State Space Models and Filtering Methods in Longitudinal Studies

by

Daniel Y.T. Fong

A thesis
presented to the University of Waterloo
in fulfilment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 1997

0-612-22201-2

The University of Waterloo requires the signatures of all persons using or photocopying this thesis. Please sign below, and give address and date.

Abstract

The main objective of this dissertation is to study the use of state space models and filtering methods in tackling several fundamental issues in longitudinal studies involving multiple subjects. These include serial dependence of a subject's responses that come naturally from time, inter-subject heterogeneity, missing values and measurement errors in subjects' responses, and non-stationary process drifts. We consider both repeated measure problems and problems involving event histories, and in particular, recurrent events. Several classes of models are introduced and filtering methods developed to implement parameter estimation. Properties of the models and methods are examined. We consider two sets of data for illustrations: a dataset from automobile manufacturing (repeated multivariate measurements), and a set of small bowel motility data (recurrent events).

We consider a class of general state space models and give a review of some common sub-models and the available tools for statistical inference. We point out the need for more efficient estimation for handling missing values and measurement errors, a careful understanding of different types of random effects models, and a tractable likelihood inference procedure.

We first discuss methods of estimating the variation in product quality characteristics measured at several stages in a manufacturing process. By determining which stages contribute most to variation one can focus variation reduction activities more effectively. A multivariate Gaussian Markov process is used to model the variation in characteristics. Methods that deal with measurement error and missing data are introduced through a state space formulation.

Then, we differentiate random effects models for recurrent events into autocorrelated and dynamic random effects models. Their similarities and key differences are discussed

in the case of Gaussian models. Numerical comparisons are provided by using the small bowel motility data and cases when the models might be used are discussed.

Thirdly, we study a dynamic proportional hazards random effects model for recurrent events with non-informative right censoring. Subject heterogeneity and potential non-stationary process drifts are handled by repeatedly updating an initial frailty as more recurrence times are observed. An arbitrary baseline hazard together with an external time-dependent covariate process are allowed. The full model is actually a non-Gaussian state space model with a multiplicative state transition process. Parametric inference on hyperparameters is carried out by maximizing the likelihood function, which can be shown to be numerically tractable. A simulation study is conducted for further insight into the model.

Finally, we conclude this dissertation with some general remarks and point to some potential future research directions.

Acknowledgements

I would like to express my most sincere gratitude to my supervisor, Professor Jerald F. Lawless, for his continuous and invaluable guidance in preparing this thesis.

I am greatly indebted to my friends from Hong Kong as well as all brothers and sisters especially in Waterloo Chinese Christian Fellowship and Kitchener-Waterloo Chinese Alliance Church for their prayers, encouragements and supports. Also, I would like to thank my precious parents and sister for their patience, understanding, and unconditional support throughout the years of my graduate study abroad. More importantly, thanks to the Lord for His continuous guidance and love.

Finally, financial support from the Croucher Foundation Scholarship and research grants from Prof. Jerald F. Lawless are also gratefully acknowledged.

Contents

List of Tables

List of Figures

Chapter 1

Motivation and Examples

1.1 Longitudinal Studies

1.1.1 Introduction

Longitudinal studies often involve analyses of specific dynamic changes of subjects in a group over a certain time period. Longitudinal behaviour can be examined by either monitoring subjects continuously over time, or examining them only at discrete time epochs. A typical dataset in a longitudinal study consists of event occurrence times or repeated measurements for each subject over time. Several recent books (e.g. Andersen *et al.*, 1993; Diggle *et al.*, 1994; Lindsey, 1993) discuss and provide comprehensive coverage of various types of studies.

It is a characteristic of longitudinal studies that measurements or events associated with individuals at different time points are related, i.e. not statistically independent. The main objective of this thesis is to consider the use of dynamic models for representing dependencies and to develop methods of inference for such models. We will consider situations with repeated measurements taken at discrete time points and also situations where events may occur repeatedly to subjects over time. Section 1.2 provides some motivating examples but first in Section 1.1.2, we consider some basic problems and objectives associated with longitudinal studies.

1.1.2 Basic Problems and Objectives

It is possible to have numerous complications in longitudinal studies but there are three basic ones. The most fundamental problem is modelling, because of the time element, the inherent stochastic dependence between a subject's measurements or event history, in particular when previous observations contain information relevant to present and future variates. Modelling dependencies can be basically classified, from Cox (1981), to be observation-driven when dependency is due directly to previous observations and parameter-driven when it is induced by a hidden stochastic process of the parameters. Choice of models will be discussed in Chapter 2.

Another problem is subject heterogeneity. This is usually handled by including observable covariates in models, but very often there remains unexplained variation. This is often called unobservable heterogeneity and it can be effectively handled by using random covariates or random effects with certain distribution assumptions (e.g. Aalen and Husebye, 1991; Pickles and Crouchley, 1994; Hougaard, 1995).

Missing data is another common feature in longitudinal studies (e.g. Little, 1992, 1995; Baker, 1995; Follmann and Wu, 1995). Subjects may drop out during surveillance or have measurements missing intermittently. The presence of missing data has several effects in longitudinal analysis. One is that a simple multivariate analysis for balanced data when we have an equal number of observations for each subject measured at equal time intervals may be made complicated when data are missing. A similar problem arises for missing data in event history analysis (e.g. Lawless and Yan, 1992). A worse complication is when data are not missing completely at random (e.g. Little and Rubin, 1987). For example, in a study of the efficacy of a new drug on lowering blood pressure, patients with higher blood pressure may tend to drop out from the experiment. Ignoring the "reasons" for dropout will give a seemingly high efficacy of the new drug and lead to a biased conclusion.

These are not the only problems in longitudinal studies. Another problem we study is non-stationary process drifts due to interventions across time. There are still other problems which include measurement errors in both responses and covariates, and data collected at irregular time intervals.

However with all these kinds of complications, a major merit of longitudinal studies is that we can differentiate the changes over time within subjects and differences among subjects. Thus two basic objectives in longitudinal analysis are to characterize the degree of heterogeneity across subjects and to assess the effects of covariates at a subject-specific level. Other objectives depend on the type of data at hand. Specifically, with repeated measurements taken at certain fixed discrete time points, we may be interested in characterizing the response profile over time (Diggle *et al.*, 1994) while with recurrent event data, we may be interested in estimation of the mean event recurrence times, prediction of the next event occurrence, and analysis of rates (Lawless, 1995).

Motivating examples which highlight different problems and objectives in longitudinal studies are discussed in the next section.

1.2 Motivating Examples

Three datasets are used to motivate later developments, and illustrate different characteristics and objectives in longitudinal studies. They will be studied in the rest of this dissertation.

1.2.1 Automobile Manufacturing

The first two examples concern processes used in the production of automobiles. In each case, certain important measurements on part of a car are taken at a sequence of several stages of the process. The objective is to determine which stages contribute most to variation in the part, and thus to help reduce variation. Lawless *et al.* (1997) discuss this area in depth. The two datasets are shown in Appendix A.1.

Piston Machining

A piston is used in engines to impart motion by means of a piston-rod. It is a short metallic cylinder which is closed at the top and open at the bottom, fitting closely inside an engine cylinder in which it vibrates up and down, pushing out exhaust on the up-stroke and intaking fuel on the down-stroke. The quality characteristics of interest were four diameters, located at heights of 4 mm, 10 mm, 36.7 mm and 58.7 mm from the bottom of the piston. The diameters were measured after each of four operations in the machining process, the measurements being in millimeters, to a precision of 1 micron (10^{-3} mm). Details of the study can be found in Agrawal *et al.* (1997).

It is clearly important to control the diameter across the body of the piston to ensure a close fit and smooth movements inside the engine cylinder. However, note that the four diameter measurements are obviously correlated and thus separate modelling for each single diameter is likely to be inappropriate. Simultaneous modelling of multiple measurements is preferred to account for the interactions between the measurements. Moreover, at each of the 4 locations on the piston, fewer than 15 distinct measured values occur. Thus, accounting for measurement errors due to heavy rounding of measurements is also desirable.

Our main interest focuses on determining the sources of variation contributing to the diameters at the final stage and the variation transmitted across different process stages. Major factors are (i) serial correlation of measurements across different process stages, (ii) the presence of multiple measurements (the four diameters), and (iii) measurement errors on the diameters.

Door Hanging

We consider an assembly process for rear doors of vehicles. There were seven stages of the process, corresponding to seven operations: (1) the door hang, (2) paint the door, (3) install door hardware, (4) striker installation, (5) striker fit, (6) install seals and chassis,

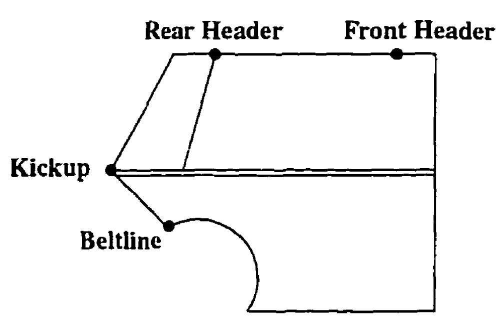


Figure 1.1: Locations of the four flushness deviation measurements of a rear door.

and (7) final fit. The quality characteristics of interest relate to the flushness of the rear door to the surrounding body of the car. This was quantified through four flushness deviation measurements at locations called kickup, beltline, front to header, and rear to header; see Figure 1.1. They were measured after each of the seven stages. A zero measurement at any location means the door is perfectly flush, and positive and negative measurements mean it is too high and too low respectively. Details of the experiment can be found in Hamada and Lawless (1994) and Fong and Lawless (1997).

A major characteristic of the data is that not all measurements are successively taken and around 46% of the data are missing. The missing data may be caused by the difficulty in taking measurements while maintaining the flow of the whole production line but their actual sources are not clear from the manufacturer. We will however assume the data are missing in a random fashion. Moreover, all measurements were taken with a special hand-held tool. Accounting for variation due to the instrument is thus desirable.

Primary objectives are to have rear doors as close to perfectly flush as possible after the final stage and to learn about the origins and transmission of variation. Major factors are (i) serial correlation across process stages, (ii) multiple measurements, (iii) the presence of missing values, and (iv) measurement errors.

These two examples are discussed in some detail in Chapter 3.

1.2.2 Small Bowel Motility

Our small bowel has both absorptive and secretory functions and the muscular activity (motility) of it is vital for gastrointestinal function in humans. In a study described by Aalen and Husebye (1991), nineteen healthy individuals with age ranging from 22 to 50 were monitored for 13 hours and 40 minutes, from 5:45pm in a day to the next morning at

7:25am. At 6:00pm, all individuals were treated with a standardized mixed meal. They then entered a fed state with irregular contractions in the small bowel which is followed by a fasting state with a regular cyclic motility pattern defined by three phases. However, only "phase III" can be easily detected and is thus used to define the fasting cycle which is termed the migrating motor complex (MMC). The first detected phase III is defined as the start of the fasting state and recurrence times of phase III were continuously tracked until the end of the experiment. Please refer to Aalen and Husebye (1991) for a detailed description of the experiment. The data are reproduced in Appendix A.2.

In a closer look at the data, we can see that there are large variations of both within and between subjects MMC periods. Subjects with different ages may have difference in frequency of MMC periods. As the age of subjects or other subject specific information are not recorded, the effects on subject heterogeneity remain unobserved. Also, accounting for time trends or non-stationary drifts of a subject's MMC periods to assess the regularity of MMC is also desirable. Moreover, removing the censored final MMC periods for each subject will lead to estimation bias while treating them as if they were complete will lead to underestimation of the overall mean of MMC period. Thus censored MMC periods have to be handled properly.

Objectives of the study are to model the distribution of recurrence times. Major factors are (i) possible correlation among recurrence times, (ii) subject heterogeneity, (iii) right censoring for the last recurrence time, and (iv) the possibility of time trends or non-stationary process drifts.

This example is discussed in some detail in Chapter 4 and 5.

1.3 Dissertation Plan

We will expand our discussions on the aforementioned topics in longitudinal studies in the coming chapters. As suggested from the title of this dissertation, we will focus on using state space models in capturing desired characteristics of longitudinal data and showing how filtering methods can assist in facilitating statistical inference.

Chapter 2 introduces a general class of statistical models called a general state space model and discusses several of its different common descendants in longitudinal studies. Then a brief survey is given of the available tools for statistical inference with emphasis on filtering methods. At the end of the chapter, we give background and motivation for three specific areas that we will study in more detail in subsequent chapters.

Chapter 3 discusses methods of estimating the variation in product quality characteristics measured in a multi-stage manufacturing process, e.g. the two automobile manufacturing examples in Section 1.2.1. A multivariate Gaussian Markov process is used to model the variation in characteristics. Methods that deal with measurement errors and missing data are introduced through a state space formulation. Estimation of model parameters is developed through a filtering approach and the use of the parametric bootstrap.

In Chapter 4, we identify two different types of Normal-based random effects models for recurrent events which are given the names: autocorrelated and dynamic random effects models. Their similarities and differences are pinpointed and guidelines for their use are provided. The Small Bowel Motility Data is analyzed using the models and filtering methodology.

Chapter 5 studies a dynamic proportional hazards model to account for subject heterogeneity and non-stationary process drifts for times between recurrent events. Parametric inference on hyperparameters is carried out by maximizing the likelihood function via filtering. This is numerically tractable, a property that is not shared by most hazard-based random effects models. Properties of the model and estimation procedures are studied.

The last chapter discusses some further potential research.

Chapter 2

State Space Models and Filtering Methods

2.1 General State Space Models

Before we introduce a general class of state space models and discuss the use of filtering methods in longitudinal studies, we need some notation to describe the anticipated data. We consider the situation where measurements are taken repeatedly on an individual at each of several distinct time points. Suppose we studied N subjects and measurements were taken at n_i time points from subject i ($i = 1, 2, \ldots, N$). Let y_{ij} be a vector of the jth ($j = 1, 2, \ldots, n_i$) set of measurements taken from subject i and x_{ij} be a corresponding vector of measured covariates. This is a standard type of longitudinal data. The automobile manufacturing data in Section 1.2.1 have items (subjects) measured at a given sequence of process stages (indexed by j), so that n_i is a constant. The recurrent event data in Section 1.2.2 can also be described in this way. It has y_{ij} as the jth recurrence time but the number of event recurrences ($= n_i$) varies with different subjects stochastically. Throughout this chapter, we will assume this basic notation and more will be specified if needed. Also, for brevity, indices i and j are assumed to run from 1 to N and 1 to n_i respectively unless otherwise specified.

Now, a general state space model (GSSM) is defined by

1. an observation model for

$$y_{ij} \mid Y_i^{j-1},\ x_{ij},\ z_{ij}$$

   where $Y_i^{j-1} = \{y_{i1}, y_{i2}, \ldots, y_{i,j-1}\}$ denotes the set of all observations of subject i up to and including the $(j-1)$th one, Y_i^0 = null set, and z_{ij}'s, called states, are unobservable random variables whose dynamics follow

2. a transition model for $z_{ij} \mid z_{i,j-1}$.

There are four basic assumptions for GSSMs by which the joint density of y_{ij}'s and z_{ij}'s can be generally written down. They are enumerated as follows.

(A1) The covariate process $\{x_{ij}\}$ is non-stochastic; otherwise we condition on its observed values.

(A2) Responses between different subjects are conditionally independent, i.e.

$$f(y_{1j}, y_{2j}, \ldots, y_{Nj} \mid Y^{j-1}, X^j, Z^j) = \prod_{i=1}^{N} f(y_{ij} \mid Y_i^{j-1}, X_i^j, Z_i^j)$$

where $Y^{j-1} = \{Y_1^{j-1}, \ldots, Y_N^{j-1}\}$, and Z^j, Z_i^j, X^j, X_i^j are similarly defined.

(A3) At occasion j and given all the past responses, Y_i^{j-1}, current responses depend only on the current state and covariates, i.e.

$$f(y_{ij} \mid Y_i^{j-1}, X_i^j, Z_i^j) = f(y_{ij} \mid Y_i^{j-1}, x_{ij}, z_{ij})$$

(A4) The transition model is first order Markovian, i.e.

$$f(z_{ij} \mid Z_i^{j-1}, Y_i^{j-1}) = f(z_{ij} \mid z_{i,j-1}).$$

Note that higher order Markov dependency can be transformed to first order by augmenting z_{ij} by its lagged variables.

Note that the independence assumption of the transition model on past responses can be relaxed and this extension is considered in Chapter 5. Under the model, measurement or response vector y_{ij} is allowed not only to depend on its past observations and some covariates but also on some unobserved effects, possibly due to measurement errors or missing covariates, governed by the transition model. This class of GSSMs is quite general and provides a unifying framework for models in longitudinal studies. However, an example which does not belong to this class will be considered in Chapter 5. This section will present several fruitful classes of commonly used longitudinal models which will be frequently referred to throughout this dissertation. Most of them assume the Normal distribution assumption for the sake of convenience only. It can be replaced by other distributions whenever plausible, as directed by the references cited in the discussion.

2.1.1 Normal (Gaussian) State Space Models

Linear state space models with Normal distribution assumptions can be derived from a GSSM as

$$\begin{aligned} y_{ij} \,|\, Y_i^{j-1}, x_{ij}, z_{ij} &\sim \mathcal{N}(H_j x_{ij} + G_j z_{ij}, \Sigma_j) \\ z_{ij} \,|\, z_{i,j-1} &\sim \mathcal{N}(B_j z_{i,j-1}, Q_j) \end{aligned}$$

where H_j, G_j and B_j are design matrices specified by some unknown parameters. The initial z_{i0} can either be defined as a constant or another independent Normal variate. This kind of model has been popular in time series forecasting (Harvey, 1989). Examples of formulating some time series models into a linear state space form can be found in Lütkepohl (1993). It has also numerous applications in longitudinal studies, e.g. growth curve analysis (Wilson, 1988), longitudinal count data (Jørgensen *et al.*, 1996a, 1996b). Other applications can be found in the books by Jones (1993) and Fahrmeir and Tutz (1994). The model assumes all responses are continuous and unrestricted, possibly after transformation in order to justify the Gaussian distribution assumption. For responses which are discrete (e.g. number of defective items in a batch in quality control), nominal (e.g. type of infection among a number of categories), or ordinal (e.g. test results that are classified as normal, borderline and abnormal) in nature, the Gaussian assumption is far from being reasonable and the following models are usually considered.

2.1.2 Dynamic Generalized Linear Models (DGLMs)

A dynamic generalized linear model has

$$g(\mu_{ij}) = C_{ij} z_{ij} + H_j x_{ij}, \qquad \mu_{ij} = E(y_{ij} \mid Y_i^{j-1}, x_{ij}, z_{ij}),$$
$$z_{ij} \mid z_{i,j-1} \sim \mathcal{N}(B_j z_{i,j-1}, Q_j), \qquad z_{i0} \sim \mathcal{N}(A_{i0}, Q_0)$$

where the design matrix C_{ij} is a function of Y_i^{j-1} and x_{ij}, and g is a monotonic and differentiable link function. It includes the Gaussian linear state space model when g is the identity function and the distribution of $y_{ij} \mid Y_i^{j-1}, x_{ij}, z_{ij}$ is Gaussian. Note that the distribution assumption in the observation model, though not specified above is usually assumed to come from the exponential family. Through this, together with the link function, discrete and categorical responses can be modelled, for example, a Poisson distribution with a logarithm link for counts, or a Multinomial distribution with a logistic link to the marginal or cumulative probabilities for nominal or ordinal responses. Furthermore, for the transition model, other dynamic processes other than the additive and Gaussian assumption are also possible (e.g Jørgensen *et al.*, 1996a; Yue and Chan, 1994).

The ancestral model of DGLM is the dynamic linear model (with g as the identity function) defined by Harrison and Stevens (1976). It was then studied by West *et al.* (1985) through a Bayesian analysis using discounting to get rid of the unknown error variance in the transition model; refer to Section 2.2.3 for more Bayesian methods on the model. Thereafter, applications on longitudinal count data (Harvey and Fernandes, 1989; Singh and Roberts, 1992; Lambert, 1996b, 1996a), competing risks models with discrete duration times (Fahrmeir and Wagenpfeil, 1996), and recurrent event data (Smith and Miller, 1986; Yue and Chan, 1994) were considered. Use of the model in handling random effects and serial correlation in longitudinal studies, especially on recurrent events, has not been yet fully studied. The model is also described in the books by Lindsey (1993) and Fahrmeir and Tutz (1994).

2.1.3 Generalized Linear Models (GLMs)

Diggle *et al.* (1994) and Lindsey (1993) described three extensions of GLMs (McCullagh and Nelder, 1989) for longitudinal studies: namely, marginal, random effects and conditional models (we use conditional model instead of "transition model" as in Diggle *et al.* (1994) to avoid confusion with the transition model in GSSMs). They belong to the class of GSSMs or DGLMs. All of them are defined by a linear regression on the mean of the responses through a known link function g but they have different domains of application.

Marginal models separate the regression of the mean response from the within-subject association. They assume

$$g(\mu_{ij}) = x'_{ij}\beta; \qquad \mu_{ij} = E(y_{ij} \,|\, x_{ij}) \tag{2.1}$$

and the within-subject covariance, $Cov(y_{ir}, y_{is})$ is assumed to be a function of μ_{ir}, μ_{is} and possibly some additional parameters. The model is appropriate when we are interested in population-averaged inference, for example, a study of the average difference between the effects of two treatments in clinical studies. In other words, we are interested in the average behaviour over the whole population at various time points.

Random (Mixed) effects models account for inter-subject heterogeneity by specifying

$$\begin{aligned} g(\mu_{ij}) &= x'_{ij}\beta + w'_{ij}z_i; \qquad \mu_{ij} = E(y_{ij} \mid x_{ij}, z_i) \\ z_i &\sim \mathcal{N}(0, \Sigma) \end{aligned} \tag{2.2}$$

where w_{ij} is usually a subset of the covariate x_{ij}. The z_i's are subject-specific effects assumed to be independent and identically distributed (i.i.d.). This class of models is also called the generalized linear mixed models (GLMMs) (Breslow and Clayton, 1993). Note that given z_i's, the responses y_{ij}'s are independent and thus within-subject association is solely induced by the random effects. These models are appropriate when we are interested in subject-specific effects or in accounting for extra inter-subject variation, perhaps due to missing covariates. There is a huge literature on these models (e.g. see McCulloch, 1997).

Conditional models, unlike (2.1) and (2.2), make the within-subject association explicit in the regression equation as

$$g(\mu_{ij}) = x'_{ij}\beta + \sum_{m=1}^{M} f_m(Y_i^{j-1}; \alpha); \qquad \mu_{ij} = E(y_{ij} \mid Y_i^{j-1}, x_{ij}) \tag{2.3}$$

where f_m's are known functions depending on some unknown parameter α. The conditional variance $Var(y_{ij} \mid Y_i^{j-1}, x_{ij})$ is assumed to be a function of μ_{ij}. Modelling stochastic dependence of a single subject's responses directly, rather than by random effects, is often desirable. A merit of using (2.3) is that all successive conditional probabilities for computing the likelihood function can be written down directly when a distribution is adopted.

Although Models (2.1)–(2.3) stand on different objectives and conceive different structural response behaviour, the fixed effects β from them have the same interpretation when g is the identity function (Diggle *et al.*, 1994). More comparisons are discussed in Diggle *et al.* (1994) and Zeger and Liang (1992). Note however that these models are only basic ingredients on which more useful models can be constructed. For example, we can combine a marginal model with an exponential correlation structure and a random effects model as

$$\begin{aligned} y_{ij} &= x'_{ij}\beta + w'_{ij}b_i + e_{ij}; \quad b_i \sim \mathcal{N}(0, \omega^2), \quad e_{i1} \sim \mathcal{N}(0, \sigma_1^2), \\ e_{ij} &= \phi e_{i,j-1} + \epsilon_{ij}; \qquad \epsilon_{ij} \sim \mathcal{N}(0, \sigma^2), \quad |\phi| < 1 \end{aligned} \tag{2.4}$$

where w_{ij} is a subset of the covariate x_{ij}. The model still falls in the class of GSSM. The b_i is the subject-specific effect and ϕ measures the intra-subject correlation. The initial variance parameter σ_1^2 is usually chosen as 0 or $\sigma^2/(1-\phi^2)$ to give an equilibrium transition process. A major model characteristic is that the marginal correlation between any two responses of a subject gets smaller exponentially as they are further apart which, in the presence of random effects, converges to a non-zero positive constant. This model will be revisited in Chapter 4. References on the model are Wilson (1988), Louis (1988), and Chan and Kuk (1997). In addition, Sutradhar (1990) considers a similar model with nested subject effects.

2.1.4 Frailty Models

Many models involving survival times or times between events are considered in terms of hazard functions (Clayton, 1994). That is, we model y_{ij} by its hazard function and often we employ a proportional hazards model (Cox, 1972)

$$h_{ij}(y_{ij}) = z_{ij} e^{\beta' x_{ij}} h_0(y_{ij}) \quad (2.5)$$

where $h_0(\cdot)$ is called the baseline hazard function. It is the hazard function when $x_{ij} = 0$ and $z_{ij} = 1$. The z_{ij} is often called the frailty because, for example when y_{ij}'s are the recurrence times of a certain circuit failure, susceptibility to failure increases with z_{ij}. One objective in the thesis is to consider dynamic frailties for (2.5). For example, to model inter-subject heterogeneity and non-stationary process drifts, we might define

$$z_{i,j+1} = \psi^{-1} z_{ij} \eta_{ij}; \quad \eta_{ij} \sim Beta(\psi \kappa_{ij}, (1-\psi)\kappa_{ij}) \quad (2.6)$$

where $\kappa_{ij} = \psi \kappa_{i,j-1} + \delta_{ij}$, δ_{ij} is 0 when $i = j$ and 1 otherwise, $\kappa_{i1} = 1 + 1/\omega^2$, $z_{i1} \sim Ga(\frac{1}{\omega^2}, \frac{1}{\omega^2})$ and $Ga(a, b)$ denotes the Gamma distribution with mean a/b and variance a/b^2. Equation (2.5) and (2.6) together define a dynamic frailty model which is clearly a sub-model of GSSMs. It is described in Yue and Chan (1994) and is fully discussed in Chapter 5. The model includes some special sub-models which have been used often in the literature. In particular, when $\omega^2 \to 0$, all survival times become independent and ordinary survival analysis methods (e.g. Lawless, 1982) can be used. When $\psi \to 1$, (2.6) becomes

$$z_{ij} = z_{i1} \sim Ga(\frac{1}{\omega^2}, \frac{1}{\omega^2})$$

which together with (2.5) defines the ordinary Gamma frailty model.

A survey of frailty models on survival and event history analysis is given in a series of review papers by Aalen (1994), Pickles and Crouchley (1994) and Hougaard (1995).

Recently, Petersen *et al.* (1996) constructed frailty models for clustered samples by letting subjects within a cluster share some frailties. For example, in survival analysis of twins, we can have

$$h_i^{(1)}(t) = (z_i^{(0)} + z_i^{(1)})h_0^{(1)}(t) \quad \text{and} \quad h_i^{(2)}(t) = (z_i^{(0)} + z_i^{(2)})h_0^{(2)}(t)$$

where $h_i^{(j)}$ and $h_0^{(j)}$ are the hazard and baseline hazard for the jth ($j = 1, 2$) one of a twin and $z_i^{(k)}$ ($k = 0, 1, 2$) are the frailty variables. Ng and Cook (1997) and Xue and Brookmeyer (1996) provide other recent examples.

The hazard-based models (2.5) are particularly useful in modelling recurrent event data when the covariates x_{ij} are time-dependent, in which case distribution based approaches are hard to use. As in the GLMMs, conditional on z_{ij}, all recurrence times are assumed to be independent for each subject, so they form a renewal process. Use of this kind of proportional hazards models has been quite popular in the literature of longitudinal studies (e.g. Aalen and Husebye, 1991). Non- or semi-parametric analysis for the models are generally pursued through a counting process approach for which details and more references can be found in the book by Andersen *et al.* (1993). For parametric analysis, the likelihood function is often intractable (e.g. Clayton, 1994). A class of dynamic frailty models with a tractable likelihood is studied in Chapter 5.

2.2 Filtering and General State Space Solutions

With reference to the basic objectives of longitudinal studies in the first chapter, we are interested in things like estimating fixed covariates effects, inter-subject variability, intra-subject correlation, etc. All of these can be parametrically modelled into the observation and transition models of a GSSM. Estimation of parameters can be assisted by filtering methods originally proposed by Kalman (1960), and Kalman and Bucy (1961) to estimate the unobserved state z_{ij} based on Y_i^T for some $T > 0$ under a Gaussian linear state space model. In general, special cases are given the names filtering ($T = j$), prediction ($T < j$) and smoothing ($T > j$). The corresponding estimates are called filters, predictors and smoothers. The filtering step evaluates, by Bayes Theorem,

$$f(z_{ij} \,|\, Y_i^j,\, x_{ij}) \propto f(y_{ij} \,|\, Y_i^{j-1},\, x_{ij},\, z_{ij}) f(z_{ij} \,|\, Y_i^{j-1},\, x_{ij}) \tag{2.7}$$

which iterates with the prediction step

$$f(z_{ij} \,|\, Y_i^{j-1},\, x_{ij}) = \int f(z_{ij} \,| z_{i,j-1}) f(z_{i,j-1} \,|\, Y_i^{j-1},\, x_{ij}) dz_{i,j-1} \tag{2.8}$$

to get all the filters and one-step predictors for later computing the smoothers. Note that we have used $f(\cdot)$ as a generic function for the probability density function and distinctions between the random variables referred to are made explicit in the function arguments. Now, the unobserved state z_{ij} is estimated by the smoothing density $f(z_{ij} \,|\, Y_i^T,\, x_{ij})$ computed recursively from $f(z_{ij} \,|\, Y_i^j,\, x_{ij})$ and $f(z_{i,j+1} \,|\, Y_i^j,\, x_{i,j+1})$. A smoothing formula is given by

$$f(z_{ij} \,|\, Y_i^T,\, x_{ij}) = f(z_{ij} \,|\, Y_i^j,\, x_{ij}) \int f(z_{i,j+1} \,|\, Y_i^T,\, x_{i,j+1}) \frac{f(z_{i,j+1} \,|\, z_{ij})}{f(z_{i,j+1} \,|\, Y_i^j,\, x_{i,j+1})} \, dz_{i,j+1};$$

see Kitagawa (1987). For estimation and for prediction of y_{ij}'s, we need to get $f(y_{ij} \,|\, Y_i^{j-1}\, x_{ij})$ by using certain formulas based on the z_{ij}'s, e.g. equation (2.12); see also Figure 2.1.

For Gaussian linear state space models, the celebrated linear Kalman filter (Kalman, 1960; Kalman and Bucy, 1961) can be easily implemented. There are several smoothing algorithms but the classical fixed interval smoothing algorithm can be found in Anderson and Moore (1979). Recently, Koopman (1993) developed a faster and more efficient smoothing algorithm when the Gaussian distribution assumption is not appropriate. For non-Gaussian linear state space models, the Kalman filter still provides the best linear predictor but not necessarily the optimal forecast in the sense of minimizing the mean square errors. In general, with non-Gaussian and nonlinear structure, integrations in (2.7) for computing the normalization constant, and (2.8) are hard to compute mathematically. Various approaches such as piecewise linear approximation of all densities when the dimension of the states is small (in Kitagawa, 1987), Gibbs sampling on the posterior density of the states, use of posterior modes under a Gaussian linear transition model (in Chapter 7 and 8 of Fahrmeir and Tutz, 1994), and estimating functions without distributional assumptions on the observation and transition models (in Naik-Nimbalkar and Rajarshi, 1995) are proposed. More approximate filtering and smoothing methods can be found in the books by Anderson and Moore (1979) and West and Harrison (1997).

The GSSMs provide a unifying framework for many important models used in longitudinal studies. An advantage of using filtering for statistical inference is, because of its recursive nature, the high efficiency in handling data with lots of measurements per subject. Different problems with specific estimation approaches tailored to different subclasses of GSSMs have been emerging in the literature. A main focus of this dissertation is to explore how filtering works for estimation under different types of state space models.

