Flexible Bent-Cable Models for Mixture Longitudinal Data

by

Shahedul Ahsan Khan

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics - Biostatistics

Waterloo, Ontario, Canada, 2010

© Shahedul Ahsan Khan 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Data showing a trend that characterizes a change due to a shock to the system are a type of changepoint data, and may be referred to as shock-through data. As a result of the shock, this type of data may exhibit one of two types of transitions: gradual or abrupt. Although shock-through data are of particular interest in many areas of study such as biological, medical, health and environmental applications, previous research has shown that statistical inference from modeling the trend is challenging in the presence of discontinuous derivatives. Further complications arise when we have (1) longitudinal data, and/or (2) samples which come from two potential populations: one with a gradual transition, and the other abrupt.

Bent-cable regression is an appealing statistical tool to model shock-through data due to the model's flexibility while being parsimonious with greatly interpretable regression coefficients. It comprises two linear segments (incoming and outgoing) joined by a quadratic bend. In this thesis, we develop extended bent-cable methodology for longitudinal data in a Bayesian framework to account for both types of transitions; inference for the transition type is driven by the data rather than a presumption about the nature of the transition. We describe explicitly the computationally intensive Bayesian implementation of the methodology. Moreover, we describe modeling only one type of transition, which is a special case of this more general model. We demonstrate our methodology by a simulation study, and with two applications: (1) assessing the transition to early hypothermia in a rat model, and (2) understanding CFC-11 trends monitored globally.

Our methodology can be further extended at the cost of both theoretical and computational extensiveness. For example, we assume that the two populations mentioned above share common intercept and slopes in the incoming and outgoing phases, an assumption that can be relaxed for instances when intercept and slope parameters could behave differently between populations. In addition to this, we discuss several other directions for future research out of the proposed methodology presented in this thesis.

Acknowledgements

I wish to express my sincere gratitude and appreciation to my supervisors Grace Chiu and Joel Dubin for their continual guidance, helpful supervision, constant support, and great patience from the initial to the final level, enabled me to develop an understanding of the subject and to prepare this thesis. Their perpetual enthusiasm in research had motivated all their advisees, including me. In addition, they were always accessible and willing to help their students. As a result, research life became smooth and rewarding for me.

I would like to thank Richard Cook and Shoja'eddin Chenouri for serving on my advisory committee, and providing me valuable suggestions and directions in my research.

I thank Dr. P. S. Reynolds for permitting me to use the rat data in this thesis. I extend my appreciation to the NOAA/GMDD and GAGE/AGAGE global network program, who made their CFC-11 data available to the public. I am grateful to Geoffrey Dutton, NOAA Earth System Research Laboratory, Global Monitoring Division, and Derek Cunnold, a member of the AGAGE team, for clarifications of their data used in this thesis.

I am grateful to the Department of Statistics and Actuarial Science, University of Waterloo, for the computer facilities, and financial support in terms of graduate assistantships. The research assistantships from Grace Chiu and Joel Dubin, and the graduate studentship from Grace Chiu are also highly appreciated. I would also like to thank Government of Ontario for awarding me the Ontario Graduate Scholarship (000113006 and 00012613), and the University of Waterloo for the President's Scholarship. I would like to show my gratitude to Peter X. Song, Professor, Department of Statistics and Actuarial Science, University of Waterloo, for his temporary support through NSERC.

It is an honor for me to be selected one of the International Biometric Society's Eastern North American Region (ENAR) Distinguished Student Paper Awards as a recognition of an article out of this thesis.

I would also like to thank all my friends for their encouragement and support.

I express my gratitude to my entire family. My parents have been a constant source of support – emotional, moral and of course financial – during my studies, and this thesis would certainly not have existed without them. My brothers and sisters were particularly supportive, and always encouraged me throughout my graduate studies.

I express my very special thanks to my wife Rifat Jahan Khan for her encouragement and being exceptionally caring and supportive.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the thesis.

Dedication

This dissertation is dedicated to my wife, Rifat, and son, Aditya.

Contents

List of Tables **xii**

List of Figures **xv**

1 Introduction **1**

2 Changepoint Modeling: A Review **9**

3 The Mixed Bent-Cable Model **14**

3.1 Hierarchical Formulation of the Model 14

3.1.1 The Model for One Individual 15

3.1.2 Assumptions 16

3.1.3 The Hierarchy 17

3.1.4 Summary of the Hierarchy 22

3.2 Rationale Behind the Assumptions and Possible Further Extension of the Model 23

3.2.1 Assumption T 23

3.2.2 Assumption A 24

3.2.3 Assumption B 25

3.2.4 Assumption C 26

3.3 Discussion 27

3.4 Chapter Appendix 28
3.4.1 Level 1 28
3.4.2 Delta Bivariate Mixed Lognormal Distribution 28

4 Bayesian Inference for the Mixed Bent-Cable Model 30

4.1 Bayesian Inference 31
4.1.1 Posterior Density 31
4.2 Bayesian Inference and MCMC Methods 33
4.2.1 MCMC Methods 34
4.3 Construction of a Markov Chain 37
4.3.1 Metropolis-Hastings Algorithm 38
4.3.2 Gibbs Sampler 39
4.3.3 Metropolis within Gibbs Algorithm 40
4.4 Mixing and Convergence 42
4.4.1 Number of Chains 43
4.4.2 Burn-in and Stopping Time 43
4.4.3 Thinning 44
4.4.4 Graphical Diagnostics 44
4.5 Software Implementation 45
4.6 Discussion 46
4.7 Chapter Appendix 48
4.7.1 Joint Density 48
4.7.2 Approximating Fitted Values 49
4.7.3 Full Conditionals 50
4.7.4 Metropolis Step for $\boldsymbol{\alpha_i}$ 63

5 Special Cases of the Mixed Bent-Cable Model **65**

5.1 Model A 66

5.2 Model G 67

5.3 Discussion 69

5.4 Chapter Appendix: Full Conditionals 70

6 Data Analyses **72**

6.1 Rat Data 72

6.1.1 Background of the Study 72

6.1.2 Data 73

6.1.3 Results 74

6.2 CFC-11 Data 84

6.2.1 Background of the Study 84

6.2.2 Data 86

6.2.3 Results 90

6.3 Discussion and Conclusion 100

7 Simulation Study **101**

7.1 Scenarios 101

7.2 Parameters in the Simulations 102

7.3 Data 102

7.4 Results 103

7.4.1 *Scenarios 1* and *2* 103

7.4.2 *Scenario 3* 107

7.5 Discussion 110

8 Concluding Remarks and Future Work **112**

8.1 Cautionary Remarks 113

8.2 Future Work 114

8.3 Publications 116

References **117**

List of Tables

6.1 Posterior summaries for the two populations of rats assuming AR(0) noise: posterior means for the population slope parameters (μ_1 and μ_2) are in "per 15 seconds" and those for the population transitions are in minutes. 79

6.2 Rat data analysis – posterior summaries of the standard deviations and correlations associated with Σ_β, Σ_α and $\sigma^2_{\tau_A}$ (priors for the random regression coefficients); posterior summaries for the standard deviations of γ_i and τ_i are in minutes. 83

6.3 CFC-11 data summary 88

6.4 Posterior summaries of the global concentrations of CFC-11 assuming AR(1) within-individual noise: posterior means for the population slope parameters (μ_1 and μ_2) are in "per 1 month". 91

6.5 Posterior summaries of the global concentrations of CFC-11 assuming AR(2) within-individual noise: posterior means for the population slope parameters (μ_1 and μ_2) are in "per 1 month". 94

6.6 Estimated station-specific concentrations of CFC-11: posterior means for the slope parameters (β_{1i} and β_{2i}) are in "per 1 month". 97

6.7 CFC-11 data analysis – posterior summaries of the standard deviations and correlations associated with Σ_β and Σ_α (priors for the random regression coefficients). 98

7.1 Simulation study with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means of the mixing proportion ω, population regression coefficients and the AR parameters; also coverage of 95% credible intervals. 104

7.2 Simulation study with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means (medians for the variance parameters) of the variances and covariances ($\sigma^2_{\tau_A}$, Σ_β and Σ_α) in the priors for the random regression coefficients; also coverage of 95% credible intervals. 105

7.3 Simulation study with $n_i = 150$ for all i and $m = 20$: average of 500 posterior medians of the innovation variances; also coverage of 95% credible intervals. 106

7.4 Simulation study assuming an AR(1) within-individual noise, and with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means of the population regression coefficients and the AR parameters; also coverage of 95% credible intervals. 107

7.5 Simulation study assuming an AR(1) within-individual noise, and with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means (medians for the variance parameters) of the variances and covariances (Σ_β and Σ_α) in the priors for the random regression coefficients; also coverage of 95% credible intervals. 108

7.6 Simulation study assuming an AR(1) within-individual noise, and with $n_i = 150$ for all i and $m = 20$: average of 500 posterior medians of the innovation variances; also coverage of 95% credible intervals. 109

List of Figures

1.1 Core body temperature (T_c) of 5 representative rats out of 38, recorded during hemorrhage in a study to collect information about the state of hypothermia and resuscitation strategy. 2

1.2 Two monthly mean profiles of CFC-11 in parts-per-trillion (ppt), monitored at different geographical locations. 3

1.3 The bent-cable function. (a) A gradual quadratic transition joining two linear segments (incoming and outgoing). The transition period ranges from $\tau - \gamma$ to $\tau + \gamma$. If the slope changes sign, then it takes place at the critical time point (CTP). (b) An abrupt transition with $\gamma = 0$ yields a broken-stick. The change in slope takes place at the CTP. 5

1.4 Observed data (black lines) and the corresponding individual-specific fitted curves (red lines) for 5 representative rats. Estimated transitions are marked by the vertical lines with estimated CTPs (for Population G) by the dotted lines; The CTP estimate is not marked for Rat 3 because the estimated slope of its cable does not change signs. 7

2.1 A graphical description of the parameters for the changepoint model (2.1). 11

6.1 Rat data analysis – trace plots for the posteriors of parameters μ_γ, μ_τ and μ_{τ_A} from two chains for AR(0), and for AR(1) with $\phi \sim \mathcal{N}(0, 0.00005)$ (Prior 1) and $\phi \sim \mathcal{N}(0, 0.0001)$ (Prior 2); poor mixing is seen for the latter two cases. 75

6.2 Rat data analysis – trace plots for the posteriors of the population parameters $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$, $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and μ_{τ_A}, and the mixture probability ω from two chains assuming AR(0) noise. 76

6.3 Rat data analysis – kernel density estimate plots for the posteriors of population parameters $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$, $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and μ_{τ_A}, and the mixture probability ω from two chains assuming AR(0) noise. 77

6.4 Rat data analysis – autocorrelation plots for the two chains for population parameters $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$, $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and μ_{τ_A}, and the mixture probability ω, assuming AR(0) noise. 78

6.5 Observed data (black lines) and fitted curves (red and green curves) for the two populations of rats assuming within-individual independence. The estimated transition for Population G $\left(\text{i.e. } \widehat{\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}}\right)$ is marked by solid red vertical lines, and that for Population A $\left(\text{i.e. } \widehat{\exp\{\mu_{\tau_A}\}}\right)$ by the green vertical line. The estimated CTP for Population G is indicated by the dotted vertical line, which virtually coincides with the estimated transition point of Population A. 80

6.6 Observed data (black lines) and the corresponding individual-specific fitted curves (red lines) for 5 representative rats. Estimated transitions $\left(\text{i.e. } \hat{\tau} \text{ and } \widehat{\tau \pm \gamma}\right)$ are marked by the vertical lines with estimated CTPs (for Population G) by the dotted lines; The CTP estimate is not marked for Rat 3 because the estimated slope of its cable does not change signs. 82

6.7 CFC-11 profiles of 8 stations (monthly mean data) 89

6.8 CFC data analysis – trace plots for the posteriors of the population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two chains assuming AR(2) noise. 92

6.9 CFC data analysis – kernel density estimate plots for the posteriors of population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two chains assuming AR(2) noise. 92

6.10 CFC data analysis – autocorrelation plots for the two chains for population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ, assuming AR(2) noise. 93

6.11 Observed data (black lines) and the corresponding population (global) fit (red) of the CFC-11 data assuming AR(2) within-individual noise. Estimate of the transition is marked by the solid vertical lines and that of the CTP by the dotted vertical line. 95

6.12 Observed data (black lines) and the corresponding station-specific fitted curves (red curves) of the CFC-11 data assuming AR(2) within-individual noise. Estimate of the transition is marked by the solid vertical lines and that of the CTP by the dotted vertical line. 96

Chapter 1

Introduction

Longitudinal data naturally arise in many areas of study, where measurements taken over time are nested within observational units drawn from some population of interest. (For convenience, in this thesis we will always use the term "individual" or "subject" to refer to the observational unit in the longitudinal study, including inanimate objects such as a geographical location.) In particular, data showing a trend that characterizes a change due to a system shock are commonly observed over time in biological, medical, health and environmental applications. In this thesis, we refer to such data as shock-through data.

An example of shock-through data is an experiment on rats (Reynolds and Chiu [63]; also see Chapter 6) conducted with an objective to collect information about the state of hypothermia and resuscitation strategy immediately after a 60% hemorrhage (the shock) that appears to best promote survival for 3 hours without conventional large-volume crystalloid support. Note that hypothermia — a fatal condition which can occur when core body temperature (T_c) falls below 35^oC — is used as a therapeutic tool for cardiac arrest, stroke and brain injury. Since T_c is typically associated with a critical threshold associated with a breakdown in the compensatory homeostatic mechanisms following severe hemorrhage (Connett et al. [17]), quantification of the transition of T_c to early hypothermia is of clinical interest.

Figure 1.1 shows 5 representative T_c profiles out of a study sample size of 38 rats. In addition to roughly linear incoming and outgoing phases at either end of each profile, we see that some rats may exhibit a gradual transition in T_c, while others, perhaps an abrupt transition. That is, we have samples potentially coming from two different populations,

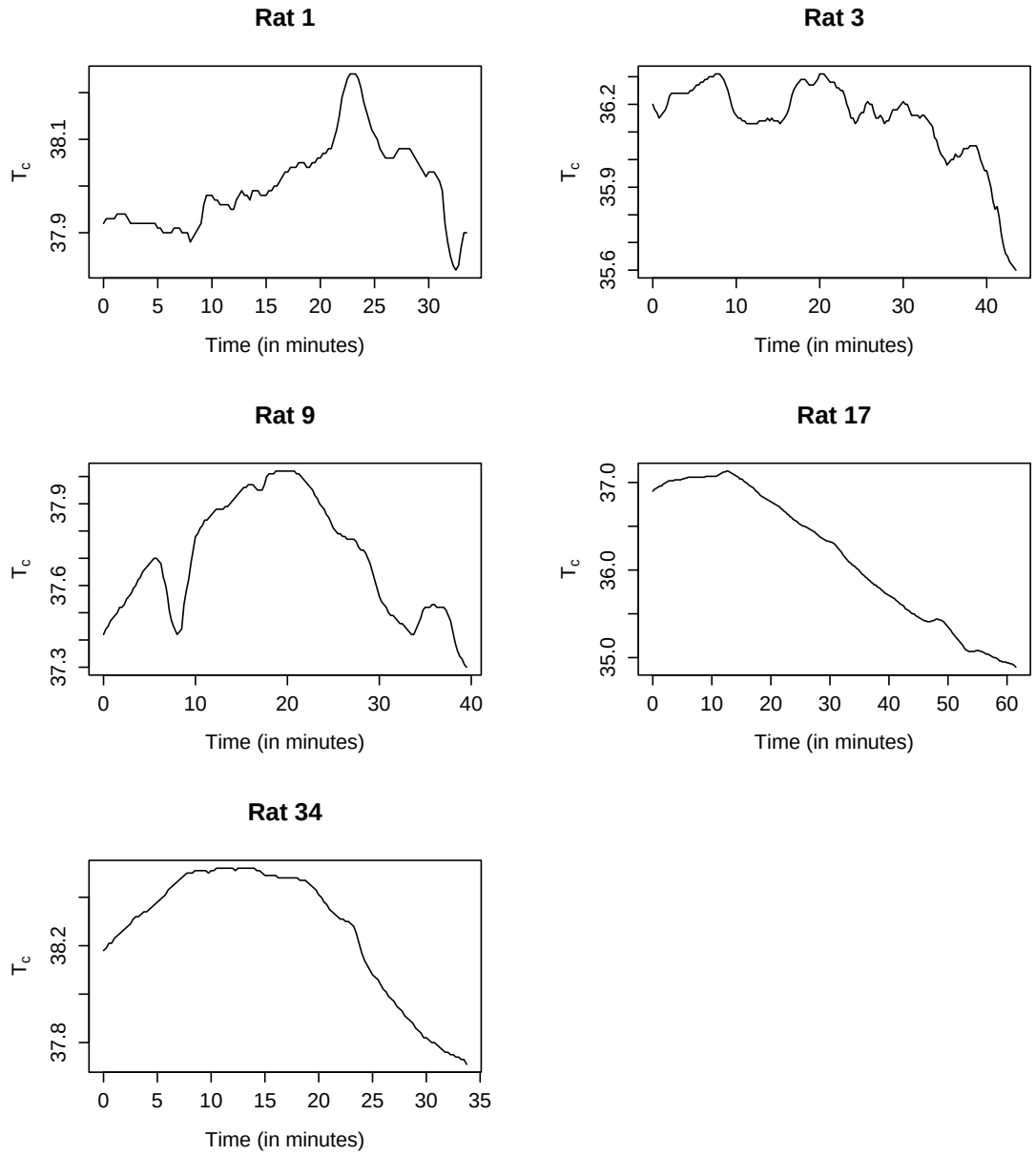

Figure 1.1: Core body temperature (T_c) of 5 representative rats out of 38, recorded during hemorrhage in a study to collect information about the state of hypothermia and resuscitation strategy.

say, G (gradual) and A (abrupt), respectively, according to the type of transition for the underlying T_c trend. Accounting for this possibility, we develop statistical methodologies

in this thesis for modeling such data to address questions of broad interest, for example,

q1. How long did it take for the trend to show an obvious change because of the shock?

q2. What were the rates of increase/decrease before and after the change?

q3. What was the time point at which the trend went from increasing to decreasing, or vice versa?

Our modeling approach, which is a substantial generalization of a special changepoint model, the bent cable (Chiu et al. [16]), will provide a flexible methodology, where inference for the type of transition is data driven, rather than pre-assumed as a specific type.

Data exhibiting only one type of transition are also common in many areas. An example is the atmospheric concentration of chloroflurocarbons (CFCs) (The Columbia Encyclopedia [72]) in response to the Montréal Protocol's (The Columbia Encyclopedia [73]) ban (the shock) on CFC products. Figure 1.2 shows a rather gradual transition in the CFC-11 concentrations monitored from 2 of many stations around the globe. One appealing feature of our new approach in this thesis is that it encompasses the case of a single transition type as illustrated in Figure 1.2.

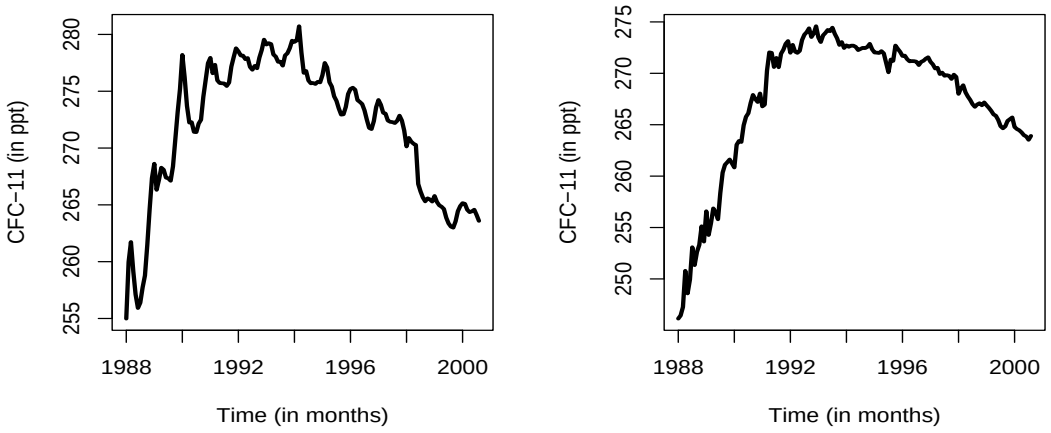

Figure 1.2: Two monthly mean profiles of CFC-11 in parts-per-trillion (ppt), monitored at different geographical locations.

The so-called broken-stick is a popular piece-wise linear model, and is a natural candidate to describe a continuous trend with an abrupt change (Hall et al. [39], Lang et al. [47], Kiuchi et al. [45], Slate and Turnbull [68]). Broken-stick models, however, may not be realistic in many applications where data may exhibit gradual changes, e.g., CFC-11 data (Figure 1.2), or are believed to exhibit gradual changes, e.g., a smooth change agrees with the clinical belief of a progressive decline of cognitive test scores in the pre-diagnosis phase of dementia (Jacqmin-Gadda et al. [41]). So, it is desirable to relax the assumption of abruptness beforehand, and to formulate a model that is flexible enough to handle either type of changes, gradual or abrupt. One such methodology is the bent-cable regression, as we now describe.

Chiu et al. [16] and Chiu and Lockhart [15] developed the bent-cable regression methodology to analyze shock-through data for a single profile showing roughly three phases: incoming and outgoing, both of which are linear, joined by a quadratic bend (Figure 1.3(a)). The model is parsimonious, and appealing due to its simple structure, great flexibility and interpretability. Because an extremely sharp bend reduces the bent-cable to a broken-stick (Figure 1.3), it also encompasses the broken-stick model as a special case.

There are other modeling approaches that handle gradual changes, such as penalized spline regression (e.g., Ruppert et al. [65]), but the added flexibility in terms of the potential shape of the fitted model can come at a cost of interpretability. For the data we are considering, we aim to provide a flexible modeling approach that additionally allows for interpretability of fitted parameters, which is important in many practical contexts. Motivated by this, our main methodological contribution presented in this thesis is to account for either type of transition through a mixture model extension of the bent-cable regression technique, as well as to incorporate a mixed-effects structure to accommodate the repeated measurements observed for each individual (see below). Specifically, our main methodological contribution is an extension of the existing single-profile bent-cable regression in two directions:

Ext 1. extension of the bent-cable regression for longitudinal data (i.e., multiple profiles); and

Ext 2. accounting for either type of transition — gradual or abrupt — in addition to **Ext 1**.

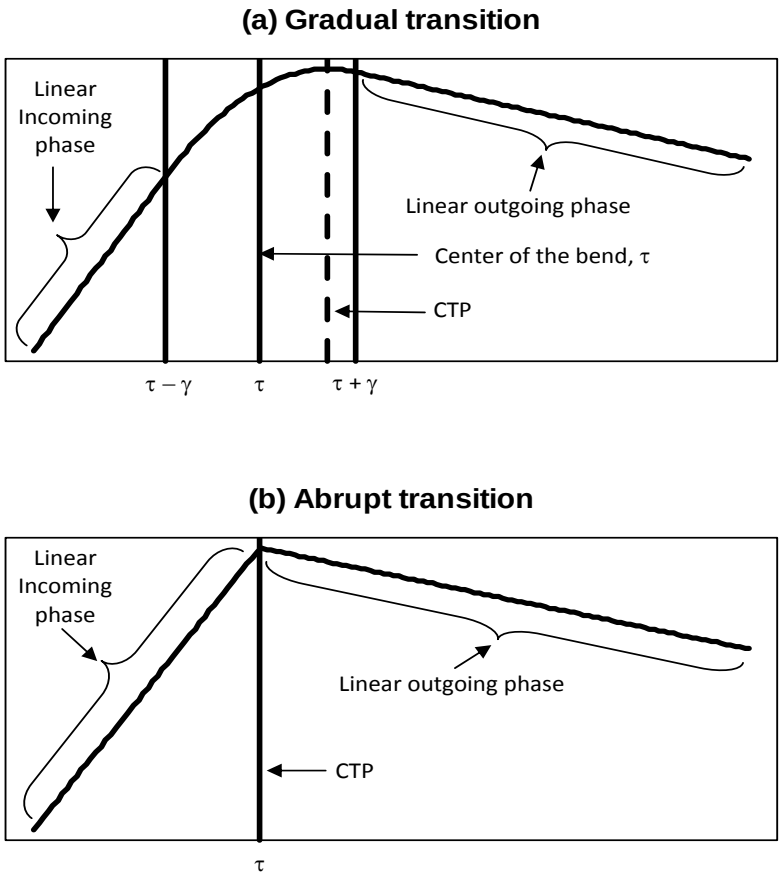

Figure 1.3: The bent-cable function. (a) A gradual quadratic transition joining two linear segments (incoming and outgoing). The transition period ranges from $\tau - \gamma$ to $\tau + \gamma$. If the slope changes sign, then it takes place at the critical time point (CTP). (b) An abrupt transition with $\gamma = 0$ yields a broken-stick. The change in slope takes place at the CTP.

From a statistical point of view, in analyzing longitudinal data, we would like to answer the questions (Singer and Willett [67]): (1) How does the response change over time? and (2) Do different individuals experience different patterns of change? The first question characterizes each individual's pattern of change over time (commonly called within-individual variation), and the second question addresses the association between patterns of change (commonly called between-individual variation). Mixed-effects models (Laird and Ware [46]), which unify information from each individual to answer the above two questions, are well suited for the analysis of longitudinal data and provide useful information regarding

the above questions.

Therefore, we accomplish **Ext 1** through a mixed-effects model extension of bent-cable regression for time series by introducing multi-level random effects to combine information available from the data from each individual. **Ext 2** is accomplished through a mixture model extension of **Ext 1** by allowing a mixture of distributions to model the abruptness parameter of bent-cable regression. We focus our work in the scenario of an autoregressive (AR) process of order p (Box et al. [8]), $p \geq 0$, for the within-individual noise. As a special case, we also describe the modeling approach for the existence of a single transition type. Henceforth, we will refer to our extended bent-cable methodology simply as *mixed bent-cable regression*.

As an illustration, Figure 1.4 shows the fitted curves/sticks of the 5 rat profiles of Figure 1.1 by an application of the mixed bent-cable regression (see Chapter 6 for details). It demonstrates that our methodology picks up the two types of transitions adequately. The estimated transitions for Rats 1, 13 and 34 reflect a slow (gradual) change in T_c and a linear decrease thereafter, while those for Rats 9 and 17 show an abrupt transition followed by a linear decrease. That is, we estimate that the former 3 are from Population G whereas the latter 2 are from Population A.

Although modeling the trend of changepoint data is challenging in the presence of discontinuous derivatives, it is common practice (e.g. Dominicus et al. [20], Lang et al. [47], Morrell et al. [52]) to overcome the non-differentiability problem at the modeling and/or estimation and inference stage(s). We precede in Chapter 2 with a review of some of these methods including the mathematical formulation of the bent-cable regression of Chiu et al. [16] and Chiu and Lockhart [15]. There we preview the difference between bent-cable regression and other works.

We introduce our mixed bent-cable model in Chapter 3. Specifically, we present the hierarchical formulation of our unifying modeling framework with its underlying assumptions, and introduce additional assumptions which are adjustable to alter the complexity of the model.

The existence of two variance components (within-individual and between-individual) in mixed-effects models complicate inference on the (fixed) population effect of interest. Chiu et al. [16] and Chiu and Lockhart [15] describe the complexity of the frequentist estimation method and asymptotics of bent-cable regression for one profile only. Since the second

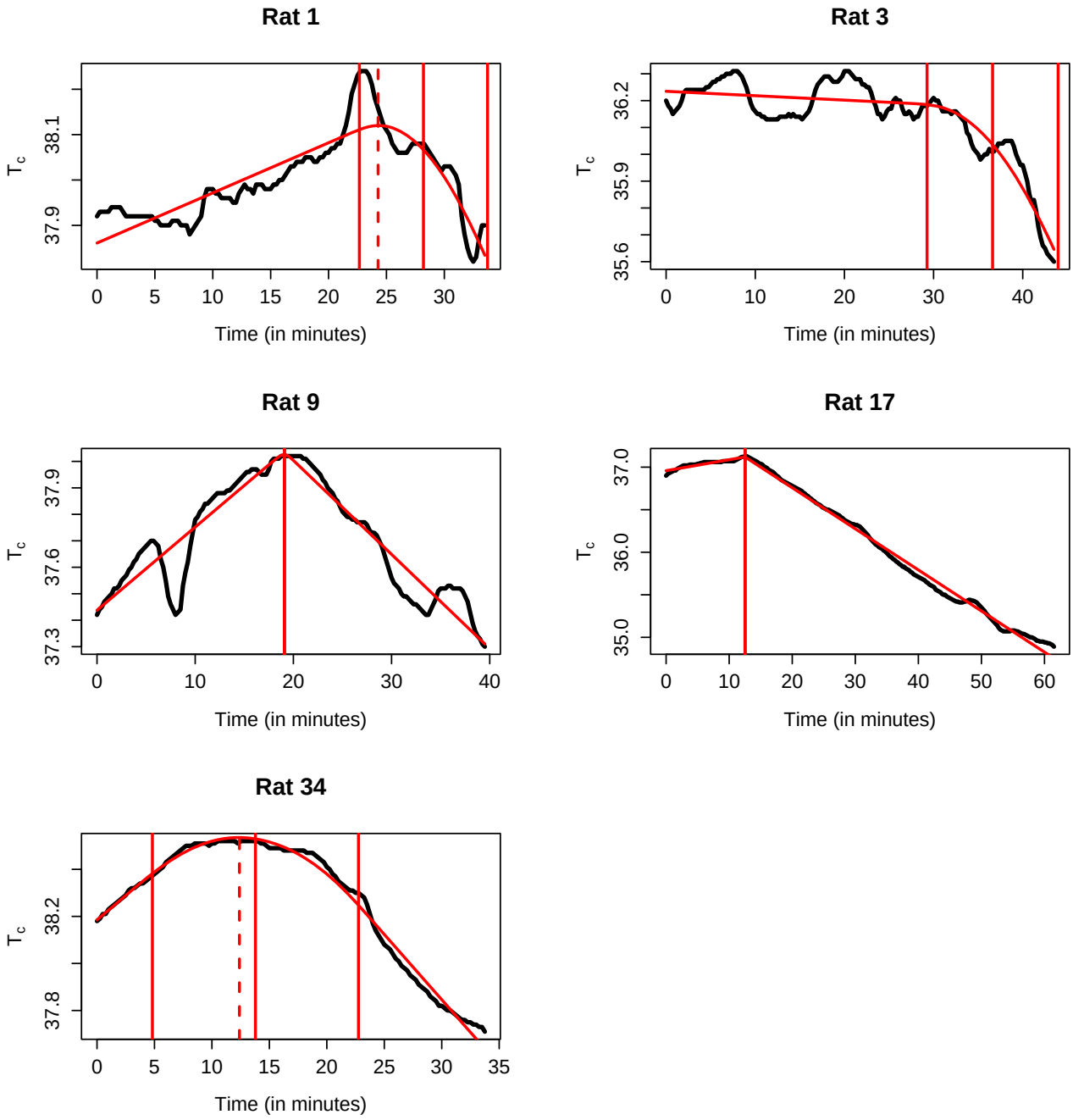

Figure 1.4: Observed data (black lines) and the corresponding individual-specific fitted curves (red lines) for 5 representative rats. Estimated transitions are marked by the vertical lines with estimated CTPs (for Population G) by the dotted lines; The CTP estimate is not marked for Rat 3 because the estimated slope of its cable does not change signs.

derivative of the likelihood function does not exist everywhere, the highly complicated asymptotics are developed under non-standard regularity conditions. In this thesis, we employ the Bayesian method of inference (Chapter 4) for our mixed bent-cable regression, which avoids the substantial additional complexity of asymptotics due to mixed-effects and mixture modeling. All the derivations and computational tools associated with the Bayesian implementation are presented in that chapter. In Chapter 5, we describe a special case of our mixed bent-cable methodology, namely, modeling one transition type only.

The next few chapters are very technical, and the readers interested in the practical aspects of our methodology may immediately go to Chapter 6, where our proposed method is illustrated with applications to the aforementioned rat and CFC-11 data. In Chapter 7, we demonstrate the performance of our methodology under various scenarios through a simulation study. We conclude the thesis in Chapter 8 by presenting some additional considerations relevant to this work and possible extensions.

Chapter 2

Changepoint Modeling: A Review

The choice of the model framework for longitudinal data often depends on the nature of the data. The most simple framework is the linear mixed-effects model (Laird and Ware [46], Verbeke and Molenberghs [80]). However, longitudinal data trajectories are frequently nonlinear in nature. In some settings, an appropriate nonlinear function may be derived on the basis of the theoretical considerations, and in other settings, a nonlinear relationship may be employed to provide an empirical description of the data (Davidian and Giltinan [19]). In either case, the use of a suitable nonlinear function is very important to appropriately describe the data.

One of the most intriguing nonlinear models seen in the environmental and biological applications is the piecewise model, also known as the segmented regression model or changepoint model, to analyze shock-through data as described in Chapter 1. A changepoint model is continuous if the segments join at the changepoint, and discontinuous if at the changepoint a sudden jump or drop occurs in the mean response (Piegrosh and Biler [57]). Moreover, the segmentation can be smooth (gradual transition) or abrupt (e.g., broken-stick) in a continuous changepoint model. In this thesis, we do not consider discontinuous changepoint model, because shock-through data as shown in Figures 1.1 and 1.2 for the rat and CFC-11 data, respectively, usually do not exhibit a sudden jump at the changepoint. This type of data can also arise very naturally in other applications where the changes can occur at known times (e.g., taxation or policy changes) or unknown times (e.g., cognition and behavior changes due to transient ischemic attack (mini-stroke)). Henceforth, we will only refer to a continuous changepoint model and simply call it a changepoint model.

In the presence of discontinuous derivatives, modeling the trend of changepoint data, even for a single profile is challenging, especially at the inference stage (Chiu et al. [16], Chiu and Lockhart [15], Kelly et al. [43], Tishler and Zang [75]). The common approach to overcome this difficulty is to reformulate the model so that the derivative with respect to each parameter is a continuous function of time (t) and/or to utilize an inference procedure that does not depend on the derivatives, e.g. Bayesian inference. We now present some of these modeling techniques common in the literature.

Broken-stick models are extensively used in many applications of changepoint data, e.g. Lang et al. [47] and Kiuchi et al. [45] to analyze longitudinal series of the number of T4 cells which is considered a marker of disease progression for persons infected with human immunodeficiency virus (HIV); Bellera et al. [5], and Slate and Turnbull [68] to analyze data sets concerning prostate specific antigen (PSA) as a serial marker for prostate cancer; and Dominicus et al. [21] to analyze cognitive functions at older ages. A general representation of such models for one profile is

$$y_j = \beta_0 + \beta_1(t_j - \tau) + \beta_2(t_j - \tau)\text{sgn}(t_j - \tau) + \epsilon_j \tag{2.1}$$

where τ is the unknown join point of the two sticks; β_0, β_1 and β_2 are the linear regression coefficients with $\beta_1 - \beta_2$ and $\beta_1 + \beta_2$ being the slopes of the incoming and outgoing phases, respectively; β_1 is the average of the two slopes; β_2 is the half of the difference of the two slopes; β_0 is the expected value of the response at τ; ϵ_j is a random variable accounting for error; and sgn(.) is a sign function where

$$\text{sgn}(t_j - \tau) = \begin{cases} -1, \text{ if } t_j < \tau \\ 0, \text{ if } t_j = \tau \\ +1, \text{ if } t_j > \tau \end{cases} . \tag{2.2}$$

A graphical description of the parameters is presented in Figure 2.1. Note that in all of the above examples, the inference is carried out by the Bayesian approach.

Although broken-stick models are used in many applications of longitudinal changepoint data, modeling gradual transition has not been studied extensively so far. Bacon and Watts [4] propose to use a smooth function to appropriately model the transition from one regime to another by replacing the sign function (2.2) of (2.1) by a transition function $\text{trn}(d_j/\gamma)$, where $d_j = t_j - \tau$, and γ is a parameter whose value determines the type of transition (abrupt or gradual). To make this function behave like a sign function, they propose the

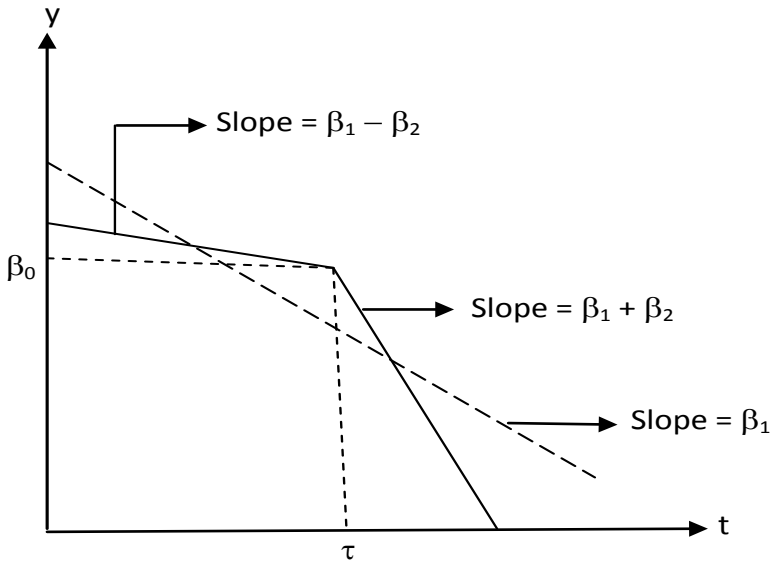

Figure 2.1: A graphical description of the parameters for the changepoint model (2.1).

following conditions to be satisfied by $\mathrm{trn}(d_j/\gamma)$.

1. $\lim_{d_j \to \infty} \mathrm{trn}(|d_j|/\gamma) = 1$ so that $\mathrm{trn}(d_j/\gamma)$ behaves like $\mathrm{sgn}(d_j)$ for large d_j;
2. $\mathrm{trn}(0) = 0$ so that $\mathrm{trn}(d_j/\gamma) = \mathrm{sgn}(d_j)$ for $d_j = 0$;
3. $\lim_{\gamma \to 0} \mathrm{trn}(d_j/\gamma) = \mathrm{sgn}(d_j)$ so that $\mathrm{trn}(d_j/\gamma)$ behaves like $\mathrm{sgn}(d_j)$ for small γ; and
4. $\lim_{d_j \to \infty} d_j \ \mathrm{trn}(d_j/\gamma) = d_j$ so that $d_j \ \mathrm{trn}(d_j/\gamma)$ behaves like $d_j \ \mathrm{sgn}(d_j) = |d_j|$ for large d_j; this condition is necessary because of the way the transition function enters into the model (2.1).

Examples of two such functions are the cumulative distribution function of any symmetric probability density function and the hyperbolic tangent. They originally propose this to analyze a single profile by an implementation of a Bayesian analysis procedure in which the variance parameter $var(\epsilon_j) = \sigma^2$ and the linear parameters β_0, β_1 and β_2 are integrated out so that attention may be focused on the transition parameters. Bacon and Watts's development is later applied by others in piece-wise nonlinear mixed models (see below).

Morrell et al. [52] propose a changepoint model with a linear incoming and an exponential outgoing phase to analyze longitudinal PSA data. To overcome the nondifferentiability

problem, they use a transition function as per Bacon and Watts to provide a smooth transition between the phases. This enables them to use frequentist's estimation and inference based on some approximations as detailed by Lindstrom and Bates [49]. Dominicus et al. [20] consider a hyperbolic tangent function to smooth out the changepoint model, and they demonstrate a Bayesian approach for inference.

Jacqmin-Gadda et al. [41] propose a slightly different approach in which they consider a linear trend before the changepoint and a polynomial trend thereafter, resulting in a smooth transition at the changepoint. The proposed model has the property that the derivative with respect to each parameter is a continuous function of t. They consider a likelihood-based approach for statistical inference. However, choosing the order of the polynomial is judgemental, and the model is not parsimonious in case of higher order polynomials.

There are also case-based methods that can handle changepoint data with complex trajectories, such as penalized splines (e.g., Ruppert et al. [65]) and local polynomial (Fan and Gijbels [23]), though the added flexibility can come at a cost of interpretability.

As introduced briefly in Chapter 1, bent-cable regression is first proposed by Chiu et al. [16] to model changepoint data for a single profile. The model is parsimonious and flexible in that under a single mathematical formulation, it combines the gradual and abrupt transitions using only 5 regression coefficients. For the types of data under consideration, in light of the apparent three phases — linear incoming and outgoing, and the adjoining curved transition — an individual profile is characterized by the bent-cable regression model as follows:

$$y_j = f(t_j, \beta_0, \beta_1, \beta_2, \gamma, \tau) + \epsilon_j \tag{2.3}$$

where
$$f(t_j, \beta_0, \beta_1, \beta_2, \gamma, \tau) = \beta_0 + \beta_1 t_j + \beta_2\ q(t_j, \gamma, \tau), \tag{2.4}$$

$$q(t_j, \gamma, \tau) = \frac{(t_j - \tau + \gamma)^2}{4\gamma}\mathbf{1}\{|t_j - \tau| \leq \gamma\} + (t_j - \tau)\mathbf{1}\{t_j - \tau > \gamma\}, \tag{2.5}$$

with β_0, β_1 and β_2 characterizing the linear incoming and outgoing phases, and γ and τ characterizing the transition. Here, β_0 and β_1 are, respectively, the intercept and slope of the incoming phase; β_2 is a scaling factor associated with the location of the transition;

$\beta_1 + \beta_2$ is the slope of the outgoing phase; and γ and τ represent the center and half-width of the bend, respectively. Note that $\gamma = 0$ reduces the bent cable to a broken-stick model for which

$$q(t_j, 0, \tau) = (t_j - \tau)\mathbf{1}\{t_j - \tau > 0\}. \tag{2.6}$$

Chiu and Lockhart [15] define the critical time point (CTP) as the time at which the slope of the bent-cable changes sign. Thus, for a gradual transition ($\gamma > 0$), the CTP is $\tau - \gamma - 2\beta_1\gamma/\beta_2$ at which $f(t_j, \beta_0, \beta_1, \beta_2, \gamma, \tau)$ has slope zero. Note that this formula does not represent anything meaningful when the slope of the cable does not change signs. When $\gamma = 0$, any sign change of the slope occurs at the point τ, and therefore τ is the CTP for an abrupt transition.

For time series data, Chiu and Lockhart [15] develop estimation theory under quite general conditions of the correlation structure of the error terms ϵ_j's. Under some design conditions, they show that the regression model (2.3)-(2.5) is regular in the sense that the usual results of consistency and asymptotic normality are valid for the least squares estimator, as well as asymptotic χ^2 distribution for the deviance statistic. The design conditions also guarantee the identifiability of the regression problem. An R (R Development Core Team [61]) library "bentcableAR" is developed by Chiu [14] to fit the bent-cable regression under autoregressive noise of order $p \geq 0$.

The three appealing features — flexibility, greatly interpretable regression coefficients and characterization of the transition period through the parameters γ, τ and CTP — distinguish the bent-cable regression model from those other works described earlier. Therefore, the extension of the bent-cable regression for longitudinal data (**Ext 1** and **Ext 2**, Chapter 1) would perceivably provide a powerful longitudinal modeling approach for changepoint data. We proceed in Chapter 3 with the modeling framework of our new approach, and with a detailed description for inference in Chapter 4.

