May 2015

Shape-Invariant Models for Non-Independent Functional Data

Wen Yang
University of Wisconsin-Milwaukee

Shape-Invariant Models for Non-Independent Functional Data

by

Wen Yang

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in

Mathematics

at

The University of Wisconsin-Milwaukee

May 2015

ABSTRACT

Shape-Invariant Models for Non-Independent Functional Data

by

Wen Yang

The University of Wisconsin-Milwaukee, 2015
Under the Supervision of Professor Daniel Gervini

Non-independent functional data frequently arise in evolutionary and biological studies. It is important to possess models that incorporate correlations between subjects and appropriately describe the relationships between response and covariates. The variation in the response curves is usually a mixture of amplitude and phase variation, both of which should be explicitly modeled for efficient statistical inference. In this dissertation we propose a shape-invariant model that explicitly addresses amplitude and phase variability. We incorporate genetic and environmental random effects for the parameters, and use the additive genetic information matrix in the representation of the covariance matrices to make the unobservable genetic components mathematically identifiable. We derive the asymptotic properties of the maximum likelihood estimators and study their finite sample behavior by simulation. Then we apply the new method to the analysis of growth curves of flour beetles.

To my husband and my parents.

Contents

List of Figures

List of Tables

Acknowledgements

First of all, I cannot express enough thanks to my advisor Dr. Daniel Gervini for his continuous guidance and encouragement during my graduate study. He taught me to think independently and solve any difficulty with patience. I could never have this dissertation finished without his help. Besides, I would like to thank the rest of my dissertation committee: Dr. Jay Beder, Dr. Vytaras Brazauskas, Dr. Dashan Fan, and Dr. Jugal Ghorai for their special attention and support.

I would also like to thank the Department of Mathematical Sciences at the University of Wisconsin-Milwaukee, who provided me good environment to study, practice, and develop. I enjoyed the wonderful time with my friends here.

Last but not least, I would like to thank my parents for their hearty support and encouragement. My deepest thanks go to my beloved husband who always stood by me when I was struggling.

Chapter 1

Introduction

In many studies, data are collected by observing a number of observations on each subject at different times or under different conditions, and this type of data is called *repeated measures*. For example, the daily weights of a subject are repeated measures which record the same type of information on the same subject over a period of time. The outcome of interest in these studies may be continuous rather than discrete, such as the shape of the response curve itself, the slope at a particular point, or the locations of extrema. For repeated measures, it is common that the relationship between the explanatory variable (e.g., time) and the response variable (e.g., growth) is nonlinear. Therefore, a nonlinear regression model can usually fit the data. However, the results sometimes may be difficult to interpret.

Self-modeling regression, introduced by Lawton *et al.* (1972), provides a way of estimating the model and its parameters at the same time. It allows the data to choose the underlying model by utilizing the information from themselves. When the knowledge of model building is limited, it would be more flexible to use self-modeling regression than the classical nonlinear regression approaches.

The model that implements the idea of the self-modeling approach is called *shape-invariant model* (SIM). It is a semi-parametric method of estimating a functional re-

lationship and has been used in many studies. SIM assumes a common shape function for all subjects and the curve of each subject is a shift or/and a scale transformation of the common shape. The estimation of the common shape function is usually non-parametric, using the entire data set, while subject-specific effects are modeled in a parametric way. The parameters in a SIM usually have a straightforward interpretation, i.e., shift or scale, and can be estimated by techniques in nonlinear regression. Lawton *et al.* (1972) implemented the shape function through a piece-wise linear spline, and Lindstrom (1995) extended it by introducing a free-knot (knots are allowed to vary) cubic spline function which is smooth and provides more flexibility of the shape function. In addition, the use of the mixed-effects structure on the parameters reduces the number of parameters and allows inclusion of subjects with few observations (Lindstrom, 1995).

In many cases, the variability of shift and scale between subjects in SIMs can be viewed as phase (usually in time) and amplitude variabilities, respectively. Taking the growth curves of insects as an example, all insects experience certain events in their life cycle. Although all growth curves look similar, each insect reaches its own growth peak at a particular time with a particular weight. These features vary from one subject to another, which produces the variation of curves both in time and amplitude. Variation of features in amplitude is usually given more attention in the statistical analysis, but it is worth noting that ignoring the time variability can lead to severe problems on the results. For example, it is natural to use the point-wise sample mean to estimate the population mean function. However, the point-wise mean may not resemble any sample curve because of the confounding effect of time and amplitude variation. Therefore, it is important to take both time and amplitude variability into account when building models. In recent years, Functional Data Analysis (FDA), which is the study of samples of functions or curves, has developed tools for curve registration (or time-warping), which aim at adjusting for random

time distortions (Kneip and Engel, 1995; Ramsay and Li, 1998; Wang and Gasser, 1999; Gervini and Gasser, 2004). Curve registration eliminates time variation by aligning salient features of the curves and gives a more sensible structural mean than the point-wise mean.

In biological studies, correlations between subjects are common. For instance, in evolutionary studies of pedigree data, subjects can have a variety of relationships, such as grandparent and grandchild, parent and child, and half/full siblings. Therefore, it is not always appropriate to assume independence between subjects. Since the relationship between subjects can be represented by genetic information, it is useful to investigate their genetic correlations.

In quantitative genetics, two types of variability are involved in the observed physical traits, genetic variability and environmental variability. The knowledge about genetic variability is important in determining how physical traits develop and evolve. The additive genetic relationship matrix, which describes relationships among individuals, is often used in quantitative genetic literature (Henderson, 1975). It is important to possess statistical tools for non-independent functional data involving time and amplitude variability. Gervini and Carter (2014) proposed a functional ANOVA approach explicitly addressing time variability through a time-warping component which sheds light on our shape-invariant model building process.

This dissertation presents a shape-invariant model for genetically correlated functional data. We introduce genetic and environmental random effects representing time and amplitude variability, and utilize the additive genetic relationship matrix to incorporate the genetic correlation among subjects. We develop a four-parameter model for a special growth pattern. For the general situations, we propose a self-modeling regression which can handle more complex data structures. These shape-invariant models solve the problems caused by time variation and inherit the flexibility of self-modeling regression and nonlinear mixed-effects modeling.

This dissertation is organized as follows. In Chapter 2, we review the background of self-modeling regression, nonlinear mixed-effects models, curve registration, and some useful results from quantitative genetics. We also introduce the Beetle's growth data as an example to describe the issues we want to solve. Chapter 3 presents the new shape-invariant model. In Chapter 4 we prove the consistency and asymptotic normality of the estimators. In Chapter 5 we conduct a simulation study, and in Chapter 6 we illustrate the model using the flour-beetle growth data.

Chapter 2

Preliminaries

In this chapter, we first present the background of nonlinear mixed-effects models for repeated measures data and self-modeling regression. Then, we review the techniques for curve registration in functional data analysis. A real data example is presented to describe the issues in model building. In the last section, some important results in quantitative genetics are provided.

2.1 Nonlinear mixed-effects model

In studies of biology and physical sciences, data commonly consist of measurements repeatedly taken on each subject. This kind of data is termed repeated measures. Mixed-effects models for repeated measures data have become prevailing because of their flexibility. They can easily handle unbalanced repeated-measures data and allow flexible specification of the covariance structure. In many applications, the relationship between the response y and the covariates $\boldsymbol{x}$ is nonlinear. Lindstrom and Bates (1990) proposed a general nonlinear mixed-effects model for repeated measures data as a hierarchical model that in some ways generalizes both the linear mixed-effects model of Laird and Ware (1982) and the standard fixed-effects nonlinear model of Bates and Watts (1988). In this section, we carefully review the main results of

Lindstrom and Bates (1990).

Let y_{ij} and $\boldsymbol{x}_{ij}$ be the jth response and covariates of the ith subject. There are two stages in this hierarchical model. In the first stage, the nonlinear mixed-effects model is defined as

$$y_{ij} = f(\boldsymbol{\phi}_i, \boldsymbol{x}_{ij}) + \epsilon_{ij}, \quad i = 1, \cdots, M, \quad j = 1, \cdots, n_i, \tag{2.1}$$

where $\boldsymbol{\phi}_i$ is a r-dimensional vector of subject-specific parameters, $\boldsymbol{x}_{ij}$ is covariates, f is a nonlinear function of $\boldsymbol{\phi}_i$ and $\boldsymbol{x}_{ij}$, and ϵ_{ij} is a normally distributed noise term. M is the total number of subjects and n_i is the number of observations on the ith subject. In the second stage, we introduce a mixed-effects structure into the subject-specific parameter vector $\boldsymbol{\phi}_i$, and it is given by

$$\boldsymbol{\phi}_i = \boldsymbol{A}_i\boldsymbol{\beta} + \boldsymbol{B}_i\boldsymbol{b}_i, \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}), \tag{2.2}$$

where $\boldsymbol{\beta}$ is a p-dimensional vector of the fixed effects and is assumed to be the same for all subjects, $\boldsymbol{b}_i$ is a q-dimensional vector of the random effects and is assumed to follow a multivariate normal distribution with zero mean and covariance matrix $\boldsymbol{D}$. The matrices $\boldsymbol{A}_i$ and $\boldsymbol{B}_i$ are the corresponding design matrices with dimensions $r \times p$ and $r \times q$ for the fixed and random effects, respectively. For example, suppose we have $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, \phi_{i3})^T$, and $\phi_{i1} = \beta_1 + b_{i1}, \phi_{i2} = \beta_2, \phi_{i3} = \beta_3$, which means only the first parameter is allowed to vary from subject to subject. Writing them into the form as in (2.2), we have $\boldsymbol{A}_i = \boldsymbol{I}, \boldsymbol{B}_i = (1, 0, 0)^T$. (2.1) and (2.2) together give the general form of the nonlinear mixed-effects model. Collecting the responses and errors for the

ith individual into vectors $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{in_i})^T$ and $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \cdots, \epsilon_{in_i})^T$, and letting

$$\boldsymbol{f}_i(\boldsymbol{\phi}_i) = \begin{pmatrix} f(\boldsymbol{\phi}_i, \boldsymbol{x}_{i1}) \\ f(\boldsymbol{\phi}_i, \boldsymbol{x}_{i2}) \\ \vdots \\ f(\boldsymbol{\phi}_i, \boldsymbol{x}_{in_i}) \end{pmatrix},$$

we obtain the model for the ith subject

$$\boldsymbol{y}_i = \boldsymbol{f}_i(\boldsymbol{\phi}_i) + \boldsymbol{\epsilon}_i. \tag{2.3}$$

Following the classical framework of model given by (2.1), we assume that $\boldsymbol{\epsilon}_i \sim N(\boldsymbol{0}, \boldsymbol{\Lambda}_i)$. $\boldsymbol{\Lambda}_i$ is a matrix that depends on i only through its dimension, and in many situations, $\boldsymbol{\Lambda}_i = \boldsymbol{I}$. We also assume that the random effects $\boldsymbol{b}_i$ are uncorrelated with the error term $\boldsymbol{\epsilon}_i$. Next, we write the M individual models as one by defining

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_M \end{pmatrix}, \quad \boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \vdots \\ \boldsymbol{\phi}_M \end{pmatrix}, \quad \boldsymbol{f}(\boldsymbol{\phi}) = \begin{pmatrix} \boldsymbol{f}_1(\boldsymbol{\phi}_1) \\ \boldsymbol{f}_2(\boldsymbol{\phi}_2) \\ \vdots \\ \boldsymbol{f}_I(\boldsymbol{\phi}_M) \end{pmatrix},$$

$\tilde{\boldsymbol{D}} = \text{diag}(\boldsymbol{D}, \boldsymbol{D}, \cdots, \boldsymbol{D})$, and $\boldsymbol{\Lambda} = \text{diag}(\boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2, \cdots, \boldsymbol{\Lambda}_M)$. Then, the whole model is

$$\boldsymbol{y}|\boldsymbol{b} \sim N(\boldsymbol{f}(\boldsymbol{\phi}), \boldsymbol{\Lambda}), \quad \boldsymbol{\phi} = \boldsymbol{A}\boldsymbol{\beta} + \boldsymbol{B}\boldsymbol{b}, \tag{2.4}$$

$$\boldsymbol{b} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{D}}), \tag{2.5}$$

where

$$\boldsymbol{B} = \text{diag}(\boldsymbol{B}_1, \boldsymbol{B}_2, \cdots, \boldsymbol{B}_M), \quad \boldsymbol{b} = \begin{pmatrix} \boldsymbol{b}_1 \\ \boldsymbol{b}_2 \\ \vdots, \\ \boldsymbol{b}_M \end{pmatrix}, \quad \boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 \\ \boldsymbol{A}_2 \\ \vdots \\ \boldsymbol{A}_M \end{pmatrix}.$$

When the variance components of $\boldsymbol{\Lambda}$ and $\boldsymbol{D}$ are known and $\boldsymbol{f}$ is a linear function of $\boldsymbol{\beta}$ and $\boldsymbol{b}$, the estimators of $\boldsymbol{\beta}$ and $\boldsymbol{b}$ can be derived by the generalized least squares approach. In the case that they are unknown and $\boldsymbol{f}$ is a nonlinear in $\boldsymbol{\beta}$ and $\boldsymbol{b}$, the estimation to the parameters usually makes use of the likelihood approach based on the marginal density of $\boldsymbol{y}$

$$p(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{D}, \tilde{\boldsymbol{\Lambda}}, \sigma^2) = \int p(\boldsymbol{y}|\boldsymbol{b}, \boldsymbol{\beta}, \boldsymbol{D}, \tilde{\boldsymbol{\Lambda}}, \sigma^2) p(\boldsymbol{b})\, d\boldsymbol{b}. \tag{2.6}$$

Unfortunately, this integral does not have a closed-form expression because the model function $\boldsymbol{f}$ is nonlinear in $\boldsymbol{\beta}$ and $\boldsymbol{b}$.

