(m|m+1)$ for the corresponding opposite move from state $(M_{m+1}, \boldsymbol{\theta}^{(m+1)})$ to state $(M_m, \boldsymbol{\theta}^{(m)})$, see (8.20).

The proposal $p(m|m+1)$ for the corresponding opposite move will also make use of the full conditional for $\boldsymbol{\beta}$ in (8.18). The opposite move would combine the two structural break components (denoted in the notation of the state $(M_m, \boldsymbol{\theta}^{(m)})$) $(\alpha_1^*, \beta_1^*, \tau_1^*)$ and $(\alpha_2^*, \beta_2^*, \tau_2^*)$ into one $(\alpha_j, \beta_j, \tau_j)$, where $\tau_j = \tau_2^*$.

The proposal for this decreasing move is

$$\begin{aligned} p(m|m+1) &= p(M_m, \boldsymbol{\theta}^{(m)}|M_{m+1}, \boldsymbol{\theta}^{(m+1)}, \mathbf{y}) \\ &= q(\boldsymbol{\theta}^{(m)}|M_{m+1}, M_m, \boldsymbol{\theta}^{(m+1)}, \mathbf{y}) p(m+1, m), \end{aligned} \tag{8.32}$$

where $p(m+1, m) = 0.5$, the probability for the decreasing proposal. Here we have

$$q(\boldsymbol{\theta}^{(m)}|M_{m+1}, M_m, \boldsymbol{\theta}^{(m+1)}, \mathbf{y}) = q(\alpha_j, \beta_j|M_{m+1}, M_m, \boldsymbol{\theta}^{(m+1)}, \mathbf{y}). \tag{8.33}$$

For the proposal $q(\alpha_j, \beta_j|M_{m+1}, M_m, \boldsymbol{\theta}^{(m+1)}, \mathbf{y})$, we can use the posterior density for $\boldsymbol{\beta}$ in (8.18) provided that $t \in [\tau_{j-1}, \tau_j) = \{t_1, t_2, \ldots, t_{\tilde{n}}\}$,

$$X = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \\ & \ddots \\ 1 & t_{\tilde{n}} \end{pmatrix}, V = \begin{pmatrix} k_v^2 & 0 \\ 0 & k_v^2 \end{pmatrix},$$

and $\mathbf{y} = (y_{t_1}, y_{t_2}, \ldots, y_{t_{\tilde{n}}})'$.

We accept the above increasing move with probability $\{1, A\}$, where

$$A = \frac{p(M_{m+1}, \boldsymbol{\theta}^{(m+1)}) q(\alpha_j, \beta_j|M_{m+1}, M_m, \boldsymbol{\theta}^{(m+1)}, \mathbf{y})}{p(M_m, \boldsymbol{\theta}^{(m)}) q(\alpha_1^*, \beta_1^*, \alpha_2^*, \beta_2^*|\tau_1^*, M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y}) q(\tau_1^*|M_{m+1}, M_m, \boldsymbol{\theta}^{(m)}, \mathbf{y})}. \tag{8.34}$$

The between model move which decreases the dimensions of the parameter space can be carried out as follows. Suppose the current state is $(M_m, \boldsymbol{\theta}^{(m)})$, we suggest a move to the state $(M_{m-1}, \boldsymbol{\theta}^{(m-1)})$. We select j randomly from $\{2, 3, \ldots, m+1\}$, and combine the two adjacent structural break components $(\alpha_{j-1}, \beta_{j-1}, \tau_{j-1})$ and $(\alpha_j, \beta_j, \tau_j)$ to obtain a new one $(\alpha_j^*, \beta_j^*, \tau_j^*)$ where $\tau_j^* = \tau_j$. We accept this move with probability $\{1, 1/A\}$, where A is obtained for the above increasing move in (8.34), with some corresponding different substitutions.

8.3.2 Within model moves

For within model moves, we make use of the full conditional distributions for $\boldsymbol{\beta}$ and σ, which are multivariate normal and inverted gamma distributions respectively, using the Gibbs sampler to update $\boldsymbol{\beta}$ and σ. For updating $\boldsymbol{\tau}$, we use the Metropolis-Hastings' sampler. We update $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_m)$ in turn. For example, we will update τ_j, the proposal replacement value is τ^*, drawn uniformly on (τ_{j-1}, τ_{j+1}), and we accept this proposal replacement value with acceptance probability

$$\min\left\{1, \text{likelihood ratio} \times \frac{(\tau_{j+1} - \tau^*)(\tau^* - \tau_{j-1})}{(\tau_{j+1} - \tau_j)(\tau_j - \tau_{j-1})}\right\}. \tag{8.35}$$

If we denote I as the set of t which are between $\tau_j T$ and $\tau^* T$ (exclusive $\tau_j T$ and $\tau^* T$ if they are integral), then we have the likelihood ratio as

$$\begin{aligned}\text{likelihood ratio} \;=\; \exp\Bigg\{-\frac{\operatorname{sgn}(\tau^* - \tau_j)}{2\sigma^2}\Bigg[&\sum_I (y_t - (\alpha_{j+1} + \beta_{j+1} t))^2 \\ &- \sum_I (y_t - (\alpha_j + \beta_j t))^2\Bigg]\Bigg\},\end{aligned}$$

where $\operatorname{sgn}(x)$ is the sign function which is 1 if $x > 0$; -1 if $x < 0$ and 0 if $x = 0$.

8.4 Empirical Applications

In this section, we will illustrate the approach we discussed above. We apply our approach to two real data sets. The first is the annual data on US real GNP for the period 1909–1970 (61 observations). This data set was used in the paper by Nelson and Plosser (1982). Busetti and Harvey (2001) fit a structural break model with one or two breaks in the trend to this data. They argued that there were quite a number of places where a break or a set of breaks can be introduced. The break in trend can be in level or slope or both. Busetti and Harvey (2001) found that two breaks happen in 1929 and 1945 respectively. In our approach, we do not fix the number of breaks *a priori* but let the reversible jump Markov chain select the number of breaks and the corresponding locations. The second data set concerns the US consumer price index data, from 1860 to 1970 (111 observations), which was also used by Nelson and Plosser (1982). Harvey and Mills (2000) used the double smooth transition model to fit this series. Nelson and Plosser (1982), Harvey and Mills (2000) and

Busetti and Harvey (2001) focus on testing the null hypothesis of a unit root. If their null was rejected, then the model would be fitted with deterministic trend and a stationary error. The trend could consist of smooth transition components (Harvey and Mills, 2000) or structural break components (Busetti and Harvey, 2001). We use our approach to fit these data sets with variable number of structural break components and white noise error.

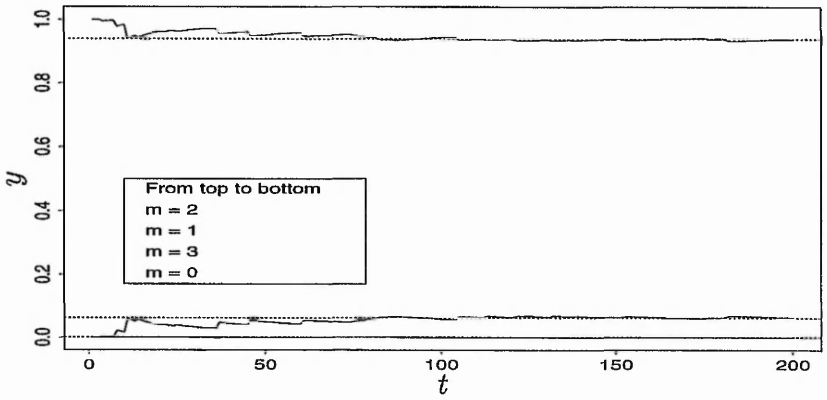
We use the second approach, that is, method II, because method I has a very much lower acceptance rate for the between model moves. For example, for the US real GNP data set, the acceptance rate for the between model move of method I is less than 0.1%, much less than that of method II, see Table 8.1 in the following.

The performance of the RJMCMC consists of two essential elements, mixing of the chains over the model indicator M_m and mixing for parameters within each model. It is difficult to make assessments of the convergence of the chains constructed by RJMCMC. In fact, the chains constructed by Richardson and Green (1997) were examined by Brooks (1997) who found that the chain for the model indicator have not reached the convergence using any of the diagnostics proposed by Gelman and Rubin (1992), Raftery and Lewis (1992), Geweke (1992) and Yu (1995). Robert, Rydén and Titterington (2000) and Fan and Brooks (2000) did not use any of the above diagnostics to examine the convergence of their chains. Nevertheless, we follow the arguments of Richardson and Green (1997) and check the stability of the chain by comparing the estimated model probabilities against the sweeping times after the burn-in and at the end of the chain.