Chapter 3

The Mixed Bent-Cable Model

In this chapter, extension of the bent-cable regression model for longitudinal data, namely, the mixed bent-cable model is introduced. In Section 3.1, we develop a hierarchical formulation of the model on the basis of the assumptions presented in Section 3.1.2. We describe the rationale behind our assumptions and possible extensions of the model by relaxing some of those assumptions in Section 3.2. A discussion of our modeling approach is presented in Section 3.3, and the chapter ends in Section 3.4 with some mathematical details underlying the development of the hierarchy.

3.1 Hierarchical Formulation of the Model

As briefly mentioned in Chapter 1, the hierarchical formulation of a mixed-effects model generally involves two levels (Davidian and Giltinan [19]). The first level, called Level 1 or within-individual level, characterizes an individual trajectory, and the second level, called Level 2 or between-individual level, specifies whether different individuals manifest different patterns and what may influence these differences. Additionally, there is a third level (Level 3) for Bayesian inference, which quantifies prior information about the random quantities in Levels 1 and 2. In this section, we present a description of these three levels for our mixed bent-cable model. In Chapter 2, we presented the bent-cable regression model to analyze data from a single individual. For completeness of this chapter, and an introduction of our notation for longitudinal data, we re-introduce in Section 3.1.1 the bent-cable model for one individual but in the context of having multiple individuals. Then,

we present assumptions of our hierarchical mixed bent-cable model in Section 3.1.2, and develop the three levels mathematically in Section 3.1.3. A summary of the hierarchical formulation is presented in Section 3.1.4.

3.1.1 The Model for One Individual

When individuals can be regarded as having been randomly selected from some population, and repeated measurements are collected for each individual, it is useful to unify information from each individual to aid the understanding of the population as well as subject-specific behavior. Suppose we have m individuals. For the i^{th} individual ($i = 1, 2, \ldots, m$), let there be n_i measurements, and let t_{ij} denote the j^{th} measurement occasion, $j = 1, 2, \ldots, n_i$, with $\mathbf{y}_i = (y_{i1}, y_{i2}, \ldots, y_{in_i})$ and $\mathbf{t}_i = (t_{i1}, t_{i2}, \ldots, t_{in_i})$. We model the corresponding response at time t_{ij}, denoted by y_{ij}, by the relationship

$$y_{ij} = f(t_{ij}, \boldsymbol{\theta}_i) + \epsilon_{ij} \tag{3.1}$$

where $\boldsymbol{\theta}_i$ is a vector of regression coefficients for the ith individual, $f(\cdot)$ is a flexible function of t_{ij} and $\boldsymbol{\theta}_i$ to characterize the trend of the subject-specific data over time, and ϵ_{ij} represents the random error component, which accounts for measurement error and possibly additional within-individual error structure.

In light of the apparent three phases as described in Chapters 1 and 2 — linear incoming and outgoing, and the adjoining curved transition — we characterize the individual profiles by the bent-cable function (Chiu et al. [16]), given by

$$f(t_{ij}, \boldsymbol{\theta}_i) = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} q(t_{ij}, \boldsymbol{\alpha}_i), \tag{3.2}$$

$$\text{where} \qquad q(t_{ij}, \boldsymbol{\alpha}_i) = \frac{(t_{ij} - \tau_i + \gamma_i)^2}{4\gamma_i} \mathbf{1}\{|t_{ij} - \tau_i| \leq \gamma_i\} + (t_{ij} - \tau_i)\mathbf{1}\{t_{ij} - \tau_i > \gamma_i\} \tag{3.3}$$

with $\boldsymbol{\beta}_i = (\beta_{0i}, \beta_{1i}, \beta_{2i})'$ and $\boldsymbol{\alpha}_i = (\gamma_i, \tau_i)'$ being the vectors of linear and transition coefficients, respectively, and $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i', \boldsymbol{\alpha}_i')'$. Henceforth, we will denote $f(t_{ij}, \boldsymbol{\theta}_i)$ and $q(t_{ij}, \boldsymbol{\alpha}_i)$ simply by f_{ij} and q_{ij}, respectively. Recall that $\gamma_i = 0$ reduces the bent cable to a broken-stick model for which

$$q_{ij} = (t_{ij} - \tau_i)\mathbf{1}\{t_{ij} - \tau_i > 0\}. \tag{3.4}$$

Moreover, the CTP (see Chapter 2) for the i^{th} individual is given by $(\tau_i - \gamma_i - 2\beta_{1i}\gamma_i/\beta_{2i})$ if $\gamma_i > 0$ (gradual transition), and τ_i if $\gamma_i = 0$ (abrupt transition).

Having specified a model to characterize an individual profile, we now focus on the assumptions upon which our hierarchical model is based.

3.1.2 Assumptions

If all individuals are assumed to exhibit the same type of transition – be it abrupt or gradual – then shrinkage towards the population may force an observed profile resembling a broken stick to take on a bent-cable fit, and vice versa. Therefore, we have the following:

T. Assumptions Regarding the Transition:

- **T1.** Each individual i potentially comes from one of two populations: Population A for which $\gamma_i = 0$ and Population G for which $\gamma_i > 0$; and
- **T2.** each individual has probability ω to have come from Population G (and, hence, probability $1 - \omega$ to come from Population A).

Besides **T**, additional assumptions are made as follows, but they can be adjusted to alter the complexity of the model:

A. Level 1 Assumptions:

- **A1.** Each profile has a unique amount of scatter around the stick/cable;
- **A2.** ϵ_{ij}'s follow a stationary AR(p) structure (Box et al. [8]) with a common p; and
- **A3.** $\boldsymbol{\theta}_i$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{in_i})'$ are independent.

B. Level 2 Assumptions:

- **B1.** The vector of responses $\mathbf{y}_i$ for the m individuals are independent of each other;
- **B2.** $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ are independently distributed random variables;

B3. between-individual variation is due entirely to unexplained phenomena;

B4. Populations A and G share common intercept and slopes, that is, $\boldsymbol{\beta}_i$ has the same distribution between the two populations; and

B5. Populations A and G do not share a common parameter for the center of the bend, that is, the distribution of τ_i varies between the populations.

C. Assumption for All 3 Levels:

C1. The relevant quantities have certain probability distributions (see Sections 3.1.3 and 3.2.4).

The hierarchical formulation of our mixed bent-cable model is based on the above assumptions, and is described in the following sub-section.

3.1.3 The Hierarchy

Specification of a model for ϵ_{ij} of (3.1) completes the description of the within-individual variation, whereas the between-individual variation is taken into account by the individual parameter vectors $\boldsymbol{\theta}_i$'s which are, by **T1**, assumed to arise from either Population A or G. Now we present below our modeling approach for ϵ_{ij} and $\boldsymbol{\theta}_i$ and the prior specifications in Level 3.

Level 1: within-individual variation

Within-individual level of the hierarchy takes into account two components: correlation and variation among the repeated measurements over time. Under Assumptions **A1**, **A3** and **C1**, the most simple model for the within-individual variation is

$$[\epsilon_{ij}|\sigma_i^2] \sim \mathcal{N}(0, \sigma_i^2) \text{ for all } j = 1, 2, \ldots, n_i, \tag{3.5}$$

which is considered a conditionally independent specification of the within-individual variation. Under this model, the correlation of the repeated measures from the same individual is induced by the individual-specific random coefficients $\boldsymbol{\theta}_i$'s. Though, for some data, $\boldsymbol{\theta}_i$'s may adequately account for this correlation, quite often there is additional serial correlation

remaining that can be accounted for by the ϵ_{ij}'s. Therefore, specific correlation structures (e.g. time series models) are often used for a description of the within-individual variation (Fitzmaurice et al. [24], Pinheiro and Bates [58]). One such model is the autoregressive process of order $p > 0$ (Box et al. [8]) which is well-suited in many types of longitudinal data where measurements are made at equal (or approximately equal) intervals of time (Fitzmaurice et al. [24]). Therefore, we consider here an AR(p) process for describing the within-individual variation.

Thus, by **A2**, we have at Level 1

$$\epsilon_{ij} = \phi_1\epsilon_{i,j-1} + \phi_2\epsilon_{i,j-2} + \ldots + \phi_p\epsilon_{i,j-p} + v_{ij} \tag{3.6}$$

where $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_p)'$ is the vector of AR(p) parameters. For **A1**, **A3** and **C1**, we assume that the innovations v_{ij}'s in (3.6) are independent and identically normal with mean 0 and variance σ_i^2, that is

$$[v_{ij}|\sigma_i^2] \sim \mathcal{N}(0, \sigma_i^2) \text{ for all } j = 1, 2, \ldots, n_i. \tag{3.7}$$

Furthermore, we consider a conditional likelihood framework, where the initial p observations for each i, $\mathbf{y}_i^{(1)} = (y_{i1}, y_{i2}, \ldots, y_{ip})'$, are treated as known, whereas $\mathbf{y}_i^{(2)} = (y_{i,p+1}, y_{i,p+2}, \ldots, y_{i,n_i})'$ are random. This assumption is common, and was considered by Chiu and Lockhart [15] for frequentist bent-cable regression of a single profile, and by Chib [11] in a Bayesian approach for linear regression.

Now, it can be verified from (3.1), (3.6) and (3.7) that the first level of the hierarchy (see appendix (Section 3.4.1)) is

$$[\mathbf{y}_i^{(2)}|\ \mathbf{y}_i^{(1)}, \boldsymbol{\theta}_i, \boldsymbol{\phi}, \sigma_i^2] \sim \mathcal{N}_{n_i-p}(\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i), \tag{3.8}$$

where $\boldsymbol{\mu}_i(\boldsymbol{\theta}_i, \boldsymbol{\phi}) \equiv \boldsymbol{\mu}_i = (\mu_{i,p+1}, \mu_{i,p+2}, \ldots, \mu_{i,n_i})'$, $\mu_{ij} = \beta_{0i} + \beta_{1i}x_{ij} + \beta_{2i}r_{ij} + \sum_{k=1}^{p}\phi_k y_{i,j-k}$, $x_{ij} = t_{ij} - \sum_{k=1}^{p}\phi_k t_{i,j-k}$, $r_{ij} = q_{ij} - \sum_{k=1}^{p}\phi_k q_{i,j-k}$, $\mathbb{I}_i$ is an identity matrix of order $n_i - p$, and $\mathcal{N}_{n_i-p}$ denotes a $(n_i - p)$-variate normal distribution.

Level 2: between-individual variation

The between-individual variation is accounted for by specifying models for the individual-specific regression coefficients $\boldsymbol{\theta}_i$'s. This, in turn, requires distributional assumptions for these parameters (Assumption **C1**). Note that $\boldsymbol{\theta}_i$ has two components: $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$, and

by **B2**, they are independent. Therefore, we model these two components separately. Moreover, we assume that the differences in the $\boldsymbol{\theta}_i$'s are due entirely to unexplained phenomena (Assumption **B3**). In Section 3.3, we describe a possible direction of relaxing this assumption.

Since (i) the dependence structure can be fully specified by the covariance matrix through assuming a multivariate normal distribution, and (ii) multivariate normal distributions are convenient theoretically and computationally, we assume

$$[\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta] \sim \mathcal{N}_3(\boldsymbol{\mu}_\beta, \Sigma_\beta) \tag{3.9}$$

for both populations A and G (Assumption **B4**), where $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$ and Σ_β are, respectively, the mean vector and covariance matrix of $\boldsymbol{\beta}_i$, and $\mathcal{N}_3$ denotes a trivariate normal distribution. Note that $\boldsymbol{\mu}_\beta$ summarizes the linear coefficients of the populations and Σ_β quantifies the association among the random coefficients β_{0i}, β_{1i} and β_{2i}.

We now consider modeling $\boldsymbol{\alpha}_i = (\gamma_i, \tau_i)'$. First note that since $\boldsymbol{\alpha}_i$ is positive, it is reasonable to assume a lognormal ($\mathcal{LN}$) distribution for **C1**. To accommodate **T1** and **T2**, we model γ_i by a delta-lognormal distribution (Aitchison and Brown [1]), which assumes a non-zero probability, $1-\omega$, that $\gamma_i = 0$, and a probability ω that $\gamma_i > 0$; and a lognormal distribution for γ_i given $\gamma_i > 0$, that is,

$$g(\gamma_i) = \mathbf{1}(\gamma_i = 0)\ (1-\omega) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}(\gamma_i) \tag{3.10}$$

where $\mathcal{LN}(\gamma_i)$ is the probability density function of a lognormal distribution. To accommodate **B5**, we assume a $\mathcal{LN}$ distribution for τ_i given γ_i with different sets of parameters for the two populations. Then, denoting a Bernoulli distribution by $\mathcal{BER}$, it can be shown that (see appendix (Section 3.4.2)) the probability density function of $\boldsymbol{\alpha}_i$ is

$$\left.\begin{array}{c} g(\boldsymbol{\alpha}_i|I_i) = (1-I_i)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + I_i\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\ [I_i|\omega] \sim \mathcal{BER}(\omega) \end{array}\right\}, \tag{3.11}$$

where $\mathcal{LN}_2$ stands for bivariate lognormal distribution; μ_{τ_A} and $\sigma^2_{\tau_A}$ are, respectively, the mean and variance of $\log(\tau_i)$ for Population A (for which $I_i = 0$); and $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and Σ_α are, respectively, the mean vector and the covariance matrix of $\log(\boldsymbol{\alpha}_i)$ for Population G (for which $I_i = 1$). Note that I_i is a latent allocation variable ($I_i = 0$ if $\gamma_i = 0$ and $I_i = 1$ if $\gamma_i > 0$), and marginalizing over it in (3.11) gives

$$g(\boldsymbol{\alpha}_i) = \mathbf{1}(\gamma_i = 0)\ (1-\omega)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \tag{3.12}$$

which is the delta bivariate mixed lognormal distribution (Shimizu [66]) for $\boldsymbol{\alpha}_i$. Writing the delta bivariate mixed lognormal distribution hierarchically as in (3.11) facilitates implementing the Bayesian approach (Chapter 4).

Level 3: prior specifications

We employ a Bayesian approach for statistical inference (Chapter 4). The main idea of Bayesian inference is to combine data and prior knowledge on a parameter to determine its *posterior* distribution (the conditional density of the parameter given the data). The prior knowledge is supplied in the form of a prior distribution of the parameter, which quantifies information (or uncertainty) about the parameter prior to any data being gathered. For example, recall **B5** which states that $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ are independent of each other for all i.

Now, our choice of distributions for **C1** leads to

$$\left.\begin{array}{c} [\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1] \sim \mathcal{N}_3(\mathbf{h}_1, \mathbb{H}_1),\ \ [\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2] \sim \mathcal{N}_2(\mathbf{h}_2, \mathbb{H}_2), \\ [\mu_{\tau_A}|a_0, a_1] \sim \mathcal{N}(a_0, a_1),\ \ \left[\Sigma_\beta^{-1}|\nu_1, \mathbb{A}_1\right] \sim \mathcal{W}\big(\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big), \\ [\Sigma_\alpha^{-1}|\nu_2, \mathbb{A}_2] \sim \mathcal{W}\big(\nu_2, (\nu_2\mathbb{A}_2)^{-1}\big),\ \ \left[\sigma_{\tau_A}^{-2}|b_0, b_1\right] \sim \mathcal{G}(\tfrac{b_0}{2}, \tfrac{b_1}{2}), \\ [\omega|c_0, c_1] \sim \mathcal{B}(c_0, c_1),\ \ [\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3] \sim \mathcal{N}_p(\mathbf{h}_3, \mathbb{H}_3), \\ \left[\sigma_i^{-2}|d_0, d_1\right] \sim \mathcal{G}(\tfrac{d_0}{2}, \tfrac{d_1}{2}) \end{array}\right\}, \tag{3.13}$$

where $\mathcal{W}$, $\mathcal{G}$ and $\mathcal{B}$ stand for Wishart, gamma and beta distributions, respectively, with the gamma parameterization in terms of the shape and rate parameters. Here, the hyperparameters are $\mathbf{h}_1$, $\mathbb{H}_1$, $\mathbf{h}_2$, $\mathbb{H}_2$, a_0, a_1, ν_1, $\mathbb{A}_1$, ν_2, $\mathbb{A}_2$, b_0, b_1, c_0, c_1, $\mathbf{h}_3$, $\mathbb{H}_3$, d_0 and d_1, all of which assumed known. Below in this section, we present the rationale behind the choices of the distributions in (3.13) along with some common choices of the hyperparameters.

The choice of the distributions for **C1** in developing the three-level hierarchical model leads to conditional conjugacy (conditional on everything else including the data) for all the parameters except for $\boldsymbol{\alpha}_i$. We prefer conjugacy here because of extensive computations involved in Bayesian inference. In general, Bayesian inference is based on the posterior distributions of the relevant parameters. However, computations of the posteriors often involve evaluation of multi-dimensional integrals that cannot be done analytically. Therefore,

special techniques are employed to approximate a posterior density. One such technique is the Monte Carlo integration which is convenient and straightforward to implement for hierarchical nonlinear models. However, this technique involves extensive computations, and works through generating random samples from the full conditionals (distribution of a parameter conditioned on all the remaining ones and the data). Conditional conjugacy leads to closed-form full conditionals from which drawing random samples is computationally easier. Therefore, conjugacy is an important consideration in Bayesian inference as it facilitates implementing the Monte Carlo integration; see Chapter 4 for details.

Values of the hyperparameters in (3.13) reflect our prior knowledge. When little is reliably known about the individual trajectories beyond its functional form of the bent-cable, it is reasonable to choose the hyperprior values that lead to fairly vague, minimally informative priors (Carlin [9]).

For us, the most convenient choices of the priors for $\boldsymbol{\mu}_\beta$, $\boldsymbol{\mu}_\alpha$, $\boldsymbol{\phi}$ and μ_{τ_A} are normal distributions: $[\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1] \sim \mathcal{N}_3(\mathbf{h}_1, \mathbb{H}_1)$, $[\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2] \sim \mathcal{N}_2(\mathbf{h}_2, \mathbb{H}_2)$, $[\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3] \sim \mathcal{N}_p(\mathbf{h}_2, \mathbb{H}_2)$ and $[\mu_{\tau_A}|a_0, a_1] \sim \mathcal{N}(a_0, a_1)$, which lead to conditional conjugacy. The choice of a mean vector (e.g., $\mathbf{h}_1$, $\mathbf{h}_2$ or $\mathbf{h}_3$) has very little effect on Bayesian estimation, as long as the respective variance parameters (diagonal elements of $\mathbb{H}_1$, $\mathbb{H}_2$ or $\mathbb{H}_3$, respectively) are taken to be very large which lead to flat priors. Therefore, a common practice is to choose a zero mean vector and a covariance matrix, say, $\mathbb{H}_1$ such that $\mathbb{H}_1^{-1} \approx \mathbb{O}$, where $\mathbb{O}$ is a matrix with all its elements zero (Davidian and Giltinan [19]). According to Song [70], the diagonals of such a covariance matrix should be as large as 10^6. Similarly, $a_0 = 0$ and $a_1^{-1} \approx 0$ lead to a flat prior for μ_{τ_A}.

The reader may refer to Gelman [27] and Spiegelhalter et al. [71] for a discussion on the prior distributions for variance parameters such as σ_i^2 and $\sigma_{\tau_A}^2$. For conjugacy, here we choose inverse gamma distributions for σ_i^2 and $\sigma_{\tau_A}^2$ (or equivalently, gamma distributions for σ_i^{-2} and $\sigma_{\tau_A}^{-2}$). We use the parameterization of the gamma distribution as given in Chib [11]. For example, $[\sigma_i^{-2}|d_0, d_1] \sim \mathcal{G}(\frac{d_0}{2}, \frac{d_1}{2})$ has mean d_0/d_1 and variance $2d_0/d_1^2$. Small values of the hyperprior parameters lead to a flat prior. According to Song [70], d_0 and d_1 can be as small as 10^{-4}. Dominicus et al. [20] used $d_0 = 0.1$ and $d_1 = 0.1$ to analyze cognitive decline by a changepoint model.

The conjugate prior of a covariance matrix such as Σ_β or Σ_α is the inverse Wishart distribution, or equivalently, Wishart for the inverse of the covariance matrix that is Σ_β^{-1}

or Σ_α^{-1}. We use the parameterization of the Wishart distribution as given in Carlin [9] and Wakefield et al. [83]. For example, $[\Sigma_\beta^{-1}|\nu_1, \mathbb{A}_1] \sim W(\nu_1, (\nu_1\mathbb{A}_1)^{-1})$ has degrees of freedom ν_1 and expectation $\mathbb{A}_1^{-1}$. Setting the degrees of freedom equal to the order of the scale matrix (e.g. 3 for the prior of Σ_β^{-1}) makes a Wishart prior nearly flat (Wakefield et al. [83]). Under the parameterization of the Wishart distribution as above, the matrix $\mathbb{A}_1$ (or $\mathbb{A}_2$) is chosen to be an approximate prior estimate of Σ_β (or Σ_α). In the absence of such prior knowledge, one may use the sample covariance matrix of the individual-specific estimates of the regression coefficients; the R (R Development Core Team [61]) library "bentcableAR" [14] for single profile bent-cable regression can be useful in this regard. Note that the choice of $\mathbb{A}_1$ (or $\mathbb{A}_2$) often has very little effect on the result except for cases with very few individuals (Lindley [48]).

Since $0 < \omega < 1$, the obvious choice of the prior for ω is either a beta distribution or an uniform distribution in the interval $(0, 1)$. We choose the beta distribution $[\omega|c_0, c_1] \sim \mathcal{B}(c_0, c_1)$ in our model as it is conditionally conjugate. In the absence of prior information, one may choose $c_0 = c_1 = 1$ which leads to $\mathcal{U}(0, 1)$ distribution.

3.1.4 Summary of the Hierarchy

We developed a 3-level hierarchical model for our mixed bent-cable model in Sections 3.1.1-3.1.3. In summary, these are

$$[\mathbf{y}_i^{(2)}|\ \mathbf{y}_i^{(1)}, \boldsymbol{\theta}_i, \boldsymbol{\phi}, \sigma_i^2] \sim \mathcal{N}_{n_i-p}(\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i), \tag{3.14}$$

$$\left.\begin{array}{c} \big[\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta\big] \sim \mathcal{N}_3(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\ g(\boldsymbol{\alpha}_i|I_i) = (1 - I_i)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma_{\tau_A}^2) + I_i\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \\ [I_i|\omega] \sim \mathcal{BER}(\omega) \end{array}\right\}, \tag{3.15}$$

$$\left.\begin{array}{c} [\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1] \sim \mathcal{N}_3(\mathbf{h}_1, \mathbb{H}_1),\ \ [\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2] \sim \mathcal{N}_2(\mathbf{h}_2, \mathbb{H}_2), \\ [\mu_{\tau_A}|a_0, a_1] \sim \mathcal{N}(a_0, a_1),\ \ \big[\Sigma_\beta^{-1}|\nu_1, \mathbb{A}_1\big] \sim \mathcal{W}\big(\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big), \\ [\Sigma_\alpha^{-1}|\nu_2, \mathbb{A}_2] \sim \mathcal{W}\big(\nu_2, (\nu_2\mathbb{A}_2)^{-1}\big),\ \ \big[\sigma_{\tau_A}^{-2}|b_0, b_1\big] \sim \mathcal{G}(\tfrac{b_0}{2}, \tfrac{b_1}{2}), \\ [\omega|c_0, c_1] \sim \mathcal{B}(c_0, c_1),\ \ [\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3] \sim \mathcal{N}_p(\mathbf{h}_3, \mathbb{H}_3), \\ \big[\sigma_i^{-2}|d_0, d_1\big] \sim \mathcal{G}(\tfrac{d_0}{2}, \tfrac{d_1}{2}) \end{array}\right\}, \tag{3.16}$$

where the first two levels are (3.14) and (3.15), and the third level is (3.16) with the hyperparameters $\mathbf{h}_1$, $\mathbb{H}_1$, $\mathbf{h}_2$, $\mathbb{H}_2$, a_0, a_1, ν_1, $\mathbb{A}_1$, ν_2, $\mathbb{A}_2$, b_0, b_1, c_0, c_1, $\mathbf{h}_3$, $\mathbb{H}_3$, d_0 and d_1. Statistical assumptions **T1**, **T2**, **A1**-**A3**, **B1**-**B5** and **C1**, together with Equations (3.14)-(3.16), constitute our mixed bent-cable model.

3.2 Rationale Behind the Assumptions and Possible Further Extension of the Model

Thus far in this chapter, we developed the mixed bent-cable regression model in a hierarchical framework. The underlying assumptions of our methodology are presented in Section 3.1.2. In this section, we discuss the rationale of those assumptions, and possible further extensions of the mixed bent-cable regression by relaxing some of the underlying assumptions. Note that we developed the mixed bent-cable model in Section 3.1 keeping in mind our motivating examples — the rat and CFC-11 data (Chapter 1). In Chapter 6, we will see that the proposed model is capable of characterizing those two data sets reasonably well. Because of that, and since relaxing the underlying assumption(s) may involve both theoretical and computational complications, we do not pursue these in this thesis. However, depending on the context, relaxation of some the assumptions may be more compatible for other data sets, and therefore, we present a discussion about such possibilities in the following sub-sections.

3.2.1 Assumption T

We take into account through **T** that the sample potentially comes from two populations. Our flexible methodology should be used when there is strong reason to believe that this assumption is reasonable. Recall from Section 3.1.3 that the Bayesian inference is based on the posterior distributions of the parameters, and we employ the Monte Carlo integration technique to approximate those via generating random samples from the full conditionals. However, drawing random samples from some of the full conditionals for the mixture model may produce an indeterminate value, and that can ultimately break down the whole computational process; see Section 4.6 for details. In contrast, data exhibiting only one type of transition is common in many areas and the above computational problem is

irrelevant in modeling such data, we present in Chapter 5 special cases of the mixed-bent cable regression where all the individuals exhibit only one of type of transition (abrupt or gradual).

3.2.2 Assumption A

Assumption **A1** implies a constant variance among the repeated measurements for an individual. Although this assumption could be reasonable in a wide range of applications, it may be somewhat restrictive for some particular types of longitudinal data. For example, a phenomenon frequently observed in pharmacokinetics studies is that variability in measured concentrations increases with the predicted concentrations, reflecting the fact that assay precision is greater at lower concentrations (Wakefield [82]). On the other hand, prostate specific antigen (PSA) laboratory assays are known to have lower precision at lower PSA concentrations (Bellera et al. [6], Carter et al. [10], Eastham et al. [22]). Therefore, a possible extension of our methodology lies in relaxing **A1** to incorporate non-constant variance among the repeated measurements at Level 1. This possibility could be modeled by a variance function which may depend on (a) t_{ij}, or the individual-specific regression coefficients $\boldsymbol{\theta}_i$ through the function $f(t_{ij}, \boldsymbol{\theta}_i)$, and (b) an additional variance parameter vector to represent an increasing/decreasing variance over time. Readers may refer to Pinheiro and Bates [58] for a description of some of these variance functions. Note that sometimes a change of scale (e.g., logarithmic scale) rather than relaxing **A1** can help to deal with increasing/decreasing variances.

Assumption **A2** accounts for additional serial correlation among the repeated measurements remaining after what has been accounted for by the individual-specific regression coefficients $\boldsymbol{\theta}_i$. Our methodology is based on a stationary AR(p) model for **A2**. Note that assuming $\{v_{ij},\ j = 1, 2, \ldots, n_i\}$ in (3.6) are independent and identically distributed random variables with mean 0 and constant variance σ_i^2 (i.e., **A1** holds), an AR(p) model is stationary if and only if the roots of the AR polynomial $\phi(w) = 1 - \phi_1 w - \phi_2 w^2 + \ldots + \phi_p w^p$ all lie outside the unit circle (Box et al. [8]), a condition which can be evaluated numerically using the Schur's theorem for any $p \geq 1$ (readers may refer to Cryer [18] for a description of the algorithm). Thus, nonstationarity can arise if

(a) **A1** is relaxed to incorporate non-constant variance as described above; or

(b) **A1** holds, but at least one root of the AR polynomial lies inside the unit circle.

We remark here that in case of nonstationarity, if ϕ_k's satisfy the unit root condition $\sum_{k=1}^{p} \phi_k = 1$, the intercepts β_{0i}'s are not identifiable (Chib [11], Zellner [84]). Addressing the non-stationarity caused by either (a) or (b) could be an area of future research. Moreover, a more general time series model (e.g., autoregressive moving average (ARMA) model (Box et al. [8])) could be employed to account for more general types of serial correlation. Readers may refer to Chib and Greenberg [12] for a Bayesian approach of modeling the within-individual noise by an ARMA process for linear regression. Note that the AR and ARMA models are appropriate when the measurements are made at equal (or approximately equal) intervals of time. Since the continuous autoregressive (CAR) process (Pinheiro and Bates [58]) can be used for non-equidistant repeated measurements, another useful extension of our methodology could be based on considering the CAR process to model within-individual noises.

Assumption **A3** is a consequence of **A1** and **A2**, which states that the within-individual noise does not depend on the individual-specific regression coefficients $\boldsymbol{\theta}_i$.

3.2.3 Assumption B

Assumption **B1** is a standard assumption for longitudinal studies (Davidian and Giltinan [19], Fitzmaurice et al. [24], Vonesh and Chinchilli [81]), and facilitates the writing down of the joint density compactly (see Section 4.7.1). This assumption simply refers to the fact that repeated measurements for a particular individual are not expected to predict or influence the responses for another individual. For example, it may be reasonable to assume in the rat experiment (Chapters 1 and 6) that effects due to hemorrhage for a particular rat do not predict or influence those for another rat. However, for the CFC-11 data, this assumption could be in question because of the possible spatial effects, if any. Extension of our methodology to incorporate spatial effects through, for example, a conditional autoregressive model (Pettitt et al. [56]) could be an area of future research.

Assumption **B2** may be reasonable in the sense that there is generally no way to infer about $\boldsymbol{\alpha}_i$ from a prior knowledge of $\boldsymbol{\beta}_i$, without first seeing the data. For example, suppose the two linear phases are very steep. This information alone is not sufficient to answer the

question "what is the chance of $\boldsymbol{\alpha}_i$ being, say, $(30, 40)'$ versus $(35, 50)'$?" Therefore, we developed our model in Section 3.1 by assuming that $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ are independent.

The degree of complexity for modeling $\boldsymbol{\theta}_i$ depends on the nature of the data. In general, $\boldsymbol{\theta}_i$ can be modeled by assuming that Level 2 variation is due to (a) entirely unexplained phenomena (Assumption **B3**), or (b) both random and systematic components. The former is reasonable when experimental procedures are consistent from individual to individual, so that it is unlikely for a systematic, identifiable basis for Level 2 variation (Davidian and Giltinan [19]). For example, as the rats are matched with respect to species, gender, age, weight and temperature, and the experiment is conducted in a laboratory setting (Chapter 6), (a) might be reasonable for the rat experimental model. On the other hand, (b) might be more reasonable for the CFC-11 data because of the potential effects due to change in the instrumentation over time (Chapter 6) to monitor the atmospheric concentration of CFC-11. Note that since Level 2 variation under (a) is accounted for, at least partially, by the random components, ignorance of the systematic components may not seriously affect the results. However, an extension via incorporating covariates (e.g., different types of instruments to monitor CFC-11 data) in a linear additive fashion in the model for $\boldsymbol{\theta}_i$ is straightforward, although the technical details are difficult (Davidian and Giltinan [19]).

Although Assumption **B5** is natural to characterize the transition differently for the two populations, **B4** can be relaxed to account for different $\boldsymbol{\beta}_i$'s for the two populations, i.e., when intercept and slope parameters could behave differently between populations. However, the extension would be at the cost of increased computational burden.

3.2.4 Assumption C

The choice of the distributions at levels 1-3 (Section 3.1) leads to conditional conjugacy for all the parameters except for $\boldsymbol{\alpha}_i$ because of its nonlinearity in the regression function. As mentioned in Section 3.1.3, we prefer conjugacy because of extensive computations involved in Bayesian inference. These types of priors under the assumptions of (i) samples coming from a single population, and (ii) AR(0) model for the ϵ_{ij}'s, were used by many other authors, e.g., Bennett et al. [7], Carlin [9], Dominicus et al. [21], Wakefield et al. [83]. Note that a multivariate t distribution can be considered for $\boldsymbol{\beta}_i$ and $\boldsymbol{\phi}$ for

theoretical modeling (Wakefield et al. [83], Chib [11]), but this would require estimation of m additional parameters for each of $\boldsymbol{\beta}_i$ and $\boldsymbol{\phi}$. This consideration might be a future topic.

For $\boldsymbol{\phi}$, the specification of prior normality can readily deal with higher-order AR processes without extra complications, and is flexible in terms of incorporating prior information about stationarity of the error process by truncating the prior distribution (see Chib [11]). For an unrestricted normal prior for $\boldsymbol{\phi}$ as we have chosen here, Chib suggests to retain a draw from the full conditional (Chapter 4) if the stationarity condition for the AR process is satisfied. The proportion of draws thus accepted provides the conditional probability that the process is stationary.

An alternative prior for σ_i^2 (and $\sigma_{\tau_A}^2$) is a uniform distribution on $\log \sigma_i$ on the real line. This is equivalent to $\pi(\sigma_i^2) \propto 1/\sigma_i^2$. However, $\sigma_i^2 = 0$ is theoretically possible, and the improper prior $\pi(\sigma_i^2) \propto 1/\sigma_i^2$ goes to infinity at zero at a rate sufficiently quickly to give a posterior distribution that is also improper (Spiegelhalter et al. [71]). For this reason, and because of conditional conjugacy, we choose gamma priors for σ_i^2 and $\sigma_{\tau_A}^2$ in our modeling approach.

3.3 Discussion

To develop our hierarchical model, we take into consideration both feasibility and practicality. Here, feasibility is accomplished by taking into account (1) the possibility of a sample being drawn from two populations: one is characterized by a gradual transition and the other by an abrupt transition, and (2) a model which is parsimonious, and appealing due to its simple structure and greatly interpretable regression coefficients. On the other hand, practicality is accomplished by specifying our priors so that they are conditionally conjugate. Bayesian inference (Chapter 4) generally involves extensive computations that may require a very long time for evaluation even with high-speed computational facilities. In this regard, conditional conjugacy is an important consideration in developing a hierarchical Bayesian model.

3.4 Chapter Appendix

3.4.1 Level 1

We can write the first level of our hierarchical model as in Equation (3.8). This is formulated from the regression model represented by Equations (3.1), (3.6) and (3.7), i.e.

$$y_{ij} = f_{ij} + \epsilon_{ij}, \quad \epsilon_{ij} = \sum_{k=1}^{p} \phi_k \ \epsilon_{i,j-k} + v_{ij}, \quad [v_{ij}|\sigma_i^2] \sim \mathcal{N}(0, \sigma_i^2).$$

Here,

$$\begin{aligned}
y_{ij} &= f_{ij} + \epsilon_{ij} \\
&= \beta_{0i} + \beta_{1i} \ t_{ij} + \beta_{2i} \ q_{ij} + \epsilon_{ij} \\
&= \beta_{0i} + \beta_{1i} \ t_{ij} + \beta_{2i} \ q_{ij} + \phi_1 \ \epsilon_{i,j-1} + \phi_2 \ \epsilon_{i,j-2} + \ldots + \phi_1 \ \epsilon_{i,j-p} + v_{ij} \\
&= \beta_{0i} + \beta_{1i} \ t_{ij} + \beta_{2i} \ q_{ij} + \phi_1 \ (y_{i,j-1} - \beta_{0i} - \beta_{1i} \ t_{i,j-1} - \beta_{2i} \ q_{i,j-1}) + \\
&\quad \phi_2 \ (y_{i,j-2} - \beta_{0i} - \beta_{1i} \ t_{i,j-2} - \beta_{2i} \ q_{i,j-2}) + \ldots + \\
&\quad \phi_p \ (y_{i,j-p} - \beta_{0i} - \beta_{1i} \ t_{i,j-p} - \beta_{2i} \ q_{i,j-p}) + v_{ij} \\
&= \beta_{0i} \ \Big(1 - \sum_{k=1}^{p} \phi_k\Big) + \beta_{1i} \ \Big(t_{ij} - \sum_{k=1}^{p} \phi_k \ t_{i,j-k}\Big) + \beta_{2i} \ \Big(q_{ij} - \sum_{k=1}^{p} \phi_k \ q_{i,j-k}\Big) + \\
&\quad \sum_{k=1}^{p} \phi_k \ y_{i,j-k} + v_{ij} \\
&= \beta_{0i} \ \Big(1 - \sum_{k=1}^{p} \phi_k\Big) + \beta_{1i} \ x_{ij} + \beta_{2i} \ r_{ij} + \sum_{k=1}^{p} \phi_k \ y_{i,j-k} + v_{ij} \\
&= \mu_{ij} + v_{ij}.
\end{aligned}$$

Now, since $[v_{ij}|\sigma_i^2| \sim \mathcal{N}(0, \sigma_i^2)$ for all j, the first level of the hierarchy is

$$[\mathbf{y}_i^{(2)}| \ \mathbf{y}_i^{(1)}, \boldsymbol{\theta}_i, \boldsymbol{\phi}, \sigma_i^2] \sim \mathcal{N}_{n_i-p}(\boldsymbol{\mu}_i, \sigma_i^2 \ \mathbb{I}_i).$$

3.4.2 Delta Bivariate Mixed Lognormal Distribution

As discussed in Section 3.1, we model γ_i by a delta-lognormal distribution (Aitchison and Brown [1]), which assumes a non-zero probability, $1 - \omega$, that $\gamma_i = 0$, and a probability ω

that $\gamma_i > 0$; and a lognormal distribution for γ_i given $\gamma_i > 0$, that is,

$$g(\gamma_i) = \mathbf{1}(\gamma_i = 0)\ (1-\omega) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}(\gamma_i).$$

Next, for the center of the bend, since τ_i may vary with the two populations (Assumption **B5**), we account for this by conditioning τ_i on γ_i. More specifically, we assume $[\tau_i|\gamma_i = 0]$ and $[\tau_i|\gamma_i > 0]$ are both lognormally distributed, but with different sets of parameters. We denote the probability density functions of these two lognormal distributions by $\mathcal{LN}(\tau_i|\gamma_i = 0)$ and $\mathcal{LN}(\tau_i|\gamma_i > 0)$, respectively. Then, the joint distribution of $\boldsymbol{\alpha}_i$ is denoted by

$$\begin{aligned} g(\boldsymbol{\alpha}_i) &= g(\gamma_i)\ g(\tau_i|\gamma_i) \\ &= \mathbf{1}(\gamma_i = 0)\ (1-\omega)\ g(\tau_i|\gamma_i = 0) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}(\gamma_i)\ g(\tau_i|\gamma_i > 0) \\ &= \mathbf{1}(\gamma_i = 0)\ (1-\omega)\ \mathcal{LN}(\tau_i|\gamma_i = 0) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}(\gamma_i)\ \mathcal{LN}(\tau_i|\gamma_i > 0). \end{aligned} \quad (3.17)$$

Here $\mathcal{LN}(\tau_i|\gamma_i = 0) = \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A})$ and $\mathcal{LN}(\gamma_i) \times \mathcal{LN}_1(\tau_i|\gamma_i > 0) = \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)$, the probability density function of a bivariate lognormal distribution. Therefore, (3.17) becomes

$$g(\boldsymbol{\alpha}_i) = \mathbf{1}(\gamma_i = 0)\ (1-\omega)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha) \quad (3.18)$$

which is the probability density function of a delta bivariate mixed lognormal distribution (Shimizu [66]). Now, (3.18) can be expressed hierarchically as in (3.11). Note that using (3.11), we can write the joint density of $\boldsymbol{\alpha}_i$ and I_i as follows:

$$g(\boldsymbol{\alpha}_i, I_i) = [(1-I_i)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + I_i\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)] \times [\omega^{I_i}\ (1-\omega)^{1-I_i}]$$

As mentioned in Section 3.1.3, marginalizing over I_i gives

$$g(\boldsymbol{\alpha}_i) = \sum_{I_i=0}^{1} [(1-I_i)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + I_i\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)] \times [\omega^{I_i}\ (1-\omega)^{1-I_i}].$$

$$= \mathbf{1}(\gamma_i = 0)\ (1-\omega)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha),$$

which is the probability density function of a delta bivariate mixed lognormal distribution.

Chapter 4

Bayesian Inference for the Mixed Bent-Cable Model

In this chapter, we describe the inference method of the mixed bent-cable regression proposed in Chapter 3. Statistical inference is carried out via the Bayesian technique. Until recently, evaluating integrals in Bayesian computation has been the source of most of the practical difficulties in its application, especially in high dimensions. In most cases, analytic evaluation of the integrals is impossible. Alternative approaches (Gilks [34]) include (1) numerical evaluation; (2) Laplace approximation; and (3) Monte Carlo integration, including Markov chain Monte Carlo (MCMC). MCMC methods are straightforward to implement and provide a unifying framework for approximate the integrals. It is therefore the preferred means of implementation in this thesis.

In Section 4.1, we describe the Bayesian inference technique. There, we introduce the posterior distribution for the mixed bent-cable model based on which the statistical inference is made. In Section 4.2, we present how the MCMC method comes into play in Bayesian inference for our model, and how it works. Generation of Markov chains to implement the Monte Carlo method is described in Section 4.3. In Section 4.4, we discuss some common issues of mixing and convergence of the Markov chain to its stationary distribution, and commonly used diagnostics for them. Computational extensiveness is a common concern to the implementation of the MCMC method. We describe our approach of carrying out the computations in Section 4.5. In Section 4.6, we discuss a few cautionary remarks about using our methodology. We present all the mathematical derivations and other technical materials related to this chapter in the chapter appendix (Section 4.7).

4.1 Bayesian Inference

Statistical inference for our mixed bent-cable model is carried out via a Bayesian approach. The main idea of Bayesian inference is to combine data and prior knowledge on a parameter (or a vector of parameters) to determine its posterior distribution (the conditional density of the parameter given the data). The prior knowledge is supplied in the form of a prior distribution (see Section 3.1.3), which quantifies information and uncertainty about the parameter prior to any data being gathered.

Bayesian inference is made based on the posterior distribution of a parameter. Ideally, one might report the entire posterior distribution, which provides the behavior of the parameter given the data. A graphical display of the posterior density could be useful in this regard. However, it is often more practical to report several numerical characteristics describing the posterior. For example, the posterior mean or median can be considered a point estimate of the parameter. For a symmetric posterior density, the mean and median are identical. For an asymmetric posterior, the median is preferred over the mean because the mean often gives too much weight to extreme values. Note that the mean minimizes the expected posterior squared-error loss function, whereas the median minimizes the expected posterior absolute-error loss function. Thus, other point estimates may be defined using different loss functions. Another important quantity is the posterior standard deviation which measures the uncertainty of the parameter a posteriori. The Bayesian analogue of a frequentist confidence interval is usually referred to as a credible interval or Bayesian confidence interval. A $100(1-2a)\%$ credible interval is (p_1, p_2), where p_1 and p_2 are the a^{th} and $(1-a)^{th}$ quantiles of the posterior density, respectively. This credible interval has a probabilistic interpretation. For example, for $a = 0.025$, the conditional probability that the parameter falls in the interval (p_1, p_2) given the data is 0.95.