Different approximation procedures have been proposed, such as LME method (Lindstrom and Bates, 1990), Laplace's approximation (Tierney and Kadane, 1986), importance sampling (Geweke, 1989) and Gaussian quadrature (Davidian and Gallant, 1992). Pinheiro and Bates (1995) compared these four different approximations based on their computational and statistical properties and concluded that LME method, Laplace's approximation, and Gaussian quadrature centered at the conditional modes of the random effects are quite accurate and computationally efficient. In this dissertation, we use the Laplace's approximation, since it has a relatively simple mathematical form which is illustrated as follows.

In general, Laplace's approximation is used to approximate integrals of the form

$$\int e^{-g(\boldsymbol{x})}\, d\boldsymbol{x},$$

where $\boldsymbol{x} \in \mathbb{R}^d$. Assume that $g(\boldsymbol{x})$ is smooth and has a global minimum point $\boldsymbol{x}^*$, then the second-order Taylor expansion of g around $\boldsymbol{x}^*$ is

$$g(\boldsymbol{x}) \approx g(\boldsymbol{x}^*) + \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \mathbf{H}_g(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*), \tag{2.7}$$

where $\mathbf{H}_g$ is the Hessian of g. The linear term vanishes because the gradient $\mathbf{D}_g(\boldsymbol{x}^*) = 0$. Then, the Laplace's approximation is

$$\begin{aligned} \int \exp\{-g(\boldsymbol{x})\}\, d\boldsymbol{x} &\approx \int \exp\left\{-g(\boldsymbol{x}^*) - \frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^*)^T \mathbf{H}_g(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*)\right\} d\boldsymbol{x} \\ &= \exp\{-g(\boldsymbol{x}^*)\} \frac{(2\pi)^{d/2}}{\sqrt{\det \mathbf{H}_g(\boldsymbol{x}^*)}}. \end{aligned} \tag{2.8}$$

2.2 Self-modeling regression

As we discussed in the previous section, nonlinear models arise frequently in studies of repeated-measures data. The relevant nonlinear relationship between the response and the covariates can be derived on physical or mechanistic backgrounds or simply to provide an empirical description of the data. However, in the case of lack of a priori specification, it is hard to find a model that can provide a good fit to the data, or the results from the proposed model are difficult to interpret.

Self-modeling regression (Lawton *et al.*, 1972) is used to model the continuous response curve for a number of individuals and provides more flexibility to the nonlinear relationship between the response and covariates. It utilizes the information of data themselves to describe the nonlinear relationship, especially when no appropriate priori information is available. In many studies, individual curves follow a similar trend, such as the growth curves of human being from born to adult which all experience a period of pubertal spurt and then become relatively stable. Self-modeling regression makes use of this characteristic. It assumes that all individual curves share a common

shape and a particular curve is a shift or/and a scale transformation of the common shape.

The model that implements the idea of self-modeling regression is called the *shape-invariant model*. For the ith subject, the shape-invariant model (Lawton *et al.*, 1972) can be represented as

$$y_i(x) = \theta_{i1} + \theta_{i2} \cdot s\left(\frac{x - \theta_{i3}}{\theta_{i4}}\right) + \epsilon_i(x), \tag{2.9}$$

where $s(x)$ denotes the shape function which describes the characteristic shape of the response curve, $\boldsymbol{\theta}_i = (\theta_{i1}, \cdots, \theta_{i4})^T$ is the parameter vector, and $\epsilon_i(x)$ is the random error. As shown in model (2.9), the parameters have direct graphical interpretation: θ_{i1} and θ_{i3} are the shift parameters for the ith curve on the y and x axes respectively, and θ_{i2} and θ_{i4} are the corresponding scale parameters on the y and x axes. Using this approach, the problem of model selection is simplified to that of selecting the appropriate shape function $s(x)$. In some cases, a simple mathematical function can be an appropriate choice of the shape function $s(x)$. For instance, the logistic model $s(x) = 1/(1 + \exp(-x))$ and the Gompertz model $s(x) = \exp(-\exp(-x))$ are often used in studies on the pubertal spurt (Marubini *et al.*, 1972). In other contexts, one can implement a piece-wise linear spline (Lawton *et al.* 1972) or cubic splines (Lindstrom 1995) for a more complex nonlinear relationship. When the shape function $s(x)$ is unknown, both $s(x)$ and the parameter vector $\boldsymbol{\theta}_i$ have to be estimated from the data simultaneously using an iterative procedure through either the nonlinear least squares regression or a likelihood based approach.

Identifiability is a pervasive issue in self-modeling regressions, that is, given all possible realizations of the underlying process, it is possible that different functions and parameter vectors generate the same model. For example, in model (2.9), one could achieve the same model by taking $s^*(x) = 2s(x)$ and $\theta_2^* = \theta_2/2$ while holding

the other parameters the same. Kneip and Gasser (1988) suggested using the shape-invariant model to find the conditions that guarantee the identifiability in practice. For instance, one can limit the family of shape functions such that it is not closed under aforementioned type of scaling. As originally addressed by Lawton *et al.* (1972), this problem was solved by forcing the range of the shape function to be $[0, 1]$ and some other relevant properties. However, the choice of required conditions will depend on the particular data analysis problem.

A natural extension of the shape-invariant model (2.9) is to introduce a mixed-effects structure on the shift and scale parameters, since these parameters vary from individual to individual. The mixed-effects shape-invariant model (Lindstrom, 1995) is defined as follows,

$$y_{ij} = (\beta_1 + b_{i1}) + (\beta_2 + b_{i2}) \cdot s\left(\frac{x_{ij} - (\beta_3 + b_{i3})}{\beta_4 + b_{i4}}\right) + \epsilon_{ij}, \quad \boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{D}), \qquad (2.10)$$

where s is the shape function, $\epsilon_{ij} \sim N(0, \sigma^2), \boldsymbol{b}_i = (b_{i1}, \cdots, b_{i4})^T$, and $\text{cov}(b_{ik}, \epsilon_{ij}) = 0$ for all i, j and k. A substantial improvement of this model is that the dimension of the parameter space is reduced and does not depend on the number of subjects anymore. In addition, this model inherits all properties of the nonlinear mixed-effects model defined by (2.1) and (2.2).

2.3 Curve registration

In many studies, the outcome of interest is about the response curve, such as its overall shape, the slope at a particular point, or the location of an turning point. The observed curves usually exhibit two types of variability: amplitude and time. For example, consider the sample curves shown in Figure 2.1 (a). All curves have a common trend with two small peaks followed by a large one, but the intensity and the timing of the peaks differ from one curve to another. The variability in amplitude

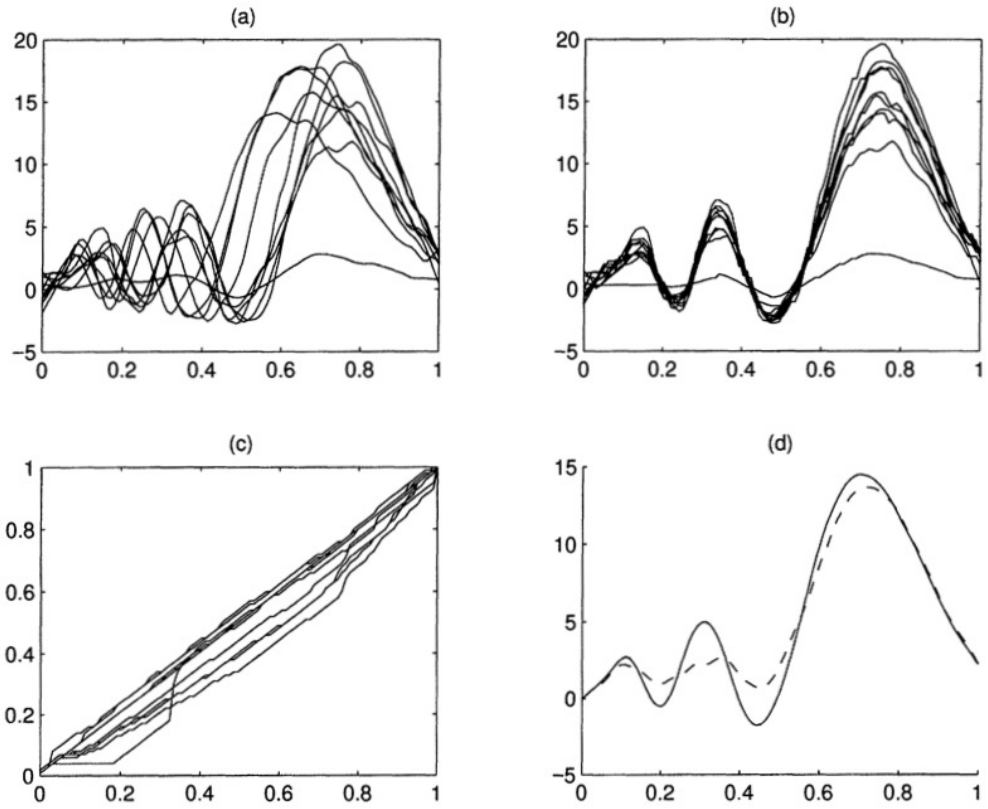

Figure 2.1: Example of curve registration from Wang and Gasser (1999). (a) 10 sample curves; (b) aligned sample curves; (c) the estimated time transformations used for the alignment; (d) the average curves [solid: average of the aligned curves in (b); dashed: average of the sample curves in (a)].

is considered in many analyses, like using a point-wise sample mean to estimate the overall curve. However, the time variability is usually neglected and this can lead to large bias in the estimates. As shown in Figure 2.1 (d), the point-wise sample mean (dashed line) has much less variation and does not resemble any of the observed curves. Thus, it is necessary to take the time variation into account when conducting statistical analysis.

A commonly used technique to handle this issue is curve registration (or time-warping) which has been frequently addressed in the literature (Gasser *et al.*, 1990; Kneip and Gasser, 1992; Kneip and Engel, 1995; Ramsay and Li, 1998; Wand and Gasser, 1999; Gervini and Gasser, 2004). Curve registration is aiming at adjusting for random time distortions by aligning various salient features (such as maxima,

minima, and zero crossings of curves), since the intensity of these salient features of different curves should be compared at their respective time rather than at any fixed time. An example is given in Figure 2.1 (b), where all curves are aligned to have the same timing for peaks and valleys. The average of aligned sample curves, as shown by the solid line in Figure 2.1 (d), resembles more of the common shape and provides a more meaningful structural mean than the unaligned sample curves.