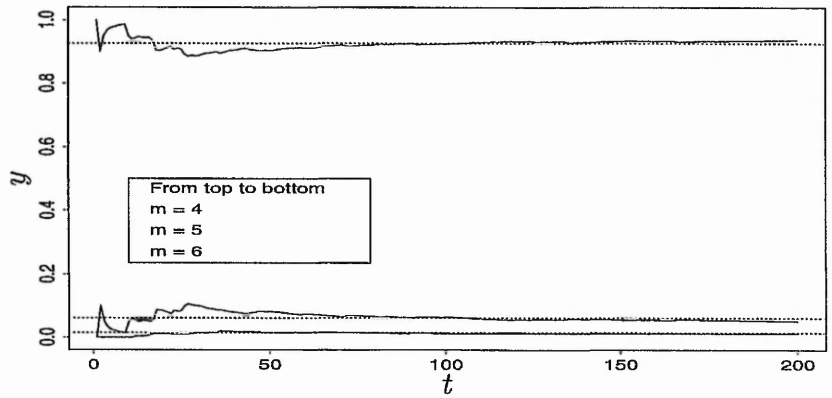
For each of the two data sets, we generate 1 million sweeps, following a burn-in period also of 1 million sweeps. We choose the maximum possible number of breaks to be $k_{max} = 10$. We believe that in sample of about 100 observations, allowing for 10 breaks should be big enough and in fact $k_{max} = 10$ is never reached in our sweeps for the two data sets.

The estimated model probabilities against the sweeping times (after the burn-in of 1 million sweeps) for both series are presented in Figure 8.1 (Figure 8.1(a) and Figure 8.1(b)). The solid lines are the estimated posterior model probabilities against the sweeping times (after the burn-in) and the dotted lines are the estimated posterior model probabilities obtained at the end of the chain. We select the 100,000

sweeps after the burn-in to estimate the posterior model probabilities against the sweeping times and plot these posterior model probabilities with lag 500 (one in every 500 sweeps). We found the obvious stability of the estimation for the posterior model probabilities after the burn-in for both time series. Notice that in Figure 8.1(a), the estimated posterior model probabilities (against the sweeping times and at the end of chain) for $m = 3$ and $m = 0$ are all so close to 0 that they are not distinguishable from each other.



(a) The US real GNP series



(b) The US consumer price index

Figure 8.1: Estimated model probabilities after the burn-in.

The estimated posterior model probabilities are listed in Table 8.1. For convenience, we denote the posterior model probability $p(M_m|\mathbf{y})$ as $p(m|\mathbf{y})$. In each of

the two data sets, there are several competing values for the number of structural break components.

From Figure 8.1(a) we notice the rapid "cut-off" of the moves between models with different orders. The reason for this is, from Table 8.1, the very low acceptance rates for the between model moves. These low acceptance rates show the Markov chains we generated are not so easy to move between different state spaces. The RJMCMC algorithm is not so efficient for updating the structural break model, or the RJMCMC algorithm we designed is not so efficient because the RJMCMC algorithm has a lot of flexibility we have not tried yet. Usually the RJMCMC algorithm is easy to be constructed for nested models or hierarchical models. For other models, especially those changing a large number of variables at one between model move, the RJMCMC is hard to achieve high efficiency (see Brooks, Catchpole, Morgan and Harris, 2002). For the structural break model, we need to increase or decrease at least 3 variables at one between model move.

Different prior or different hyperparameters in the "diffuse" prior could be considered. For example, we could use economists' opinion to elicit prior for model M_m, which is dependent of which economic time series we would like to make inference. For example, for the US real GNP series, economists usually believe that there are one or two structural breaks. (See Busetti and Harvey, 2001.) We might allocate bigger prior probabilities for the models with 1 or 2 structural break points than the others. However, since the posterior model probabilities with 1 or 2 structural break points are much bigger than the others, this change of priors for models would have no effect on the "mixture" of the Markov chains. Other consideration such as using uniform distributions for $\boldsymbol{\tau}$ made no significant difference on the posteriors.

For the US real GNP data, the posterior distribution for m favours 1 or 2 structural break components, which coincides with the discussion of Busetti and Harvey (2001). The posterior model probabilities for US real GNP data concentrate on $m = 2$, which means we favour two structural breaks rather than any other number, with a probability of about 0.94.

For the US consumer price data, the posterior distribution for m favours between 4 and 6 structural break components. The posterior probability for $m = 5$ is much higher than the others, so our approach strongly favours 5 structural break

components in the US consumer data set.

Data set	Sample size	$p(m\|\mathbf{y})$		Acceptance ratio
real	61	$p(0) = 0.000048$	$p(1) = 0.060929$	0.2490%
GNP		$p(2) = 0.938760$	$p(3) = 0.000263$	
consumer	111	$p(4) = 0.060147$	$p(5) = 0.925994$	0.5135%
price		$p(6) = 0.013859$		

Table 8.1: Posterior distribution of m for two data sets.

The predictive means of $\tilde{y}_t$, $E(\tilde{y}_t|\mathbf{y})$ for the US real GNP data, conditional on different numbers of structural breaks are plotted in Figure 8.2. It is difficult to distinguish between predictive means for $m = 2$ and $m = 3$ from the graph since both follow the actual data very closely.

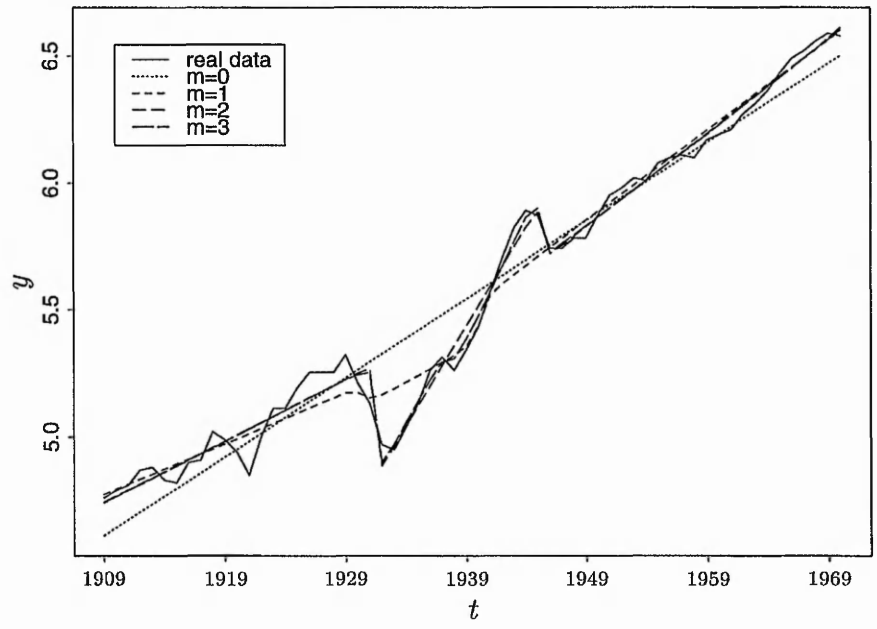


Figure 8.2: US Real GNP and fitted trend with different number of structural breaks.

We also plotted the predictive means of $\tilde{y}_t$, $E(\tilde{y}_t|\mathbf{y})$, for the US consumer price index series as estimates of the trend conditional on different dimensionalities of model together with the US consumer price index series in Figure 8.3. The predictive

means follows the real data closely for $m = 5$ and $m = 6$ and it is difficult to distinguish between them. The plot of the predictive means for $m = 4$ is a little away from the real data.

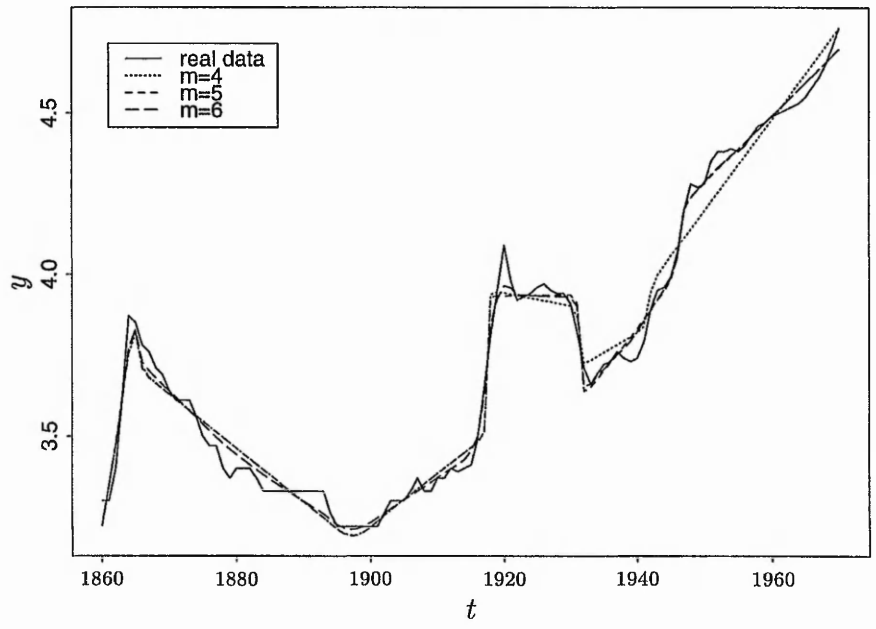


Figure 8.3: US Consumer price and fitted trend with different number of structural breaks.