Our main interest is in inference procedures for longitudinal models. We now discuss approaches to estimation of parameters in state space models.

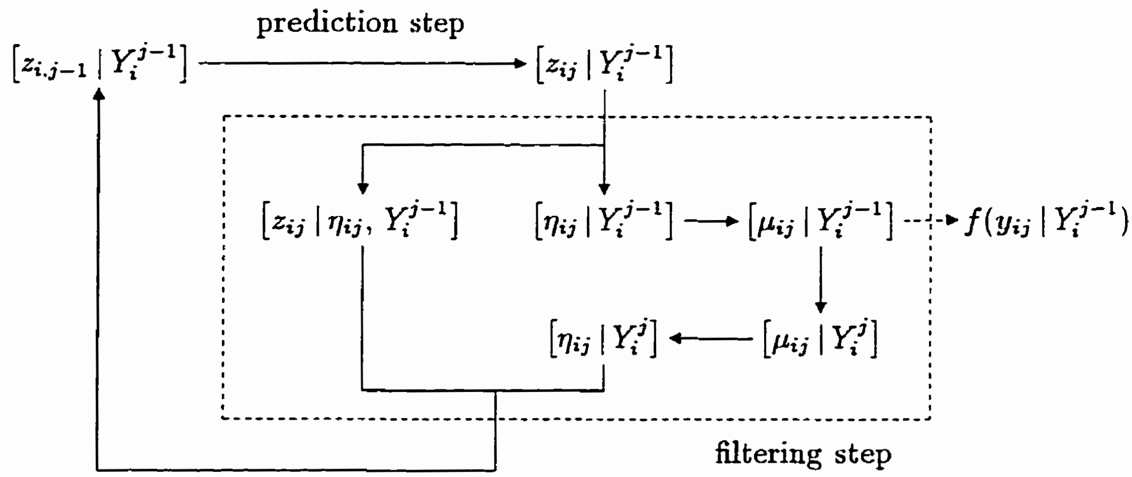


Figure 2.1: A pictorial outline of the Bayesian scheme in West and Harrison (1997) for DGLMs ($\eta_{ij} = g(\mu_{ij})$).

2.2.1 The Expectation-Maximization (EM) algorithm

As the states of a GSSM are unobserved, it is natural to consider the well-known EM algorithm proposed by Dempster *et al.* (1977) to tackle problems with unobserved or missing values. In our applications, observed data refers to $\{y_{ij}\}$ and the "complete" data refers to $\{y_{ij}, z_{ij}\}$. We assume that all covariates x_{ij} in the model are observed. Let θ be a vector of the unknown model parameters. Suppressing the dependence on x_{ij}'s, the log-likelihood based on the complete data is

$$l_c(\theta\,;\, y_{ij}\text{'s},\, z_{ij}\text{'s}) = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{n_i} \left[\log f(y_{ij} \mid Y_i^{j-1},\, x_{ij},\, z_{ij}) + \log f(z_{ij} \mid z_{i,j-1})\right] + \log f(z_{i0}) \right\} \tag{2.9}$$

The EM algorithm is a recursion consisting of a E-step and a M-step. The E-step computes the conditional expectation

$$M(\theta \mid \hat{\theta}^{(k)}) = E(l_c \mid y_{ij}\text{'s};\, \hat{\theta}^{(k)}) \tag{2.10}$$

where $\hat{\theta}^{(k)}$ is the estimate of θ at the kth iteration. The M-step then maximizes $M(\theta \mid \hat{\theta}^{(k)})$ at $\theta = \hat{\theta}^{(k+1)}$, the next iterated estimate of θ. The recursion then continues until convergence. The M-step is usually easy to handle but the E-step is the most critical concern for deciding whether the EM algorithm is applicable. In (2.10), the expectation may involve some functions of the unobserved states. Accordingly, we are concerned with the posterior density of z_{ij}'s which relates with l_c by

$$\log f(z_{ij}\text{'s} \mid y_{ij}\text{'s}) = l_c(\theta\,;\, y_{ij}\text{'s},\, z_{ij}\text{'s}) + l_o(\theta\,;\, y_{ij}\text{'s}) \tag{2.11}$$

where l_o is the log-likelihood based on the observed data obtained by integrating out the z_{ij}'s in l_c. There are two main approaches in the literature. We can use either, if z_{ij}'s appear linearly in l_c, the posterior means of the z_{ij}'s which are the official requirement of the E-step, or, more generally, the posterior modes by maximizing (2.11) directly with respect to the z_{ij}'s with θ fixed.

Posterior means for the simplest Gaussian linear state space model are easily obtained from standard fixed interval smoothing and the linear Kalman filter (Harvey, 1989; Jørgensen *et al.*, 1996a) as $M(\theta \mid \hat{\theta}^{(k)})$ is a linear function of the first two moments of the states z_{ij}'s. Extension to incorporate measurements taken at irregular time intervals is straightforward and discussed in Jazwinski (1970) and Jones (1993). For DGLMs and frailty models, conjugate-prior posterior Bayesian analysis is possible resulting in the same form as the standard Kalman filter recursion (West *et al.*, 1985; Smith and Miller, 1986). In general, when l_c is non-linear in the z_{ij}'s, computing (2.10) resorts to numerical integration such as the Gauss-Hermite quadrature technique (Schnatter, 1992) but numerical effort increases exponentially with the dimension of the states. Instead, Monte

Carlo methods, e.g. Gibbs sampling, are used (Clayton, 1991; Fahrmeir and Tutz, 1994; Chan and Kuk, 1997). Estimation of standard errors can be approached by bootstrapping (e.g. Stoffer and Wall, 1991; Efron and Tibshirani, 1993), the supplemented Expectation-Maximization (SEM) algorithm which uses the convergence rate of the EM algorithm to estimate the "missing information" from using the Fisher information computed from l_c, (Meng and Rubin, 1991), or Monte Carlo approximation to the complete and missing information matrices from which the sum leads to the observed information matrix (Louis, 1982; Chan and Kuk, 1997).

Alternatively, when working with DGLMs, integration in (2.10) for computing the posterior means and covariance matrix can be avoided by approximating them with the posterior modes and curvatures (defined as the negative inverse of the second derivative of l_c) respectively. They are obtained by maximizing l_c in (2.9). However, direct maximization is inefficient when n_i is large and several recursive posterior mode filtering and smoothing algorithms are derived by using Gauss-Newton (Fisher scoring) iteration to l_c (Fahrmeir and Kaufmann, 1991; Fahrmeir and Tutz, 1994). Clearly, the posterior modes coincide with the posterior means under the special case of a Gaussian linear observation model. For GLMs with random effects (GLMMs), the resulting covariance estimate of Σ corresponds to the restricted maximum likelihood (REML) estimate (which will be discussed in the next section). However, the resulting EM-type algorithm from posterior mode filtering and smoothing relies on the appropriateness of the Gaussian linear transition model. For a highly skewed transition model, e.g. Gamma transition as in Jørgensen *et al.* (1996a), there will be great discrepancies between the posterior modes and means, and no guarantee that the recursion will converge.

In view of our own applications, preference will be given to the official posterior means as it is more natural and covers a wider range of distribution assumptions. But the main disadvantage is that the integration may be hard to perform and Monte Carlo approximation within each EM iteration may make it extremely slow to converge. However, on the other hand, estimation can be directed to the likelihood based on observed data l_o obtained as a by-product of filtering (Figure 2.1). This will be discussed in the next section. But the main advantages of using the EM algorithm over direct maximization of the likelihood based on observed data are that we only need to manipulate (2.9) which is usually much simpler as we do not need to integrate out the z_{ij}'s in l_c, and that solutions of the M-step can often be performed with standard statistical software.

2.2.2 Direct Likelihood Methods

By direct likelihood methods, we mean methods that work directly on the likelihood to be maximized. The EM algorithm is an indirect method as we work on the likelihood based on complete data with the aim to maximize the observed data likelihood. Now, the likelihood can be the one based on either observed or complete data.

To compute the observed log-likelihood l_o, the successive predictive densities needed are

$$f(y_{ij} \mid Y_i^{j-1}, x_{ij}) = \int f(y_{ij} \mid Y_i^{j-1}, x_{ij}, z_{ij}) f(z_{ij} \mid Y_i^{j-1}, x_{ij}) dz_{ij} \tag{2.12}$$

which can be obtained as by-products of the filtering recursion in (2.7) and (2.8) (Figure 2.1). If all the densities in (2.12) can be at least numerically evaluated, maximum likelihood estimates can be obtained by using common optimization algorithms, e.g. Quasi-Newton Raphson algorithm which has a fast convergence rate if the corresponding first derivative is tractable; otherwise derivative-free optimization algorithms such as Nelder-Mead Simplex method (Press *et al.*, 1986) are often more feasible. Availability of standard errors depends on the effort in evaluating the second derivative of the log-likelihood. This is usually high so we wish to resort to simulation methods such as parametric bootstrapping. However, as mentioned earlier in this section, integrations in (2.7), (2.8) and (2.12) may be hard to pursue. For DGLMs, numerical integration techniques or Monte Carlo methods have been studied (e.g. Chapter 7 and 8 of Fahrmeir and Tutz, 1994; Chapter 15 of West and Harrison, 1997), or we can put appropriate conjugate prior and posterior distributions assumptions on the model from which successive predictive densities in (2.12) can be written down mathematically (Smith and Miller, 1986; Harvey and Fernandes, 1989). A Bayesian approach from Chapter 4 and 14 of the book by West and Harrison (1997) for DGLMs is depicted in Figure 2.1.

For GLMMs in (2.2), the likelihood based on complete observation (assuming all z_i's are known) is sometimes maximized with respect to the fixed effects β and random effects z_i's to get the so-called best linear unbiased predictors (BLUPs) for variance components (McGilchrist, 1994). This is in contrast to the indirect posterior mode estimation when the likelihood based on complete observation is maximized with all variance components fixed in each M-step (Fahrmeir and Tutz, 1994). However, the BLUPs are asymptotically biased and inconsistent. Adjustment can be made to the BLUPs to approximate the REML estimates which have the variance components estimates corrected by an appropriate degrees of freedom resulting in estimates with smaller bias (Schall, 1991; McGilchrist, 1994; Breslow and Clayton, 1993). Direct bias adjustment of BLUPs is also considered by Kuk (1995) and McCullagh and Tibshirani (1990) using Monte Carlo iteration and bootstrapping respectively. Kalman filtering can also be used for "prewhitening" to obtain

REML estimates under linear mixed effects models (Wilson, 1988; Tsimikas and Ledolter, 1994). The validity and properties of most of these methods are not clear, but have been investigated for a few models; the longitudinal problems have not been studied much.

2.2.3 Bayesian Methods

In Bayesian analysis of longitudinal data, known prior distribution is imposed on each unknown parameter, and we want to compute the posterior density. Except under some rather restrictive assumptions, the posterior density is intractable and Monte Carlo methods are used. A popular one is the Gibbs sampler which is an iterative resampling scheme in a complete set of conditional posterior densities to approximate a marginal posterior density. An overview on the Gibbs sampler and other sampling methods is given by Gelfand and Smith (1990).

For GLMMs in (2.2) when the observation model assumes an exponential family distribution, the marginal joint posterior density of β and Σ can be approximated by the Gibbs sampler (Zeger and Karim, 1991). Carlin *et al.* (1992) considered the same basic technique on a special class of non-Gaussian and non-linear state space models but the computing time may not be reasonably affordable. Carter and Kohn (1994, 1996) developed more efficient Gibbs sampler based sampling schemes on a state space model which is Gaussian and linear when conditioned on a set of indicator variables. Another elegant Gibbs sampler based sampling scheme has recently been proposed on a Bayesian version of DGLMs (Section 2.1.2) when $g(\mu_{ij})$ is treated as random and follows a Gaussian distribution, i.e.

$$g(\mu_{ij}) \sim \mathcal{N}(C_{ij}z_{ij}, \Sigma_j);$$

see Cargnoni *et al.* (1997). However, even with current computing capacity, reducing convergence time of Monte Carlo methods remains a challenging issue. More efficient algorithm on broader class of models is still desirable.

2.2.4 Estimating Functions

We will mainly focus on maximum likelihood estimation in this dissertation but we briefly mention the use of estimating functions due to their numerous applications in the statistical literature. An estimating function is a function of observations and unknown parameters which is said to be unbiased if its marginal expectation is zero (Godambe, 1985; Thavaneswaran and Thompson, 1986, for discrete and continuous stochastic processes respectively). Inference for parameters is pursued by searching for the optimal estimating function among a class of unbiased estimating functions. Some optimality criteria are given in Godambe and Thompson (1989) which, roughly speaking, amounts to having the tightest confidence bounds for the estimates. In usual maximum likelihood analysis, optimal estimating functions often coincide with the score functions. In cases when iteration is needed to solve the score functions, good initial guesses can usually be easily obtained from the class of unbiased estimating functions. Optimal estimating functions also have promising uses in semi-parametric models when we do not have strong distribution assumptions. Some examples include non-linear time series estimation (Thavaneswaran and Abraham, 1988), and obtaining filtering and smoothing algorithms generally for non-Gaussian and nonlinear state space models (Naik-Nimbalkar and Rajarshi, 1995). Thompson and Kaseke (1995) has a brief review of unbiased estimating functions, with motivation from the EM algorithm, for estimation in GSSMs.

Another similar class of estimation methods which is proposed by Liang and Zeger

(1986) and Zeger and Liang (1986) is often called generalized estimating equations (GEEs). It has also been popularly entertained to estimate fixed effects in GLMs with correlated responses and possibly in the presence of random effects. A nice overview of using GEE in GLMs can be found in Zeger and Liang (1992). For GLMs with independent responses, the GEE reduces to a "quasi-likelihood" equation which corresponds to an optimal estimating function. A more general definition of quasi-likelihood equations for dependent responses and its application in stochastic processes are given in Godambe and Heyde (1987). However, GEEs are only optimal estimating functions under some restrictive situation on the marginal covariance structure (Liang *et al.*, 1992; McCullagh and Nelder, 1989, Chapter 9). For GLMMs, apart from estimating fixed effects, predicting random effects and estimating between subject variability can be performed through a three-stage iteration scheme using GEE and estimating functions (Waclawiw and Liang, 1993). More references on GEE can be found in Diggle *et al.* (1994).

2.3 Applications in Longitudinal Studies

In this dissertation, we will focus on three main areas in longitudinal studies: missing values and measurement errors in multiple responses, modelling recurrent events with random effects, and differentiating between different random effects models. The following sections will give a brief background and introductory discussion on each of these topics.

2.3.1 Missing Values and Measurement Error in Multiple Responses

Missing values is an important issue in longitudinal studies which brings problems that would not exist in cross-sectional studies. Let y be a vector representing all responses as if they were all observed, and partition $y = (y^{(o)}, y^{(m)})$ where $y^{(o)}$ are the observed responses while $y^{(m)}$ are those which are actually missing. Then three types of missing data mechanisms can be distinguished according to Little and Rubin (1987), namely, (i) *missing completely at random* (MCAR) when the missing data mechanism, R, does not depend on $y^{(o)}$ and $y^{(m)}$; (ii) *missing at random* (MAR) when R depends on $y^{(o)}$ only; and (iii) *informative* when R depends on both $y^{(o)}$ and $y^{(m)}$. MCAR and MAR are also collectively called *ignorable* or *non-informative* missing data mechanisms wherein likelihood based inference is unaffected due to the decomposition of the likelihood function separately into one based on the observed responses and the other based on the missing data mechanism. Only the likelihood based on the observed responses is used in statistical analysis.

Throughout this dissertation, we assume all missing responses are ignorable or non-informative. For the two sets of automobile manufacturing data mentioned in Section 1.2.1 of Chapter 1, the chief aim is to model production variation, added and transmitted, across different process stages while incorporating missing values and measurement errors in the multiple responses. For univariate responses without missing values, a first order autoregressive model can be used to analyze the variation transmission process with measurement errors (Lawless *et al.*, 1997; Agrawal *et al.*, 1997). For multiple responses with some or all values not measured, we can use the EM algorithm under a first order multivariate autoregressive model with the E-step carried out by directly taking conditional expectations on the vector of responses from each vehicle (Hamada and Lawless. 1994). However, computational effort increases exponentially with the size of the multiple measurements at each process stage and the total number of process stages. More efficient estimation while handling missing values and measurement errors is desirable and is studied, on Gaussian linear models, in Chapter 3.

2.3.2 Modelling Recurrent Event Data

The small bowel motility data mentioned in Section 1.2.2 of Chapter 1 is a typical set of recurrent event data. The last recurrence time for each subject is censored at the planned end of surveillance. That is, the last recurrence time is the time to end of surveillance instead of the time to next event recurrence. Renewal processes, in which the times between successive occurrences are independent and identically distributed, are often used to analyze such data. Inter-subject heterogeneity or random effects come naturally in longitudinal studies when the subjects are a random sample from some larger population and some important covariates are missing or there are measurement errors incurred in some time-independent covariates (Pickles and Crouchley. 1994). Modelling within-subject correlation in observed measurements is another fundamental objective and is also a consequence of using subject-level random effects.

Common parametric regression models for lifetime data (Lawless, 1982) can be classified into accelerated life models and proportional hazards models. One of their main distinctions is that the effect of explanatory variables is directed to a function of the recurrence time in accelerated life models and to the hazard function in proportional hazards models. Aalen and Husebye (1991) compared the use of a Normal-based (GLMM) and a hazard-based (Gamma frailty model) model on recurrent events in their extension of renewal processes. These models introduce inter-subject variability and intra-subject covariability (that results in dependencies between a subject's recurrence times). Also, correlation between recurrence times of a subject is induced by subject-level random effects. Following these, we will discuss, in a more general framework, the Normal-based and hazard-based models in accounting for inter-subject heterogeneity and within-subject correlation.

A merit of using hazard-based models for recurrent events is the convenience of incorporating time-dependent covariates. For a proportional hazards model with a Gamma frailty, maximum likelihood estimates can be easily obtained (Aalen and Husebye, 1991). However, with a log-Normal frailty, the likelihood is no longer tractable and estimation strategies typically resort to numerical integrations or Monte Carlo methods (Clayton, 1994; Ng and Cook, 1997; Xue and Brookmeyer, 1996). Thus, with emphasis on proportional hazards models and additionally allowing non-stationary drifts, we would like to study the use of filtering and smoothing type methods by which the likelihood function and subsequent event recurrence times can be easily evaluated and predicted. This is investigated in Chapter 5.

Chapter 3

Missing Data and Measurement Error in a Multivariate AR(1) Model

3.1 Introduction

In order to reduce variation in manufacturing processes consisting of several discrete stages it is often worthwhile to study the variation that is added at different stages. and whether that variation is transmitted downstream to subsequent stages. In particular. there may be certain stages where considerable variation originates, and other stages that filter out variation introduced upstream. By understanding how variation is added and transmitted across the stages of a process we can decide where to concentrate variation reduction efforts. The piston machining and door hanging processes taken from automobile manufacturing in Section 1.2.1 are two examples.

Lawless *et al.* (1997) present methods for analyzing the transmission of variation in a univariate characteristic, based on a first order autoregressive model. In order to carry out such analysis it is necessary to be able to track units (in our examples these are vehicles) through the manufacturing process so that measurements may be taken on the same unit at different stages. Lawless *et al.* (1997) assume that a univariate quality characteristic y_t is measured at each of T process stages $t = 1, \ldots, T$, and consider the model

$$y_1 = \mu_1 + e_1 \tag{3.1}$$

$$y_t = \alpha_t + \beta_t y_{t-1} + e_t \quad t = 2, \ldots, T \tag{3.2}$$

where $e_t \sim N(0, \sigma^2_{e_t})$ and are independent. This first order Markov, or autoregressive AR(1) model can often be justified in manufacturing processes, and it leads to the following variation transmission formula for $\sigma^2_t = \mathrm{Var}(y_t)$:

$$\sigma^2_t = \beta^2_t \sigma^2_{t-1} + \sigma^2_{e_t} . \tag{3.3}$$

The first term on the right side of (3.3) represents variation transmitted from stage $t-1$ to stage t, and the second term represents variation added at stage t. Lawless *et al.* (1997) fit models (3.1) and (3.2) to process data and discuss how to use (3.3) recursively to assess variation transmission across stages $t = 1, \ldots, T$ of a process.

In this chapter, we extend the techniques of Lawless *et al.* (1997) in several directions. First, we consider multivariate measurements, and in particular, deal with a multivariate version of (3.1) and (3.2). We will refer to the model as an AR(1) model, but it should be noted that T is generally small and the model is non-stationary, unlike many applications involving AR(1) models. Second, we deal with missing data; this is important since it is often difficult to measure all characteristics on every unit in a study that is undertaken on-line, i.e. while the manufacturing process is operating. Finally, we incorporate measurement error into the multivariate AR(1) model; this is important because, as discussed by Agrawal *et al.* (1997) and Lawless *et al.* (1997), if substantial measurement error is ignored the results of the AR(1)-based variance transmission analysis are misleading.

Section 3.2 of the chapter introduces the multivariate AR(1) model and incorporates measurement error. Section 3.3 is the core of the chapter and presents methodology for fitting the model to process data; this is done by using a state space formulation that leads to efficient computational procedures. Section 3.4 illustrates the methodology on the piston machining and door hanging processes, and Section 3.5 concludes with comments and points that deserve further study.

3.2 An AR(1) Variation Transmission Model

The methods that we are considering are designed for use on a stable process. That is, the model (3.4)-(3.5) applies to units manufactured over time, and the parameter values in the model do not change over time. We assume that sequential measurements on a random sample of n units from the process are available. As discussed by Lawless *et al.* (1997) for the univariate case, we consider a (non-stationary) first order autoregressive, or AR(1), model for the $C \times 1$ vector of multivariate measurements on z_{it} on unit i at stage t ($t = 1, \ldots, T$; $i = 1, \ldots, n$). This can be expressed as

$$\boldsymbol{z}_{i1} = \boldsymbol{\mu}_1 + \boldsymbol{e}_{i1} \tag{3.4}$$

$$\boldsymbol{z}_{it} = \boldsymbol{A}_t + \boldsymbol{B}_t \boldsymbol{z}_{i,t-1} + \boldsymbol{e}_{it} \quad t = 2, \ldots, T \tag{3.5}$$

where $e_{it} \sim N_C(0, \sum_{e_t})$, $t = 1, \ldots, T$; the notation $y \sim N_p(\mu, \sum)$ means that y has a p-variate normal distribution with mean vector μ and covariance matrix $\sum$. The dimensions of A_t and B_t are $C \times 1$ and $C \times C$, respectively. It is assumed that the measurements for different units are independent.

The marginal means and covariance matrices for the z_{it}'s are given by

$$E(z_{i1}) = \mu_1, \ E(z_{it}) = \mu_t = A_t + B_t\mu_{t-1}, \ t = 2, \ldots, T \tag{3.6}$$

$$\mathrm{Var}(z_{i1}) = \sum_{e_1}, \ \mathrm{Var}(z_{it}) = \sum_t = B_t \sum_{t-1} B_t' + \sum_{e_t}, \ t = 2, \ldots, T \tag{3.7}$$

In addition

$$\mathrm{Cov}(z_{is}, z_{it}) = \sum_{st} = \sum_s B_{s+1}' \ldots B_t' \ (s < t) \tag{3.8}$$

The vector e_t and its covariance matrix $\sum_{e_t}$ represent variation added at stage t, whereas $B_t \sum_{t-1} B_t'$ represents variation transmitted from stage $t-1$; in this regard the right hand portion of (3.7) is the multivariate generalization of (3.3). The intercept A_t allows the means $\mu_t = E(z_{it})$ to vary across $t = 1, \ldots, T$. In a case like that in Example 2, for instance, a stage may reduce the diameters from the preceding stage substantially. An alternative but equivalent parameterization is $E(z_{it}|z_{i,t-1}) = \mu_t + B_t(z_{i,t-1} - \mu_{t-1})$.

In practice there may be significant measurement error, that is, variation in the process by which the $z_{it} = (z_{it1}, \ldots, z_{itc})'$ are measured. As discussed in Section 3.5, this can invalidate the methods described herein if it is ignored, so we consider it explicitly. We let y_{it} represent the measurement of z_{it} and assume that

$$y_{it} = z_{it} + \delta_{it}, \ t = 1, \ldots, T \tag{3.9}$$

where the δ_{it}'s are mutually independent $N_c(0, \sum_{\delta_t})$ random vectors and are independent of the e_{it}'s in (3.4) and (3.5). It should be noted that the y_{it}'s do not follow an AR(1) model.

The motivation for considering the model (3.2) is to examine the sources of variation in the measurements z_{iT} at the final stage. This may be done by working backwards from the final stage: (3.2) for $t = T$ indicates that the covariance matrix $\sum_T$ may be decomposed into variation transmitted from stage $T-1$ and variance added at stage T.

$$\sum_T = B_T \sum_{T-1} B_T' + \sum_{e_T} . \quad (3.10)$$

Similarly, $\sum_{T-1}$ may be decomposed and, working backwards, we may ascertain the contribution of the variation added at any stage t (i.e. $\sum_{e_t}$) to $\sum_T$. Multivariate covariance matrices may admittedly be hard to interpret, and it is important to relate them to the physical properties of the units under consideration. The example of Section 3.5 illustrates and discusses this further.

Care should be taken to assess the appropriateness of the model (3.1)-(3.5), possibly with measurement error accounted for by (3.9). Section 3.4 discusses model checking and Section 3.5 comments on the robustness of the methods to departures from the model.

3.3 Parameter Estimation

It is important to have estimation procedures that deal with missing data, since it is often impossible to measure all the characteristics on every unit at every stage. We therefore suppose that some arbitrary subset of the CT univariate measurements on unit i may be missing, and that observations are missing at random in the terminology of Rubin (1976) and Little and Rubin (1987). This means that the probability a particular set of measurements on a unit is missing does not depend on the values of the measurements for that or other units, and implies that the likelihood function may be based on the joint distribution of the measurements available for each unit.

We assume that the covariance matrices $\sum_{\delta_t}(t = 1, \ldots, T)$ for the measurement errors are known. In practice these should be estimated from measurement studies. The set of unknown parameters then includes μ_1, the $\sum_{e_t}$'s $(t = 1, \ldots, T)$ and the A_t's and B_t's $(t = 2, \ldots, T)$. Since the observed measurements $y_{itc}(t = 1, \ldots, T; c = 1, \ldots, C)$ for unit i jointly follow a multivariate normal distribution of dimension CT or less, it would be possible in principle to write the mean and covariance matrix for each i in terms of the unknown parameters and to maximize the likelihood by a search algorithm. In particular, we note that, under (3.4), (3.5) and (3.9), the complete data y_{it}'s have means μ_t given by (3.6) and covariance matrices

$$\text{Var}(y_{it}) = \sum\nolimits_t + \sum\nolimits_{\delta_t}, \quad \text{Cov}(y_{is}, y_{it}) = \sum\nolimits_{st} \ (s < t), \tag{3.11}$$

where $\sum_t$ and $\sum_{st}$ are given in (3.7) and (3.8), respectively. This brute force approach encounters matrices of large dimension if CT is large, and is computationally slow; the latter is a drawback for the use of bootstrap methods for obtaining variance estimates or confidence intervals, as described in Section 3.4. Consequently we will express the model in state space form (e.g. Harvey, 1989; Harvey and McKenzie, 1984; Shumway, 1988), and utilize the EM algorithm (Dempster *et al.*, 1977) to obtain maximum likelihood estimates.

The model given by (3.4), (3.5) and (3.9) with arbitrary measurements missing at random can be expressed in the following form, where y_{it} now stands for the vector of observed measurements on unit i at stage t:

$$z_{it} = A_t + B_t z_{i,t-1} + e_{it} \tag{3.12}$$

$$y_{it} = H_{it} z_{it} + H_{it}\delta_{it} \tag{3.13}$$

where $i = 1, \dots, n$; $t = 1, \dots, T$, we define $A_1 = \mu_1$, $B_1 = 0$, $z_{i0} = 0$, and where H_{it} is a matrix obtained by taking the $C \times C$ identity matrix and deleting rows which correspond to missing observations on unit i at stage t. This belongs to the Normal linear state space models mentioned in Section 2.1.1 of Chapter 2.

The log-likelihood function based on the observed data is computed by a product of all successive predictive densities, $f(y_{ij} \mid Y_i^{j-1})$ which may be written in the form of an arbitrary constant plus

$$\ell = -\frac{1}{2}\sum_{i,t} \log \left|\sum\nolimits_{iy}(t|t-1)\right| - \frac{1}{2}\sum_{i,t} (y_{it} - y_{i,t|t-1})'\sum\nolimits_{iy}^{-1}(t|t-1)(y_{it} - y_{i,t|t-1}), \tag{3.14}$$

where we introduce the notation

$$y_{ip|q} = E[y_{ip}|y_{i1}, \dots, y_{iq}], \ \sum\nolimits_{iy}(p|q) = \mathrm{Var}[y_{ip}|y_{i1} \dots, y_{iq}] \tag{3.15}$$

and where the range for i and t in the sum $\sum_{i,t}$ is over $i = 1, \dots, n$ and $t = 1, \dots, T$. Expression (3.14) assumes there is at least one measurement at each stage for each unit. If all measurements at a stage t happen to be missing for unit i, then (3.14) is modified to omit terms involving $\sum_{iy}(t|t-1)$, $\sum_{iy}(t+1|t)$ and to add a term involving $\sum_{iy}(t+1|t-1)$.

The terms $y_{i,t|t-1}$ and $\sum_{iy}(t|t-1)$ needed to calculate (3.15) may be computed recursively using the following state space, or Kalman *filtering formulas* as derived from (2.7), (2.8) and (2.12) in Section 2.2. They have closed form expressions for Normal linear models as we have here. Define, following (3.15),

$$z_{ip|q} = E[z_{ip}|y_{i1}, \ldots, y_{iq}], \ \sum\nolimits_{iz}(p|q) = \mathrm{Var}[z_{ip}|y_{i1}, \ldots, y_{iq}],$$

and set $z_{i0|0} = 0$, $\sum_{iz}(0|0) = 0$. Then for $t = 1, 2, \ldots, T$,

$$z_{it|t-1} = A_t + B_t z_{i,t-1|t-1} \tag{3.16}$$

$$\sum\nolimits_{iz}(t|t-1) = B_t \sum\nolimits_{iz}(t-1|t-1)B_t' + \sum\nolimits_{e_t} \tag{3.17}$$

$$y_{it|t-1} = H_{it} z_{it|t-1} \tag{3.18}$$

$$\sum\nolimits_{iy}(t|t-1) = H_{it} \sum\nolimits_{iz}(t|t-1)H_{it}' + H_{it} \sum\nolimits_{\delta_t} H_{it}', \tag{3.19}$$

where $z_{it|t}$ and $\sum_{iz}(t|t)$ are computed via

$$P_{it} = \sum\nolimits_{iz}(t|t-1)H_{it}' \sum\nolimits_{iy}(t|t-1)^{-1}$$

$$z_{it|t} = z_{it|t-1} + P_{it}(y_{it} - y_{it|t-1})$$

$$\sum\nolimits_{iz}(t|t) = \sum\nolimits_{iz}(t|t-1) - P_{it} \sum\nolimits_{iy}(t|t-1)P_{it}'.$$

Derivation of these formulas is outlined in Appendix B. These calculations involve only square matrices of dimension C or smaller.

Now that we can compute (3.14), we could maximize it by using a derivative-free procedure such as the simplex search algorithm (Nelder and Mead, 1965; Press *et al.*, 1986, Section 10.4). An attractive alternative, which also allows easier access to model-checking and to handling cases where entire stages are missing on some units, is an EM algorithm. This has been well-discussed for use with missing data in normal models (e.g. Little and Rubin, 1987, Chapter 8) and is adapted here to deal with both missing data and measurement error. A brief discussion on EM algorithm was given in Section 2.2.1 of Chapter 2.