4.1.1 Posterior Density

Recall from Chapter 3 that we let $\mathbf{y}_i^{(1)} = (y_{i1}, y_{i2}, \ldots, y_{ip})'$ and $\mathbf{y}_i^{(2)} = (y_{i,p+1}, y_{i,p+2}, \ldots, y_{i,n_i})'$ be the initial p and the next $n_i - p$ observations for the i^{th} individual, respectively, where p is the order of the AR process to model the within-individual noises ϵ_{ij}'s. Then, we denote the observed data from all m individuals by $\mathbf{y} = \left(\mathbf{y}^{(1)'}, \mathbf{y}^{(2)'}\right)'$, where $\mathbf{y}^{(1)} = \left(\mathbf{y}_1^{(1)'}, \mathbf{y}_2^{(1)'}, \ldots, \mathbf{y}_m^{(1)'}\right)'$ and $\mathbf{y}^{(2)} = \left(\mathbf{y}_1^{(2)'}, \mathbf{y}_2^{(2)'}, \ldots, \mathbf{y}_m^{(2)'}\right)'$. Also recall we consider a con-

ditional likelihood framework, where the initial p observations for each i, $\mathbf{y}_i^{(1)}$, are treated as known, whereas $\mathbf{y}_i^{(2)}$ are random.

Next, recall the model parameters from page 22; letting $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \ldots, \boldsymbol{\beta}_m')'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \boldsymbol{\alpha}_2', \ldots, \boldsymbol{\alpha}_m')'$ and $\boldsymbol{\sigma}^{-2} = (\sigma_1^{-2}, \sigma_2^{-2}, \ldots, \sigma_m^{-2})'$, we denote all the model parameters collectively by $\Theta = \big(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big)$. Now, denoting a density function by $\pi(.)$, the joint density of the model parameters and the data can be written as

$$\begin{aligned}
\pi\big(\Theta, \mathbf{y}^{(2)}\big) &= \pi\big(\mathbf{y}^{(2)}|\Theta\big)\pi\big(\Theta\big) \\
&= \pi\big(\mathbf{y}^{(2)}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big) \\
&\times \pi\big(\boldsymbol{\beta}|\boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big) \\
&\times \pi\big(\boldsymbol{\alpha}|\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big) \\
&\times \pi\big(\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big)
\end{aligned} \tag{4.1}$$

where the last term on the right-hand side of Equation (4.1) represents the prior specifications at Level 3 for our mixed bent-cable regression (see Section 3.1.3). The expression of this joint density (4.1) under our assumptions for the probability distributions of the relevant quantities (Assumption **C1**, Section 3.1.2) is given in the chapter appendix (Section 4.7.1).

Now, the posterior density for our model specified in Section 3.1.4 can be written using Bayes' theorem as follows:

$$\pi\big(\Theta|\mathbf{y}^{(2)}\big) = \frac{\pi\big(\Theta, \mathbf{y}^{(2)}\big)}{\pi\big(\mathbf{y}^{(2)}\big)} = \frac{\pi\big(\mathbf{y}^{(2)}|\Theta\big)\pi\big(\Theta\big)}{\pi\big(\mathbf{y}^{(2)}\big)} \tag{4.2}$$

where $\pi\big(\mathbf{y}^{(2)}|\Theta\big)\pi\big(\Theta\big)$ is given by (4.1), and $\pi\big(\mathbf{y}^{(2)}\big) = \int\int\ldots\int \pi\big(\mathbf{y}^{(2)}|\Theta\big)\pi\big(\Theta\big)d\Theta$ is the normalizing factor in Bayes' theorem. Note that calculation of $\pi\big(\mathbf{y}^{(2)}\big)$ requires evaluation of multi-dimensional integrals as Θ represents all the model parameters as described above.

To make inference about a particular parameter (or a vector of parameters), we now need to work out its marginal posterior distribution. For example, inference about the population slope parameter $\boldsymbol{\mu}_\beta$ is based on its marginal posterior density given by

$$\pi\big(\boldsymbol{\mu}_\beta|\mathbf{y}^{(2)}\big) = \frac{\int\int\ldots\int \pi\big(\mathbf{y}^{(2)}|\Theta\big)\pi\big(\Theta\big)d\boldsymbol{\beta}\; d\boldsymbol{\alpha}\; d\boldsymbol{\mu}_\alpha\; d\mu_{\tau_A}\; d\Sigma_\beta^{-1}\; d\Sigma_\alpha^{-1}\; d\sigma_{\tau_A}^{-2}\; d\omega\; d\boldsymbol{\sigma}^{-2}\; d\boldsymbol{\phi}}{\int\int\ldots\int \pi\big(\mathbf{y}^{(2)}|\Theta\big)\pi\big(\Theta\big)d\boldsymbol{\beta}\; d\boldsymbol{\alpha}\; d\boldsymbol{\mu}_\beta\; d\boldsymbol{\mu}_\alpha\; d\mu_{\tau_A}\; d\Sigma_\beta^{-1}\; d\Sigma_\alpha^{-1}\; d\sigma_{\tau_A}^{-2}\; d\omega\; d\boldsymbol{\sigma}^{-2}\; d\boldsymbol{\phi}}.$$

Thus, the main obstacle of Bayesian inference is to evaluate multi-dimensional integrals as illustrated above. However, development of MCMC methods and computational advances in recent decades designed to overcome this obstacle have led to considerable interest in the application of Bayesian techniques to complex modeling problems. The rest of this chapter describes the MCMC methods for the mixed bent-cable regression model, and our approach to implement those techniques.

4.2 Bayesian Inference and MCMC Methods

Though our prior specifications lead to conditional conjugacy for all the parameters except for $\boldsymbol{\alpha}_i$, conjugate priors in the sense that the marginal posteriors are in the same family of distributions as the respective priors, in general, do not exist for a three-level hierarchical nonlinear model. Therefore, the following typically intractable integration problems arise in Bayesian inference, and can be solved by the MCMC techniques (Andrieu et al. [3]):

(a) *Normalization*: to obtain the posterior (4.2) given the prior $\pi\big(\Theta\big)$ and conditional likelihood $\pi\big(\mathbf{y}^{(2)}|\Theta\big)$, the normalizing constant $\pi\big(\mathbf{y}^{(2)}\big)$, which involves evaluation of multi-dimensional integrals, needs to be computed.

(b) *Marginalization*: we need to compute the marginal posterior distribution for a particular parameter (or a vector of parameter) of interest from the posterior (4.2) to make inference about it.

(c) *Expectation*: one of the main objectives in any statistical problem is to obtain summary statistics to summarize a distribution of interest. As mentioned in Section 4.1, the posterior mean can be one such summary measure in Bayesian statistics, and the distribution of interest may be for a certain function of a parameter. Thus, we may wish to evaluate an integral to compute the expectation of this function of the parameter. For example, for $\boldsymbol{\mu}_\beta$, we would evaluate the integral of the form

$$E\big[h(\boldsymbol{\mu}_\beta)|\mathbf{y}^{(2)}\big] = \int h(\boldsymbol{\mu}_\beta)\pi(\boldsymbol{\mu}_\beta|\mathbf{y}^{(2)})d\boldsymbol{\mu}_\beta,$$

where, for instance, $h(\boldsymbol{\mu}_\beta) = \mu_1$ (population incoming slope), or $h(\boldsymbol{\mu}_\beta) = \mu_1 + \mu_2$ (population outgoing slope), or $h(\boldsymbol{\mu}_\beta) = \big\{\mu_1 - E\big[\mu_1|\mathbf{y}^{(2)}\big]\big\}^2$ (posterior variance of the population incoming slope).

Note that once we have a Monte Carlo simulated version of the posterior, we need not do anything analytic, and it is straightforward to carry out the statistical inference through the MCMC principles as we will now present in the subsequent sections.

4.2.1 MCMC Methods

The key idea of MCMC methods is as follows. Let $\{\Theta^{(s)},\ s = 1, 2, \ldots, T, \ldots\}$ be a realization from an appropriately constructed Markov chain with stationary distribution $\pi(.)$. Also, let $\{\Theta^{(s)}\}$ be (1) irreducible, (2) positive recurrent, and (3) aperiodic. Then, under these three regularity conditions,

$$\Theta^{(s)} \xrightarrow{d} \Theta \sim \pi(\Theta|\mathbf{y}^{(2)}) \quad \text{and} \quad \frac{1}{T}\sum_{s=1}^{T} h\big(\Theta^{(s)}\big) \xrightarrow{a.s.} E\big[h(\Theta)|\mathbf{y}^{(2)}\big] \quad as\ T \longrightarrow \infty$$

where the latter result is known as the ergodic theorem. If regularity holds, then after a sufficiently long burn-in (see Section 4.4), say S iterations, the chain gradually "forgets" the initial state, and eventually converges to the stationary distribution, which does not depend on s. Note that, by the ergodic theorem, the sample mean computed from the Markov chain can be considered a point estimate of $h(\Theta)$. Moreover, since the chain converges in distribution to its stationary distribution, it is also legitimate to use the sample median as an estimate of $h(\Theta)$. As described at the beginning of this section (Page 31), the mean and the median are identical for a symmetric distribution, whereas the median is preferred for an asymmetric distribution.

The regularity conditions as mentioned above are required for $\{\Theta^{(s)}\}$ to converge to its stationary distribution because of the following reasons as described by Roberts [64].

1. The irreducible condition ensures that the Markov chain can reach any state in a finite number of iterations from any starting point with a positive probability;

2. the positive recurrent condition ensures that if the starting value $\Theta^{(0)}$ is sampled from $\pi(.)$, then all subsequent iterations will be generated from $\pi(.)$; and

3. the aperiodic condition prevents the Markov chain from oscillating between different sets of states in a regular periodic movement.

Note that although a Markov chain converges under regularity conditions, in practice, such conditions are hard to verify but convergence diagnostics help us identify nonconvergence (see Section 4.4).

In summary, one needs to first construct a Markov chain that necessarily converges to $\pi(.)$. To summarize $\pi(.)$, one can produce the marginal posterior density plot using kernel density estimation; then, the posterior mean, median, standard deviation and other summaries can be approximated by their sample equivalents in the MCMC output $\{\Theta^{(s)},\ s = 1, 2, \ldots, T\}$; and the $100(1-2a)\%$ credible interval (p_1, p_2), where p_1 and p_2 are the a^{th} and $(1-a)^{th}$ quantiles of the marginal posterior density, respectively, is produced also using the MCMC quantiles.

For example, the population incoming and outgoing slopes, and the CTP for Population G are common parameters of interest, for which $h(\Theta) = \mu_1$, $h(\Theta) = \mu_1 + \mu_2$, and $h(\Theta) = \exp\{\mu_\tau\} - \exp\{\mu_\gamma\} - 2\mu_1 \exp\{\mu_\gamma\}/\mu_2$, respectively. Note that $\exp\{\mu_\gamma\}$ and $\exp\{\mu_\tau\}$ are the medians of the individual-specific random coefficients γ_i and τ_i, respectively, as our Level 2 assumption for the $\boldsymbol{\alpha}_i$'s involves lognormal distributions. Now, discarding the burn-in samples, marginal posterior means of these quantities are approximated by

$$\hat{\mu}_1 = \frac{1}{T-S} \sum_{s=S+1}^{T} \mu_1^{(s)},$$

$$\widehat{\mu_1 + \mu_2} = \frac{1}{T-S} \sum_{s=S+1}^{T} \left(\mu_1^{(s)} + \mu_2^{(s)}\right),$$

$$\text{and} \qquad \widehat{\text{CTP}} = \frac{1}{T-S} \sum_{s=S+1}^{T} \left(\exp\{\mu_\tau^{(s)}\} - \exp\{\mu_\gamma^{(s)}\} - 2\mu_1^{(s)} \exp\{\mu_\gamma^{(s)}\}/\mu_2^{(s)}\right).$$

Similarly, we can use Level 2 theoretical medians, $\exp\{\mu_\gamma\}$ and $\exp\{\mu_\tau\}$ for Population G and $\exp\{\mu_{\tau_A}\}$ for Population A, to describe the transition locations. For Population

G, we can also use Level 2 theoretical standard deviations of γ_i and τ_i with the formulas $\big(\exp\{2\mu_\gamma + (\Sigma_\alpha)_{11}\} \times [\exp\{(\Sigma_\alpha)_{11}\} - 1]\big)^{1/2}$ and $\big(\exp\{2\mu_\tau + (\Sigma_\alpha)_{22}\} \times [\exp\{(\Sigma_\alpha)_{22}\} - 1]\big)^{1/2}$, respectively, and the theoretical correlation between γ_i and τ_i with the formula $\big(\exp\{(\Sigma_\alpha)_{12}\} - 1\big)/\big([\exp\{(\Sigma_\alpha)_{11}\} - 1] \times [\exp\{(\Sigma_\alpha)_{22}\} - 1]\big)^{1/2}$ to describe the between-individual variability of these transition parameters, where $(\Sigma_\alpha)_{11}$, $(\Sigma_\alpha)_{22}$ and $(\Sigma_\alpha)_{12}$ are, respectively, the $(1,1)$, $(2,2)$ and $(1,2)$ elements of Σ_α. All these quantities can be estimated easily from the MCMC output $\{\Theta^{(s)},\ s = 1, 2, \ldots, T\}$.

Keeping in mind that each parameter has its own posterior distribution in Bayesian statistics, and thus there are many ways to produce point and interval estimates for each parameter, we now summarize our approach of making inference for the mixed bent-cable model:

(a) use posterior means/medians as estimates of the parameters of interest, and the a^{th} and $(1-a)^{th}$ quantiles of the posteriors for the corresponding $100(1-2a)\%$ credible intervals;

(b) since the estimates cannot be worked out analytically, approximate those by the MCMC counterparts; and

(c) produce individual-specific and population fitted values.

To explain our approach for (c), first consider the i^{th} individual. Note that the parameter vectors $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ have their own posterior distributions, so the bent-cable function $f(t_{ij}, \boldsymbol{\beta}_i, \boldsymbol{\alpha}_i) \equiv f_{ij}$ itself has a posterior distribution at each observed time point t_{ij}, $j = 1, 2, \ldots, n_i$. Now, the question is: are we interested in the posterior of f_{ij} at every observed t_{ij}, or in a general impression of the function $f(t, \boldsymbol{\beta}_i, \boldsymbol{\alpha}_i)$ over $t \in [0, \infty)$ based on the posteriors of $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$? There is certainly no single right answer to this question. In this thesis, we consider the posterior of the bent-cable function to produce the fitted values by taking the MCMC sample means of the bent-cable function (we describe the other approach based on the posteriors of $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$ in the chapter appendix in Section 4.7.2). In our approach, the bent cable for the i^{th} individual at observed time t_{ij} is $h(\Theta) = f_{ij} = \beta_{0i} + \beta_{1i} t_{ij} + \beta_{2i} q_{ij}$, where q_{ij} is given by (3.3), and the corresponding fitted values are

$$\hat{f}_{ij} = \frac{1}{T-S} \sum_{s=S+1}^{T} \Big(\beta_{0i}^{(s)} + \beta_{1i}^{(s)}\ t_{ij} + \beta_{2i}^{(s)}\ q_{ij}^{(s)} \Big), \quad j = 1, 2, \ldots, n_i, \tag{4.3}$$

where

$$q_{ij}^{(s)} = \frac{\left(t_{ij} - \tau_i^{(s)} + \gamma_i^{(s)}\right)^2}{4\gamma_i^{(s)}} \mathbf{1}\Big\{|t_{ij} - \tau_i^{(s)}| \leq \gamma_i^{(s)}\Big\} + \Big(t_{ij} - \tau_i^{(s)}\Big)\mathbf{1}\Big\{t_{ij} - \tau_i^{(s)} > \gamma_i^{(s)}\Big\}.$$

Similarly, the population fitted value at time $t \in [0, C]$, where C is the maximum time point observed in the data set, is

$$\hat{f}_t = \frac{1}{T-S} \sum_{s=S+1}^{T} \Big(\mu_0^{(s)} + \mu_1^{(s)}\ t + \mu_2^{(s)}\ q_t^{(s)}\Big),$$

where for Population G

$$\begin{aligned} q_t^{(s)} =& \frac{\left(t - \exp\{\mu_\tau^{(s)}\} + \exp\{\mu_\gamma^{(s)}\}\right)^2}{4\exp\{\mu_\gamma^{(s)}\}}\ \mathbf{1}\Big\{|t - \exp\{\mu_\tau^{(s)}\}| \leq \exp\{\mu_\gamma^{(s)}\}\Big\} \\ &+ \Big(t - \exp\{\mu_\tau^{(s)}\}\Big)\ \mathbf{1}\Big\{t - \exp\{\mu_\tau^{(s)}\} > \exp\{\mu_\gamma^{(s)}\}\Big\}, \end{aligned}$$

and for Population A

$$q_t^{(s)} = \Big(t - \exp\{\mu_\tau^{(s)}\}\Big)\ \mathbf{1}\Big\{t - \exp\{\mu_\tau^{(s)}\} > 0\Big\}.$$

Individual-specific and population fitted curves can then be interpolated based on the $\hat{f}_{ij}$ and $\hat{f}_t$ values, respectively.

4.3 Construction of a Markov Chain

We now describe our approach to form a Markov chain for the mixed bent-cable model. We employ the Metropolis within Gibbs algorithm (Smith and Roberts [69]), a form of the Markov chain schemes which is convenient for nonlinear hierarchical regression models (Davidian and Giltinan [19], Wakefield et al. [83]). We first give a preview of the Metropolis-Hastings algorithm in a general context, and then describe the Gibbs sampler in the context of our mixed bent-cable model pointing out the problem in its implementation, and finally the Metropolis within Gibbs algorithm to complete the description of our technique. Note that the Gibbs and Metropolis within Gibbs samplers are based

on the full conditionals of each of the components of Θ, where a full conditional refers to the distribution of a parameter conditioned on all the remaining ones and the data. For example, the full conditional for $\boldsymbol{\beta}_i$, denoted by $\pi\big(\boldsymbol{\beta}_i|\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_{i-1}, \boldsymbol{\beta}_{i+1}, \ldots, \boldsymbol{\beta}_m, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}, \mathbf{y}^{(2)}\big)$, is the distribution of $\boldsymbol{\beta}_i$ given $\mathbf{y}^{(2)}$ and all the remaining components of Θ.

4.3.1 Metropolis-Hastings Algorithm

Suppose that $\boldsymbol{\zeta} \sim k(.)$, where $k(.)$ can be evaluated but not easily sampled. Here, $k(.)$ is often referred to as the target distribution, and could be a full conditional in implementing the Gibbs algorithm that may or may not have a closed form expression. The Metropolis-Hastings algorithm can be used to generate a draw from a distribution that approximates $k(.)$. Description of this algorithm can be found in Hastings [40] and Metropolis et al. [51].

To implement the algorithm, a proposal distribution (see below in this section) is needed from which to sample a candidate point. Let $\boldsymbol{\zeta}^{(0)}$ be an arbitrary starting point and g denote the proposal distribution. Given $\boldsymbol{\zeta}^{(s)}$ at the s^{th} iteration, the algorithm generates $\boldsymbol{\zeta}^{(s+1)}$ as follows:

1. Draw a candidate point $\boldsymbol{\zeta}^*$ from $g\big(.|\boldsymbol{\zeta}^{(s)}\big)$;
2. compute the Metropolis-Hastings ratio
$$R\big(\boldsymbol{\zeta}^{(s)}, \boldsymbol{\zeta}^*\big) = \frac{k\big(\boldsymbol{\zeta}^*\big)g\big(\boldsymbol{\zeta}^{(s)}|\boldsymbol{\zeta}^*\big)}{k\big(\boldsymbol{\zeta}^{(s)}\big)g\big(\boldsymbol{\zeta}^*|\boldsymbol{\zeta}^{(s)}\big)};$$
3. if $U \leq \min\big\{1,\ R\big(\boldsymbol{\zeta}^{(s)}, \boldsymbol{\zeta}^*\big)\big\}$, accept $\boldsymbol{\zeta}^*$ with probability $\min\big\{1,\ R\big(\boldsymbol{\zeta}^{(s)}, \boldsymbol{\zeta}^*\big)\big\}$ and set $\boldsymbol{\zeta}^{(s+1)} = \boldsymbol{\zeta}^*$, otherwise $\boldsymbol{\zeta}^{(s+1)} = \boldsymbol{\zeta}^{(s)}$, where U is a random variable from $\mathcal{U}(0, 1)$;
4. increment s and return to step 1.

Clearly, a chain constructed in such a way is Markov as $\boldsymbol{\zeta}^{(s+1)}$ depends on the history only through $\boldsymbol{\zeta}^{(s)}$. A proposal distribution can have any form. The rate of convergence to the stationary distribution, however, depends largely on the choice of the proposal distribution. A good proposal distribution should well approximate the target distribution (the full conditional for a Gibbs sampler), so that it produces candidate values that (a) cover the

support of the stationary distribution in a reasonable number of iterations, and (b) are not accepted or rejected too frequently (Chib and Greenberg [13]). A reasonable acceptance rate is in the range $[0.15, 0.50]$ for high-dimensional target distributions (Roberts [64]).

There are several variants of the Metropolis-Hastings algorithm. The Metropolis algorithm (Metropolis et al. [51]) considers only symmetric proposal distributions of the form $g\big(\boldsymbol{\zeta}^*|\boldsymbol{\zeta}^{(s)}\big) = g\big(\boldsymbol{\zeta}^{(s)}|\boldsymbol{\zeta}^*\big)$ so that the Metropolis-Hastings ratio reduces to

$$R\big(\boldsymbol{\zeta}^{(s)}, \boldsymbol{\zeta}^*\big) = \frac{k\big(\boldsymbol{\zeta}^*\big)}{k\big(\boldsymbol{\zeta}^{(s)}\big)}.$$

A special case of the Metropolis algorithm is the random-walk Metropolis (see Gilks et al. [35], Givens and Hoeting [37]), for which the candidate point is generated from $\boldsymbol{\zeta}^* = \boldsymbol{\zeta}^{(s)} + u$ where u is a point drawn from a symmetric proposal density $g(.)$. In this case, $g(\boldsymbol{\zeta}^*|\boldsymbol{\zeta}^{(s)}) = g(|\boldsymbol{\zeta}^* - \boldsymbol{\zeta}^{(s)}|)$ (Gilks et al. [35]).

Tierney [74] and Gelman [25] propose another variant of the Metropolis-Hastings algorithm to sample from approximate full conditionals whilst maintaining exactly the required stationary distribution of the Markov chain. Their approach uses an approximate full conditional as a proposal distribution $g(.)$ in an independence-type Metropolis-Hastings algorithm, for which $g\big(\boldsymbol{\zeta}^*|\boldsymbol{\zeta}^{(s)}\big) = g(\boldsymbol{\zeta}^*)$. The Metropolis-Hastings ratio for this method reduces to

$$R\big(\boldsymbol{\zeta}^{(s)}, \boldsymbol{\zeta}^*\big) = \frac{k(\boldsymbol{\zeta}^*)g(\boldsymbol{\zeta}^{(s)})}{k(\boldsymbol{\zeta}^{(s)})g(\boldsymbol{\zeta}^*)}.$$

4.3.2 Gibbs Sampler

Recall that Θ denotes the model parameters for our mixed bent-cable regression. It is often convenient and computationally efficient to break down Θ into components of possibly differing dimensions, and then update each of these components one by one. This framework was originally proposed by Metropolis et al. [51] for MCMC, and often referred to as the single-component Metropolis-Hastings algorithm. The Gibbs sampler (Geman and Geman [31]), a method for obtaining the marginals of interest from the set of full conditionals, is a special case of this single-component Metropolis-Hastings algorithm where the acceptance probability (probability of accepting the candidate in a Metropolis step) is one (Gilks et al. [34]).

The breakdown of Θ into components depends on practical considerations. In some applications, it may be convenient to break down Θ into its scalar components, while in other cases, the components could be vectors or matrices. From a practical point of view, one important consideration in choosing the components is to have full conditionals that have closed form expressions, since the MCMC method is computationally expensive, especially when generating samples from a distribution that can be expressed only up to a proportionality constant. Another important consideration is the correlation structure in the sequence of $\Theta^{(s)}$ at different iteration lags. Smith and Roberts [69] indicated that if highly correlated components are treated individually, there could be painfully slow convergence of the chain to equilibrium as a result of very little movement at each conditional random variate generation step. They also pointed out that if, on the other hand, correlated components are blocked together, the problem is avoided, but perhaps at the expense of having to perform a draw from a multivariate conditional distribution.

The Gibbs algorithm proceeds as follows. Pick an arbitrary starting point $\Theta^{(0)}$. Then generate an instance from the full conditional of each of the components of Θ conditional on the current values of the remaining components. This completes a transition from $\Theta^{(0)}$ to $\Theta^{(1)}$. Iterating through this cycle of generating data from the full conditionals gives a sequence $\Theta^{(0)}, \Theta^{(1)}, \ldots, \Theta^{(T)}$. A sequence thus generated is a Markov chain with stationary distribution $\pi\big(\Theta|\mathbf{y}^{(2)}\big)$ under the regularity conditions as described in Section 4.2.1 (Davidian and Giltinan [19]). Monte Carlo integration can now be applied to the converged Markov chain for Bayesian inference. See Section 4.4 for a discussion of the convergence issues of a Markov chain.

Note that the usual context of a Gibbs sampler refers to all full conditionals having closed form. We work out the full conditionals in Section 4.7.3 (appendix) for all the components of Θ, and show that all of those have a closed form expression except for $\boldsymbol{\alpha}_i$. For this reason, the usual context of Gibbs sampler does not hold for the mixed bent-cable model.

4.3.3 Metropolis within Gibbs Algorithm

Recall that each step in the Gibbs sampler is a Metropolis-Hastings step with acceptance probability one. So, it is permissible to use different Metropolis-Hastings variants in a Gibbs step (Givens and Hoeting [37]). Such an algorithm is called Metropolis within Gibbs. This

algorithm is particularly useful when any of the full conditionals does not have a closed form expression like the one for $\boldsymbol{\alpha}_i$ for our mixed bent-cable model. In implementing the algorithm, we therefore employ the Gibbs steps for all the components of Θ except for $\boldsymbol{\alpha}_i$ for which a Metropolis step is considered.

Now, we present the steps involved in the construction of the Markov chain. Noting that the full conditional $\pi(.|.)$ for a particular component is a function of only some components of Θ but not all, we omit the irrelevant components from the condition in the expression of $\pi(.|.)$. For example, in addition to the data $\mathbf{y}_i^{(2)}$, the full conditional for $\boldsymbol{\beta}_i$ depends only on $\boldsymbol{\alpha}_i$, σ_i^{-2}, $\boldsymbol{\mu}_\beta$, Σ_β^{-1} and $\boldsymbol{\phi}$, and hence we use the notation $\pi\big(\boldsymbol{\beta}_i|\boldsymbol{\alpha}_i, \sigma_i^{-2}, \boldsymbol{\mu}_\beta, \Sigma_\beta^{-1}, \boldsymbol{\phi}, \mathbf{y}_i^{(2)}\big)$ to denote it. Now, recall from Section 3.1.3 that I_i is a latent allocation variable such that $I_i = 0$ when $\gamma_i = 0$, and $I_i = 1$ when $\gamma_i > 0$. Letting $\mathbf{I} = (I_1, I_2, \ldots, I_m)$, suppose that $\{\Theta^{(0)}, \mathbf{I}^{(0)}\}$ is an arbitrary starting point, and $\{\Theta^{(s)}, \mathbf{I}^{(s)}\}$ is its update at iteration s. Given $\{\Theta^{(s)}, \mathbf{I}^{(s)}\}$, our choice of order to update the components leads to the following sequence to achieve the new set $\{\Theta^{(s+1)}, \mathbf{I}^{(s+1)}\}$ in one iteration:

1. For $i = 1$, generate $\boldsymbol{\beta}_i^{(s+1)} \sim \pi\big(\boldsymbol{\beta}_i|\ \boldsymbol{\alpha}_i^{(s)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\beta^{(s)}, \Sigma_\beta^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\beta}_i|\ \boldsymbol{\alpha}_i^{(s)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\beta^{(s)}, \Sigma_\beta^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ is a trivariate normal distribution;

2. for $i = 1$, generate $\boldsymbol{\alpha}_i^{(s+1)} \sim \pi\big(\boldsymbol{\alpha}_i|\ \boldsymbol{\beta}_i^{(s+1)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\alpha^{(s)}, \mu_{\tau_A}^{(s)}, \Sigma_\alpha^{-1(s)}, \sigma_{\tau_A}^{-2(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ $\big($and hence $I_i^{(s+1)}\big)$ via a Metropolis step, where $\pi\big(\boldsymbol{\alpha}_i|\ \boldsymbol{\beta}_i^{(s+1)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\alpha^{(s)}, \mu_{\tau_A}^{(s)}, \Sigma_\alpha^{-1(s)}, \sigma_{\tau_A}^{-2(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ can be expressed only up to a proportionality constant;

3. for $i = 1$, generate $\sigma_i^{-2(s+1)} \sim \pi\big(\sigma_i^{-2}|\ \boldsymbol{\beta}_i^{(s+1)}, \boldsymbol{\alpha}_i^{(s+1)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a Gibbs step, where $\pi\big(\sigma_i^{-2}|\ \boldsymbol{\beta}_i^{(s+1)}, \boldsymbol{\alpha}_i^{(s+1)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ is a gamma distribution;

4. repeat 1–3 for $i = 2, 3, \ldots, m$; these complete updating the individual-specific parameters, that is, we now have $\boldsymbol{\beta}^{(s+1)}$, $\boldsymbol{\alpha}^{(s+1)}$ $\big($and hence $\mathbf{I}^{(s+1)}\big)$ and $\boldsymbol{\sigma}^{-2(s+1)}$;

5. generate $\Sigma_\beta^{-1(s+1)} \sim \pi\big(\Sigma_\beta^{-1}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}_\beta^{(s)}\big)$ via a Gibbs step, where $\pi\big(\Sigma_\beta^{-1}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}_\beta^{(s)}\big)$ is a Wishart distribution;

6. generate $\Sigma_\alpha^{-1(s+1)} \sim \pi\big(\Sigma_\alpha|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \boldsymbol{\mu}_\alpha^{(s)}\big)$ via a Gibbs step, where $\pi\big(\Sigma_\alpha|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \boldsymbol{\mu}_\alpha^{(s)}\big)$ is a Wishart distribution;

7. generate $\sigma_{\tau_A}^{-2(s+1)} \sim \pi\big(\sigma_{\tau_A}^{-2}|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \mu_{\tau_A}^{(s)}\big)$ via a Gibbs step, where $\pi\big(\sigma_{\tau_A}^{-2}|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \mu_{\tau_A}^{(s)}\big)$ is a gamma distribution;

8. generate $\boldsymbol{\mu}_\beta^{(s+1)} \sim \pi\big(\boldsymbol{\mu}_\beta|\ \boldsymbol{\beta}^{(s+1)}, \Sigma_\beta^{-1(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\mu}_\beta|\ \boldsymbol{\beta}^{(s+1)}, \Sigma_\beta^{-1(s+1)}\big)$ is a trivariate normal distribution;

9. generate $\boldsymbol{\mu}_\alpha^{(s+1)} \sim \pi\big(\boldsymbol{\mu}_\alpha|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \Sigma_\alpha^{-1(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\mu}_\alpha|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \Sigma_\alpha^{-1(s+1)}\big)$ is a bivariate normal distribution;

10. generate $\mu_{\tau_A}^{(s+1)} \sim \pi\big(\mu_{\tau_A}|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \sigma_{\tau_A}^{-2(s+1)}\big)$ via a Gibbs step, where $\pi\big(\mu_{\tau_A}|\ \boldsymbol{\alpha}^{(s+1)}, \mathbf{I}^{(s+1)}, \sigma_{\tau_A}^{-2(s+1)}\big)$ is normal distribution;

11. generate $\boldsymbol{\phi}^{(s+1)} \sim \pi\big(\boldsymbol{\phi}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\alpha}^{(s+1)}, \boldsymbol{\sigma}^{-2(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\phi}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\alpha}^{(s+1)}, \boldsymbol{\sigma}^{-2(s+1)}\big)$ is a p-variate normal distribution;

12. generate $\omega^{(s+1)} \sim \pi\big(\omega|\ \mathbf{I}^{s+1}\big)$ via a Gibbs step, where $\pi\big(\omega|\ \mathbf{I}^{s+1}\big)$ is a beta distribution.

The Bayesian inference can now be made using the monte carlo integration described in Section 4.2.

4.4 Mixing and Convergence

The efficiency of an MCMC algorithm is reflected by good mixing of a chain. The mixing property of a chain includes how quickly a chain "forgets" its initial values and how quickly it fully explores the support and the shape of the target distribution. Besides good mixing, we must also be concerned about the convergence of the chain, that is, whether the chain has approximately reached its stationary distribution.

Since there is substantial overlap between the goals of diagnosing convergence to the stationary distribution, and investigating the mixing property of a chain (Givens and Hoeting [37]), we combine the discussion of mixing and convergence based on existing works on these topics in the following sub-sections. In this regard, although the discussion is given in the context of our mixed bent-cable regression model, the underlying material in this section is mainly review – much of the material can be found in Gilks et al. [36] and Givens and Hoeting [37].

4.4.1 Number of Chains

When a chain at some point reaches around the mode of the target distribution, it is possible that it may stay there forever. If this is the case, even though the chain does not fully explore the support and the shape of the target distribution, a convergence diagnostic may indicate that the chain has converged to the stationary distribution. A partial solution to this problem is to run several independent chains, and then investigate the within-chain and between-chain behavior. However, recommendation concerning running one chain or multiple chains is conflicting in the literature ranging from several long chains (Gelman and Rubin [29], [30]) to one very long one (Geyer [32]), each with its own merits and demerits (see Gilks et al. [35], Givens and Hoeting [37]).

The motivation for running multiple chains is the hope that at least one of them will explore the features (e.g., modes) of the target distribution, and to detect the wash out of the influence of the starting values (Givens and Hoeting [37]). On the other hand, as pointed out by Gilks et al. [35], running several very long chains is not practical because (1) one very long run has the best chance to explore the features of the target distribution, and (2) comparison between chains can never prove convergence. Moreover, some of the chains may carry the same information as that of some of the others. To balance between running a single very long chain and running multiple long chains, we consider two very long independent chains (5,000,000 MCMC iterations for each) to analyze both the rat and CFC-11 data introduced in Chapter 1. However, in our simulation study, due to computational extensiveness, we consider one chain with 100,000 MCMC iterations to analyze each simulated data set.

4.4.2 Burn-in and Stopping Time

The MCMC method can yield marginal posterior distributions only in the limits; thus, in practice, the dependence of the chain on the starting value may remain strong even after running the chain for a sufficiently long time. If a chain starts with an initial value that is far from the posterior mode, this dependence may make the chain converge slowly. To reduce the severity of this problem, an initial S iterations are discarded as a burn-in period from a chain.

For stopping time T, in general, the aim is to run a chain sufficiently long to obtain good mixing. One informal way (Gilks et al. [35]) of deciding the stopping time is to run

several long chains and to compare the estimates (posterior means/medians) from each chain. If the estimates from the different chains do not agree adequately, then the run length, $L = T - S$, should be increased.

A popular technique for determining the burn-in and run length is the Gelman-Rubin statistic R [30]. The technique is based on a comparison of within-chain and between-chain variances (readers may refer to the reference manual for the "coda" package (Plummer et al. [59]) in R for a description of this technique). Values of R substantially above 1 indicate lack of convergence (some authors suggest that $R < 1.2$ is acceptable, e.g., Gelman et al. [28]). Now, if the chosen S does not yield an acceptable R, then S or L or preferably both should be increased.

To analyze the rat and CFC-11 data, we examine both the Gelman-Rubin statistic and the graphical approach to choose the burn-in and stopping time.

4.4.3 Thinning

A chain that has poor mixing properties generally exhibits slow decay of autocorrelation. Therefore, it is good practice for the inference to be based on every l^{th} iteration of a chain, with l set to some value high enough that successive draws are approximately independent (Gelman [26]). This strategy is known as thinning. Gelman [26] also points out that thinning can be useful when the set of simulated MCMC values is so large that reducing the number of simulations by a factor of l gives important savings in storage and computation time. Examining the autocorrelation plots, we take $l = 200$ in analyzing both the rat and CFC-11 data, and $l = 20$ in the simulation study.

4.4.4 Graphical Diagnostics

In Sections 4.4.2, we describe the Gelman-Rubin statistic, a formal statistical tool for examining the mixing and convergence of a Markov chain. In this section, we present three widely used graphical diagnostics, which we take into consideration in addition to the Gelman-Rubin statistic to examine the mixing and convergence for the rat and CFC-11 data.

Recall that good mixing implies that the chain quickly "forgets" its initial values and fully explores the support and shape of the target distribution. One obvious diagnostic tool

is a trace plot, where one plots the realization of the chain versus the iteration number. A trace plot can be useful to see if a chain is mixing rapidly. In addition, a clear trend in the trace plot indicates that stationarity has not been achieved. This in turn suggests that a longer run is necessary. A chain that is mixing well will quickly move away from its starting value, no matter where it started, and the samples will wiggle about vigorously in the region supported by the posterior density (Givens and Hoeting [37]). Visual inspection of a trace plot can also be useful to determine the burn-in as well as stopping time.

Another useful graphical diagnostic is the autocorrelation plot. An autocorrelation plot displays the serial correlation in the chain at different lags of iteration. For a highly autocorrelated chain, the Gibbs sampler is slow to explore the entire support of the posterior distribution. Typically, the autocorrelation declines as the lag increases. If this is not the case, thinning can be explored.

The kernel density plot of the chain is another useful diagnostic tool. In the absence of high autocorrelation, nonconvergence is sometimes reflected in multimodal distributions, which is especially true if the density plot displays not only multiple modes but also lumpiness rather than a smooth curve. Note that high autocorrelation within a converged chain can also reveal such behavior in the density plot. In such cases, it may be necessary to run the chain for longer and/or with heavier thinning.

4.5 Software Implementation

We wrote our computer code in R (R Development Core Team [61]) and MATLAB, but found them inefficient in terms of computational speed. Then, we wrote our own code in C to generate MCMC samples. The functions for random sample generation from standard univariate distributions such as uniform, normal, chi-square, gamma, beta and Bernoulli, and matrix algebra such as Cholesky decomposition and inversion of a symmetric positive definite matrix, are countered by interfacing C with R. We have written other functions required to generate samples from the full conditionals such as calculations of $\mathbf{z}_i$, $\mathbf{x}_i$, $\mathbf{r}_i$, residual sum of squares and likelihood function, data generation from multivariate normal and Wishart distributions, extracting a column from a matrix, and so on.

In implementing the Metropolis within Gibbs algorithm, we follow the updating order of Θ to maximize the computational speed, which is a common concern for MCMC methods.

That is, we first update the individual-specific components ($\boldsymbol{\beta}_i$, $\boldsymbol{\alpha}_i$, and σ_i^{-2}) one by one, and calculate the required terms to update the population components within a single loop over all individuals i's ($i = 1, 2, \ldots, m$). These calculations are done through steps 1-4 of our chosen updating algorithm described in Section 4.3.3. After completing the loop, we update the population components in an appropriate order by noting the dependency of a particular component on the remaining ones in the full conditionals. This procedure saves a considerable amount of time when running and constructing the desired Markov chain.

We write a function "update" for one MCMC iteration. Then, another function uses this "update" to perform the required number of iterations with given burn-in and thinning criteria, and stores the MCMC samples in external text files, which we subsequently analyze using the "coda" package (Plummer et al. [59]) in R.

4.6 Discussion

Since some standard regularity conditions do not hold, frequentist inference is complicated for bent-cable regression, even for a single profile (Chiu et al. [16], Chiu and Lockhart [15]). Our proposed Bayesian approach for inference makes no special regularity assumptions; the trade-off is the need to evaluate high-dimensional integrals. Although computationally intensive, its implementation is straightforward through the use of MCMC numerical integration. Moreover, in Bayesian inference, the full behavior of a parameter can be readily investigated via its posterior distribution, rather than relying on the parameter estimator's asymptotic distribution that may be far off from the actual distribution in a finite-sample setting.

In practice, some caution is required for the following reason. If a particular MCMC iteration yields $I_i = 0$ or 1 for all i, then the full conditionals for some parameters depend only on their priors; in the case of very diffuse priors, a draw in the Gibbs algorithm may yield an indeterminate value, rendering a computational breakdown of the MCMC. As an example, we see that the full conditional for $\sigma_{\tau_A}^{-2}$ (Section 4.7.3) reduces simply to its prior $\mathcal{G}\big(\frac{b_0}{2}, \frac{b_1}{2}\big)$ when $I_i = 1$ for all i. So, if $I_i = 1$ for all i at a particular iteration, a diffuse prior for $\sigma_{\tau_A}^{-2}$ (e.g., $b_0 = b_1 = 0.01$) may lead to a near-zero sample point, and this may break down the whole process of generating samples to construct a Markov chain. Similarly, if $I_i = 0$ for all i, the full conditional for $\boldsymbol{\mu}_\alpha$ reduces to its prior, and therefore, a similar problem may arise. To avoid this problem, one may specify informative priors if prior

information is available. Note that this problem is irrelevant for certain special cases of our methodology: Model A for which $I_i = 0$ for all i, and Model G for which $I_i = 1$ for all i.