We can also obtain the posterior densities for parameters of different models with different dimensionalities. We presented the posterior means, standard deviations and 95% posterior probability intervals for parameters of our preferred model (which has the largest posterior model probability) for both the US real GNP and the US consumer price index series in Table 8.2. For the US real GNP series, the posterior means of τ_1 and τ_2 in the preferred model (with 2 structural break points) are 0.3686 and 0.5965, corresponding to the years 1931 and 1945 respectively. For the US consumer price index series, the preferred model has 5 structural break points. The posterior means for τ_i, $i = 1, 2, \ldots, 5$ are 0.04467, 0.3386, 0.5224, 0.653 and 0.7873, corresponding to 1864, 1897, 1917, 1931 and 1946 respectively.

We plotted the posterior densities for parameters of the preferred model with two structural breaks in Figure 8.4, for the US real GNP data. We notice that the support of the posterior density for τ_1 concentrates on the interval $[0.361, 0.377]$,

	US real GNP			US Consumer price		
	mean	sd	probability intervals (95%)	mean	sd	probability intervals (95%)
α_1	4.743	0.03308	(4.68, 4.81)	3.223	0.05581	(3.12, 3.34)
α_2	3.087	0.1745	(2.76, 3.44)	3.795	0.03716	(3.72, 3.87)
α_3	4.342	0.1172	(4.12, 4.58)	2.578	0.1547	(2.26, 2.87)
α_4				3.908	0.4022	(3.22, 4.75)
α_5				1.543	0.5889	(-0.287, 2.38)
α_6				2.424	0.2381	(1.97, 2.84)
β_1	0.02405	0.002592	(0.019, 0.0292)	0.1262	0.03119	(0.045, 0.173)
β_2	0.07822	0.005927	(0.0664, 0.0894)	-0.01664	0.001723	(-0.0201, -0.0136)
β_3	0.03705	0.002352	(0.0323, 0.0416)	0.01613	0.003208	(0.0101, 0.0226)
β_4				0.0004104	0.006358	(-0.013, 0.0111)
β_5				0.02863	0.007279	(0.018, 0.0508)
β_6				0.02066	0.002368	(0.0164, 0.0252)
τ_1	0.3686	0.005511	(0.361, 0.377)	0.04467	0.009509	(0.0231, 0.0596)
τ_2	0.5965	0.01113	(0.575, 0.611)	0.3386	0.02112	(0.301, 0.378)
τ_3				0.5224	0.00411	(0.515, 0.527)
τ_4				0.653	0.01657	(0.639, 0.721)
τ_5				0.7873	0.01556	(0.773, 0.839)
σ	0.08127	0.007656	(0.068, 0.098)	0.06879	0.004915	(0.06, 0.0793)

Table 8.2: Posterior means and standard deviations for the parameters in the preferred models for two US series.

corresponding to 1930 to 1931.

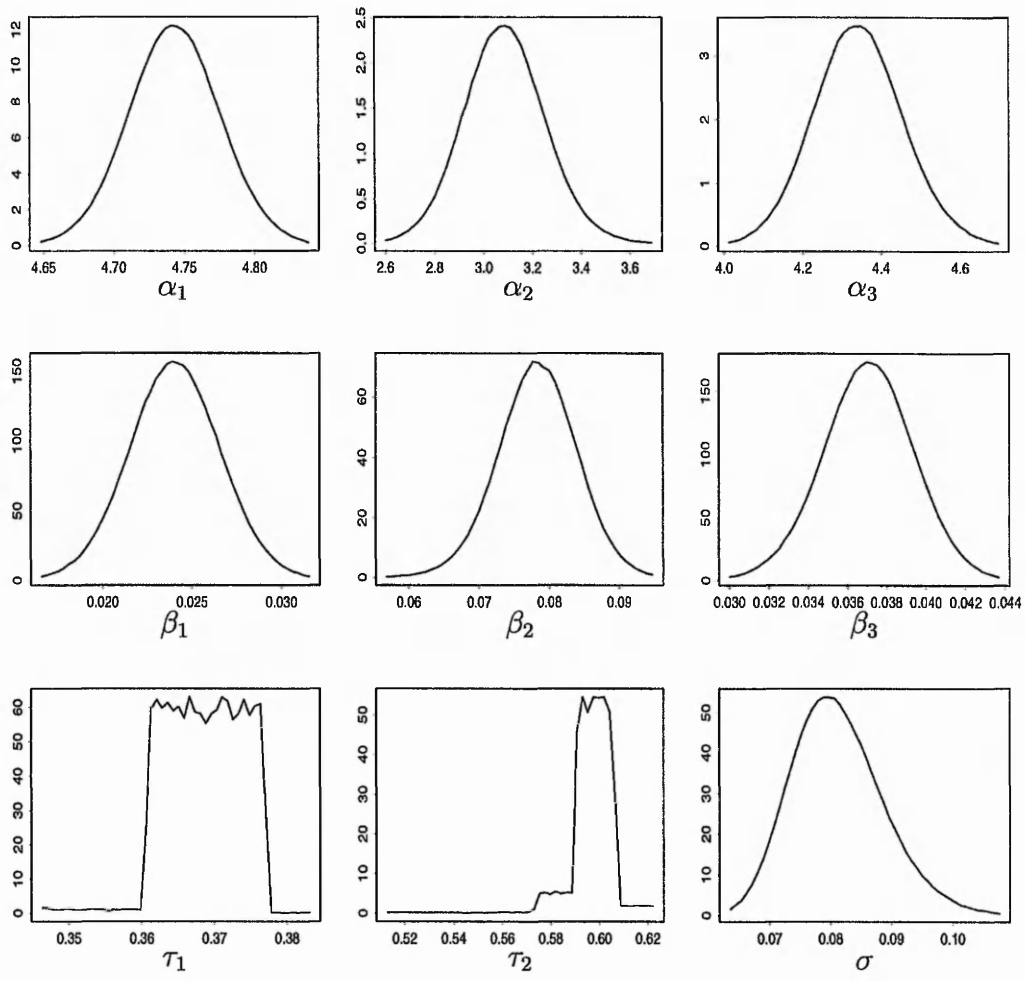
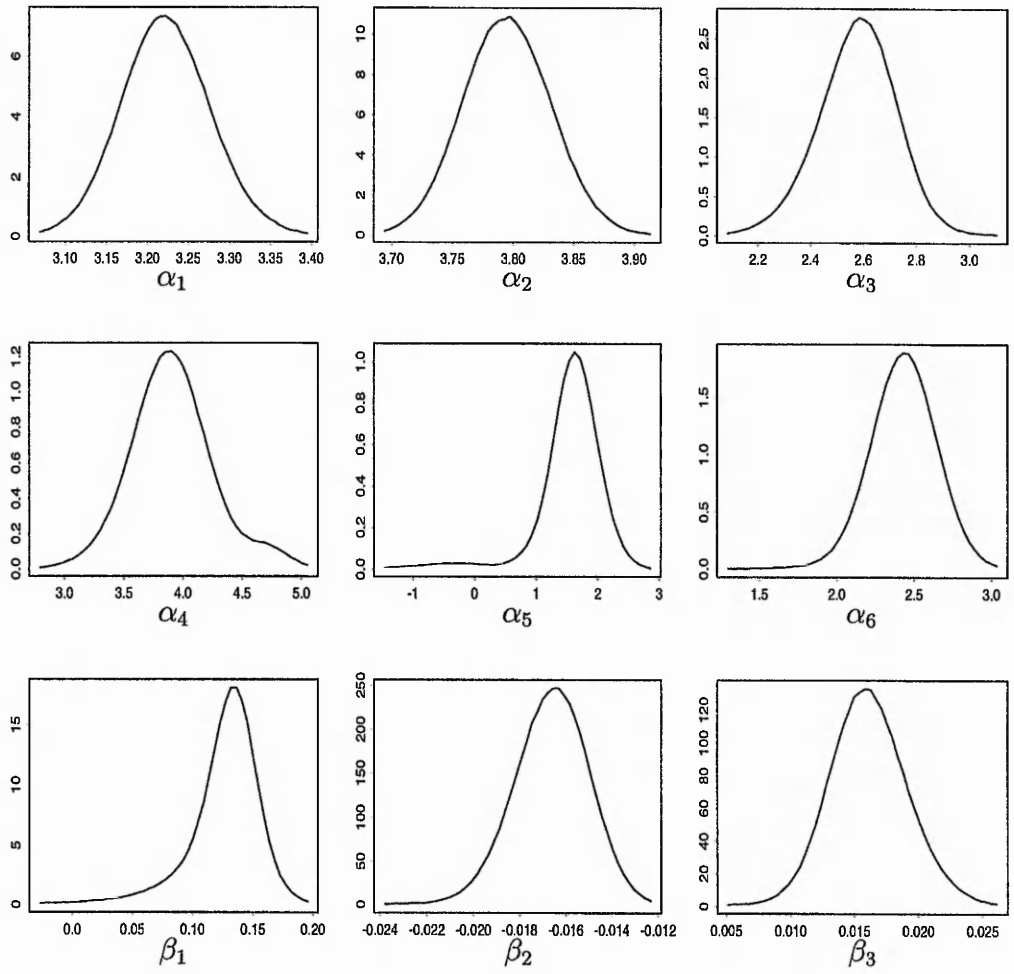