Referring to (3.12), we consider the "complete data" log-likelihood as that based on the z_{it}'s, which may be written as an arbitrary constant plus

$$\ell_c = -\frac{nC}{2}\sum_{t=1}^{T}\log\left|\sum\nolimits_{e_t}\right| - \frac{1}{2}\sum_{i=1}^{n}\sum_{t=1}^{T} e_{it}'\sum\nolimits_{e_t}^{-1} e_{it}\,. \tag{3.20}$$

The model (3.12) is AR(1), and maximum likelihood estimates are easily found to be (e.g. Mardia *et al.*, 1979, Chapter 6)

$$\begin{gathered}\hat{A}_1 = \bar{z}_1, \qquad \hat{\sum}\nolimits_{e_1} = S_{1.1}\\ \hat{B}_t = S_{t.t-1}(S_{t-1.t-1})^{-1},\ \hat{A}_t = \bar{z}_t - \hat{B}_{t-1}\bar{z}_{t-1}\\ \hat{\sum}\nolimits_{e_t} = S_{t.t} - \hat{B}_t S_{t-1.t}\end{gathered} \tag{3.21}$$

for $t = 2, \ldots, T$, where

$$\bar{z}_t = \frac{1}{n}\sum_{i=1}^{n} z_{it},\ S_{u.t} = \frac{1}{n}\sum_{i=1}^{n} z_{iu}z_{it}' - \bar{z}_u\bar{z}_t' \tag{3.22}$$

The M-step in the EM algorithm is given by (3.21). The E-step consists of computing the expectations of the complete data, conditional on the observed data, that are needed to compute the conditional expectation of (3.20). This may be done using the state-space *smoothing formulas* for $t = 1, \ldots, T-1$:

$$R_{it} = \sum\nolimits_{iz}(t|t) B'_{t+1} \sum\nolimits_{iz}(t+1|t)^{-1}$$

$$z_{it|T} = z_{it|t} + R_{it}(z_{i,t+1|T} - z_{i,t+1|t}) \tag{3.23}$$

$$\sum\nolimits_{iz}(t|T) = \sum\nolimits_{iz}(t|t) - R_{it}[\sum\nolimits_{iz}(t+1|t) - \sum\nolimits_{iz}(t+1|T)]R'_{it} \tag{3.24}$$

Derivations are outlined in Appendix B. The E-step is now carried out by replacing $\bar{z}_t$ and $z_{i,t-1}z'_{it}$ in the expressions (3.22) with (compare Little and Rubin, 1987, page 143)

$$\frac{1}{n}\sum_{i=1}^{n} z_{it|T} \tag{3.25}$$

$$z_{i,t-1|T}z'_{it|T} + \mathrm{Cov}(z_{i,t-1}, z_{it}|y_{i1} \ldots, y_{iT}) \tag{3.26}$$

respectively, evaluated at the most recent parameter estimates from the M-step (3.21).

In the case where there is no measurement error, $\mathrm{Cov}(z_{i,t-1}, z_{it}|y_{i1}, \ldots, y_{iT}) = \sum_{iz}(t-1|T)B'_t$. More generally, however, it must be obtained from the smoothing formula (3.24) for the augmented model

$$z^*_{it} = \begin{pmatrix} z_{it} \\ z_{i,t-1} \end{pmatrix} = \begin{pmatrix} A_t \\ 0 \end{pmatrix} + \begin{pmatrix} B_t & 0 \\ I & 0 \end{pmatrix} z^*_{i,t-1} + \begin{pmatrix} e_{it} \\ 0 \end{pmatrix},$$

where I represents an identity matrix.

The EM algorithm proceeds by alternating E and M steps until convergence is achieved. Initial estimates that can be used to start the process can be obtained by the following simple procedure: compute empirical means $\bar{y}_t$ and cross-product matrices $S_{t,t}$ and $S_{t-1,t}$ using units with no missing measurements at stage t (for $\bar{y}_t$ and $S_{t,t}$) and at stages $t-1$

and t (for $S_{t-1,t}$), respectively. Then, compute the estimates

$$\tilde{\mu}_1 = \bar{y}_1 \; \tilde{\sum}_{e_1} = S_{1,1} - \sum_{\delta_1} \tag{3.27}$$
$$\tilde{B}_t = S_{t,t-1}(S_{t-1,t-1} - \sum_{\delta_{t-1}})^{-1} \; t = 2, \ldots, T$$
$$\tilde{A}_t = \bar{y}_t - \tilde{B}_t \bar{y}_{t-1} \; t = 2, \ldots, T$$
$$\tilde{\sum}_{e_t} = (S_{t,t} - \sum_{\delta_t}) - \tilde{B}_t S_{t-1,t} \; t = 2 \ldots, T.$$

When there is no missing data, these are the estimates that would be obtained by maximum likelihood if the process had only $T = 2$ stages. Agrawal *et al.* (1997) study these estimates in the univariate case.

There are many $(CT + C^2(T-1) + C(C+1)T/2)$ parameters in the model, and we are primarily interested in components of variance as epitomized in (3.3) and (3.7). In these circumstances it does not make sense to develop estimates of the asymptotic variances and covariances of all parameter estimates. In order to assess variation in estimates and to obtain confidence intervals for quantities of interest, we use a parametric bootstrap (Efron and Tibshirani, 1993). The procedure is as follows: treating the maximum likelihood estimates as if they were the true parameter values and the H_{it}'s as given by the pattern of missingness in the original data, we generate B sets of data from the model (3.12)-(3.13). For each of the B sets of data we obtain maximum likelihood estimates $\hat{\theta}_b^*$ (where θ stands for the vector of all parameters). Estimates of functions $\psi = g(\theta)$ that are of interest are then calculated for each sample. Variance estimates for $\hat{\psi} = g(\hat{\theta})$ (where $\hat{\theta}$ is the maximum likelihood estimate from the original data) or confidence intervals for ψ may then be calculated in various standard ways (see Efron and Tibshirani, 1993).

An example of the bootstrap methods is given in Section 3.4.

3.4 Applications to Car Manufacturing Processes

Here we consider two car manufacturing data as described in Section 1.2.1 of Chapter 1. The data are given in Appendix A.1.

3.4.1 Piston Machining

We consider data on $96(=n)$ randomly selected pistons from the piston machining process mentioned in Section 1.2.1 of Chapter 1. Four $(=C)$ diameter measurements were taken at each of $4(=T)$ process stages.

The model represented by (3.4), (3.5) and (3.9) was fitted. There are no missing observations here and the measurement error covariance matrix is assumed to be $\sigma_\delta^2\ I_4$, where I_4 is the 4×4 identity matrix. The measurements are discrete, diameters being measured to the nearest micron $(10^{-3}mm)$, and at each of the 4 locations on the piston fewer than 15 distinct values occur; see Section 1.2.1. Nevertheless we will work with the assumed normal model, which seems to provide a reasonable picture of variation.

Models were fitted with $\sigma_\delta^2 = .04167$ microns^2 and also with $\sigma_\delta^2 = .1$ microns^2. The former corresponds to the variance of a triangular distribution on $(-.5, .5)$ and the latter is slightly larger than the variance of a uniform distribution on $(-.5, .5)$. The latter seems a more realistic value but we wanted to assess the effect of measurement error on estimated variance components.

The EM algorithm based on the filtering and smoothing procedures was iterated until the increase in the log-likelihood (3.14) was less than .1; the maximum value at convergence was 8017.0. Maximum likelihood estimates of B_t, $\sum_t$ and $\sum_{e_t}$, as in (3.7), are shown in Table 3.1 for the case where $\sigma_\delta^2 = .10$. Estimates of μ_t are also shown. The units for all variances and covariances are microns^2. Parametric bootstrap methods (Efron and Tibshirani, 1993) were used to generate standard errors and confidence limits for variance components. Standard errors for estimates of variance tended to be about 10-20% of the size of the estimate. The entire procedure, including 1000 bootstrap replications, used under 7 minutes of CPU time on a DEC OSF/1 V3.2 system when programmed in C++. The estimates obtained when $\sigma_\delta^2 = .04167$ was used were a little different, but the qualitative picture was similar to that in Table 3.1. The main feature was that $\hat{\sum}_{e_t}$ tended to be about 10% larger than in Table 3.1, whereas $\hat{\sum}_t$ was more or less the same.

Table 3.1 suggests that roughly 30-60% of the variation in diameters at each stage is added at that stage and the rest is transmitted from the preceding stage. By using (3.10) recursively we can express $\hat{\sum}_4$ as a sum of four components, one representing the variation at each stage. This indicates that attempts to reduce variation at the final stage should be directed at stages 3 or 4; little variation is transmitted from stages 1 and 2. We remark that it is also of interest with multivariate measurements to examine their correlation structure. Table 3.1 indicates a moderate degree of correlation for adjacent diameters in both the total variance and in the variance added at each stage. The examination of principal components or other linear functions of measurement variables is also of general interest but we will not pursue this here.

The model (3.4), (3.5) and (3.9) can be checked informally by examining residuals

$$r_{it} = y_{it} - \hat{y}_{it|t-1}$$

or standardized versions of the same by using $\hat{\sum}_{iy}(t|t-1)$. Standardized residuals should look roughly like $N(0,1)$ variables. Figure 3.1 shows plots of standardized residuals versus

Stage (t)	$\hat{\sum}_t^1$				$\hat{\sum}_{e_t}^1$				$\hat{B}_t$				$\hat{\mu}_t$ (mm)
1	4.67	0.64	0.21	0.05	4.67	0.64	0.21	0.05	0	0	0	0	88.96
		1.94	0.54	0.38		1.94	0.54	0.38	0	0	0	0	88.97
			2.29	0.51			2.29	0.51	0	0	0	0	88.94
				3.56				3.56	0	0	0	0	88.16
2	5.10	0.56	0.27	-0.10	1.71	0.74	0.33	0.08	0.73	0.16	0.22	-0.22	88.96
		3.09	0.55	0.16		1.36	0.35	0.18	-0.24	1.02	0.25	-0.20	88.97
			5.94	0.37			3.20	0.41	-0.18	0.31	1.00	-0.09	88.93
				3.24				1.07	-0.08	-0.01	-0.06	0.79	88.16
3	4.69	0.42	0.08	-0.07	2.28	0.44	0.28	0.13	0.75	-0.19	-0.07	0.09	88.96
		1.43	0.44	0.01		0.81	0.44	0.37	0.04	0.45	-0.09	0.05	88.97
			3.42	0.17			2.20	0.31	-0.18	0.24	0.32	0.11	88.93
				4.64				1.65	0.01	-0.35	-0.21	0.92	88.16
4	5.91	0.64	0.18	-0.23	2.20	0.67	0.51	0.30	0.75	0.22	-0.11	-0.29	88.96
		2.13	0.46	-0.09		1.25	0.60	0.49	0.05	0.53	0.07	-0.24	88.97
			4.79	0.22			2.25	0.55	-0.04	-0.37	0.91	-0.27	88.93
				4.20				1.92	-0.07	-0.34	0.12	0.63	88.16

[1]The off-diagonal elements are the correlations; the diagonal elements are the variances.

Table 3.1: Estimated Covariance Matrices for Piston Diameters

predictors $\hat{y}_{it|t-1}$ across all stages measurements ($i = 1, \ldots, 96$; $t = 1, \ldots, 4$) for $c = 1, 2, 3, 4$ (corresponding to 4, 10, 36.7 and 58.7 mm). The banded appearance in each plot is due to the fact that for each diameter there are only 10-15 distinct values of y_{it}, and that the estimated variance for r_{it} does not depend on i and varies slightly with t. Figure 3.2 shows a normal probability plot of standardized residuals. These are reasonably linear, though a single extreme observation is noted at each of 4 mm and 36.7 mm. More exhaustive checks not shown here likewise do not indicate substantial departures from the working model.

3.4.2 Door Hanging

We now examine the door hanging process in Section 1.2.1 of Chapter 1. The data consists of $42 (= n)$ vehicles passing through $7 (= T)$ process stages. At each stage, $4 (= C)$ characteristics of a rear door of each vehicle are of interest.

The model represented by (3.4), (3.5) and (3.9) was again fitted. There are no measurement errors assumed here. Initial estimates were obtained by maximizing the likelihood based on vehicles with complete measurements at all stages. The EM iteration was stopped when the increase in log-likelihood (3.14) was less than .1. The recursion stopped in ten iterations with converged maximum log-likelihood at -119.118 within 40 seconds of CPU time on a DEC OSF/1 V3.2 system when programmed in C++. Standard errors of the estimates, obtained from 1000 bootstrap samples, again tended to be around 10% of the estimates. Maximum likelihood estimates of $\sum_t$ and $\sum_{e_t}$ are shown in Table 3.2 and those of B_t are shown in Table 3.3.

Table 3.2 shows little variation in door exterior fitness is transmitted from stage 3 to subsequent stages. For the first fitness measure at stage 4, there is over 80% of variation added and a fair portion of them are transmitted to the later stages. The same is observed for the last fitness measure at stage 6. More than 20% of variation added at the last stage is also observed. Thus, attempts to reduce variability at the final stage should be directed to stage 4, 6 and 7. Again, moderate correlations are observed.

The model was again checked by examining the residuals. Plots of standardized residuals against the predictors $y_{it|t-1}$ for each characteristics are shown in Figure 3.3. Figure 3.4 shows normal quantile plots of standardized residuals for each characteristics. No substantial departures from the working model are observed from further checks likewise.

3.5 Concluding Remarks

The methods in this chapter depend on the approximate validity of a normal $AR(1)$ model for the true measurements. This assumption should be realistic in many contexts. but it would be of interest to consider the implications of model departures. One topic which is readily assessed is the effect of ignoring measurement error. If the model (3.4). (3.5) is assumed correct but there is in fact measurement error as expressed by (3.9). then the maximum likelihood estimates $\hat{B}_t$ derived under (3.4), (3.5) alone converge in probability in large samples not to B_t but to

$$B_t^* = B_t \sum_{t-1} (\sum_{t-1} + \sum_{\delta_{t-1}})^{-1}$$

This underestimation of regression parameters is well known when measurement error in covariates is ignored (e.g. Fuller. 1987). A consequence of this in the present circumstances is that the variation transmitted to each stage is underestimated and the variation added is overestimated. This has serious consequences when there are several stages in the process. Agrawal *et al.* (1997) give a detailed discussion of measurement error for the univariate ($C = 1$) case. They have shown for the case with measurement error but no missing data that the use of simple estimates (3.27) combined with bootstrap confidence intervals provide good procedures. Extension of these methods to the multivariate case is worth considering. For example, sensitive analysis by trying different values for the variability of measurement errors.

In practical situations one must decide which measurements to consider. This choice can affect whether or not an AR(1) model is satisfactory. For example, if we include a pair of measurements but omit a third which is highly correlated with the other two, we may find an AR(1) model for the two measurements is inadequate.

The analysis here is based on the assumption that the missing mechanism does not depend on the missing measurements. A likelihood ratio test can be used for testing "informative" drop-out processes (Diggle and Kenward, 1994). However, in our applications, developing testing procedures for whether the intermittent missing values are informative is desirable.

Further work on ways to interpret multivariate analyses of variation in special contexts is desirable. In particular, one would hope to expose significant relationships among variables and to relate them to the geometry of the units being manufactured. With the piston data there do not appear to be important systematic effects but one could imagine situations in which, for example, the deviations in diameters at opposite ends of a cylinder were negatively correlated after certain stages. The present chapter has developed efficient procedures for model fitting and assessment which should make it feasible to undertake further studies with relative ease.

Finally, the methods here deal with processes in which the same variables are measured on parts at each stage. However, as mentioned by Lawless *et al.* (1997), the general ideas of variation transmission also apply to studies of the effect of upstream process variables on downstream measurements. This area requires further development in practical situations.

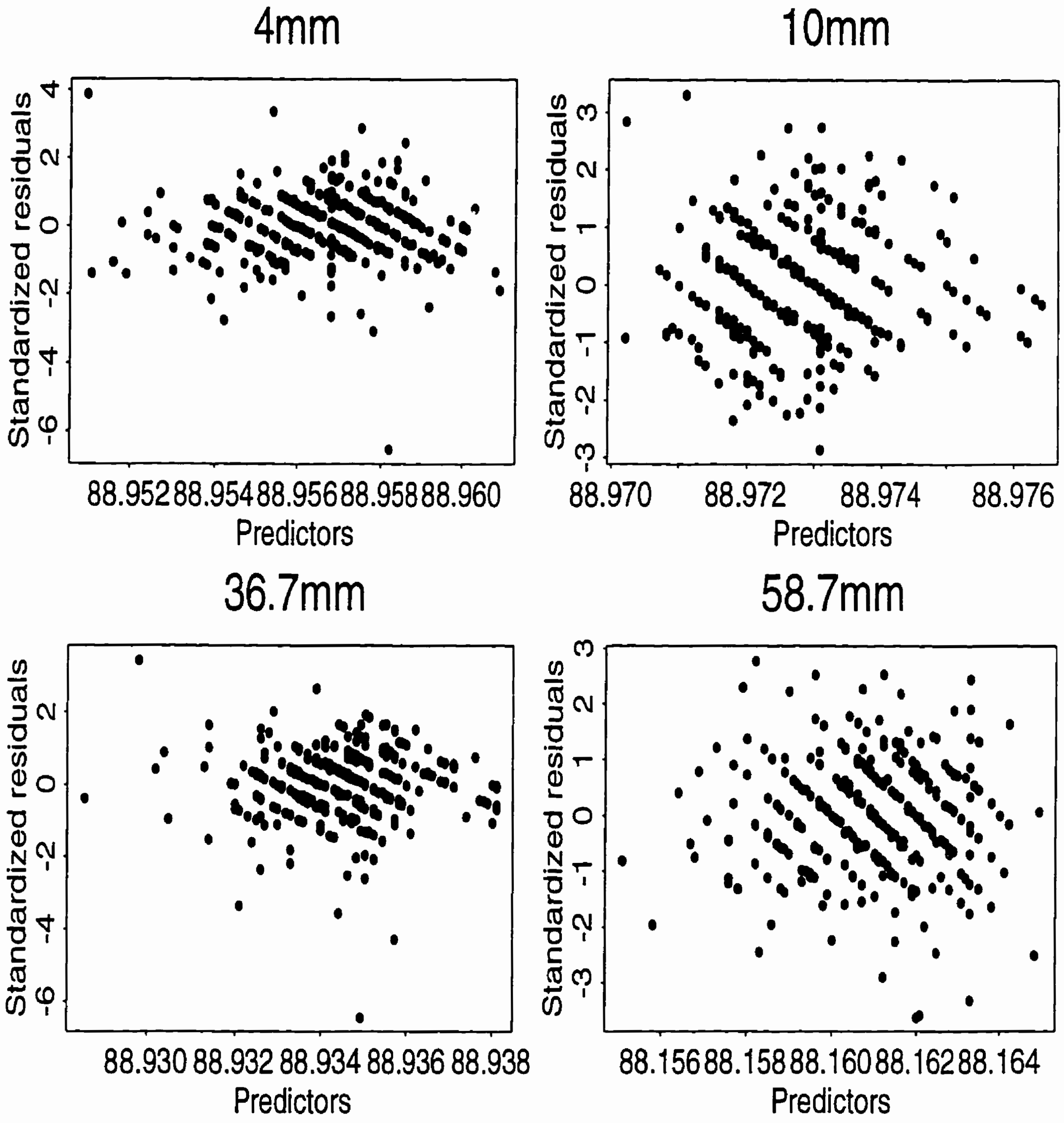


Figure 3.1: Piston Machining: Plots of residuals against the predictors $\hat{y}_{it|t-1}$.

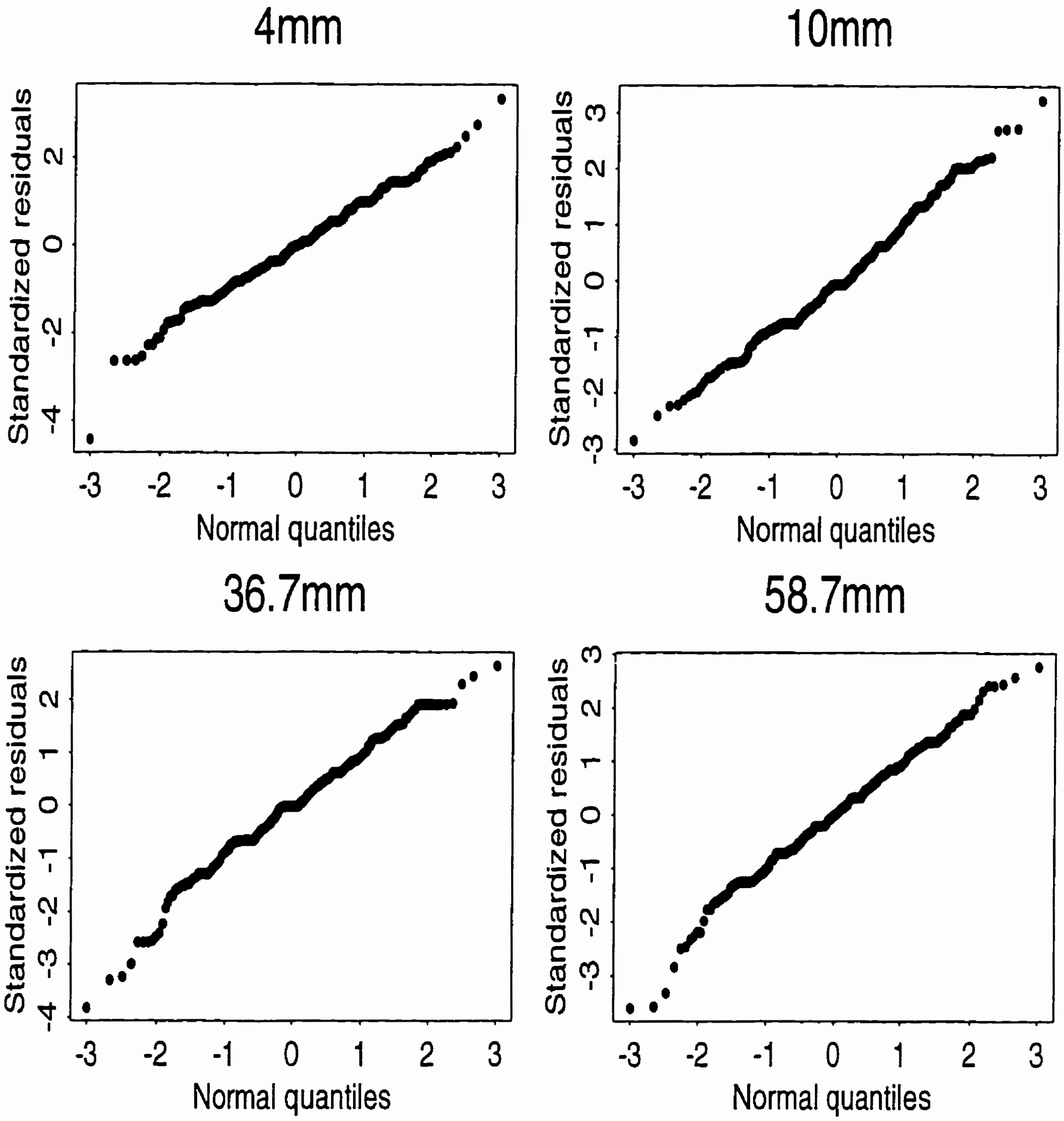


Figure 3.2: Piston Machining: Q-Q plots of standardized residuals.

Stage (t)	$\hat{\Sigma}_{e_t}$[1]				$\hat{\Sigma}_t$[1]			
1	1.4028	-0.8439	0.6333	0.7728	1.4028	-0.8439	0.6333	0.7728
		0.0580	-0.6605	-0.7519		0.0580	-0.6605	-0.7519
			0.7983	0.7112			0.7983	0.7112
				1.4832				1.4832
2	0.1264	0.1683	0.6000	0.4552	1.2872	-0.5381	0.7034	0.8001
		0.0345	-0.0202	0.5865		0.0654	-0.4677	-0.4611
			0.0330	0.4338			0.8829	0.7359
				0.1335				1.4565
3	0.0797	-0.2282	0.2830	0.1446	1.7144	-0.8559	0.5397	0.6718
		0.0109	-0.1428	-0.5370		0.0713	-0.5424	-0.7894
			0.1297	0.2739			0.8791	0.7419
				0.3703				1.1138
4	0.3537	0.7997	0.5604	0.9362	0.3929	0.7514	-0.0676	0.4559
		0.1496	0.4936	0.7183		0.2252	-0.4059	0.0947
			0.0495	0.4971			0.6293	0.5049
				0.1899				0.8429
5	0.0237	0.4207	0.7228	0.9260	0.3590	0.7894	-0.0380	0.4461
		0.0097	0.3300	0.3096		0.2257	-0.2614	0.2419
			0.3233	0.8709			1.0790	0.6394
				0.4474				1.4483
6	0.0235	-0.2698	0.4256	0.3348	0.3880	0.6117	-0.2502	-0.1066
		0.0389	0.0241	-0.2133		0.1931	-0.4874	-0.3543
			0.1015	0.0004			0.5250	0.2304
				0.3614				0.4541
7	0.0451	0.4965	0.3212	-0.0509	0.1974	0.5156	0.1512	-0.2005
		0.0391	-0.3275	-0.0845		0.1298	0.3909	-0.1345
			0.2239	0.5090			0.6159	0.4743
				0.3982				0.8874

[1] The off-diagonal elements are the correlations; the diagonal elements are the variances.

Table 3.2: Estimated Covariance Matrices for Door Fitness

Stage (t)	$\hat{B}_t$			
1	0	0	0	0
	0	0	0	0
	0	0	0	0
	0	0	0	0
2	0.6948	-0.6962	-0.0009	0.1108
	0.0244	0.5413	0.0374	-0.0887
	-0.0225	-0.4437	0.9523	0.0166
	0.0412	0.2400	0.0580	0.9182
3	1.0587	-0.3374	0.0900	-0.0334
	-0.0693	0.2450	-0.0355	-0.0921
	-0.0079	0.8442	1.1550	-0.1721
	0.0931	-0.2171	0.7547	0.0487
4	-0.1133	-0.2108	-0.2472	0.1615
	0.0123	-0.5841	-0.3534	0.0055
	-0.2285	-0.3414	0.9668	-0.0854
	-0.5094	-0.3259	-0.0958	1.0472
5	0.8028	0.0922	-0.0405	0.0900
	0.0110	0.8789	-0.0875	0.0973
	-0.7602	0.8216	1.0964	0.1874
	-0.9288	1.0344	0.0395	1.2550
6	0.8894	0.2472	0.0235	-0.0497
	-0.0592	0.9317	0.0308	-0.1126
	0.2703	-0.4535	0.7106	-0.2670
	-0.2931	-0.1537	0.0045	0.2528
7	0.1980	0.5970	0.0732	-0.1973
	-0.3943	0.6602	-0.0367	-0.1557
	-1.1818	1.7945	0.1370	0.2591
	-0.5606	1.3000	0.7695	0.6986

Table 3.3: Estimates of B_t for Door Fitness.

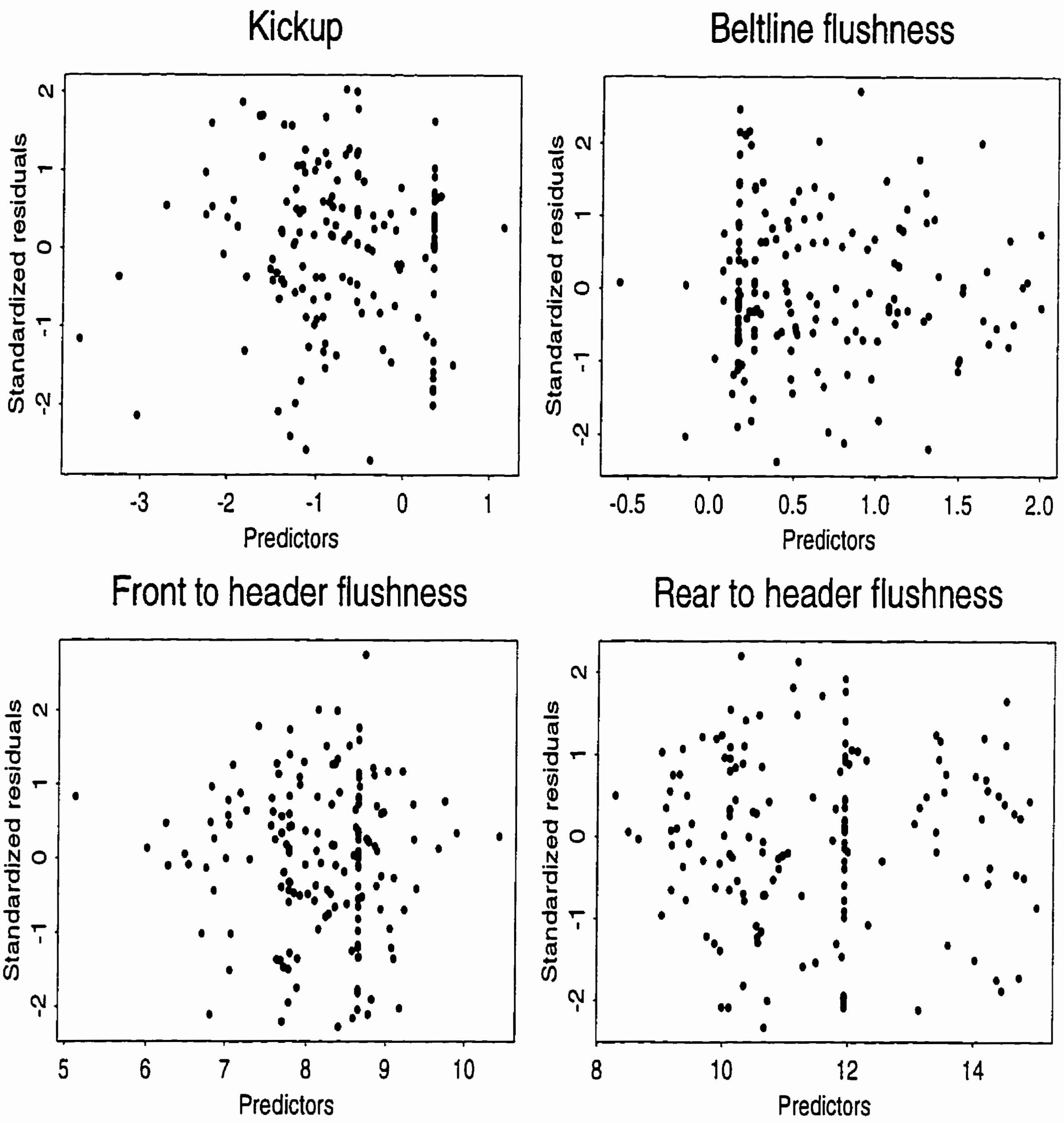


Figure 3.3: Door Hanging: Plots of residuals against the predictors $\hat{y}_{it|t-1}$.