To be more specific, let $\{x_i : \mathbb{R} \to \mathbb{R}\}$ be a sample of curves on $[a, b] \in \mathbb{R}$. The goal of curve registration is to construct a transformation $w_i : [a, b] \to [a, b]$ for each curve such that the aligned curves with values

$$x_i^*(t) = x_i\{w_i(t)\}$$

have identical arguments values for any given features. The functions $\{w_i\}$ are called time-warping functions. Let $\boldsymbol{\tau}_0 = (\tau_{01}, \cdots, \tau_{0p})^T$ be a sequence of the salient features of the population mean curve such that $a < \tau_{01} < \cdots < \tau_{0p} < b$, and $\boldsymbol{\tau}_i = (\tau_{i1}, \cdots, \tau_{ip})^T$ be the corresponding features of the ith curve with $a < \tau_{i1} < \cdots < \tau_{ip} < b$. Ramsay and Silverman (2005) presents the properties of the time-warping functions: (1) $w_i(a) = a$ and $w_i(b) = b$; (2) $w_i(\tau_{0k}) = \tau_{ik}$, for $k = 1, \cdots, p$; (3) w_i is strictly monotonic, i.e., $w_i(t_1) < w_i(t_2)$ for any $t_1 < t_2$. The first property means that the warping functions keep the endpoints fixed. The second one implies that the salient features of the aligned curves each occur at the same time as in the mean function. The third one indicates the warping-functions are invertible so that for the same event the time points before and after warping have a one-to-one mapping.

There are different types of warping functions available. The simplest method is linear interpolation, but the resulting warping functions are not differentiable everywhere. Differentiable families of warping functions are given by penalized B-splines

(Telesca and Inoue, 2008), the smooth monotone family of Ramsay and Li (1998) and monotone interpolating Hermite splines (Gervini and Carter, 2014). Figure 2.1 (c) illustrates (non-smooth) estimated time-warping functions. If the plot is given using the warped time versus the actual time on the vertical and horizontal axes respectively, any estimated time-warping function below the diagonal line indicates the warped time is earlier than the actual time of features, and vice versa.

2.4 Flour-beetle growth data

In evolutionary biology, researchers are particularly interested in patterns of growth. The growth trajectories provide important information on the growth rate, the development time of some events, and the adult fecundity. Consider, for example, the growth curve of the flour-beetle which is a common insect model for population genetic and development studies (Irwin and Carter, 2013). The whole data set consists of 1124 larvae, representing 134 full-sib families nested with 29 half-sib families. Figure 2.2 (a) shows these growth trajectories describing how body masses change over time of larvae from hatching to pupation. For better visualization, we only plotted 100 randomly chosen curves from the whole data set.

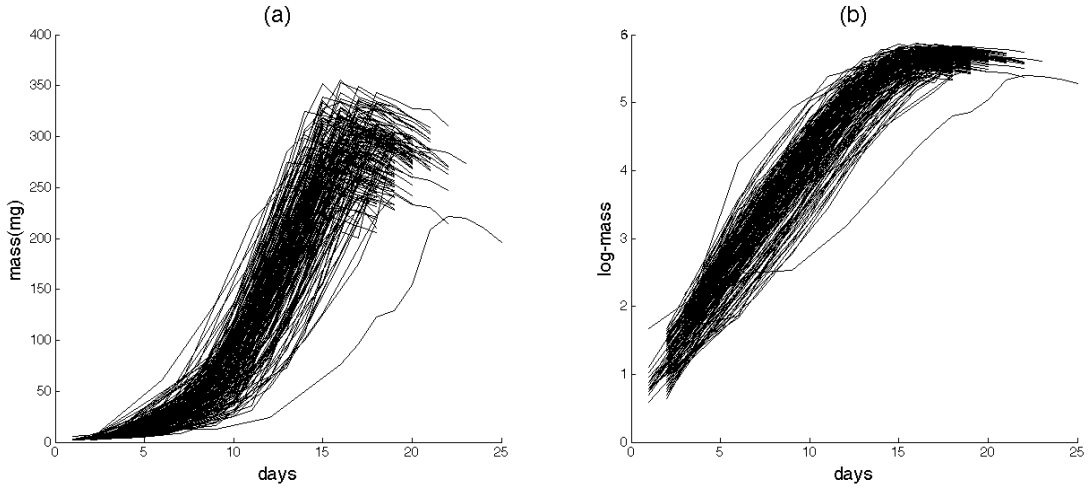

Figure 2.2: Beetle-growth curves. (a) Raw mass trajectories; (b) Trajectories of log-mass.

The mass measurements were collected every 1-4 days during the larval period for each individual and more frequent, usually every day, around the anticipation of pupation period. Therefore, the time points of taking records for different individuals vary from one to another. From Figure 2.2 (a), we see that all curves share a common shape, increasing exponentially at the beginning and then decreasing after a specific time point. These turning points are around day 15 and indicate the timing when larvae stop eating and begin pupating which causes the loss of body mass. This process is caused by hormonal mechanisms and varies among individuals. Since there are two orders of magnitude in the body size between younger and older ages of each flour-beetle, a log-transformation is taken to the mass measurements to stabilize the error variance. Figure 2.2 (b) shows the log-transformed growth trajectories.

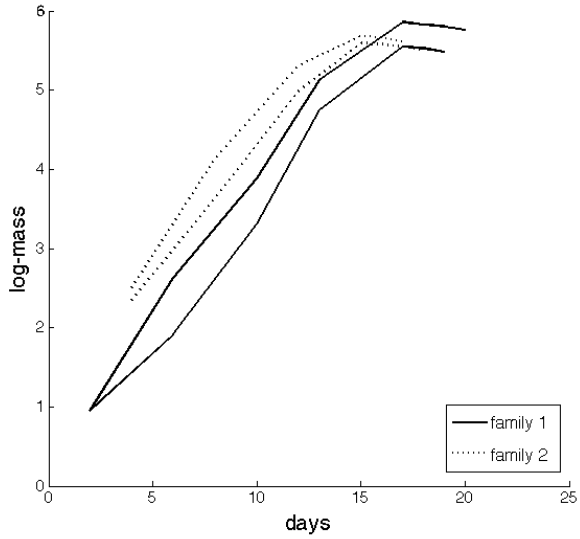

Figure 2.3: Beetle-growth trajectories of two different families.

All log-transformed growth trajectories follow a common piece-wise linear trend, but they still differ in many features, such as the growth rate, the peak of the body size, and the time of the turning point. However, as shown in Figure 2.3, these features share more in common for subjects within the same family because of their genetic correlation. Therefore, it would be important to take the genetic information into

account to increase the accuracy of prediction on the evolution of traits in response to selection. In the last section, we review some useful results of quantitative genetics which gives methods of utilizing the genetic correlation between subjects.

2.5 Some results of quantitative genetics

In evolutionary studies, biologists are dedicated to investigate how phenotypes (or physical traits) change across generations, including both the similarities across all life and the diversity of a particular life. A phenotype can be qualitative (e.g., hair color and blood type) or quantitative (e.g., height and weight). It is the result of interactions between the genes and the environment. In quantitative genetics, the evolution of phenotypes consists of two components: a selection mechanism and genetic variability. The selection mechanism can provide the change of the distribution of the phenotype across generations. For example, birds can easily spot insects of bright colors, and this can cause fewer bright-colored individuals in the next generation. However, without the genetic variation a selection mechanism cannot operate, e.g., if all insects in the population are of the same color, selection does not have any effects because the genetic make-up in the population could not change. Thus, knowledge about the genetic variability of the phenotype is important in determining how the phenotype evolves.

A basic idea in the study of genetic variability is its partitioning into components attributable to different causes. Let z denote the observable phenotype. The decomposition of z is

$$z = g + e, \tag{2.11}$$

where g is called the additive genetic component, and e is the environmental component (Kirkpatrick and Heckman, 1989; Heckman, 2003). It is commonly assumed that the unobservable g and e are uncorrelated. Since all subjects can be raised in

independent but similar environments, the environment effects for all subjects are assumed uncorrelated and identically distributed. However, we cannot make such assumption on the genetic effects, because they may be genetically correlated.

Using the standard genetic arguments (Bürger, 2000), we can obtain the correlation between all kinds of relatives. Some of these relationships are shown in Table (2.1). A good way to describe the correlation structure of the genetic effects

Table 2.1: Correlation of some relationship between relatives

Relationship	**Correlation**
identical twins	1
parent-offspring	0.5
full siblings	0.5
half siblings	0.25
grandparent-grandchild	0.25
uncle/aunt-nephew/niece	0.25
great grandparent-great grandchild	0.125

in a pedigree is to use the *additive genetic relationship matrix*. The additive genetic relationship matrix $\boldsymbol{A}$ is symmetric. The diagonal element a_{ii} is equal to $1+F_i$, where F_i is the inbreeding coefficient of the ith subject; the off-diagonal element a_{ij} equals $r_{ij}\sqrt{a_{ii}a_{jj}}$, where r_{ij} is the coefficient of relationship between subject i and subject j (Wright, 1922).

The cost of computing the matrix $\boldsymbol{A}$ by the path coefficient method (Wright, 1922) is computationally expensive as the number of subjects increases and the structure of the relationship becomes complex. Henderson (1976) proposed a simple and fast recursive method to compute $\boldsymbol{A}$. As an example, suppose that a data set contains six subjects and their relationship is shown in Table 2.2 (Mrode, 2014).

Table 2.2: Pedigree for six subjects.

Subjects	Sire	Dam
3	1	2
4	1	unknown
5	4	3
6	5	2

Using Henderson's method, we obtain the additive genetic relationship matrix

$$\boldsymbol{A} = \begin{pmatrix} 1.00 & 0.00 & 0.50 & 0.50 & 0.50 & 0.25 \\ 0.00 & 1.00 & 0.50 & 0.00 & 0.25 & 0.625 \\ 0.50 & 0.50 & 1.00 & 0.25 & 0.625 & 0.563 \\ 0.50 & 0.00 & 0.25 & 1.00 & 0.625 & 0.313 \\ 0.50 & 0.25 & 0.625 & 0.625 & 1.125 & 0.688 \\ 0.25 & 0.625 & 0.563 & 0.313 & 0.688 & 1.125 \end{pmatrix}.$$

In particular, if the n subjects in the data set are equally related, such as full-siblings or half-siblings, the additive genetic relationship matrix becomes

$$\boldsymbol{A} = \begin{pmatrix} 1 & p & p & \cdots & p \\ p & 1 & p & \cdots & p \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p & p & p & \cdots & 1 \end{pmatrix}_{n \times n},$$

where $p = 0.5$ for full-siblings and $p = 0.25$ for half-siblings.

Next, we explain how the additive relationship matrix $\boldsymbol{A}$ is involved in the covariance of the genetic effects. Consider sample data of n subjects. To formalize the discussion, let $\boldsymbol{z}_i \in \mathbb{R}^k$ be phenotype of the ith individual and $\boldsymbol{z}_i = \boldsymbol{g}_i + \boldsymbol{e}_i$ by (2.11). Denote $\boldsymbol{\Sigma}_g = \mathrm{cov}(\boldsymbol{g}_i)$ the covariance of the additive genetic effects and $\boldsymbol{\Sigma}_e = \mathrm{cov}(\boldsymbol{e}_i)$

the covariance of the environmental effects. Then, define

$$\boldsymbol{z} = \begin{pmatrix} \boldsymbol{z}_1 \\ \boldsymbol{z}_2 \\ \vdots \\ \boldsymbol{z}_n \end{pmatrix}, \quad \boldsymbol{g} = \begin{pmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_n \end{pmatrix}, \quad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix},$$

we have the following decomposition of the covariance matrix of $\boldsymbol{z}$,