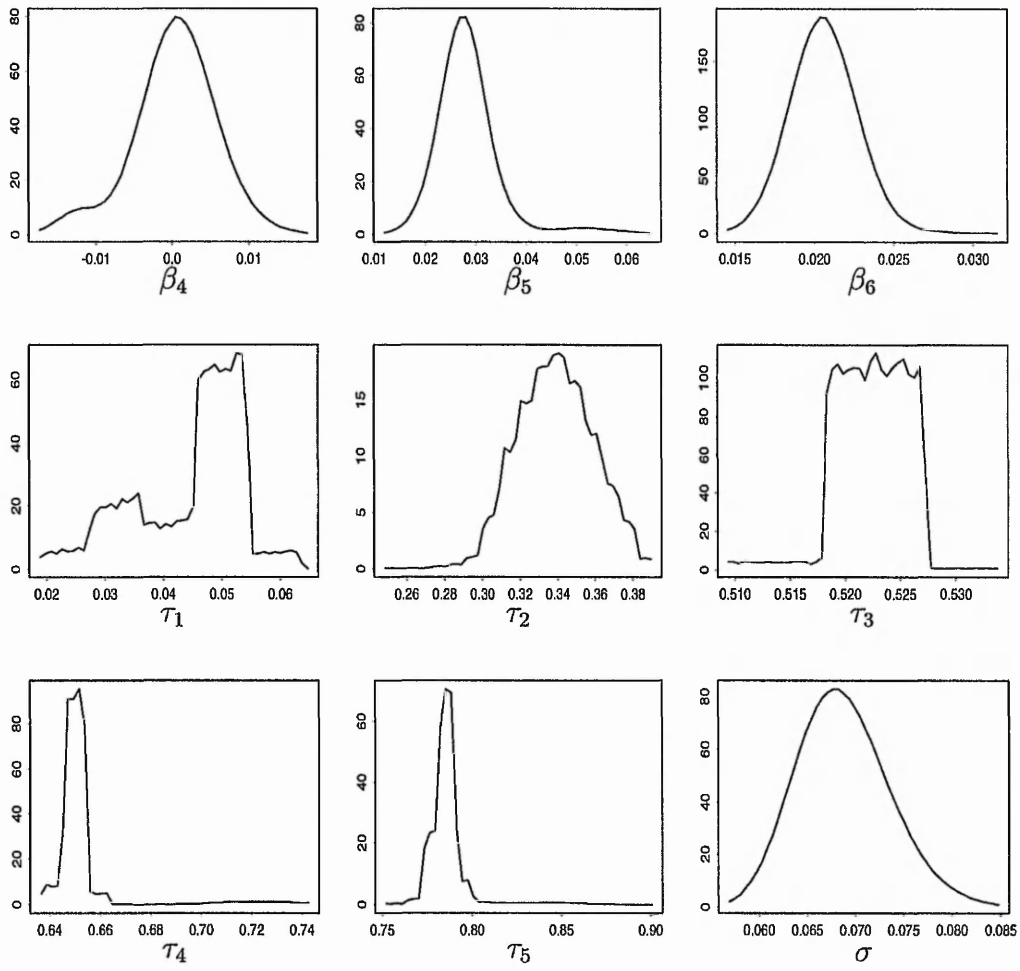
Figure 8.4: Posterior densities of 2 structural break model for US real GNP series.

Finally we obtain the posterior densities for parameters in the model for the US consumer price series, with the largest posterior model probability, that is $m = 5$, and present plots in Figure 8.5 (Figure 8.5(a) and Figure 8.5(b)).



(a) first part

Figure 8.5: Posterior densities of model with 5 structural breaks for US consumer price index.



(b) second part

Figure 8.5: Posterior densities of model with 5 structural breaks for US consumer price index.

Chapter 9

Conclusion

In this thesis we have discussed two kind of structural change models, the two-phase model and the structural break model for the abrupt structural change at some point(s); the smooth transition model for the gradual structural change over a period of time. We obtained posterior densities for parameters in these models, which were used to make inferences about parameters, and posterior model probabilities which were used to select between models. Different priors were adopted as well as noise terms and different Markov chain Monte Carlo sampling methods were used such as the Gibbs sampler, the Metropolis-Hastings sampler, the slice sampler, the adaptive rejection Metropolis sampling and the reversible jump Markov chain Monte Carlo.

In Chapters 2 and 3, we discussed the two-phase models. We adopted different priors for parameters and considered the noise term both as a white noise and an AR(1) process.

From Chapter 4 to Chapter 6, we discussed the smooth transition model, starting with the simple one with one smooth transition component. We considered the noise term as both a white noise and an AR(1) process. We have obtained the posterior model probabilities which were used for model selection and posterior densities for the parameters in the model which were used for parameter estimation. As an alternative to obtaining a closed expression for the posterior densities for all of the parameters, we used Markov chain Monte Carlo methods to sample directly from the posterior joint density of the parameters. The slice sampler was used for the sampling and both Raftery and Lewis and Gelman and Rubin diagnostics were used to make an assessment of the convergence of the Markov chains. Other

supplementary tools including autocorrelation functions and the sample paths of the chains were also considered to assist the diagnostics. We fitted the smooth transition model to two real time series, the British industrial production index and the US quarterly GDP series.

In Chapter 7, we discussed the double smooth transition model, which has two smooth transition components. In order to apply our approach to fit the model to some real data (which might have a wide range of values for the parameter γ), we developed an Adaptive Rejection Metropolis Sampling (ARMS) to sample from the marginal posterior densities for the two speed parameters and the two midpoint parameters. A reparameterisation of the speed parameters was adopted in order not to impose any limit on the range of speeds.

In Chapter 8, we applied the reversible jump Markov chain Monte Carlo (RJMCMC) to the structural break model with an unknown number of structural break points. The structural break model consists of segments which change their levels and slopes between the structural break points. In our approach, several methods of MCMC were adopted including the Gibbs sampler for parameters whose full conditionals are standard distributions, the Metropolis-Hastings sampler for the break point parameter τ and the reversible jump Markov chain Monte Carlo sampler for updating the number of breaks in the model. So our approach can be called a hybrid sampler in the terminology of Tierney (1994). We fitted the structural break model to the US real GNP series and the US consumer price index series. Our approach automatically selects the order of the dimensions of the model for each of these real time series by providing the estimated posterior model probabilities. We also obtained the posterior density, mean, standard deviation and 95% probability interval for each parameter in the preferred model (the one with the largest estimated posterior model probability). The predictive means are used to fit the model to the two real time series.

Further work could include the following: An AR(p) process as noise term could be explored for both the structural break model and the smooth transition model. Although our approach for the double smooth transition model in Chapter 7 was successful, a further consideration would be to address the problem of an efficient Metropolis-Hastings sampler for the parameters (γ, τ). The convergence properties

of the Metropolis-Hastings sampler will not automatically apply to those of the ARMS sampler because the ARMS sampler lacks time homogeneity. A possible justification of the ARMS, suggested in Robert and Casella (1999), pp250–251, is to revert to homogeneous chains by fixing the proposal distributions after a burn-in period. Other details that could be addressed are: how long a burn-in could be needed and how to fix the proposal distribution because in an ARMS sampler, the proposal distributions vary from step to step. A good Metropolis-Hastings sampler for γ, τ could make it possible to apply the reversible jump Markov chain Monte Carlo method to the smooth transition model with unknown number of components. Choosing between the structural break model and smooth transition model with unknown number of components for both models could also be addressed since we have found that the smooth transition model can behave like a structural break model when the speed γ is very large (see Chapter 4).