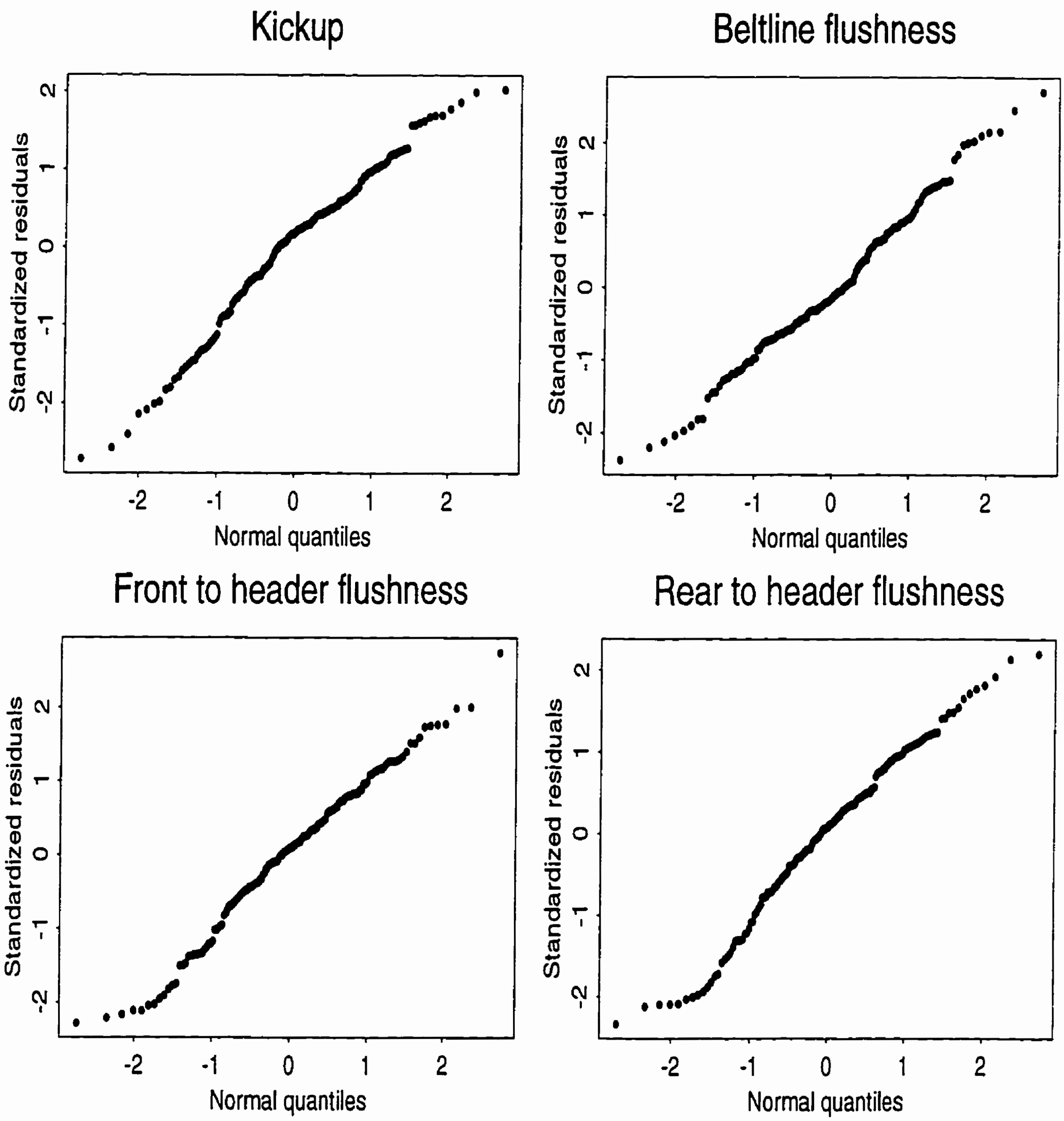


Figure 3.4: Door Hanging: Q-Q plots of standardized residuals.

Chapter 4

Random Effects Models for Recurrent Event Data

4.1 Introduction

Recurrent events arise when a number of subjects experience repeated occurrence of an event of interest. This kind of data has been frequently studied in the literature of longitudinal studies (Lawless, 1995; Clayton, 1994). The small bowel motility data described in Section 1.2.2 of Chapter 1 is a typical example with an additional feature of right censoring. Objectives in analyzing recurrent event data include estimation of the mean recurrence time (Aalen and Husebye, 1991), assessing the effects of covariates (e.g. treatment and control), estimation of the cumulative mean number of event recurrences (Lawless, 1995), prediction of next event recurrences (Chapter 5).

There are several approaches to the analysis of recurrent events (Lawless, 1995) but we will focus on modelling the recurrence times between events. Lawless and Fong (1997) review and discuss different choices of both modelling and analysis of inter-event times and point out the main difficulties that are encountered. We consider two common issues in modelling recurrent event data namely: inter-subject heterogeneity and within-subject dependence. Heterogeneity between subjects may be related to observable covariates or to unobservable random effects (often referred to as 'frailty'). Sources of these unobservable subject-level effects include unobserved subject-specific covariates, and measurement errors in time-independent covariates. Clearly, random effects induce correlation between a subject's recurrence times. Aalen and Husebye (1991) considered models where recurrence times are independent when conditioned on the random effects. Specifically, suppose there are N subjects and each subject i $(= 1, 2, \ldots, N)$ is observed over some time interval, say $(0, \tau_i)$. Let t_{ij} $(j = 1, 2, \ldots, n_i)$ be the jth recurrence time of subject i and t_{i,n_i+1} be the last recurrence time which is censored due to the planned end of surveillance. Also, we assume the censoring mechanism for τ_i is non-informative (Section 2.3.1 of Chapter 2) and only covariates which are constant between successive event recurrences are considered. Then, if u_i is the ith subject-specific effect, one model in Aalen and Husebye (1991) (A-H model) is

$$g(t_{ij})|_{u_i} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu + u_i, \sigma^2), \quad u_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega^2) \tag{4.1}$$

where g is some one-to-one function and "i.i.d." means independent and identically dis-

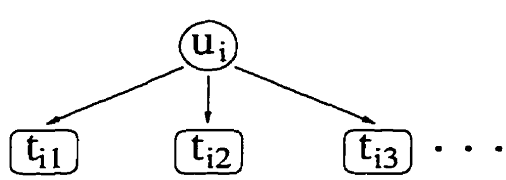
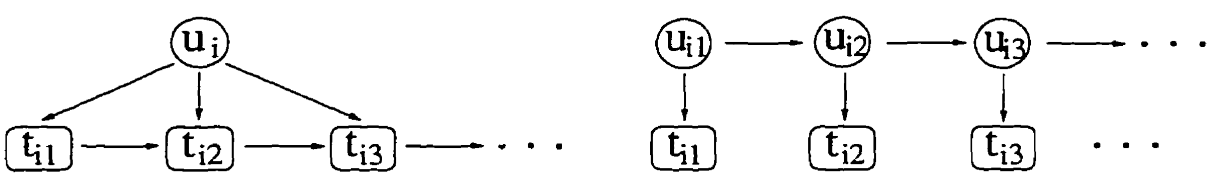


Figure 4.1: Independent recurrence times when conditioned on random effects u_i.

tributed. It is depicted (by borrowing the symbols from Clayton, 1994) in Figure 4.1. In cases with no time-dependent covariates, not only the marginal means and variances of the recurrence times are constant but also the correlation among them are the same. This kind of model structure is often unrealistic in practice as measurements closer in time are likely to be more strongly related. Hence modelling stochastic dependence between recurrence times of a subject by other models than that in Figure 4.1 is desirable.

Two general approaches to implanting non-constant correlation structure will be considered. One is to adopt certain dependence structure on the recurrence times, e.g. conditioned on u_i, a first order autoregressive process (AR(1)) on t_{ij}'s as shown in Figure 4.2(a). We will refer to this group of models as autocorrelated random effects models (AREMs). Another approach is to allow dynamic random effects where the random effects themselves follow an AR(1) process as shown in Figure 4.2(b). This group of models is also commonly called dynamic generalized linear models where we regress on a function of the mean measurements other than the identity function. These dynamic random effects models (DREMs) can be pushed further to have the transition process of the random effects depend on the past recurrence times. We will however delay discussing models of this type in the context of hazard-based models to Chapter 5. The AREMs have been popularly used in modelling longitudinal data. Wilson (1988) used them in parametric



(a) Autocorrelated random effects models. (b) Dynamic random effects models.

Figure 4.2: Two types of random effects models.

growth curve analysis and it is also more recently mentioned in Chan and Kuk (1997). Sutradhar (1990) has also consider a similar model with nested subject-specific effects. The DREMs have been entertained by Singh and Roberts (1992) and Jørgensen *et al.* (1996a) in modelling longitudinal counts data. However, there has not been any work in the literature to directly address the relationships between the two types of models. They are, though share some similarities, are quite distinct in nature. In this chapter, we will study their properties and differentiate their uses in longitudinal studies.

In the sequel, to model recurrence times, we can specify either the distribution or hazard function (e.g. Lawless and Fong, 1997). The hazard-based method will naturally lead to Cox's proportional hazard model (Cox, 1972) which is treated in Chapter 5. For the sake of easy discussion, we will put our attention in this Chapter on Normal-based models for which a Normal distribution is assumed on a certain suitably transformed value of t_{ij} as in (4.1). In the next section, we will first study the Normal-based approach and contrast the properties of autocorrelated and dynamic random effects models. Then, they are further studied by looking at the small bowel motility example from Aalen and Husebye (1991) in Section 4.3. Some concluding remarks and discussions are given in the last section.

4.2 Normal-Based Models

One of the main characteristics of recurrent event data is that the recurrence times are all non-negative and most likely positive. Thus, it may be necessary to assume that some transformation of t_{ij}, denoted as y_{ij}, is Normal if we wish to use the models here. Let the overall mean be $E(y_{ij}) = \mu_{ij}$, which may depend on some covariates which are time-independent during the jth event recurrence of subject i. Then an *autocorrelated random effects model* is specified as a variance components model with autocorrelated errors

$$\begin{aligned} y_{ij} &= \mu_{ij} + u_i + e_{ij}, \quad u_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega^2), \quad e_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_1^2), \qquad j = 1, 2, \ldots \\ e_{ij} &= \phi e_{i,j-1} + \epsilon_{ij}, \quad \epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad |\phi| < 1 \qquad j = 2, 3, \ldots \end{aligned} \tag{4.2}$$

where u_i is the subject-specific effect and ϕ measures the autocorrelation not explained by the u_i's. Thus, a AREM is composed of "autocorrelated errors" e_{ij} to impart within-subject correlation and "random effects" u_i component to account inter-subject heterogeneity. The model was also given in (2.4) of Chapter 2. The initial dispersion parameter σ_1^2 can be set to σ^2 (e.g. as in Chapter 3) resulting in a non-stationary process, or $\sigma^2/(1-\phi^2)$ when the recurrence times from a subject are stationary. However, it cannot be left arbitrary; otherwise it will be confounded with ω^2. Model (4.2) includes several commonly used sub-models. When $(\phi = 0, \omega = 0)$, it reduces to the ordinary renewal process (RP) model where all recurrence times (both between and within subjects) are independent. When $(\phi = 0, \omega > 0)$, we get back to the A-H model when there is only inter-subject heterogeneity and each subject forms a renewal process conditional on u_i. When $(0 < |\phi| < 1, \omega = 0)$, we have independent and identically structured AR(1) processes for subjects.

A *dynamic random effects model* is specified as

$$\begin{aligned} y_{ij} &= \mu_{ij} + u_{ij} + e_{ij}, \quad e_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_e^2), \quad u_{i1} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \omega^2), \qquad j = 1, 2, \ldots \\ u_{ij} &= \phi u_{i,j-1} + \epsilon_{ij}, \quad \epsilon_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad |\phi| < 1 \qquad j = 2, 3, \ldots. \end{aligned} \tag{4.3}$$

where the random effects u_{ij}'s are now not only subject-specific but also specific to the jth event recurrence and σ_e^2 measures the conditional response variability not explained by the u_{ij}'s (e.g. measurement errors).

Thus a DREM has the "dynamic random effects" u_{ij} to account for inter-subject heterogeneity as well as non-constant within-subject correlation. The DREMs also include the sub-models mentioned above. Specifically, we get the RP model when ($\omega = \sigma = 0$, $\sigma_e > 0$), the A-H model when ($\phi \to 1$, $\sigma = 0$, $\sigma_e > 0$), and the independent AR(1) model when ($0 < |\phi| < 1$, $\sigma_e = 0$). However, when $\phi = 0$, only $\sigma_e^2 + \omega^2$ and $\sigma_e^2 + \sigma^2$ are the estimable variance components. Note that, to avoid too much notation, except for μ_{ij} and y_{ij}, other symbols in (4.2) and (4.3) do not share exactly the same interpretation although they are consistent. For example, ω^2 in (4.2) is measuring the variability of the overall effect from subject heterogeneity, while in (4.3), it refers to the variability of the effect from subject heterogeneity on the first event recurrence time (see Table 4.1).

Both the AREMs and DREMs are natural extensions to the A-H model in (4.1) to accommodate non-constant within-subject correlation through a dynamic process (e.g. an AR(1)) to the errors (AREMs) and random effects (DREMs). Note that both models belong to the family of GSSMs defined in Section 2.1 of Chapter 2. Specifically, AREM in (4.2) can be formulated as

$$\begin{aligned} y_{ij} &= \mu_{ij} + (1 \quad 1) z_{ij} \\ z_{ij} | z_{i,j-1} &\sim \mathcal{N}\left(\begin{pmatrix} 1 & 0 \\ 0 & \phi \end{pmatrix} z_{i,j-1}, \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right) \end{aligned}$$

where $z_{i1} \sim \mathcal{N}\left(\left(\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right), \left(\begin{smallmatrix} \omega^2 & 0 \\ 0 & \sigma_1^2 \end{smallmatrix}\right) \right)$ and $z_{ij} = (u_i \quad e_{ij})^T$. The DREM in (4.3) can also be easily verified by writing the model as

$$\begin{aligned} y_{ij} \mid Y_i^{j-1}, \mu_{ij}, u_{ij} &\sim \mathcal{N}(\mu_{ij} + u_{ij}, \sigma_e^2) \\ u_{ij} \mid u_{i,j-1} &\sim \mathcal{N}(\phi u_{i,j-1}, \sigma^2) \end{aligned}$$

where $u_{i1} \sim \mathcal{N}(0, \omega^2)$.

Compared with the AREM in (4.2) which, without counting the μ_{ij}, has three parameters $\{\phi, \omega, \sigma\}$, the DREM in (4.3) has an extra parameter σ_e^2 to account for response-specific variability not explained by u_{ij} in (4.3). It is also interesting to see that the autocorrelated process for the errors e_{ij}'s in the AREM parallel to the dynamic process of the random effects u_{ij}'s in the DREM. In other words, there is some ambiguity about what we call autocorrelated errors in (4.2) and what we call dynamic random effects in (4.3). A key property of the random effects is that they are subject-specific (only indexed by i). This is opposed to response-specific effects (indexed by both i and j). The autocorrelated errors in (4.2) and dynamic random effects in (4.3) can be treated as compromise between random effects and response-specific effects. The ambiguity can be cleared by looking at the corresponding complementary components

$$y_{ij} - \mu_{ij} - e_{ij} = u_i \sim \mathcal{N}(0, \omega^2) \tag{4.4}$$

$$\text{and} \quad y_{ij} - \mu_{ij} - u_{ij} = e_{ij} \sim \mathcal{N}(0, \sigma_e^2) \tag{4.5}$$

for AREM and DREM respectively. The u_i's from AREM as viewed in (4.4) are constant for a single subject and hence they represent the subject-specific random effects. The e_{ij}'s from DREM as viewed in (4.5) are independent for each measurement of all subjects and hence they represent the response-specific effects not explained by u_{ij}.

If the DREM in (4.3) is extended to have certain correlation assumptions on $\{e_{ij}\}$, e.g. $Corr(e_{ij}, e_{i,j+s}) = \rho^s$ $(s > 0)$, then as $\rho \to 1$, the AREM can be viewed as a sub-model of DREM. However, basically, the two models are not nested although they intersect at some sub-models. To see this, we can look at their marginal properties which are summarized in Table 4.1. Both models have stationary and non-stationary versions of their autoregressive counterpart. In both cases, they share the same marginal means but different variances and lagged correlations. Influence from (initial) inter-subject heterogeneity ω^2 persists under the AREM but keeps diminishing under the DREM with rate controlled by the corresponding ϕ. Moreover, the limiting correlation shows that recurrence times which are infinitely apart are uncorrelated under the DREM but still mutually related under the AREM. Hence, choice between the autocorrelated and dynamic random effects models relies on whether the influence due to inter-subject heterogeneity will persist consistently over time. For example, AREMs are more appropriate when sources of inter-subject heterogeneity are missing important subject-specific and time-independent covariates, or there are measurement errors of some time-independent covariates. On the other hand, DREMs are desirable when (initial) inter-subject heterogeneity dilutes over time.

4.3 Application to Small Bowel Motility Data

Fitting both autocorrelated and dynamic random effects models can easily proceed by computing the mean and variance of y_{ij} conditional on its current past history. We denote them as $y_{ij|j-1}$ and $\sigma^2_{iy}(j|j-1)$ respectively, where we use the same notation as in Section 3.3. Then, with the assumption of non-informative right censoring, the log-likelihood function can be decomposed as

$$
\begin{aligned}
l &= \sum_{i=1}^{N}\left\{-\frac{n_i}{2}\log(2\pi)-\frac{1}{2}\sum_{j=1}^{n_i}\left[\log\sigma_{iy}^2(j|j-1)+\frac{(y_{ij}-y_{ij|j-1})^2}{\sigma_{iy}^2(j|j-1)}\right]\right. \\
&\qquad\left.+\log\left[1-\Phi\left(\frac{y_{i,n_i+1}-y_{i,n_i+1|n_i}}{\sigma_{iy}(n_i+1|n_i)}\right)\right]\right\} \qquad (4.6)
\end{aligned}
$$

where Φ is the cumulative distribution function of the standard Normal. Maximum likelihood estimates are obtained by maximizing l. The conditional moments for AREMs exist in closed form but are more efficiently computed from a modified Kalman filter recursion as described in Appendix C. The DREMs are already in a linear state-space form and the celebrated linear Kalman filter recursion can be conveniently applied. One could also use an EM algorithm (Dempster *et al.*, 1977) with the "complete data" as all the recurrence times as well as the random effects. However, we prefer direct maximization of the log-likelihood (as discussed in Section 2.2.2 of Chapter 2) (4.6) which is more efficient and convenient with standard maximization routines in common computing software (e.g. SAS/IML, MATLAB and GAUSS).

We consider the small bowel motility example as described in Section 1.2.2 of Chapter 1 for illustration. The complete dataset is reproduced from Aalen and Husebye (1991) in Appendix A.2. All computations were programmed in SAS/IML version 6.10 under Digital UNIX V3.2C. Optimization subroutine NLPNMS using the Nelder-Mead Simplex method was employed to maximize (4.6) with a fast convergence rate. Standard errors were obtained by inverting the observed Fisher's information matrix approximated by finite differences using subroutine NLPFDD. Both the identity and logarithmic transformation of t_{ij} (i.e. $y_{ij} = t_{ij}$ and $y_{ij} = \log t_{ij}$) were considered. Here there are no covariates present and we assume the recurrence times are identical in mean. Estimates and standard errors from the autocorrelated random effects model (4.2) and some of its sub-models are summarized in Table 4.2 while those from the dynamic random effects model (4.3) are presented in Table 4.3.

From Table 4.2 where l_{max} is the maximum log-likelihood value, we see that, in all cases, neither the random effects nor extra autocorrelation between recurrence times or their logarithmic version is significant. Hence, an ordinary renewal process model is sufficient for the data. This agrees with the results of Aalen and Husebye (1991) who fitted only the frailty model (i.e. $\phi = 0$) with $y_{ij} = t_{ij}$. From Table 4.3, both ω^2 and σ^2 are highly insignificant (no evidence they are not zero) and have again resulted in the same conclusion. Also, the estimates of ϕ are all close to 1 which reflects that the initial random effect (though insignificant) tends to persist over time and an AREM is more appropriate in this case.

A look at the data suggests the possibility of a longer first recurrence time. It is also reflected from the difference between the Kaplan-Meier estimates for the survivor functions of the first recurrence times and the others (Figure 4.3). Thus, we re-fitted the data by an A-H model with a different initial mean (μ_1) and variance (σ_1^2). Results are summarized in Table 4.4. The likelihood ratio statistics values are 3.82 ($y_{ij} = t_{ij}$) and 7.08 ($y_{ij} = \log t_{ij}$) which have p-values 0.15 and 0.03 respectively from a χ^2-distribution with 2 degrees of freedom. Thus, there is no significant difference based on $y_{ij} = t_{ij}$ and a marginally significant difference when based on $y_{ij} = \log t_{ij}$, between the first recurrence time and the rest in terms of the mean and variance. The two-sample non-parametric log-rank test statistic is 2.5 which gives a p-value of 0.12 (insignificant) from χ^2 with 1 degree of freedom.

Note that the distribution of all estimates, especially the variance estimates, may not

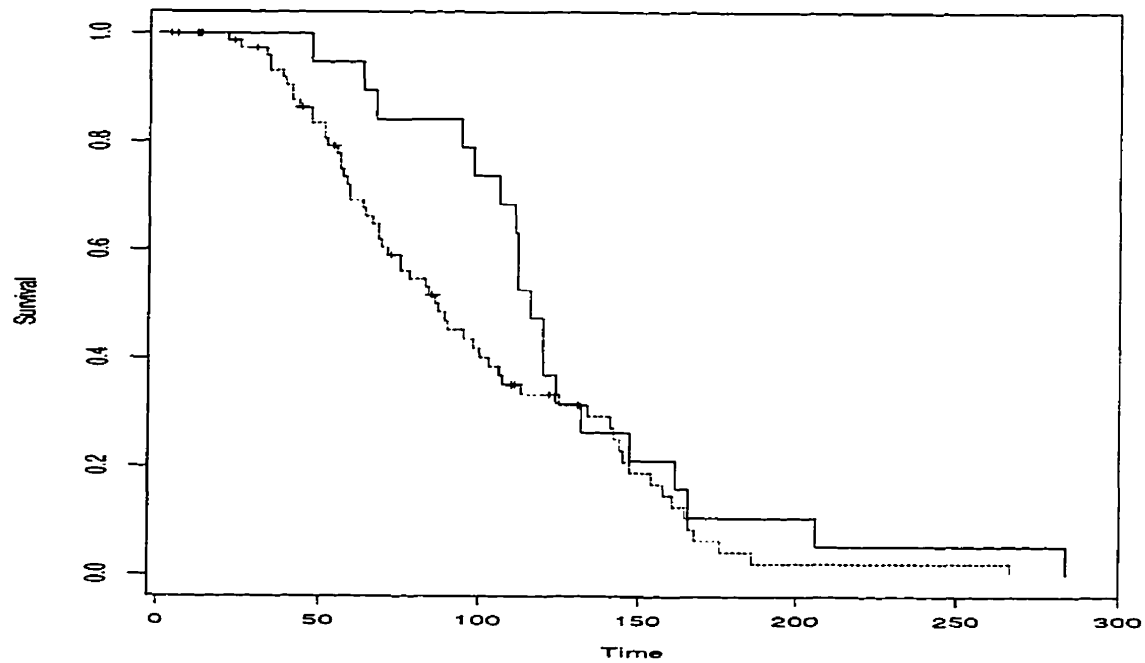
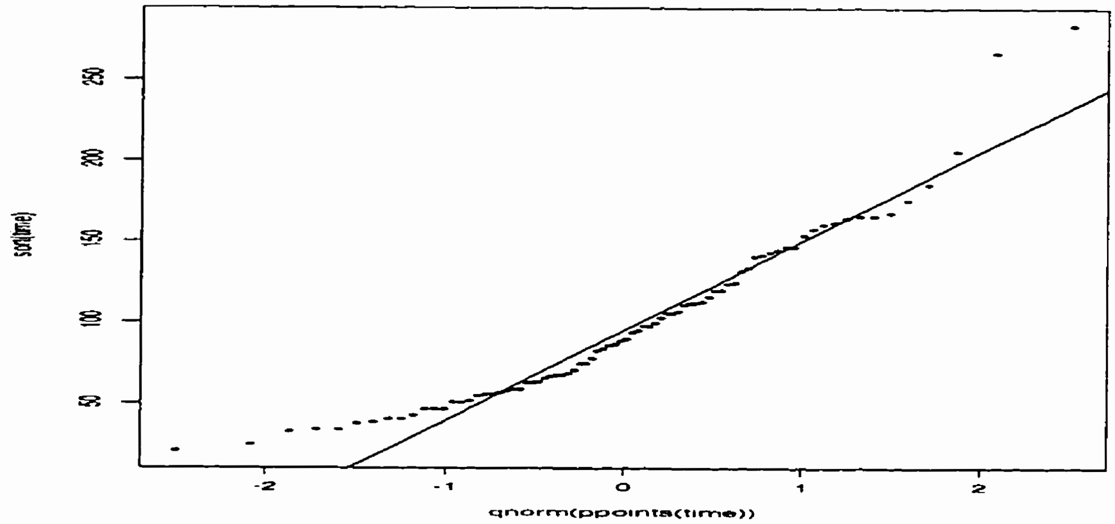
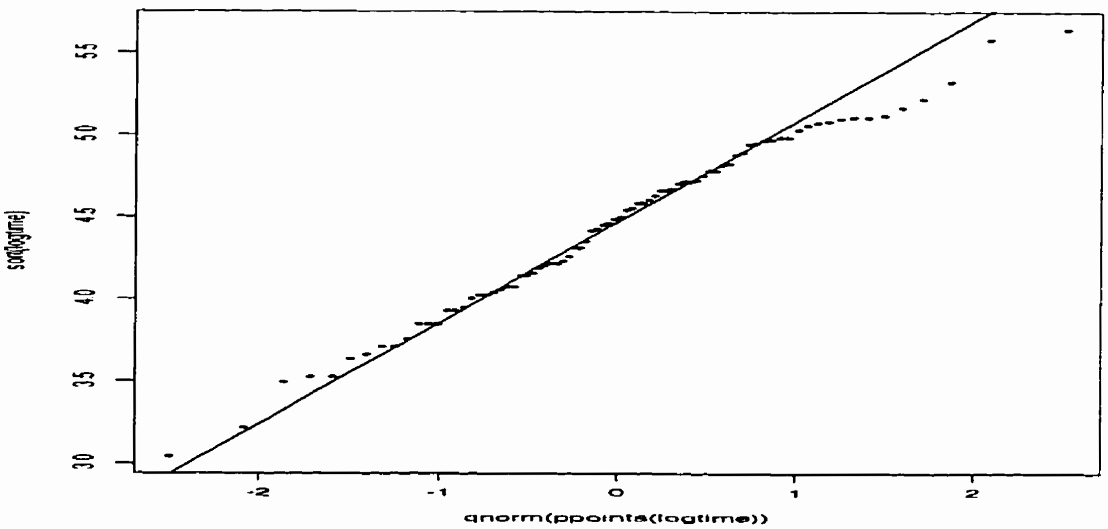


Figure 4.3: Plots of Kaplan-Meier estimates for the survivor functions of the first recurrence times (denoted by the solid line) and the others (denoted by the dotted line).

be close to Normal with only 19 subjects and a small number of recurrences for those models. So, if we need precise significance levels or confidence intervals, parametric bootstrapping (e.g. the end of Section 3.3 of Chapter 3) is more useful and feasible. Moreover, although preliminary analysis from Aalen and Husebye (1991) suggested that $y_{ij} = t_{ij}$ is a reasonable assumption, we find by looking at plots of non-parametric estimates that $\log t_{ij}$ is closer to Normal and that t_{ij} departs from Normality (Figure 4.4). With the small number of subjects and event recurrences, there is not a lot of power to detect lack of fit, however.



(a) $y_{ij} = t_{ij}$



(b) $y_{ij} = \log t_{ij}$

Figure 4.4: Q-Q plots of y_{ij} without censored periods. The straight line is the ideal case that the data are exactly Normal.

4.4 Conclusions and Discussion

We have studied and numerically illustrated both the autocorrelated and dynamic random effects models in longitudinal studies. Properties and comparisons of the two types of models have not been thoroughly examined in the literature. The AREMs are attractive by the fact that they "orthogonally" separate persistent inter-subject heterogeneity (u_i in (4.2)) and non-constant within-subject correlation (through e_{ij} in (4.3)). This is not shared by DREMs and the dynamic random effects (u_{ij} in (4.3)) account for both inter-subject heterogeneity and non-constant within-subject correlation. The key distinction of the two models is the persistent effect from initial inter-subject heterogeneity across time in AREMs while the effect keeps decreasing with time in DREMs. Thus, AREMs are used as strong derivative tracking models (e.g. Taylor *et al.*, 1994). For example, in the AIDS-related study of the natural history of CD4 T-cell counts, an immunologically weak subject who has an initial fast rate of decline of CD4 counts relative to other HIV-infected people will persist with a more rapid rate of decline of CD4 counts than will the others. Taylor *et al.* (1994) has indicated the desire for random effects which are dynamically changing with time to study measurements of the human immune system. They considered, instead of dynamic random effects, an AREM with the autocorrelated errors replaced by the sum of an integrated Ornstein-Uhlenbeck and independent error processes. Aalen (1994) has also a brief discussion of the need for dynamic random effects, for example, because of the induced weakness that results from the stresses of life. In these cases, DREMs are more appealing.

Generally, fitting both types of models is straightforward and convenient with the maximization routines in SAS/IML. In our applications with the Nelder-Mead Simplex method to maximize the log-likelihood function, different but rather arbitrary initial estimates were used to ensure a global maximum is attained. In maximizing the log-likelihood functions from DREMs, the Simplex method did not converge with certain initial guesses but only several tries were needed to obtain the estimates. On the whole, we did not encounter serious difficulties in fitting the models.

Note that we have not mentioned the very important issue of model checking. Assessing the fitness of both types of models can be generally pursued through the conditional residuals $r_{ij} = y_{ij} - y_{ij|j-1}$, which are independently distributed as Normal with mean 0 and variance $\sigma^2_{iy}(j|j-1)$ under the models. More work is also needed on testing and confidence interval procedures. The bootstrap seems to be the most appealing method but the usual likelihood ratio methods would also be applicable for large enough samples in both the number of subjects and event recurrences per subject. The bootstrap is illustrated in Chapter 5.

Finally, also note that the discussion in Section 4.3 depends on what is assumed about the μ_{ij} (we used $\mu_{ij} = \mu$) in looking at variance components. For example, a time trend may be confounded with the variance components when only a constant mean is modelled. However, with not too many event recurrences per subject in the small bowel motility data, it is hard to speculate on the mean profile. A model which adopts non-stationary drifts is considered in next chapter.

	Autocorrelated random effects model	Dynamic random effects model
	non-stationary	
	$\sigma_1^2 = \sigma^2$	$\omega^2 > 0$
$E(y_{ij})$	μ_{ij}	
$Var(y_{ij})$	$\omega^2 + \phi^{2(j-1)}\sigma_1^2 + \frac{1-\phi^{2(j-1)}}{1-\phi^2}\sigma^2$	$\phi^{2(j-1)}\omega^2 + \frac{1-\phi^{2(j-1)}}{1-\phi^2}\sigma^2 + \sigma_\epsilon^2$
$Cov(y_{ij}, y_{i,j+s})$: $s > 0$	$\omega^2 + \phi^s\left(\phi^{2(j-1)}\sigma_1^2 + \frac{1-\phi^{2(j-1)}}{1-\phi^2}\sigma^2\right)$	$\phi^s\left(\phi^{2(j-1)}\omega^2 + \frac{1-\phi^{2(j-1)}}{1-\phi^2}\sigma^2\right)$
limiting correlation[†]	$\frac{\omega^2}{\omega^2 + \frac{\sigma^2}{1-\sigma^2}}$	0
	stationary	
	$\sigma_1^2 = \frac{\sigma^2}{1-\phi^2}$	$\omega^2 = \frac{\sigma^2}{1-\phi^2}$
$E(y_{ij})$	μ_{ij}	
$Var(y_{ij})$	$\omega^2 + \frac{\sigma^2}{1-\phi^2}$	$\frac{\sigma^2}{1-\phi^2} + \sigma_\epsilon^2$
$Cov(y_{ij}, y_{i,j+s})$: $s > 0$	$\omega^2 + \phi^s\frac{\sigma^2}{1-\phi^2}$	$\phi^s\frac{\sigma^2}{1-\phi^2}$
limiting correlation[†]	$\frac{\omega^2}{\omega^2 + \frac{\sigma^2}{1-\sigma^2}}$	0

[†] limiting correlation $= \lim_{s\to\infty} Corr(y_{ij}, y_{i,j+s})$.

Table 4.1: Marginal properties of the autocorrelated and dynamic random effects models.