$$\begin{aligned}
&\mathrm{cov}(\boldsymbol{z}) = \mathrm{cov}(\boldsymbol{g}) + \mathrm{cov}(\boldsymbol{e}) \\
&= \begin{bmatrix} \sigma(\boldsymbol{g}_1,\boldsymbol{g}_1) & \sigma(\boldsymbol{g}_1,\boldsymbol{g}_2) & \cdots & \sigma(\boldsymbol{g}_1,\boldsymbol{g}_n) \\ \sigma(\boldsymbol{g}_2,\boldsymbol{g}_1) & \sigma(\boldsymbol{g}_2,\boldsymbol{g}_2) & \cdots & \sigma(\boldsymbol{g}_2,\boldsymbol{g}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(\boldsymbol{g}_n,\boldsymbol{g}_1) & \sigma(\boldsymbol{g}_n,\boldsymbol{g}_2) & \cdots & \sigma(\boldsymbol{g}_n,\boldsymbol{g}_n) \end{bmatrix} + \begin{bmatrix} \sigma(\boldsymbol{e}_1,\boldsymbol{e}_1) & \sigma(\boldsymbol{e}_1,\boldsymbol{e}_2) & \cdots & \sigma(\boldsymbol{e}_1,\boldsymbol{e}_n) \\ \sigma(\boldsymbol{e}_2,\boldsymbol{e}_1) & \sigma(\boldsymbol{e}_2,\boldsymbol{e}_2) & \cdots & \sigma(\boldsymbol{e}_2,\boldsymbol{e}_n) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma(\boldsymbol{e}_n,\boldsymbol{e}_1) & \sigma(\boldsymbol{e}_n,\boldsymbol{e}_2) & \cdots & \sigma(\boldsymbol{e}_n,\boldsymbol{e}_n) \end{bmatrix} \\
&= \begin{bmatrix} a_{11}\boldsymbol{\Sigma}_g & a_{12}\boldsymbol{\Sigma}_g & \cdots & a_{1n}\boldsymbol{\Sigma}_g \\ a_{21}\boldsymbol{\Sigma}_g & a_{22}\boldsymbol{\Sigma}_g & \cdots & a_{2n}\boldsymbol{\Sigma}_g \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}\boldsymbol{\Sigma}_g & a_{n2}\boldsymbol{\Sigma}_g & \cdots & a_{nn}\boldsymbol{\Sigma}_g \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Sigma}_e & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_e & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_e \end{bmatrix} \\
&\equiv \boldsymbol{A} \otimes \boldsymbol{\Sigma}_g + \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_e, \qquad (2.12)
\end{aligned}$$

where $\boldsymbol{I}_n$ is the n-dimensional identity matrix. As we can see, the additive genetic relationship matrix $\boldsymbol{A}$ plays a key role in equation (2.12), because it makes the covariance of the unobservable genetic characteristics estimable. In the following chapters, our models are built on this equation when introducing the random genetic and environmental effects. Another thing we should note is that many statistical methods involving $\boldsymbol{A}$ also require the calculation of its inverse in the estimation procedures. A recursive algorithm is also available in Henderson (1976) when $\boldsymbol{A}^{-1}$ is needed.

Chapter 3

Shape-Invariant Model

In this chapter, we develop a shape-invariant model for a special growth pattern which is motivated by the flour-beetle growth curves. We present the model specification of the two-factor model, i.e., considering father and mother factors in the genetic effects, and derive its maximum likelihood estimators in Sections 3.1 and 3.2, respectively. The balanced one-factor four-parameter model, i.e., only considering the father factor in the genetic effects, and its maximum likelihood estimators which are used to establish the asymptotic properties are presented in Section 3.3. In Sections 3.4 and 3.5, we propose a general shape-invariant model and derive the maximum likelihood estimators of the two-factor model.

3.1 Two-factor model specification

Suppose there are n subjects in the data. For each subject i, we observe its responses at m_i time points. Assume that the jth observation of the ith subject follows

$$y_{ij} = f_i(t_{ij}) + \epsilon_{ij},$$

where $\{f_i(t)\}$ are the individual growth curves, $\{\epsilon_{ij}\}$ are random errors. We assume that $\{\epsilon_{ij}\}$ are i.i.d. from $N(0, \sigma^2)$ and they are independent of $\{f_i(t)\}$. Motivated by the log-transformed flour-beetle growth curves, we propose the following four-parameter shape-invariant model

$$f_i(t) := f(\boldsymbol{\phi}_i, t) = \begin{cases} \phi_{i2} + \phi_{i1}(t - \phi_{i4}) & t \leq \phi_{i4}, \\ \phi_{i2} + \phi_{i3}(t - \phi_{i4}) & t \geq \phi_{i4}, \end{cases} \tag{3.1}$$

where $\boldsymbol{\phi}_i = (\phi_{i1}, \phi_{i2}, \phi_{i3}, \phi_{i4})^T$ are the parameters of the ith subject such that ϕ_{i4} is the location of the turning point, ϕ_{i2} represents the response at ϕ_{i4}, and ϕ_{i1} and ϕ_{i3} are the constant growth rates before and after the turning point. The key of the four-parameter model (3.1) is that it not only considers the amplitude variability through ϕ_{i1}, ϕ_{i2}, and ϕ_{i3}, but also takes the time variability into account by ϕ_{i4}. Although this model is of a special growth pattern, it can be easily generalized to many others, such as a joint curve of a spline and an exponential function or a curve with more than two pieces of functions. The choice of functions in the common shape is based on the structure of the curves of interest.

Since most estimation procedures require the derivative of f_i with respect to the parameters to be continuous, we introduce a transition function (Morrell *et al.*, 1995) to smooth out the connection of the two linear functions. The transition function $\mathrm{trn}(x)$ is defined as

$$\mathrm{trn}(x) = \frac{e^{cx}}{1 + e^{cx}}, \tag{3.2}$$

where c is a fixed positive number which controls the speed and smoothness of the transition and satisfies the following conditions: (a) $\lim_{x\to\infty} \mathrm{trn}(x) = 1$, (b) $\lim_{x\to-\infty} \mathrm{trn}(x) = 0$, and (c) $\mathrm{trn}(0) = 0.5$. With the transition function $\mathrm{trn}(x)$,

the proposed shape-invariant model becomes

$$f_i(t) = [\phi_{i2} + \phi_{i1}(t - \phi_{i4})]\text{trn}(\phi_{i4} - t) + [\phi_{i2} + \phi_{i3}(t - \phi_{i4})]\text{trn}(t - \phi_{i4}), \qquad (3.3)$$

and we can see that $f_i(t)$ is a weighted average of the two linear functions. Morrell *et al.* (1995) used $c = 2$ in $\text{trn}(x)$ which provides a fairly rapid transition between the two functions.

The parameter vector $\boldsymbol{\phi}_i$ in the model (3.3) varies from individual to individual. For data from pedigrees, this randomness comes not only from the environment but also from the genetic factor, e.g., beetles from one family may have larger born mass and earlier occurrence of the pupation period than those from another. As we discussed in Section 2.5, the parameters $\boldsymbol{\phi}_i$ can be partitioned into the genetic and environmental random effects. Thus, we assume

$$\boldsymbol{\phi}_i = \boldsymbol{g}_i + \boldsymbol{e}_i,$$
$$\boldsymbol{g}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g),\ \boldsymbol{e}_i \overset{iid}{\sim} N(\boldsymbol{0}, \boldsymbol{\Sigma}_e),$$

where $\boldsymbol{g}_i$ is the genetic random effect, $\boldsymbol{e}_i$ is the environmental effect, and $\{\boldsymbol{g}_i\}$ and $\{\boldsymbol{e}_i\}$ are assumed uncorrelated. Although the genetic effects $\{\boldsymbol{g}_i\}$ are not independent, their covariance structure can be captured using the additive genetic relationship matrix $\boldsymbol{A}$. We write the n individual models as one by letting

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}, \qquad \boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \vdots \\ \boldsymbol{\phi}_n \end{pmatrix}, \qquad \boldsymbol{f}(\boldsymbol{\phi}) = \begin{pmatrix} \boldsymbol{f}_1(\boldsymbol{\phi}_1) \\ \boldsymbol{f}_2(\boldsymbol{\phi}_2) \\ \vdots \\ \boldsymbol{f}_n(\boldsymbol{\phi}_n) \end{pmatrix},$$

where $\boldsymbol{y}_i = (y_{i1}, \cdots, y_{im_i})^T$, $\boldsymbol{f}_i(\boldsymbol{\phi}_i) = [f(\boldsymbol{\phi}_i, t_{i1}), \cdots, f(\boldsymbol{\phi}_i, t_{im_i})]^T$. Then, according

to (2.12), we have

$$\boldsymbol{y}|(\boldsymbol{g},\boldsymbol{e}) \sim N(\boldsymbol{f}(\boldsymbol{\phi}), \sigma^2 \boldsymbol{I}_N), \quad \boldsymbol{\phi} = \boldsymbol{g} + \boldsymbol{e}, \tag{3.4}$$

$$\boldsymbol{g} \sim N(\tilde{\boldsymbol{\mu}}, \boldsymbol{A} \otimes \boldsymbol{\Sigma}_g), \quad \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_e), \tag{3.5}$$

where $N = \sum_{i=1}^{n} m_i$, and

$$\boldsymbol{g} = \begin{pmatrix} \boldsymbol{g}_1 \\ \boldsymbol{g}_2 \\ \vdots \\ \boldsymbol{g}_n \end{pmatrix}, \qquad \boldsymbol{e} = \begin{pmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_n \end{pmatrix}, \qquad \tilde{\boldsymbol{\mu}} = \begin{pmatrix} \boldsymbol{\mu} \\ \boldsymbol{\mu} \\ \vdots \\ \boldsymbol{\mu} \end{pmatrix}.$$

The model defined by (3.4) and (3.5) is a nonlinear mixed-effects model which uses ideas similar to those described in Lindstrom and Bates (1995), but the major difference is the non-independence of subjects due to the genetic effects. The random effects in Lindstrom and Bates (1995) have a diagonal structure in the covariance matrix, while (3.5) involves a more complex covariance matrix through the additive genetic relationship matrix $\boldsymbol{A}$ which complicates the process of parameter estimation.

3.2 Maximum likelihood estimators of the parameters

In this section, we derive the maximum likelihood estimators of the four-parameter shape-invariant model. First, we have the marginal density of $\boldsymbol{y}$ which is defined by

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e}) p(\boldsymbol{g}) p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}. \tag{3.6}$$

This integral does not have an explicit form. To find maximum likelihood estimators for the parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}_g, \boldsymbol{\Sigma}_e$ and σ^2, we use the log-likelihood function

$$\begin{aligned} l &= \ln p(\boldsymbol{y}) \\ &= \ln \int (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\phi})\right]^T\left[\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\phi})\right]\right\} \\ &\quad \times (2\pi)^{-nk/2}|\boldsymbol{A}\otimes\boldsymbol{\Sigma}_g|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{g}-\mathbb{1}_n\otimes\boldsymbol{\mu})^T(\boldsymbol{A}\otimes\boldsymbol{\Sigma}_g)^{-1}(\boldsymbol{g}-\mathbb{1}_n\otimes\boldsymbol{\mu})\right\} \\ &\quad \times (2\pi)^{-nk/2}|\boldsymbol{I}_n\otimes\boldsymbol{\Sigma}_e|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{e}^T(\boldsymbol{I}_n\otimes\boldsymbol{\Sigma}_e)^{-1}\boldsymbol{e}\right\}\, d\boldsymbol{g}\, d\boldsymbol{e}, \end{aligned} \tag{3.7}$$

where k is the dimension of the subject-specific parameter vector $\boldsymbol{\phi}_i$ and $\mathbb{1}_n$ is the $n \times 1$ vector whose elements are all 1. Define the $k \times n$ dimensional matrices $\boldsymbol{G} = (\boldsymbol{g}_1, \cdots, \boldsymbol{g}_n)$, $\boldsymbol{V} = (\boldsymbol{g}_1 - \boldsymbol{\mu}, \cdots, \boldsymbol{g}_n - \boldsymbol{\mu})$, and $\boldsymbol{E} = (\boldsymbol{e}_1, \cdots, \boldsymbol{e}_n)$. Then, we take the partial derivatives of the log-likelihood l with respect to $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_g$, $\boldsymbol{\Sigma}_e$ and σ^2, and obtain

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\mu}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\boldsymbol{\Sigma}_g^{-1}\boldsymbol{G}\boldsymbol{A}^{-1}\mathbb{1}_n - \mathbb{1}_n^T\boldsymbol{A}^{-1}\mathbb{1}_n\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\mu}\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\ \frac{\partial l}{\partial \boldsymbol{\Sigma}_g^{-1}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\frac{n}{2}\boldsymbol{\Sigma}_g - \frac{1}{2}\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{V}^T\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\ \frac{\partial l}{\partial \boldsymbol{\Sigma}_e^{-1}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\frac{n}{2}\boldsymbol{\Sigma}_e - \frac{1}{2}\boldsymbol{E}\boldsymbol{E}^T\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\ \frac{\partial l}{\partial \sigma^2} &= \frac{1}{p(\boldsymbol{y})}\int \left\{-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\left[\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\phi})\right]^T\left[\boldsymbol{y}-\boldsymbol{f}(\boldsymbol{\phi})\right]\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \end{aligned}$$

Note that $p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})/p(\boldsymbol{y})$ is the posterior density of the parameters $\boldsymbol{g}$ and $\boldsymbol{e}$ given $\boldsymbol{y}$. Therefore, setting the equations to be zero and solving for the parameters, we have the following maximum likelihood estimating equations:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \frac{1}{(\mathbb{1}_n^T\boldsymbol{A}^{-1}\mathbb{1}_n)}\mathbb{E}\{\boldsymbol{G}\boldsymbol{A}^{-1}\mathbb{1}_n|\boldsymbol{y}\}, \\ \hat{\boldsymbol{\Sigma}}_g &= \frac{1}{n}\mathbb{E}\{\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{V}^T|\boldsymbol{y}\}, \\ \hat{\boldsymbol{\Sigma}}_e &= \frac{1}{n}\mathbb{E}\{\boldsymbol{E}\boldsymbol{E}^T|\boldsymbol{y}\}, \end{aligned}$$

$$\hat{\sigma}^2 = \frac{1}{N}\mathbb{E}\{[\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\phi})]^T [\boldsymbol{y} - \boldsymbol{f}(\boldsymbol{\phi})] \,|\boldsymbol{y}\}.$$

The difficulty for calculating these estimators is caused by the integrals in the formulae, since there is no closed form of the density $p(\boldsymbol{y})$. As mentioned in Section 2.1, we use Laplace's method to approximate the marginal density $p(\boldsymbol{y})$ and the expectations in the calculation of these maximum likelihood estimators.