4.7 Chapter Appendix

4.7.1 Joint Density

Under the assumption that the priors are independent, and using **B1**, the joint density of the model parameters and the data as given in (4.1) can be rewritten as

$$\begin{aligned}\pi\big(\Theta, \mathbf{y}^{(2)}\big) &= \pi\big(\mathbf{y}^{(2)}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big) \\ &\times \prod_{i=1}^{m} \pi\big(\boldsymbol{\beta}_i|\boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big) \\ &\times \prod_{i=1}^{m} \pi\big(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \mu_{\tau_A}, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \sigma_{\tau_A}^{-2}, \omega, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big) \\ &\times \big\{\pi\big(\boldsymbol{\mu}_\beta\big)\ \pi\big(\boldsymbol{\mu}_\alpha\big)\ \pi\big(\mu_{\tau_A}\big)\ \pi\big(\Sigma_\beta^{-1}\big)\ \pi\big(\Sigma_\alpha^{-1}\big)\ \pi\big(\sigma_{\tau_A}^{-2}\big)\ \pi\big(\omega\big)\ \pi\big(\boldsymbol{\sigma}^{-2}\big)\ \pi\big(\boldsymbol{\phi}\big)\big\}. \end{aligned} \tag{4.4}$$

For our model specifications in (3.14)-(3.16), and under the assumptions **T1**, **T2**, **A1**-**A3**, **B1**-**B5** and **C1** (Section 3.1.2), the joint density (4.4) becomes

$$\begin{aligned}\pi\big(\Theta, \mathbf{y}^{(2)}\big) &= \prod_{i=1}^{m} \mathcal{N}_{n_i-p}(\mathbf{y}^{(2)}|\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i) \times \prod_{i=1}^{m} \mathcal{N}_3\big(\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta\big) \\ &\times \prod_{i=1}^{m} \Delta\big(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha, \mu_{\tau_A}, \sigma_{\tau_A}^2, \omega\big) \times \mathcal{N}_2\big(\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1\big) \times \mathcal{N}_2\big(\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2\big) \\ &\times \mathcal{N}\big(\mu_{\tau_A}|a_0, a_1\big) \times \mathcal{W}\big(\Sigma_\beta^{-1}|\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big) \times \mathcal{W}\big(\Sigma_\alpha^{-1}|\nu_2, (\nu_2\mathbb{A}_2)^{-1}\big) \\ &\times \mathcal{G}\Big(\sigma_{\tau_A}^{-2}\Big|\frac{b_0}{2}, \frac{b_1}{2}\Big) \times \mathcal{B}(\omega|c_0, c_1) \times \prod_{i=1}^{m} \mathcal{G}\Big(\sigma_i^{-2}\Big|\frac{d_0}{2}, \frac{d_1}{2}\Big) \times \mathcal{N}_p\big(\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3\big), \end{aligned}$$

where $\Delta\big(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha, \mu_{\tau_A}, \sigma_{\tau_A}^2, \omega\big)$ denotes the pdf of the delta bivariate mixed lognormal distribution given in (3.12). Now, recalling the latent allocation variables $\mathbf{I} = (I_1, I_2, \ldots, I_m)'$, we get

$$\begin{aligned}\pi\big(\Theta, \mathbf{I}, \mathbf{y}^{(2)}\big) &= \prod_{i=1}^{m} \mathcal{N}_{n_i-p}(\mathbf{y}^{(2)}|\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i) \times \prod_{i=1}^{m} \mathcal{N}_3\big(\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta\big) \\ &\times \prod_{i=1}^{m} \big\{(1 - I_i)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma_{\tau_A}^2) + I_i\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)\big\} \end{aligned}$$

$$
\begin{aligned}
&\times \prod_{i=1}^{m} \mathcal{BER}(I_i|\omega) \times \mathcal{N}_3\big(\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1\big) \times \mathcal{N}_2\big(\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2\big) \\
&\times \mathcal{N}\big(\mu_{\tau_A}|a_0, a_1\big) \times \mathcal{W}\big(\Sigma_\beta^{-1}|\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big) \times \mathcal{W}\big(\Sigma_\alpha^{-1}|\nu_2, (\nu_2\mathbb{A}_2)^{-1}\big) \\
&\times \mathcal{G}\Big(\sigma_{\tau_A}^{-2}\Big|\frac{b_0}{2}, \frac{b_1}{2}\Big) \times \mathcal{B}(\omega|c_0, c_1) \times \prod_{i=1}^{m} \mathcal{G}\Big(\sigma_i^{-2}\Big|\frac{d_0}{2}, \frac{d_1}{2}\Big) \times \mathcal{N}_p\big(\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3\big) \\
&= \prod_{i=1}^{m} \mathcal{N}_{n_i-p}(\mathbf{y}^{(2)}|\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i) \times \prod_{i=1}^{m} \mathcal{N}_3\big(\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta\big) \\
&\times \prod_{i=1}^{m} \Big\{ \big[\mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma_{\tau_A}^2)\big]^{1-I_i} \times \big[\mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)\big]^{I_i} \Big\} \\
&\times \prod_{i=1}^{m} \mathcal{BER}(I_i|\omega) \times \mathcal{N}_3\big(\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1\big) \times \mathcal{N}_2\big(\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2\big) \\
&\times \mathcal{N}\big(\mu_{\tau_A}|a_0, a_1\big) \times \mathcal{W}\big(\Sigma_\beta^{-1}|\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big) \times \mathcal{W}\big(\Sigma_\alpha^{-1}|\nu_2, (\nu_2\mathbb{A}_2)^{-1}\big) \\
&\times \mathcal{G}\Big(\sigma_{\tau_A}^{-2}\Big|\frac{b_0}{2}, \frac{b_1}{2}\Big) \times \mathcal{B}(\omega|c_0, c_1) \times \prod_{i=1}^{m} \mathcal{G}\Big(\sigma_i^{-2}\Big|\frac{d_0}{2}, \frac{d_1}{2}\Big) \times \mathcal{N}_p\big(\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3\big),
\end{aligned}
\tag{4.5}
$$

which is the basis of deriving the full conditionals to implement the Metropolis within Gibbs algorithm for our mixed bent-cable regression.

4.7.2 Approximating Fitted Values

We described our approach to approximate the individual-specific and population fitted values based on the posteriors of the respective bent-cable functions in Section 4.2.1. As mentioned there, the fitted values can also be produced based on the posterior means/medians of the regression coefficients, denoted by $\hat{\boldsymbol{\beta}}_i$ and $\hat{\boldsymbol{\alpha}}_i$ for $i = 1, 2, \ldots, m$. For example, the fitted values for the i^{th} individual are then

$$
\hat{f}_{ij} = \frac{1}{T-S} \sum_{s=S+1}^{T} \big(\hat{\beta}_{0i} + \hat{\beta}_{1i}\ t_{ij} + \hat{\beta}_{2i}\ \hat{q}_{ij}\big), \quad j = 1, 2, \ldots, n_i, \tag{4.6}
$$

where

$$
\hat{q}_{ij} = \frac{\big(t_{ij} - \hat{\tau}_i + \hat{\gamma}_i\big)^2}{4\hat{\gamma}_i} \mathbf{1}\big\{|t_{ij} - \hat{\tau}_i| \le \hat{\gamma}_i\big\} + \big(t_{ij} - \hat{\tau}_i\big)\mathbf{1}\big\{t_{ij} - \hat{\tau}_i > \hat{\gamma}_i\big\}.
$$

Note that the first two terms of (4.3) and (4.6) are exactly the same. However, there can be small differences between the two approaches due to averaging q_{ij} as a quantity over the MCMC iterations as in (4.3), which is different from treating $q_{ij}()$ as a function, and plugging in the means/medians of the other quantities into $q_{ij}()$ as in (4.6).

Continuing with the idea of (4.6), the population fitted value at time $t \in [0, C]$ is

$$\hat{f}_t = \frac{1}{T-S} \sum_{s=S+1}^{T} \big(\hat{\mu}_0 + \hat{\mu}_1\ t + \hat{\mu}_2\ \hat{q}_t\big),$$

where for Population G

$$\hat{q}_t = \frac{\big(t - \exp\{\hat{\mu}_\tau\} + \exp\{\hat{\mu}_\gamma\}\big)^2}{4\exp\{\hat{\mu}_\gamma\}}\ \mathbf{1}\big\{|t - \exp\{\hat{\mu}_\tau\}| \le \exp\{\hat{\mu}_\gamma\}\big\}$$
$$+ \big(t - \exp\{\hat{\mu}_\tau\}\big)\ \mathbf{1}\big\{t - \exp\{\hat{\mu}_\tau\} > \exp\{\hat{\mu}_\gamma\}\big\},$$

and for Population A

$$\hat{q}_t = \big(t - \exp\{\hat{\mu}_\tau\}\big)\ \mathbf{1}\big\{t - \exp\{\hat{\mu}_\tau\} > 0\big\}.$$

4.7.3 Full Conditionals

In general, full conditionals are derived from the joint distribution (4.5). To construct the full conditional for, say, $\boldsymbol{\beta}_i$, we need only to pick out the terms in the joint density which involves $\boldsymbol{\beta}_i$. Note that any term which does not depend on $\boldsymbol{\beta}_i$ can be taken as a proportionality constant in the full conditional.

An appealing feature of the bent-cable function is that it is partially linear – given $\boldsymbol{\alpha}_i$, f_{ij} as defined by Equation (3.2) is linear – and we can exploit this fact to derive a closed-form full conditional for $\boldsymbol{\beta}_i$. However, the full conditional of $\boldsymbol{\alpha}_i$ can be expressed only up to a proportionality constant, and is given by

$$\pi(\boldsymbol{\alpha}_i|.) \propto \exp\Big\{-\frac{1}{2\sigma_i^2}(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)\Big\} \times \Big[\frac{1}{\tau_i}\exp\Big\{-\frac{1}{2\sigma_{\tau_A}^2}(\kappa_i - \mu_{\tau_A})^2\Big\}\Big]^{1-I_i}$$
$$\times \Big[\frac{1}{\gamma_i\tau_i}\exp\Big\{-\frac{1}{2}(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)'\ \Sigma_\alpha^{-1}\ (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\Big\}\Big]^{I_i} \tag{4.7}$$

where $z_{ij} = y_{ij} - \sum_{k=1}^{p} \phi_k \; y_{i,j-k}$, $x_{ij} = t_{ij} - \sum_{k=1}^{p} \phi_k \; t_{i,j-k}$, $r_{ij} = q_{ij} - \sum_{k=1}^{p} \phi_k \; q_{i,j-k}$, $\mathbf{z}_i = (z_{i,p+1}, z_{i,p+2}, \ldots, z_{i,n_i})'$ and

$$\mathbb{X}_i = \begin{pmatrix} 1 - \sum_{k=1}^{p} \phi_k & x_{i,p+1} & r_{i,p+1} \\ 1 - \sum_{k=1}^{p} \phi_k & x_{i,p+2} & r_{i,p+2} \\ \vdots & \vdots & \vdots \\ 1 - \sum_{k=1}^{p} \phi_k & x_{i,n_i} & r_{i,n_i} \end{pmatrix}.$$

For the full conditionals of the remaining components of Θ, let

- $m_A = \sum_{i=1}^{m} (1 - I_i)$ and $m_G = \sum_{i=1}^{m} I_i$;
- $\boldsymbol{\xi}_i = \log \boldsymbol{\alpha}_i = (\log \gamma_i, \log \tau_i)'$ and $\kappa_i = \log \tau_i$;
- $\tilde{\boldsymbol{\beta}} = \sum_{i=1}^{m} \boldsymbol{\beta}_i$, $\tilde{\boldsymbol{\xi}} = \sum_{i=1}^{m} I_i \; \boldsymbol{\xi}_i$, and $\tilde{\kappa} = \sum_{i=1}^{m} (1 - I_i) \; \kappa_i$;
- $\mathbb{M}_i^{-1} = \sigma_i^{-2} \; \mathbb{X}_i' \; \mathbb{X}_i + \Sigma_\beta^{-1}$, $\mathbb{U}_1^{-1} = m \; \Sigma_\beta^{-1} + \mathbb{H}_1^{-1}$, and $\mathbb{U}_2^{-1} = m_G \; \Sigma_\alpha^{-1} + \mathbb{H}_2^{-1}$; and
- $\epsilon_{ij} = y_{ij} - \beta_{0i} - \beta_{1i} \, t_{ij} - \beta_{2i} \, q_{ij}$ for $j = p+1, p+2, \ldots, n_i$, $\boldsymbol{\epsilon}_i = (\epsilon_{i,p+1}, \epsilon_{i,p+2}, \ldots, \epsilon_{i,n_i})'$, and $\mathbb{V}^{-1} = \sum_{i=1}^{m} \sigma_i^{-2} \; \mathbb{W}_i' \; \mathbb{W}_i + \mathbb{H}_3^{-1}$, where $\mathbb{W}_i$ is a $(n_i - p) \times p$ matrix given by

$$\mathbb{W}_i = \begin{pmatrix} \epsilon_{i,p} & \epsilon_{i,p-1} & \epsilon_{i,1} \\ \epsilon_{i,p+1} & \epsilon_{i,p} & \epsilon_{i,2} \\ \vdots & \vdots & \vdots \\ \epsilon_{i,n_i-1} & \epsilon_{i,n_i-2} & \epsilon_{i,n_i-p} \end{pmatrix}.$$

Then one can verify (see below) that the full conditionals are

$$[\boldsymbol{\beta}_i | .] \sim \mathcal{N}_3\Big(\mathbb{M}_i \; \big(\sigma_i^{-2} \; \mathbb{X}_i' \; \mathbf{z}_i + \Sigma_\beta^{-1} \; \boldsymbol{\mu}_\beta\big), \; \mathbb{M}_i\Big),$$

$$[\boldsymbol{\mu}_\beta | .] \sim \mathcal{N}_3\Big(\mathbb{U}_1 \; \big(\Sigma_\beta^{-1} \; \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1} \; \mathbf{h}_1\big), \; \mathbb{U}_1\Big),$$

$$[\boldsymbol{\mu}_\alpha|.] \sim \mathcal{N}_2\Big(\mathbb{U}_2\ \big(\Sigma_\alpha^{-1}\ \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1}\ \mathbf{h}_2\big),\ \mathbb{U}_2\Big),$$

$$[\mu_{\tau_A}|.] \sim \mathcal{N}\bigg(\frac{\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}},\ \frac{1}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}\bigg),$$

$$[\Sigma_\beta^{-1}|.] \sim \mathcal{W}\bigg(m + \nu_1,\ \Big[\sum_{i=1}^{m}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\ (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_1 \mathbb{A}_1\Big]^{-1}\bigg),$$

$$[\Sigma_\alpha^{-1}|.] \sim \mathcal{W}\bigg(m_G + \nu_2,\ \Big[\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\ (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' + \nu_2 \mathbb{A}_2\Big]^{-1}\bigg),$$

$$[\sigma_{\tau_A}^{-2}|.] \sim \mathcal{G}\bigg(\frac{m_A + b_0}{2}, \frac{\sum_{i=1}^{m}(1 - I_i)(\kappa_i - \mu_{\tau_A})^2 + b_1}{2}\bigg),$$

$$[\sigma_i^{-2}|.] \sim \mathcal{G}\bigg(\frac{n_i - p + d_0}{2},\ \frac{(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i) + d_1}{2}\bigg),$$

$$[\boldsymbol{\phi}|.] \sim \mathcal{N}_p\Big(\mathbb{V}\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big),\ \mathbb{V}\Big),$$

$$[\omega|.] \sim \mathcal{B}(m_G + c_0, m_A + c_1).$$

Derivations of the above results are presented as follows.

Full Conditional for $\boldsymbol{\beta}_i$

Picking out the terms in the joint density (4.5) which involves $\boldsymbol{\beta}_i$, we get

$$\pi(\boldsymbol{\beta}_i|.) \propto \exp\Big\{-\frac{1}{2\sigma_i^2}\big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)'\big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big) - \frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\Sigma_\beta^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\Big\}.$$

From Section 3.4, we have

$$\begin{aligned} y_{ij} - \mu_{ij} &= y_{ij} - \beta_{0i}\ \Big(1 - \sum_{k=1}^{p}\phi_k\Big) - \beta_{1i}\ x_{ij} - \beta_{2i}\ r_{ij} - \sum_{k=1}^{p}\phi_k\ y_{i,j-k} \\ &= z_{ij} - \beta_{0i}\ \Big(1 - \sum_{k=1}^{p}\phi_k\Big) - \beta_{1i}\ x_{ij} - \beta_{2i}\ r_{ij} \quad \text{for} \quad j = p+1, p+2, \ldots, n_i. \end{aligned}$$

Therefore, In vector-matrix notation, $\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i = \mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i$. Using this result, we get

$$\begin{aligned}
\pi(\boldsymbol{\beta}_i|\ .) &\propto \exp\Big\{ -\frac{1}{2}\ \big[\sigma_i^{-2}\ \big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big)'\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big) + (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\ \Sigma_\beta^{-1}\ (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\big]\Big\} \\
&= \exp\Big\{ -\frac{1}{2}\ \big[\sigma_i^{-2}\ \mathbf{z}_i'\ \mathbf{z}_i - \sigma_i^{-2}\ \mathbf{z}_i'\ \mathbb{X}_i\ \boldsymbol{\beta}_i - \sigma_i^{-2}\ \boldsymbol{\beta}_i'\ \mathbb{X}_i'\ \mathbf{z}_i + \sigma_i^{-2}\ \boldsymbol{\beta}_i'\ \mathbb{X}_i'\ \mathbb{X}_i\ \boldsymbol{\beta}_i + \\
&\quad \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i - \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i + \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big]\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}\ \big[-\sigma_i^{-2}\ \mathbf{z}_i'\ \mathbb{X}_i\ \boldsymbol{\beta}_i - \sigma_i^{-2}\ \boldsymbol{\beta}_i'\ \mathbb{X}_i'\ \mathbf{z}_i + \sigma_i^{-2}\ \boldsymbol{\beta}_i'\ \mathbb{X}_i'\ \mathbb{X}_i\ \boldsymbol{\beta}_i + \\
&\quad \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i - \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i\big]\Big\}
\end{aligned}$$

$\big[$*proportionality follows because* $\sigma_i^{-2}\ \mathbf{z}_i'\ \mathbf{z}_i$ *and* $\boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta$ *do not depend on* $\boldsymbol{\beta}_i\big]$

$$= \exp\Big\{ -\frac{1}{2}\ \big[-2\ \sigma_i^{-2}\ \boldsymbol{\beta}_i'\ \mathbb{X}_i'\ \mathbf{z}_i + \sigma_i^{-2}\ \boldsymbol{\beta}_i'\ \mathbb{X}_i'\ \mathbb{X}_i\ \boldsymbol{\beta}_i + \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i - 2\ \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big]\Big\}$$

$\big[\sigma_i^{-2}\ \mathbf{z}_i'\ \mathbb{X}_i\ \boldsymbol{\beta}_i$ *and* $\boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i$ *are scalars*$\big]$

$$\begin{aligned}
&= \exp\Big\{ -\frac{1}{2}\ \big[-2\boldsymbol{\beta}_i'\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big) + \boldsymbol{\beta}_i'\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbb{X}_i + \Sigma_\beta^{-1}\big)\ \boldsymbol{\beta}_i\big]\Big\} \\
&= \exp\Big\{ -\frac{1}{2}\ \big[-2\boldsymbol{\beta}_i'\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big) + \boldsymbol{\beta}_i'\ \mathbb{M}_i^{-1}\ \boldsymbol{\beta}_i\big]\Big\} \\
&\propto \exp\Big\{ -\frac{1}{2}\ \big[-\boldsymbol{\beta}_i'\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big) - \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)'\ \boldsymbol{\beta}_i + \\
&\quad \boldsymbol{\beta}_i'\ \mathbb{M}_i^{-1}\ \boldsymbol{\beta}_i + \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)'\ \mathbb{M}_i\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)\big]\Big\}
\end{aligned}$$

$\Big[\boldsymbol{\beta}_i'\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)$ *is a scalar, and so is* $\boldsymbol{\beta}_i'\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big) =$

$\big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)'\ \boldsymbol{\beta}_i$*; proportionality follows because* $\big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)'\ \mathbb{M}_i$

$\big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)$ *does not depend on* $\boldsymbol{\beta}_i\Big]$

$$= \exp\Big\{ -\frac{1}{2}\ \big[\boldsymbol{\beta}_i - \mathbb{M}_i\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)\big]'\ \mathbb{M}_i^{-1}\ \big[\boldsymbol{\beta}_i - \mathbb{M}_i\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)\big]\Big\},$$

which is proportional to the pdf of a trivariate normal distribution with mean vector $\mathbb{M}_i\big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big)$ and covariance matrix $\mathbb{M}_i$. Therefore,

$$[\boldsymbol{\beta}_i|.] \sim \mathcal{N}_3\Big(\mathbb{M}_i\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big),\ \mathbb{M}_i\Big).$$

Full Conditional for $\boldsymbol{\alpha}_i$

Picking out the terms in the joint density (4.5) which involves $\boldsymbol{\alpha}_i$, it is easy to see that the full conditional for $\boldsymbol{\alpha}_i$ is given by (4.7).

Full Conditional for $\boldsymbol{\mu}_\beta$

Picking out the terms in the joint density (4.5) which involve $\boldsymbol{\mu}_\beta$, we get

$$\begin{aligned}
\pi(\boldsymbol{\mu}_\beta|.) &\propto \prod_{i=1}^{m} \exp\Big\{-\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\ \Sigma_\beta^{-1}\ (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\Big\} \exp\Big\{-\frac{1}{2}(\boldsymbol{\mu}_\beta - \mathbf{h}_1)'\ \mathbb{H}_1^{-1}\ (\boldsymbol{\mu}_\beta - \mathbf{h}_1)\Big\} \\
&= \exp\Big\{-\frac{1}{2}\Big[\sum_{i=1}^{m}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\ \Sigma_\beta^{-1}\ (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) + (\boldsymbol{\mu}_\beta - \mathbf{h}_1)'\ \mathbb{H}_1^{-1}\ (\boldsymbol{\mu}_\beta - \mathbf{h}_1)\Big]\Big\} \\
&= \exp\Big\{-\frac{1}{2}\Big[\sum_{i=1}^{m}(\boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i - \boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i + \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta) + \\
&\quad (\boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \mathbf{h}_1 - \mathbf{h}_1'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta + \mathbf{h}_1'\ \mathbb{H}_1^{-1}\ \mathbf{h}_1)\Big]\Big\} \\
&\propto \exp\Big\{-\frac{1}{2}\Big[\sum_{i=1}^{m}(-\boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i + \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta) + \\
&\quad \boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta - \boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \mathbf{h}_1 - \mathbf{h}_1'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta\Big]\Big\}
\end{aligned}$$

$\big[$*proportionality follows because* $\boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i$ *and* $\mathbf{h}_1'\ \mathbb{H}_1^{-1}\ \mathbf{h}_1$ *do not depend on* $\boldsymbol{\mu}_\beta\big]$

$$= \exp\Big\{-\frac{1}{2}\Big[-2\boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \sum_{i=1}^{m}\boldsymbol{\beta}_i + m\ \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta + \boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta - 2\boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \mathbf{h}_1\Big]\Big\}$$

$\big[\boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta$ *and* $\mathbf{h}_1'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta$ *are scalars, and so are* $\boldsymbol{\beta}_i'\ \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta = \boldsymbol{\mu}_\beta'\ \Sigma_\beta^{-1}\ \boldsymbol{\beta}_i$ *and*

$\mathbf{h}_1'\ \mathbb{H}_1^{-1}\ \boldsymbol{\mu}_\beta = \boldsymbol{\mu}_\beta'\ \mathbb{H}_1^{-1}\ \mathbf{h}_1\big]$

$$= \exp\Big\{-\frac{1}{2}\Big[-2\ \boldsymbol{\mu}_\beta'\ \Big(\Sigma_\beta^{-1}\ \sum_{i=1}^{m}\boldsymbol{\beta}_i + \mathbb{H}_1^{-1}\ \mathbf{h}_1\Big) + \boldsymbol{\mu}_\beta'\ (m\ \Sigma_\beta^{-1} + \mathbb{H}_1^{-1})\ \boldsymbol{\mu}_\beta\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\big[-2\ \boldsymbol{\mu}_\beta'\ (\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1) + \boldsymbol{\mu}_\beta'\ \mathbb{U}_1^{-1}\ \boldsymbol{\mu}_\beta\big]\Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\big[-\boldsymbol{\mu}_\beta'\ (\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1) - (\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)'\ \boldsymbol{\mu}_\beta + \boldsymbol{\mu}_\beta'\ \mathbb{U}_1^{-1}\ \boldsymbol{\mu}_\beta +$$

$$(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)'\ \mathbb{U}_1\ (\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)\big]\Big\}$$

$\Big[\boldsymbol{\mu}_\beta'\ (\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)$ *is a scalar, and so is* $\boldsymbol{\mu}_\beta'(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1) =$

$(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)'\boldsymbol{\mu}_\beta$*; proportionality follows because* $(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)'\ \mathbb{U}_1$

$(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)$ *does not depend on* $\boldsymbol{\mu}_\beta\Big]$

$$= \exp\Big\{ -\frac{1}{2}\big[\boldsymbol{\mu}_\beta - \mathbb{U}_1(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)\big]'\ \mathbb{U}_1^{-1}\ \big[\boldsymbol{\mu}_\beta - \mathbb{U}_1(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1)\big]\Big\},$$

which is proportional to the pdf of a trivariate normal distribution with mean vector $\mathbb{U}_1\ \big(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1\big)$ and covariance matrix $\mathbb{U}_1$. Therefore,

$$[\boldsymbol{\mu}_\beta|.] \sim \mathcal{N}_3\Big(\mathbb{U}_1\ \big(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1\big),\ \mathbb{U}_1\Big).$$

Full Conditional for $\boldsymbol{\mu}_\alpha$

Picking out the terms in the joint density (4.5) which involve $\boldsymbol{\mu}_\alpha$, and recalling $\boldsymbol{\xi}_i = (\log\gamma_i, \log\tau_i)'$, we get

$$\pi(\boldsymbol{\mu}_\alpha|.) \propto \prod_{i=1}^{m} \exp\Big\{ -\frac{I_i}{2}(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)'\ \Sigma_\alpha^{-1}\ (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\Big\} \exp\Big\{ -\frac{1}{2}(\boldsymbol{\mu}_\alpha - \mathbf{h}_2)'\ \mathbb{H}_2^{-1}\ (\boldsymbol{\mu}_\alpha - \mathbf{h}_2)\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\Big[\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)'\ \Sigma_\alpha^{-1}\ (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) + (\boldsymbol{\mu}_\alpha - \mathbf{h}_2)'\ \mathbb{H}_2^{-1}\ (\boldsymbol{\mu}_\alpha - \mathbf{h}_2)\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\Big[\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i'\ \Sigma_\alpha^{-1}\ \boldsymbol{\xi}_i - \boldsymbol{\xi}_i'\ \Sigma_\alpha^{-1}\ \boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha'\ \Sigma_\alpha^{-1}\ \boldsymbol{\xi}_i + \boldsymbol{\mu}_\alpha'\ \Sigma_\alpha^{-1}\ \boldsymbol{\mu}_\alpha) +$$

$$(\boldsymbol{\mu}_\alpha'\ \mathbb{H}_2^{-1}\ \boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha'\ \mathbb{H}_2^{-1}\ \mathbf{h}_2 - \mathbf{h}_2'\ \mathbb{H}_2^{-1}\ \boldsymbol{\mu}_\alpha + \mathbf{h}_2'\ \mathbb{H}_2^{-1}\ \mathbf{h}_2)\Big]\Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\Big[\sum_{i=1}^{m} I_i(-\boldsymbol{\xi}_i'\ \Sigma_\alpha^{-1}\ \boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha'\ \Sigma_\alpha^{-1}\ \boldsymbol{\xi}_i + \boldsymbol{\mu}_\alpha'\ \Sigma_\alpha^{-1}\ \boldsymbol{\mu}_\alpha) +$$

$$\boldsymbol{\mu}_\alpha'\ \mathbb{H}_2^{-1}\ \boldsymbol{\mu}_\alpha - \boldsymbol{\mu}_\alpha'\ \mathbb{H}_2^{-1}\ \mathbf{h}_2 - \mathbf{h}_2'\ \mathbb{H}_2^{-1}\ \boldsymbol{\mu}_\alpha\Big]\Big\}$$

$\big[$*proportionality follows because* $\boldsymbol{\xi}_i' \, \Sigma_\alpha^{-1} \, \boldsymbol{\xi}_i$ *and* $\mathbf{h}_2' \, \mathbb{H}_2^{-1} \, \mathbf{h}_2$ *do not depend on* $\boldsymbol{\mu}_\alpha\big]$

$$= \exp\Big\{ -\frac{1}{2}\Big[-2\boldsymbol{\mu}_\alpha' \, \Sigma_\alpha^{-1} \, \sum_{i=1}^{m} I_i \boldsymbol{\xi}_i + m_G \, \boldsymbol{\mu}_\alpha' \, \Sigma_\alpha^{-1} \, \boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\alpha' \, \mathbb{H}_2^{-1} \, \boldsymbol{\mu}_\alpha - 2\boldsymbol{\mu}_\alpha' \, \mathbb{H}_2^{-1} \, \mathbf{h}_2 \Big] \Big\}$$

$\big[\boldsymbol{\xi}_i' \, \Sigma_\alpha^{-1} \, \boldsymbol{\mu}_\alpha$ *and* $\mathbf{h}_2' \, \mathbb{H}_2^{-1} \, \boldsymbol{\mu}_\alpha$ *are scalars, and so are* $\boldsymbol{\xi}_i' \, \Sigma_\alpha^{-1} \, \boldsymbol{\mu}_\alpha = \boldsymbol{\mu}_\alpha' \, \Sigma_\alpha^{-1} \, \boldsymbol{\xi}_i$

and $\mathbf{h}_2' \, \mathbb{H}_2^{-1} \, \boldsymbol{\mu}_\alpha = \boldsymbol{\mu}_\alpha' \, \mathbb{H}_2^{-1} \, \mathbf{h}_2\big]$

$$= \exp\Big\{ -\frac{1}{2}\Big[-2 \, \boldsymbol{\mu}_\alpha' \, \Big(\Sigma_\alpha^{-1} \, \sum_{i=1}^{m} I_i \boldsymbol{\xi}_i + \mathbb{H}_2^{-1} \, \mathbf{h}_2 \Big) + \boldsymbol{\mu}_\alpha' \, (m_G \, \Sigma_\alpha^{-1} + \mathbb{H}_2^{-1}) \, \boldsymbol{\mu}_\alpha \Big] \Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\big[-2 \, \boldsymbol{\mu}_\alpha' \, (\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2) + \boldsymbol{\mu}_\alpha' \, \mathbb{U}_2^{-1} \, \boldsymbol{\mu}_\alpha \big] \Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\big[-\boldsymbol{\mu}_\alpha' \, (\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2) - (\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)' \, \boldsymbol{\mu}_\alpha + \boldsymbol{\mu}_\alpha' \, \mathbb{U}_2^{-1} \, \boldsymbol{\mu}_\alpha +$$

$$(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)' \, \mathbb{U}_2 \, (\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)\big] \Big\}$$

$\Big[\boldsymbol{\mu}_\alpha' \, (\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)$ *is a scalar, and so is* $\boldsymbol{\mu}_\alpha'(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2) =$

$(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)' \boldsymbol{\mu}_\alpha$*; proportionality follows because* $(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)' \, \mathbb{U}_2$

$(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)$ *does not depend on* $\boldsymbol{\mu}_\alpha\Big]$

$$= \exp\Big\{ -\frac{1}{2}\big[\boldsymbol{\mu}_\alpha - \mathbb{U}_2(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)\big]' \, \mathbb{U}_2^{-1} \, \big[\boldsymbol{\mu}_\alpha - \mathbb{U}_2(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2)\big] \Big\},$$

which is proportional to the pdf of a bivariate normal distribution with mean vector $\mathbb{U}_2 \, \big(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2\big)$ and covariance matrix $\mathbb{U}_2$. Therefore,

$$[\boldsymbol{\mu}_\alpha|.] \sim \mathcal{N}_2\Big(\mathbb{U}_2 \, \big(\Sigma_\alpha^{-1} \, \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1} \, \mathbf{h}_2\big), \, \mathbb{U}_2\Big).$$

Full Conditional for μ_{τ_A}

Picking out the terms in the joint density (4.5) which involve μ_{τ_A}, and letting $\kappa_i = \log \tau_i$, we get

$$\pi(\mu_{\tau_A}|.) \propto \prod_{i=1}^{m} \exp\Big\{ -\frac{1 - I_i}{2\sigma_{\tau_A}^2}(\kappa_i - \mu_{\tau_A})^2 \Big\} \exp\Big\{ -\frac{1}{2a_1}(\mu_{\tau_A} - a_0)^2 \Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\Big[\sum_{i=1}^{m}(1-I_i)\ \sigma_{\tau_A}^{-2}\ (\kappa_i-\mu_{\tau_A})^2 + a_1^{-1}\ (\mu_{\tau_A}-a_0)^2\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\Big[\sum_{i=1}^{m}(1-I_i)(\sigma_{\tau_A}^{-2}\ \kappa_i^2 - 2\sigma_{\tau_A}^{-2}\ \kappa_i\ \mu_{\tau_A} + \sigma_{\tau_A}^{-2}\ \mu_{\tau_A}^2)+$$

$$(a_1^{-1}\ \mu_{\tau_A}^2 - 2a_1^{-1}\ a_0\ \mu_{\tau_A} + a_1^{-1}\ a_0)\Big]\Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\Big[\sum_{i=1}^{m}(1-I_i)(-2\sigma_{\tau_A}^{-2}\ \kappa_i\ \mu_{\tau_A} + \sigma_{\tau_A}^{-2}\ \mu_{\tau_A}^2) + (a_1^{-1}\ \mu_{\tau_A}^2 - 2a_1^{-1}\ a_0\ \mu_{\tau_A})\Big]\Big\}$$

$\big[$*proportionality follows because* $\sigma_{\tau_A}^{-2}\ \kappa_i^2$ *and* $a_1^{-1}\ a_0$ *are constants with respect to* $\mu_{\tau_A}\big]$

$$= \exp\Big\{ -\frac{1}{2}\big(m_A\ \sigma_{\tau_A}^{-2}\ \mu_{\tau_A}^2 - 2\sigma_{\tau_A}^{-2}\ \tilde{\kappa}\ \mu_{\tau_A} + a_1^{-1}\ \mu_{\tau_A}^2 - 2a_1^{-1}\ a_0\ \mu_{\tau_A}\big)\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\big[(m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1})\mu_{\tau_A}^2 - 2\mu_{\tau_A}(\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0)\big]\Big\}$$

$$\propto \exp\Big\{ -\frac{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}{2}\ \Big[\mu_{\tau_A}^2 - 2\mu_{\tau_A}\Big(\frac{\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}\Big) + \Big(\frac{\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}\Big)^2\Big]\Big\}$$

$\big[$*proportionality follows because* $(\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0)/(m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1})$ *is constant*$\big]$

$$= \exp\Big\{ -\frac{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}{2}\ \Big(\mu_{\tau_A} - \frac{\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}\Big)^2\Big\},$$

which is proportional to the pdf of a normal distribution with mean $(\sigma_{\tau_A}^{-2}\ \tilde{\kappa}+a_1^{-1}\ a_0)/(m_A\ \sigma_{\tau_A}^{-2}+ a_1^{-1})$ and variance $1/(m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1})$. Therefore,

$$[\mu_{\tau_A}|.] \sim \mathcal{N}\Big(\frac{\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}},\ \frac{1}{m_A\ \sigma_{\tau_A}^{-2} + a_1^{-1}}\Big).$$

Full Conditional for Σ_β^{-1}

Picking out the terms in the joint density (4.5) which involve Σ_β, we get

$$\pi(\Sigma_\beta^{-1}|.) \propto \prod_{i=1}^{m}\mathcal{N}_3(\boldsymbol{\beta}_i|\ \boldsymbol{\mu}_\beta, \Sigma_\beta)\ \mathcal{W}(\Sigma_\beta^{-1}|\ \nu_1, (\nu_1\mathbb{A}_1)^{-1})$$

$$\propto \prod_{i=1}^{m}\frac{1}{|\Sigma_\beta|^{1/2}}\exp\Big\{ -\frac{1}{2}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)'\Sigma_\beta^{-1}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\Big\}|\Sigma_\beta^{-1}|^{\frac{\nu_1-3-1}{2}}\ \exp\Big\{ -\frac{\nu_1}{2}tr(\mathbb{A}_1\Sigma_\beta^{-1})\Big\}$$

$$= \frac{1}{|\Sigma_\beta|^{m/2}} \exp\Big\{ -\frac{1}{2} \sum_{i=1}^{m} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' \Sigma_\beta^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\Big\} |\Sigma_\beta^{-1}|^{\frac{\nu_1-3-1}{2}} \exp\Big\{ -\frac{\nu_1}{2} tr(\mathbb{A}_1\Sigma_\beta^{-1})\Big\}$$

$$= |\Sigma_\beta^{-1}|^{\frac{m+\nu_1-3-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[\sum_{i=1}^{m} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' \Sigma_\beta^{-1} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) + tr(\nu_1\mathbb{A}_1\Sigma_\beta^{-1})\Big]\Big\}$$

$$= |\Sigma_\beta^{-1}|^{\frac{(m+\nu_1)-3-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[\sum_{i=1}^{m} tr\big((\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' \Sigma_\beta^{-1}\big) + tr(\nu_1\mathbb{A}_1\Sigma_\beta^{-1})\Big]\Big\}$$

$\big[$*if we let* $\mathbf{d}_{1i} = (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' \Sigma_\beta^{-1}$ *and* $\mathbf{d}_{2i} = (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)$, *then* $\mathbf{d}_{1i}\mathbf{d}_{2i} = tr(\mathbf{d}_{i1}\mathbf{d}_{i2}) =$

$tr(\mathbf{d}_{2i}\mathbf{d}_{1i})$ *by the property of the trace of a matrix*$\big]$

$$= |\Sigma_\beta^{-1}|^{\frac{(m+\nu_1)-3-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[tr\Big(\sum_{i=1}^{m} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' \Sigma_\beta^{-1}\Big) + tr(\nu_1\mathbb{A}_1\Sigma_\beta^{-1})\Big]\Big\}$$

$$= |\Sigma_\beta^{-1}|^{\frac{(m+\nu_1)-3-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[tr\Big(\sum_{i=1}^{m} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_1\mathbb{A}_1\Big) \Sigma_\beta^{-1}\Big]\Big\},$$

which is proportional to a Wishart pdf with degrees of freedom $m + \nu_1$ and scale matrix $\big[\sum_{i=1}^{m} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_1\mathbb{A}_1\big]^{-1}$. Therefore,

$$[\Sigma_\beta^{-1}|.] \sim \mathcal{W}\Big(m + \nu_1, \Big[\sum_{i=1}^{m} (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta) (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_1\mathbb{A}_1\Big]^{-1}\Big).$$

Full Conditional for Σ_α^{-1}

Picking out the terms in the joint density (4.5) which involve Σ_α, we get

$$\pi(\Sigma_\alpha^{-1}|.) \propto \prod_{i=1}^{m} [\mathcal{LN}_2(\boldsymbol{\alpha}_i| \boldsymbol{\mu}_\alpha, \Sigma_\alpha)]^{I_i} \mathcal{W}(\Sigma_\alpha^{-1}| \nu_2, (\nu_1\mathbb{A}_2)^{-1})$$

$$\propto \prod_{i=1}^{m} \frac{1}{|\Sigma_\alpha|^{I_i/2}} \exp\Big\{ -\frac{I_i}{2}(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)'\Sigma_\alpha^{-1}(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\Big\} |\Sigma_\alpha^{-1}|^{\frac{\nu_2-2-1}{2}} \exp\Big\{ -\frac{\nu_2}{2} tr(\mathbb{A}_2\Sigma_\alpha^{-1})\Big\}$$

$$= \frac{1}{|\Sigma_\alpha|^{m_G/2}} \exp\Big\{ -\frac{1}{2} \sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' \Sigma_\alpha^{-1} (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\Big\}$$

$$|\Sigma_\alpha^{-1}|^{\frac{\nu_2-2-1}{2}} \exp\Big\{ -\frac{\nu_2}{2} tr(\mathbb{A}_2\Sigma_\alpha^{-1})\Big\}$$

$$= |\Sigma_\alpha^{-1}|^{\frac{m_G+\nu_2-2-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' \Sigma_\alpha^{-1} (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) + tr(\nu_2 \mathbb{A}_2 \Sigma_\alpha^{-1}) \Big] \Big\}$$

$$= |\Sigma_\alpha^{-1}|^{\frac{(m_G+\nu_2)-2-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[\sum_{i=1}^{m} tr\big(I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' \Sigma_\alpha^{-1}\big) + tr(\nu_2 \mathbb{A}_2 \Sigma_\alpha^{-1}) \Big] \Big\}$$

$\big[$*if we let* $\mathbf{d}_{1i} = (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' \Sigma_\alpha^{-1}$ *and* $\mathbf{d}_{2i} = I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)$, *then* $\mathbf{d}_{1i}\mathbf{d}_{2i} = tr(\mathbf{d}_{i1}\mathbf{d}_{i2}) =$

$tr(\mathbf{d}_{2i}\mathbf{d}_{1i})$ *by the property of the trace of a matrix*$\big]$

$$= |\Sigma_\alpha^{-1}|^{\frac{(m_G+\nu_2)-2-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[tr\Big(\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' \Sigma_\alpha^{-1} \Big) + tr(\nu_2 \mathbb{A}_2 \Sigma_\alpha^{-1}) \Big] \Big\}$$

$$= |\Sigma_\alpha^{-1}|^{\frac{(m_G+\nu_2)-2-1}{2}} \exp\Big\{ -\frac{1}{2} \Big[tr\Big(\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' + \nu_2 \mathbb{A}_2 \Big) \Sigma_\alpha^{-1} \Big] \Big\},$$

which is proportional to a Wishart pdf with degrees of freedom $m_G + \nu_2$ and scale matrix $\big[\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' + \nu_2 \mathbb{A}_2\big]^{-1}$. Therefore,

$$[\Sigma_\alpha^{-1}|.] \sim \mathcal{W}\Big(m_G + \nu_2, \ \Big[\sum_{i=1}^{m} I_i(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha) (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' + \nu_2 \mathbb{A}_2 \Big]^{-1} \Big).$$

Full Conditional for $\sigma_{\tau_A}^{-2}$

Picking out the terms in the joint density (4.5) which involve σ_{τ_A}, we get

$$\begin{aligned}
\pi(\sigma_{\tau_A}^{-2}|.) &\propto \prod_{i=1}^{m} [\mathcal{LN}(\tau_i|\ \mu_{\tau_A}, \sigma_{\tau_A}^2)]^{1-I_i}\ \mathcal{G}\Big(\sigma_{\tau_A}^{-2} \Big|\ \frac{b_0}{2}, \frac{b_1}{2}\Big) \\
&\propto \prod_{i=1}^{m} \frac{1}{(\sigma_{\tau_A}^2)^{(1-I_i)/2}} \exp\Big\{ -\frac{1-I_i}{2\sigma_{\tau_A}^2} (\kappa_i - \mu_{\tau_A})^2 \Big\}\ (\sigma_{\tau_A}^{-2})^{b_0/2-1} \exp\Big\{ -\frac{b_1}{2}\sigma_{\tau_A}^{-2} \Big\} \\
&= (\sigma_{\tau_A}^{-2})^{m_A/2} \exp\Big\{ -\frac{1}{2\sigma_{\tau_A}^2} \sum_{i=1}^{m} (1-I_i)(\kappa_i - \mu_{\tau_A})^2 \Big\}\ (\sigma_{\tau_A}^{-2})^{b_0/2-1} \exp\Big\{ -\frac{b_1}{2}\sigma_{\tau_A}^{-2} \Big\} \\
&= (\sigma_{\tau_A}^{-2})^{\frac{m_A+b_0}{2}-1} \exp\Big\{ -\frac{\sum_{i=1}^{m} (1-I_i)(\kappa_i - \mu_{\tau_A})^2 + b_1}{2}\ \sigma_{\tau_A}^{-2} \Big\}
\end{aligned}$$

which is proportional to a gamma pdf with shape and rate parameters $(m_A + b_0)/2$ and $\big\{\sum_{i=1}^{m} (1-I_i)(\kappa_i - \mu_{\tau_A})^2 + b_1\big\}/2$, respectively. Therefore,

$$[\sigma_{\tau_A}^{-2}|.] \sim \mathcal{G}\bigg(\frac{m_A + b_0}{2}, \frac{\sum_{i=1}^{m}(1-I_i)(\kappa_i - \mu_{\tau_A})^2 + b_1}{2}\bigg).$$