Model	μ	ϕ	ω^2	σ^2	l_{max}
		$y_{ij} = t_{ij}$			
RP	104.05	-	-	2699.22	-437.12
($\phi = 0, \omega = 0$)	(5.58)			(419.24)	
AR(1)	104.58	0.05	-	2694.96	-437.04
($0 < \lvert\phi\rvert < 1, \omega = 0$)	(5.96)	(0.12)		(418.69)	
A-H	106.82	-	262.47	2434.52	-436.41
($\phi = 0, \omega > 0$)	(6.89)		(277.83)	(426.22)	
non-stationary	106.97	-0.09	353.99	2327.96	-436.24
AREM	(6.87)	(0.16)	(317.72)	(439.64)	
stationary	106.93	-0.10	358.23	2316.60	-436.23
AREM	(6.86)	(0.16)	(316.72)	(444.70)	
		$y_{ij} = \log t_{ij}$			
RP	4.512	-	-	0.302	-73.12
($\phi = 0, \omega = 0$)	(0.059)			(0.048)	
AR(1)	4.531	0.141	-	0.296	-72.40
($0 < \lvert\phi\rvert < 1, \omega = 0$)	(0.068)	(0.116)		(0.047)	
A-H	4.542	-	0.029	0.271	-72.17
($\phi = 0, \omega > 0$)	(0.072)		(0.027)	(0.047)	
non-stationary	4.543	0.068	0.022	0.277	-72.08
AREM	(0.073)	(0.165)	(0.032)	(0.051)	
stationary	4.543	0.060	0.023	0.276	-72.09
AREM	(0.073)	(0.152)	(0.031)	(0.050)	

Table 4.2: Estimates and standard errors (in parenthesis) of the autocorrelated random effects model (4.2) and its sub-models.

Model	μ	ϕ	σ_ϵ^2	ω^2	σ^2	l_{max}
			$y_{ij} = t_{ij}$			
non-stationary	107.47	1.00	2418.87	195.77	33.60	-436.36
DREM	(7.09)	(0.36)	(478.21)	(392.86)	(219.24)	
stationary	106.79	0.879	2357.42	-	77.04	-436.37
DREM	(6.90)	(0.41)	(509.82)		(310.98)	
			$y_{ij} = \log t_{ij}$			
non-stationary	4.581	1.000	0.246	0.000	0.022	-70.46
DREM	(0.076)	(0.227)	(0.053)	(0.000)	(0.034)	
stationary	4.546	0.877	0.255	-	0.010	-71.90
DREM	(0.074)	(0.358)	(0.068)		(0.040)	

Table 4.3: Estimates and standard errors (in parenthesis) of the dynamic random effects model (4.3).

y_{ij}	μ_1	μ	ω^2	σ_1^2	σ^2	l_{max}
t_{ij}	125.63	100.25	179.53	2475.94	2297.74	-434.50
	(11.82)	(7.22)	(242.54)	(887.07)	(441.71)	
$\log t_{ij}$	4.754	4.472	0.020	0.144	0.293	-68.63
	(0.093)	(0.078)	(0.024)	(0.054)	(0.058)	

Table 4.4: Estimates and standard errors (in parenthesis) of the A-H model when the first recurrence time has different moments.

Chapter 5

A Dynamic Hazard-Based Model for Recurrent Event Data

5.1 Introduction

Suppose a recurrent event of interest is studied among N subjects. For each subject, the waiting times between successive event occurrences are recorded until a certain stopping time is reached and thus the last recurrence time may be censored. An example is the study of muscular activity (motility) of the small bowel discussed earlier in Section 1.2.2 of Chapter 1 and in Chapter 4. Modelling inter-subject variability and stochastic dependence between subject recurrence times using random effects is an important statistical issue in longitudinal studies, as discussed in Chapter 4. This chapter will address these two issues through hazard-based models. We focus on analysis in terms of the inter-event times. Other methods of analyzing recurrent events are given by various authors (Wei *et al.*, 1989, e.g.); Lawless (1995) gives a review.

Following from Cox (1972), we consider a class of proportional hazards models defined as

$$h_{ij}(t) = z_{ij} h_0(t) e^{\beta' x_{ij}(t)} \tag{5.1}$$

where $h_{ij}(t)$ and $x_{ij}(t)$ are the hazard function and time-dependent covariates for the jth event recurrence of subject i respectively and $h_0(t)$ is called the baseline hazard function. The variable z_{ij} specific to the jth event recurrence of subject i is called the dynamic frailty or dynamic random effect as it changes with the number of event recurrence. The ordinary frailty model (A-H model) considered in Aalen and Husebye (1991) is a special case of (5.1) when $z_{ij} = z_i$; see also Section 2.1.4 of Chapter 2. As mentioned in the last section of Chapter 4, dynamic frailty is a desired feature in some longitudinal studies.

Hazard-based dynamic frailty models are generally difficult to handle in terms of frequency-based inference. Even in the "static" case when z_{ij} are identical to e^z and z is Normally distributed, the likelihood is no longer tractable (Clayton, 1994). Our work here represents one of the first tractable developments. For example, a class of stationary dynamic frailty models can be obtained by taking $z_{i1} \sim Ga(\omega^{-2}, \omega^{-2})$ and

$$z_{ij} = \phi z_{i,j-1} + (1-\phi) z^*_{ij}, \quad 0 < \phi < 1; \quad j = 2, 3, \ldots \tag{5.2}$$

where $z^*_{i2}, z^*_{i2}, \ldots$ are independent (and of z_{i1}) with $Ga\left(\frac{1-\phi}{1+\phi}\omega^{-2}, \frac{1-\phi}{1+\phi}\omega^{-2}\right)$ distributions. Note that $Ga(\kappa, \nu)$ denotes the Gamma distribution with mean κ/ν and variance κ/ν^2. We note (5.2) is a stationary process up to the first two moments of z_{ij} which are 1 and ω^2 respectively. The lag s correlation of the z_{ij}'s is ϕ^s and the model defined by (5.1) and (5.2) gives the A-H model when $\phi = 1$. Petersen *et al.* (1996) discussed the fitting of a similar kind of frailty models with the frailty composed of a sum of independent Gamma variates, in another context. The likelihood function for these models can be expressed in closed form as a sum but the number of terms increases exponentially with the number of event recurrences per subject (Lawless and Fong, 1997). Models with log z_{ij}'s following a Gaussian distribution are also often proposed, but hard to handle computationally.

In general, likelihood based inference for dynamic random effects models outside the linear Normal framework is often computationally intractable (Aalen, 1994, Section 5). Various methods of approximation and other estimation approaches have thus been used. These include generalized estimating equations which solely depend on the first two moments of observations (Zeger and Liang, 1992), linearization of the transition component of a state space model (Jørgensen *et al.*, 1996a), Monte Carlo simulation (Carlin *et al.*, 1992), and posterior mode estimation (Fahrmeir and Tutz, 1994). Smith and Miller (1986) developed a class of non-Gaussian state space models with a multiplicative state transition process by assuming the observation process is Exponential after a 1-1 transformation. Under their model, all the predictive distributions (see below) can be numerically evaluated and thus the likelihood function can be readily maximized. Harvey and Fernandes (1989) considered an equivalent form of the model for count data without getting into the state space form on which the full model is actually based. This model was also adopted by Yue and Chan (1994) for recurrent event data, and we study it further in this chapter. In particular, we consider an extension of the model for recurrent events proposed by Yue and Chan (1994), and investigate its properties. The models in question have the ability to incorporate both inter-subject heterogeneity and non-stationary intra-subject variability in recurrence times. However, we will find that the applicability of the models in multi-subject studies is somewhat limited, and that a fairly large number of subjects

(N) and number of event recurrences per subject may be needed to estimate all model parameters adequately.

In the following exposition, we first introduce and discuss the use of the modelling scheme from Harvey and Fernandes (1989) when applied for recurrent events. Then an intensity based model is proposed in Section 5.3 together with an updating scheme for the random effects z_{ij}. The link with the model above is thus made explicit. Then, construction and computation of likelihood functions for censored recurrent event data are discussed in Section 5.4. The score and the Hessian matrix are seen to be easily computable and hence maximum likelihood estimates and standard errors of the estimates may be obtained. In Section 5.5, the set of small bowel motility data in Aalen and Husebye (1991) is used for illustration. A simulation study is given in Section 5.6 for further insight on the model. Finally, conclusions and some further remarks are given in Section 5.7.

5.2 Harvey and Fernandes Model

For convenience, we will introduce the model in a general non-state space form which allows the calculation of likelihood contributions. The state space formulation is given in Section 5.3. Let t_{ij} ($i = 1, 2, \ldots, N$, $j = 1, 2, \ldots, n_i$) be the first n_i uncensored recurrence times for subject i (i.e. times between successive events) and denote the last censored recurrence time as t_{i,n_i+1}. For brevity, indices i and j are assumed to run from 1 to N and 1 to $n_i + 1$ respectively unless otherwise specified. Then, the model can be characterized by

1. an observation model, $f(t_{ij}|z_{ij}, T_i^{j-1})$ where $T_i^j = \{t_{i1}, t_{i2}, \ldots, t_{ij}\}$, T_i^0 is the null set, and the z_{ij}'s are random effects whose dynamics are controlled by

2. a starting distribution, $g(z_{i1})$, and

3. a sequential updating scheme for the "priors", or random effects distributions from $g(z_{ij}|T_i^{j-1})$ to $g(z_{i,j+1}|T_i^j)$ after t_{ij} ($j = 1, 2, \ldots, n_i$) is observed.

These are what we need in order to compute the predictive densities $f(t_{ij}|T_i^{j-1})$ and hence to evaluate the likelihood function. The choice for $g(z_{i1})$ and $g(z_{ij}|T_i^{j-1})$ discussed below is motivated by the fact that

$$f(t_{ij}|T_i^{j-1}) = \int_0^\infty f(t_{ij}|z_{ij}, T_i^{j-1})g(z_{ij}|T_i^{j-1})dz_{ij} \tag{5.3}$$

which suggests the use of a natural conjugate family in the priors $g(z_{ij}|T_i^{j-1})$ for the sampling distributions $f(t_{ij}|z_{ij}, T_i^{j-1})$, in order to get a closed form for $f(t_{ij}|T_i^{j-1})$.

Now, the information gained from t_{ij} for updating $g(z_{ij}|T_i^{j-1})$ to $g(z_{i,j+1}|T_i^j)$ can be first utilized in the "posterior" $g(z_{ij}|T_i^j)$ and then linked to the next prior $g(z_{i,j+1}|T_i^j)$. In other words, in the updating scheme, we have an information update through the posterior as well as a non-stationarity update by linking the posterior to the next prior. For instance, if $f(t_{ij}|z_{ij}, T_i^{j-1})$ is Exponential with rate z_{ij}, the corresponding conjugate prior, $g(z_{ij}|T_i^{j-1})$, is Gamma. In this Harvey-Fernandes scheme, the updating controls the underlying mean and variance of the z_{ij} process, as follows:

$$\begin{aligned} E[z_{i,j+1}|T_i^j] &= E[z_{ij}|T_i^j] \\ \text{and } Var[z_{i,j+1}|T_i^j] &= \gamma^{-1}Var[z_{ij}|T_i^j] \end{aligned} \tag{5.4}$$

for $j = 1, 2, \ldots, n_i$ where $0 < \gamma < 1$ is a parameter that possibly depends on the past recurrence times.

Previous applications have mainly focused on a single time series of data and neglected the fact that γ can be time-dependent (Smith and Miller, 1986; Harvey and Fernandes, 1989). Lambert (1996b) considered a Poisson observation model and extended it to repeated count data allowing irregular sampling intervals by having a time-dependent parameter γ. Lambert (1996a) considered a version of the model robust to extreme values and included the special case of having non-informative prior $g(z_{i1})$. Yue and Chan (1994) considered a proportional hazards model with dynamic random effects designed to incorporate both inter- and intra-subject variability in recurrence times. Our model is essentially the same as theirs, except that the parameter γ can be time-dependent. This model is given in the next section.

5.3 An Intensity Based Model for Recurrent Events

We propose an intensity based model to account for intra-subject covariability as well as inter-subject heterogeneity. Suppose that, in addition to the recurrence times t_{ij}, we also observed a time-dependent covariate process, $x_{ij}(t)$. The model is characterized through the hazard function of T_{ij}, denoted by $h_{ij}(t)$, as

$$h_{ij}(t) = z_{ij} h_0(t) e^{\beta' x_{ij}(t)} \tag{5.5}$$

$$\text{with} \quad z_{ij}|_{T_i^{j-1}} \sim Ga(\kappa_{ij|j-1}, \nu_{ij|j-1}) \quad \text{and} \quad z_{i1} \sim Ga(1/\omega^2, 1/\omega^2) \tag{5.6}$$

where β is a vector of covariate parameters with the same dimension as $x_{ij}(t)$, and $h_0(t)$ is a baseline hazard function. Note that we have implicitly defined $\kappa_{i1|0} = \nu_{i1|0} = 1/\omega^2$.

The posteriors $g(z_{ij}|T_i^j)$ are found to be $Ga(\kappa_{ij}, \nu_{ij})$ where

$$\begin{aligned} \kappa_{ij} &= \kappa_{ij|j-1} + \delta_{ij} \\ \text{and} \qquad \nu_{ij} &= \nu_{ij|j-1} + \int_0^{t_{ij}} h_0(t) e^{\beta' x_{ij}(t)} dt \end{aligned} \tag{5.7}$$

where $\delta_{ij} = 0$ when $j = n_i + 1$ and 1 otherwise; see Appendix D.1. The "non-stationarity" update is taken as in (5.6), with

$$\kappa_{i,j+1|j} = \Psi(T_i^j)\kappa_{ij} \quad \text{and} \quad \nu_{i,j+1|j} = \Psi(T_i^j)\nu_{ij} \tag{5.8}$$

for $j = 1, 2, \ldots, n_i$ where $\Psi(T_i^j)$ can be any time-dependent positive-valued function taking values less than 1. Note that, through (5.8), the mean of z_{ij} is kept unchanged while the variability is increased; this allows a non-stationary process drift as for (5.4).

The set of model parameters includes β, ω^2, $\Psi(\cdot)$, as well as any parameters in $h_0(t)$. The initial z_{i1}'s are independent and identically distributed with mean 1 and variance ω^2. Hence ω^2 controls the initial variability due to subject heterogeneity. The function $\Psi(\cdot)$ reflects the within-subject stability. The closer $\Psi(\cdot)$ is to 1, the more stable is the process while the closer $\Psi(\cdot)$ is to 0, the less stable is the process. The limiting behaviour at the boundaries of ω^2 and $\Psi(\cdot)$ is better understood by looking at the state space form of the model which we now describe.

Consider a state space model based on the model (5.5) for the distribution of $T_{ij}|_{T_i^{j-1}, z_{i1}, \ldots, z_{ij}}$, and sequence of z_{ij}'s which follows the multiplicative transition process,

$$z_{i,j+1} = \Psi(T_i^j)^{-1} z_{ij} \eta_{ij}; \qquad j = 1, 2, \ldots, n_i \tag{5.9}$$

where $\eta_{ij} \sim Beta(\Psi(T_i^j)\kappa_{ij}, (1-\Psi(T_i^j))\kappa_{ij})$. Relationship (5.9) together with (5.5) defines a full state space model which is equivalent to the model represented by formulas (5.5) - (5.8); see Appendix D.2. With no subject heterogeneity, i.e. $\omega^2 \to 0$, $\kappa_{i1|0} \to \infty$ for all i and, by (5.7) and (5.8), we have $\kappa_{ij} \to \infty$ for all i, j. Thus $\eta_{ij} = \Psi(T_i^j)$ and (5.9) implies

$$z_{i,j+1} = z_{ij} = z_{i1} = 1 \quad \text{for all } i,\ j \geq 1.$$

Hence all between and within subject recurrence times are uncorrelated no matter the value of $\Psi(T_i^j)$. In other words, intra-subject correlation is triggered by the random effects. The functionality of $\Psi(\cdot)$ is best seen by noting, from (5.9), that

$$\begin{aligned} E[z_{i,j+1}|z_{ij}, T_i^j] &= z_{ij} \\ \text{and } \ Var[z_{i,j+1}|z_{ij}, T_i^j] &= (\Psi(T_i^j)^{-1} - 1)z_{ij}^2/(\kappa_{ij}+1). \end{aligned}$$

When $\Psi(T_i^j) \to 1$, $z_{ij} = z_{i1}$ for all $j \geq 1$ and Model (5.5) and (5.9) reduces to the Gamma frailty model considered by Aalen and Husebye (1991). When $\Psi(T_i^j) \to 0$, $Var[z_{i,j+1}|z_{ij}, T_i^j] \to \infty$ which, from (5.5), means a high instability of the hazard due to large process drifts. The random effect z_{i1} induces within-subject covariability which is adjusted by $\Psi(\cdot)$ to give a non-stationary process drift as more recurrence times are observed.

5.4 The Likelihood

A merit of using the conjugate-prior type model in (5.5) and (5.9) is the availability of the likelihood without the effort of numerical integration or the expense of inaccuracies from approximations. The likelihood function for recurrent event data with censoring can be constructed through (5.3) and the usual decomposition rule as

$$L = \prod_{i=1}^{N} \left(\prod_{j=1}^{n_i} f(t_{ij}|T_i^{j-1}) \right) Pr\{T_{i,n_i+1} \geq t_{i,n_i+1}|T_i^{n_i}\}$$

where each individual predictive density is computed, using (5.5) and (5.6), by integrating over z_{ij} in $E_{z_{ij}|T_i^{j-1}}[\, f(t_{ij}\,|\,z_{ij}, T_i^{j-1})\,]$ and similarly for the last term with the censoring time. This gives

$$f(t_{ij}|T_i^{j-1}) = h_0(t_{ij}) e^{\beta' x_{ij}(t_{ij})} \kappa_{ij|j-1} \frac{\nu_{ij|j-1}^{\kappa_{ij|j-1}}}{\nu_{ij}^{\kappa_{ij|j-1}+1}} \tag{5.10}$$

$$\text{and} \quad Pr\{T_{i,n_i+1} \geq t_{i,n_i+1}|T_i^{n_i}\} = \left(\frac{\nu_{i,n_i+1|n_i}}{\nu_{i,n_i+1}} \right)^{\kappa_{i,n_i+1|n_i}}. \tag{5.11}$$

Thereupon, the log-likelihood function can be written as

$$l = \sum_{i=1}^{N} \left\{ \sum_{j=1}^{n_i} \left[\log h_0(t_{ij}) + \beta' x_{ij}(t_{ij}) + \log(\kappa_{ij|j-1}) - \log \nu_{ij} \right] + \sum_{j=1}^{n_i+1} \kappa_{ij|j-1} \left[\log \nu_{ij|j-1} - \log \nu_{ij} \right] \right\} \tag{5.12}$$

which can be evaluated numerically by computing $\kappa_{ij|j-1}$, $\nu_{ij|j-1}$ and ν_{ij} recursively using (5.7) and (5.8). Note that when there are no random effects, i.e. $\omega^2 \to 0$, (5.10) and (5.11) reduces to the densities from independent recurrence times (see Appendix D.3), but the degeneracy does not cause much problem in our applications (see Section 5.5). The score function and Hessian matrix can be routinely evaluated; see Appendix D.4 for the case of a Weibull baseline hazard function with a time-independent discounting constant $\Psi(T_i^j) = \psi$ when there is no covariate process. Common optimization algorithms such as the Downhill Simplex Method (which does not require the first and second derivatives) and the Newton-Raphson Method are usually sufficient in searching for maximum likelihood estimates.

5.5 Application to Small Bowel Motility Data

The model in Section 5.3 was fitted to a set of small bowel motility data from Aalen and Husebye (1991). There were 19 subjects with no covariates. Successive MMC periods were recorded over a fixed time period. As in the Gamma frailty model in Aalen and Husebye (1991), we considered a Weibull baseline hazard function, i.e.

$$h_0(t) = bt^k; \quad b > 0, \; k > -1.$$

We assumed $\Psi(\cdot) = \psi$. Initial estimates for (b, k, ω^2) were obtained by fitting a Gamma frailty model as in Aalen and Husebye (1991) and ψ was initially taken as 0.5. To avoid boundary value problems and highly correlated estimates, the set of parameters $\theta = (b, k, \omega^2, \psi)$ was transformed to $\theta_U = (u, \delta, \gamma, \tau)$ where

$$u = \frac{1}{k+1}\log(\frac{k+1}{b}), \quad \delta = \log(k+1), \quad \gamma = \log(\omega^2) \quad \text{and} \quad \tau = \gamma - \log(\frac{\psi}{1-\psi}). \tag{5.13}$$

The corresponding log-likelihood, score and Hessian are given in Appendix D.4. We programmed in SAS/IML Version 6.10 in a DEC alpha, Digital UNIX (OSF/1) V3.2 system and a nonlinear optimization subroutine, NLPNMS (Nelder-Mead Simplex method), was employed for likelihood maximization.

The log-likelihood was maximized at -429.13 and maximum likelihood estimates together with their asymptotic standard errors and correlation coefficients are shown in Table 5.1. The sampling distribution of the estimates was examined by 500 bootstrap samples. Figure 5.2 exhibits plots of histograms for various estimates and shows a fairly symmetric empirical distribution for u and δ. Using the Normal assumption, the 95% confidence intervals for u and δ are 4.75 ± 0.13 and 0.83 ± 0.19 respectively and hence the confidence interval for b and k are

$$(3.16 \times 10^{-5}, 5.74 \times 10^{-5}) \quad \text{and} \quad (0.90, 1.77)$$

respectively. The seemingly bi-modal behaviour for the estimates of τ is due to the flatness of the likelihood as τ gets small when the hypothesized value of ψ is close to 1. In a careful look, estimates of τ smaller than -9 usually have scores greater than -10^{-3} which keep increasing when the estimates are pushed smaller. Figure 5.3 shows the increasing score for a typical iterated estimate of -13.31 for τ. Thus the left cluster of the estimates for τ should actually spread over towards $-\infty$ and the empirical distribution of both γ and τ have a long left tail. Indeed, as can be seen in the next simulation study, the bi-modal behaviour disappears for small values of ψ.

The likelihood ratio statistic for the null hypothesis $\omega^2 = 0$ is $R = 2.58$. However, since $\omega^2 = 0$ lies on the parameter space boundary, R is not distributed as a simple chi-square. The empirical significance level of R is 0.09. It was computed by bootstrapping 1,000 samples with $(b, k, \omega^2) = (0.000044, 1.28, 0)$ and calculating the proportion of likelihood ratio statistics for testing $\omega^2 = 0$ that are greater than R. Thus, we arrive at the conclusion as in Aalen and Husebye (1991) that the data do not exhibit strong evidence of subject heterogeneity. Indeed, a graphical test of the Weibull model does not reveal a

Parameters	Estimates	[a] Asymptotic correlation matrix			
u	4.7525	0.0658	0.0789	-0.2031	0.1760
δ	0.8261		0.0991	0.3409	0.2237
γ	-1.9304			0.9084	-0.3668
τ	-6.2818				52.9457
b	0.000044	0.000044	-0.9887	-0.3026	0.2683
k	1.2844		0.2265	0.3409	-0.2165
ω^2	0.1451			0.1318	0.3815
ψ	0.9873				0.6694

[a] The off-diagonal elements are the asymptotic correlations; the diagonal elements are the asymptotic standard errors.

Table 5.1: Maximum likelihood estimates for a set of small bowel motility data.

serious model departure (Figure 5.1). Consequently, the value of ψ becomes irrelevant as mentioned in the last paragraph of Section 5.3. Indeed, the likelihood ratio statistic for testing the null hypothesis $\omega^2 = 0$ against the alternative $\omega^2 > 0$ but $\psi = 1$ (the frailty model) is only slightly smaller ($\approx 1.9 \times 10^{-4}$) than R.

5.6 Simulation Study

To determine the efficacy of the estimators and enhance our understanding of the model, we performed a simulation study at some hypothetical but plausible values of the parameters. We assumed the same number of subjects (19) and censoring times as in the small bowel motility data. The baseline hazard is taken from a Weibull distribution, i.e. $h_o(t) = bt^k$; $b > 0$, $k > -1$ and the discounting function is taken as a constant, i.e. $\Psi(\cdot) = \psi$. Then, with some specified value of $\theta = (b, k, \omega^2, \psi)$, a set of recurrent event times with censoring was generated from the following algorithm. For each subject i,

1. Take the censoring time, s_i, from the ith subject in the small bowel motility data.

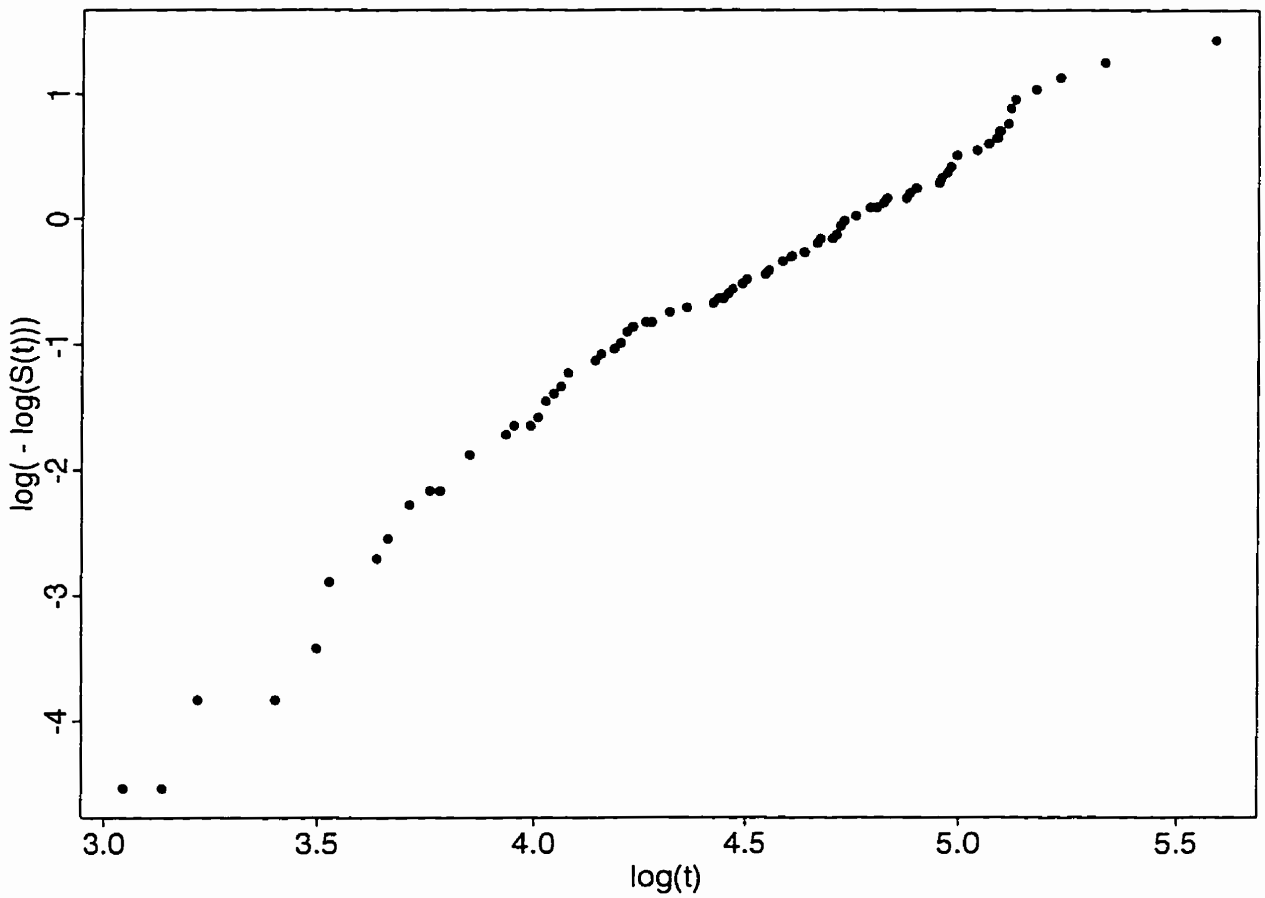


Figure 5.1: A graphical check of the Weibull model. A correct model should give a linear graph.

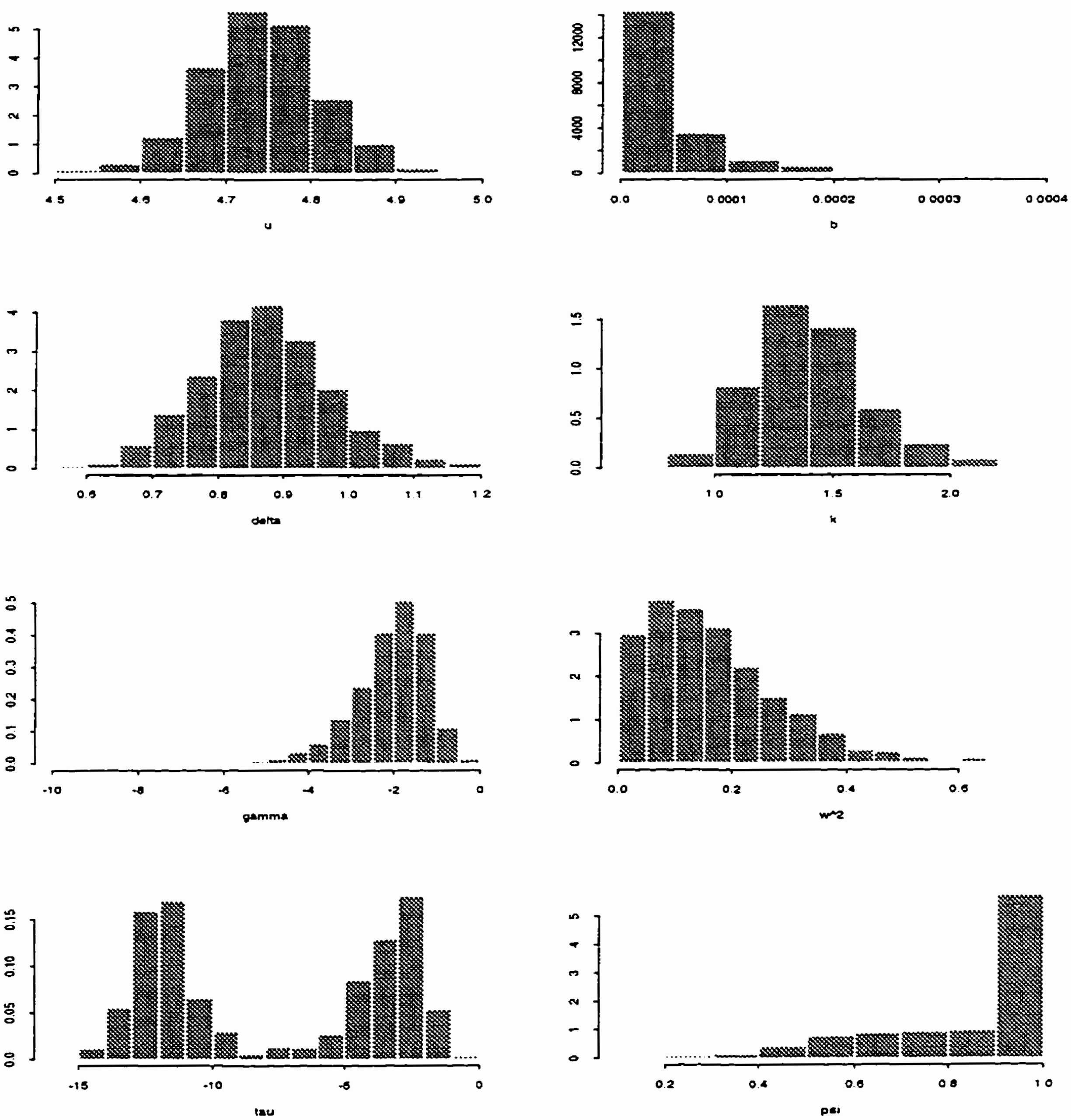


Figure 5.2: Histograms for estimates from a bootstrap sample of size 500.