3.3 Balanced one-factor model

Suppose that we have I fathers and each father has J descendants. Assume that the observations of all subjects are obtained at the same time grids $(t_1, \cdots, t_m)$. Thus, a total of $N = I \times J \times m$ observations are available. Let y_{ijk} denote the kth observation of the jth descendant of father i, and assume it follows the four-parameter shape-invariant model

$$y_{ijk} = f_{ij}(t_{ijk}) + \epsilon_{ijk},$$
$$f_{ij}(t) = [\phi_{ij2} + \phi_{ij1}(t - \phi_{ij4})]\text{trn}(\phi_{ij4} - t) + [\phi_{ij2} + \phi_{ij3}(t - \phi_{ij4})]\text{trn}(t - \phi_{ij4}),$$

where $f_{ij}(t)$ is the growth curve of the jth descendant of father i, $\{\epsilon_{ijk}\}$ are i.i.d. $N(0, \sigma^2)$ random errors, and $\{f_{ij}(t_{ijk})\}$ are independent of $\{\epsilon_{ijk}\}$. We propose the one-factor design of this model as follows. Assume that the subject-specific parameters $\boldsymbol{\phi}_{ij} = (\phi_{ij1}, \cdots, \phi_{ij4})^T$ have a structure

$$\boldsymbol{\phi}_{ij} = \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij},$$

where $\{\boldsymbol{\alpha}_i\}$ are i.i.d. from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$ representing the father effects, $\{\boldsymbol{e}_{ij}\}$ are i.i.d. from $N(\boldsymbol{0}, \boldsymbol{\Sigma}_e)$ indicating the environmental effects and the genetic contribution of

the mothers. We also assume $\{\boldsymbol{\alpha}_i\}$ and $\{\boldsymbol{e}_{ij}\}$ are uncorrelated. Then

$$\boldsymbol{y}_{ij}|(\boldsymbol{\alpha}_i, \boldsymbol{e}_{ij}) \sim N(\boldsymbol{f}(\boldsymbol{\phi}_{ij}), \sigma^2 \boldsymbol{I}_m), \quad \boldsymbol{\phi}_{ij} = \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij}, \tag{3.8}$$

where $\boldsymbol{f}(\boldsymbol{\phi}_{ij}) = [f_{ij}(t_{ij1}, \boldsymbol{\phi}_{ij}), \cdots, f_{ij}(t_{ijm}, \boldsymbol{\phi}_{ij})]^T$. Denote $p(\boldsymbol{y}_{i\cdot}) = p(\boldsymbol{y}_{i1}, \cdots, \boldsymbol{y}_{iJ})$. Since subjects from different fathers are independent, the log-likelihood function is

$$l = \sum_{i=1}^{I} \log\, p(\boldsymbol{y}_{i\cdot}).$$

Under all assumptions, $\{\boldsymbol{y}_{ij}\}$ are identically distributed and they are conditionally independent given $(\boldsymbol{\alpha}_i, \boldsymbol{e}_{ij})$. Therefore, the component $p(\boldsymbol{y}_{i\cdot})$ in l is

$$p(\boldsymbol{y}_{i\cdot}) = \int \left\{ \prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{ij}) \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{e}_{ij}) \right\} p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}).$$

Similar to Section 3.2, the maximum likelihood estimators of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_g$, $\boldsymbol{\Sigma}_e$ and σ^2 can be obtained by taking the partial derivatives of l with respect to the corresponding parameters, and they are:

$$\hat{\boldsymbol{\mu}} = \frac{1}{I} \sum_{i=1}^{I} \mathbb{E}\{\boldsymbol{\alpha}_i|\boldsymbol{y}_{i\cdot}\},$$

$$\hat{\boldsymbol{\Sigma}}_g = \frac{1}{I} \sum_{i=1}^{I} \mathbb{E}\{(\boldsymbol{\alpha}_i - \boldsymbol{\mu})(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T \big| \boldsymbol{y}_{i\cdot}\},$$

$$\hat{\boldsymbol{\Sigma}}_e = \frac{1}{n} \sum_{i=1}^{I} \mathbb{E}\left\{ \sum_{j=1}^{J} \boldsymbol{e}_{ij} \boldsymbol{e}_{ij}^T | \boldsymbol{y}_{i\cdot} \right\},$$

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{i=1}^{I} \mathbb{E}\left\{ \sum_{j=1}^{J} [\boldsymbol{y}_{ij} - \boldsymbol{f}(\boldsymbol{\phi}_{ij})]^T [\boldsymbol{y}_{ij} - \boldsymbol{f}(\boldsymbol{\phi}_{ij})] \big| \boldsymbol{y}_{i\cdot} \right\}.$$

3.4 General self-modeling regression

Now we propose a more general shape-invariant model than the four-parameter model discussed above. Suppose there are n subjects in the data set. Assume that the jth observation of the i subject follows the shape-invariant model

$$y_{ij} = a_i f(v_i(t_{ij})) + \epsilon_{ij}, \quad \text{for } j = 1, \cdots, m_i, \tag{3.9}$$

where a_i is the amplitude factor, $f : T \to \mathbb{R}$ is the common shape function, $\{v_i : T \to T\}$ are the inverse functions of the warping functions $\{w_i\}$, and $\{\epsilon_{ij}\}$ are random errors which are i.i.d. from $N(0, \sigma^2)$. From the expression of model (3.9), we see that it separates the amplitude variability and the time variability explicitly through $\{a_i\}$ and $\{v_i\}$: the aligned curves

$$y_i \circ w_i(t) = a_i f\{v_i(w_i(t))\} = a_i f(t), \quad \text{for } i = 1, \cdots, n,$$

only exhibit amplitude variation, which makes the comparison on the intensity of salient features of different curves more sensible. It is also worth noting that we must take model identifiability into account whenever using self-modeling regression. In fact, without further conditions model (3.9) is not identifiable. Consider the following situations:

Situation 1: let h be a monotone increasing function, then

$$y_i(t) = a_i f(v_i(t)) = a_i f(h^{-1}(h(v_i(t))) = a_i f^*(v_i^*(t)),$$

$$\text{where } v_i^*(t) = h(v_i(t)) \text{ and } f^*(t) = f(h^{-1}(t)).$$

Situation 2: let $c \neq 0$, then

$$y_i(t) = a_i f(v_i(t)) = \frac{a_i}{c}\{cf(v_i(t)\} = \tilde{a}_i \tilde{f}(v_i(t)),$$

where $\tilde{a}_i = \frac{a_i}{c}$ and $\tilde{f}(t) = cf(t)$.

Thus, $\{a_i, v_i^*(t), f^*(t)\}$ and $\{\tilde{a}_i, v_i(t), f^*(t)\}$ define the same model as $\{a_i, v_i(t), f(t)\}$ does. This means model (3.9) is not identifiable, since there exist different functions and parameter vectors that generate the same model. Therefore, it is necessary to make some further assumptions on the amplitude factor and the warping functions. For identifiability (which is proved in Section 4.3), we assume that $a_i \neq 0$, $\bar{a} = 1$, and $\bar{w}(t) = t$ for all t. Although the assumptions on the amplitude variability make the model rather limited, it is appropriate for many real data sets.

For the shape function in model (3.9), there are a number of choices available, such as linear splines (Lawton *et al.*, 1972) and cubic splines (Lindstrom, 1995). Define a knot sequence of the shape function as $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_p)^T$ on the interval $T = [a, b]$ such that $a < \tau_1 < \tau_2 < \cdots < \tau_p < b$ (the knot sequence is usually a sequence of salient features of the curve). In general, a spline function of order r is a linear combination of the basis functions $\{1, x, \cdots, x^{r-1}, (x-\tau_1)_+^{r-1}, \cdots, (x-\tau_p)_+^{r-1}\}$. This basis system is called the truncated power basis. A disadvantage of it is lack of numerical stability. To overcome this problem, we use the equivalent B-spline basis (see Eubank (1999) for the definition), denoted by $b_1(x), \cdots, b_m(x)$, where $m = p + r$. A direct derivation of B-spline from the truncated power basis can be found in de Boor (1978). The B-spline representation of the shape function f on $[a, b]$ is

$$f(t) = \sum_{k=1}^{m} d_k b_k(t) := \boldsymbol{d}^T \boldsymbol{B}(t),$$

where $\boldsymbol{d}$ is a vector of the B-spline coefficients and $\boldsymbol{B}(t)$ is the B-spline basis introduced by the knot sequence $\boldsymbol{\tau}$.

Next, we consider the subject-specific parameters. Let $\boldsymbol{\tau}_i = (\tau_{i1}, \tau_{i2}, \cdots, \tau_{ip})^T$ is a strictly increasing knot sequence of the ith subject on (a, b) which represents

the locations of the salient features. Then, the parameters of each curve consist of the amplitude parameter a_i and the knot sequence $\boldsymbol{\tau}_i$. These parameters vary from subject to subject and produce the randomness of each individual curve. Note that $\{\boldsymbol{\tau}_i\}$ are increasing and constrained on $[a, b]$, it is difficult to assume an appropriate distribution on them. The Jupp transformation (Jupp, 1978) is widely applied in the estimation of series of increasing parameters. Define $\boldsymbol{\theta}_i = \text{Jupp}(\boldsymbol{\tau}_i)$, where

$$\begin{aligned}
\theta_{i1} &= \log\left(\frac{\tau_{i2} - \tau_{i1}}{\tau_{i1} - a}\right), \\
\theta_{ik} &= \log\left(\frac{\tau_{i,k+1} - \tau_{ik}}{\tau_{ik} - \tau_{i,k-1}}\right), \quad \text{for } k = 2, \cdots, p-1, \\
\theta_{ip} &= \log\left(\frac{b - \tau_{ip}}{\tau_{ip} - \tau_{i,p-1}}\right).
\end{aligned}$$

This transformation is invertible which makes $\boldsymbol{\theta}_i$ uniquely defining the location of the the knot sequence $\boldsymbol{\tau}_i$. Using this transformation, $\{\boldsymbol{\theta}_i\}$ are not constrained any more. Therefore, we define $\boldsymbol{\phi}_i = (a_i, \boldsymbol{\theta}_i)^T \in \mathbb{R}^{p+1}$ the subject-specific parameters and assume that

$$\boldsymbol{\phi}_i = \boldsymbol{g}_i + \boldsymbol{e}_i,$$

where $\boldsymbol{g}_i$ is the vector of the genetic effects and $\boldsymbol{e}_i$ is the vector of the environmental effects. We assume that $\{\boldsymbol{g}_i\}$ are identically distributed from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$, $\{\boldsymbol{e}_i\}$ are i.i.d. from $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$, and the genetic and environmental effects are independent. By (2.12), $\boldsymbol{g} \sim N(\tilde{\boldsymbol{\mu}}, \boldsymbol{A} \otimes \boldsymbol{\Sigma}_g)$ with $\boldsymbol{g} = (\boldsymbol{g}_1^T, \cdots, \boldsymbol{g}_n^T)^T$, $\tilde{\boldsymbol{\mu}} = (\boldsymbol{\mu}^T, \cdots, \boldsymbol{\mu}^T)^T$, and $\boldsymbol{A}$ the additive genetic relationship matrix. Next, we write the n individual models as one by defining

$$\boldsymbol{y}_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{im_i} \end{pmatrix}, \qquad \tilde{\boldsymbol{f}}_i(\boldsymbol{\phi}_i) = \begin{pmatrix} a_i f\{v_i(t_{i1}, \boldsymbol{\theta}_i)\} \\ a_i f\{v_i(t_{i2}, \boldsymbol{\theta}_i)\} \\ \vdots \\ a_i f\{v_i(t_{im_i}, \boldsymbol{\theta}_i)\} \end{pmatrix}.$$