Full Conditional for σ_i^{-2}

Picking out the terms in the joint density (4.5) which involve σ_i, we get

$$\begin{aligned}
\pi(\sigma_i^{-2}|.) &\propto \mathcal{N}_{n_i-p}(\mathbf{y}_i^{(2)}|\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i)\ \ \mathcal{G}\Big(\sigma_i^{-2}\Big|\ \frac{d_0}{2},\ \frac{d_1}{2}\Big) \\
&\propto \frac{1}{|\sigma_i^2\ \mathbb{I}_i|^{1/2}}\ \exp\Big\{-\frac{1}{2\sigma_i^2}\ \big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)'\big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)\Big\}\ \ \mathcal{G}\Big(\sigma_{ui}^{-2}\Big|\ \frac{d_0}{2},\ \frac{d_1}{2}\Big) \\
&= \frac{1}{|\sigma_i^2\ \mathbb{I}_i|^{1/2}}\ \exp\Big\{-\frac{1}{2\sigma_i^2}\ \big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big)'\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big)\Big\}\ \ \mathcal{G}\Big(\sigma_{ui}^{-2}\Big|\ \frac{d_0}{2},\ \frac{d_1}{2}\Big) \\
&\propto \big(\sigma_i^{-2}\big)^{(n_i-p)/2}\ \exp\Big\{-\frac{1}{2\sigma_i^2}\big(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\big)'\big(\mathbf{z}_i - \mathbb{X}_i\boldsymbol{\beta}_i\big)\Big\}\big(\sigma_i^{-2}\big)^{d_0/2-1}\ \exp\big(-\frac{d_1}{2}\sigma_i^{-2}\big) \\
&= \big(\sigma_i^{-2}\big)^{\frac{n_i-p+d_0}{2}-1}\ \exp\Big\{-\frac{\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big)'\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big) + d_1}{2}\ \sigma_i^{-2}\Big\},
\end{aligned}$$

which is proportional to a gamma pdf with shape parameter $(n_i - p + d_0)/2$ and rate parameter $\big\{\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big)'\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big) + d_1\big\}/2$. Therefore,

$$[\sigma_i^{-2}|.] \sim \mathcal{G}\bigg(\frac{n_i - p + d_0}{2},\ \frac{\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big)'\big(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i\big) + d_1}{2}\bigg).$$

Full Conditional for $\boldsymbol{\phi}$

Picking out the terms in the joint density (4.5) which involve $\boldsymbol{\phi}$, we get

$$\begin{aligned}
\pi(\boldsymbol{\phi}|.) &\propto \prod_{i=1}^{m} \mathcal{N}_{n_i-p}(\mathbf{y}_i^{(2)}|\ \boldsymbol{\mu}_i,\ \sigma_i^2\ \mathbb{I}_i)\ \mathcal{N}_p(\boldsymbol{\phi}|\ \mathbf{h}_3,\ \mathbb{H}_3) \\
&\propto \prod_{i=1}^{m} \exp\Big\{-\frac{1}{2\sigma_i^2}\ \big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)'\ \big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)\Big\}\ \exp\Big\{-\frac{1}{2}\ (\boldsymbol{\phi} - \mathbf{h}_3)'\ \mathbb{H}_3^{-1}\ (\boldsymbol{\phi} - \mathbf{h}_3)\Big\}.
\end{aligned}$$

Here, for $j = p+1, p+2, \ldots, n_i$,

$$y_{ij} - \mu_{ij} = y_{ij} - \beta_{0i}\ \Big(1 - \sum_{k=1}^{p}\phi_k\Big) - \beta_{1i}\ x_{ij} - \beta_{2i}\ r_{ij} - \sum_{k=1}^{p}\phi_k\ y_{i,j-k}$$

$$= (y_{ij} - \beta_{0i} - \beta_{1i}\ t_{ij} - \beta_{2i}\ q_{ij}) - \phi_1\ (y_{i,j-1} - \beta_{0i} - \beta_{1i}\ t_{i,j-1} - \beta_{2i}\ q_{i,j-1}) -$$

$$\cdots - \phi_p\ (y_{i,j-p} - \beta_{0i} - \beta_{1i}\ t_{i,j-p} - \beta_{2i}\ q_{i,j-p})$$

$$= \epsilon_{ij} - \phi_1\ \epsilon_{i,j-1} - \phi_2\ \epsilon_{i,j-2} - \ \ \ldots\ \ - \phi_p\ \epsilon_{i,j-p}$$

which can expressed in vector-matrix notation as $\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i = \boldsymbol{\epsilon}_i - \mathbb{W}_i\boldsymbol{\phi}$. Therefore,

$$\pi(\boldsymbol{\phi}|.) \propto \prod_{i=1}^{m} \exp\Big\{ -\frac{1}{2\sigma_i^2}\ \big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)'\ \big(\mathbf{y}_i^{(2)} - \boldsymbol{\mu}_i\big)\Big\}\ \exp\Big\{ -\frac{1}{2}\ (\boldsymbol{\phi} - \mathbf{h}_3)'\ \mathbb{H}_3^{-1}\ (\boldsymbol{\phi} - \mathbf{h}_3)\Big\}$$

$$= \prod_{i=1}^{m} \exp\Big\{ -\frac{1}{2\sigma_i^2}\ \big(\boldsymbol{\epsilon}_i - \mathbb{W}_i\ \boldsymbol{\phi}\big)'\ \big(\boldsymbol{\epsilon}_i - \mathbb{W}_i\ \boldsymbol{\phi}\big)\Big\}\ \exp\Big\{ -\frac{1}{2}\ (\boldsymbol{\phi} - \mathbf{h}_3)'\ \mathbb{H}_3^{-1}\ (\boldsymbol{\phi} - \mathbf{h}_3)\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\ \Big[\sum_{i=1}^{m} \sigma_i^{-2}\ \big(\boldsymbol{\epsilon}_i - \mathbb{W}_i\ \boldsymbol{\phi}\big)'\ \big(\boldsymbol{\epsilon}_i - \mathbb{W}_i\ \boldsymbol{\phi}\big) + (\boldsymbol{\phi} - \mathbf{h}_3)'\ \mathbb{H}_3^{-1}\ (\boldsymbol{\phi} - \mathbf{h}_3)\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\ \Big[\sum_{i=1}^{m} \sigma_i^{-2}\ \big(\boldsymbol{\epsilon}_i'\ \boldsymbol{\epsilon}_i - \boldsymbol{\epsilon}_i'\ \mathbb{W}_i\ \boldsymbol{\phi} - \boldsymbol{\phi}'\ \mathbb{W}_i'\ \boldsymbol{\epsilon}_i + \boldsymbol{\phi}'\ \mathbb{W}_i'\ \mathbb{W}_i\ \boldsymbol{\phi}\big) +$$

$$\big(\boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \boldsymbol{\phi}\ - \boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \mathbf{h}_3\ - \mathbf{h}_3'\ \mathbb{H}_3^{-1}\ \boldsymbol{\phi} + \mathbf{h}_3'\ \mathbb{H}_3^{-1}\ \mathbf{h}_3\big)\Big]\Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\ \Big[\sum_{i=1}^{m} \sigma_i^{-2}\ \big(-2\ \boldsymbol{\phi}'\ \mathbb{W}_i'\ \boldsymbol{\epsilon}_i + \boldsymbol{\phi}'\ \mathbb{W}_i'\ \mathbb{W}_i\ \boldsymbol{\phi}\big) + \boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \boldsymbol{\phi} - 2\boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big]\Big\}$$

$\big[\boldsymbol{\epsilon}_i'\ \mathbb{W}_i\ \boldsymbol{\phi}$ *and* $\mathbf{h}_3'\ \mathbb{H}_3^{-1}\ \boldsymbol{\phi}$ *are sclars, and so are* $\boldsymbol{\epsilon}_i'\ \mathbb{W}_i\ \boldsymbol{\phi} = \boldsymbol{\phi}'\ \mathbb{W}_i'\ \boldsymbol{\epsilon}_i$ *and*

$\mathbf{h}_3'\ \mathbb{H}_3^{-1}\ \boldsymbol{\phi} = \boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \mathbf{h}_3$*; proportionality follows because* $\sigma_i^{-2}\ \boldsymbol{\epsilon}_i'\ \boldsymbol{\epsilon}_i$ *and* $\mathbf{h}_3'\ \mathbb{H}_3^{-1}\ \mathbf{h}_3$

do not depend on $\boldsymbol{\phi}\big]$

$$= \exp\Big\{ -\frac{1}{2}\ \Big[-2\ \boldsymbol{\phi}'\sum_{i=1}^{m} \sigma_i^{-2}\ \mathbb{W}_i'\ \boldsymbol{\epsilon}_i + \boldsymbol{\phi}'\Big(\sum_{i=1}^{m} \sigma_i^{-2}\ \mathbb{W}_i'\ \mathbb{W}_i\Big)\boldsymbol{\phi} +$$

$$\boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \boldsymbol{\phi} - 2\boldsymbol{\phi}'\ \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\ \Big[-2\ \boldsymbol{\phi}'\Big(\sum_{i=1}^{m} \sigma_i^{-2}\ \mathbb{W}_i'\ \boldsymbol{\epsilon}_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big) + \boldsymbol{\phi}'\Big(\sum_{i=1}^{m} \sigma_i^{-2}\ \mathbb{W}_i'\ \mathbb{W}_i + \mathbb{H}_3^{-1}\Big)\boldsymbol{\phi}\Big]\Big\}$$

$$= \exp\Big\{ -\frac{1}{2}\ \Big[-2\ \boldsymbol{\phi}'\Big(\sum_{i=1}^{m} \sigma_i^{-2}\ \mathbb{W}_i'\ \boldsymbol{\epsilon}_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big) + \boldsymbol{\phi}'\ \mathbb{V}^{-1}\ \boldsymbol{\phi}\Big]\Big\}$$

$$\propto \exp\Big\{ -\frac{1}{2}\Big[-\boldsymbol{\phi}'\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big) - \Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)'\boldsymbol{\phi} +$$

$$\boldsymbol{\phi}'\ \mathbb{V}^{-1}\ \boldsymbol{\phi} + \Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)'\ \mathbb{V}\ \Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)\Big]\Big\}$$

$$\Big[\boldsymbol{\phi}'\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big) = \Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)'\boldsymbol{\phi}\ \ \textit{because}$$

$$\boldsymbol{\phi}'\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)\ \textit{is a scalar; proportionality follows because}$$

$$\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)'\ \mathbb{V}\ \Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)\ \textit{does not depend on}\ \boldsymbol{\phi}\Big]$$

$$= \exp\Big\{ -\frac{1}{2}\Big[\boldsymbol{\phi} - \mathbb{V}\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)\Big]'\ \mathbb{V}^{-1}\ \Big[\boldsymbol{\phi} - \mathbb{V}\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big)\Big]\Big\},$$

which is proportional to the pdf of a p-variate normal distribution with mean vector $\mathbb{V}\big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\big)$ and covariance matrix $\mathbb{V}$. Therefore,

$$[\boldsymbol{\phi}|.] \sim \mathcal{N}_p\Big(\mathbb{V}\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big),\ \mathbb{V}\Big).$$

Full Conditional for ω

Picking out the terms in the joint density (4.5) which involve ω, we get

$$\begin{aligned}\pi(\omega|.) &\propto \prod_{i=1}^{m}\big\{\omega^{I_i}\ (1-\omega)^{1-I_i}\big\}\ \omega^{c_0-1}\ (1-\omega)^{c_1-1}\\ &= \omega^{\sum_{i=1}^{m} I_i}\ (1-\omega)^{\sum_{i=1}^{m}(1-I_i)}\ \omega^{c_0-1}\ (1-\omega)^{c_1-1}\\ &= \omega^{m_G+c_0-1}\ (1-\omega)^{m_A+c_1-1}\end{aligned}$$

which is proportional to the pdf of a beta distribution with parameters m_G+c_0 and m_A+c_1. Therefore

$$[\omega|.] \sim \mathcal{B}(m_G + c_0, m_A + c_1).$$

4.7.4 Metropolis Step for $\boldsymbol{\alpha_i}$

As shown in Section 4.7.3, the full conditional for $\boldsymbol{\alpha}_i$ can be expressed only up to a proportionality constant. Letting

$$L(\boldsymbol{\alpha}_i) = \exp\Big\{ -\frac{1}{2\sigma_i^2}(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)\Big\},$$

the full conditional for $\boldsymbol{\alpha}_i$ can be re-written as follows:

$$\pi(\boldsymbol{\alpha}_i|.) \propto L(\boldsymbol{\alpha}_i) \times \big[\mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A})\big]^{1-I_i} \times \big[\mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha)\big]^{I_i}. \tag{4.8}$$

Generating a candidate point $\boldsymbol{\alpha}_i^* = (\gamma_i^*, \tau_i^*)'$ involves two steps (Johnson et al. [42]): (1) draw a candidate I_i^* from a Bernoulli distribution with parameter ω; and (2)(a) if $I_i^* = 0$, set $\gamma_i^* = 0$ and draw τ_i^* from the proposal density, but (2)(b) if $I_i^* = 1$, draw $\boldsymbol{\alpha}_i^*$ from the proposal density. Our choices of the proposal densities depend on two cases, and are described next.

Let $\big(\boldsymbol{\alpha}_i^{(s)}, I_i^{(s)}\big)$ be the draw at the s^{th} MCMC iteration. We consider two variants of the Metropolis-Hastings algorithm to update $\big(\boldsymbol{\alpha}_i^{(s)}, I_i^{(s)}\big)$: an independence Metropolis-Hastings step when $I_i^{(s)} = 1$ and $I_i^* = 0$ or vice versa, i.e., when there is a jump from one population to another; and a random-walk Metropolis step when $I_i^{(s)} = I_i^* = 0$ (or 1), i.e., when the proposed draw remains in the same population. The algorithmic procedure is described as follows.

Case I: $I_i^{(s)} = 0$ *and* $I_i^* = 1$ *or vice versa*

We take the delta bivariate mixed lognormal distribution $\Delta(\boldsymbol{\mu}_\alpha, \Sigma_\alpha, \mu_{\tau_A}, \sigma^2_{\tau_A}, \omega)$ as our proposal density with pdf

$$g(\boldsymbol{\alpha}_i) = \mathbf{1}(\gamma_i = 0)\ (1-\omega)\ \mathcal{LN}(\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}) + \mathbf{1}(\gamma_i > 0)\ \omega\ \mathcal{LN}_2(\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha),$$

where $\boldsymbol{\mu}_\alpha$, Σ_α, μ_{τ_A}, $\sigma^2_{\tau_A}$, ω are at their current states. Note that drawing a random point from $\Delta(.)$ involves two steps: draw $I_i^* \sim \mathcal{BER}(\omega)$ first, and then set $\gamma_i^* = 0$ and draw $\tau_i^* \sim \mathcal{LN}(\mu_{\tau_A}, \sigma^2_{\tau_A})$ if $I_i^* = 0$, but draw $\boldsymbol{\alpha}_i^* \sim \mathcal{LN}_2(\boldsymbol{\mu}_\alpha, \Sigma_\alpha)$ if $I_i^* = 1$.

First, consider the case $I^{(s)} = 0$ and $I^* = 1$. The Metropolis-Hastings ratio for the

independence Metropolis-Hastings step is

$$R_{01} = \frac{\left\{L\left(\boldsymbol{\alpha}_i^*\right) \times \mathcal{LN}_2\left(\boldsymbol{\alpha}_i^* | \boldsymbol{\mu}_\alpha, \Sigma_\alpha\right)\right\} \times \left\{(1-\omega) \times \mathcal{LN}\left(\tau_i^{(s)} | \mu_{\tau_A}, \sigma^2_{\tau_A}\right)\right\}}{\left\{L\left(\boldsymbol{\alpha}_i^{(s)}\right) \times \mathcal{LN}\left(\tau_i^{(s)} | \mu_{\tau_A}, \sigma^2_{\tau_A}\right)\right\} \times \left\{\omega \times \mathcal{LN}_2\left(\boldsymbol{\alpha}_i^* | \boldsymbol{\mu}_\alpha, \Sigma_\alpha\right)\right\}}$$
$$= \frac{L\left(\boldsymbol{\alpha}_i^*\right)}{L\left(\boldsymbol{\alpha}_i^{(s)}\right)} \frac{1-\omega}{\omega}.$$

Next, consider $I^{(s)} = 1$ and $I^* = 0$. The Metropolis-Hastings ratio for this case is

$$R_{10} = \frac{\left\{L\left(\boldsymbol{\alpha}_i^*\right) \times \mathcal{LN}\left(\tau_i^* | \mu_{\tau_A}, \sigma^2_{\tau_A}\right)\right\} \times \left\{\omega \times \mathcal{LN}_2\left(\boldsymbol{\alpha}_i^{(s)} | \boldsymbol{\mu}_\alpha, \Sigma_\alpha\right)\right\}}{\left\{L\left(\boldsymbol{\alpha}_i^{(s)}\right) \times \mathcal{LN}_2\left(\boldsymbol{\alpha}_i^{(s)} | \boldsymbol{\mu}_\alpha, \Sigma_\alpha\right)\right\} \times \left\{(1-\omega) \times \mathcal{LN}\left(\tau_i^* | \mu_{\tau_A}, \sigma^2_{\tau_A}\right)\right\}}$$
$$= \frac{L\left(\boldsymbol{\alpha}_i^*\right)}{L\left(\boldsymbol{\alpha}_i^{(s)}\right)} \frac{\omega}{1-\omega}.$$

Case II: $I_i^{(s)} = I_i^* = 0$ *or* 1

For this case, we choose a random-walk Metropolis with proposal densities $\mathcal{N}\left(\tau_i^{(s)}, c\ \sigma^2_{\tau_A}\right)$ and $\mathcal{N}_2\left(\boldsymbol{\alpha}_i^{(s)}, c\ \Sigma_\alpha\right)$ for $I_i^{(s)} = I_i^* = 0$ and $I_i^{(s)} = I_i^* = 1$, respectively, where $\sigma^2_{\tau_A}$ and Σ_α are at their current states in the MCMC simulation, and c is called a *tuning* parameter which is adjusted to a value in such a way to achieve a reasonable acceptance probability (may need to do so by trial-and-error). Metropolis-Hastings ratios for these two cases are, respectively,

$$R_{00} = \frac{L\left(\boldsymbol{\alpha}_i^*\right) \times \mathcal{LN}\left(\tau_i^* | \mu_{\tau_A}, \sigma^2_{\tau_A}\right)}{L\left(\boldsymbol{\alpha}_i^{(s)}\right) \times \mathcal{LN}\left(\tau_i^{(s)} | \mu_{\tau_A}, \sigma^2_{\tau_A}\right)}$$

and

$$R_{11} = \frac{L\left(\boldsymbol{\alpha}_i^*\right) \times \mathcal{LN}_2\left(\boldsymbol{\alpha}_i^* | \boldsymbol{\mu}_\alpha, \Sigma_\alpha\right)}{L\left(\boldsymbol{\alpha}_i^{(s)}\right) \times \mathcal{LN}_2\left(\boldsymbol{\alpha}_i^{(s)} | \boldsymbol{\mu}_\alpha, \Sigma_\alpha\right)}.$$

Now, For both **Case I** and **Case II**, if the candidate is accepted, we take $\boldsymbol{\alpha}_i^{(s+1)} = \boldsymbol{\alpha}_i^*$ and $I_i^{(s+1)} = I_i^*$, otherwise $\boldsymbol{\alpha}_i^{(s+1)} = \boldsymbol{\alpha}_i^{(s)}$ and $I_i^{(s+1)} = I_i^{(s)}$.

Chapter 5

Special Cases of the Mixed Bent-Cable Model

An AR(0) process for the within-individual noise is a special case of the methodology described in Chapters 3 and 4, where $p = 0$ and $\boldsymbol{\phi}$ is degenerate. Statistical inference for this model can be carried out following the same lines as for the general mixed bent-cable model: the full conditionals take the same form as given in Section 4.7.3, but with $p = 0$ and $\boldsymbol{\phi}$ degenerate so that $\mathbf{z}_i = \mathbf{y}_i$, $\mathbf{x}_i = \mathbf{t}_i$ and $\mathbf{r}_i = \mathbf{q}_i$; and omitting Step 11 in the updating order of constructing a Markov chain described in Section 4.3.3. There are two other special cases irrespective of the value of p: a model for an abrupt transition (Population A) only, and another for a gradual transition (Population G) only. Henceforth, we will refer to these two by Model A and Model G, respectively. The underlying assumption for each of these two models is that the sample comes only from that population, and therefore the mixture probability ω is degenerate. It is easy to reduce the mixed bent-cable model to involve only one type of transition. Moreover, the statistical inference for these two models can be carried out following the same lines as for the general model (Chapter 4). We describe the models and inference techniques for Models A and G in Sections 5.1 and 5.2, respectively. We conclude this chapter with a discussion in Section 5.3, and the derivations of the full conditionals in the chapter appendix (Section 5.4).

5.1 Model A

For Model A, the transition parameter $\boldsymbol{\alpha}_i = (\gamma_i, \tau_i)$ has $\gamma_i = 0$ for all i, and the bent cable function for the i^{th} individual is given by Equations (3.2) and (3.4). Though Level 1 and modeling $\boldsymbol{\beta}_i$ at Level 2 are as in Section 3.1, we require only a lognormal distribution to model τ_i in this case since $\gamma_i = 0$. Therefore, $\boldsymbol{\mu}_\alpha$ and Σ_α are degenerate as these are associated only with gradual transition for which $\gamma_i > 0$. The three levels of the hierarchy from Section 3.1.4 are

$$[\mathbf{y}_i^{(2)}|\ \mathbf{y}_i^{(1)}, \boldsymbol{\theta}_i, \boldsymbol{\phi}, \sigma_i^2] \sim \mathcal{N}_{n_i-p}(\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i), \tag{5.1}$$

$$\left.\begin{array}{c} \big[\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta\big] \sim \mathcal{N}_3(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\ \big[\tau_i|\mu_{\tau_A}, \sigma^2_{\tau_A}\big] \sim \mathcal{LN}(\mu_{\tau_A}, \sigma^2_{\tau_A}), \end{array}\right\}, \tag{5.2}$$

$$\left.\begin{array}{c} [\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1] \sim \mathcal{N}_3(\mathbf{h}_1, \mathbb{H}_1),\ \ [\mu_{\tau_A}|a_0, a_1] \sim \mathcal{N}(a_0, a_1), \\ \big[\Sigma_\beta^{-1}|\nu_1, \mathbb{A}_1\big] \sim \mathcal{W}\big(\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big),\ \ \big[\sigma_{\tau_A}^{-2}|b_0, b_1\big] \sim \mathcal{G}(\tfrac{b_0}{2}, \tfrac{b_1}{2}) \\ [\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3] \sim \mathcal{N}_p(\mathbf{h}_3, \mathbb{H}_3),\ \ \big[\sigma_i^{-2}|d_0, d_1\big] \sim \mathcal{G}(\tfrac{d_0}{2}, \tfrac{d_1}{2}) \end{array}\right\}, \tag{5.3}$$

where the first two levels are (5.1) and (5.2), and the third level is (5.3) with the hyperparameters $\mathbf{h}_1$, $\mathbb{H}_1$, a_0, a_1, ν_1, $\mathbb{A}_1$, b_0, b_1, $\mathbf{h}_3$, $\mathbb{H}_3$, d_0 and d_1.

Letting $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_m)'$, we denote the parameters for Model A by $\Theta_A = \big(\boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\mu}_\beta, \mu_{\tau_A}, \Sigma_\beta^{-1}, \sigma_{\tau_A}^{-2}, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big)$. The full conditionals to implement the Metropolis within Gibbs algorithm can be derived following the same lines as for the general mixed bent-cable model, and are given in the chapter appendix (Section 5.4). Now, given $\Theta_A^{(s)}$ at iteration s, we update using the following sequence to achieve the new set $\Theta_A^{(s+1)}$ in one iteration:

1. For $i = 1$, generate $\boldsymbol{\beta}_i^{(s+1)} \sim \pi\big(\boldsymbol{\beta}_i|\ \tau_i^{(s)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\beta^{(s)}, \Sigma_\beta^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\beta}_i|\ \tau_i^{(s)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\beta^{(s)}, \Sigma_\beta^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ is a trivariate normal distribution;

2. for $i = 1$, generate $\tau_i^{(s+1)} \sim \pi\big(\tau_i|\ \boldsymbol{\beta}_i^{(s+1)}, \sigma_i^{-2(s)}, \mu_{\tau_A}^{(s)}, \sigma_{\tau_A}^{-2(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a random-walk Metropolis step with the proposal distribution $\mathcal{N}\big(\tau_i^{(s)}, c\ \sigma_{\tau_A}^{2(s)}\big)$ with c being the tuning parameter, where $\pi\big(\tau_i|\ \boldsymbol{\beta}_i^{(s+1)}, \sigma_i^{-2(s)}, \mu_{\tau_A}^{(s)}, \sigma_{\tau_A}^{-2(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ can be expressed only up to a proportionality constant;

3. for $i = 1$, generate $\sigma_i^{-2(s+1)} \sim \pi\big(\sigma_i^{-2} |\ \boldsymbol{\beta}_i^{(s+1)}, \tau_i^{(s+1)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a Gibbs step, where $\pi\big(\sigma_i^{-2} |\ \boldsymbol{\beta}_i^{(s+1)}, \tau_i^{(s+1)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ is a gamma distribution;

4. repeat 1–3 for $i = 2, 3, \ldots, m$; these complete the update for the individual-specific parameters, that is, we now have $\boldsymbol{\beta}^{(s+1)}$, $\boldsymbol{\tau}^{(s+1)}$ and $\boldsymbol{\sigma}^{-2(s+1)}$;

5. generate $\Sigma_\beta^{-1(s+1)} \sim \pi\big(\Sigma_\beta^{-1} |\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}_\beta^{(s)}\big)$ via a Gibbs step, where $\pi\big(\Sigma_\beta^{-1} |\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}_\beta^{(s)}\big)$ is a Wishart distribution;

6. generate $\sigma_{\tau_A}^{-2(s+1)} \sim \pi\big(\sigma_{\tau_A}^{-2} |\ \boldsymbol{\tau}^{(s+1)}, \mu_{\tau_A}^{(s)}\big)$ via a Gibbs step, where $\pi\big(\sigma_{\tau_A}^{-2} |\ \boldsymbol{\tau}^{(s+1)}, \mu_{\tau_A}^{(s)}\big)$ is a gamma distribution;

7. generate $\boldsymbol{\mu}_\beta^{(s+1)} \sim \pi\big(\boldsymbol{\mu}_\beta |\ \boldsymbol{\beta}^{(s+1)}, \Sigma_\beta^{-1(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\mu}_\beta |\ \boldsymbol{\beta}^{(s+1)}, \Sigma_\beta^{-1(s+1)}\big)$ is a trivariate normal distribution;

8. generate $\mu_{\tau_A}^{(s+1)} \sim \pi\big(\mu_{\tau_A} |\ \boldsymbol{\tau}^{(s+1)}, \sigma_{\tau_A}^{-2(s+1)}\big)$ via a Gibbs step, where $\pi\big(\mu_{\tau_A} |\ \boldsymbol{\tau}^{(s+1)}, \sigma_{\tau_A}^{-2(s+1)}\big)$ is a normal distribution;

9. generate $\boldsymbol{\phi}^{(s+1)} \sim \pi\big(\boldsymbol{\phi} |\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\tau}^{(s+1)}, \boldsymbol{\sigma}^{-2(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\phi} |\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\tau}^{(s+1)}, \boldsymbol{\sigma}^{-2(s+1)}\big)$ is a p-variate normal distribution.

The Bayesian inference can now be made by applying Monte Carlo integration on the Markov chain constructed via the above steps, and the same diagnostics discussed in Section 4.4 can be employed to check the convergence of the Markov chain.

5.2 Model G

For Model G, the transition parameter $\boldsymbol{\alpha}_i = (\gamma_i, \tau_i)$ has $\gamma_i > 0$ for all i, and the bent cable function for the i^{th} individual is given by Equations (3.2) and (3.3). Like Model A, Level 1 and modeling $\boldsymbol{\beta}_i$ at Level 2 are as before (Section 3.1). However, we require a bivariate lognormal distribution to model $\boldsymbol{\alpha}_i$ in this case. Note that μ_{τ_A} and $\sigma_{\tau_A}^2$ are degenerate as

these are associated only with an abrupt transition for which $\gamma_i = 0$. The three levels of the hierarchy are

$$[\mathbf{y}_i^{(2)}|\ \mathbf{y}_i^{(1)}, \boldsymbol{\theta}_i, \boldsymbol{\phi}, \sigma_i^2] \sim \mathcal{N}_{n_i-p}(\boldsymbol{\mu}_i, \sigma_i^2\ \mathbb{I}_i), \tag{5.4}$$

$$\left.\begin{array}{c} \big[\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta\big] \sim \mathcal{N}_3(\boldsymbol{\mu}_\beta, \Sigma_\beta), \\ [\boldsymbol{\alpha}_i|\boldsymbol{\mu}_\alpha, \Sigma_\alpha] \sim \mathcal{LN}_2(\boldsymbol{\mu}_\alpha, \Sigma_\alpha), \end{array}\right\}, \tag{5.5}$$

$$\left.\begin{array}{c} [\boldsymbol{\mu}_\beta|\mathbf{h}_1, \mathbb{H}_1] \sim \mathcal{N}_3(\mathbf{h}_1, \mathbb{H}_1),\ \ [\boldsymbol{\mu}_\alpha|\mathbf{h}_2, \mathbb{H}_2] \sim \mathcal{N}_2(\mathbf{h}_2, \mathbb{H}_2), \\ \big[\Sigma_\beta^{-1}|\nu_1, \mathbb{A}_1\big] \sim \mathcal{W}\big(\nu_1, (\nu_1\mathbb{A}_1)^{-1}\big),\ \ [\Sigma_\alpha^{-1}|\nu_2, \mathbb{A}_2] \sim \mathcal{W}\big(\nu_2, (\nu_2\mathbb{A}_2)^{-1}\big) \\ [\boldsymbol{\phi}|\mathbf{h}_3, \mathbb{H}_3] \sim \mathcal{N}_p(\mathbf{h}_3, \mathbb{H}_3),\ \ \big[\sigma_i^{-2}|d_0, d_1\big] \sim \mathcal{G}(\frac{d_0}{2}, \frac{d_1}{2}) \end{array}\right\}, \tag{5.6}$$

where the first two levels are (5.4) and (5.5), and the third level is (5.6) with the hyperparameters $\mathbf{h}_1$, $\mathbb{H}_1$, $\mathbf{h}_2$, $\mathbb{H}_2$, ν_1, $\mathbb{A}_1$, ν_2, $\mathbb{A}_2$, $\mathbf{h}_3$, $\mathbb{H}_3$, d_0 and d_1.

We denote the parameters for Model G by $\Theta_G = \big(\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\mu}_\beta, \boldsymbol{\mu}_\alpha, \Sigma_\beta^{-1}, \Sigma_\alpha^{-1}, \boldsymbol{\sigma}^{-2}, \boldsymbol{\phi}\big)$, and their full conditionals are given in the appendix (Section 5.4). Now, given $\Theta_G^{(s)}$ at iteration s, we use the following sequence to achieve the new set $\Theta_G^{(s+1)}$ in one iteration:

1. For $i = 1$, generate $\boldsymbol{\beta}_i^{(s+1)} \sim \pi\big(\boldsymbol{\beta}_i|\ \boldsymbol{\alpha}_i^{(s)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\beta^{(s)}, \Sigma_\beta^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\beta}_i|\ \boldsymbol{\alpha}_i^{(s)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\beta^{(s)}, \Sigma_\beta^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ is a trivariate normal distribution;

2. for $i = 1$, generate $\boldsymbol{\alpha}_i^{(s+1)} \sim \pi\big(\boldsymbol{\alpha}_i|\ \boldsymbol{\beta}_i^{(s+1)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\alpha^{(s)}, \Sigma_\alpha^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a random-walk Metropolis step with the proposal distribution $\mathcal{MVN}\big(\boldsymbol{\alpha}_i^{(s)}, c\ \Sigma_\alpha^{(s)}\big)$ with c being the tuning parameter, where $\pi\big(\boldsymbol{\alpha}_i|\ \boldsymbol{\beta}_i^{(s+1)}, \sigma_i^{-2(s)}, \boldsymbol{\mu}_\alpha^{(s)}, \Sigma_\alpha^{-1(s)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ can be expressed only up to a proportionality constant;

3. for $i = 1$, generate $\sigma_i^{-2(s+1)} \sim \pi\big(\sigma_i^{-2}|\ \boldsymbol{\beta}_i^{(s+1)}, \boldsymbol{\alpha}_i^{(s+1)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ via a Gibbs step, where $\pi\big(\sigma_i^{-2}|\ \boldsymbol{\beta}_i^{(s+1)}, \boldsymbol{\alpha}_i^{(s+1)}, \boldsymbol{\phi}^{(s)}, \mathbf{y}_i^{(2)}\big)$ is a gamma distribution;

4. repeat 1–3 for $i = 2, 3, \ldots, m$; these complete the update for the individual-specific parameters, that is, we now have $\boldsymbol{\beta}^{(s+1)}$, $\boldsymbol{\alpha}^{(s+1)}$ and $\boldsymbol{\sigma}^{-2(s+1)}$;

5. generate $\Sigma_\beta^{-1(s+1)} \sim \pi\big(\Sigma_\beta^{-1}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}_\beta^{(s)}\big)$ via a Gibbs step, where $\pi\big(\Sigma_\beta^{-1}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}_\beta^{(s)}\big)$ is a Wishart distribution;

6. generate $\Sigma_\alpha^{-1(s+1)} \sim \pi\big(\Sigma_\alpha^{-1}|\ \boldsymbol{\alpha}^{(s+1)}, \boldsymbol{\mu}_\alpha^{(s)}\big)$ $\big($and hence $\Sigma_\alpha^{(s+1)}\big)$ via a Gibbs step, where $\pi\big(\Sigma_\alpha^{-1}|\ \boldsymbol{\alpha}^{(s+1)}, \boldsymbol{\mu}_\alpha^{(s)}\big)$ is a Wishart distribution;

7. generate $\boldsymbol{\mu}_\beta^{(s+1)} \sim \pi\big(\boldsymbol{\mu}_\beta|\ \boldsymbol{\beta}^{(s+1)}, \Sigma_\beta^{-1(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\mu}_\beta|\ \boldsymbol{\beta}^{(s+1)}, \Sigma_\beta^{-1(s+1)}\big)$ is a trivariate normal distribution;

8. generate $\boldsymbol{\mu}_\alpha^{(s+1)} \sim \pi\big(\boldsymbol{\mu}_\alpha|\ \boldsymbol{\alpha}^{(s+1)}, \Sigma_\alpha^{-1(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\mu}_\alpha|\ \boldsymbol{\alpha}^{(s+1)}, \Sigma_\alpha^{-1(s+1)}\big)$ is a bivariate normal distribution;

9. generate $\boldsymbol{\phi}^{(s+1)} \sim \pi\big(\boldsymbol{\phi}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\alpha}^{(s+1)}, \boldsymbol{\sigma}^{-2(s+1)}\big)$ via a Gibbs step, where $\pi\big(\boldsymbol{\phi}|\ \boldsymbol{\beta}^{(s+1)}, \boldsymbol{\alpha}^{(s+1)}, \boldsymbol{\sigma}^{-2(s+1)}\big)$ is a p-variate normal distribution;

The Bayesian inference can now be carried out using the techniques described in Chapter 4.

5.3 Discussion

In this chapter, we have presented the special cases of the mixed bent-cable regression. We recommend using the flexible methodology (Chapters 3 and 4), with $0 < \omega < 1$, only when there is a strong reason to believe that the sample potentially comes from two populations (Populations A and G), a scenario that exists for the rat study (Chapters 1 and 6); otherwise, depending on the context, either Model A or G (ω is 0 or 1, respectively) should be used. For example, since the CFC-11 profiles exhibit gradual transition all over the globe (Chapter 1), Model G should be the preferred method to analyze this data set.

5.4 Chapter Appendix: Full Conditionals

Recall that $I_i = 0$ and 1 (for all i) for Models A and G, respectively, and that $f_{ij} = \beta_{0i} + \beta_{1i}t_{ij} + \beta_{2i}q_{ij}$ is the bent-cable function, with $q_{ij} = (t_{ij} - \tau_i)\mathbf{1}\{t_{ij} - \tau_i > 0\}$ for Model A, and $q_{ij} = (t_{ij} - \tau_i + \gamma_i)^2\ \mathbf{1}\{|t_{ij} - \tau_i| \leq \gamma_i\}/4\gamma_i + (t_{ij} - \tau_i)\mathbf{1}\{t_{ij} - \tau_i > \gamma_i\}$ for Model G. Also recall our notation from Page 51. Note that $\tilde{\boldsymbol{\xi}}$ and $\mathbb{U}_2^{-1}$ are associated with Model G only for which $I_i = 1$ for all i, so that $m_G = \sum_{i=1}^m I_i = m$, $\mathbb{U}_2^{-1} = m_G\ \Sigma_\alpha^{-1} + \mathbb{H}_2^{-1} = m\ \Sigma_\alpha^{-1} + \mathbb{H}_2^{-1}$ and $\tilde{\boldsymbol{\xi}} = \sum_{i=1}^m I_i\ \boldsymbol{\xi}_i = \sum_{i=1}^m \boldsymbol{\xi}_i$. Also, since $\tilde{\kappa}$ is associated with Model A only for which $I_i = 0$ for all i, $\tilde{\kappa} = \sum_{i=1}^m (1 - I_i)\ \kappa_i = \sum_{i=1}^m \kappa_i$

Now, following the same lines as for the general mixed bent-cable regression in Section 4.7.3, one can verify that the full conditionals for Model A are

$$[\boldsymbol{\beta}_i|.] \sim \mathcal{N}_3\Big(\mathbb{M}_i\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big),\ \mathbb{M}_i\Big),$$

$$\pi(\tau_i|.) \propto \exp\Big\{-\frac{1}{2\sigma_i^2}(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)\Big\} \times \Big[\frac{1}{\tau_i}\exp\Big\{-\frac{1}{2\sigma_{\tau_A}^2}(\kappa_i - \mu_{\tau_A})^2\Big\}\Big]$$

$$[\boldsymbol{\mu}_\beta|.] \sim \mathcal{N}_3\Big(\mathbb{U}_1\ \big(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1\big),\ \mathbb{U}_1\Big),$$

$$[\mu_{\tau_A}|.] \sim \mathcal{N}\Big(\frac{\sigma_{\tau_A}^{-2}\ \tilde{\kappa} + a_1^{-1}\ a_0}{m\ \sigma_{\tau_A}^{-2} + a_1^{-1}},\ \frac{1}{m\ \sigma_{\tau_A}^{-2} + a_1^{-1}}\Big),$$

$$[\Sigma_\beta^{-1}|.] \sim \mathcal{W}\Big(m + \nu_1,\ \Big[\sum_{i=1}^m (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\ (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_1\mathbb{A}_1\Big]^{-1}\Big),$$

$$[\sigma_{\tau_A}^{-2}|.] \sim \mathcal{G}\Big(\frac{m_A + b_0}{2}, \frac{\sum_{i=1}^m (\kappa_i - \mu_{\tau_A})^2 + b_1}{2}\Big),$$

$$[\sigma_i^{-2}|.] \sim \mathcal{G}\Big(\frac{n_i - p + d_0}{2},\ \frac{(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i) + d_1}{2}\Big),$$

$$[\boldsymbol{\phi}|.] \sim \mathcal{N}_p\Big(\mathbb{V}\Big(\sum_{i=1}^m \sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big),\ \mathbb{V}\Big),$$

and those for Model G are

$$[\boldsymbol{\beta}_i|.] \sim \mathcal{N}_3\Big(\mathbb{M}_i\ \big(\sigma_i^{-2}\ \mathbb{X}_i'\ \mathbf{z}_i + \Sigma_\beta^{-1}\ \boldsymbol{\mu}_\beta\big),\ \mathbb{M}_i\Big),$$

$$\pi(\boldsymbol{\alpha}_i|.) \propto \exp\Big\{ -\frac{1}{2\sigma_i^2}(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)\Big\} \times \Big[\frac{1}{\gamma_i\tau_i}\exp\Big\{ -\frac{1}{2}(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)'\ \Sigma_\alpha^{-1}\ (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\Big\}\Big],$$

$$[\boldsymbol{\mu}_\beta|.] \sim \mathcal{N}_3\Big(\mathbb{U}_1\ \big(\Sigma_\beta^{-1}\ \tilde{\boldsymbol{\beta}} + \mathbb{H}_1^{-1}\ \mathbf{h}_1\big),\ \mathbb{U}_1\Big),$$

$$[\boldsymbol{\mu}_\alpha|.] \sim \mathcal{N}_2\Big(\mathbb{U}_2\ \big(\Sigma_\alpha^{-1}\ \tilde{\boldsymbol{\xi}} + \mathbb{H}_2^{-1}\ \mathbf{h}_2\big),\ \mathbb{U}_2\Big),$$

$$[\Sigma_\beta^{-1}|.] \sim \mathcal{W}\Big(m + \nu_1,\ \Big[\sum_{i=1}^{m}(\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)\ (\boldsymbol{\beta}_i - \boldsymbol{\mu}_\beta)' + \nu_1\mathbb{A}_1\Big]^{-1}\Big),$$

$$[\Sigma_\alpha^{-1}|.] \sim \mathcal{W}\Big(m + \nu_2,\ \Big[\sum_{i=1}^{m}(\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)\ (\boldsymbol{\xi}_i - \boldsymbol{\mu}_\alpha)' + \nu_2\mathbb{A}_2\Big]^{-1}\Big),$$

$$[\sigma_i^{-2}|.] \sim \mathcal{G}\Big(\frac{n_i - p + d_0}{2},\ \frac{(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i)'(\mathbf{z}_i - \mathbb{X}_i\ \boldsymbol{\beta}_i) + d_1}{2}\Big),$$

$$[\boldsymbol{\phi}|.] \sim \mathcal{N}_p\Big(\mathbb{V}\Big(\sum_{i=1}^{m}\sigma_i^{-2}\ \mathbb{W}_i'\ \epsilon_i + \mathbb{H}_3^{-1}\ \mathbf{h}_3\Big),\ \mathbb{V}\Big).$$

Chapter 6

Data Analyses

Two applications of our methodology are presented in this chapter. In the first example (Section 6.1), we apply our flexible mixed bent-cable methodology to data from an experimental rat model. The second example (Section 6.2) uses Model G to address the global concern of atmospheric concentration of CFC-11. Any parameter estimate in the following two examples is based on the posterior mean or median, depending on the extent of asymmetry of the corresponding marginal posterior density; a fitted curve is then produced based on the instances of the regression coefficients in the Markov chain as described in Section 4.2.1. For example, since the CTP for Population G is $\exp\{\mu_\tau\} - \exp\{\mu_\gamma\} - 2\mu_1 \exp\{\mu_\gamma\}/\mu_2$ (see Section 4.2.1), we use the mean from the posterior of this expression for inference. Estimates for the other parameters for Level 2 Population A/G medians/standard deviations are produced similarly. In both the examples, we choose flat priors (see Section 3.1.3). We conclude this chapter with a few general remarks about the performance of our methodology in Section 6.3.