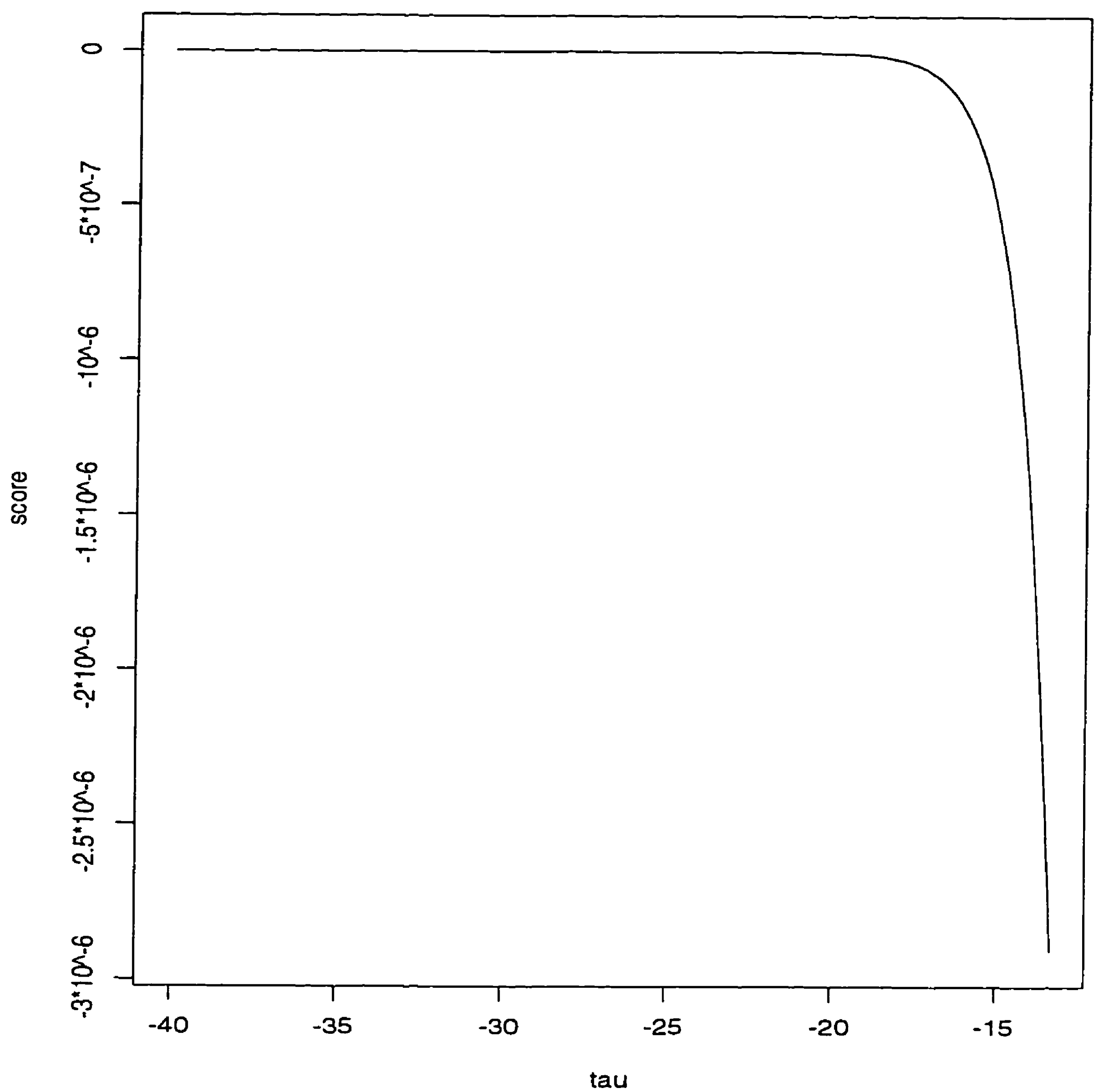


Figure 5.3: Plot of scores against τ for an iterated estimate of -13.31 for τ in a simulated data.

2. Set $\kappa_{i1|0} = \nu_{i1|0} = 1/\omega^2$ and generate z_{i1} from $Ga(\kappa_{i1|0}, \nu_{i1|0})$.

3. Generate t_{i1} from its intensity function $z_{i1}h_o(t)$.

4. Set $j = 1$.

5. Compute $\kappa_{ij} = \kappa_{ij|j-1} + 1$.

6. Generate η_{ij} from $Beta(\psi\kappa_{ij}, (1-\psi)\kappa_{ij})$ and evaluate $z_{i,j+1} = z_{ij}\eta_{ij}/\psi$.

7. Generate $t_{i,j+1}$ from its intensity function $z_{i,j+1}h_o(t)$.

8. If the total time span, $\sum_{k=1}^{j+1} t_{ik}$, is less than s_i, compute $\kappa_{i,j+1|j} = \psi\kappa_{ij}$, set j = j +1 and go to step 5.

9. Put $n_i = j$ and $t_{i,n_i+1} = s_i - \sum_{j=1}^{n_i} t_{ij}$.

From this, we generated 1,000 samples. For each simulated sample m ($m = 1, 2, \ldots, 1000$), the log-likelihood (D.1) was maximized with respect to $\theta_U = (u, \delta, \gamma, \tau)$, the transformed form of θ from (5.13), and the maximum likelihood estimates $\hat{\theta}_U^{(m)}$ as well as the asymptotic correlation matrix with the diagonal elements replaced by the asymptotic standard errors $\hat{\Sigma}_U^{(m)}$ were obtained. The efficacy of the estimators was assessed by computing the following summary statistics,

$$\bar{\hat{\theta}}_U = \frac{1}{1000}\sum_{m=1}^{1000}\hat{\theta}_U^{(m)}, \quad \bar{\hat{\Sigma}}_U = \frac{1}{1000}\sum_{m=1}^{1000}\hat{\Sigma}_U^{(m)}, \quad \Sigma_U = \text{sample correlation matrix with diagonal elements replaced by the sample standard deviations}$$

and the 95% coverage which is the proportion of the 95% confidence intervals computed using the Normal assumption, i.e. $\hat{\theta}_U^{(m)} \pm 1.96 \times$ standard error$(\hat{\theta}_U^{(m)})$, that includes the hypothesized value θ_U. With obvious notation, $\bar{\hat{\theta}}$, $\bar{\hat{\Sigma}}$, Σ, and the 95% coverage were also computed.

Based on the previous example, the values of b and k are taken to be 0.00005 and 1.5 respectively. The random effects variance ω^2 is taken at two values: a small value 0.1 and a large value 1.5. In either case, ψ takes values in $\{0.1, 0.5, 0.9\}$. All summary statistics at different values of the parameters are tabulated in Table 5.2 for $\omega^2 = 0.1$ and Table 5.3 for $\omega^2 = 1.5$.

In Table 5.2 with $\omega^2 = 0.1$, except for τ, there is a fair agreement between the average of estimates and the true values as well as between the estimated standard errors and the corresponding sample estimates. The coverage for θ_U agrees very much with 0.95. The finite sample approximation by Normal distribution for the sampling distribution of $\hat{\theta}_U$ can be reasonably assumed despite the long tail distribution for $\hat{\gamma}$ and $\hat{\tau}$ as in the numerical example. Again, the great discrepancy between the standard error of $\hat{\tau}$ and its small finite sample standard deviation is due to the flatness of the likelihood when ψ is close to 1. But this does not create any serious disagreement for $\hat{\psi}$. Indeed, further study of the empirical distribution of $\hat{\tau}$ for smaller values of ψ, say < 0.5 shows the seemingly bi-modal behaviour does not appear. This can also be seen from a much better agreement between the standard errors and the corresponding finite sample estimates in Table 5.2(a). In Table 5.3 with $\omega^2 = 1.5$, similar phenomena are observed with better overall agreements of the estimates.

Thus with only a small number of subjects and around 10 to 20 recurrence times per subject, the asymptotic approximations perform reasonably well when there are large random effects. Standard errors of the estimates tend to be smaller for large values of ψ which corresponds to a more stable process. Interval estimates should be computed by a

Normal approximation for $\hat{\theta}_U$, which gives closer to nominal coverage.

5.7 Concluding remarks

We have employed a proportional hazards model with dynamic random effects in modelling inter-subject heterogeneity and non-stationary intra-subject variability in recurrent events with censoring. It is flexible enough to incorporate random effects and pick up non-stationary process drifts through z_{i1} and $\Psi(\cdot)$ (as also discussed in Harvey and Fernandes, 1989, for the case with count data). The likelihood function, which is usually intractable outside the linear Normal framework, can be easily evaluated and differentiated from (5.12).

Now, prediction of the next event recurrence is based on the mean of $T_{i,n_i+1}|_{T_i^{n_i},T_{i,n_i+1}\geq t_{i,n_i+1}}$ which can be easily shown, from (5.11), to be $t^*_{i,n_i+1} = t_{i,n_i+1} + t_{iW}$ where

$$t_{iW} = \nu_{i,n_i+1}^{\kappa_{i,n_i+1|n_i}} \int_{t_{i,n_i+1}}^{\infty} \frac{dt}{\left(\nu_{i,n_i+1|n_i} + \int_0^t h_0(u) e^{\beta' x_{i,n_i+1}(u)} du\right)^{\kappa_{i,n_i+1|n_i}}}. \tag{5.14}$$

Thus the predicted waiting time until the next event recurrence for subject i is t_{iW} and the $(m+1)$th mean recurrence time $(m \geq n_i + 1)$ can be similarly deduced, from (5.10), as

$$t^*_{i,m+1|m} = \kappa_{i,m+1|m} (\nu^*_{i,m+1|m})^{\kappa_{i,m+1|m}} \int_0^{\infty} \frac{h_0(t) e^{\beta' x_{i,m+1}(t)}}{\left(\nu^*_{i,m+1|m} + \int_0^t h_0(u) e^{\beta' x_{i,m+1}(u)} du\right)^{\kappa_{i,m+1|m}}} dt \tag{5.15}$$

where $\nu^*_{i,m+1|m} = \Psi(T_i^{*m}) \left(\nu^*_{i,m|m-1} + \int_0^{t^*_{i,m|m-1}} h_0(t) e^{\beta' x_m(t)} dt\right)$, $\nu^*_{i,n_i+1|n_i} = \nu_{i,n_i+1|n_i}$ and $T_i^{*m} = \{T_i^{n_i}\} \cup \{t^*_{i,n_i+1}, \ldots, t^*_{im}\}$. Note that our results assumed the covariate process is known before prediction is made. Integration in (5.14) and (5.15) depends on the complexity of $h_0(t)$ which is manageable in most applications.

Note that we have not looked at the very important issue of model diagnostics. Some thoughts on diagnostic checking are to perform "post-sample" diagnostics where we shorten the surveillance time of each subject. That is, we are discarding some observations (post-sample) but the last retained one is still censored. Then the present model is fitted to the retained dataset and predicted values, from (5.14) and (5.15), are computed and compared with those recurrence times in the post-sample. However, in the case when we do not have too many observations for each subject, discarding observations may result in a too small sample which is not informative enough for the model to be well-fitted. Another approach would be by using parametric bootstrapping to generate samples with values of the parameters taken as the estimates from fitting the original data. The bootstrap samples are then compared with the original data to assess the fit of the model. Sufficiency of ordinary proportional hazards models when subjects forming renewal processes can be assessed by using hazard-based residuals (Lawless, 1982) defined as

$$e_{ij} = \begin{cases} \hat{H}(t_{ij}) & \text{if } t_{ij} \text{ is not censored,} \\ \hat{H}(t_{ij}) + 1 & \text{if } t_{ij} \text{ is censored} \end{cases}$$

where $\hat{H} = -\log \hat{S}(t_{ij})$ and $\hat{S}$ are the estimated cumulative hazard and survivor function under the ordinary proportional hazards model with a chosen baseline hazard function. Then, the model is sufficient if a plot of the logarithm of the Kaplan Meier estimate of the e_{ij}'s versus e_{ij} is roughly a straight line with slope -1. However, residual analysis to assess proportional hazards models with a dynamic frailty, for example, the validity of the baseline hazard function, is still desirable.

It is also worth studying the possible use of semi-parametric methods when the baseline hazard function h_0 is not specified. For example, semi-parametric analysis can be generally pursued through the EM algorithm as in Petersen *et al.* (1996). In our model, the logarithm of the complete data (assuming the frailties are known) likelihood can be written as

$$\begin{aligned} l_c(\beta, h_0, \omega^2, \Psi) &= \log f(z_{i1}) + \sum_{j=2}^{n_i} \log f(z_{i,j+1} \mid z_{ij}, T_i^{j-1}) + \sum_{j=1}^{n_i} \log(z_{ij}) \\ &+ \sum_{j=1}^{n_i} \log \left[h_0(t_{ij}) e^{\beta' x_{ij}(t_{ij})} \right] - \sum_{j=1}^{n_i+1} z_{ij} \int_0^{t_{ij}} h_0(u) e^{\beta' x_{ij}(u)} \, du \end{aligned} \quad (5.16)$$

Assuming ω^2 is known and noting that the last two terms of l_c in (5.16) only involve z_{ij} linearly, the E-step requires

$$E(z_{ij} \mid T_i^{n_i+1}\text{'s}). \quad (5.17)$$

Then, in the M-step, we maximize (5.16) with respect to the baseline hazard function h_0 and β, after substituting z_{ij} as (5.17). Note that only the last two terms in (5.16) are needed to be maximized and this is equivalent to the usual Cox regression analysis (z_{ij}'s in the M-step are now known) which estimates the baseline hazard function through the Nelson-Aalen estimator. Estimates of ω^2 and Ψ can be obtained by maximizing the observed data log-likelihood as given by (5.10) and (5.11). The EM-step together with the estimation of ω^2 and Ψ iterates until convergence. The key is to compute (5.17) which can be generally approximated by the Gibbs sampler (e.g. Gelfand and Smith, 1990) by noting $f(z_{ij}\text{'s} \mid t_{ij}\text{'s})$ is proportional to (5.16). However, further study of the convergence properties is needed.

θ_U	$\bar{\hat{\theta}}_U$	$\bar{\hat{\Sigma}}_U$				Σ_U				95% coverage
4.3279	4.3250	0.0922	-0.1075	-0.3740	-0.3844	0.1051	-0.1956	-0.4154	-0.4533	0.92
0.9163	0.9459		0.1332	0.3550	0.5128		0.1446	0.3899	0.5360	0.94
-2.3026	-2.5601			0.8657	0.8940			0.9184	0.9057	0.96
-0.1054	-0.2356				0.5997				0.7289	0.95
θ	$\bar{\hat{\theta}}$	$\bar{\hat{\Sigma}}$				Σ				95% coverage
0.00005	0.00008	0.00009	-0.9841	-0.2912	0.0941	0.00023	-0.4339	-0.1294	0.3285	0.77
1.5	1.6023		0.3482	0.3550	-0.0621		0.3830	0.4150	-0.1066	0.95
0.1	0.1070			0.0743	0.7010			0.0851	0.4965	0.82
0.1	0.0942				0.0403				0.0353	0.91

(a) $(b, k, \omega^2, \psi) = (0.00005, 1.5, 0.1, 0.1)$.

θ_U	$\bar{\hat{\theta}}_U$	$\bar{\hat{\Sigma}}_U$				Σ_U				95% coverage
4.3279	4.3259	0.0684	-0.1097	-0.3465	-0.3841	0.0707	-0.1287	-0.3491	-0.3577	0.93
0.9163	0.9313		0.0909	0.3180	0.4796		0.0933	0.2781	0.4293	0.94
-2.3026	-2.5622			0.9178	0.8811			0.9604	0.7352	0.96
-2.3026	-2.5572				1.2650				0.9418	0.97
θ	$\bar{\hat{\theta}}$	$\bar{\hat{\Sigma}}$				Σ				95% coverage
0.00005	0.00006	0.00005	-0.9813	-0.2531	0.2010	0.00005	-0.8309	-0.1698	0.2227	0.84
1.5	1.5489		0.2321	0.3180	-0.1897		0.2409	0.2893	-0.2244	0.95
0.1	0.1072			0.0801	0.5779			0.0788	0.4907	0.86
0.5	0.4913				0.0937				0.1038	0.93

(b) $(b, k, \omega^2, \psi) = (0.00005, 1.5, 0.1, 0.5)$.

Table 5.2: Summary statistics from the dynamic proportional hazards model with $\omega^2 = 0.15$.

θ_U	$\hat{\theta}_U$	$\hat{\Sigma}_U$				Σ_U				95% coverage
4.3279	4.3239	0.0516	0.0204	-0.1915	-0.1477	0.0505	0.0503	-0.1189	-0.1109	0.95
0.9163	0.9364		0.0793	0.3283	0.2588		0.0786	0.2272	0.1807	0.95
-2.3026	-2.6262			1.4511	0.2112			1.1497	-0.0735	0.96
-4.4998	-9.001				396.87				5.81	0.95
θ	$\hat{\theta}$	$\hat{\Sigma}$				Σ				95% coverage
0.00005	0.00005	0.00004	-0.9865	-0.2918	0.1067	0.00004	-0.8520	-0.2312	0.1208	0.832
1.5	1.5588		0.2032	0.3283	-0.1128		0.2035	0.3212	-0.1475	0.96
0.1	0.1080			0.0877	0.3174			0.0816	0.3604	0.93
0.9	0.8649				0.1374				0.1573	0.53

(c) $(b, k, \omega^2, \psi) = (0.00005, 1.5, 0.1, 0.9)$.

θ_U	$\bar{\hat{\theta}}_U$	$\bar{\hat{\Sigma}}_U$				Σ_U				95% coverage
4.3279	4.3334	0.2217	-0.4514	-0.5451	-0.5189	0.2217	-0.4402	-0.5886	-0.5980	0.94
0.9163	0.9497		0.1975	0.6169	0.7334		0.2007	0.6450	0.7455	0.96
0.4055	0.5144			0.4881	0.8285			0.4822	0.8686	0.94
2.6027	2.4724				0.6368				0.6110	0.97
θ	$\bar{\hat{\theta}}$	$\bar{\hat{\Sigma}}$				Σ				95% coverage
0.00005	0.00010	0.00014	-0.9584	-0.5089	0.4549	0.00020	-0.5321	-0.2624	0.4014	0.75
1.5	1.6390		0.5292	0.6169	-0.4643		0.5607	0.6856	-0.4347	0.97
1.5	1.8767			0.8890	-0.1110			0.9630	-0.1742	0.96
0.1	0.1275				0.0396				0.0357	0.99

(a) $(b,\ k,\ \omega^2,\ \psi) = (0.00005,\ 1.5,\ 1.5,\ 0.1)$.

θ_U	$\bar{\hat{\theta}}_U$	$\bar{\hat{\Sigma}}_U$				Σ_U				95% coverage
4.3279	4.3417	0.1716	-0.3521	-0.4239	-0.4457	0.1789	-0.4284	-0.4507	-0.3298	0.94
0.9163	0.9335		0.1265	0.4986	0.6811		0.1343	0.5026	0.5553	0.95
0.4055	0.3234			0.4232	0.7806			0.4515	0.5551	0.96
0.4055	0.2731				1.1875				0.9716	0.95
θ	$\bar{\hat{\theta}}$	$\bar{\hat{\Sigma}}$				Σ				95% coverage
0.00005	0.00007	0.00007	-0.9363	-0.3778	0.4933	0.00008	-0.7074	-0.2754	0.4156	0.81
1.5	1.5668		0.3264	0.4986	-0.5487		0.3521	0.5133	-0.5512	0.95
1.5	1.5207			0.6187	-0.1572			0.6613	-0.1640	0.90
0.5	0.5018				0.0947				0.1064	0.92

(b) $(b,\ k,\ \omega^2,\ \psi) = (0.00005,\ 1.5,\ 1.5,\ 0.5)$.

Table 5.3: Summary statistics from the dynamic proportional hazards model with $\omega^2 = 1.5$.

θ_U	$\bar{\hat{\theta}}_U$	$\bar{\hat{\Sigma}}_U$				Σ_U				95% coverage
4.3279	4.3379	0.1305	-0.1778	-0.1906	-0.1692	0.1395	-0.2153	-0.2160	-0.1734	0.93
0.9163	0.9344		0.0948	0.3780	0.3751		0.0988	0.3966	0.3030	0.93
0.4055	0.3192			0.3602	0.2359			0.3836	0.1547	0.94
-1.7918	-4.2859				33.1671				3.9415	0.93
θ	$\bar{\hat{\theta}}$	$\bar{\hat{\Sigma}}$				Σ				95% coverage
0.00005	0.00006	0.00005	-0.9370	-0.3132	0.2850	0.00005	-0.7787	-0.2846	0.2796	0.82
1.5	1.5582		0.2433	0.3780	-0.3253		0.2574	0.4035	-0.4143	0.94
1.5	1.4757			0.5202	-0.0406			0.5465	-0.0847	0.90
0.9	0.8939				0.1023				0.1121	0.62

(c) $(b, k, \omega^2, \psi) = (0.00005, 1.5, 1.5, 0.9)$.

Chapter 6

Conclusion and Further Research

6.1 Summary of Results

In this dissertation, we have used, in Chapter 3, a state space model to deal with multivariate longitudinal measurements with missing values and measurement errors. The linear Kalman filter demonstrated its efficiency, especially when we have a number of long series of multivariate measurements. Next, we identified and discussed, in Chapter 4, two classes of Gaussian random effects models for recurrent event data; namely autocorrelated and dynamic random effects models. Dynamic random effects models are more appropriate when the initial inter-subject heterogeneity does not persist over time, otherwise autocorrelated random effects models are preferred. In Chapter 5, we extended the Cox proportional hazards models for recurrent event data to allow inter-subject heterogeneity and non-stationary process drifts by using a dynamic Gamma frailty process. The resulting model is somewhat similar to the dynamic random effects models discussed in Chapter 4 but is distinct in the fact that each dynamic frailty effect also accounts for the past recurrence times (a slight modification of GSSMs). The model enjoys tractability of the likelihood function by which scores and Hessian matrix can be easily numerically evaluated, a property not shared by most general state space models with random effects.

6.2 Further Research

We would like, in the future, to embark on the use of filtering methods for longitudinal data with different characteristics, e.g. missing responses and covariates, measurement errors in responses and covariates, measurements taken at irregular time epochs, inter-subject heterogeneity, and more. Some potential topics are described in the following sections.

6.2.1 Missing Data in Conditional Models

In Chapter 3, we dealt with multivariate and continuous measurements at specific time epochs with values missing at random. Lipsitz *et al.* (1994) considered a marginal approach for categorical responses with time-dependent covariates. They estimated covariate effects when responses are allowed to be missing at random. With illustration on binary responses, they stratified subjects according to their covariate values. A two-stage estimation procedure was adopted with the first stage used to estimate the marginal probabilities of a subject's responses and the second stage to estimate the covariate effects by regressing a known function of the marginal probabilities on the covariates. The first stage was carried out by maximizing the likelihood using EM or a Newton-Raphson method and the second stage proceeded by using ordinary weighted least squares. However, estimation especially in the first stage is cumbersome when the number of responses of a subject is large. Moreover, the method of stratification may not be appropriate when the resulting stratum size is small.

We are interested in the possible use of similar model and filtering methods as in Chapter 3 to improve the efficiency of estimation while accommodating time-dependent covariates and responses which can be missing at random and/or measured with errors.

6.2.2 Measurement Errors in Longitudinal Studies

Measurement errors can occur in both responses and covariates. They may also produce identifiability problems, e.g. whether the variability is due to measurement errors or inherent variations (Chapter 3). Ignoring measurement errors can lead to inconsistent estimates. However, most previous studies focused on examining the effects of measurement errors on survival (Tsiatis *et al.*, 1995; Raboud *et al.*, 1993) or ordinary GLM (Haukka, 1995; Sepanski *et al.*, 1994) type data. Methods that account for errors in covariates are mainly through imputation by assuming a certain measurement error model, or by the bootstrap. It is worth studying the effects of measurement errors (in both responses and covariates) in repeated measurements and exploring the applicability of filtering methods.

6.2.3 Combining Missing Values and Measurement Errors

We have discussed, in Chapter 3, missing values and measurement errors under Gaussian linear models only. It is also worth extending this to non-Gaussian models such as the exponential family models. There have been separate studies on measurement errors and missing values. For example, Sutradhar and Rao (1996) studied the correction of bias in regression parameters' estimates from solving GEEs under GLMs as a result of measurement errors on covariates. For partially missing covariates in GLMs, Ibrahim

(1990) considered the use of an EM algorithm through the "method of weights". More general discussions on missing data in longitudinal studies can be found in Laird (1988). However, it is in general hard to combine measurement errors and missing values on responses and covariates, and further research is highly desirable.

6.2.4 Irregularly Spaced Measurements

Irregularly spaced time data arises when subjects are measured at arbitrary time intervals. Sometimes, they can be treated as equally spaced time data with missing values but this may not be plausible when there is no basic sampling interval. Thus, it is more natural to consider an underlying continuous time process which govern the observed responses. The use of linear Gaussian state space models and filtering methods for irregularly spaced data are well described and discussed in Jones (1993). Elliott *et al.* (1995) considered, in a more general framework, the use of optimal filtering for estimation under both discrete and continuous time Hidden Markov Models (HMMs). The HMMs can be treated as another type of GSSMs. For example, a continuous time AR(1) process $\{X_t : t \in [0, \infty)\}$ with measurement errors can be formulated as a continuous time HMM by

$$\begin{aligned} Y_t &= X_t + W_t \\ \text{and} \quad X_t &= X_0 + \int_0^t \alpha_u X_u \, du + V_t \end{aligned}$$

where Y_t is the observable process while V_t and W_t are independent zero mean martingale processes. The key technique used by Elliott *et al.* (1995) is a change of measure through the Girsanov Theorem to work on a "fictitious world" where well-developed and straightforward tools can be employed. Results are then transformed back to the "real world" by a reverse change of measure. The mathematics is neat and complete but its use in the actual fitting of irregularly spaced measurements remains to be investigated.

Appendix A

Datasets

A.1 The Two Automobile Datasets

Both the Piston Machining and the Door Hanging Data are described in Section 1.2.1 of Chapter 1 and analyzed by a multivariate AR(1) variation transmission model in Section 3.4 of Chapter 3. They are printed in the following two subsections.

A.1.1 Piston Machining Data

The table shown below gives the four diameter measurements located at heights of 4 mm, 10 mm, 36.7 mm and 58.7 mm (the four values from top to bottom of each cell of the table) from the bottom of 96 pistons at four process stages.

Piston	Stage			
number	1	2	3	4
1	88.960	88.959	88.957	88.959
	88.976	88.975	88.973	88.975
	88.936	88.935	88.935	88.936
	88.167	88.163	88.161	88.163
2	88.955	88.955	88.955	88.958
	88.972	88.972	88.972	88.974
	88.935	88.934	88.934	88.937
	88.163	88.159	88.157	88.160
3	88.958	88.959	88.957	88.960
	88.974	88.975	88.973	88.975
	88.936	88.938	88.935	88.938
	88.163	88.164	88.160	88.163

4	88.956	88.959	88.958	88.959
	88.973	88.976	88.973	88.974
	88.935	88.937	88.935	88.937
	88.162	88.160	88.159	88.160
5	88.958	88.958	88.958	88.956
	88.972	88.972	88.972	88.970
	88.935	88.935	88.935	88.933
	88.163	88.162	88.161	88.159
6	88.954	88.954	88.952	88.952
	88.972	88.973	88.971	88.970
	88.935	88.934	88.933	88.933
	88.164	88.162	88.162	88.159
7	88.958	88.960	88.948	88.957
	88.973	88.973	88.973	88.972
	88.935	88.933	88.931	88.931
	88.163	88.161	88.160	88.160
8	88.954	88.955	88.956	88.953
	88.973	88.972	88.973	88.971
	88.934	88.934	88.935	88.933
	88.163	88.163	88.163	88.161
9	88.957	88.958	88.957	88.959
	88.974	88.977	88.975	88.976
	88.936	88.938	88.936	88.938
	88.165	88.163	88.162	88.162
10	88.957	88.959	88.957	88.959
	88.974	88.975	88.973	88.975
	88.937	88.938	88.936	88.938
	88.161	88.159	88.157	88.158
11	88.960	88.960	88.957	88.956
	88.977	88.976	88.973	88.973
	88.938	88.937	88.936	88.936
	88.168	88.162	88.160	88.162
12	88.957	88.958	88.955	88.956
	88.973	88.974	88.972	88.972
	88.935	88.936	88.935	88.935
	88.162	88.160	88.162	88.157
13	88.960	88.958	88.959	88.957
	88.974	88.973	88.973	88.971
	88.935	88.933	88.936	88.935
	88.164	88.161	88.161	88.157
14	88.955	88.955	88.955	88.953
	88.973	88.973	88.972	88.972
	88.934	88.933	88.933	88.933
	88.163	88.163	88.162	88.162
15	88.959	88.958	88.957	88.956
	88.973	88.972	88.973	88.972
	88.935	88.934	88.935	88.933
	88.161	88.159	88.160	88.159
16	88.955	88.955	88.956	88.953
	88.972	88.972	88.973	88.972
	88.933	88.933	88.934	88.932
	88.162	88.161	88.161	88.160
17	88.960	88.961	88.959	88.959
	88.975	88.977	88.976	88.974
	88.938	88.939	88.938	88.937
	88.165	88.164	88.162	88.161
18	88.957	88.961	88.957	88.960
	88.974	88.976	88.973	88.975
	88.936	88.939	88.932	88.938
	88.161	88.161	88.156	88.158
19	88.958	88.957	88.957	88.957
	88.976	88.975	88.974	88.974
	88.938	88.937	88.936	88.938
	88.167	88.166	88.161	88.162
20	88.958	88.958	88.956	88.957
	88.973	88.974	88.973	88.973
	88.935	88.933	88.933	88.934
	88.161	88.160	88.157	88.159
21	88.959	88.959	88.960	88.957
	88.972	88.971	88.972	88.970
	88.935	88.934	88.935	88.933
	88.163	88.161	88.162	88.160
22	88.951	88.954	88.953	88.950
	88.971	88.976	88.972	88.970
	88.934	88.935	88.934	88.934
	88.163	88.163	88.159	88.160
23	88.960	88.960	88.961	88.958
	88.974	88.974	88.974	88.972
	88.935	88.935	88.932	88.932
	88.162	88.161	88.159	88.159
24	88.954	88.954	88.955	88.952
	88.972	88.973	88.972	88.971
	88.936	88.935	88.935	88.934
	88.163	88.163	88.162	88.162
25	88.957	88.957	88.958	88.959
	88.974	88.975	88.975	88.975
	88.937	88.937	88.937	88.937
	88.165	88.162	88.163	88.162
26	88.957	88.959	88.956	88.957
	88.972	88.974	88.973	88.973
	88.936	88.936	88.935	88.937
	88.163	88.162	88.157	88.160
27	88.957	88.958	88.957	88.958
	88.975	88.975	88.973	88.974
	88.938	88.938	88.938	88.937
	88.166	88.164	88.160	88.161