Then, let

$$\boldsymbol{y} = \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_n \end{pmatrix}, \qquad \boldsymbol{\phi} = \begin{pmatrix} \boldsymbol{\phi}_1 \\ \boldsymbol{\phi}_2 \\ \vdots \\ \boldsymbol{\phi}_n \end{pmatrix}, \qquad \tilde{\boldsymbol{f}}(\boldsymbol{\phi}) = \begin{pmatrix} \tilde{\boldsymbol{f}}_1(\boldsymbol{\phi}_1) \\ \tilde{\boldsymbol{f}}_2(\boldsymbol{\phi}_2) \\ \vdots \\ \tilde{\boldsymbol{f}}_n(\boldsymbol{\phi}_n) \end{pmatrix},$$

we obtain

$$\boldsymbol{y}|(\boldsymbol{g}, \boldsymbol{e}) \sim N(\tilde{\boldsymbol{f}}(\boldsymbol{\phi}), \sigma^2 \boldsymbol{I}_N), \quad \boldsymbol{\phi} = \boldsymbol{g} + \boldsymbol{e}, \tag{3.10}$$

$$\boldsymbol{g} \sim N(\tilde{\boldsymbol{\mu}}, \boldsymbol{A} \otimes \boldsymbol{\Sigma}_g), \quad \boldsymbol{e} \sim N(\boldsymbol{0}, \boldsymbol{I}_n \otimes \boldsymbol{\Sigma}_e), \quad \text{and} \tag{3.11}$$

$$f(t) = \boldsymbol{d}^T \boldsymbol{B}(t), \tag{3.12}$$

where $N = \sum_{i=1}^{n} m_i$ is the total number of observations in the data set.

In this general shape-invariant model, we are interested in the estimation of the amplitude factor and the locations of the salient features. The variability of these parameters is decomposed into the genetic and environmental variabilities, and warping functions are used to address the time variability. For the warping functions $\{w_i\}$, we recommend to use the family of monotone interpolating cubic Hermite splines. The coefficients of this spline family can be directly related to features of the sample curves and the model identifiability has been proved (Gervini and Carter, 2014). Another difference from the four-parameter shape-invariant model is that the use of cubic splines in the shape function provides more flexibility to various growth patterns.

3.5 Maximum likelihood estimators of the general model

The parameters that need to be estimated are $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_g$, $\boldsymbol{\Sigma}_e$, σ^2, and $\boldsymbol{d}$. The marginal density function of $\boldsymbol{y}$ is the same as that defined by (3.6),

$$p(\boldsymbol{y}) = \int p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e},$$

and the log-likelihood function is

$$\begin{aligned} l = &\ln \int (2\pi\sigma^2)^{-N/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{y}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi})\right]^T\left[\boldsymbol{y}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi})\right]\right\} \\ &\times (2\pi)^{-n(p+1)/2}|\boldsymbol{A}\otimes\boldsymbol{\Sigma}_g|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{g}-\mathbb{1}_n\otimes\boldsymbol{\mu})^T(\boldsymbol{A}\otimes\boldsymbol{\Sigma}_g)^{-1}(\boldsymbol{g}-\mathbb{1}_n\otimes\boldsymbol{\mu})\right\} \\ &\times (2\pi)^{-n(p+1)/2}|\boldsymbol{I}_n\otimes\boldsymbol{\Sigma}_e|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{e}^T(\boldsymbol{I}_n\otimes\boldsymbol{\Sigma}_e)^{-1}\boldsymbol{e}\right\}\, d\boldsymbol{g}\, d\boldsymbol{e}. \end{aligned}$$

Define the $(p+1)\times n$ dimensional matrices $\boldsymbol{G}=(\boldsymbol{g}_1,\cdots,\boldsymbol{g}_n)$, $\boldsymbol{V}=(\boldsymbol{g}_1-\boldsymbol{\mu},\cdots,\boldsymbol{g}_n-\boldsymbol{\mu})$, and $\boldsymbol{E}=(\boldsymbol{e}_1,\cdots,\boldsymbol{e}_n)$. Taking the partial derivatives of l with respect to the parameters, we obtain

$$\begin{aligned}
\frac{\partial l}{\partial \boldsymbol{\Sigma}_g^{-1}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\boldsymbol{\Sigma}_g^{-1}\boldsymbol{G}\boldsymbol{A}^{-1}\mathbb{1}_n - \mathbb{1}_n^T\boldsymbol{A}^{-1}\mathbb{1}_n\boldsymbol{\Sigma}_g^{-1}\boldsymbol{\mu}\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\
\frac{\partial l}{\partial \boldsymbol{\Sigma}_g^{-1}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\frac{n}{2}\boldsymbol{\Sigma}_g - \frac{1}{2}\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{V}^T\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\
\frac{\partial l}{\partial \boldsymbol{\Sigma}_e^{-1}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\frac{n}{2}\boldsymbol{\Sigma}_e - \frac{1}{2}\boldsymbol{E}\boldsymbol{E}^T\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\
\frac{\partial l}{\partial \sigma^2} &= \frac{1}{p(\boldsymbol{y})}\int \left\{-\frac{N}{2\sigma^2} + \frac{1}{2\sigma^4}\left[\boldsymbol{y}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi})\right]^T\left[\boldsymbol{y}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi})\right]\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e}, \\
\frac{\partial l}{\partial \boldsymbol{d}} &= \frac{1}{p(\boldsymbol{y})}\int \left\{\frac{1}{\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\left[y_{ij}-a_i\boldsymbol{d}\boldsymbol{B}(t_{ij},\boldsymbol{\theta}_i)\right]\left[a_i\boldsymbol{B}(t_{ij},\boldsymbol{\theta}_i)\right]\right\} p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})\, d\boldsymbol{g}\, d\boldsymbol{e},
\end{aligned}$$

Note that $p(\boldsymbol{y}|\boldsymbol{g},\boldsymbol{e})p(\boldsymbol{g})p(\boldsymbol{e})/p(\boldsymbol{y})$ is the posterior density of the parameter $\boldsymbol{g}$ and $\boldsymbol{e}$ given $\boldsymbol{y}$. Therefore, setting the equations to be zero and solving for the parameters, we obtain the maximum likelihood estimators as follows:

$$\hat{\boldsymbol{\mu}} = \frac{1}{(\mathbb{1}_n^T \boldsymbol{A}^{-1}\mathbb{1}_n)}\mathbb{E}\{\boldsymbol{G}\boldsymbol{A}^{-1}\mathbb{1}_n|\boldsymbol{y}\},$$

$$\hat{\boldsymbol{\Sigma}}_g = \frac{1}{n}\mathbb{E}\{\boldsymbol{V}\boldsymbol{A}^{-1}\boldsymbol{V}^T|\boldsymbol{y}\},$$

$$\hat{\boldsymbol{\Sigma}}_e = \frac{1}{n}\mathbb{E}\{\boldsymbol{E}\boldsymbol{E}^T|\boldsymbol{y}\},$$

$$\hat{\sigma}^2 = \frac{1}{N}\mathbb{E}\left\{\left[\boldsymbol{y}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi})\right]^T\left[\boldsymbol{y}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi})\right]\Big|\boldsymbol{y}\right\},$$

$$\hat{\boldsymbol{d}} = \left[\sum_{i=1}^{n}\mathbb{E}\left\{a_i^2\sum_{j=1}^{m_i}\boldsymbol{B}(t_{ij},\boldsymbol{\theta}_i)\boldsymbol{B}^T(t_{ij},\boldsymbol{\theta}_i)\big|\boldsymbol{y}_i\right\}\right]^{-1}\times\left[\sum_{i=1}^{n}\mathbb{E}\left\{a_i\sum_{j=1}^{m_i}\boldsymbol{B}(t_{ij},\boldsymbol{\theta}_i)y_{ij}\big|\boldsymbol{y}_i\right\}\right]$$

As before, the calculation of these maximum likelihood estimators involves approximations to the marginal density $p(\boldsymbol{y})$ and the expectations which can be obtained by Laplace's method.

Chapter 4

Asymptotics

In this chapter we prove, in detail, consistency and asymptotic normality of the maximum likelihood estimators for the four-parameter shape-invariant model, and indicate the necessary modifications that would extend the results to the general model.

4.1 Consistency

Based on the model assumptions in Section 3.3, $\{\boldsymbol{y}_{i\cdot}\}$ are independently and identically distributed from the following distribution

$$\begin{aligned} p(\boldsymbol{y}_{i\cdot}) &= \int p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p(\boldsymbol{\alpha}_i) p(\boldsymbol{e}_{i\cdot})\, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) \\ &= \int (2\pi\sigma^2)^{-Jm/2} \exp\left\{ -\frac{1}{2\sigma^2} \left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T \left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right] \right\} \\ &\quad \times (2\pi)^{-1/2} |\boldsymbol{\Sigma}_g|^{-1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{\alpha}_i - \boldsymbol{\mu}) \right\} \\ &\quad \times (2\pi)^{-Jk/2} |\boldsymbol{\Sigma}_e|^{-J/2} \exp\left\{ -\frac{1}{2} \boldsymbol{e}_{i\cdot}^T (\boldsymbol{I}_J \otimes \boldsymbol{\Sigma}_e)^{-1} \boldsymbol{e}_{i\cdot} \right\} d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}), \end{aligned} \tag{4.1}$$

where $\boldsymbol{e}_{i\cdot} = (\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ})$. Let $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ be the row vectors containing the unique elements of the covariance matrices $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_e$, respectively. Then, define

$$\boldsymbol{\theta} = (\boldsymbol{\mu}^T, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \sigma^2)$$

as the unknown parameter vector. We shall show the consistency of the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ by verifying that the balanced one-factor four-parameter model satisfies the assumptions of Wald's consistency theorem (Wald, 1949). Let $X_1, \cdots, X_n$ be a sequence of independently and identically distributed random variables. For any parameter point θ in the parameter space Θ, let $F(x, \theta)$ denote the corresponding cumulative distribution function of X_i. If $F(x, \theta)$ is absolutely continuous, $p(x, \theta)$ denotes the density at x. If $F(x, \theta)$ is discrete, $p(x, \theta)$ is equal to the probability that $X_i = x$.