6.1 Rat Data

6.1.1 Background of the Study

Recall from Chapter 1 that hypothermia is a fatal condition which can occur when core body temperature (T_c) falls below 35^oC. Recall also that hypothermia is used as a therapeutic tool for cardiac arrest, stroke and brain injury. By reducing a patient's T_c, systematic

metabolism, and therefore tissue oxygen demand, can be lowered. This is believed to attenuate the effects of ischemia at the cellular level (Gordon [38]). In contrast, when the body becomes very cold, all physiological systems begin to slow down, eventually to the point that threatens survival. Treatment priorities in such situations include prevention of further cooling and resuscitation. Therefore, as mentioned in Chapter 1, quantification of the transition of T_c to early hypothermia is of great clinical interest.

6.1.2 Data

Motivated by the above, 38 approximately 8-week-old male Long-Evans rats were used in an experiment conducted by Reynolds et al. [62]. Rats were anesthetized, and their core temperature T_c was logged by remote data collection every 15 seconds for the duration of the trial. We denote time for subject i by t_{ij}, $j = 1, 2, \ldots, n_i$, where $t_{i1} = 0$ refers to the starting point of the study for rat i, and each subsequent time increment is 15 seconds. Eight representative T_c profiles were shown in Figure 1.1. There we pointed out that in addition to roughly linear incoming and outgoing phases at either end of each profile, some rats may exhibit a gradual transition in T_c, while others, seemingly an abrupt transition. That is, we have samples potentially coming from two different populations, G (gradual) and A (abrupt), respectively, according to the type of transition for the underlying T_c trend. Therefore, we analyze this data set using our flexible mixed bent-cable methodology.

Note that there are 4 rat profiles that exhibit neither a gradual nor an abrupt transition, but rather a linearly decreasing trend. Extension of our methodology to incorporate those profiles as coming from another population is straightforward (see Section 8.2). However, since only 4 profiles are not expected to provide enough information to make inference about the underlying population, taking into account that population in analyzing the rat data may not be worthwhile. Also note that those 4 profiles are not expected to create any problem in model fitting because of shrinkage towards the population. However, since they must belong to one of the two populations A and G by Assumption **T1** (Section 3.1.2), the mixture probability ω is no longer the actual proportion of profiles that belong to G under the assumption of the existence of a third population. Nonetheless, 4 profiles only are not expected to contribute anything significant to the estimation, and hence, interpretation of the actual ω.

6.1.3 Results

Reynolds and Chiu [63] use broken sticks and bent cables to model the data, treating each rat profile as an individual time series. Here, we unify the inference from all 38 rats with our more general mixed bent-cable methodology incorporating random effects. Like Reynolds and Chiu, we consider data from the start of hemorrhage until resuscitation intervention. Note that our longitudinal context allows pooling of information from multiple individuals, overcoming apparent violation of linearity of the incoming and outgoing phases that prevented model convergence for certain rat profiles in the case of Reynolds and Chiu, who truncate "problematic" profiles to achieve convergence.

We construct two Markov chains each of 5,000,000 iterations to approximate the posterior density. The initial 200,000 iterations are discarded as burn-in, and the inferences are based on every 200^{th} iteration of the chains (thinning), resulting in a total of 24,000 iterations per chain.

Preliminary analysis reveals that the data exhibit nonstationarity when assuming AR(p), $p > 0$, for the ϵ_{ij}'s. As our current methodology is intended only for stationary AR processes (see Section 4.6), we analyze the data in two ways: assuming AR(0), and assuming AR(1) by imposing stationarity through a restrictive prior for ϕ: we consider $\phi \sim \mathcal{N}(0, 0.00005)$ (Prior 1) and $\phi \sim \mathcal{N}(0, 0.0001)$ (Prior 2). However, imposing stationarity leads to poor mixing for some parameters as illustrated by the trace plots. Moreover, the estimates of ϕ vary considerably depending on the actual prior (the estimates are 0.52 and 0.91, respectively, from Priors 1 and 2). On the other hand, we observe reasonable mixing for all the parameters under an AR(0) assumption (Figure 6.1 displays the trace plots for population transition parameters μ_γ, μ_τ and μ_{τ_A} as examples). Therefore, we report here the results for AR(0). Note that our main goal is to make inference about the populations, and our simulation study (Section 7.4.1) reveals that the flexible methodology can perform well with respect to the population regression coefficients even for a misspecified correlation structure for the ϵ_{ij}'s. Note also that in Section 7.4.1, we consider more general p (up to $p = 2$) in our simulation study.

The trace, density and autocorrelation plots of each of the population parameters $\boldsymbol{\mu}_\beta$, $\boldsymbol{\mu}_\alpha$ and μ_{τ_A}, and the mixture probability ω are shown in Figures 6.2, 6.3 and 6.4, respectively. Here, the lack of any trend in the trace plot and low autocorrelations in the two chains indicate good mixing. The density plots in Figure 6.3 display no signs of multi-

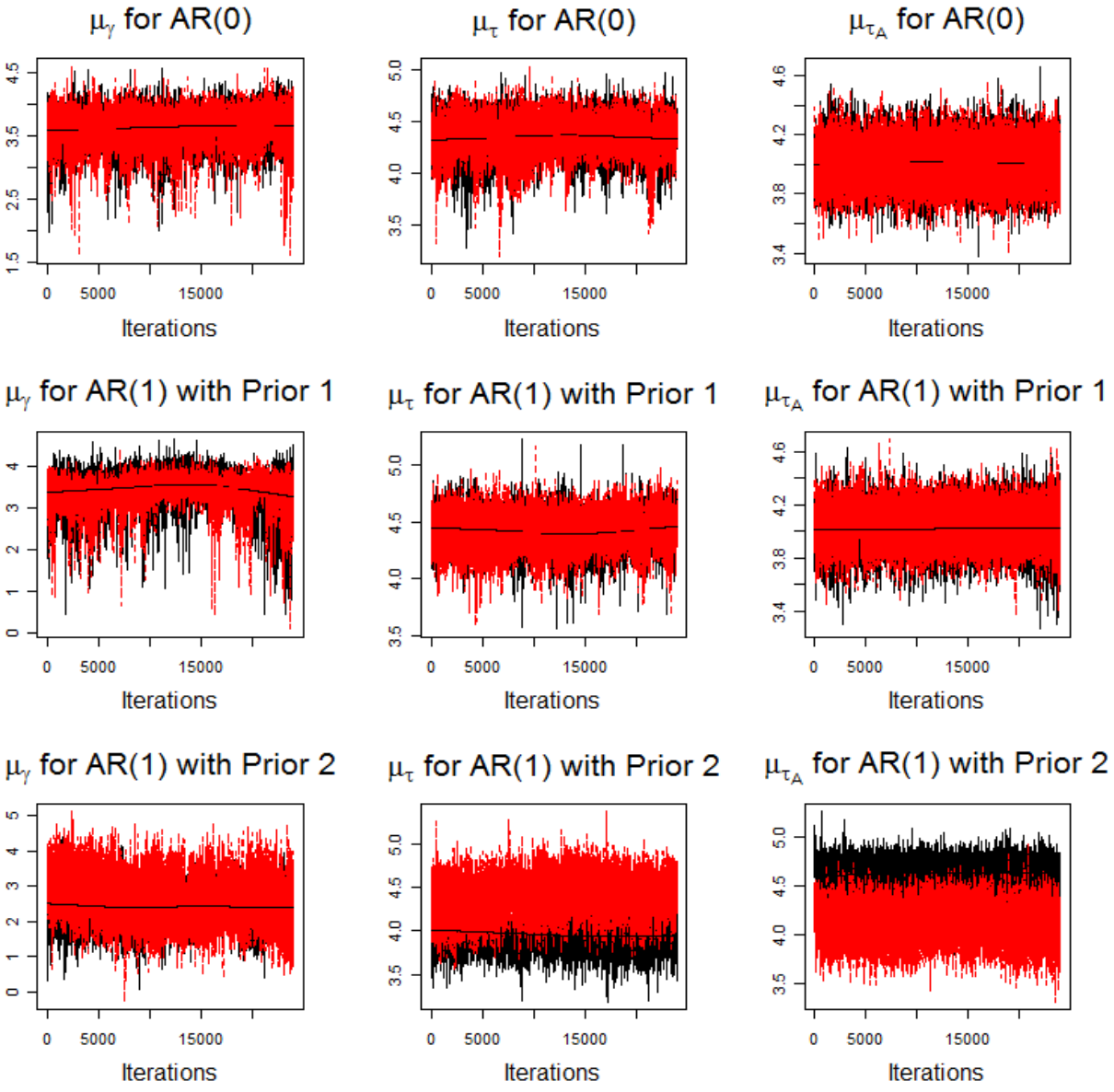

Figure 6.1: Rat data analysis – trace plots for the posteriors of parameters μ_γ, μ_τ and μ_{τ_A} from two chains for AR(0), and for AR(1) with $\phi \sim \mathcal{N}(0, 0.00005)$ (Prior 1) and $\phi \sim \mathcal{N}(0, 0.0001)$ (Prior 2); poor mixing is seen for the latter two cases.

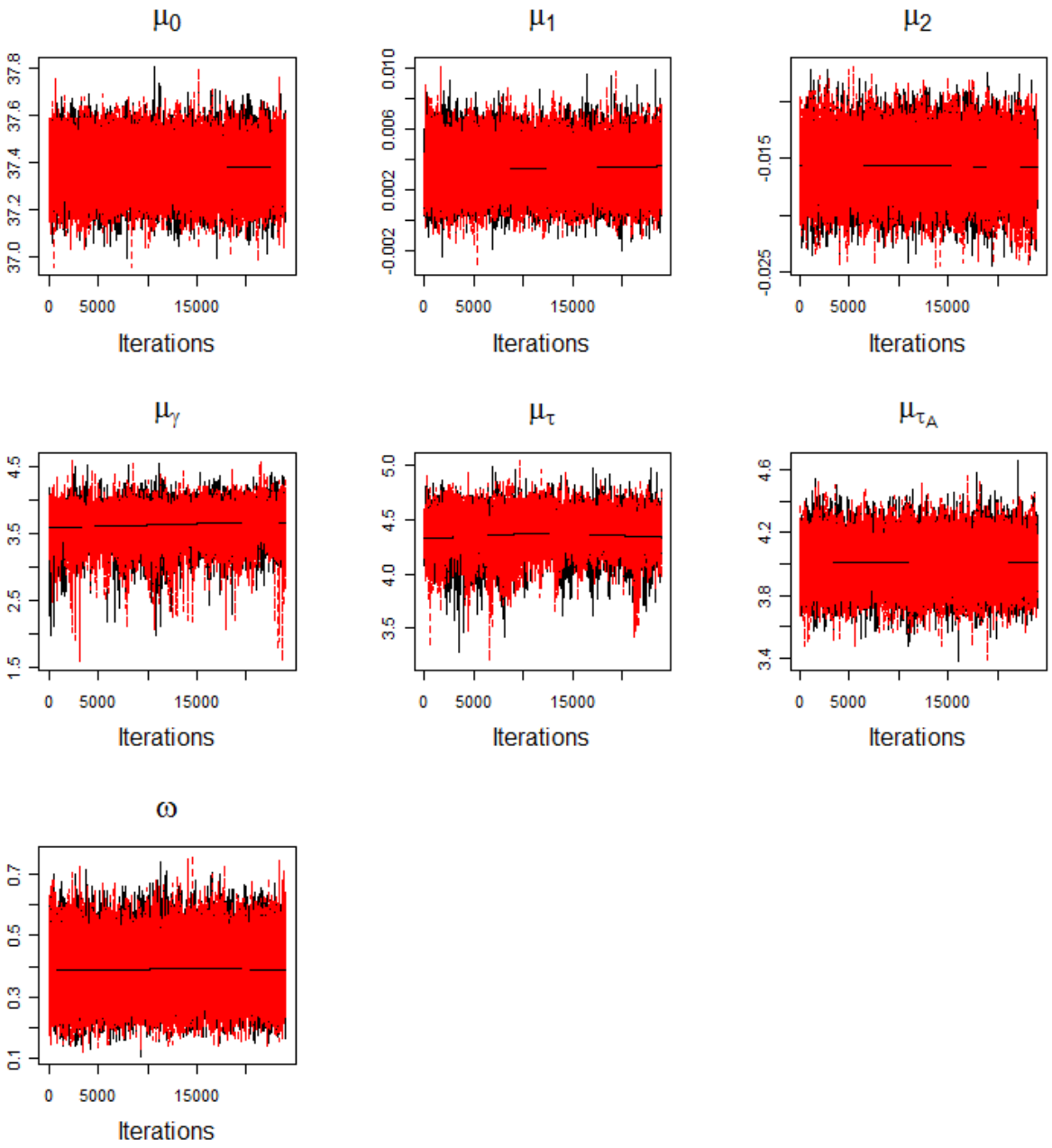

Figure 6.2: Rat data analysis – trace plots for the posteriors of the population parameters $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$, $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and μ_{τ_A}, and the mixture probability ω from two chains assuming AR(0) noise.

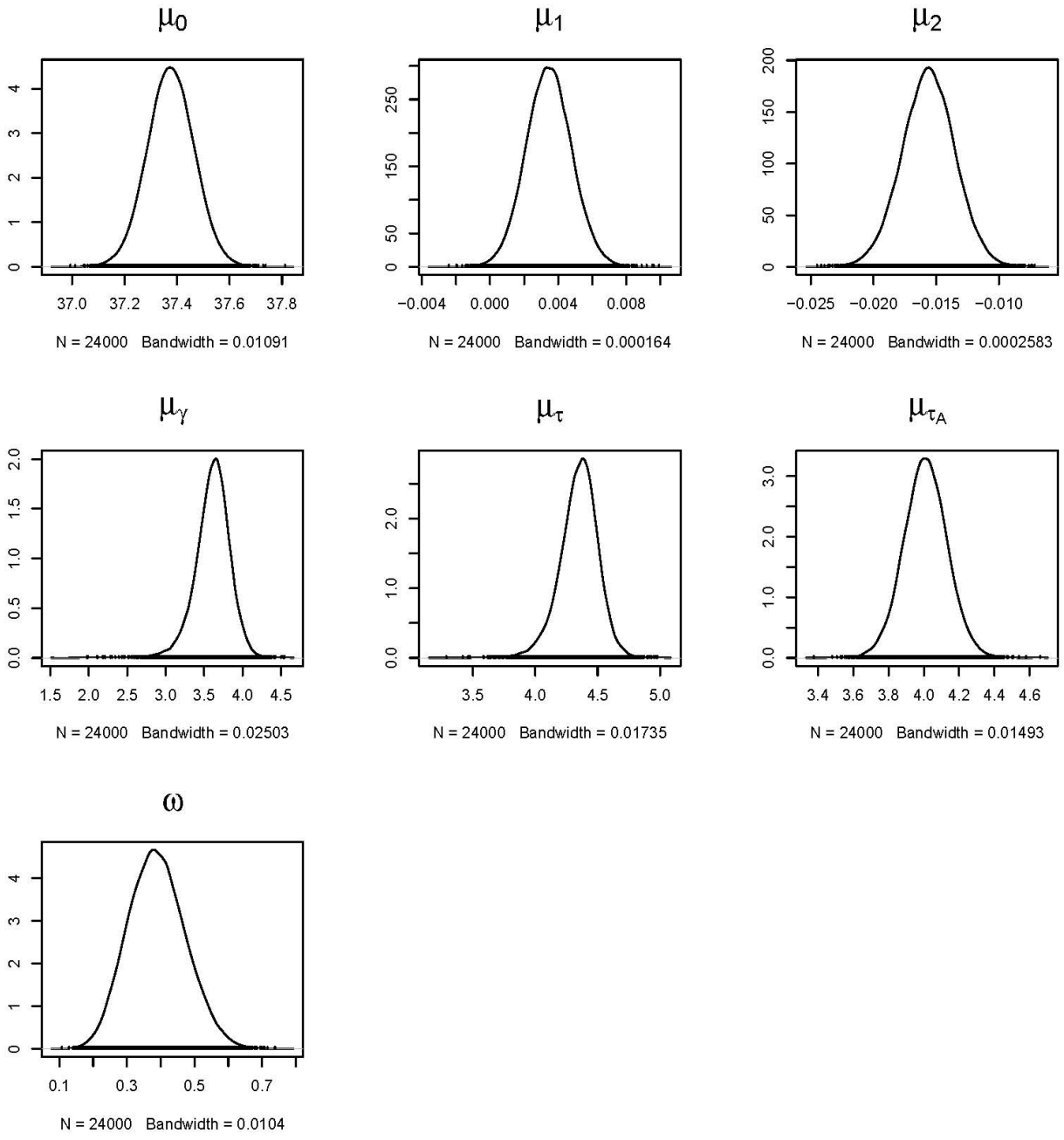

Figure 6.3: Rat data analysis – kernel density estimate plots for the posteriors of population parameters $\boldsymbol{\mu}_\beta = (\mu_0, \mu_1, \mu_2)'$, $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and μ_{τ_A}, and the mixture probability ω from two chains assuming AR(0) noise.

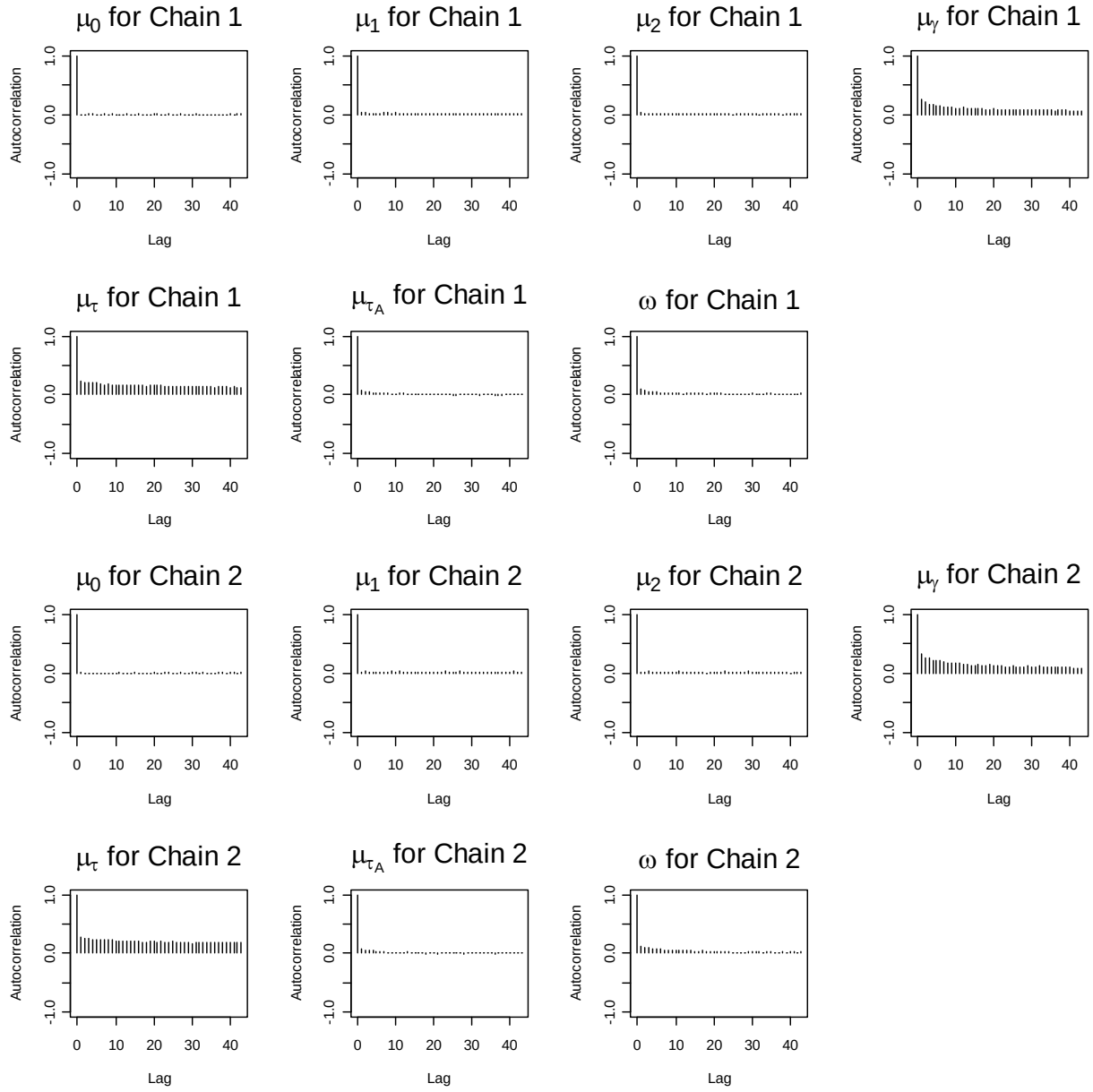

Figure 6.4: Rat data analysis – autocorrelation plots for the two chains for population parameters $\boldsymbol{\mu}_\beta = (\mu_0,\ \mu_1, \mu_2)'$, $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$ and μ_{τ_A}, and the mixture probability ω, assuming AR(0) noise.

modality. These three sets of plots show stationarity of the Markov chains. We also test the stationarity by the Gelman-Rubin statistic (Section 4.4.2), which provides no evidence against stationarity of the chains. Having reasonable mixing and convergence, we next proceed to interpret the results.

Table 6.1: Posterior summaries for the two populations of rats assuming AR(0) noise: posterior means for the population slope parameters (μ_1 and μ_2) are in "per 15 seconds" and those for the population transitions are in minutes.

	Posterior mean	95% credible interval
ω	0.39	$(0.23, 0.55)$
μ_0	37.38	$(37.21, 37.56)$
μ_1	0.003	$(0.001, 0.006)$
μ_2	-0.016	$(-0.020, -0.011)$
$\exp\{\mu_{\tau_A}\}$ (Population CTP for A)	13.89	(10.59, 17.34)
$\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}$ (Transition period for Population G)	10.11 to 29.03	–
$\exp\{\mu_\tau\} - \exp\{\mu_\gamma\} - 2\mu_1 \exp\{\mu_\gamma\}/\mu_2$ (Population CTP for G)	14.28	(6.33, 21.84)

Some posterior characteristics of parameters for the two populations are given in Table 6.1, and the population fitted curves are displayed in Figure 6.5. We see a significant linear increase in population T_c at the rate of 0.003^oC per 15 seconds in the incoming phase $\big($95% credible interval of the incoming slope is $(0.001, 0.006)$ which excludes $0\big)$. We also see virtually identical metabolic thresholds associated with a breakdown in the compensatory mechanisms for the two populations: posterior means for population CTP and $\exp\{\mu_{\tau_A}\}$ are 14.28 and 13.89 minutes, respectively. Thus, for G, the drop in T_c started at approximately 14.28 minutes after hemorrhage, and 13.89 minutes for A. Moreover, the posterior mean for ω is 0.39, which suggests that about 39% of the rats belong to Population G who

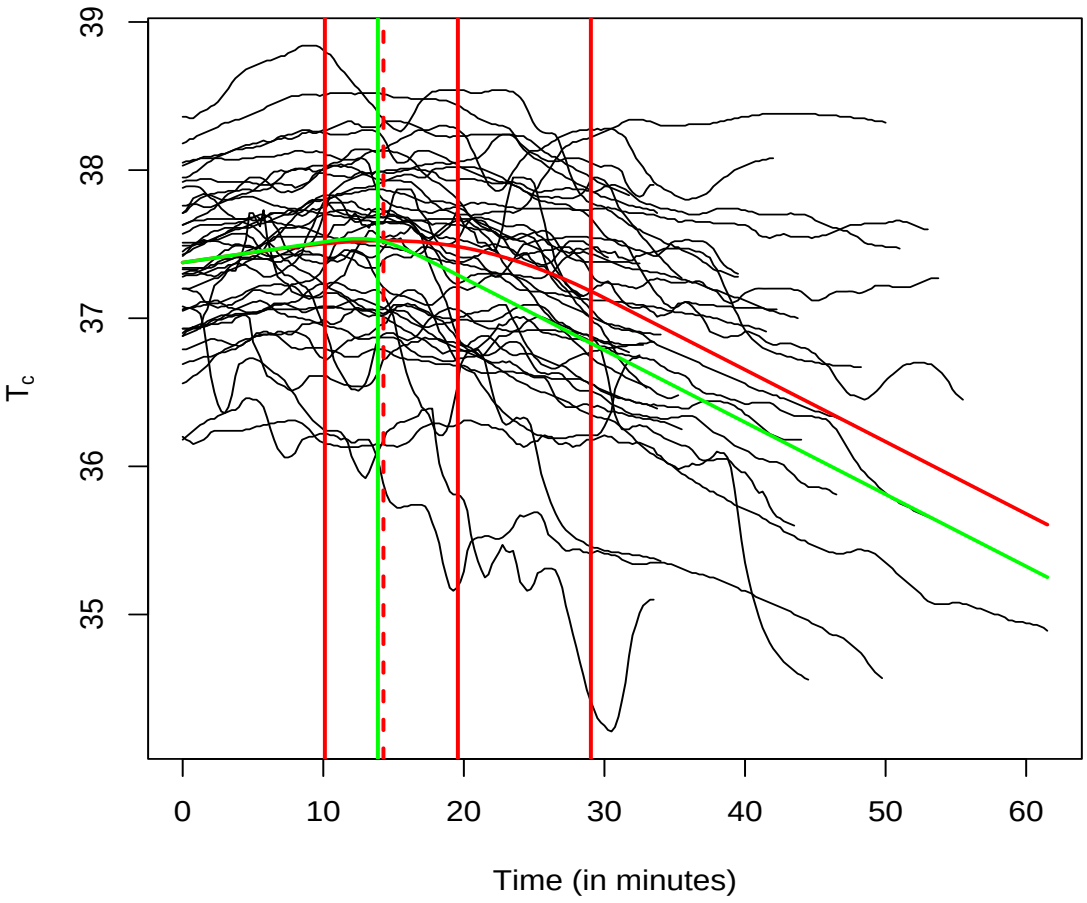

Figure 6.5: Observed data (black lines) and fitted curves (red and green curves) for the two populations of rats assuming within-individual independence. The estimated transition for Population G $\left(\text{i.e. } \widehat{\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}}\right)$ is marked by solid red vertical lines, and that for Population A $\left(\text{i.e. } \widehat{\exp\{\mu_{\tau_A}\}}\right)$ by the green vertical line. The estimated CTP for Population G is indicated by the dotted vertical line, which virtually coincides with the estimated transition point of Population A.

exhibit a slow (gradual) change in T_c, and a significant linear decrease thereafter at the rate of 0.013^oC per 15 seconds. Since the posterior means for $\exp\{\mu_\tau\} - \exp\{\mu_\gamma\}$ and $\exp\{\mu_\tau\} + \exp\{\mu_\gamma\}$ are 10.11 and 29.03 minutes, respectively, the population transition for those rats begins approximately 10.11 minutes from the time of hemorrhage and lasts for about 18.92 minutes. The remaining 61% of the rats are estimated to arise from Population

A, and exhibit an abrupt linear decrease, assumed through Assumption **B4** (Section 3.1.2) to be at the same rate as that for the linear decrease for Population G.

Individual fitted curves for 5 representative rats were presented in Figure 1.4. For convenience of reference, we reproduce the plot here again as Figure 6.6. In general, the fits look reasonable as the observed data closely agree with the respective fitted lines. The estimated transitions also demonstrate that our methodology picks up the two types of transition adequately.

The posterior characteristics of the standard deviations and correlations associated with Σ_β, Σ_α and $\sigma^2_{\tau_A}$ (priors for the random regression coefficients) are given in Table 6.2. We see small posterior variations in the linear regression coefficients, especially for the slope parameters β_{1i} and β_{2i}. Since the biological conditions of different rats should vary to some extent, we can expect some variation in the core body temperatures at the time of administering hemorrhage. This is reflected in the estimate of the standard deviation for the intercept β_{0i}, which is 0.535. After administering hemorrhage, we see very little variation in the slope parameters (the estimated standard deviations for β_{1i}, β_{2i} and $\beta_{1i} + \beta_{2i}$ are 0.008, 0.123 and 0.011, respectively), that is, all the rats exhibit very similar rates of increase/decrease in the core body temperatures before/after the transition period. Significant negative correlation between β_{1i} and β_{2i} $\big(\widehat{corr(\beta_{1i}, \beta_{2i})} = -0.476$ with 95% credible interval $(-0.711, -0.204)$ which excludes $0\big)$ indicates that the transition starts to take place sooner if the rate of increase in T_c in the incoming phase is high, and vice versa.

Let us turn now to the behavior of the transition in the T_c profiles. We see considerable variability in the times to maximal T_c and the variability in the times to transition zones (Table 6.2). This fact is reflected in the posterior medians for the standard deviations of γ_i and τ_i for Population G (7.643 and 10.172 minutes, respectively), and of τ_i for Population A (9.312 minutes). We also see significant negative correlation between γ_i and τ_i $\big(\widehat{corr(\gamma_i, \tau_i)} = -0.132$ with 95% credible interval $(-0.186, -0.025)$ which excludes $0\big)$, which implies that the wider the transition zone, the sooner the transition takes place, and vice versa. Relating this finding to the negative correlation between β_{1i} and β_{2i}, we can infer that if the rate of increase in T_c in the incoming phase is higher, then the transition takes place sooner with a wider transition zone so that there will be a delayed linear drop in the outgoing phase. On the other hand, if the rate of increase in T_c in the incoming phase is lower, then the transition takes place later with a narrower transition zone so that there will be an early linear drop in the outgoing phase.

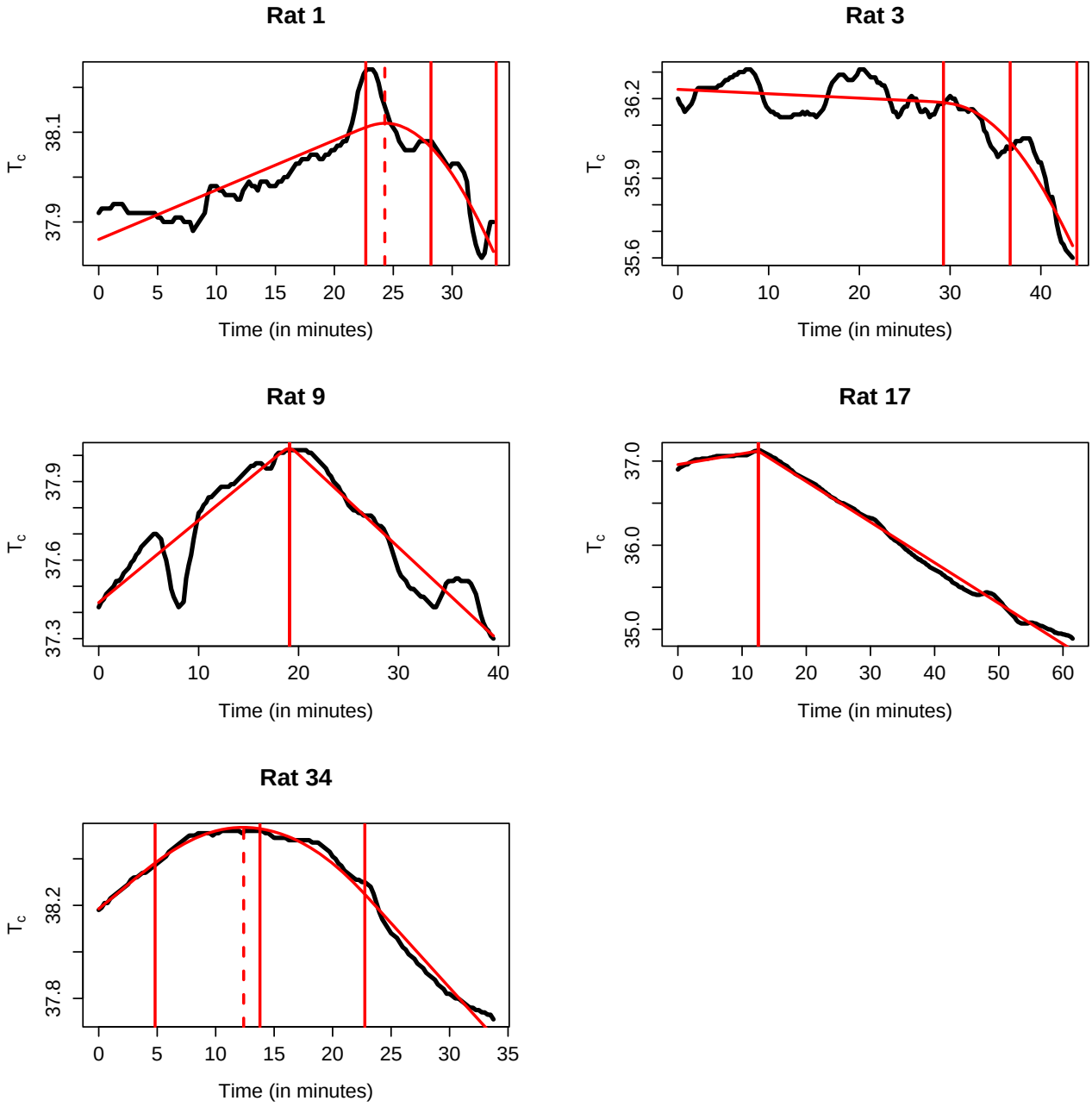

Figure 6.6: Observed data (black lines) and the corresponding individual-specific fitted curves (red lines) for 5 representative rats. Estimated transitions $\left(\text{i.e. } \hat{\tau} \text{ and } \widehat{\tau \pm \gamma}\right)$ are marked by the vertical lines with estimated CTPs (for Population G) by the dotted lines; The CTP estimate is not marked for Rat 3 because the estimated slope of its cable does not change signs.

Table 6.2: Rat data analysis – posterior summaries of the standard deviations and correlations associated with Σ_β, Σ_α and $\sigma^2_{\tau_A}$ (priors for the random regression coefficients); posterior summaries for the standard deviations of γ_i and τ_i are in minutes.

	Posterior median	95% credible interval
$\sqrt{(\Sigma_\beta)_{11}}$ (S.D. of the intercept β_{0i})	0.535	$(0.423, 0.669)$
$\sqrt{(\Sigma_\beta)_{22}}$ (S.D. of the incoming slope β_{1i})	0.008	$(0.006, 0.010)$
$\sqrt{(\Sigma_\beta)_{33}}$ (S.D. of the scaling factor β_{2i})	0.123	$(0.010, 0.016)$
$\sqrt{(\Sigma_\beta)_{22} + (\Sigma_\beta)_{33} + 2(\Sigma_\beta)_{23}}$ (S.D. of the outgoing slope $\beta_{1i} + \beta_{2i}$)	0.011	$(0.009, 0.014)$
$(\Sigma_\beta)_{12}/\sqrt{(\Sigma_\beta)_{11} \times (\Sigma_\beta)_{22}}$ (Correlation between β_{0i} and β_{1i})	0.023	$(-0.296, 0.343)$
$(\Sigma_\beta)_{13}/\sqrt{(\Sigma_\beta)_{11} \times (\Sigma_\beta)_{33}}$ (Correlation between β_{0i} and β_{2i})	-0.001	$(-0.323, 0.319)$
$(\Sigma_\beta)_{23}/\sqrt{(\Sigma_\beta)_{22} \times (\Sigma_\beta)_{33}}$ (Correlation between β_{1i} and β_{2i})	-0.476	$(-0.711, -0.204)$
$\sqrt{\exp\{2\mu_\gamma + (\Sigma_\alpha)_{11}\} \times [\exp\{(\Sigma_\alpha)_{11}\} - 1]}$ (S.D. of the half-width parameter γ_i for Population G)	7.643	$(2.967, 19.068)$
$\sqrt{\exp\{2\mu_\tau + (\Sigma_\alpha)_{22}\} \times [\exp\{(\Sigma_\alpha)_{22}\} - 1]}$ (S.D. of the center of the bend τ_i for Population G)	10.172	$(4.728, 20.928)$
$[\exp\{(\Sigma_\alpha)_{12}\} - 1]/\sqrt{[\exp\{(\Sigma_\alpha)_{11}\} - 1] \times [\exp\{(\Sigma_\alpha)_{22}\} - 1]}$ (Correlation between γ_i and τ_i for Population G)	-0.132	$(-0.186, -0.025)$
$\sqrt{\exp\{2\mu_{\tau_A} + \sigma^2_{\tau_A}\} \times [\exp\{\sigma^2_{\tau_A}\} - 1]}$ (S.D. of the CTP τ_i for Population A)	9.312	$(5.030, 16.200)$

In summary, under the assumptions that both populations share the same linear slopes and that each rat's T_c measurements are independent over time, our analysis yields the following points of clinical interest: (i) about 61% of the rats exhibit an abrupt linear drop in T_c during hemorrhage, whereas the remaining 39% exhibit a gradual transition followed by a linear drop; (ii) all rats are from populations that show approximately the same metabolic threshold (about 14 minutes after hemorrhage) associated with a breakdown in the compensatory mechanisms; (iii) during hemorrhage, either population shows a significant increase of T_c followed by a significant decrease; (iv) all the rats exhibit very similar rates of increase and decrease in T_c before and after the transition period, respectively (the estimated standard deviations for β_{1i}, β_{2i} and $\beta_{1i} + \beta_{2i}$ are 0.008, 0.123 and 0.011, respectively); (v) there is a considerable amount of between-rat variability in the times to maximal T_c and transition zones – the posterior medians for the standard deviations are large for γ_i and τ_i for both populations (all three values are between 7.6 and 10.2 minutes); (vi) there are significant negative correlations between the slope parameters β_{1i} and β_{2i} and the transition parameters γ_i and τ_i.

6.2 CFC-11 Data

Much of the material of this section can be found in our published article (Khan et al. [44]), but it is repeated in this thesis for the sake of completeness. Specifically, the background of the study, data description and analysis assuming AR(1) within-individual noise are taken from that article, and then we extend the analysis by assuming a higher-order AR process for the within-individual noises.

6.2.1 Background of the Study

Many biological consequences such as skin cancer and cataracts, irreversible damage to plants, and reduction of drifting organisms (animals, plants, archaea, bacteria) in the ocean's photic zone may result from the increased ultraviolet (UV) exposure due to ozone depletion. According to the *U.S. Environmental Protection Agency* (U.S. EPA) [79], each natural reduction in ozone levels has been followed by a recovery, though there is convincing scientific evidence that the ozone shield is being depleted well beyond changes due to

natural processes. In particular, ozone depletion due to human activities is a major concern, and may be controlled. One such human activity is the use of CFCs. As cited in "The Ozone Hole Tour" by the *University of Cambridge* [76], the catalytic destruction of ozone by atomic chlorine and bromine is the major cause of the forming of polar ozone holes, and photodissociation of CFC compounds is the main reason for these atoms to be in the stratosphere.

CFCs are nontoxic, nonflammable chemicals containing atoms of carbon, chlorine and fluorine. CFC-11 is one such compound. CFCs were extensively used in air conditioning/cooling units, and as aerosol propellants prior to the 1980's. While CFCs are safe to use in most applications and are inert in the lower atmosphere, they do undergo significant reaction in the upper atmosphere. Chlorine from CFCs is one of the most important free radical catalysts to destroy ozone. The destruction process continues over the atmospheric lifetime of the chlorine atom (one or two years), during which an average of 100,000 ozone molecules are broken down (The Columbia Encyclopedia [72]). Because of this, CFCs were banned globally by the 1987 Montréal Protocol on Substances That Deplete the Ozone Layer (The Columbia Encyclopedia [73]). Since this protocol came into effect, the atmospheric concentration of CFCs has either leveled off or decreased.

The effects of CFCs in ozone depletion is a global concern. Although exploratory data analyses reveal a decrease of CFCs in the earth's atmosphere since the early 1990's, so far no sophisticated statistical analysis has been conducted to evaluate the global trend. In addition, there are several other important questions regarding the CFC concentration in the atmosphere that could be useful not only to policy makers, but also for human awareness. For example, (q1) How long did it take for the CFC concentration to show an obvious decline? (q2) What were the rates of change (increase/decrease) in CFCs before and after the transition period? (q3) What was the critical time point (CTP) at which the CFC trend went from increasing to decreasing? In Section 6.2.3, we will address these questions statistically by fitting Model G for CFC-11 data. We focus on CFC-11, because it is considered one of the most dangerous CFCs to reduce the ozone layer in the atmosphere. In fact, it has the shortest lifetime of common CFCs, and is regarded as a reference substance in the definition of the ozone depletion potential (ODP). The ODP of a chemical is the ratio of its impact on ozone compared to the impact of a similar mass of CFC-11. Thus, the ODP is 1 for CFC-11 and ranges from 0.6 to 1 for other CFCs. (These facts about CFCs are taken from the U.S. EPA websites [77], [78].)

In a broader sense, we will comment in Section 6.2.3 on (1) the global trend of CFC-11, and (2) the effectiveness of the Montréal Protocol on preserving the ozone level by reducing the use of CFC-11. Our findings will also provide a rough idea of how long it may take to diminish CFC-11 from the earth's atmosphere.

6.2.2 Data

CFCs are monitored from different stations all over the globe by the Global Monitoring Division of the National Oceanic and Atmospheric Administration (NOAA/GMD [54]), and Atmospheric Lifetime Experiment/Global Atmospheric Gases Experiment/Advanced Global Atmospheric Gases Experiment (ALE/ GAGE/AGAGE [2]) sponsored by NASA. Henceforth, we will refer to these two programs simply as NOAA and AGAGE.

In general, the sites for the stations were chosen to ensure as much as possible that clean air is sampled to determine the concentrations of CFCs in clean air in the lower troposphere, where the troposphere begins at the surface and extends to between 7 km at the poles and 17 km at the equator. Gas chromatographs with Electron Capture Detectors (ECDs) were installed in each station to measure CFCs from air drawn through a sampling line using a pump (readers may refer to the NOAA website [55] and Prinn et al. [60] for details about the analytic techniques to measure CFCs).

Under the Radiatively Important Trace Species (RITS) program, NOAA began measuring CFCs using in situ gas chromatographs at their four baseline observatories — Pt. Barrow (Alaska) located on the northern most point of the U.S. with its intake about 11 meters above sea level (masl), Cape Matatula (American Samoa) located in the middle of the South Pacific with its intake about 42 masl and 50 meters from the shore, Mauna Loa (Hawaii) located around the Big Island of Hawaii with its intake about 3397 masl, and South Pole (Antarctica) located at the geographic South Pole on the Antarctic plateau with its intake about 2810 masl — and, in collaboration with the University of Colorado, at Niwot Ridge (Colorado) located in the alpine research area of the Institute for Arctic and Alpine Research with the site's intake about 3013 masl. We will label these five stations from 1 to 5 respectively. During the period of 1988-1999, a new generation of gas chromatography called Chromatograph for Atmospheric Trace Species (CATS) was developed and has been used to measure CFC concentrations ever since.