28	88.958	88.958	88.956	88.957
	88.975	88.975	88.974	88.974
	88.937	88.937	88.936	88.937
	88.163	88.161	88.158	88.160
29	88.960	88.959	88.959	88.958
	88.974	88.973	88.973	88.972
	88.936	88.935	88.935	88.933
	88.164	88.162	88.161	88.160
30	88.956	88.956	88.957	88.955
	88.973	88.974	88.974	88.973
	88.935	88.933	88.934	88.934
	88.164	88.160	88.162	88.160
31	88.958	88.959	88.958	88.958
	88.973	88.973	88.973	88.972
	88.937	88.937	88.936	88.936
	88.164	88.163	88.162	88.160
32	88.956	88.955	88.955	88.954
	88.973	88.973	88.972	88.971
	88.936	88.935	88.935	88.935
	88.164	88.163	88.161	88.160
33	88.957	88.960	88.958	88.959
	88.973	88.975	88.974	88.973
	88.935	88.937	88.935	88.935
	88.165	88.165	88.162	88.163
34	88.956	88.957	88.955	88.955
	88.972	88.973	88.971	88.971
	88.935	88.931	88.930	88.933
	88.161	88.160	88.158	88.158
35	88.958	88.958	88.956	88.957
	88.975	88.974	88.973	88.973
	88.938	88.937	88.934	88.936
	88.166	88.164	88.161	88.162
36	88.956	88.958	88.956	88.956
	88.972	88.974	88.972	88.973
	88.936	88.937	88.935	88.935
	88.162	88.160	88.159	88.158
37	88.959	88.958	88.959	88.959
	88.973	88.972	88.973	88.973
	88.934	88.933	88.933	88.933
	88.161	88.159	88.160	88.160
38	88.951	88.950	88.950	88.949
	88.969	88.969	88.973	88.969
	88.933	88.933	88.933	88.932
	88.164	88.162	88.163	88.163
39	88.959	88.959	88.959	88.958
	88.973	88.972	88.971	88.970
	88.935	88.934	88.933	88.932
	88.163	88.161	88.161	88.161
40	88.954	88.953	88.952	88.952
	88.971	88.970	88.970	88.970
	88.934	88.933	88.932	88.932
	88.162	88.162	88.160	88.161
41	88.958	88.956	88.955	88.957
	88.976	88.975	88.974	88.975
	88.931	88.931	88.929	88.929
	88.157	88.156	88.154	88.153
42	88.955	88.957	88.955	88.956
	88.972	88.973	88.972	88.973
	88.933	88.935	88.934	88.935
	88.160	88.158	88.158	88.158
43	88.953	88.955	88.953	88.953
	88.971	88.972	88.971	88.971
	88.935	88.936	88.936	88.937
	88.165	88.164	88.162	88.164
44	88.956	88.958	88.956	88.957
	88.973	88.975	88.973	88.973
	88.936	88.938	88.936	88.937
	88.162	88.161	88.155	88.159
45	88.961	88.959	88.960	88.960
	88.975	88.973	88.973	88.972
	88.932	88.932	88.931	88.932
	88.160	88.160	88.159	88.159
46	88.954	88.960	88.953	88.953
	88.973	88.973	88.971	88.972
	88.935	88.934	88.933	88.935
	88.163	88.161	88.161	88.162
47	88.959	88.962	88.958	88.959
	88.973	88.974	88.972	88.973
	88.934	88.930	88.932	88.933
	88.164	88.158	88.161	88.161
48	88.954	88.954	88.953	88.951
	88.972	88.972	88.970	88.970
	88.934	88.934	88.934	88.933
	88.164	88.161	88.162	88.162
49	88.958	88.960	88.957	88.959
	88.975	88.977	88.973	88.974
	88.936	88.938	88.934	88.937
	88.166	88.165	88.160	88.163
50	88.956	88.958	88.956	88.957
	88.974	88.975	88.973	88.974
	88.935	88.937	88.936	88.936
	88.161	88.159	88.156	88.158
51	88.956	88.957	88.955	88.955
	88.974	88.976	88.972	88.974
	88.935	88.938	88.935	88.937
	88.166	88.164	88.161	88.163

52	88.958	88.959	88.956	88.956
	88.973	88.975	88.973	88.973
	88.936	88.937	88.934	88.936
	88.163	88.160	88.156	88.160
53	88.958	88.954	88.957	88.958
	88.972	88.971	88.971	88.972
	88.934	88.933	88.930	88.934
	88.165	88.163	88.161	88.162
54	88.954	88.955	88.955	88.954
	88.972	88.973	88.973	88.973
	88.935	88.935	88.935	88.936
	88.165	88.162	88.162	88.163
55	88.959	88.960	88.961	88.959
	88.973	88.973	88.974	88.973
	88.934	88.934	88.933	88.932
	88.164	88.162	88.162	88.160
56	88.956	88.955	88.956	88.954
	88.973	88.970	88.972	88.971
	88.934	88.933	88.934	88.933
	88.166	88.163	88.163	88.162
57	88.958	88.959	88.956	88.957
	88.976	88.976	88.973	88.975
	88.938	88.938	88.935	88.937
	88.166	88.164	88.160	88.162
58	88.957	88.959	88.958	88.958
	88.973	88.975	88.974	88.974
	88.935	88.923	88.935	88.929
	88.161	88.158	88.156	88.156
59	88.958	88.959	88.956	88.958
	88.975	88.974	88.972	88.974
	88.937	88.937	88.933	88.936
	88.166	88.162	88.158	88.161
60	88.958	88.958	88.956	88.958
	88.974	88.976	88.972	88.974
	88.937	88.938	88.935	88.937
	88.162	88.160	88.157	88.158
61	88.959	88.959	88.959	88.957
	88.974	88.974	88.975	88.973
	88.933	88.931	88.931	88.929
	88.164	88.162	88.162	88.160
62	88.954	88.955	88.955	88.954
	88.971	88.973	88.972	88.970
	88.935	88.934	88.934	88.934
	88.164	88.162	88.162	88.161
63	88.960	88.960	88.958	88.958
	88.973	88.973	88.972	88.972
	88.935	88.935	88.933	88.934
	88.163	88.163	88.162	88.162
64	88.955	88.954	88.954	88.953
	88.973	88.972	88.972	88.972
	88.935	88.934	88.934	88.932
	88.165	88.163	88.163	88.164
65	88.956	88.958	88.956	88.957
	88.974	88.976	88.974	88.975
	88.934	88.935	88.935	88.935
	88.166	88.162	88.162	88.162
66	88.956	88.957	88.954	88.957
	88.972	88.973	88.970	88.973
	88.934	88.935	88.932	88.937
	88.160	88.159	88.156	88.158
67	88.956	88.959	88.956	88.958
	88.973	88.976	88.974	88.975
	88.935	88.937	88.935	88.936
	88.164	88.163	88.161	88.162
68	88.956	88.957	88.957	88.957
	88.973	88.973	88.973	88.973
	88.935	88.936	88.933	88.935
	88.161	88.159	88.157	88.158
69	88.960	88.960	88.958	88.959
	88.973	88.973	88.973	88.972
	88.934	88.933	88.933	88.934
	88.162	88.161	88.160	88.159
70	88.953	88.954	88.953	88.952
	88.972	88.972	88.971	88.971
	88.934	88.932	88.932	88.932
	88.164	88.162	88.162	88.162
71	88.960	88.961	88.959	88.960
	88.973	88.973	88.971	88.973
	88.931	88.932	88.927	88.928
	88.162	88.161	88.161	88.161
72	88.954	88.956	88.955	88.956
	88.971	88.972	88.972	88.972
	88.933	88.934	88.933	88.934
	88.163	88.163	88.162	88.164
73	88.959	88.959	88.957	88.958
	88.976	88.976	88.973	88.975
	88.937	88.938	88.935	88.936
	88.168	88.165	88.162	88.164
74	88.956	88.958	88.957	88.957
	88.972	88.974	88.972	88.973
	88.934	88.935	88.933	88.934
	88.162	88.159	88.157	88.158
75	88.959	88.959	88.957	88.958
	88.976	88.976	88.974	88.974
	88.938	88.936	88.935	88.937
	88.166	88.163	88.161	88.163

76	88.955	88.957	88.956	88.957
	88.972	88.973	88.973	88.974
	88.936	88.937	88.935	88.936
	88.161	88.160	88.158	88.158
77	88.958	88.956	88.956	88.957
	88.971	88.970	88.970	88.975
	88.934	88.933	88.933	88.935
	88.162	88.160	88.161	88.161
78	88.954	88.954	88.955	88.954
	88.971	88.971	88.972	88.971
	88.933	88.933	88.934	88.934
	88.164	88.163	88.162	88.163
79	88.959	88.960	88.959	88.958
	88.972	88.973	88.972	88.972
	88.931	88.931	88.931	88.932
	88.162	88.162	88.160	88.162
80	88.955	88.954	88.954	88.954
	88.973	88.972	88.972	88.973
	88.933	88.931	88.933	88.932
	88.164	88.164	88.163	88.164
81	88.955	88.957	88.954	88.956
	88.973	88.975	88.971	88.974
	88.934	88.936	88.933	88.937
	88.163	88.164	88.160	88.163
82	88.955	88.958	88.956	88.956
	88.972	88.974	88.972	88.972
	88.934	88.936	88.933	88.935
	88.161	88.160	88.157	88.159
83	88.958	88.958	88.956	88.956
	88.976	88.975	88.973	88.973
	88.938	88.938	88.937	88.938
	88.167	88.164	88.163	88.163
84	88.956	88.957	88.955	88.956
	88.972	88.973	88.971	88.973
	88.936	88.934	88.935	88.937
	88.161	88.160	88.157	88.160
85	88.960	88.959	88.958	88.959
	88.974	88.971	88.971	88.973
	88.934	88.933	88.932	88.934
	88.164	88.162	88.161	88.163
86	88.954	88.954	88.952	88.953
	88.972	88.971	88.970	88.971
	88.935	88.934	88.933	88.936
	88.164	88.162	88.162	88.164
87	88.957	88.956	88.956	88.955
	88.972	88.970	88.970	88.971
	88.934	88.934	88.932	88.936
	88.163	88.161	88.159	88.164
88	88.955	88.954	88.955	88.954
	88.972	88.972	88.972	88.972
	88.934	88.933	88.934	88.934
	88.163	88.161	88.160	88.161
89	88.956	88.958	88.957	88.956
	88.972	88.975	88.973	88.973
	88.933	88.935	88.935	88.933
	88.163	88.163	88.161	88.162
90	88.956	88.957	88.955	88.957
	88.973	88.973	88.972	88.973
	88.935	88.935	88.935	88.934
	88.162	88.160	88.159	88.160
91	88.957	88.957	88.956	88.956
	88.974	88.974	88.971	88.973
	88.937	88.937	88.934	88.937
	88.164	88.163	88.159	88.162
92	88.956	88.958	88.955	88.957
	88.972	88.975	88.971	88.973
	88.935	88.936	88.929	88.933
	88.162	88.161	88.158	88.160
93	88.960	88.959	88.958	88.959
	88.973	88.972	88.971	88.972
	88.935	88.934	88.933	88.934
	88.162	88.160	88.158	88.161
94	88.953	88.951	88.952	88.952
	88.971	88.970	88.971	88.971
	88.935	88.930	88.934	88.935
	88.165	88.162	88.164	88.164
95	88.960	88.959	88.958	88.962
	88.972	88.971	88.970	88.972
	88.935	88.934	88.932	88.933
	88.163	88.161	88.160	88.161
96	88.955	88.955	88.955	88.956
	88.973	88.973	88.972	88.972
	88.934	88.933	88.934	88.935
	88.164	88.163	88.163	88.163

==

A.1.2 Door Hanging Data

The table shown below contains the four measurements taken from 42 vehicles at seven process stages. Values marked "NA" refer to missing data.

Stage Number	Car ID	Characteristic 1	2	3	4
1 Body	NA	NA	NA	NA	NA
1 Body	36672	NA	NA	NA	NA
1 Body	37465	NA	NA	NA	NA
1 Body	38788	0.4300	0.1200	8.8600	12.4900
1 Body	38792	0.7200	0.1300	8.5200	12.0600
1 Body	41178	0.6300	0.2600	8.6700	12.4600
1 Body	46764	NA	NA	NA	NA
1 Body	46813	NA	NA	NA	NA
1 Body	46982	NA	NA	NA	NA
1 Body	47051	NA	NA	NA	NA
1 Body	47157	NA	NA	NA	NA
1 Body	47170	NA	NA	NA	NA
1 Body	47176	NA	NA	NA	NA
1 Body	47208	0.5000	0.1000	9.3600	13.3100
1 Body	47226	1.1500	0.1100	7.5900	12.1500
1 Body	47228	-1.0800	0.5100	7.7600	11.2000
1 Body	47233	0.7000	0.2900	8.3000	12.0000
1 Body	47236	0.7900	0.1900	6.8100	10.9800
1 Body	47243	-0.3500	0.4500	9.6100	12.3600
1 Body	47247	0.8300	0.0100	9.2900	12.0700
1 Body	47256	1.1900	-0.2900	8.9700	14.0900
1 Body	47269	1.4300	0.0700	8.5200	11.7500
1 Body	47274	2.2700	-0.0800	10.0600	13.6400
1 Body	47280	-1.3700	0.3800	7.5600	11.5600
1 Body	47284	0.8400	0.0150	9.6650	12.3700
1 Body	47299	-1.7750	0.3700	8.1500	9.5200
1 Body	47318	0.7700	-0.0100	8.0600	12.0900
1 Body	47322	0.0400	0.0300	8.7000	12.0200
1 Body	47325	0.3800	0.0000	9.3800	12.1700
1 Body	47328	-1.8200	0.3200	7.9100	10.8200
1 Body	47332	0.3400	0.0500	8.9300	11.8400
1 Body	47335	NA	NA	NA	NA
1 Body	47353	NA	0.6100	7.0600	9.3800
1 Body	47355	0.6300	0.1600	8.6400	13.0400
1 Body	47356	-1.6300	0.7600	7.0100	9.4600
1 Body	47358	1.5600	-0.0100	8.5500	11.4400
1 Body	47365	1.5600	-0.0800	10.2050	11.4500
1 Body	47369	1.1900	-0.1000	10.2100	13.1200
1 Body	47372	-1.5300	0.5200	8.7200	10.7200
1 Body	47395	1.0600	0.1250	9.3500	14.2750
1 Body	47401	NA	NA	NA	NA
1 Body	47481	-2.0350	0.6850	7.4450	9.5700
2 Paint	NA	NA	NA	NA	NA
2 Paint	36672	-0.4900	0.2750	7.8450	10.4950
2 Paint	37465	1.4800	0.0450	9.0950	11.9750
2 Paint	38788	NA	NA	NA	NA
2 Paint	38792	-0.7050	0.1750	7.4250	10.0300
2 Paint	41178	-0.5650	0.2200	7.4150	10.3700
2 Paint	46764	0.8600	0.1800	9.4100	11.4200

2	Paint	46813	0.5450	0.6100	8.5650	11.2500
2	Paint	46982	-1.0700	0.6200	7.3700	9.3200
2	Paint	47051	0.0275	0.1150	8.1700	11.0600
2	Paint	47157	-0.0150	0.2850	8.3400	10.4050
2	Paint	47170	0.5050	0.0950	7.9000	10.4800
2	Paint	47176	-0.4700	0.3600	7.2200	9.8450
2	Paint	47208	-0.1325	-0.1350	8.7900	11.5825
2	Paint	47226	NA	NA	NA	NA
2	Paint	47228	-1.1850	0.5300	6.9850	9.5950
2	Paint	47233	-0.3875	0.5100	7.7175	10.5075
2	Paint	47236	-0.2650	0.2350	6.0250	8.9150
2	Paint	47243	-1.3600	0.2750	8.7200	10.9100
2	Paint	47247	-0.1550	0.5900	8.4250	11.0600
2	Paint	47256	NA	NA	NA	NA
2	Paint	47269	-0.1400	-0.0900	7.6100	9.4500
2	Paint	47274	1.2500	0.0450	9.1000	11.9050
2	Paint	47280	NA	0.3100	6.5050	9.5800
2	Paint	47284	0.2300	-0.0050	9.0400	10.0850
2	Paint	47299	NA	NA	NA	NA
2	Paint	47318	-0.3650	-0.0800	7.2950	10.3600
2	Paint	47322	-0.5900	0.0400	8.0850	10.0550
2	Paint	47325	-1.3400	0.1600	8.3350	9.6700
2	Paint	47328	-1.9300	0.5150	7.1300	9.3700
2	Paint	47332	-0.1700	0.1400	8.0400	10.0200
2	Paint	47335	-0.0600	0.1800	7.4750	10.5100
2	Paint	47353	NA	NA	NA	NA
2	Paint	47355	NA	NA	NA	NA
2	Paint	47356	NA	NA	NA	NA
2	Paint	47358	0.5950	0.6000	7.7700	10.1050
2	Paint	47365	0.6550	-0.0250	9.3900	9.3050
2	Paint	47369	NA	NA	NA	NA
2	Paint	47372	NA	0.6200	7.4550	8.6600
2	Paint	47395	0.2000	0.1200	8.3850	11.9300
2	Paint	47401	NA	0.2467	6.3700	7.5867
2	Paint	47481	NA	NA	NA	NA
3	Before_Striker	NA	NA	NA	NA	NA
3	Before_Striker	36672	-1.3500	0.4400	7.8900	10.2300
3	Before_Striker	37465	NA	NA	NA	NA
3	Before_Striker	38788	NA	NA	NA	NA
3	Before_Striker	38792	-1.1500	0.4500	7.7000	9.9000
3	Before_Striker	41178	-1.5500	0.3500	7.8000	11.2000
3	Before_Striker	46764	-0.0900	0.2400	9.7000	12.6800
3	Before_Striker	46813	NA	NA	NA	NA
3	Before_Striker	46982	NA	NA	NA	NA
3	Before_Striker	47051	-0.7200	0.3900	7.9500	10.9000
3	Before_Striker	47157	-0.7000	0.3600	8.5000	12.1800
3	Before_Striker	47170	NA	NA	NA	NA
3	Before_Striker	47176	NA	NA	NA	NA
3	Before_Striker	47208	-0.8000	0.0700	9.0000	12.6000
3	Before_Striker	47226	NA	NA	NA	NA
3	Before_Striker	47228	NA	NA	NA	NA
3	Before_Striker	47233	-0.9500	0.4500	7.8000	10.7000
3	Before_Striker	47236	-1.2600	0.7500	6.4000	9.3000
3	Before_Striker	47243	-2.1500	0.4500	8.7000	10.3000
3	Before_Striker	47247	NA	NA	NA	NA
3	Before_Striker	47256	NA	NA	NA	NA
3	Before_Striker	47269	-1.2500	0.5500	6.9000	9.8000
3	Before_Striker	47274	0.1500	0.1500	9.0000	11.0000
3	Before_Striker	47280	-3.4000	0.8500	6.7000	9.4000
3	Before_Striker	47284	NA	NA	NA	NA

3	Before_Striker	47299	NA	NA	NA	NA
3	Before_Striker	47318	-1.1500	0.1500	7.5000	11.0000
3	Before_Striker	47322	-1.3500	0.4500	8.1000	10.7000
3	Before_Striker	47325	-1.7500	0.6500	9.7000	10.8000
3	Before_Striker	47328	-3.6500	0.7500	7.8000	9.5000
3	Before_Striker	47332	NA	NA	NA	NA
3	Before_Striker	47335	NA	NA	NA	NA
3	Before_Striker	47353	NA	NA	NA	NA
3	Before_Striker	47355	NA	NA	NA	NA
3	Before_Striker	47356	NA	NA	NA	NA
3	Before_Striker	47358	-0.2500	0.4500	8.8000	10.5000
3	Before_Striker	47365	NA	NA	NA	NA
3	Before_Striker	47369	NA	NA	NA	NA
3	Before_Striker	47372	-4.1500	0.8500	8.0000	10.6000
3	Before_Striker	47395	-0.7500	0.4500	8.4000	10.8000
3	Before_Striker	47401	NA	NA	NA	NA
3	Before_Striker	47481	NA	NA	NA	NA
4	After_Striker	NA	NA	NA	NA	NA
4	After_Striker	36672	-0.5700	1.0100	7.9000	10.7000
4	After_Striker	37465	NA	NA	NA	NA
4	After_Striker	38788	NA	NA	NA	NA
4	After_Striker	38792	-0.5500	1.1500	7.9000	10.4000
4	After_Striker	41178	-1.4500	0.6500	7.8000	10.8000
4	After_Striker	46764	-0.3700	0.5400	9.4900	12.5800
4	After_Striker	46813	NA	NA	NA	NA
4	After_Striker	46982	NA	NA	NA	NA
4	After_Striker	47051	-2.6500	-0.0100	7.5800	9.8300
4	After_Striker	47157	-0.4500	1.4200	8.5900	12.3800
4	After_Striker	47170	NA	NA	NA	NA
4	After_Striker	47176	NA	NA	NA	NA
4	After_Striker	47208	-1.2500	0.5500	8.9000	12.4000
4	After_Striker	47226	NA	NA	NA	NA
4	After_Striker	47228	NA	NA	NA	NA
4	After_Striker	47233	-1.8500	0.5500	7.5000	10.1000
4	After_Striker	47236	-0.3500	1.2500	6.5000	9.8000
4	After_Striker	47243	-0.9500	0.9500	9.1000	10.9000
4	After_Striker	47247	NA	NA	NA	NA
4	After_Striker	47256	NA	NA	NA	NA
4	After_Striker	47269	-1.5500	0.9500	7.2000	9.3000
4	After_Striker	47274	-1.6500	0.5500	8.3000	10.0000
4	After_Striker	47280	-0.4500	1.1500	7.4000	10.6000
4	After_Striker	47284	NA	NA	NA	NA
4	After_Striker	47299	NA	NA	NA	NA
4	After_Striker	47318	-0.2500	1.0500	7.9000	11.8000
4	After_Striker	47322	-1.6500	-0.0500	8.1500	10.5500
4	After_Striker	47325	-1.7500	-0.3500	9.9000	10.7000
4	After_Striker	47328	-0.9500	0.8500	8.6000	10.6000
4	After_Striker	47332	NA	NA	NA	NA
4	After_Striker	47335	NA	NA	NA	NA
4	After_Striker	47353	NA	NA	NA	NA
4	After_Striker	47355	NA	NA	NA	NA
4	After_Striker	47356	NA	NA	NA	NA
4	After_Striker	47358	-1.6500	0.3500	8.7000	9.8000
4	After_Striker	47365	NA	NA	NA	NA
4	After_Striker	47369	NA	NA	NA	NA
4	After_Striker	47372	-1.2500	0.1500	8.1000	11.9000
4	After_Striker	47395	-0.8000	1.1500	8.3500	11.2000
4	After_Striker	47401	NA	NA	NA	NA
4	After_Striker	47481	NA	NA	NA	NA
5	Striker_Fit	NA	NA	NA	NA	NA

5 Striker_Fit	36672	-1.1400	0.7100	7.0800	8.8900
5 Striker_Fit	37465	NA	NA	NA	NA
5 Striker_Fit	38788	NA	NA	NA	NA
5 Striker_Fit	38792	-0.8500	0.8500	7.0000	9.2000
5 Striker_Fit	41178	-1.4500	0.6500	7.8000	10.8000
5 Striker_Fit	46764	-0.3700	0.5400	9.4900	12.5800
5 Striker_Fit	46813	NA	NA	NA	NA
5 Striker_Fit	46982	NA	NA	NA	NA
5 Striker_Fit	47051	-2.6500	-0.3500	7.3600	9.6800
5 Striker_Fit	47157	-0.4500	1.4200	8.5900	12.3800
5 Striker_Fit	47170	NA	NA	NA	NA
5 Striker_Fit	47176	NA	NA	NA	NA
5 Striker_Fit	47208	-1.4500	0.3500	8.3000	11.7000
5 Striker_Fit	47226	NA	NA	NA	NA
5 Striker_Fit	47228	NA	NA	NA	NA
5 Striker_Fit	47233	-1.9500	0.4500	6.6000	9.1000
5 Striker_Fit	47236	-0.6500	1.0500	5.6000	8.6000
5 Striker_Fit	47243	-1.5500	0.5500	7.1000	8.6000
5 Striker_Fit	47247	NA	NA	NA	NA
5 Striker_Fit	47256	NA	NA	NA	NA
5 Striker_Fit	47269	-1.8500	0.7500	6.5000	8.5000
5 Striker_Fit	47274	-1.8500	0.4500	7.8000	9.3000
5 Striker_Fit	47280	-0.6500	1.0500	6.2000	9.8000
5 Striker_Fit	47284	NA	NA	NA	NA
5 Striker_Fit	47299	NA	NA	NA	NA
5 Striker_Fit	47318	-0.6500	0.8500	5.6000	9.8000
5 Striker_Fit	47322	-1.8500	-0.1500	7.0000	9.5000
5 Striker_Fit	47325	-2.0700	-0.5500	8.9000	9.1000
5 Striker_Fit	47328	-1.2500	0.5500	7.6000	9.6000
5 Striker_Fit	47332	NA	NA	NA	NA
5 Striker_Fit	47335	NA	NA	NA	NA
5 Striker_Fit	47353	NA	NA	NA	NA
5 Striker_Fit	47355	NA	NA	NA	NA
5 Striker_Fit	47356	NA	NA	NA	NA
5 Striker_Fit	47358	-1.8500	0.1500	7.6000	8.6000
5 Striker_Fit	47365	NA	NA	NA	NA
5 Striker_Fit	47369	NA	NA	NA	NA
5 Striker_Fit	47372	-1.7500	0.2500	6.2000	9.1000
5 Striker_Fit	47395	-1.1500	0.9500	8.0000	10.4000
5 Striker_Fit	47401	NA	NA	NA	NA
5 Striker_Fit	47481	NA	NA	NA	NA
6 Final	NA	NA	NA	NA	NA
6 Final	36672	-0.3700	1.6200	8.7500	14.8800
6 Final	37465	NA	NA	NA	NA
6 Final	38788	NA	NA	NA	NA
6 Final	38792	-0.4050	1.6450	8.8250	14.2600
6 Final	41178	-1.1250	1.2950	8.2050	14.8750
6 Final	46764	-0.0700	1.1200	9.9900	15.1600
6 Final	46813	NA	NA	NA	NA
6 Final	46982	NA	NA	NA	NA
6 Final	47051	-2.1100	0.6600	8.7100	NA
6 Final	47157	0.1700	2.1500	9.1000	13.7000
6 Final	47170	NA	NA	NA	NA
6 Final	47176	NA	NA	NA	NA
6 Final	47208	-1.1800	1.1900	8.6800	14.5000
6 Final	47226	NA	NA	NA	NA
6 Final	47228	NA	NA	NA	NA
6 Final	47233	-1.2700	1.2700	8.7950	14.7200
6 Final	47236	-0.0950	1.9500	7.3150	13.5800
6 Final	47243	-1.0000	2.0300	9.1150	13.8850

6	Final	47247	NA	NA	NA	NA
6	Final	47256	NA	NA	NA	NA
6	Final	47269	-1.0950	1.7350	8.1400	14.0250
6	Final	47274	-1.3400	1.3100	9.5700	15.1800
6	Final	47280	-0.0650	1.8900	7.9500	14.6050
6	Final	47284	NA	NA	NA	NA
6	Final	47299	NA	NA	NA	NA
6	Final	47318	-0.0950	1.5300	8.0050	14.5550
6	Final	47322	-1.5300	1.4300	8.8950	14.8200
6	Final	47325	-1.4000	0.4150	10.5200	14.4050
6	Final	47328	-0.7700	1.5100	8.6600	14.6950
6	Final	47332	NA	NA	NA	NA
6	Final	47335	NA	NA	NA	NA
6	Final	47353	NA	NA	NA	NA
6	Final	47355	NA	NA	NA	NA
6	Final	47356	NA	NA	NA	NA
6	Final	47358	-1.6650	1.4750	9.2450	13.3250
6	Final	47365	NA	NA	NA	NA
6	Final	47369	NA	NA	NA	NA
6	Final	47372	-1.0800	1.1950	8.7150	13.3250
6	Final	47395	-0.2500	1.9400	9.3800	15.5000
6	Final	47401	NA	NA	NA	NA
6	Final	47481	NA	NA	NA	NA
7	Enhanced	NA	-0.3700	1.5700	9.7050	13.2100
7	Enhanced	36672	-0.7950	0.8650	8.4150	14.0250
7	Enhanced	37465	NA	NA	NA	NA
7	Enhanced	38788	NA	NA	NA	NA
7	Enhanced	38792	-0.6700	1.2950	7.6100	13.5250
7	Enhanced	41178	-1.1150	1.0150	8.2900	11.7850
7	Enhanced	46764	NA	NA	NA	NA
7	Enhanced	46813	NA	NA	NA	NA
7	Enhanced	46982	NA	NA	NA	NA
7	Enhanced	47051	-2.1200	0.6250	8.6600	14.3050
7	Enhanced	47157	-0.4550	0.8800	8.5650	14.1600
7	Enhanced	47170	NA	NA	NA	NA
7	Enhanced	47176	NA	NA	NA	NA
7	Enhanced	47208	-1.2250	1.1750	8.1500	13.3400
7	Enhanced	47226	NA	NA	NA	NA
7	Enhanced	47228	NA	NA	NA	NA
7	Enhanced	47233	-1.5400	1.3100	8.0700	13.8600
7	Enhanced	47236	-0.4250	1.4050	6.9800	10.9800
7	Enhanced	47243	-0.6600	1.7150	8.2100	13.0650
7	Enhanced	47247	NA	NA	NA	NA
7	Enhanced	47256	NA	NA	NA	NA
7	Enhanced	47269	-1.0700	1.5300	8.0250	13.1350
7	Enhanced	47274	NA	NA	NA	NA
7	Enhanced	47280	NA	NA	NA	NA
7	Enhanced	47284	NA	NA	NA	NA
7	Enhanced	47299	NA	NA	NA	NA
7	Enhanced	47318	NA	NA	NA	NA
7	Enhanced	47322	-0.8800	1.5300	8.8150	14.4800
7	Enhanced	47325	-1.5850	0.5450	6.5900	12.7500
7	Enhanced	47328	-0.8150	1.0600	8.1600	14.0200
7	Enhanced	47332	NA	NA	NA	NA
7	Enhanced	47335	NA	NA	NA	NA
7	Enhanced	47353	NA	NA	NA	NA
7	Enhanced	47355	NA	NA	NA	NA
7	Enhanced	47356	NA	NA	NA	NA
7	Enhanced	47358	-0.7150	1.5600	9.3600	13.4150
7	Enhanced	47365	NA	NA	NA	NA

7	Enhanced	47369	NA	NA	NA	NA
7	Enhanced	47372	-0.5600	1.6000	7.9450	12.8650
7	Enhanced	47395	-0.4350	1.3450	8.5900	14.4900
7	Enhanced	47401	NA	NA	NA	NA
7	Enhanced	47481	NA	NA	NA	NA

A.2 The Small Bowel Motility Data

The table shown below is reproduced from Aalen and Husebye (1991) which contains the observed migrating motor complex (MMC) periods (in minutes) for 19 subjects. The data are described in Section 1.2.2 of Chapter 1 and analyzed by Normal-based models in Section 4.3 of Chapter 4 and a Hazard-based model in Section 5.5 of Chapter 5.

Subject	Complete observed periods									Censored
1	112	145	39	52	21	34	33	51		54
2	206	147								30
3	284	59	186							4
4	94	98	84							87
5	67									131
6	124	34	87	75	43	38	58	142	75	23
7	116	71	83	68	125					111
8	111	59	47	95						110
9	98	161	154	55						44
10	166	56								122
11	63	90	63	103	51					85
12	47	86	68	144						72
13	120	106	176							6
14	112	25	57	166						85
15	132	267	89							86
16	120	47	165	64	113					12
17	162	141	107	69						39
18	106	56	158	41	41	168				13
19	147	134	78	66	100					4

Appendix B

Derivation of Filtering and Smoothing Formulas

The filtering formulas (3.16)-(3.19) follow from straightforward conditional mean and variance calculations.

For example,

$$\begin{aligned} z_{it|t-1} &= E\left\{E[z_{it}|y_{is}, z_{is}, s=1,\dots,t-1]\right\} \\ &= E\left\{A_t + B_t z_{i,t-1}|y_{is}, s=1,\dots,t-1\right\} \\ &= A_t + B_t z_{i,t-1|t-1} \end{aligned}$$

$$\begin{aligned} \sum\nolimits_{iz}(t|t-1) &= E\left\{\sum\nolimits_{e_t}|y_{is}, s=1,\dots,t-1\right\} + \mathrm{Var}\left\{A_t + B_t z_{i,t-1}|y_{is}, s=1,\dots,t-1\right\} \\ &= \sum\nolimits_{e_t} + B_t \sum\nolimits_{iz}(t-1|t-1) B_t' \,. \end{aligned}$$

Formulas for $z_{it|t}$, $\sum_{iz}(t|t)$ and the smoothing formulas (3.23), (3.24) are a little bit more complicated, but may be obtained from standard results about multivariate normal variables. In particular, if x, y and z are random vectors with

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} \sim \text{Normal}\left[\begin{pmatrix} \mu_x \\ \mu_y \\ \mu_z \end{pmatrix}, \begin{pmatrix} \sum_{xx} & \sum_{xy} & \sum_{xz} \\ \sum_{yx} & \sum_{yy} & \sum_{yz} \\ \sum_{zx} & \sum_{zy} & \sum_{zz} \end{pmatrix}\right] .$$

Then

$$\begin{aligned} E(x|z) &= \mu_x + \textstyle\sum_{xz}\sum_z^{-1}(z-\mu_z) &&= \mu_{x|z} \\ \mathrm{Var}(x|z) &= \textstyle\sum_x - \sum_{xz}\sum_z^{-1}\sum_{zx} &&= \sum_{x|z} \\ \mathrm{Cov}(x,y|z) &= \textstyle\sum_{xy} - \sum_{xz}\sum_z^{-1}\sum_{zy} &&= \sum_{xy|z} \end{aligned}$$

and so also, for example,

$$E(x|y,z) = \mu_{x|z} + \sum_{xy|z}\sum_{y|z}^{-1}(y-\mu_{y|z}) .$$

Then, for example, letting $y_{i,t-1}^{*} = (y_{i1}, \ldots, y_{i,t-1})'$, we have

$$\begin{aligned} z_{it|t} &= E(z_{it}|y_{it}, y_{i,t-1}^{*}) \\ &= E(z_{it}|y_{i,t-1}^{*}) + \mathrm{Cov}(z_{it}, y_{it}|y_{i,t-1}^{*})\mathrm{Var}(y_{it}|y_{i,t-1}^{*})^{-1}(y_{it} - y_{it|t-1}) \\ &= z_{it|t-1} + \sum_{iz}(t|t-1)H_{it}'\sum_{iy}(t|t-1)^{-1}(y_{it} - y_{it|t-1}) . \end{aligned}$$

These formulas are standard in state space models; see for example Harvey and McKenzie (1984) or Koopman and Shepherd (1992).