Appendix A

Appendix

A.1 Probability Distribution Functions

Here we provide some probability distribution functions (pdf) used in this work for reference.

A.1.1 Univariate Probability Distribution Functions

The Beta PDF

A random variate, x, is distributed in the beta form, if and only if it has the following probability density function (pdf)

$$p(x;a,b) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}, \tag{A.1}$$

where $x \in (0,1)$, $a,b > 0$ and $B(a,b)$ is the beta function as

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx. \tag{A.2}$$

The mode is $z_{\text{mod}} = \frac{a-1}{a+b-2}$, if $a,b > 1$. The mean is $\frac{a}{a+b}$ and the variance is $\frac{ab}{(a+b)^2(a+b+1)}$.

The Bilateral Chi PDF

A random variate, x, is distributed in the bilateral chi form, if and only if, it has the following pdf:

$$p(x;a,b,c) = \left[b2^{c/2}\Gamma(c/2)\right]^{-1} \left|\frac{x-1}{b}\right|^{c-1} e^{-\frac{1}{2}\left(\frac{x-a}{b}\right)^2}, \tag{A.3}$$

where $b, c > 0$, and $E(x) = a$, $\text{var}(x) = cb^2$.

The Inverted Gamma PDF

A random variate, σ, is distributed in the inverted gamma form, if and only if, it has the following pdf:

$$p(\sigma|\nu, s) = \frac{2}{\Gamma(\nu/2)} \left(\frac{\nu s^2}{2}\right)^{\nu/2} \frac{1}{\sigma^{\nu+1}} e^{-\frac{\nu s^2}{2\sigma^2}}, \tag{A.4}$$

where $\sigma > 0$, $\nu, s > 0$ are two parameters and denoted by $IG(\nu, \nu s^2)$. The first two moments about zero of (A.4) are $\mu_1 = \frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)} \left(\frac{\nu}{2}\right)^{1/2} s, \quad \nu > 1$ and $\mu_2 = \frac{\nu s^2}{\nu - 2}, \quad \nu > 2$. Generally we have the moments about zero of order r as $\mu_r = \frac{\Gamma[(\nu-r)/2]}{\Gamma(\nu/2)} \left(\frac{\nu s^2}{2}\right)^{r/2}, \quad \nu > r.$

We sometimes use another form of the inverted gamma distribution by letting $x = \sigma$, $a = \frac{2}{\nu s^2}$ and $b = \frac{\nu}{2}$,

$$p(x|a, b) = \frac{2}{\Gamma(b) a^b x^{2b+1}} e^{-\frac{1}{ax^2}}, \tag{A.5}$$

where $x > 0$, $a, b > 0$. The first two moments about zero of (A.5) are $\mu_1 = \frac{\Gamma(b-1/2)}{\Gamma(b)\sqrt{a}}, \quad b > 1/2$ and $\mu_2 = \frac{1}{a(b-1)}, \quad b > 1.$

The Gamma PDF

A random variate, γ, is distributed in the gamma form, if and only if, it has the following pdf:

$$p(\gamma|a, b) = \frac{\gamma^{a-1} e^{-\gamma/b}}{b^a \Gamma(a)}, \tag{A.6}$$

where $\gamma > 0$ and the two parameters $a > 0, b > 0$. The mean is $E(\gamma|a, b) = ab$, the variance is $\text{var}(\gamma|a, b) = ab^2$ and the mode $\text{mode}(\gamma|a, b) = (a-1)b$, for $a \geq 1$. We denote it as $G(a, b)$.

The gamma distribution has a simple property which is sometimes used in generating sample data from a gamma distribution. If we have $X \sim G(a, 1)$, then $Y = bX \sim G(a, b)$. So we can always generate data from $G(a, 1)$ and then multiply the data by b.

The relationship between gamma distribution and inverse gamma distribution is simple. If we have X is distributed in gamma, $X \sim G(a, b)$, then $1/X$ is distributed

in inverse gamma. In order to obtain the inverse gamma form in equation (A.4), we have to set the transformation of the variable and two parameters as $\sigma = 1/\sqrt{X}$, $\nu = 2a$, $\nu s^2 = 2/b$, then we have $\sigma \sim IG(\nu, \nu s^2)$.

The Logistic PDF

A random variate, t, is distributed in the logistic form, if and only if, it has the following pdf:

$$p(t; a, b) = \frac{be^{-b(t-a)}}{(1 + e^{-b(t-a)})^2}, \tag{A.7}$$

where $-\infty < t < \infty$, $b > 0$. $E(t) = a$ and $\text{var}(t) = \frac{\pi^2}{3b^2}$.

A.1.2 Multivariate Probability Distribution Functions

The Multivariate Normal PDF

A random vector, $\mathbf{x} = (x_1, x_2, \ldots, x_m)$ is distributed in the multivariate normal form if, and only if, it has the following pdf:

$$p(\mathbf{x}|\boldsymbol{\theta}, \Sigma) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{m/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\theta})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\theta})\right\}, \tag{A.8}$$

where $-\infty < x_i < \infty$, $i = 1, 2, \ldots, m$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_m)'$, with $-\infty < \theta_i < \infty$, $i = 1, 2, \ldots, m$ and $\Sigma = (\sigma_{ij})$ is an $m \times m$ positive definite symmetric (PDS) matrix. We have $E(\mathbf{x}|\boldsymbol{\theta}, \Sigma) = \boldsymbol{\theta}$ and $\text{Cov}(\mathbf{x}|\boldsymbol{\theta}, \Sigma) = \sigma_{ij}$. If we set $\Sigma^{-1} = \frac{1}{\sigma^2}V^{-1}$, σ is a constant, then we have

$$p(\mathbf{x}|\sigma, \boldsymbol{\theta}, V) = \frac{|V|^{-1/2}}{\left(\sqrt{2\pi}\sigma\right)^m} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{x} - \boldsymbol{\theta})'V^{-1}(\mathbf{x} - \boldsymbol{\theta})\right\}. \tag{A.9}$$

We denote $V = (v_{ij})$, $\Sigma = \sigma^2 V$, so $\sigma_{ij} = \sigma^2 v_{ij}$. We have $\text{Cov}(\mathbf{x}|\sigma, \boldsymbol{\theta}, V) = \sigma^2 v_{ij}$.

The Multivariate Student t PDF

A random vector, $\mathbf{x}' = (x_1, x_2, \ldots, x_m)$ is distributed in the multivariate student t form if, and only if, it has the following pdf:

$$p(\mathbf{x}|\boldsymbol{\theta}, V, \nu, m) = \frac{\nu^{\nu/2}\Gamma[(v + m)/2]|V|^{1/2}}{\pi^{m/2}\Gamma(\nu/2)}[\nu + (\mathbf{x} - \boldsymbol{\theta})'V(\mathbf{x} - \boldsymbol{\theta})]^{-(m+\nu)/2}, \tag{A.10}$$

where $\nu > 0$, V is an $m \times m$ positive definite symmetric (PDS) matrix, and $\boldsymbol{\theta}' = (\theta_1, \theta_2, \ldots, \theta_m)$, with $-\infty < \theta_i < \infty$, $i = 1, 2, \ldots, m$.

A.2 Results of Multivariate Normal Regression

Here we list some results of multivariate normal regression as formulae because we will use them frequently. These formulae can be found in Zellner (1987), p66.

With the multivariate normal regression model

$$\mathbf{y} = X\boldsymbol{\beta} + \mathbf{u}, \tag{A.11}$$

where X is an $n \times k$ matrix, with rank k, of observations on k independent variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients, $\mathbf{u}$ is an $n \times 1$ vector of disturbance (error terms) which are *iid* $\mathrm{N}(0, \sigma^2)$.

The likelihood function for the model is

$$p(\mathbf{y}|X, \boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\beta})'(\mathbf{y} - X\boldsymbol{\beta})\right\}. \tag{A.12}$$

Sometimes in order to highlight the structure of $\boldsymbol{\beta}$, we can rewrite the likelihood as

$$p(\mathbf{y}|X, \boldsymbol{\beta}, \sigma) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'X'X(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right]\right\}. \tag{A.13}$$

where $\nu = n - k$, $\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{y}$ and $\nu s^2 = (\mathbf{y} - X\hat{\boldsymbol{\beta}})'(\mathbf{y} - X\hat{\boldsymbol{\beta}})$.