Assumption 1. *Let $F(x, \theta)$ denote the corresponding cumulative distribution function of the random variables X_i, then $F(x, \theta)$ is either discrete for all θ or is absolutely continuous for all θ.*

Proof. This is satisfied because the probability density function $p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})$ exists under our model specification. ■

Assumption 2. *If $\lim_{n\to\infty} \theta_n = \theta$, then $\lim_{n\to\infty} p(x, \theta_n) = p(x, \theta)$ for all x except perhaps on a set which may depend on the limit point θ (but not on the sequence $\{\theta_n\}$) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point θ_0.*

Proof. This is immediate given the continuity of $p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})$ in $\boldsymbol{\theta}$. ■

Assumption 3. *If $\lim_{k\to\infty} |\theta_k| = \infty$, then $\lim_{k\to\infty} p(x, \theta_k) = 0$ for any x except on a fixed set (independent of the sequence $\{\theta_k\}$) whose probability is zero according to the true parameter point θ_0.*

Proof. If any elements of $\boldsymbol{\theta}$ tends to be infinity, the corresponding density function $p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})$, $p(\boldsymbol{\alpha}_i)$, and $p(\boldsymbol{e}_{i\cdot})$ tends to be zero. This implies the marginal density $p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})$ tends to be zero, which shows the assumption. ■

Assumption 4. *For the true parameter point θ_0 we have*

$$\int |\log p(x, \theta_0)| \, dF(x, \theta_0) < \infty.$$

Proof. Without ambiguity, we write $p(\boldsymbol{y}_{i\cdot}) := p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)$. Then,

$$p(\boldsymbol{y}_{i\cdot}) = \int p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p(\boldsymbol{\alpha}_i) p(\boldsymbol{e}_{i\cdot}) \, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) = \mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}}[p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})].$$

Since $-\log(\cdot)$ is a convex function, by Jensen's inequality, we have

$$-\log p(\boldsymbol{y}_{i\cdot}) = -\log \mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}}[p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})] \leq \mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}}[-\log p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})] := M(\boldsymbol{y}_{i\cdot}).$$

Note that

$$\begin{aligned} -\log p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) =& C + \frac{1}{2\sigma^2} \left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T \left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right], \\ =& C + \frac{1}{2\sigma^2} \left[\boldsymbol{y}_{i\cdot}^T \boldsymbol{y}_{i\cdot} - 2\boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})^T \boldsymbol{y}_{i\cdot} + \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})^T \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right], \end{aligned}$$

where $C = (Jm/2)\log(2\pi\sigma^2)$. Then

$$\begin{aligned} M(\boldsymbol{y}_{i\cdot}) =& \mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}}[-\log p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})] \\ =& \mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}} \left[C + \frac{1}{2\sigma^2} \left\{\boldsymbol{y}_{i\cdot}^T \boldsymbol{y}_{i\cdot} - 2\boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})^T \boldsymbol{y}_{i\cdot} + \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})^T \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right\}\right] \\ =& C + \frac{1}{2\sigma^2} \left\{\boldsymbol{y}_{i\cdot}^T \boldsymbol{y}_{i\cdot} - 2\mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}}^T[\boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})] \boldsymbol{y}_{i\cdot} + \mathbb{E}_{\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}}[\boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})^T \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})]\right\}. \end{aligned}$$

Thus, $M(\boldsymbol{y}_{i\cdot})$ is a quadratic expression in $\boldsymbol{y}_{i\cdot}$. By the law of total expectation,

$$\begin{aligned}\mathbb{E}_{\boldsymbol{y}_{i\cdot}}[M(\boldsymbol{y}_{i\cdot})] =& \mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}\left[\mathbb{E}_{\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}\{M(\boldsymbol{y}_{i\cdot})|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}\}\right] \\ =& \mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}\left[\mathbb{E}_{\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}\left\{C+1/(2\sigma^2)\left(\boldsymbol{y}_{i\cdot}^T\boldsymbol{y}_{i\cdot}-2\mathbb{E}^T_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}[\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})]\boldsymbol{y}_{i\cdot}\right.\right.\right. \\ & \left.\left.\left.+\mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}[\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})^T\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})]\right)\Big|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}\right\}\right].\end{aligned}$$

Since $p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})$ is normally distributed, $\mathbb{E}_{\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}[h(\boldsymbol{y}_{i\cdot})|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}]$ is finite for any polynomial function $h(\boldsymbol{y}_{i\cdot})$. The elements of $\mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}[\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})]$ and $\mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}[\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})^T\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})]$ are all finite, because the shape function is bounded on the closed interval $T\in\mathbb{R}$. Therefore, $\mathbb{E}[M(\boldsymbol{y}_{i\cdot})]$ is finite, which implies $\mathbb{E}[-\log p(\boldsymbol{y}_{i\cdot})]$ has an upper bound, or equivalently, $\mathbb{E}[\log p(\boldsymbol{y}_{i\cdot})]$ has a lower bound.

On the other hand, $p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})$ is a normal density which is bounded by $(2\pi\sigma^2)^{-Jm/2}$, because

$$\exp\{-\frac{1}{2\sigma^2}[\boldsymbol{y}_{i\cdot}-\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})]^T[\boldsymbol{y}_{i\cdot}-\boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})]\}\le 1.$$

Thus,

$$\begin{aligned}\mathbb{E}[\log p(\boldsymbol{y}_{i\cdot})] =& \mathbb{E}_{\boldsymbol{y}_{i\cdot}}[\log \mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}\{p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})\}] \\ \le& \mathbb{E}_{\boldsymbol{y}_{i\cdot}}[\log \mathbb{E}_{\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}}\{(2\pi\sigma^2)^{-Jm/2}\}] \\ =& -\frac{Jm}{2}\log(2\pi\sigma^2)<\infty,\end{aligned}$$

which provides an upper bound of $\mathbb{E}[\log p(\boldsymbol{y}_{i\cdot})]$. Therefore, $\int|\log p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}_0)|\,dF(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}_0)=E_{\boldsymbol{y}_{i\cdot}}[|\log p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}_0)|]<\infty$. ■

For any θ and for any positive value ρ, let $p(x,\theta,\rho)$ be the supremum of $p(x,\theta')$ with respect to θ' when $|\theta-\theta'|\le\rho$. For any positive r, let $\psi(x,r)$ be the supremum of $p(x,\theta)$ with respect to θ when $|\theta|>r$. Furthermore, let $p^*(x,\theta,\rho)=p(x,\theta,\rho)$ when $p(x,\theta,\rho)>1$, and $=1$ otherwise. Similarly, let $\psi^*(x,r)=\psi(x,r)$ when $\psi(x,r)>1$, and $=1$ otherwise.

Assumption 5. *For sufficiently small ρ and for sufficiently large r, the expected values*

$$\int \log p^*(x,\theta,\rho)\, dF(x,\theta_0) \;\; and \;\; \int \log \psi^*(x,r)\, dF(x,\theta_0)$$

are finite.

Proof. For $\|\boldsymbol{\theta}\| > r > 0$

$$\begin{aligned}
p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}) &= \int p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\, d(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}) \\
&= \int (2\pi\sigma^2)^{-Jm/2} \exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})\right]^T \left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})\right]\right\} \\
&\quad \times p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\, d(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}) \\
&\leq \int (2\pi\sigma^2)^{-Jm/2} p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\, d(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}) \\
&= (2\pi\sigma^2)^{-Jm/2} \\
&\leq (2\pi r)^{-Jm/2} := M_r,
\end{aligned}$$

and

$$\mathbb{E}_{\boldsymbol{\theta}_0}\{\log M_r\} = -\frac{Jm}{2}\log(2\pi) - \frac{Jm}{2}\log r < \infty,$$

therefore,

$$\mathbb{E}_{\boldsymbol{\theta}_0}\{\log \psi^*(\boldsymbol{y}, r)\} \leq \mathbb{E}_{\boldsymbol{\theta}_0}\{\log M_r\} < \infty.$$

Next, we show that $\int \log p^*(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta},\rho)\, dF(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}_0) < \infty$. First, note that $E_{\boldsymbol{\theta}_0}[\log p^*(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta},\rho)]$ is finite if and only if $E_{\boldsymbol{\theta}_0}[|\log p^*(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta},\rho)|] < \infty$. For any $\rho > 0$, $\{\boldsymbol{\theta}' : \|\boldsymbol{\theta} - \boldsymbol{\theta}'\| \leq \rho\}$ is closed. Since $\log p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, by the Mean Value Theorem, we have

$$\log p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}') - \log p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta}) = D(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta} + \xi(\boldsymbol{y}_{i\cdot}))^T(\boldsymbol{\theta}' - \boldsymbol{\theta}),$$

with $\|\xi(\boldsymbol{y}_{i\cdot})\| < \|\boldsymbol{\theta}' - \boldsymbol{\theta}\|$, where $D(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta})$ is the gradient of $\log p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta})$ with respect

to $\boldsymbol{\theta}$. Then,

$$|\log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}') - \log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})| \leq \|D(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta} + \xi(\boldsymbol{y}_{i\cdot}))^T\| \cdot \|\boldsymbol{\theta}' - \boldsymbol{\theta}\|.$$

Therefore, for $\|\boldsymbol{\theta}' - \boldsymbol{\theta}\| < \rho$,

$$|\log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}, \rho)| \leq \|D(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta} + \xi(\boldsymbol{y}_{i\cdot}))^T\| \cdot \rho + |\log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})|.$$

Since $\mathbb{E}_{\boldsymbol{\theta}_0}\{|\log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})|\}$ is finite (by Assumption 4), it suffices to show $\mathbb{E}_{\boldsymbol{\theta}_0}\{|D(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta} + \xi(\boldsymbol{y}_{i\cdot}))|\}$ is finite. Note that $p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}) = \int p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})$, then

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})}{\partial \theta_j} =& \frac{\int \frac{\partial}{\partial \theta_j}\{p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\}\, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})}{p(\boldsymbol{y}, \boldsymbol{\theta})} \\ =& \frac{\int \frac{\partial}{\partial \theta_j} \log\{p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\} p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})}{p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})},\end{aligned}$$

Thus, it remains to show that $|\frac{\partial}{\partial \theta_j} \log\{p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\}|$ is bounded. Recall that

$$\begin{aligned}(*) :=& \log\{p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}) p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i) p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})\} \\ =& -\frac{Jm}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right] \\ & - \frac{1}{2}\log(2\pi) - \frac{1}{2}\log|\boldsymbol{\Sigma}_g| - \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}) \\ & - \frac{Jk}{2}\log(2\pi) - \frac{J}{2}\log|\boldsymbol{\Sigma}_e| - \frac{1}{2}\sum_{j=1}^{J}\boldsymbol{e}_{ij}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{e}_{ij}.\end{aligned}$$

For σ^2, we have

$$\frac{\partial(*)}{\partial\sigma^2} = -\frac{Jm}{2\sigma^2} + \frac{1}{2\sigma^4}\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right],$$

then, for $|\xi| < \|\boldsymbol{\theta}' - \boldsymbol{\theta}\| \leq \rho$,

$$\begin{aligned}
|\frac{\partial(*)}{\partial\sigma^2}(\sigma^2+\xi)| =& |-\frac{Jm}{2(\sigma^2+\xi)} + \frac{1}{2(\sigma^2+\xi)^2}\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]| \\
\leq& |-\frac{Jm}{2(\sigma^2+\rho)} + \frac{1}{2(\sigma^2-\rho)^2}\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]| \\
\leq& \frac{Jm}{2(\sigma^2+\rho)} + \frac{1}{2(\sigma^2-\rho)^2}\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right]^T\left[\boldsymbol{y}_{i\cdot} - \boldsymbol{f}(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot})\right].
\end{aligned}$$

As discussed in Assumption 4, once we integrate $\boldsymbol{\alpha}_i, \boldsymbol{e}_{i\cdot}$ out from this upper bound, it becomes a quadratic function of $\boldsymbol{y}_{i\cdot}$, which has finite expectation.