Appendix C

A Modified Kalman Filter Recursion for AREMs

Here we describe a modified Kalman filter recursion for computing the conditional moments of the responses under Model (4.2) of Chapter 4. Further define

$$e_{is|t} = E(e_{is} \mid Y_i^t) \quad \text{and} \quad \sigma_{ie}^2(s|t) = Var(e_{is} \mid Y_i^t)$$

where $Y_i^t = \{y_{i1}, y_{i2}, \ldots, y_{it}\}$. Now, for $i = 1, 2, \ldots, N$,

1. Initialize $y_{i1|0} = \mu$, $\sigma_{iy}^2(1|0) = \omega^2 + \sigma_1^2$, $e_{i1|1} = \frac{\sigma_1^2}{\omega^2+\sigma_1^2}(y_{i1} - \mu)$ and $\sigma_{ie}^2(1|1) = \frac{\omega^2\sigma_1^2}{\omega^2+\sigma_1^2}$.

2. Set $j = 2$.

3. (Filtering) Compute

$$\begin{aligned} y_{i,j|j-1} &= y_{i,j-1} - (1-\phi)e_{i,j-1|j-1} \\ \text{and} \quad \sigma_{iy}^2(j|j-1) &= (1-\phi)^2\sigma_{ie}^2(j-1|j-1) + \sigma^2. \end{aligned}$$

4. (Correcting) Compute

$$\begin{aligned} e_{ij|j} &= \phi e_{i,j-1|j-1} + [\phi(\phi-1)\sigma_{ie}^2(j-1|j-1) + \sigma^2]\sigma_{iy}^{-2}(j|j-1)(y_{ij} - y_{ij|j-1}). \\ \sigma_{ie}^2(j|j) &= \phi^2\sigma_{ie}^2(j-1|j-1) + \sigma^2 - [-\phi(1-\phi)\sigma_{ie}^2(j-1|j-1) + \sigma^2]^2\sigma_{iy}^{-2}(j|j-1). \end{aligned}$$

5. Set $j = j + 1$.

6. Goto Step 3 until $j > n_i + 1$.

The recursion can be derived as follows. For each i, when $j = 1$, we can directly observe that

$$y_{i1|0} = E(y_{i1}) = \mu \quad \text{and} \quad \sigma_{iy}^2(1|0) = Var(y_{i1}) = \sigma_\epsilon^2 + \omega^2.$$

Now,

$$\begin{aligned} \begin{pmatrix} e_{i1} \\ y_{i1} \end{pmatrix} &= \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u_i + \begin{pmatrix} 1 \\ 1 \end{pmatrix} e_{i1} \\ &= \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} u_i \\ \epsilon_{i1} \end{pmatrix} \\ &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} \omega^2 & 0 \\ 0 & \sigma_1^2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \right). \end{aligned}$$

Thus, from standard multivariate Normal theory,

$$\begin{aligned} e_{i1|1} &= E(e_{i1}\,|\,y_{i1}) = \frac{\sigma_1^2}{\omega^2+\sigma_1^2}(y_{i1}-\mu) \\ \text{and} \quad \sigma_{ie}^2(1|1) &= Var(e_{i1}\,|\,y_{i1}) = \frac{\omega^2\sigma_1^2}{\sigma_1^2+\omega^2}. \end{aligned}$$

Hence, the initialization step is true. Now, for $j > 1$,

$$\begin{aligned} \begin{pmatrix} e_{ij} \\ y_{ij} \end{pmatrix} &= \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u_i + \begin{pmatrix} 1 \\ 1 \end{pmatrix} e_{ij} \\ &= \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u_i + \begin{pmatrix} \phi \\ \phi \end{pmatrix} e_{i,j-1} + \begin{pmatrix} 1 \\ 1 \end{pmatrix} \epsilon_{ij} \\ &= \begin{pmatrix} 0 \\ \mu \end{pmatrix} + \begin{pmatrix} 0 & \phi & 1 \\ 1 & \phi & 1 \end{pmatrix} \begin{pmatrix} e_{i,j-1} \\ \epsilon_{ij} \end{pmatrix}. \end{aligned}$$

Then, by noting that

$$\begin{aligned} E(u_i\,|\,Y_i^{j-1}) &= y_{i,j-1} - \mu - e_{i,j-1|j-1} \\ \text{and} \quad Cov(u_i,\, e_{i,j-1}\,|\,Y_i^{j-1}) &= -Var(u_i\,|\,Y_i^{j-1}) \ \text{ or } \ -\sigma_{ie}^2(j-1|j-1). \end{aligned}$$

and assuming the recursion is true at $j-1$, we have

$$\left.\begin{pmatrix} e_{ij} \\ y_{ij} \end{pmatrix}\right| Y_i^{j-1} \sim \mathcal{N}\left(\begin{pmatrix} \phi e_{i,j-1|j-1} \\ y_{i,j-1} - (1-\phi)e_{i,j-1|j-1} \end{pmatrix}, \begin{pmatrix} \phi^2\sigma^2_{ie}(j-1|j-1) + \sigma^2_\epsilon & -\phi(1-\phi)\sigma^2_{ie}(j-1|j-1) + \sigma^2_\epsilon \\ -\phi(1-\phi)\sigma^2_{ie}(j-1|j-1) + \sigma^2_\epsilon & (1-\phi)^2\sigma^2_{ie}(j-1|j-1) + \sigma^2_\epsilon \end{pmatrix}\right)$$

and hence the filtering and correcting step follow immediately. ∎

Appendix D

Derivations of Formulas in Chapter 5

D.1 Getting the Posterior Densities

For $j = 1, 2, \ldots, n_i$, the posterior densities are

$$\begin{aligned} g(z_{ij}|T_i^j) &\propto f(t_{ij}|z_{ij}, T_i^{j-1})g(z_{ij}|T_i^{j-1}) \\ &\propto z_{ij}h_0(t_{ij})e^{\beta' x_{ij}(t_{ij})} \exp\left\{-z_{ij}\int_0^{t_{ij}} h_0(t)e^{\beta' x_{ij}(t)}dt\right\} \\ &\quad\ \exp\left\{-\nu_{ij|j-1}z_{ij}\right\} \nu_{ij|j-1}^{\kappa_{ij|j-1}} z_{ij}^{\kappa_{ij|j-1}-1} \\ &\propto h_0(t_{ij})e^{\beta' x_{ij}(t_{ij})} z_{ij}^{\kappa_{ij|j-1}} \exp\left\{-z_{ij}\left(\nu_{ij|j-1} + \int_0^{t_{ij}} h_0(t)e^{\beta' x_{ij}(t)}dt\right)\right\}. \end{aligned}$$

Similarly,

$$g(z_{i,n_i+1}|T_i^{n_i+1}) \propto z_{i,n_i+1}^{\kappa_{i,n_i+1|n_i}-1} \exp\left\{-z_{i,n_i+1}\left(\nu_{i,n_i+1|n_i} + \int_0^{t_{i,n_i+1}} h_0(t)e^{\beta' x_{i,n_i+1}(t)}dt\right)\right\}.$$

Thus, $z_{ij}|_{T_i^j}$ is distributed as $Ga(\kappa_{ij}, \nu_{ij})$ where κ_{ij} and ν_{ij} are as given in (5.7). ∎

D.2 Getting the Multiplicative Transition Process

Note that

$$\begin{aligned} z_{ij}|_{T_i^{j-1}} &\sim Ga(\kappa_{ij|j-1}, \nu_{ij|j-1}) = Ga(\Psi(T_i^{j-1})\kappa_{i,j-1}, \Psi(T_i^{j-1})\nu_{i,j-1}) \\ \text{and} \quad \Psi(T_i^{j-1})^{-1} z_{i,j-1}|_{T_i^{j-1}} &\sim Ga(\kappa_{i,j-1}, \Psi(T_i^{j-1})\nu_{i,j-1}). \end{aligned}$$

By considering the decomposition,

$$\kappa_{i,j-1} = \Psi(T_i^{j-1})\kappa_{i,j-1} + (1 - \Psi(T_i^{j-1}))\kappa_{i,j-1},$$

it follows from standard results, e.g. Rao (1965), that

$$\left. \frac{z_{ij}}{\Psi(T_i^{j-1})^{-1} z_{i,j-1}} \right|_{T_i^{j-1}} \sim Beta(\Psi(T_i^{j-1})\kappa_{i,j-1}, (1 - \Psi(T_i^{j-1}))\kappa_{i,j-1}).$$

Hence, (5.9) follows. ■

D.3 Getting back to Independent Processes

Observe that, for $i = 1, 2, \ldots, N$ and $j = 1, 2, \ldots, n_i + 1$,

$$\begin{aligned} \kappa_{ij|j-1} &= \frac{a(T_i^{j-1})}{w^2} + c_1(\Psi) \\ \text{and} \quad \nu_{ij|j-1} &= \frac{a(T_i^{j-1})}{w^2} + c_2(\Psi, T_i^{j-1}) \end{aligned}$$

where $a(T_i^{j-1}) = \prod_{l=1}^{j-1} \Psi(T_i^l)$, and $c_1(\cdot)$ are $c_2(\cdot)$ are some functions not depending on w^2. Then

$$\begin{aligned}
\frac{\kappa_{ij|j-1}}{\nu_{ij|j-1}} &= \frac{a + c_1 w^2}{a + c_2 w^2} \to 1 \qquad \text{as} \quad w^2 \to 0 \\
\text{and} \quad \left(\frac{\nu_{ij|j-1}}{\nu_{ij}}\right)^{\kappa_{ij|j-1}} &= \left(1 + \frac{\int_0^{t_{ij}} h_0(t) e^{\beta' x_{ij}(t)} dt}{\nu_{ij|j-1}}\right)^{-\kappa_{ij|j-1}} \\
&= \left(1 + \frac{\int_0^{t_{ij}} h_0(t) e^{\beta' x_{ij}(t)} dt}{\frac{a}{w^2} + c_2}\right)^{-(\frac{a}{w^2} + c_1)} \\
&\to \; exp\left\{-\int_0^{t_{ij}} h_0(t) e^{\beta' x_{ij}(t)} dt\right\} \qquad \text{as} \quad w^2 \to 0.
\end{aligned}$$

Hence, as $\omega^2 \to 0$, (5.10) and (5.11) become

$$h_0(t_{ij}) e^{\beta' x_{ij}(t_{ij})} \quad \text{and} \quad exp\left\{-\int_0^{t_{i,n_i+1}} h_0(t) e^{\beta' x_{ij}(t)} dt\right\}$$

respectively which do not depend on their corresponding past history. ■

D.4 Getting the Scores and Hessian Matrix

With $h_0(t) = bt^k$; $b > 0$, $k > -1$, (a Weibull intensity function) and $\Psi(T_i^j) = \psi$, and the transformation of θ to θ_U in (5.13), the log-likelihood function, from (5.12), is

$$\begin{aligned}
l(\theta_U) = \sum_{i=1}^{N} \Bigg\{ & \sum_{j=1}^{n_i} \left[\delta - ue^{\delta} + (e^{\delta} - 1)\log t_{ij} + \log \kappa_{ij|j-1} - \log \nu_{ij}\right] \\
& + \sum_{j=1}^{n_i+1} \kappa_{ij|j-1} \left[\log \nu_{ij|j-1} - \log \nu_{ij}\right] \Bigg\}. \qquad \text{(D.1)}
\end{aligned}$$

The corresponding score function and Hessian matrix among to compute the first and second derivatives of $\kappa_{ij|j-1}$, $\nu_{ij|j-1}$ and ν_{ij} which are evaluated recursively by the followings.

For evaluating the score function, we need

$$\begin{aligned} \frac{\partial \kappa_{i,j+1|j}}{\partial \theta_U} &= \frac{e^{\gamma-\tau}}{1+e^{\gamma-\tau}} \frac{\partial \kappa_{ij|j-1}}{\partial \theta_U} + \frac{\kappa_{i,j+1|j}}{1+e^{\gamma-\tau}} (0, 0, 1, -1), \\ \frac{\partial \nu_{i,j+1|j}}{\partial \theta_U} &= \frac{e^{\gamma-\tau}}{1+e^{\gamma-\tau}} \frac{\partial \nu_{ij}}{\partial \theta_U} + \frac{\nu_{i,j+1|j}}{1+e^{\gamma-\tau}} (0, 0, 1, -1), \\ \text{and} \quad \frac{\partial \nu_{ij}}{\partial \theta_U} &= \frac{\partial \nu_{ij|j-1}}{\partial \theta_U} - \exp(\delta - ue^{\delta}) t_{ij}^{e^{\delta}} (1, u - \log(t_{ij}), 0, 0) \end{aligned}$$

with starting values $\frac{\partial \kappa_{i1|0}}{\partial \theta_U} = \frac{\partial \nu_{i1|0}}{\partial \theta_U} = (0, 0, -e^{-\gamma}, 0)$. With

$$M(\epsilon) = \left(\underline{0}_4 \vdots \underline{0}_4 \vdots \frac{\partial \epsilon}{\partial \theta_U} \vdots - \frac{\partial \epsilon}{\partial \theta_U} \right), \qquad \underline{0}_p = p \times 1 \text{ vector of } 0,$$

$$B(\epsilon) = Block\left(0, 0, \epsilon \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix} \right), \qquad Block(a, b, c) = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix},$$

and $m_{ij} = (1, u - \log(t_{ij}))$,

the Hessian matrix is evaluated using

$$\begin{aligned} \frac{\partial^2 \kappa_{i,j+1|j}}{\partial \theta_U \theta_U^T} &= \frac{e^{\gamma-\tau}}{1+e^{\gamma-\tau}} \frac{\partial^2 \kappa_{ij|j-1}}{\partial \theta_U \theta_U^T} + \frac{1}{1+e^{\gamma-\tau}} \left[M(\kappa_{i,j+1|j}) + M^T(\kappa_{i,j+1|j}) + B(\kappa_{i,j+1|j}) \right], \\ \frac{\partial^2 \nu_{i,j+1|j}}{\partial \theta_U \theta_U^T} &= \frac{e^{\gamma-\tau}}{1+e^{\gamma-\tau}} \frac{\partial^2 \nu_{ij}}{\partial \theta_U \theta_U^T} + \frac{1}{1+e^{\gamma-\tau}} \left[M(\nu_{i,j+1|j}) + M^T(\nu_{i,j+1|j}) + B(\nu_{i,j+1|j}) \right] \\ \text{and} \quad \frac{\partial^2 \nu_{ij}}{\partial \theta_U \theta_U^T} &= \frac{\partial^2 \nu_{ij|j-1}}{\partial \theta_U \theta_U^T} + Block\left(\exp(\delta - ue^{\delta}) t_{ij}^{e^{\delta}} \left[e^{\delta} m_{ij} m_{ij}^T - \begin{pmatrix} 0 & \vdots & \\ & \vdots & m_{ij} \\ 1 & \vdots & \end{pmatrix} \right], 0, 0 \right) \end{aligned}$$

with starting values $\frac{\partial^2 \kappa_{i1|0}}{\partial\theta_U \theta_U^T} = \frac{\partial^2 \nu_{i1|0}}{\partial\theta_U \theta_U^T} = Block\left(0, 0, \begin{pmatrix} e^{-\gamma} & 0 \\ 0 & 0 \end{pmatrix}\right)$.

Then the inverted Hessian matrix for θ can be computed from

$$\left(\frac{\partial^2 l}{\partial\theta\theta^T}\right)^{-1} = \left(\frac{\partial\theta}{\partial\theta_U^T}\right)\left(\frac{\partial^2 l}{\partial\theta_U\theta_U^T}\right)^{-1}\left(\frac{\partial\theta}{\partial\theta_U^T}\right)^T$$

where $\frac{\partial\theta}{\partial\theta_U^T} = Block\left(\begin{pmatrix} -\exp(2\delta - ue^\delta) & \exp(\delta - ue^\delta)(1 - ue^\delta) \\ 0 & e^\delta \end{pmatrix}, \begin{pmatrix} e^\gamma & 0 \\ \frac{e^{\gamma-\tau}}{(1+e^{\gamma-\tau})^2} & -\frac{e^{\gamma-\tau}}{(1+e^{\gamma-\tau})^2} \end{pmatrix}\right)$.

■

Bibliography

Aalen, O. O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, **3**, 227–243.

Aalen, O. O. and Husebye, E. (1991). Statistical Analysis of Repeated Events Forming Renewal Processes. *Statistics in Medicine*, **10**, 1227–1240.

Agrawal, R., Lawless, J. F., and MacKay, R. J. (1997). Analysis of variation transmission for manufacturing processes in the presence of measurement error. *Submitted for publication.*

Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes.* Springer-Verlag, New York.

Anderson, B. D. O. and Moore, J. B. (1979). *Optimal Filtering.* Prentice Hall, New Jersey.

Baker, S. G. (1995). Marginal Regression for Repeated Binary Data with Outcome Subject to Non-ignorable Non-response. *Biometrics*, **51**, 1042–1052.

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

Cargnoni, C., Müller, P., and West, M. (1997). Bayesian Forecasting of Multinomial Time Series Through Conditionally Gaussian Dynamic Models. *Journal of the American Statistical Association*, **92**, 640–647.

Carlin, B. P., Polson, N. G., and Stoffer, D. S. (1992). A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, **87**, 493–500.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*. **81**, 541–553.

Carter, C. K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika*, **3**, 589–601.

Chan, J. S. K. and Kuk, A. Y. C. (1997). Maximum Likelihood Estimation for Probit-Linear Mixed Models with Correlated Random Effects. *Statistics in Medicine*. **53**, 86–97.

Clayton, D. G. (1991). A Monte Carlo Method for Bayesian Inference in Frailty Models. *Biometrics*, **47**, 467–485.

Clayton, D. G. (1994). Some approaches to the analysis of recurrent event data. *Statistical Methods in Medical Research*, **3**, 244–262.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society, Series B*, **34**, 187–220.

Cox, D. R. (1981). Statistical analysis of time series, some recent developments (with discussion). *Scandinavian Journal of Statistics*, **8**, 93–115.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood estimation with incomplete data via the EM algorithm (with discussion). *Journal of the Royal Statistical Society B*, **39**, 1–38.

Diggle, P. and Kenward, M. G. (1994). Informative Drop-out in Longitudinal Data Analysis. *Applied Statistics*, **43**, 49–93.

Diggle, P. J., Liang, K. Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford University Press, Oxford.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap.* Chapman and Hall, London.

Elliott, R. J., Aggoun, L., and Moore, J. B. (1995). *Hidden Markov Models: Estimation and Control.* Springer-Verlag, New York.

Fahrmeir, L. and Kaufmann, H. (1991). On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression. *Metrika*, **38**, 37–60.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate Statistical Modelling Based on Generalized Linear Models.* Springer-Verlag, New York.

Fahrmeir, L. and Wagenpfeil, S. (1996). Smoothing Hazard Funstions and Time-Varying Effects in Discrete Duration and Competing Risks Models. *Journal of the American Statistical Association*, **91**, 1584–1594.

Follmann, D. and Wu, M. (1995). An Approximate Generalized Linear Model With Random Effects For Informative Missing Data. *Biometrics*, **51**, 151–168.

Fong, D. Y. T. and Lawless, J. F. (1997). The Analysis of Process Variation Transmission with Multivariate Measurements. *To Appear in Statistica Sinica.*

Fuller, W. A. (1987). *Measurement Error Models.* John Wiley and Sons, New York.

Gelfand, A. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, **85**, 398–409.

Godambe, V. P. (1985). The foundations of finite sample estimation in stochastic processes. *Biometrika*, **72**, 419–428.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and Optimal Estimation. *International Statistical Review*, **55**, 231–244.

Godambe, V. P. and Thompson, M. E. (1989). An Extension of Quasi-likelihood Estimation. *Journal of Statistical Planning and Inference*, **22**, 137–152.

Hamada, M. S. and Lawless, J. F. (1994). Multivariate Methods for the Assessment of Process Variation Transmission. Technical report, The Institute for Improvement in Quality and Productivity, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian Forecasting (with disscussion). *Journal of the Royal Statistical Society, Series B*, **38**, 205–247.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models, and the Kalman Filter.* Cambridge University Press, Cambridge.

Harvey, A. C. and Fernandes, C. (1989). Time Series Models for Count or Qualitative Observations. *Journal of Business & Economic Statistics*, **7**, 407–423.

Harvey, A. C. and McKenzie, C. R. (1984). Missing observations in dynamic econometric models: A partial synthesis. In E. Parzen, editor, *Time Series Analysis of Irregularly Observed Data, Proceedings, College Station 1983*, volume 25, pages 108–133, New York. Springer-Verlag.

Haukka, J. K. (1995). Correction for Covariate Measurement Error in Generalized Linear Models-A Bootstrap Approach. *Biometrics*, **51**, 1127–1132.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis*, **1**, 255–273.

Ibrahim, J. G. (1990). Incomplete Data in Generalized Linear Models. *Journal of the American Statistical Association*, **85**, 765–769.

Jazwinski, A. H. (1970). *Stochastic Processes and Filtering Theory*. Academic Press, New York.

Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: A State-space Approach*. Chapman and Hall, London.

Jørgensen, B., Lundbye-Christensen, S., Song, X., and Sun, L. (1996a). A Longitudinal Study of Emergency Room Visits and Air Pollution for Prince George, Bristish Columbia. *Statistics in Medicine*, **15**, 823–836.

Jørgensen, B., Lundbye-Christensen, S., Song, X., and Sun, L. (1996b). State-space models for multivariate longitudinal data of mixed types. *Canadian Journal of Statistics*, **24**, 385–402.

Kalman, R. E. (1960). A New Approach to Linear Filtering and Prediction Problems. *Journal of Basic Engineering*, **82**, 35–45.

Kalman, R. E. and Bucy, R. S. (1961). New Results in Linear Filtering and Prediction Theory. *Journal of Basic Engineering*, pages 95–108.

Kitagawa, G. (1987). Non-Gaussian State-Space Modelling of Non-stationary Time Series (with Comments). *Journal of the American Statistical Association*, **82**, 1032–1063.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika*, **80**, 117–126.

Koopman, S. J. and Shepherd, N. (1992). Exact score for time series models in state space form. *Biometrika*, **79**, 823–826.

Kuk, A. Y. C. (1995). Asymptotically Unbiased Estimation in Generalized Linear Models with Random Effects. *Journal of the Royal Statistical Society, Series B*, **57**, 395–407.

Laird, N. M. (1988). Missing Data in Longitudinal Studies. *Statistics in Medicine*, **7**, 305–315.

Lambert, P. (1996a). Modeling of Nonlinear Growth Curve on Series of Correlated Count Data Measured at Unequally Spaced Times: A Full Likelihood Based Approach. *Biometrics*, **52**, 50–55.

Lambert, P. (1996b). Modelling of Repeated Series of Count Data Measured at Unequally Spaced Times. *Applied Statistics*, **45**, 31–38.

Lawless, J. F. (1982). *Statistical Models and Methods for Lifetime Data*. John Wiley & Sons, Toronto.

Lawless, J. F. (1995). The Analysis of Recurrent Events for Multiple Subjects. *Applied Statistics*, **44**, 487–498.

Lawless, J. F. and Fong, D. Y. T. (1997). State Duration Models in Clinical and Observational Studies. *To be Submitted.*

Lawless, J. F. and Yan, P. (1992). Some Statistical Methods for Followup Studies of Disease with Intermittent Monitoring. In F. M. Hoppe, editor, *Multiple Comparisons, Selection, and Applications in Biometry*, volume 134, pages 427–446, New York. Marcel Dekker, Inc.

Lawless, J. F., MacKay, R. J., and Robinson, J. A. (1997). Analysis of variation transmission in manufacturing processes. *Submitted for publication.*

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Liang, K. Y., Zeger, S. L., and Qaqish, B. (1992). Multivariate regression analyses for categorical data (with Discussion). *Journal of Royal Statistical Society*, **B 54**, 3–40.

Lindsey, J. K. (1993). *Models for Repeated Measurements.* Oxford University Press, Oxford.

Lipsitz, S. R., Laird, N. M., and Harrington, D. P. (1994). Weighted Least Squares Analysis of Repeated Categorical Measurements with Outcomes Subject to Nonresponse. *Biometrics*, pages 11–24.

Little, R. J. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of the American Statistical Association*, **90**, 1112–1121.

Little, R. J. A. (1992). Incomplete Data in Event History Analysis. In *Chapter 8 of Trussell J., Hankinson, R. and Titon J. (1992). Demographic Applications of Event History Analysis*, pages 209–230. Oxford.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. John Wiley and Sons, New York.

Louis, T. A. (1982). Finding Observed Information Using the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, **44**, 98–130.

Louis, T. A. (1988). General Methods for Analyzing Repeated Measures. *Statistics in Medicine*, **7**, 29–45.

Lütkepohl, H. (1993). *Introduction to Multiple Time Series Analysis*. Springer-Verlag, New York, 2 edition.

Mardia, K. V., Kent, J., and Bibby, J. M. (1979). *Multivariate Analysis*. Academic Press, London.

McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. Chapman and Hall, London, 2nd edition.

McCullagh, P. and Tibshirani, R. (1990). A simple method for the adjustment of profile likelihoods. *Journal of the Royal Statistical Society, Series B*, **52**, 325–344.

McCulloch, C. E. (1997). Maximum Likelihood Algorithms for Generalized Linear Mixed Models. *Journal of the American Statistical Association*, **92**, 162–170.

McGilchrist, C. A. (1994). Estimation in Generalized Mixed Models. *Journal of the Royal Statistical Society, Series B*, **56**, 61–69.

Meng, X. L. and Rubin, D. B. (1991). Using EM to Obtain Asymptotic Variance Covariance Matrices: the SEM Algorithm. *Journal of the American Statistical Association*, **86**, 899–909.

Naik-Nimbalkar, U. V. and Rajarshi, M. B. (1995). Filtering and Smoothing Via Estimating Functions. *Journal of the American Statistical Association*, **90**, 301–306.

Nelder, J. A. and Mead, R. (1965). A simplex method for function minimization. *Computer Journal*, **7**, 308–313.

Ng, E. T. M. and Cook, R. J. (1997). Modeling Two-state Disease Processes with Random Effects. *To Appear in Lifetime Data Analysis.*

Petersen, J. H., Andersen, P. K., and Gill, R. D. (1996). Variance components models for survival data. *Statistica Neerlandica*, **50**, 193–211.

Pickles, A. and Crouchley, R. (1994). Generalization and applications of frailty models for survival and event data. *Statistical Methods in Medical Research*, **3**, 263–278.

Press, W. H., Flannery, B., Teukolsky, S., and Vetterling, W. T. (1986). *Numerical Recipes.* Cambridge University Press, Cambridge.

Raboud, J., Reid, N., Coates, R., and Farewell, V. T. (1993). Estimating Risks of Progressing to Aids when Covariates are Measured with Error. *Journal of Royal Statistical Society*, **A 156**, 393–406.

Rao, C. R. (1965). *Linear Statistical Inference and Its Applications.* John Wiley & Sons, Inc.,.

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, **63**, 581–592.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, **78**, 719–727.

Schnatter, S. (1992). Integration-Based Kalman-Filtering for a Dynamic Generalized Linear Trend Model. *Computational Statistics and Data Analysis*, **13**, 447–459.

Sepanski, J. H., Kinickerbockier, R., and Carroll, R. J. (1994). A Semiparametric Correction for Attenuation. *Journal of the American Statistical Association*, **89**, 1366–1373.

Shumway, R. H. (1988). *Applied Statistical Time Series Analysis*. Prentice-Hall, Englewood Cliffs, N.J.

Singh, A. C. and Roberts, G. (1992). State Space Modelling of Cross-Classified Time Series of Counts. *International Statistical Review*, **60**, 321–335.

Smith, R. L. and Miller, J. E. (1986). A Non-Gaussian State Space Model and Application to Prediction of Records. *Journal of Royal Statistical Society*, **Ser B, 48**, 79–88.

Stoffer, D. S. and Wall, K. D. (1991). Bootstrapping State-Space Models: Gaussian Maximum Likelihood Estimation and the Kalman Filter. *Journal of the American Statistical Association*, **86**, 1024–1033.

Sutradhar, B. C. (1990). Exact maximum likelihood estimation for the mixed analysis of variance model with autocorrelated errors. *The Statistician*, **39**, 3–9.

Sutradhar, B. C. and Rao, J. N. K. (1996). Estimation of regression parameters in generalized linear models for cluster correlated data with measurement error. *The Canadian Journal of Statistics*, **24**, 177–192.

Taylor, J. M. G., Cumberland, W. G., and Sy, J. P. (1994). A Stochastic Model for Analysis of Longitudinal AIDS Data. *Journal of the American Statistical Association*, **89**, 727–736.

Thavaneswaran, A. and Abraham, B. (1988). Estimation for Non-linear Time Series Models using Estimating Equations. *Journal of Time Series Analysis*, **9**, 99–108.

Thavaneswaran, A. and Thompson, M. E. (1986). Optimal Estimation for Semimartingales. *Journal of Applied Probability*, **23**, 409–417.

Thompson, M. E. and Kaseke, T. N. (1995). Estimation for partially observed Markov processes. *Stochastic Hydrology and Hydraylics*, **9**, 33–47.

Tsiatis, A. A., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the Relationship of Survival to Longitudinal Data Measured with Error. Applications to Survival and CD4 Counts in Patients with AIDS. *Journal of the American Statistical Association*, **90**, 27–37.

Tsimikas, J. and Ledolter, J. (1994). REML and Best Linear Unbiased Prediction in State Space Models. *Communications in Statistics: Theory and Methods*, **23**, 2253–2268.

Waclawiw, M. A. and Liang, K. Y. (1993). Prediction of random effects in the generalized linear model. *Journal of the American Statistical Association*, **88**, 171–178.

Wei, L. J., Lin, D. Y., and Weissfeld, L. (1989). Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions. *Journal of the American Statistical Association*, **84**, 1065–1073.

West, M. and Harrison, J. (1997). *Bayesian Forecasting and Dynamic Models.* Springer-Verlag, New York, 2nd edition.

West, M., Harrison, P. J., and Migon, H. S. (1985). Dynamic Generalized Linear Models and Bayesian Forecasting (with comment). *Journal of the American Statistical Association*, **80**, 73–97.

Wilson, P. D. (1988). Autoregressive Growth Curves and Kalman Filtering (with Discussions). *Statistics in Medicine*, **7**, 73–92.

Xue, X. and Brookmeyer, R. (1996). Bivarite frailty model for the analysis of multivarite failure time. *Lifetime Data Analysis*, **2**, 277–289.

Yue, H. and Chan, K. S. (1994). A Dynamic Frailty Model for Multivariate Survival Data. Unpublished manuscript.

Zeger, S. L. and Karim, M. R. (1991). Generalized linear models with random effects: a Gibbs sampling approach. *Journal of the American Statistical Association*, **86**, 79–86.

Zeger, S. L. and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

Zeger, S. L. and Liang, K. Y. (1992). An Overview of Methods for the Analysis of Longitudinal Data. *Statistics in Medicine*, **11**, 1825–1839.