A.3 Formulae

When we use Bayesian method to analysis model, we always need to calculate the posterior density which is proportional to the likelihood times the prior. In a lot of cases, the posterior densities for some of the parameters, if not all of them, conditional on the others, are multivariate normal distribution. In the following, we write down the formulae which have been used frequently in this thesis and give a derivation for one of them. For simplicity, we only write out the exponential part of the conduction, since complete formulae are easy to be given out with these results. We have two expressions of the likelihood functions above, so we give out two cases of the conduction accordingly. These two formulae are used in this work frequently.

A.3.1 Formula 1

We have the following denotes as $\mathbf{y}_1$, $\mathbf{y}_2$ are $n_1 \times 1$, $n_2 \times 1$ vector respectively, $\boldsymbol{\beta}$ is a $k \times 1$ vector and X_1, X_2 are $n_1 \times k$, $n_2 \times k$ matrix respectively. Then we have the

first formula as

$$\begin{aligned}&(\mathbf{y}_1 - X_1\beta)'(\mathbf{y}_1 - X_1\beta) + (\mathbf{y}_2 - X_2\beta)'(\mathbf{y}_2 - X_2\beta)\\&= \nu s^2 + (\beta - \hat{\beta})'V(\beta - \hat{\beta}),\end{aligned} \tag{A.14}$$

where

$$\begin{aligned}V &= X_1'X_1 + X_2'X_2,\\\hat{\beta} &= V^{-1}(X_1'\mathbf{y}_1 + X_2'\mathbf{y}_2),\\\nu s^2 &= (\mathbf{y}_1 - X_1\hat{\beta})'(\mathbf{y}_1 - X_1\hat{\beta}) + (\mathbf{y}_2 - X_2\hat{\beta})'(\mathbf{y}_2 - X_2\hat{\beta}),\\\nu &= n_1 + n_2 - k.\end{aligned}$$

This formula can be found in Zellner (1987), p71.

A.3.2 Formula 2.

Here we provide another form of the above formula as formula 2. Usually, if a prior for parameter vector is a multivariate normal distribution, the density function is presented in the second form of the likelihood function in (A.13) so that formula 2 presented below sometimes is more frequently to be used, for example, in my works. Formula 2 is as

$$\begin{aligned}&(\beta - \beta_1)'V_1(\beta - \beta_1) + (\beta - \beta_2)'V_2(\beta - \beta_2)\\&= \nu s^2 + (\beta - \beta_0)'V(\beta - \beta_0).\end{aligned} \tag{A.15}$$

where V_1 and V_2 are $k \times k$ positive definite symmetric matrix, β is a k dimension vector variable, β_1 and β_2 are two k dimension constant vectors and

$$\begin{aligned}V &= V_1 + V_2,\\\beta_0 &= V^{-1}(V_1\beta_1 + V_2\beta_2),\\\nu s^2 &= (\beta_0 - \beta_2)'V_2(\beta_1 - \beta_2)\\&= (\beta_0 - \beta_1)'V_1(\beta_2 - \beta_1).\end{aligned}$$

We will derive the expressions for V, β_0 and νs^2, in the mean time, derive formula 2.

We have

$$\begin{aligned}&(\beta - \beta_1)'V_1(\beta - \beta_1)\\&= (\beta - \beta_0 + \beta_0 - \beta_1)'V_1(\beta - \beta_0 + \beta_0 - \beta_1)\\&= (\beta - \beta_0)'V_1(\beta - \beta_0) + 2(\beta_0 - \beta_1)'V_1(\beta - \beta_0)\\&\quad +(\beta_0 - \beta_1)'V_1(\beta_0 - \beta_1),\end{aligned}$$

because $V_1' = V_1$ is a symmetric matrix. Notice that β only appears in the first two terms in the last equation. Similarly, we have

$$\begin{aligned}
&(\beta - \beta_2)'V_2(\beta - \beta_2) \\
&= (\beta - \beta_0 + \beta_0 - \beta_2)'V_2(\beta - \beta_0 + \beta_0 - \beta_2) \\
&= (\beta - \beta_0)'V_2(\beta - \beta_0) + 2(\beta_0 - \beta_2)'V_2(\beta - \beta_0) \\
&\quad +(\beta_0 - \beta_2)'V_2(\beta_0 - \beta_2).
\end{aligned}$$

In order for the formula to be established, we need to select suitable value of β_0 so that

$$2(\beta_0 - \beta_1)'V_1(\beta - \beta_0) + 2(\beta_0 - \beta_2)'V_2(\beta - \beta_0) \equiv 0,$$

for any vector variable β. This is equivalent to

$$[(\beta_0 - \beta_1)'V_1 + (\beta_0 - \beta_2)'V_2](\beta - \beta_0) \equiv 0.$$

The following condition is sufficient for the above equivalence to be valid

$$(\beta_0 - \beta_1)'V_1 + (\beta_0 - \beta_2)'V_2 = 0,$$

which leads to

$$\beta_0 = (V_1 + V_2)^{-1}(V_1\beta_1 + V_2\beta_2).$$

So we have

$$\begin{aligned}
&(\beta - \beta_1)'V_1(\beta - \beta_1) + (\beta - \beta_2)'V_2(\beta - \beta_2) \\
&= (\beta - \beta_0)'(V_1 + V_2)(\beta - \beta_0) \\
&\quad +(\beta_0 - \beta_1)'V_1(\beta_0 - \beta_1) + (\beta_0 - \beta_2)'V_2(\beta_0 - \beta_2).
\end{aligned}$$

Substitute $(\beta_0 - \beta_1)'V_1$ with $-(\beta_0 - \beta_2)'V_2$, we obtain

$$\begin{aligned}
&(\beta - \beta_1)'V_1(\beta - \beta_1) + (\beta - \beta_2)'V_2(\beta - \beta_2) \\
&= (\beta - \beta_0)'(V_1 + V_2)(\beta - \beta_0) + (\beta_0 - \beta_2)'V_2(\beta_1 - \beta_2).
\end{aligned}$$

Or substitute $(\beta_0 - \beta_2)'V_2$ with $-(\beta_0 - \beta_1)'V_1$, we obtain

$$\begin{aligned}
&(\beta - \beta_1)'V_1(\beta - \beta_1) + (\beta - \beta_2)'V_2(\beta - \beta_2) \\
&= (\beta - \beta_0)'(V_1 + V_2)(\beta - \beta_0) + (\beta_0 - \beta_1)'V_1(\beta_2 - \beta_1).
\end{aligned}$$

This is the formula we want to obtain.

A.4 British industrial production index

Here is the British industrial production index, from year 1700 to year 1913, with 214 observations, described in the text which incorporates the Hartley-Lewis revisions.[1] Sometimes it is not so easy to obtain actual data especially those historical data. For convenience of the reader, the author list the data in Table A.1.

seq.	index									
1	1.82	2.24	1.53	1.87	1.92	2.03	1.65	1.69	1.92	1.93
2	1.82	1.82	1.82	2.01	2.10	2.24	2.46	2.58	2.55	2.38
3	2.47	2.43	2.76	2.72	2.19	2.59	2.52	2.47	2.29	2.32
4	2.48	2.38	2.43	2.72	2.39	2.67	2.64	2.46	2.86	2.70
5	2.47	2.27	2.50	2.39	2.74	2.62	2.81	2.90	3.16	2.83
6	3.12	3.14	3.13	3.18	3.14	3.18	2.88	2.98	2.96	2.90
7	2.86	3.02	2.90	2.93	3.11	3.05	3.27	3.54	3.39	3.59
8	3.51	3.60	3.72	3.60	3.21	3.33	3.50	3.51	3.65	3.33
9	3.46	3.63	3.56	3.78	4.08	4.06	4.03	4.22	4.21	4.34
10	4.36	4.52	4.85	4.55	4.45	4.55	4.71	4.64	4.75	5.38
11	5.32	4.98	5.25	5.30	5.48	5.60	5.65	5.87	5.58	5.72
12	6.22	6.54	6.15	6.20	6.31	6.85	6.70	7.32	7.62	7.37
13	7.56	7.88	8.30	8.75	9.22	10.10	9.19	10.40	11.10	10.7
14	11.7	11.9	11.8	12.5	13.2	13.7	15.1	14.3	15.6	16.9
15	16.6	16.8	16.0	16.9	19.1	20.0	20.0	19.0	21.0	21.7
16	21.7	22.6	24.0	26.0	26.4	26.3	28.1	29.1	28.5	30.0
17	31.7	31.7	32.4	32.5	35.0	37.3	38.7	36.4	36.4	35.8
18	40.2	43.5	44.8	45.3	46.4	46.7	47.5	47.4	47.3	45.6
19	50.3	53.5	55.7	56.5	54.4	52.1	51.0	55.1	58.3	62.4
20	63.3	64.1	61.0	60.0	63.5	66.5	71.4	73.4	77.0	80.1
21	80.1	80.3	81.7	80.0	81.0	85.7	82.5	85.6	83.8	86.0
22	88.6	92.1	96.1	100.0						

Table A.1: British industrial production index, from 1700 to 1913.