The AGAGE program consists of three stages corresponding to advances and upgrades in instrumentation (Prinn et al. [60]). The first stage (ALE) began in 1978, the second (GAGE) began during 1981-1985, and the third (AGAGE) began during 1993-1996. The current AGAGE stations are located in Mace Head (Ireland) with its intake about 25 masl and 10 meters from the shore, Cape Grim (Tasmania) with two intakes about 110 and 160 masl and 50 and 100 meters from the shoreline, respectively, Ragged Point (Barbados) with its intake about 42 masl and 20 meters from the ocean, Cape Matatula (American Samoa) located at the NOAA site, and Trinidad Head (California) with its intake about 140 masl and 50 meters from the shoreline. These five stations will be labeled from 6 to 10 respectively.

We consider monthly mean data for our statistical analysis. Ideally, we wish to have (a) full data for all stations, (b) a long enough period to capture all three phases of the CFC trend, and (c) no change in instrumentation to avoid the elements of non-stationarity and biased measurement, if any. However, we do not have the same duration of consecutive observations for all stations. Moreover, data were recorded by instrumentation that switched from one type to another. Table 6.3 summarizes the availability of the consecutive observations, and the instrumentations used to record data.

Thus, ideal statistical conditions are not achievable in this case. As a compromise, we remove Stations 9 and 10 from our analysis due to insufficient data, and choose a study period in such a way that it can reflect the changing behavior of the CFC-11 concentration in the atmosphere. The Montréal Protocol came into force on Jan 1, 1989. So, we expect an increasing trend in CFC-11 prior to 1989 because of its extensive use during that period. After the implementation of the protocol, we expect a change (either decreasing or leveling off) in the CFC-11 trend. To characterize this change, we wish to have a study period starting from some point before the implementation of the protocol. Moreover, we must have sufficient data to observe the change, if any. Thus, we settle for a relatively long study period of 152 months from Jan, 1988 to Aug, 2000, which is perhaps the most reasonable to satisfy (a)-(c) as much as possible. In particular, it covers Stations 3-4 with a single measuring device, RITS. Stations 2 and 5 have RITS data until April, 2000, at which point we truncate their data so that only RITS is present for all of Stations 2-5 throughout the study period. Data for the remaining four stations during this period were recorded by two measuring devices – RITS and CATS for Station 1, and GAGE and AGAGE for Stations 6-8 – each device occupying a substantial range of the 152

Table 6.3: CFC-11 data summary

Station	Available consecutive observations	Instrumentation
Barrow, Alaska	Nov, 1987 - Feb, 1999	RITS
(Station 1)	Jun, 1998 - Aug, 2008	CATS
Cape Matatula, American Samoa	May, 1989 - Apr, 2000	RITS
(Station 2)	Dec, 1998 - Aug, 2008	CATS
Mauna Loa, Hawaii	Jul, 1987 - Aug, 2000	RITS
(Station 3)	Jun, 1999 - Aug, 2008	CATS
South Pole, Antarctica	Jun, 1990 - Nov, 2000	RITS
(Station 4)	Feb, 1998 - Aug, 2008	CATS
Niwot Ridge, Colorado	Feb, 1990 - Apr, 2000	RITS
(Station 5)	May, 2001 - Jul, 2006	CATS
Mace Head, Ireland	Feb, 1987 - Jun, 1994	GAGE
(Station 6)	Mar, 1994 - Sep, 2007	AGAGE
Cape Grim, Tasmania	Dec, 1981 - Dec, 1994	GAGE
(Station 7)	Aug, 1993 - Sep, 2007	AGAGE
Ragged Point, Barbados	Aug, 1985 - Jun, 1996	GAGE
(Station 8)	Jun, 1996 - Sep, 2007	AGAGE
Cap Matatula, American Samoa	Jun, 1991 - Sep, 1996	GAGE
(Station 9)	Aug, 1996 - Sep, 2007	AGAGE
Trinidad Head, California	–	GAGE
(Station 10)	Oct, 1995 - Sep, 2007	AGAGE

months. Figure 6.7 shows the 8 profiles of the corresponding CFC-11 data. Specifically, each station constitutes an individual curve showing an initial increasing trend (incoming phase), a gradual transition period, and a decreasing trend after the transition period (outgoing phase), which is different from the others due to actual CFC-11 levels during measurement, exposure to wind and other environmental variables, sampling techniques,

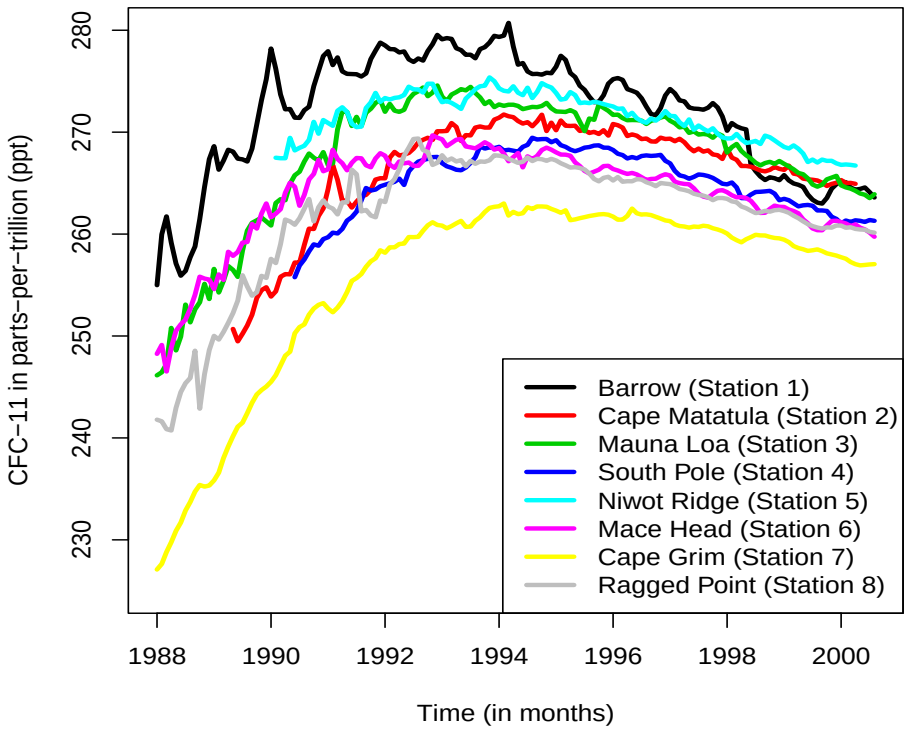

Figure 6.7: CFC-11 profiles of 8 stations (monthly mean data)

and so on. Our objective is to assess the global CFC-11 concentration in the atmosphere, as well as station-specific characterization of the trends.

Some remarks are required to explain t_{ij}. Recall that some stations do not have data for all 152 months. We employ the following system for defining t_{ij}. For example, the first and last months with recorded data by Station 3 are Jan, 1988 and Aug, 2000, respectively; thus, $t_{3,1} = 1, t_{3,2} = 2, \ldots, t_{3,152} = 152$. In contrast, Station 2 had its first and last recordings in May, 1989 and Apr, 2000, respectively; hence, $t_{2,1} = 17, t_{2,2} = 18, \ldots, t_{2,132} = 148$. The same approach is used to define t_{ij}'s for other i's. We remark here that stations with complete data can help with inferences for the stations with incomplete data by sharing information under a longitudinal modeling setup. Note that a few observations (from 1 to 5) were missing between the first and last months for a given station. We

replace them by observations from another data set (e.g. CATS or AGAGE) or by mean imputation based on neighboring time points if not available from another data set. As noted by McKnight et al. [50], if just a few missing values are replaced by the mean, the deleterious effect of mean substitution is reduced. So, we expect our findings to be minimally affected by this replacement for so few time points.

6.2.3 Results

Since we observe a gradual transition in the CFC-11 profiles, we use Model G (Section 5.2) to analyze this data set. Like the rat data analysis, we construct two Markov chains each of 5,000,000 iterations with the initial 200,000 iterations as burn-in. Inference is based on every 200^{th} iteration of the chains (thinning), resulting in a total of 24,000 iterations per chain.

Analysis Assuming AR(1) Within-Individual Noise

Table 6.4 shows the posterior characteristics of the global concentrations of CFC-11. The global drop took place between Jan, 1989 and Sep, 1994 approximately. The 95% credible intervals for μ_1 and μ_2 indicate significant slopes (neither interval includes 0) for the global incoming and outgoing phases, for which the posterior means are 0.65/month and -0.12/month, respectively. Thus, the average increase in CFC-11 was about 0.65 ppt for a one-month increase during the incoming phase (Jan, 1988 - Dec, 1988), and the average decrease was about 0.12 ppt during the outgoing phase (Oct, 1994 - Aug, 2000). The posterior mean for the global CTP is Nov, 1993, which may be interpreted as, overall CFC-11 went from increasing to decreasing around this time across all stations. The corresponding 95% credible interval ranged from Aug, 1992 to Jan, 1995.

The posterior mean for the AR(1) parameter ϕ is 0.81 with 95% credible interval $(0.77, 0.85)$. To see if a higher order AR process could be more appropriate to account for serial correlation in the repeated measures from the same individual, we next fit a model for each value of $p \in \{1, 2, \dots\}$ until the p^{th} AR coefficient is statistically insignificant in the sense that the associated 95% credible interval includes zero (also see simulations in Section 7.4.1). This process leads us to a model with AR(2) within-individual noise; we now present our results assuming AR(2).

Table 6.4: Posterior summaries of the global concentrations of CFC-11 assuming AR(1) within-individual noise: posterior means for the population slope parameters (μ_1 and μ_2) are in "per 1 month".

	Posterior mean	95% credible interval
μ_0	243.66	$(234.10, 253.16)$
μ_1	0.65	$(0.50, 0.80)$
μ_2	-0.76	$(-0.93, -0.60)$
$\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}$ (Transition period)	Jan, 1989 to Sep, 1994	–
$\exp\{\mu_\tau\} - \exp\{\mu_\gamma\} - 2\mu_1 \exp\{\mu_\gamma\}/\mu_2$ (CTP)	Nov, 1993	(Aug, 1992, Jan, 1995)

Analysis Assuming AR(2) Within-Individual Noise

Our analysis assuming AR(2) within-individual noise gives the posterior means of the AR(2) parameters to be $\hat{\phi}_1 = 0.87$ and $\hat{\phi}_2 = -0.08$ with 95% credible intervals $(0.81, 0.94)$, and $(-0.14, -0.01)$, respectively.

The trace, density and autocorrelation plots of each of the population coefficients μ_0, μ_1, μ_2, μ_γ and μ_τ are given in Figures 6.8, 6.9 and 6.10, respectively. The lack of any trend in the trace plot and low autocorrelations in the two chains indicate good mixing. We also see no signs of multimodality in the density plots. That is, those three sets of plots show stationarity of the Markov chain. Moreover, the Gelman-Rubin statistics provide no evidence against the stationarity of the chains.

Table 6.5 shows the posterior characteristics of the global concentrations of CFC-11, whereas Figure 6.11 displays the global fitted curve. Findings for the global concentrations of CFC-11 assuming an AR(2) within-individual noise are very similar to those reported above assuming an AR(1): posterior means for the population slopes are virtually identical, whereas those for the transitions are very close — a global drop occurred between Jan, 1989 and Aug, 1994, while we found between Jan, 1989 and Sep, 1994 assuming AR(1).

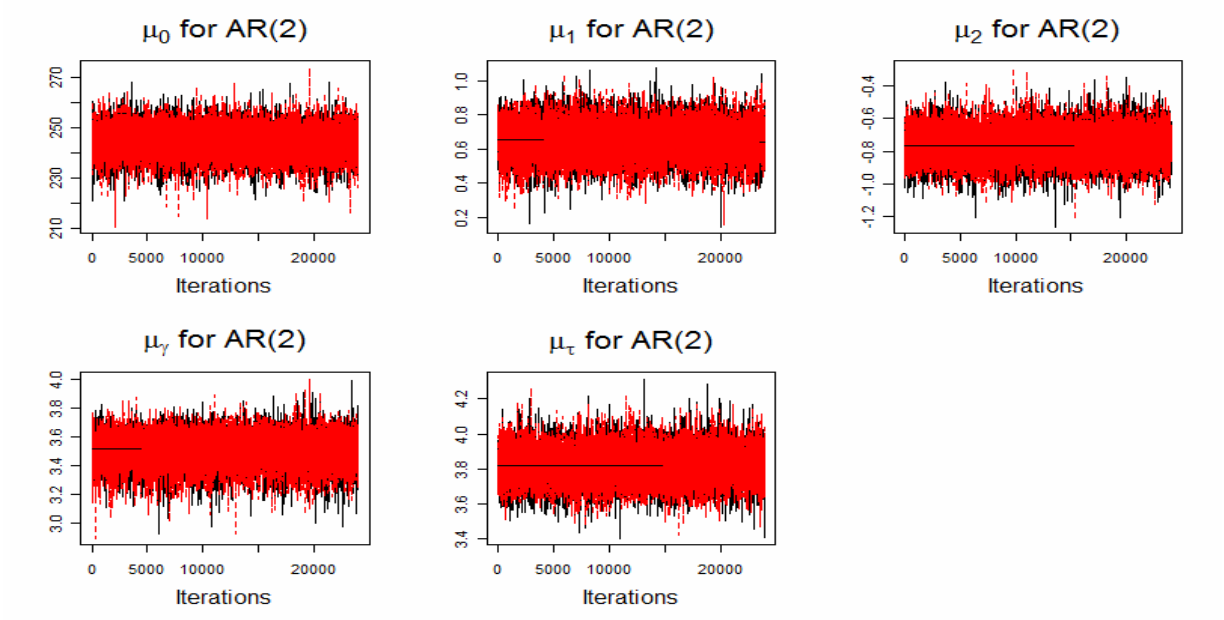

Figure 6.8: CFC data analysis – trace plots for the posteriors of the population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two chains assuming AR(2) noise.

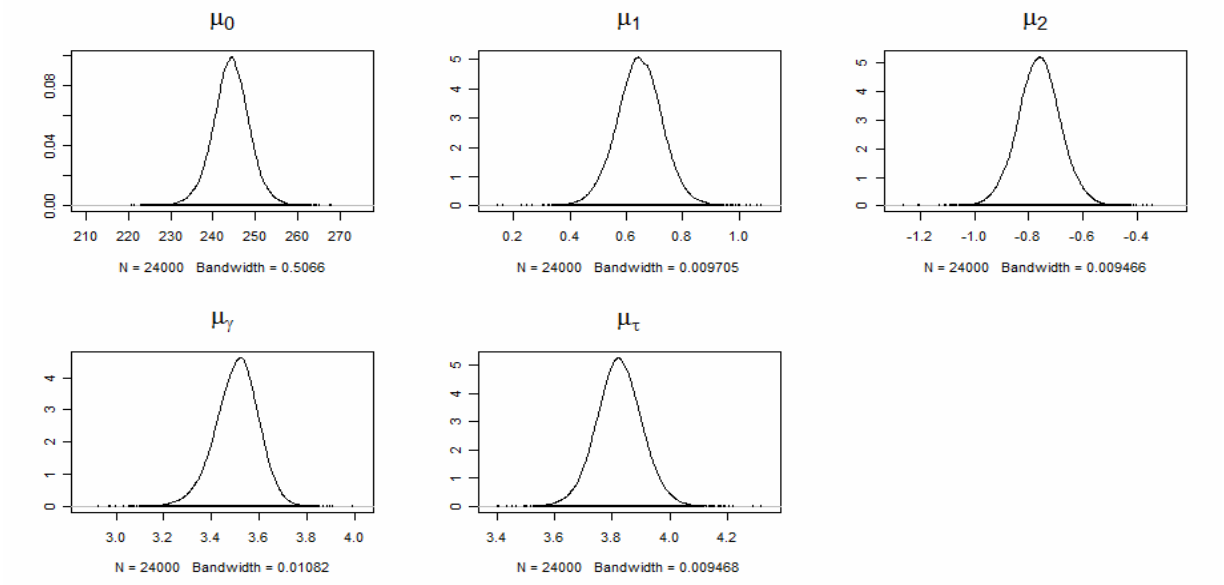

Figure 6.9: CFC data analysis – kernel density estimate plots for the posteriors of population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ from two chains assuming AR(2) noise.

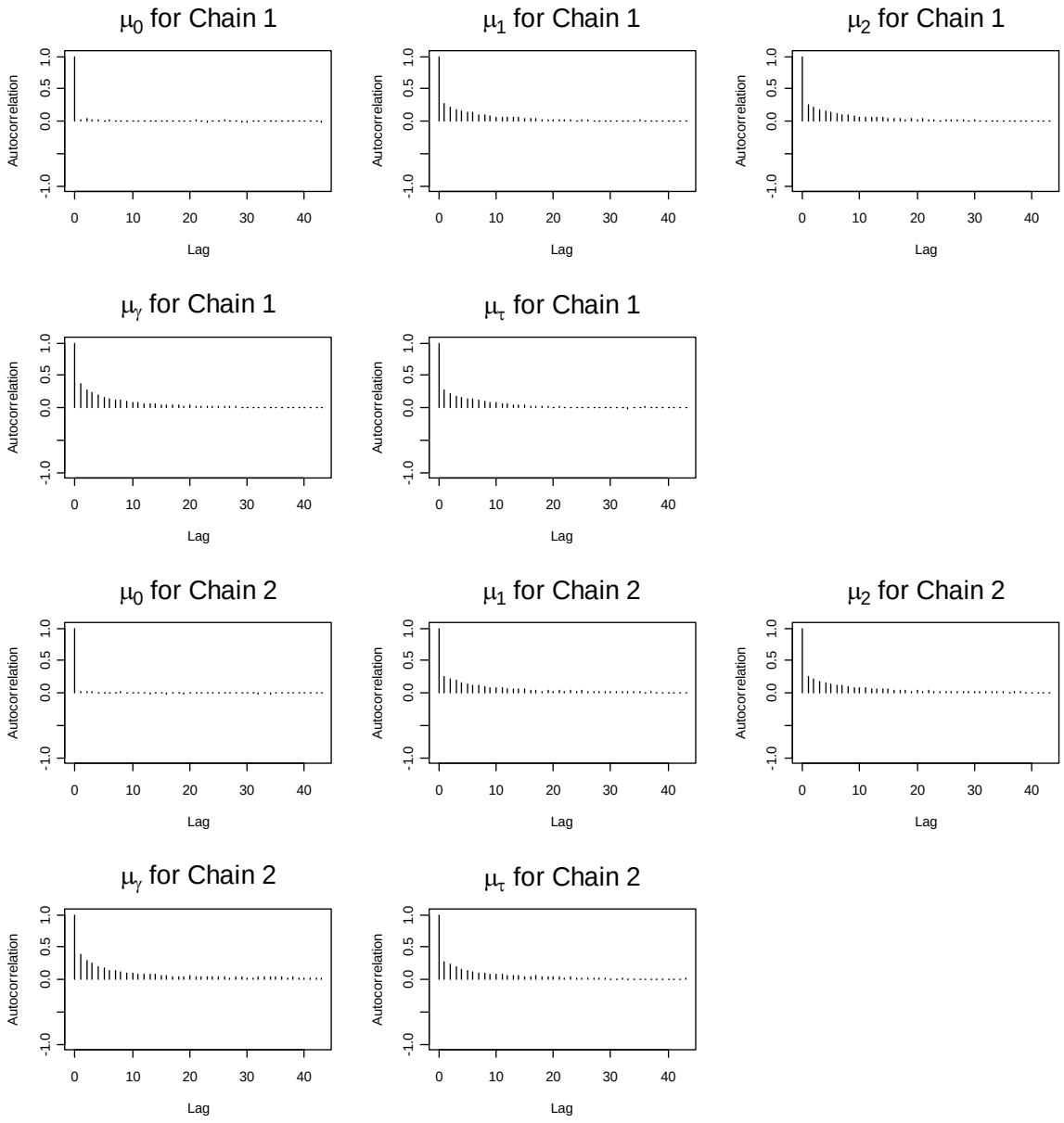

Figure 6.10: CFC data analysis – autocorrelation plots for the two chains for population parameters μ_0, μ_1, μ_2, μ_γ and μ_τ, assuming AR(2) noise.

The station-specific fits are displayed in Figure 6.12. It shows that our Model G fits the data well, with the observed data and individual fits agreeing quite closely. Table 6.6

Table 6.5: Posterior summaries of the global concentrations of CFC-11 assuming AR(2) within-individual noise: posterior means for the population slope parameters (μ_1 and μ_2) are in "per 1 month".

	Posterior mean	95% credible interval
μ_0	244.30	$(235.34, 253.37)$
μ_1	0.65	$(0.48, 0.81)$
μ_2	-0.76	$(-0.92, -0.60)$
$\exp\{\mu_\tau\} \pm \exp\{\mu_\gamma\}$ (Transition period)	Jan, 1989 to Aug, 1994	–
$\exp\{\mu_\tau\} - \exp\{\mu_\gamma\} - 2\mu_1 \exp\{\mu_\gamma\}/\mu_2$ (CTP)	Oct, 1993	(Aug, 1992, Dec, 1994)

summarizes the fits numerically, while Table 6.7 gives the posterior characteristics of the standard deviations and correlations associated with Σ_β and Σ_α (priors for the station-specific random regression coefficients). From Table 6.6, we see some variation in the estimates of the intercepts across the stations. We also observe significant increase/decrease of CFC-11 in the incoming/outgoing phases for all the stations separately. The rates at which these changes occur (Columns 3 and 4) agree closely for the stations. These phenomena are also evident in the estimate of Σ_β (Table 6.7), showing large variation of the deviations between the global and station-specific intercept parameters, while small variation for the slope parameters (the estimated standard deviations for β_{0i}, β_{1i}, β_{2i} and $\beta_{1i} + \beta_{2i}$ are 11.190, 0.165, 0.161 and 0.182, respectively). Note that unlike the rat data, no significant association between any pair of linear coefficients is observed for the CFC-11 data. This suggests that the transition location does not depend on the CFC-11 concentrations around the globe, as well as on the rates of increase and decrease in the incoming and outgoing phases, respectively.

The above findings support the notion of constant rates of increase and decrease, respectively, before and after the enforcement of the Montréal Protocol, observable despite a geographically spread-out detection network. They also point to the success of the

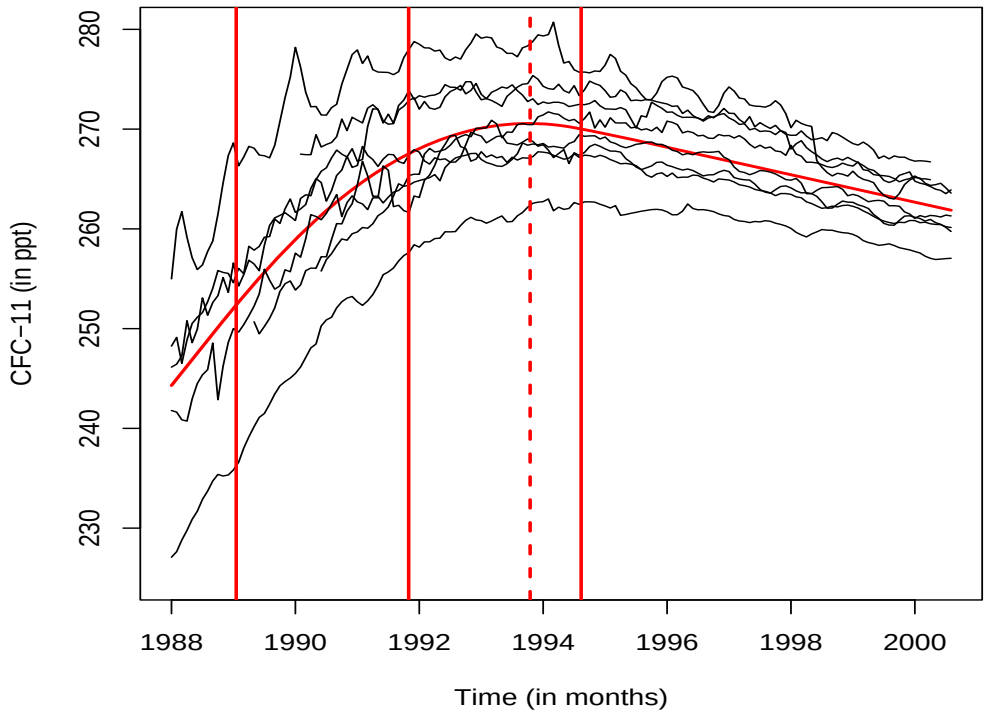

Figure 6.11: Observed data (black lines) and the corresponding population (global) fit (red) of the CFC-11 data assuming AR(2) within-individual noise. Estimate of the transition is marked by the solid vertical lines and that of the CTP by the dotted vertical line.

widespread adoption and implementation of the Montréal Protocol across the globe. However, the rate by which CFC-11 has been decreasing (about 0.12 ppt per month, globally) suggests that it will remain in the atmosphere throughout the 21st century, should current conditions prevail.

Let us turn now to the behavior of the transition of CFC-11 over time. Although γ_i and τ_i are not significantly associated ($\widehat{corr}(\gamma_i, \tau_i) = 0.245$ with 95% credible interval $(-0.484, 0.847)$ which excludes 0), the transition periods and critical time points varied somewhat across stations (Table 6.6), a fact which is also reflected in the estimate of Σ_α given in Table 6.7 (the estimated standard deviations for γ_i and τ_i are 5.191 and 7.646, respectively). This may be due to the extended CFC-11 phase-out schedules contained in

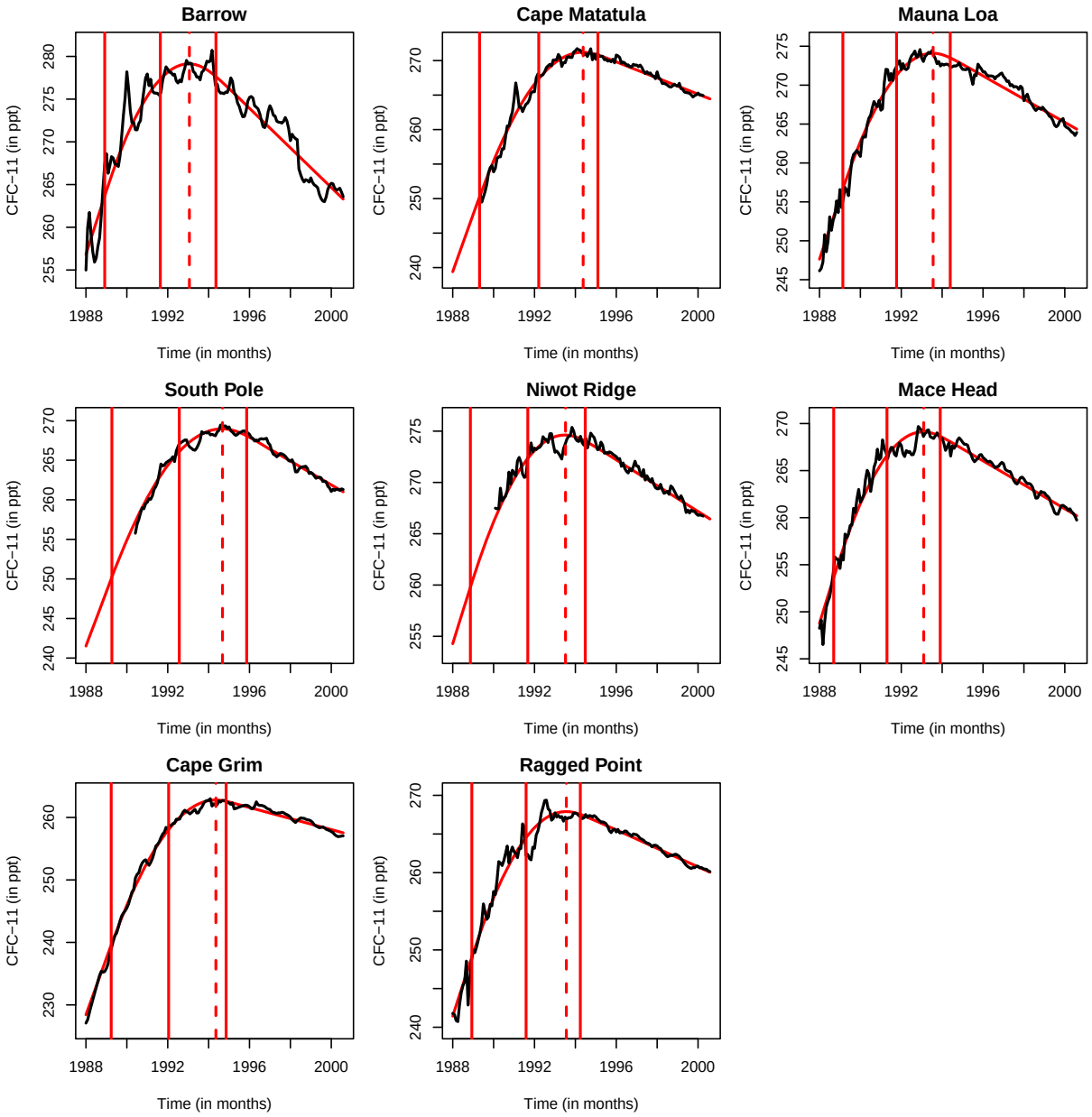

Figure 6.12: Observed data (black lines) and the corresponding station-specific fitted curves (red curves) of the CFC-11 data assuming AR(2) within-individual noise. Estimate of the transition is marked by the solid vertical lines and that of the CTP by the dotted vertical line.

Table 6.6: Estimated station-specific concentrations of CFC-11: posterior means for the slope parameters (β_{1i} and β_{2i}) are in "per 1 month".

	β_{0i} (95% credible interval)	β_{1i} (95% credible interval)	β_{2i} (95% credible interval)	Transition period[†] (Duration)	CTP[‡] (95% credible interval)
Barrow, Alaska ($\hat{\sigma}_1^2 = 0.88$)	256.83 (253.18, 260.46)	0.63 (0.45, 0.81)	−0.82 (−1.00, −0.65)	Dec, 1988 - May, 1994 (65 months)	Jan, 1993 (Aug, 1992 to Jul, 1993)
Cap Matatula, American Samoa ($\hat{\sigma}_1^2 = 0.33$)	239.40 (234.56, 244.40)	0.70 (0.50, 0.90)	−0.79 (−0.99, −0.60)	Apr, 1989 - Feb, 1995 (70 months)	May, 1994 (Nov, 1993 to Nov, 1994)
Mauna Loa, Hawaii ($\hat{\sigma}_1^2 = 0.66$)	247.64 (244.38, 250.89)	0.67 (0.50, 0.84)	−0.79 (−0.96, −0.63)	Feb, 1989 - May, 1994 (63 months)	Jul, 1993 (Jan, 1993 to Feb, 1994)
South Pole, Antarctica ($\hat{\sigma}_1^2 = 0.10$)	241.52 (236.32, 246.85)	0.58 (0.39, 0.76)	−0.71 (−0.88, −0.51)	Apr, 1989 - Nov, 1995 (79 months)	Sep, 1994 (May, 1994 to Dec, 1994)
Niwot Ridge, Colorado ($\hat{\sigma}_1^2 = 0.29$)	254.28 (246.86, 261.44)	0.54 (0.30, 0.79)	−0.64 −0.89, −0.41	Nov, 1988 - Jun, 1994 (67 months)	Jul, 1993 (Dec, 1992 to Jan, 1994)
Mace Head, Ireland ($\hat{\sigma}_1^2 = 0.43$)	248.77 (245.90, 251.63)	0.59 (0.43, 0.76)	−0.70 (−0.86, −0.54)	Sep, 1988 - Nov, 1993 (62 months)	Feb, 1993 (Jul, 1992 to Aug, 1993)
Cape Grim, Tasmania ($\hat{\sigma}_1^2 = 0.10$)	228.42 (226.69, 230.18)	0.76 (0.65, 0.90)	−0.83 (−0.97, −0.72)	Mar, 1989 - Nov, 1994 (68 months)	May, 1994 (Feb, 1994 to Aug, 1994)
Ragged Point, Barbados ($\hat{\sigma}_1^2 = 0.81$)	241.44 (238.04, 244.91)	0.69 (0.51, 0.86)	−0.79 (−0.96, −0.62)	Dec, 1988 - Mar, 1994 (63 months)	Jul, 1993 (Dec, 1992 to Feb, 1994)

[†] $\tau_i \pm \gamma_i$
[‡] $\tau_i - \gamma_i - 2\beta_{1i}\gamma_i/\beta_{2i}$

Table 6.7: CFC-11 data analysis – posterior summaries of the standard deviations and correlations associated with Σ_β and Σ_α (priors for the random regression coefficients).

	Posterior median	95% credible interval
$\sqrt{(\Sigma_\beta)_{11}}$ (S.D. of the intercept β_{0i})	11.190	$(6.416, 19.073)$
$\sqrt{(\Sigma_\beta)_{22}}$ (S.D. of the incoming slope β_{1i})	0.165	$(0.089, 0.285)$
$\sqrt{(\Sigma_\beta)_{33}}$ (S.D. of the scaling factor β_{2i})	0.161	$(0.086, 0.278)$
$\sqrt{(\Sigma_\beta)_{22} + (\Sigma_\beta)_{33} + 2(\Sigma_\beta)_{23}}$ (S.D. of the outgoing slope $\beta_{1i} + \beta_{2i}$)	0.182	$(0.107, 0.306)$
$(\Sigma_\beta)_{12}/\sqrt{(\Sigma_\beta)_{11} \times (\Sigma_\beta)_{22}}$ (Correlation between β_{0i} and β_{1i})	-0.340	$(-0.871, 0.363)$
$(\Sigma_\beta)_{13}/\sqrt{(\Sigma_\beta)_{11} \times (\Sigma_\beta)_{33}}$ (Correlation between β_{0i} and β_{2i})	0.210	$(-0.494, 0.815)$
$(\Sigma_\beta)_{23}/\sqrt{(\Sigma_\beta)_{22} \times (\Sigma_\beta)_{33}}$ (Correlation between β_{1i} and β_{2i})	-0.382	$(-0.891, 0.297)$
$\sqrt{\exp\{2\mu_\gamma + (\Sigma_\alpha)_{11}\} \times [\exp\{(\Sigma_\alpha)_{11}\} - 1]}$ (S.D. of the half-width parameter γ_i)	5.191	$(2.294, 10.154)$
$\sqrt{\exp\{2\mu_\tau + (\Sigma_\alpha)_{22}\} \times [\exp\{(\Sigma_\alpha)_{22}\} - 1]}$ (S.D. of the center of the bend τ_i)	7.646	$(3.886, 14.265)$
$[\exp\{(\Sigma_\alpha)_{12}\} - 1]/\sqrt{[\exp\{(\Sigma_\alpha)_{11}\} - 1] \times [\exp\{(\Sigma_\alpha)_{22}\} - 1]}$ (Correlation between γ_i and τ_i)	0.245	$(-0.484, 0.847)$

the Montréal Protocol – 1996 for developed countries and 2010 for developing countries. Thus, many countries at various geographical locations continued to contribute CFCs to

the atmosphere during the 152 months in our study period, while those at other locations had stopped. Overall, the eight transitions began between Sep, 1988 and Apr, 1989, a period of only 8 months. This reflects the success and acceptability of the protocol all over the globe. Durations of the transition periods are very similar among stations except for South Pole. Thus, it took almost the same amount of time in different parts of the world for CFC-11 to start dropping linearly with an average rate of about 0.12 ppt per month.

Transition for South Pole is estimated to take place over 79 months, an extended period compared to the other stations. This could be due to the highly unusual weather conditions specific to the location. CFCs are not disassociated during the long winter nights in the South Pole. Only when sunlight returns in October does ultraviolet light break the bond holding chlorine atoms to the CFC molecule ("Ozone Hole Watch", NASA [53]). For this reason, it may be expected for CFCs to remain in the atmosphere over the South Pole for a longer period of time, and hence, an extended transition period. Indeed, our findings for South Pole are very similar to those reported by Ghude et al. [33]. To evaluate the trend, the authors used the NASA EdGCM model – a deterministic global climate model wrapped in a graphical user interface. They found the average growth rate to be 9 ppt per year for 1983 – 1992, and about -1.4 ppt per year for 1993 – 2004, turning to negative in the mid 1990's. With our statistical modeling approach, we estimate a linear growth rate of 0.58 ppt per month (6.96 ppt per year) prior to Apr, 1989, a transition between Apr, 1989 and Nov, 1995, and a negative linear phase (-0.13 ppt per month, or -1.56 ppt per year) after November, 1995.

The estimates of the innovation variances (σ_i^2's) are given in the first column of Table 6.6. One noticeable fact from the profile plot (Figure 6.7) is that Barrow measurements are more variable, whereas Cape Grim and South Pole show little variation over time. This is reflected in their innovation variance estimates of 0.88, 0.10, and 0.10, respectively.

In summary, since the Montréal Protocol came into effect, a global decrease in the CFC-11 has been observed, a finding confirmed by our analysis. A gradual transition in the CFC-11 profiles makes scientific sense due to the fact that CFC molecules can stay in the upper atmosphere for about a century, and their breakdown does not take place instantaneously. The substantial decrease in global CFC-11 levels after the gradual change shown by our analysis suggests that the Montréal Protocol can be regarded as a successful international agreement to reduce the negative impact of CFCs on the ozone layer.

6.3 Discussion and Conclusion

In this chapter, we illustrated our methodology via two examples. In general, our methodology performed well for the two examples as the respective fitted lines closely resemble the observed data. Moreover, the population fit well represented the individual profiles and provided useful information about the study of interest. In particular, the general methodology (mixed bent-cable regression) used to analyze the rat data picked up the two types of transitions adequately as desired.

We have noticed from the two analyses that the general methodology requires relatively long Markov chains for good mixing and convergence to the stationary distribution. This may be due to the fact that when the sample potentially comes from two populations, the type of transition (abrupt or gradual) for some individual profiles may not be so obvious, and can almost be equally well-fitted by both the broken stick and bent cable models. This ultimately necessitates longer runs for such profiles to converge.

Chapter 7

Simulation Study

In this chapter, we present three scenarios with simulated data to illustrate the efficacy of our mixed bent-cable methodology. In Section 7.1, we describe the three scenarios, and in Section 7.2, our choices of the parameters for the simulations. Our approach of generating data is presented in Section 7.3, and simulation results in Section 7.4. We conclude this chapter with a discussion in Section 7.5.

7.1 Scenarios

We present two scenarios (*Scenarios 1* and *2*) that differ in the within-individual dependence structure. In *Scenario 1*, we consider an AR(1) process with $\phi = 0.7$. In *Scenario 2*, we have an AR(2) process with $\boldsymbol{\phi} = (0.8, -0.1)'$. In each of these scenarios, we then analyze the data assuming AR(0), AR(1) and AR(2) to evaluate the performance of our methodology under misspecified assumptions for the within-individual dependence structure.

It may not be possible to be absolutely certain that a sample arises from a single population (either A or G). Therefore, to evaluate the performance of our methodology under a misspecified assumption regarding the type of transition, we present here one more scenario (*Scenarios 3*), where we model the data assuming only Population G (i.e., Model G), when in reality, both populations A and G exist with $\omega = 0.90$ (*Scenario 3a*), and $\omega = 0.95$ (*Scenario 3b*). Note that ω equals, say, 0.90 implies that each individual has

probability 0.90 to have come from Population G (and, hence, probability 0.10 to come from Population A). In both *Scenarios 3a* and *3b*, we consider an AR(1) process with $\phi = 0.7$ for the within-individual noise.

7.2 Parameters in the Simulations

Parameters in the simulations are chosen to approximately mimic the CFC-11 data. However, we assume here that the samples come from two potential populations (A and G): $\omega = 0.50$ in *Scenarios 1* and *2*, and $\omega = 0.90$ and 0.95 in *Scenarios 3a* and *3b*, respectively. For all the scenarios, we consider

$$m = 20, \quad n \equiv n_i = 150 \text{ for } i = 1, 2, \ldots, m, \quad t_{ij} = j - 1, \quad \text{for } j = 1, 2, \ldots, n,$$
$$\boldsymbol{\mu}_\beta = (244, 0.5, -0.75)', \quad \boldsymbol{\mu}_\alpha = (3.00, 4.00)', \quad \mu_{\tau_A} = 4.50, \quad \sigma^2_{\tau_A} = 0.05,$$

$$\Sigma_\beta = \begin{bmatrix} 125.00 & -1.00 & 0.50 \\ -1.00 & 0.03 & -0.01 \\ 0.50 & -0.01 & 0.03 \end{bmatrix}, \quad \Sigma_\alpha = \begin{bmatrix} 0.020 & 0.005 \\ 0.005 & 0.030 \end{bmatrix},$$

and σ_i^2's are 0.34, 1.12, 1.75, 0.42, 0.74, 2.06, 1.16, 1.28, 0.16, 0.77, 0.04, 0.03, 0.91, 1.95, 0.32, 2.02, 0.89, 0.90, 0.82 and 2.89 (randomly drawn from the $\mathcal{G}(1, 1)$ distribution).

7.3 Data

Given the above parameters, we generate y_{ij}'s for each i according to the following steps:

1. draw $\boldsymbol{\beta}_i$ from $\mathcal{MVN}(\boldsymbol{\mu}_\beta, \Sigma_\beta)$;
2. draw I_i from $\mathcal{BER}(\omega)$; if $I_i = 0$, take $\gamma_i = 0$ and τ_i to be a random point from $\mathcal{LN}(\mu_{\tau_A}, \sigma^2_{\tau_A})$, otherwise draw $\boldsymbol{\alpha}_i$ from $\mathcal{LN}_2(\boldsymbol{\mu}_\alpha, \Sigma_\alpha)$;
3. given t_{ij}, $\boldsymbol{\beta}_i$ and $\boldsymbol{\alpha}_i$, calculate $\mathbf{f}_i = (f_{i1}, f_{i2}, \ldots, f_{in_i})'$, where f_{ij} is as in Equation (3.2);

4. given $\boldsymbol{\phi}$ and σ_{ui}^2, simulate $\boldsymbol{\epsilon}_i$ according to an AR(p) process;

5. then, set $\mathbf{y}_i = \mathbf{f}_i + \boldsymbol{\epsilon}_i$, where $\mathbf{f}_i$ and $\boldsymbol{\epsilon}_i$ are from steps 3 and 4, respectively.

7.4 Results

For each simulation, 500 data sets are generated, and 100,000 MCMC iterations with post-burnin 90,000 (an initial 10,000 iterations are discarded) are used to analyze each set. We calculate the posterior characteristics for each of the 500 data sets, after which posterior summaries are averaged over the 500 sets for each parameter, and the coverage probability of 95% credible intervals (the proportion of such credible intervals out of 500 that capture the truth) is calculated.

7.4.1 *Scenarios 1* and *2*

Results for *Scenarios 1* and *2* are summarized in Tables 7.1 to 7.3. Table 7.1 suggests that our methodology performs well for both scenarios with respect to the population characteristics: the average of posterior means for each parameter is close to the true parameter value, and coverage probabilities (from 0.92 to 0.99) are all reasonably close to the nominal 0.95.