[1]Source: Crafts, Leybourne and Mills (1989).

Appendix B

Bibliography

1. J.H. Ahrens and U. Dieter. (1974) Computer methods for sampling from gamma, beta, poisson and binomial distribution. *Computing, 12, 223–246.*

2. D. W. Bacon and D. G. Watts. (1971) Estimating the Transition between Two Intersecting Straight Lines. *Biometrika 58, 525–34.*

3. J. M. Bernardo and A. F.M. Smith. (1994) *Bayesian Theory.* Wiley.

4. J. Besag. (2000) Markov Chain Monte Carlo for Statistical Inference. *Working Paper, No. 9, University of Washington, USA.*

5. J. Besag and P. J. Green. (1993) Spatial Statistics and Bayesian Computation. *Journal of the Royal Statistical Society, Series B., 55, 25–37.*

6. D. J. Best. (1978) Letter to the Editor. *Applied Statistics (Ser. C), 27, 181.*

7. G. Box and G. Jenkins. (1976) *Time series analysis: forecasting and control.* Holden-Day.

8. L. D. Broemeling. (1985) *Bayesian Analysis of Linear models.* Marcel Dekker, INC..

9. L. D. Broemeling and H. Tsurum. (1987) *Econometrics and Structural Change.* Marcel Dekker, INC..

10. S. Brooks. (1997) Discussion of the Paper by Richardson and Green. *J. R. Statist. Soc. B, Vol. 59, No. 4, 731–792.*

11. S. Brooks, E. Catchpole, B. Morgan and M. Harris. (2002) Bayesian Methods for Analysing Ringing Data. *Journal of Applied Statistics, Vol. 29, Nos. 1–4, 187–206.*

12. S. Brooks and A. Gelman. (1998) General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics, Vol. 7, No. 4, 434–455.*

13. S. Brooks, P. Giudici and G. Roberts. (2000) Efficient Construction of Reversible Jump MCMC Proposal Distributions. *Preprint.*

14. S. Brooks and G. O. Roberts. (1998) Assessing Convergence of Markov Chain Monte Carlo Algorithms. *Statistics and Computing 8, 319–335.*

15. F. Busetti and A. Harvey. (2001) Testing for the Presence of a Random Walk in Series with Structural Breaks. *Journal of Times Series Analysis. Vol. 22, No. 2, 127–150*

16. G. Casella and E. I. George. (1992) Explaining the Gibbs Sampler. *The American Statistician, Vol. 46, No. 3, 167–174.*

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

18. R. C. H. Cheng and G. Feast. (1979) Some simple gamma variate generators. *Appl. Statist. 28, 290–295.*

19. R. Christensen. (1984) *Data Distributions—A Statistical Handbook.* Entropy Ltd..

20. S. Chib and E. Greenberg. (1995) Understanding the Metropolis-Hastings Algorithm. *The American Statistician, Vol. 49, No. 4, 327–335.*

21. D. Cook and S. Weisberg. (1982) *Residuals and Influence in Regression.* London: Chapman and Hall.

22. M. K. Cowles, G. O. Roberts and J. S. Rosenthal. (2000) Possible Biases Induced by MCMC Convergence Diagnostics. *Statistical Computation and Simulation, Vol. 1, No. 5, 1–18.*

23. N. F. R. Crafts, S. J. Leybourne and T. C. Mills (1989) Trends and Cycles in British Industrial Production, 1700–1913. *J. R. Statist. Soc., Ser. A, 152, 43–60.*

24. P. Damien, J. Wakefield and S. Walker. (1999) Gibbs Sampling for Bayesian Non-conjugate and Hierarchical Models by Using Auxiliary Variables. *J. R. Statist. Soc. B, Vol. 61, Part 2, 331–344.*

25. R. B. Davies. (1977) Hypothesis Testing When a nuisance Parameter is Present only Under the Alternatives. *Biometrika, 64, 247–254.*

26. R. B. Davies. (1987) Hypothesis Testing When a nuisance Parameter is Present only Under the Alternatives. *Biometrika, 74, 33–43.*

27. P. J. Davis and P. Rabinowitz. (1984) *Methods of Numerical Integration.* Academic Press, INC..

28. P. R. Edwards and A. D. Sokal. (1988) Generalisation of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm. *Phys. Rev. Lett., Vol. 38, 2009–2012.*

29. Y. Fan and S. Brooks. (2000) Bayesian Modelling of Prehistoric Corbelled Domes. *J. R. Statist. Soc, D, 49, 339–354.*

30. M. Ferelli and G. Tunnicliffe Wilson. (1990) Robust Estimation of Level and Trend. *Journal of Forecasting, Vol. 9, 151–172.*

31. E. B. Fowlkes. (1987) *A Folio of Distributions.* Marcel Dekker, INC..

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

33. A. Gelman and D. Rubin. (1992) Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science, Vol. 7, 457–511.*

34. J. Geweke. (1986) Bayesian Inference in Econometric Models Using Monte Carlo Integrations. *Econometrica Vol. 24, 1317–1339.*

35. J. Geweke. (1992) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments (with discussion). In *Bayesian Statistics 4*, J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith (Eds.), 169–193, Oxford University Press, Oxford.

36. C. J. Geyer. (1992) Practical Markov Chain Monte Carlo. *Statistical Science Vol. 7, No. 4, 473–511.*

37. W. R. Gilks and P. Wild. (1992) Adaptive Rejection Sampling for Gibbs Sampling. *Appl. Statist., Vol. 41, No. 2, 337–348.*

38. W. R. Gilks, N. G. Best and K. K. C. Tan. (1995) Adaptive Rejection Metropolis Sampling within Gibbs Sampling. *Appl. Statist., Vol. 44, No. 4, 455–472.*

39. W. R. Gilks, S. Richardson and D. J. Spiegelhalter. (1996) *Markov Chain Monte Carlo in Practice.* Chapman & Hall.

40. S. J. Godsill. (1998) On the Relationship between MCMC Model Uncertainty Methods. *Technical Report, CUED/F-INFENG/TR.305, Signal Processing and Communications Laboratory, Department of Engineering, University of Cambridge.*

41. C. W. J. Granger and T. Terasvirta. (1993) *Modelling Nonlinear Economic Relationships.* Oxford: Oxford University Press.

42. P. J. Green. (1995) Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika, 82, 4, 711–32.*

43. D. Greenaway, S. Leybourne and D. Sapsford. (1997) Modelling Growth (and Liberalisation) Using Smooth Transitions Analysis. *Economic Inquiry, Vol. 35, 798–814.*

44. D.V. Griffiths and I. M. Smith. (1991) *Numerical Methods for Engineers.* CRC Press.

45. D. I. Harvey and T. C. Mills. (2000) Unit Roots and Double Smooth Transitions. *Economic Research Paper No. 00/10, University of Loughborough.*

46. P. J. Harrison and C. F. Stevens. (1976) Bayesian Forecasting. *J. R. Statist. Soc. B, Vol. 38, 205–47.*

47. W. K. Hastings. (1970) Monte Carlo Sampling Methods Using Markov Chain and their applications. *Biometrika, Vol. 57, 97–109.*

48. D. M. Higdon. (1999) Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications. *J. Am. Statist. Ass. 93, 585–595.*

49. S. E. Hills and A. F. M. Smith. (1992) Parameterisation issues in Bayesian inference. *Bayesian Statistics 4, 227–246. Oxford University Press.*

50. D. V. Hinkley. (1969) Inference about the Intersection in Two-Phase Regression. *Biometrika, Vol. 56, 495–504.*

51. D. V. Hinkley. (1971) Inference in Two-Phase Regression. *Journal of the American Statistical Association, 736–743.*

52. J. G. Ibrahim and Ming-Hui Chen. (2000). Power Prior Distributions for Regression Models. *Statistical Science, Vol. 15, No. 1, 46–60.*

53. H. Jeffreys. (1961) *Theory of Probability (3rd edition).* Oxford University Press, Oxford.

54. A. Jennings, J. J. McKeown. (1992) *Matrix Computation, Second Edition.* John Wiley & Sons.

55. P. D. Jones, T. J. Osborn, and K. R. Briffa. (1997) Estimating Sampling Errors in Large-Scale Temperature Averages. *Journal of Climate 10, 2548–2568.*

56. P. D. Jones. (1994) Hemisphere surface air temperature variations: a reanalysis and an update to 1993. *J. Climate Vol. 7, 1794–1802.*

57. P. D. Jones, T. J. Osborn and K. R. Briffa. (1997) Estimating sampling errors in large-scale temperature averages. *J. Climate, Vol. 10, 2548–2568*

58. S. Leybourne, P. Newbold and D. Vougas. (1998) Unit Roots and Smooth Transitions. *Journal of Time Series Analysis Vol. 19, No. 1, 83–97.*

59. C. J. Lin and T. Terasvirta. (1994) Testing the Constancy of Regression Parameters against Continuous Structural Change. *Journal of Econometrics Vol. 62, 211–228.*

60. J. B. Kadane and L. J. Wolfson. (1998) Experiences in Elicitation. *The Statistician, 47, Part 1, 3–19.*

61. G. S. Maddla. (1977) *Econometrics.* New York: McGraw-Hill.

62. J. Marriott and P. Newbold. (1998) Bayesian Comparison of ARIMA and Stationary ARMA Models. *International Statistical Review Vol. 66, No. 3, 323–336.*

63. J. Marriott and P. Newbold. (2000) The Strength of Evidence for Unit Autoregressive Roots and Structural Breaks: A Bayesian Perspective. *Journal of Econometrics Vol. 98, No. 1, 1–25.*

64. MCMC Preprint Service. Web site: *http://www.statslab.cam.ac.uk/~mcmc/*, S. Brooks (administrator), Statistical Laboratory, University of Cambridge.

65. K. L. Mengerson, C. P. Robert and C. Guihenneuc-Jouyaux (1999) MCMC Convergence diagnostics: a review. In *Bayesian Statistics 6, 415–440.* Oxford University Press, Oxford.

66. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller. (1953) Equations of state calculations by fast computing machine. *J. Chem. Phys., Vol. 21, 1087–1091.*

67. A. Mira and L. Tierney. (1998) On the Use of Auxiliary Variables in Markov Chain Monte Carlo Sampling. *Tech. Report, University of Minnesota.*

68. J. F. Monahan. (1983) Fully Bayesian Analysis of ARMA Time Series Models. *Journal of Econometrics Vol. 21, 307–331.*

69. NAG Ltd. (1993) *NAG Fortran Library Manual, Mark 16, vol.1.*

70. J. C. Naylor and J. M. Marriott. (1996) A Bayesian Analysis of Non-Stationary AR Series. In *Bayesian Statistics 5, 705–712.* Oxford University Press.

71. R. M. Neal. (2000) Slice sampler. *Technical Report No. 2005, Department of Statistics, University of Toronto, Canada.*

72. C. R. Nelson and C. I. Plosser. (1982) Trends and random walks in macroeconomic time series; Some evidence and implications. *Journal of Monetary Economics, Vol.10, 139–162.*

73. N. Nicholls, G. V. Gruza, J. Jouzel, T. R. Karl, L. A. Ogallo and D. E. Parker. (1996) Observed climate variability and change. *Climate Change 1995: The IPCC Second Assessment, Cambridge University Press, Cambridge, 133–192.*

74. A. Nobile. (1998) A hybrid Markov chain for the Bayesian analysis of the multinomial probit model. *Statistics and Computing Vol. 8, 229–242.*

75. A. O'Hagan. (1998) Eliciting Expert Beliefs in Substantial Practical Applications. *The Statistician, 47, Part 1, 21–35.*

76. D. E. Parker, P. D. Jones, A. Bevan and C. K. Folland. (1994) Interdecade changes of surface temperature since the late 19th century. *Journal of Geophysical Research Vol.99, 14373–14399.*

77. D. E. Parker, C. K. Folland and M. Jackson. (1995) Marine surface temperature: observed variations and data requirements. *Climate Change Vol.31, 559–600.*

78. A. Raftery and S. Lewis. (1992a) How Many Iterations in the Gibbs Sampler? In *Bayesian Statistics 4.* J. O. Berger, J. M. Bernardo, A. P. Dawid and A. F. M. Smith (Eds). 763–773. Oxford University Press, Oxford.

79. A. Raftery and S. Lewis. (1992b) One Long Run with Diagnostics: implementation Strategies for Markov Chain Monte Carlo. *Statist. Sci., 7, 493–497.*

80. A. Raftery and S. Lewis. (1996) Implementing MCMC. In *Markov Chain Monte Carlo in Practice.* W. R. Gilks, S. T. Richardson and D. J. Spiegelhalter (Eds.) 115–130. Chapman & Hall, London.

81. S. Richardson and P. J. Green. (1997) On Bayesian Analysis of Mixtures with an Unknown Number of Components. *J. R. Statist. Soc. B, Vol. 59, No.4, 731–92.*

82. C. P. Robert and G. Casella. (1999) *Monte Carlo Statistical Methods.* Springer.

83. C. P. Robert, T. Rydén and D. M. Titterington. (2000) Bayesian Inference in Hidden Markov Models through the Reversible Jump Markov Chain Monte Carlo Method. *J. R. Statist. Soc. B, Vol. 62, Part 1, 57–75.*

84. G. O. Roberts and J. S. Rosenthal. (1999) Convergence of Slice Sampler Markov Chains. *J. R. Statist. Soc. B, Vol. 61, Part 3, 643–660.*

85. G. O. Roberts, A. Gelman and W. R. Gilks. (1997) Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. *The Annals of Applied Probability, Vol. 7, No. 1, 110–120.*

86. R. Y. Rubinstein. (1981) *Simulation and the Monte Carlo Method.* John Wiley & Sons.

87. Statlib, Department of Statistics at Carnegie Mellon University. Web site: *http://lib.stat.cmu.edu/.*

88. A. F. M. Smith and G. O. Roberts. (1993) Bayesian Computation *via* Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *J. R. Statist. Soc., B, 55, 3–23.*

89. R. H. Swendson and J. S. Wang. (1987) Nonuniversal Critical Dynamics in Monte Carlo Simulations. *Phys. Rev. Lett., Vol. 58, 86–88.*

90. M. A. Tanner. (1991) *Tools for Statistical Inference.* Springer-Verlag.

91. T. Teräsvirta. (1994) Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association, Vol. 89, No. 425, 208–218.*

92. T. Teräsvirta and H. M. Anderson. (1992) Characterising Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics, Vol. 7, S119–S136.*

93. P. A. Thomson and R. B. Miller. (1986) Sampling the future: A Bayesian approach to forecasting. *Journal of Business and Economic Statistics Vol.4, 427–436.*

94. L. Tierney. (1991) Exploring Posterior Distributions Using Markov Chains. In *Computer Science and Statistics: Proc 23rd Symp. Inreface (ed. E. Keramidas), 563–570.* Fairfax Station: interface Foundation.

95. L. Tierney. (1994) Markov Chains for Exploring Posterior Distributions (with Discussion). *Annal of Statistics, Vol. 22, 1701–1762.*

96. P. T. Troughton and S. J. Godsill. (1997) A Reversible Jump Sampler for Autoregressive Time Series, Employing Full Conditionals to Achieve Efficient Model Space Moves. *Technical Report CUED/F-INFENG /TR.304, Signal Processing and Communications Laboratory, Department of Engineering, University of Cambridge.*

97. R. Waagepetersen and D. Suzuki. (2001) A Tutorial on Reversible Jump MCMC with a View toward Applications in QTL-mapping. *Int. Stat. Rev., Vol. 69, (1), 49–61.*

98. J. C. Wakefield. (1992) Discussion on parameterisation issues in Bayesian inference. *Bayesian Statistics 4, 243–244. Oxford University Press.*

99. J. Wang and E. Zivot. (2000) A Bayesian Time Series Model of Multiple Structural Changes in Level, Trend and Variance. *Journal of Business & Economic Statistics, July, vol. 18, No. 3, 374–386.*

100. L. Wasserman. (2000a) Asymptotic Inference for Mixture Models Using Data-dependent Priors. *J. R. Statist. Soc. B, 62, part 1, 159–180.*

101. L. Wasserman. (2000b) Bayesian Model Selection and Model Averaging. *Journal of Mathematical Psychology, 44, 92–107.*

102. B. Yu. (1995) Discussion to Besag *et al* (1995) *Statistical Science, Vol. 10, 3–66.*

103. A. Zellner. (1987) *An Introduction to Bayesian Inference in Econometrics.* Robert E. Krieger Publishing Co..