Consider the case of analyzing a data set generated with AR(1) noise ($\phi = 0.7$) but assuming AR(2) in the inference. Note that the true value of ϕ_2 is zero. We see from Table 7.1 that the average of the posterior means for ϕ_2 comes out close to zero with coverage probability 0.92 (calculated assuming the true value of ϕ_2 is 0), whereas the average of the posterior means for ϕ_1 is 0.7, coinciding with the true value of ϕ for the AR(1) process, with coverage probability 0.94 (calculated assuming the true value of ϕ_1 is 0.7). The above results indicate that analyzing a data set generated with AR(p) noise but assuming AR($p+1$) in the inference gives (1) the averages of the posterior means for the AR parameters are all close to their true values, that is, the averages for the first p parameters are close to their respective values used to generate the data, and the average for the $(p+1)^{th}$ parameter is close to zero; and (2) coverage probabilities are all reasonably close to the nominal 0.95. These two results support the technique we employed in selecting p for the CFC-11 data in Section 6.2.3.

Table 7.1: Simulation study with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means of the mixing proportion ω, population regression coefficients and the AR parameters; also coverage of 95% credible intervals.

		Simulated ϵ_{ij}'s: AR(1)			Simulated ϵ_{ij}'s: AR(2)		
		Analysis assuming			Analysis assuming		
		AR(2)	AR(1)	AR(0)	AR(2)	AR(1)	AR(0)
	True	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage
ω	0.50	$0.52, 0.98$	$0.52, 0.99$	$0.52, 0.97$	$0.52, 0.97$	$0.51, 0.95$	$0.52, 0.96$
μ_0	244.00	$243.94, 0.97$	$244.45, 0.94$	$244.37, 0.96$	$244.64, 0.96$	$244.30, 0.95$	$244.45, 0.96$
μ_1	0.50	$0.50, 0.97$	$0.48, 0.94$	$0.49, 0.95$	$0.48, 0.94$	$0.48, 0.95$	$0.48, 0.93$
μ_2	-0.75	$-0.75, 0.95$	$-0.77, 0.92$	$-0.78, 0.92$	$-0.77, 0.95$	$-0.78, 0.93$	$-0.77, 0.92$
μ_γ	3.00	$2.95, 0.95$	$2.94, 0.95$	$2.95, 0.93$	$2.97, 0.96$	$2.95, 0.97$	$2.96, 0.93$
μ_τ	4.00	$4.02, 0.98$	$4.02, 0.97$	$4.04, 0.96$	$4.01, 0.99$	$4.02, 0.96$	$4.04, 0.92$
μ_{τ_A}	4.50	$4.50, 0.94$	$4.50, 0.96$	$4.47, 0.94$	$4.49, 0.95$	$4.49, 0.97$	$4.47, 0.94$
ϕ for AR(1)	0.70	–	$0.71, 0.93$	–	–	$0.74, -$ -	–
ϕ_1 for AR(2)	0.80	$0.70, 0.94$ -	–	–	$0.80, 0.95$	–	–
ϕ_2 for AR(2)	-0.10	$0.005, 0.92$	–	–	$-0.10, 0.95$	–	–

Table 7.2 refers to the results for the variances and covariances in the priors of the random regression coefficients. Coverage probabilities are all close to 0.99 (over-coverage) except for poor coverage for $(\Sigma_\alpha)_{11}$ (variance of γ_i for Population G): 0.78 and 0.80 for data sets generated from AR(1) and AR(2), respectively, but using an AR(0) fit. For such model misspecification, we also see a slight under coverage for $(\Sigma_\alpha)_{22}$ (variance of τ_i for Population G) and overestimation of $(\Sigma_\alpha)_{11}$, $(\Sigma_\alpha)_{22}$ and $\sigma^2_{\tau_A}$ (variability of τ_i or γ_i).

Table 7.2: Simulation study with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means (medians for the variance parameters) of the variances and covariances ($\sigma^2_{\tau_A}$, Σ_β and Σ_α) in the priors for the random regression coefficients; also coverage of 95% credible intervals.

		Simulated ϵ_{ij}'s: AR(1)			Simulated ϵ_{ij}'s: AR(2)		
		Analysis assuming			Analysis assuming		
		AR(2)	AR(1)	AR(0)	AR(2)	AR(1)	AR(0)
	True	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage
$(\Sigma_\beta)_{11}$	125.00	126.13, 0.98	126.78, 0.97	124.32, 0.98	124.39, 0.97	123.11, 0.98	125.52, 0.98
$(\Sigma_\beta)_{22}$	0.03	0.03, 0.99	0.03, 0.99	0.03, 0.98	0.03, 0.97	0.03, 0.98	0.03, 0.98
$(\Sigma_\beta)_{33}$	0.03	0.03, 0.98	0.03, 0.99	0.03, 0.97	0.03, 0.98	0.03, 0.97	0.03, 0.99
$(\Sigma_\beta)_{12}$	−1.00	−1.05, 0.98	−0.97, 0.95	−0.98, 0.97	−0.97, 0.96	−0.93, 0.97	−0.97, 0.97
$(\Sigma_\beta)_{13}$	0.50	0.49, 0.99	0.60, 0.98	0.53, 0.99	0.54, 0.98	0.53, 0.99	0.49, 0.99
$(\Sigma_\beta)_{23}$	−0.01	−0.01, 0.97	−0.01, 0.98	−0.01, 0.98	−0.01, 0.98	−0.01, 0.99	−0.01, 0.98
$(\Sigma_\alpha)_{11}$	0.020	0.021, 1.00	0.020, 0.99	0.059, 0.78	0.021, 0.99	0.036, 1.00	0.058, 0.80
$(\Sigma_\alpha)_{22}$	0.030	0.031, 0.99	0.031, 0.99	0.045, 0.91	0.032, 0.99	0.032, 0.99	0.043, 0.93
$(\Sigma_\alpha)_{12}$	0.005	0.0002, 1.00	0.001, 1.00	−0.007, 0.95	0.001, 0.99	0.0003, 1.00	−0.006, 0.95
$\sigma^2_{\tau_A}$	0.050	0.57, 0.97	0.059, 0.98	0.069, 0.98	0.059, 0.95	0.059, 0.97	0.073, 0.97

This suggests that underspecifying p as zero may result in overestimation of transition parameter prior variances. Though, in practice over-coverage, as we have observed in our simulation study, is of much less concern than under-coverage; investigating the source of it could be a future research topic.

We see in Table 7.3 noticeable differences in the estimates (average of the posterior medians) of σ^2_i's for different p's. In general, an underspecified p leads to overestimation of σ^2_i. We observe a very poor coverage for σ^2_i if we incorrectly analyze a data set by an AR(0) assumption when, in reality, it exhibits serial correlation over time. However, the problem is much less severe for an underspecified p that is positive.

Table 7.3: Simulation study with $n_i = 150$ for all i and $m = 20$: average of 500 posterior medians of the innovation variances; also coverage of 95% credible intervals.

		Simulated ϵ_{ij}'s: AR(1)			Simulated ϵ_{ij}'s: AR(2)		
		Analysis assuming			Analysis assuming		
		AR(2)	AR(1)	AR(0)	AR(2)	AR(1)	AR(0)
	True	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage	Mean, Coverage
σ_1^2	0.34	0.35, 0.96	0.35, 0.96	0.58, 0.11	0.35, 0.96	0.36, 0.95	0.65, 0.02
σ_2^2	1.12	1.12, 0.94	1.13, 0.97	1.89, 0.10	1.12, 0.94	1.16, 0.95	2.09, 0.02
σ_3^2	1.75	1.77, 0.95	1.76, 0.95	2.95, 0.09	1.76, 0.94	1.80, 0.94	3.27, 0.03
σ_4^2	0.42	0.41, 0.94	0.42, 0.95	0.71, 0.08	0.42, 0.96	0.43, 0.92	0.78, 0.02
σ_5^2	0.74	0.75, 0.95	0.74, 0.95	1.26, 0.08	0.74, 0.94	0.76, 0.96	1.43, 0.02
σ_6^2	2.06	2.08, 0.96	2.08, 0.94	3.52, 0.09	2.09, 0.95	2.10, 0.94	3.86, 0.03
σ_7^2	1.16	1.16, 0.94	1.16, 0.95	1.95, 0.09	1.16, 0.94	1.18, 0.94	2.19, 0.03
σ_8^2	1.28	1.29, 0.95	1.30, 0.96	2.17, 0.08	1.28, 0.95	1.30, 0.96	2.43, 0.03
σ_9^2	0.16	0.16, 0.94	0.16, 0.94	0.27, 0.08	0.16, 0.93	0.16, 0.95	0.30, 0.02
σ_{10}^2	0.77	0.78, 0.97	0.79, 0.95	1.32, 0.08	0.78, 0.95	0.79, 0.95	1.44, 0.03
σ_{11}^2	0.04	0.04, 0.94	0.04, 0.95	0.06, 0.07	0.04, 0.95	0.04, 0.94	0.07, 0.03
σ_{12}^2	0.03	0.03, 0.95	0.03, 0.96	0.06, 0.09	0.03, 0.94	0.03, 0.95	0.07, 0.02
σ_{13}^2	0.91	0.92, 0.96	0.92, 0.95	1.55, 0.09	0.91, 0.95	0.93, 0.95	1.70, 0.04
σ_{14}^2	1.96	1.99, 0.94	1.95, 0.95	3.36, 0.06	1.96, 0.96	1.99, 0.94	3.62, 0.04
σ_{15}^2	0.32	0.32, 0.96	0.33, 0.95	0.55, 0.07	0.33, 0.95	0.33, 0.94	0.61, 0.03
σ_{16}^2	2.02	2.03, 0.97	2.03, 0.94	3.40, 0.07	2.04, 0.96	2.05, 0.95	3.85, 0.03
σ_{17}^2	0.89	0.90, 0.94	0.89, 0.94	1.52, 0.09	0.89, 0.95	0.91, 0.94	1.68, 0.04
σ_{18}^2	0.90	0.90, 0.94	0.90, 0.94	1.53, 0.09	0.91, 0.96	0.92, 0.96	1.70, 0.04
σ_{19}^2	0.82	0.83, 0.93	0.84, 0.95	1.41, 0.07	0.82, 0.96	0.84, 0.94	1.54, 0.04
σ_{20}^2	2.89	2.92, 0.93	2.92, 0.94	4.86, 0.10	2.93, 0.96	2.99, 0.95	5.47, 0.03

7.4.2 *Scenario 3*

Table 7.4 refers to the results for the population regression coefficients and the AR parameter for *Scenario 3*. We see that Model G performs well with respect to all but one parameter: the average of the posterior means for each parameter except μ_γ is close to the true parameter value, and the corresponding coverage probabilities are all reasonably close to the nominal 0.95. When $\omega = 0.90$, we see underestimation and under coverage for μ_γ. We can explain this fact by noting that the average of the posterior means for each γ_i are expected to be close to zero for profiles that originate for Population A, which, in turn, leads to underestimation of the population counterpart μ_γ. Note that if we would model this data set using our flexible methodology, μ_γ would represent only the profiles that originate from Population G, and hence, as we observed in *Scenarios 1* and *2*, underestimation for μ_γ would not occur, and coverage for μ_γ would be close to the nominal 0.95; these facts are also evident from the simulation results for $\omega = 0.95$ (i.e., fewer abrupt profiles than for the case $\omega = 0.90$): the average of the posterior means for μ_γ which is 2.94 is closer to the true value 3.00, and also the coverage which is 0.92 is closer to the nominal 0.95.

Table 7.4: Simulation study assuming an AR(1) within-individual noise, and with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means of the population regression coefficients and the AR parameters; also coverage of 95% credible intervals.

		Simulated $\omega = 0.90$	Simulated $\omega = 0.95$
		Analysis using Model G	Analysis using Model G
	True	Mean, Coverage	Mean, Coverage
μ_0	244.00	$244.33, 0.95$	$244.51, 0.96$
μ_1	0.50	$0.49, 0.93$	$0.49, 0.94$
μ_2	-0.75	$-0.78, 0.91$	$-0.78, 0.92$
μ_γ	3.00	$2.88, 0.86$	$2.94, 0.92$
μ_τ	4.00	$4.04, 0.92$	$4.02, 0.95$
μ_{τ_A}	4.50	–	–
ϕ	0.70	$0.71, 0.93$	$0.71, 0.93$

Table 7.5: Simulation study assuming an AR(1) within-individual noise, and with $n_i = 150$ for all i and $m = 20$: average of 500 posterior means (medians for the variance parameters) of the variances and covariances (Σ_β and Σ_α) in the priors for the random regression coefficients; also coverage of 95% credible intervals.

		Simulated $\omega = 0.90$	Simulated $\omega = 0.95$
		Analysis using Model G	Analysis using Model G
	True	Mean, Coverage	Mean, Coverage
$(\Sigma_\beta)_{11}$	125.00	$123.65, 0.99$	$123.38, 0.98$
$(\Sigma_\beta)_{22}$	0.03	$0.03, 0.98$	$0.03, 0.97$
$(\Sigma_\beta)_{33}$	0.03	$0.03, 0.99$	$0.03, 0.99$
$(\Sigma_\beta)_{12}$	-1.00	$-0.95, 0.96$	$-0.94, 0.97$
$(\Sigma_\beta)_{13}$	0.50	$0.57, 0.99$	$0.57, 0.99$
$(\Sigma_\beta)_{23}$	-0.01	$-0.01, 0.99$	$-0.01, 0.99$
$(\Sigma_\alpha)_{11}$	0.020	$0.114, 0.69$	$0.067, 0.82$
$(\Sigma_\alpha)_{22}$	0.030	$0.054, 0.62$	$0.043, 0.78$
$(\Sigma_\alpha)_{12}$	0.005	$-0.051, 0.68$	$-0.024, 0.83$
$\sigma^2_{\tau_A}$	0.050	–	–

Simulation results for Σ_β and Σ_α are summarized in Table 7.5. Like *Scenarios 1* and *2*, we see that coverage probabilities for the elements of Σ_β are all close to 0.99 (over-coverage). Now, since Σ_α takes into account both abrupt and gradual transitions, large variabilities among the γ_i's and τ_i's are expected – to explain why we expect large variability among the γ_i's, consider one set of simulated γ_i's in *Scenario 3a* (i.e., $\omega = 0.90$): 21.45, 0.00, 19.34, 17.10, 21.24, 18.63, 21.30, 0.00, 0.00, 0.00, 18.28, 17.56, 18.75, 14.43, 14.78, 17.64, 19.79, 21.16, 18.13, 16.18. For Model G, $(\Sigma_\alpha)_{11}$ represents the variability among all the γ_i's including the zeros that originate from Population A. Now, in the presence of those

Table 7.6: Simulation study assuming an AR(1) within-individual noise, and with $n_i = 150$ for all i and $m = 20$: average of 500 posterior medians of the innovation variances; also coverage of 95% credible intervals.

		Simulated $\omega = 0.90$	Simulated $\omega = 0.95$
		Analysis using Model G	Analysis using Model G
	True	Mean, Coverage	Mean, Coverage
σ_1^2	0.34	0.35, 0.97	0.35, 0.94
σ_2^2	1.12	1.14, 0.95	1.12, 0.94
σ_3^2	1.75	1.78, 0.95	1.76, 0.96
σ_4^2	0.42	0.42, 0.95	0.42, 0.96
σ_5^2	0.74	0.76, 0.94	0.74, 0.94
σ_6^2	2.06	2.08, 0.95	2.08, 0.95
σ_7^2	1.16	1.16, 0.94	1.16, 0.93
σ_8^2	1.28	1.29, 0.93	1.27, 0.93
σ_9^2	0.16	0.16, 0.95	0.16, 0.96
σ_{10}^2	0.77	0.78, 0.96	0.77, 0.94
σ_{11}^2	0.04	0.04, 0.96	0.04, 0.95
σ_{12}^2	0.03	0.03, 0.96	0.03, 0.94
σ_{13}^2	0.91	0.92, 0.96	0.92, 0.95
σ_{14}^2	1.96	1.97, 0.94	1.96, 0.95
σ_{15}^2	0.32	0.33, 0.96	0.32, 0.96
σ_{16}^2	2.02	2.02, 0.95	2.05, 0.95
σ_{17}^2	0.89	0.90, 0.95	0.90, 0.96
σ_{18}^2	0.90	0.90, 0.94	0.91, 0.95
σ_{19}^2	0.82	0.83, 0.95	0.83, 0.95
σ_{20}^2	2.89	2.93, 0.97	2.91, 0.96

zeros, we have large variability in the γ_i's, and therefore, the actual $(\Sigma_\alpha)_{11}$ which excludes zeros from Population A would be overestimated if we analyze this data set using Model G. Note that if we would use our flexible methodology to analyze this data set, $(\Sigma_\alpha)_{11}$ would represent the variability among the non-zero γ_i's only, and therefore, we would not expect overestimation. Now, recall the true values of $\boldsymbol{\mu}_\alpha = (\mu_\gamma, \mu_\tau)'$, μ_{τ_A}, $\sigma^2_{\tau_A}$ and Σ_α from Section 7.2. Since τ_i's are simulated from $\mathcal{N}(\mu_{\tau_A}, \sigma^2_{\tau_A})$ and $\mathcal{N}_2(\boldsymbol{\mu}_\alpha, \Sigma_\alpha)$ for Populations A and G, respectively, we expect more variability among the τ_i's as well. However, we do not expect this variability to be as large as that for the γ_i's, since μ_{τ_A} and μ_τ, as well as $\sigma^2_{\tau_A}$ and $(\Sigma_\alpha)_{22}$ are not very different. In our simulation study, large variabilities among the γ_i's and τ_i's are indeed reflected through the overestimation for each of the variance parameters $(\Sigma_\alpha)_{11}$ and $(\Sigma_\alpha)_{22}$: the true values of $(\Sigma_\alpha)_{11}$ and $(\Sigma_\alpha)_{22}$ are 0.020 and 0.030, respectively, whereas, for $\omega = 0.90$, the averages of the posterior medians for these two parameters are 0.114 and 0.054, respectively. Note that because of the overestimation, we see low coverage probabilities for each of those two parameters. We also see that the average of the posterior medians for each of $(\Sigma_\alpha)_{11}$ and $(\Sigma_\alpha)_{22}$ gets closer to the true value as ω approaches 1.00 (e.g., the true value of $(\Sigma_\alpha)_{11}$ is 0.020, whereas the average of the posterior medians for $(\Sigma_\alpha)_{11}$ is 0.114 when $\omega = 0.90$, and is 0.067 when $\omega = 0.95$).

We see in Table 7.6 that the average of the posterior medians for each σ^2_i is close to the true parameter value, and coverage probabilities are all close the the nominal 0.95. That is, misspecifying the model as Model G, when in reality both populations A and G exist, does not affect the estimates of the innovation variances σ^2_i's.

7.5 Discussion

Of course, our simulation study cannot be used to demonstrate the efficiency of our methodology in all cases. They are based on only a small number of models. *Scenarios 1* and *2*, however, reveal that the proposed methodology can perform well with respect to the population regression coefficients even for a misspecified correlation structure for the ϵ_{ij}'s. Also, since the mean of the mixture probabilities over the simulated data sets agrees closely with the true value $\omega = 0.50$ in each of those two scenarios, we can infer that our methodology can pick up both types of transitions adequately, at least for the two scenarios we considered. However, an underspecified model with respect to the order of the AR process

tends to overestimate the innovation variances and give biased estimates for other variance parameters ($(\Sigma_\alpha)_{11}$, $(\Sigma_\alpha)_{22}$ and $\sigma^2_{\tau_A}$) in the priors of the random regression coefficients.

Simulation results for *Scenarios 3* demonstrate that an incorrect choice of the model as Model G, when in reality both populations A and G exist, affects the estimate of μ_γ: in the presence of Population A, the true μ_γ describes Population G only, for whose the γ_i's are positive; yet, by mistakenly ignoring the presence of A, the μ_γ in Model G necessarily incorporates the profiles with $\gamma_i = 0$ that arise from Population A. This type of misspecification also tends to overestimate $(\Sigma_\alpha)_{11}$ and $(\Sigma_\alpha)_{22}$ as described in Section 7.4.2. However, the biases observed for all those parameters become smaller as ω approaches 1.00.

Chapter 8

Concluding Remarks and Future Work

The most appealing feature of the bent-cable model may be its greatly interpretable parameters, and that useful information can be obtained at the cost of estimating only five (four in the case of an abrupt change) regression coefficients. Moreover, pooling information from many individuals leads to shrinkage, so that mild to moderate deviations of observed profiles from the assumed stick/cable structure does not hinder model fitting; in contrast, deviations considered mild to moderate can render the regression method infeasible when analyzing single profiles (e.g., see Reynolds and Chiu [63]). Therefore, our extension of single-profile bent-cable regression to model longitudinal data for many units provides a desirable statistical tool to characterize a special type of continuous temporal trend — one showing a change due to a shock, which is common in many areas of science. Moreover, modeling gradual and abrupt transitions for longitudinal data under a single flexible framework makes our methodology applicable in a wide variety of situations. We have also shown how the approach can be further simplified for the important special case of gradual transition only or abrupt transition only.

Since some standard regularity conditions do not hold, frequentist inference is complicated for bent-cable regression, even for a single profile (Chiu et al. [16], Chiu and Lockhart [15]). Our proposed Bayesian approach for inference makes no special regularity assumptions; the trade-off is the need to evaluate high-dimensional integrals. Although computationally intensive, its implementation is straightforward through the use of MCMC

numerical integration. Moreover, in Bayesian inference, the full behavior of a parameter can be readily investigated via its posterior distribution, rather than relying on the parameter estimator's asymptotic distribution that may be far off from the actual distribution in a finite-sample setting.

We mentioned some cautionary remarks and potential future research topics in various places in this thesis. In addition to adding a few more remarks, we summarize those in Sections 8.1 and 8.2, respectively. We present in Section 8.3 a brief description of journal articles out of this thesis.

8.1 Cautionary Remarks

Because of the flexibility and appealing features (see above) of our methodology, it could serve as a powerful statistical tool in analyzing shock-through data, which may arise in many areas such as biological, medical, health and environmental applications. However, some caution is required for the following reasons.

1. The bent-cable methodology is intended for data which exhibit only one transition period over time. Note that our methodology can be used for multiple transition periods by separating the data into portions so that each portion has only one transition period, although it is unclear how one should combine the inference from the individual portions.

2. As described in Section 4.6, the flexible mixture methodology should be used when there is strong reason to believe that the sample potentially comes from the two populations mentioned above. Diffuse priors may result in a computational breakdown in constructing a Markov chain. However, this problem is irrelevant to the special cases of the mixture methodology: Models A and G.

3. Our methodology is intended for only stationary AR(p), $p \geq 0$, process for within-individual noise (Section 3.2.2). Nonstationarity can result if (a) Assumption **A1** is relaxed to incorporate non-constant variance, or (b) **A1** holds, but at least one root of the AR polynomial $\phi(w) = 1-\phi_1 w-\phi_2 w^2+\ldots+\phi_p w^p$ lies inside the unit circle. Note that in case of (b), forcing stationarity by specifying restrictive priors for $\boldsymbol{\phi}$ may result

in poor mixing (see Section 6.1.3). However, the rat data analysis (Section 6.1.3) and the simulation study (Section 7.4) suggested that under **A1** but a misspecified correlation structure for the within-individual noise, the proposed methodology can perform well with respect to characterizing the populations. Therefore, if the main goal is to make inference for the populations, one may assume AR(0) to analyze a data set for which **A1** is reasonable but exhibits nonstationarity. Proper handling of nonstationary processes could be a topic for future research.

4. Slow-mixing may result from a poor choice of starting values $\Theta^{(0)}$ for a Markov chain. Therefore, starting values may need to be chosen carefully to avoid lengthy runs.

5. The flexible methodology compared it its special cases – Models A and G – may require relatively long Markov chains to overcome poor mixing and to achieve convergence to the stationary distribution.

8.2 Future Work

There is scope to extend the mixed bent-cable regression and its special cases (Models A and G) to a more general framework, presented as follows.

1. One possibility is to incorporate non-constant variance among the repeated measurements (a type of nonstationarity) at Level 1.

2. A more general time series model (e.g., ARMA, CAR) could be employed to account for serial correlation.

3. Extension of our methodology to incorporate spatial effects through, for example, a conditional autoregressive model (Pettitt et al. [56]) could be an area of future research; this would be useful for the CFC study and others that involve spatial data.

4. The framework described in this thesis can be extended to account for different $\boldsymbol{\beta}_i$'s for the two populations, i.e. when intercept and slope parameters could behave differently between populations.

5. Some profiles in a particular data set may follow neither a bent cable nor broken stick, but exhibit another population trend, for example a strictly linearly increasing or decreasing trend. If we have very few such profiles in the data set, those can be regarded as outliers. Note that shrinkage towards the population is expected to overcome any problem in model fitting that may arise due to their existence. We can exploit this aspect and use our methodology as a diagnostic for outliers if our main goal is to make inference for the populations. However, if our main goal is to make inference at the individual level, we should take into account those profiles as coming from another population, that is, we would consider samples potentially coming from three populations: Populations A and G, and a third, say, L which is characterized by a linear trend over time. We can account for L by taking $\beta_{2i} = 0$ in the bent-cable function (3.2). This extension can be achieved by modeling $\boldsymbol{\beta}_i$ using a mixture of normal distributions as follows:

$$\begin{aligned}\pi(\boldsymbol{\beta}_i|\boldsymbol{\mu}_{\beta_L}, \Sigma_{\beta_L}, \boldsymbol{\mu}_\beta, \Sigma_\beta) = \mathbf{1}(\beta_{2i} = 0)\ (1-\delta)\ \mathcal{N}_2(\boldsymbol{\beta}_{iL}|\boldsymbol{\mu}_{\beta_L}, \Sigma_{\beta_L}) \\ + \mathbf{1}(\beta_{2i} \neq 0)\ \delta\ \mathcal{N}_3(\boldsymbol{\beta}_i|\boldsymbol{\mu}_\beta, \Sigma_\beta)\end{aligned} \tag{8.1}$$

where $\boldsymbol{\beta}_{iL} = (\beta_{0i}, \beta_{1i})'$, and $\boldsymbol{\mu}_{\beta_L}$ and Σ_{β_L} are the parameters that characterize Population L. Note that (8.1) assumes a non-zero probability, $1-\delta$, that $\beta_{2i} = 0$, and a probability δ that $\beta_{2i} \neq 0$.

6. Variation in the individual profiles could be due to both random and systematic components (covariates). This can be accounted for by modeling the individual-specific regression coefficients as a function of covariates at Level 2. This extension is straightforward for Models A and G as long as the covariates enter in a linear fashion (see Davidian and Giltinan [19]). However, since the flexible model involves a mixture distribution, technical difficulties may arise when incorporating covariates, and can be another direction in which to extend our methodology.

7. Developing a diagnostic tool for goodness of fit of our hierarchical mixture model can be another area of future research.

8.3 Publications

As mentioned above, our article based on the application of Model G to the CFC-11 data has been published in *CHANCE* (Khan et al. [44]). We submitted another article describing our methodological contribution. We intend to write two more articles out of this thesis. One is on the computational aspects and software implementation of the MCMC algorithm, possibly in a journal in computational statistics. The other one is based on an application of our methodology to PSA profiles (for prostate cancer patients), possibly in a journal of applied statistics.

References

[1] Aitchison, J. and Brown, J. A. C. (1957). *The Lognormal Distribution*. Cambridge: Cambridge University Press. 19, 28

[2] ALE/GAGE/AGAGE (2008). Advanced Global Atmospheric Gases Experiment. URL: `http://agage.eas.gatech.edu/`. 86

[3] Andrieu, C., Freitas, N. D., Doucet, A. and Jordan, M. I. (2003). An Introduction to MCMC for Machine Learning. *Machine Learning*, **50,** 5–43. 33

[4] Bacon, D.W., Watts, D. G. (1971). Estimating the Transition between Two Intersecting Straight Lines. *Biometrika*, **58,** 525–534. 10

[5] Bellera, C. A., Hanley, J. A., Joseph, L. and Albertsen, P. C. (2008). Hierarchical Changepoint Models for Biochemical Markers Illustrated by Tracking Postradiotherapy Prostate-Specific Antigen Series in Men with Prostate Cancer. *Annals of Epidemiology*, **18,** 270–282. 10

[6] Bellera, C. A., Hanley, J. A., Joseph, L. and Albertsen, P. C. (2008). Detecting Trends in Noisy Data Series: Application to Biomarker Series. *American Journal of Epidemiology*, **167,** 1130–1139. 24

[7] Bennett, J. E., Racine-Poon, A., Wakefield, J. C. (1995). MCMC for Nonlinear Hierarchical Models. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson, D. J. Spiegelhalter), pp. 339-357. London: Chapman and Hall. 26

[8] Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (2008). *Time Series Analysis: Forecasting and Control*, 4^{th} ed. Hoboken, N.J.: Wiley. 6, 16, 18, 24, 25

[9] Carlin, B. P. (1995). Hierarchical Longitudinal Modelling. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 303-319. London: Chapman & Hall. 21, 22, 26

[10] Carter, H. B., Pearson J. D., Metter J., Brant, L. J., Chan, D. W., Andres, R., Fozard, J. L. and Walsh, P. C. (1992). Longitudinal Evaluation of Prostate-Specific Antigen Levels in Men With and Without Prostate Disease. *Journal of the American Medical Association* **267,** 2215–2220. 24

[11] Chib, S. (1993). Bayes Regression with Autoregressive Errors: A Gibbs Sampling Approach. *Journal of Econometrics* **58,** 275–294. 18, 21, 25, 27

[12] Chib, S. and Greenberg, E. (1994). Bayes Inference in Regression Models with ARMA (p, q) Errors. *Journal of Econometrics* **64,** 183–206. 25

[13] Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician* **49,** 327–335. 39

[14] Chiu, G. (2009). bentcableAR: Bent-Cable Regression for Independent Data or Autoregressive Time Series. R package version 0.2.1. URL: `http://CRAN.R-project.org/package=bentcableAR`. 13, 22

[15] Chiu, G. and Lockhart, R. (in press). Bent-Cable Regression with Autoregressive Noise. *The Canadian Journal of Statistics*. 4, 6, 10, 13, 18, 46, 112

[16] Chiu, G., Lockhart, R., Routledge, R. (2006). Bent-Cable Regression Theory and Applications. *Journal of the American Statistical Association*, **101**, 542–553. 3, 4, 6, 10, 12, 15, 46, 112

[17] Connett, R. J., Pearce, F. J. and Drucker, W. R. (1986). Scaling of Physiological Responses: A new Approach for Hemorrhage Shock. *American Journal of Physiology (Regulatory and Integrative Comparative Physiology)* **250,** R951–R959. 1

[18] Cryer, J. D. (1986). *Time Series Analysis*. Boston: Duxbury Press. 24

[19] Davidian, M., Giltinan, D. M. (1995). *Nonliner Models for Repeated Mesurement Data*. New York: Chapman and Hall. 9, 14, 21, 25, 26, 37, 40, 115

[20] Dominicus, A., Ripatti, S., Pedersen, N. L., Palmgren, J. (2006). Modelling Variability in Longitudinal Data Using Random Change Point Models. *Research Report 2006: 1*, Mathematical Statistics, Stockholm University. URL: `http://www2.math.su.se/matstat/reports/seriea/2006/rep1/report.pdf`. 6, 12, 21

[21] Dominicus, A., Ripatti, S., Pedersen, N. L., Palmgren, J. (2008). A Random Change Point Model for Assessing Variability in Repeated Measures of Cognitive Function. *Statistics in Medicine* **27,** 5786–5798. 10, 26

[22] Eastham, J. A., Riedel, E., Scardino, P. T., Shike, M., Fleisher, M., Schatzkin, A., Lanza, E., Latkany, L. and Begg, C. B. (2003). Variation of Serum Prostate-Specific Antigen Levels: An Evaluation of Year-to-Year Fluctuations. *Journal of the American Medical Association* **289,** 2695–2700. 24

[23] Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications*. London: Chapman and Hall. 12

[24] Fitzmaurice, G. M., Laird, N. M., Ware, J. W. (2004). *Applied Longitudinal Analysis*. New Jersey: Wiley. 18, 25

[25] Gelman, A. (1992). Iterative and Non-Iterative Simulation Algorithms. In *Computer Science and Statistics* (eds H. J. Newton), pp. 433–438. Fairfax Station: Interface Foundation of North America. 39

[26] Gelman, A. (1995). Inference and Monitoring Convergence. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson, D. J. Spiegelhalter), pp. 131–143. London: Chapman and Hall. 44

[27] Gelman, A. (2006). Prior Distributions for Variance parameters in Hierarchical Models (Comment on an Article by Brown and Draper). *Bayesian Analysis*, **1,** 515–534. 21

[28] Gelman, A., Carlin, J. B., Stern, H. S. and Rubin, D. B. (2004). *Bayesian Data Analysis*. London: Chapman and Hall. 44

[29] Gelman, A., Rubin, D. B. (1992). A Single Series from the Gibbs Sampler Provides a False Sense of Security. In *Bayesian Statistics 4* (eds J. M. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith), pp. 625–631. Oxford: Oxford University Press. 43

[30] Gelman, A., Rubin, D. B. (1992). Inference from Iterative Simulation using Multiple Sequences. *Statistical Science*, **7,** 457–472. 43, 44

[31] Geman, S., Geman, D. (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **6,** 721–241. 39

[32] Geyer, C. J. (1992). Practical Markov Chain Monte Carlo. *Statistical Science*, **7,** 473–511. 43

[33] Ghude, S. D., Jain, S. L., and Arya, B. C. (2009). Temporal Evolution of Measured Climate Forcing Agents at South Pole, Antarctica. *Current Science*, **96,** 49–57. 99

[34] Gilks, W. R. (1995). Full Conditional Distributions. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson, D. J. Spiegelhalter), pp. 75–88. London: Chapman and Hall. 30, 39

[35] Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (1995). Introducing Markov Chain Monte Carlo. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson, D. J. Spiegelhalter), pp. 1–19. London: Chapman and Hall. 39, 43

[36] Gilks, W. R., Richardson, S., Spiegelhalter, D. J. (1996). Markov Chain Monte Carlo in Practice. (eds W. R. Gilks, S. Richardson, D. J. Spiegelhalter). London: Chapman and Hall. 42

[37] Givens, G. H., Hoeting, J. A. (2005). *Computational Statistics*, New Jersey: John Wiley & Sons. 39, 40, 42, 43, 45

[38] Gordon, C. J. (2001). The Therapeutic Potential of Regulated Hypothermia. *Emergency Medicine Journal* **13,** 167–174. 73

[39] Hall, C. B., Ying, J., Kuo, L. and Lipton, R. B. (2003). Bayesian and Profile Likelihood Change Point Methods for Modeling Cognitive Function Over Time. *Computational Statistics and Data Analysis* **42,** 91–109. 4

[40] Hastings, W. K. (1970). Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, **57,** 97–109. 38

[41] Jacqmin-Gadda, H., Commenges, D., Dartigues, J. (2006). Random Changepoint Model for Joint Modeling of Cognitive Decline and Dementia. *Biometrics*, **62,** 254–260. 4, 12

[42] Johnson, T. D., Taylor, J. M. G., Haken, R. K. T. and Eisbruch, A. (2005). A Bayesian Mixture Model Relating Dose to Critical Organs and Functional Complication in 3D Conformal Radiation Therapy. *Biometrics*, **6,** 615–632. 63

[43] Kelly, G.E., Lindsey, J. K. and Thin A. G. (2004). Models for Estimating the Change-Point in Gas Exchange Data. *Physiological Measurement* **25,** 1425–1436. 10

[44] Khan, S. A., Chiu, G. and Dubin, J. A. (2009). Atmospheric Concentration of Chloroflurocarbons: Addressing the Global Concern with the Longitudinal Bent-Cable Model. *CHANCE* **22(3),** 8–17. 84, 116

[45] Kiuchi, A. S., Hartigan, J. A., Holford, T. R., Rubinstein, P. and Stevens, C. E. (1995). Change Points in the Series of T4 Counts Prior to Aids. *Biometrics* **51,** 236–248. 4, 10

[46] Laird, N. M., Ware, J. H. (1982). Random Effects Models for Longitudinal Data. *Biometrics*, **38,** 963–974. 5, 9

[47] Lang, N., Carlin, B. P. and Helfand, A. E. (1992). Hierarchical Bayes Models for the Progression of HIV Infection Using Longitudinal CD4 T-Cell Numbers. *Journal of the American Statistical Association* **87,** 615–626. 4, 6, 10

[48] Lindley, D. (1970). The Estimation of Many Parameters. In *Foundation of Statistical Inference* (eds Godambe, V. P., Sprott, D. A.). Toronto: Holt, Rinehart and Winston. 22

[49] Lindstrom, M. J., and Bates, D. M. (1990). Nonlinear Mixed-Effects Models for Repeated-Measures Data. *Biometrics* **46,** 673–687. 12

[50] McKnight, P. E., Figueredo, A. J., Sidani, S. (2007). *Missing Data: A Gentle Introduction*. New York: Guilford Press. 90

[51] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., Teller, E. (1953). Equations of State Calsulations by Fast Computing Machine. *Journal of Chemical Physics*, **21,** 1087–1091. 38, 39

[52] Morrell, C. H., Pearson, J. D., Carter, H. B., Brant, L. J. (1995). Estimating Unknown Transition Times Using a Piecewise Nonlinear Mixed-Effects Model in Men with Prostate Cancer. *Journal of the American Statistical Association*, **90,** 45–53. 6, 11

[53] NASA (2009). Ozone Hole Watch – Ozone Facts: What is the Ozone Hole? URL: `http://ozonewatch.gsfc.nasa.gov/facts/hole.html`. 99

[54] NOAA/ESRL Global Monitoring Division (2010). Earth System Research Laboratory: Global Monitoring Division. URL: `http://www.esrl.noaa.gov/gmd/`. 86

[55] NOAA/ESRL Global Monitoring Division (2010). HATS In Situ Monitoring Program. URL: `http://www.esrl.noaa.gov/gmd/hats/insitu/insitu.html#history`. 86

[56] Pettitt, A. N., Weir, I. S. and Hart, A. G. (2002). A Conditional Autoregressive Gaussian Process for Irregularly Spaced Multivariate Data With Application to Modelling Large Sets of Binary Data. *Statistics and Computing*, **12,** 353–367. 25, 114

[57] Piegorsch, W. W., Bailer, A. J. (2005). *Analyzing Environmental Data*. New Jersey: John Wiley & Sons. 9

[58] Pinheiro, J. C., Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. New York: Springer-Verlag. 18, 24, 25

[59] Plummer, M., Best, N., Cowles, K. and Vines, K. (2009). coda: Output Analysis and Diagnostics for MCMC. R package version 0.13-4. URL: `http://CRAN.R-project.org/package=coda`. 44, 46

[60] Prinn, R.G., Weiss, R.F., Fraser, P.J., Simmonds P.G., Cunnold, D.M., Alyea, F.N., O'Doherty, S., Salameh, P., Miller, B.R., Huang, J., Wang, R.H.J., Hartley, D.E., Harth, C., Steele, L.P., Sturrock, G., Midgley, P.M., McCulloch, A. (2000). A History of Chemically and Radiatively Important Gases in Air deduced from ALE/GAGE/AGAGE, *Journal of Geophysical Research*, **105 (D14),** 17751–17792. 86, 87

[61] R Development Core Team (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL: `http://www.R-project.org`. 13, 22, 45

[62] Reynolds, P. S., Barbee, R. W., Skaflen, M. S., and Ward, K. R. (2007). Low-Volume Resuscitation Cocktail Extends Survival After Severe Hemorrhagic Shock. *Shock* **28,** 45-52. 73

[63] Reynolds, P. S. and Chiu, G. S. (submitted). Early Warning of Hemorrhage-Induced Hypothermia: Estimating the Thermoregulatory Transition Using Bent Cable Regression. 1, 74, 112

[64] Roberts, G. O. (1995). Markov Chain Concepts Related to Sampling Algorithms. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson, D. J. Spiegelhalter), pp. 45–57. London: Chapman and Hall. 34, 39

[65] Ruppert, D., Wand, M. P. and Carroll, R. J. (2003). *Semiparametric Regression*. Cambridge: Cambridge University Press. 4, 12

[66] Shimizu, K. (1993). A Bivariate Mixed Lognormal Disribution with an Analysis of Rainfall Data. *Journal of Applied Meteorology* **32,** 161–171. 20, 29

[67] Singer, J. D., Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press. 5

[68] Slate, E. H., Turnbull, B. W. (2000). Statistical Models for Longitudinal Biomarkers of Disease Onset. *Statistics in Medicine*, **19,** 617–637. 4, 10

[69] Smith, A. F. M., Roberts, G. O. (1993). Bayesian Computation Via Gibbs Sampler and Related Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society. Series B*, **55,** 3–23. 37, 40

[70] Song, P. X.-K. (2007). *Correlated Data Analysis: Modeling, Analytics, and Applications*. New York: Springer. 21

[71] Spiegelhalter, D. J., Thomas, A., Best, N. G., Gilks, W. R., Lunn, D. (1994, 2003). *BUGS: Bayesian Inference using Gibbs Sampling*, MRC Biostatistics Unit, Cambridge, England. URL: `http://www.mrc-bsu.cam.ac.uk/bugs/`. 21, 27

[72] The Columbia Encyclopedia (2000). Chlorofluorocarbons. *The Columbia Encyclopedia*, 6^{th} ed. New York: Columbia University Press. URL: `http://www.answers.com/library/Columbia+Encyclopedia-letter-1C-first-1051`. 3, 85

[73] The Columbia Encyclopedia (2000). Montreal Protocol. *The Columbia Encyclopedia*, 6^{th} ed. New York: Columbia University Press. URL: `http://www.answers.com/library/Columbia+Encyclopedia-letter-1M-first-1401`. 3, 85

[74] Tierney, L. (1991). Exploring Posterior Distributions using Markov Chains. In *Computer Science and Statistics: Proc.* 23^{rd} *Symp. Interface* (eds E. Keramidas), pp. 563–570. Fairfax Station: Interface Foundation. 39

[75] Tishler, A. and Zang, I. (1981). A New Maximum Likelihood Algorithm for Piecewise Regression. *Journal of the American Statistical Association*, **76,** 980–987. 10

[76] University of Cambridge (1998). The Ozone Hole Tour: Part III. The Science of the Ozone Hole. URL: `http://www.atm.ch.cam.ac.uk/tour/part3.html`. 85

[77] U.S. Environmental Protection Agency (2009). Ozone Layer Depletion: Ozone Depletion Glossary. URL: `http://www.epa.gov/ozone/defns.html`. 85

[78] U.S. Environmental Protection Agency (2009). Ozone Layer Depletion – Science: Class I Ozone-depleting Substances. URL: `http://www.epa.gov/ozone/science/ods/classone.html`. 85

[79] U.S. Environmental Protection Agency (2008). Ozone Science: The Facts Behind the Phaseout. URL: `http://www.epa.gov/ozone/science/sc_fact.html`. 84

[80] Verbeke, G., Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data*. New York: Springer. 9

[81] Vonesh, E. F., Chinchilli, V. M. (1997). *Linear and NONlinear Models for the Analysis of Repeated Measurements*. New York: Marcel Dekker. 25

[82] Wakefield, J. (1996). The Bayesian Analysis of Population Pharmacokinetic Models. *Journal of the American Statistical Association*, **91,** 62–75. 24

[83] Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian Analysis of Linear and Non-linear Population Models by using the Gibbs Sampler. *Applied Statistics* **43,** 201–221. 22, 26, 27, 37

[84] Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: Wiley. 25