Next, we find the upper bound for the partial derivative of $(*)$ with respect to $\boldsymbol{\mu}$. We have

$$\frac{\partial(*)}{\partial\boldsymbol{\mu}} = (\boldsymbol{\Sigma}_g^{-1})(\boldsymbol{\alpha}_i - \boldsymbol{\mu}),$$

then,

$$\frac{\partial(*)}{\partial\boldsymbol{\mu}}(\boldsymbol{\mu}+\boldsymbol{\xi}) = \boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu} - \boldsymbol{\xi}),$$

and, for $\|\boldsymbol{\xi}\| \leq \rho$,

$$\|\frac{\partial(*)}{\partial\boldsymbol{\mu}}(\boldsymbol{\mu}_l+\boldsymbol{\xi})\| \leq \|\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu} - \boldsymbol{\xi})\| \leq \|\boldsymbol{\Sigma}_g^{-1}\|(\|\boldsymbol{\alpha}_i\| + \|\boldsymbol{\mu}\| + \rho). \tag{4.2}$$

For the partial derivative of $(*)$ with respect to $\boldsymbol{\Sigma}_g$, we have

$$\frac{\partial(*)}{\partial\boldsymbol{\Sigma}_g} = -\frac{1}{2}\boldsymbol{\Sigma}_g^{-1} + \frac{1}{2}\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_g^{-1}.$$

Note that, for any matrices G (G is invertible) and E such that $\|EG^{-1}\| < 1$, we have

$$\|(G+E)^{-1}\| = \|[(I+EG^{-1})G]^{-1}\| \leq \|G^{-1}\|\cdot\|(I+EG^{-1})^{-1}\| \leq \|G^{-1}\|\cdot\frac{1}{1-\|EG^{-1}\|}.$$

Thus, for sufficient small ρ such that $\|E\| < \rho$ and $\|E\boldsymbol{\Sigma}_g^{-1}\| < 1$, we have

$$\begin{aligned}
\|\frac{\partial(*)}{\partial\boldsymbol{\Sigma}_g}(\boldsymbol{\Sigma}_g+E)\| =& \| -\frac{1}{2}(\boldsymbol{\Sigma}_g+E)^{-1}+\frac{1}{2}(\boldsymbol{\Sigma}_g+E)^{-1}(\boldsymbol{\alpha}_i-\boldsymbol{\mu})(\boldsymbol{\alpha}_i-\boldsymbol{\mu})^T(\boldsymbol{\Sigma}_g+E)^{-1}\| \\
\leq& \frac{1}{2}\|(\boldsymbol{\Sigma}_g+E)^{-1}\|+\frac{1}{2}\|(\boldsymbol{\Sigma}_g+E)^{-1}\|^2\cdot\|(\boldsymbol{\alpha}_i-\boldsymbol{\mu})(\boldsymbol{\alpha}_i-\boldsymbol{\mu})^T\| \\
\leq& \frac{1}{2}\|\boldsymbol{\Sigma}_g^{-1}\|\cdot\frac{1}{1-\|E\boldsymbol{\Sigma}_g^{-1}\|} \\
&+\frac{1}{2}\left\{\|\boldsymbol{\Sigma}_g^{-1}\|\cdot\frac{1}{1-\|E\boldsymbol{\Sigma}_g^{-1}\|}\right\}^2\cdot\|(\boldsymbol{\alpha}_i-\boldsymbol{\mu})(\boldsymbol{\alpha}_i-\boldsymbol{\mu})^T\| \\
\leq& \frac{1}{2}\|\boldsymbol{\Sigma}_g^{-1}\|\cdot\frac{1}{1-\rho\|\boldsymbol{\Sigma}_g^{-1}\|} \\
&+\frac{1}{2}\left\{\|\boldsymbol{\Sigma}_g^{-1}\|\cdot\frac{1}{1-\rho\|\boldsymbol{\Sigma}_g^{-1}\|}\right\}^2\cdot\|(\boldsymbol{\alpha}_i-\boldsymbol{\mu})(\boldsymbol{\alpha}_i-\boldsymbol{\mu})^T\|. \qquad (4.3)
\end{aligned}$$

Finally, for the partial derivative of $(*)$ with respect to $\boldsymbol{\Sigma}_e$, we have

$$\frac{\partial(*)}{\partial\boldsymbol{\Sigma}_e}=-\frac{J}{2}\boldsymbol{\Sigma}_e^{-1}+\frac{1}{2}\boldsymbol{\Sigma}_e^{-1}\left(\sum_{j=1}^{J}\boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T\right)\boldsymbol{\Sigma}_e^{-1}.$$

Similarly, for sufficient small ρ such that $\|E\| < \rho$ and $\|E\boldsymbol{\Sigma}_e^{-1}\| < 1$, we have

$$\begin{aligned}
\|\frac{\partial(*)}{\partial\boldsymbol{\Sigma}_e}(\boldsymbol{\Sigma}_e+E)\| =& \| -\frac{J}{2}\boldsymbol{\Sigma}_e^{-1}+\frac{1}{2}\boldsymbol{\Sigma}_e^{-1}\left(\sum_{j=1}^{J}\boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T\right)\boldsymbol{\Sigma}_e^{-1}\| \\
\leq& \frac{J}{2}\|\boldsymbol{\Sigma}_e^{-1}\|\cdot\frac{1}{1-\rho\|\boldsymbol{\Sigma}_e^{-1}\|}+\frac{1}{2}\left\{\|\boldsymbol{\Sigma}_e^{-1}\|\cdot\frac{1}{1-\rho\|\boldsymbol{\Sigma}_e^{-1}\|}\right\}^2\cdot\|\sum_{j=1}^{J}\boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T\|. \qquad (4.4)
\end{aligned}$$

For the upper bounds in (4.2), (4.3), and (4.4), they all have finite expectations with respect to $\boldsymbol{\alpha}_i$ and $\boldsymbol{e}_{i\cdot}$, since $\boldsymbol{\alpha}_i$ and $\boldsymbol{e}_{i\cdot}$ are normally distributed. Then, the expectations with respect to $\boldsymbol{y}_{i\cdot}$ are also finite and this completes the proof. ■

Assumption 6. *$p(x,\theta,\rho)$ is a measurable function of x for any θ and ρ.*

Proof. Since $p_{\boldsymbol{\theta}}(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})$, $p_{\boldsymbol{\theta}}(\boldsymbol{\alpha}_i)$, and $p_{\boldsymbol{\theta}}(\boldsymbol{e}_{i\cdot})$ are measurable, $p(\boldsymbol{y}_{i\cdot},\boldsymbol{\theta})$ is measur-

able because the limit of measurable functions is measurable. Then, the supremum $p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}, \rho)$ is measurable. ■

Assumption 7. *If θ_1 is a parameter point different from the true parameter point θ_0, then $F(x, \theta_1) \neq F(x, \theta_0)$ for at least one value of x.*

Proof. We show the identifiability directly on the models, which will imply the identifiability of the distribution. For the kth observation on the jth subject of father i, the four-parameter model is

$$y_{ijk} = f_{ij}(t_{ijk}) + \epsilon_{ijk},$$
$$f_{ij}(t) = [\phi_{ij2} + \phi_{ij1}(t - \phi_{ij4})]\mathrm{trn}(\phi_{ij4} - t) + [\phi_{ij2} + \phi_{ij3}(t - \phi_{ij4})]\mathrm{trn}(t - \phi_{ij4}),$$

where $\mathrm{trn}(x) = e^{cx}/(1 + e^{cx})$ and $c > 0$. We assume that the parameters $\phi_{ij1} \neq \phi_{ij3}$, otherwise f_{ij} is just a linear function. We shall show the model is identifiable. Since $E\{\epsilon_{ij}(t)\} = 0$, we have that $\epsilon_{ij}(t) = y_{ij}(t) - E\{y_{ij}(t)\}$, so there is no ambiguity about the error term. Suppose that $E\{y_{ij}(t)\} = f(\boldsymbol{\phi}_{ij}, t) = f(\boldsymbol{\phi}^*_{ij}, t)$ for all i, j. We first show the identifiability of ϕ_{ij1} and ϕ_{ij3}. Note that

$$\mathrm{trn}(-x) = \frac{e^{-cx}}{1 + e^{-cx}} = \frac{1}{1 + e^{cx}} = 1 - \mathrm{trn}(x),$$

so we have

$$\begin{aligned} f_{ij}(t) =& [\phi_{ij2} + \phi_{ij1}(t - \phi_{ij4})][1 - \mathrm{trn}(t - \phi_{ij4})] + [\phi_{ij2} + \phi_{ij3}(t - \phi_{ij4})]\mathrm{trn}(t - \phi_{ij4}) \\ =& \phi_{ij2} + \phi_{ij1}(t - \phi_{ij4}) + (\phi_{ij3} - \phi_{ij1})(t - \phi_{ij4})\mathrm{trn}(t - \phi_{ij4}). \end{aligned}$$

Then,

$$f'_{ij}(t) = \phi_{ij1} + (\phi_{ij3} - \phi_{ij1})\mathrm{trn}(t - \phi_{ij4}) + (\phi_{ij3} - \phi_{ij1})(t - \phi_{ij4})\mathrm{trn}'(t - \phi_{i4}).$$

Since $\text{trn}'(x) = ce^{cx}/(1+e^{cx})^2$, we have

$$\lim_{x\to\infty} x\text{trn}'(x) = \lim_{x\to\infty} \frac{cxe^{cx}}{(1+e^{cx})^2} = \lim_{x\to\infty} \frac{cx}{e^{-cx}+2+e^{cx}} = \lim_{x\to\infty} \frac{c}{-ce^{-cx}+ce^{cx}} = 0,$$

similarly,

$$\lim_{x\to-\infty} x\text{trn}'(x) = \lim_{x\to-\infty} \frac{c}{-ce^{-cx}+ce^{cx}} = 0.$$

Therefore,

$$\lim_{t\to-\infty} f'_{ij}(t) = \phi_{ij1},$$
$$\lim_{t\to\infty} f'_{ij}(t) = \phi_{ij3}.$$

Since the linear functions $f(\boldsymbol{\phi}_{ij}, t)$ and $f(\boldsymbol{\phi}^*_{ij}, t)$ have the same derivatives at infinities, we obtain $\phi^*_{ij1} = \phi_{ij1}$ and $\phi^*_{ij3} = \phi_{ij3}$. Thus, ϕ_{ij1} and ϕ_{ij3} are identifiable. Next, we prove the identifiability of ϕ_{ij2} and ϕ_{ij4}. We can show that

$$f''_{ij}(t) = (\phi_{ij3}-\phi_{ij1})[2\text{trn}'(t-\phi_{ij4}) + (t-\phi_{ij4})\text{trn}''(t-\phi_{ij4})],$$

and

$$f'''_{ij}(t) = (\phi_{ij3}-\phi_{ij1})[3\text{trn}''(t-\phi_{ij4}) + (t-\phi_{ij4})\text{trn}'''(t-\phi_{ij4})].$$

The turning point of the function $f_{ij}(t)$ can be characterized as the maximum of its second derivative of $f''_{ij}(t)$, which is given by the root of the third derivative $f'''_{ij}(t)$. Equivalently, we have to find the root of the following function

$$h(x) = 3\text{trn}''(x) + x\text{trn}'''(x).$$

Since

$$\text{trn}''(x) = \frac{c^2 e^{cx}(1-e^{cx})}{(1+e^{cx})^3},$$
$$\text{trn}'''(x) = \frac{c^3 e^{cx}(e^{2cx}-4e^{cx}+1)}{(1+e^{cx})^4},$$

we obtain

$$\begin{aligned} h(x) =& 3\text{trn}''(x) + x\text{trn}'''(x) \\ =& \frac{3c^2 e^{cx}(1-e^{cx})}{(1+e^{cx})^3} + \frac{c^3 x e^{cx}(e^{2cx}-4e^{cx}+1)}{(1+e^{cx})^4} \\ =& \frac{3c^2 e^{cx}(1-e^{cx})(1+e^{cx}) + c^3 x e^{cx}(e^{2cx}-4e^{cx}+1)}{(1+e^{cx})^4} \\ =& \frac{c^2 e^{cx}[3-3e^{2cx}+cxe^{2cx}-4cxe^{cx}+cx]}{(1+e^{cx})^4} \end{aligned}$$

It can be shown that this function has three roots. $x = 0$, or equivalently, $t = \phi_{ij4}$ is the root such that $f''_{ij}(t)$ attaining its maximum. Thus, ϕ_{ij4} is the only turning point of $f_{ij}(t)$. Since $f(\boldsymbol{\phi}_{ij}, t)$ and $f(\boldsymbol{\phi}^*_{ij}, t)$ must have the same turning point and the values at the turning point are equal, we obtain the identifiability of ϕ_{ij4} and ϕ_{ij2}. This completes the proof of model identifiability. ■

Assumption 8. *The parameter space Θ is a closed subset of the k-dimensional Cartesian space.*

Assumption 8 is actually not true as stated because the parameter space of this model is open (the variances have to be strictly positive). As mentioned in Wald (1949), Assumptions 2, 3, and 8 can be replaced by the following one:

Assumption 9. *It is possible to introduce a distance $\delta(\theta_1, \theta_2)$ in the space Θ such that the following four conditions hold:*

(i) The distance $\delta(\theta_1, \theta_2)$ makes Θ to a metric space.

(ii) If $\lim_{k\to\infty} \theta_k = \theta$, then $\lim_{k\to\infty} p(x, \theta_k) = p(x, \theta)$ for all x except perhaps on a

set which may depend on the limit point θ (but not on the sequence $\{\theta_k\}$) and whose probability measure is zero according to the probability distribution corresponding to the true parameter point θ_0. (This is Assumption 2).

(iii) If θ_0 is a fixed point in Θ and $\lim_{k\to\infty}\delta(\theta_k,\theta_0)=\infty$, then $\lim_{k\to\infty}p(x,\theta_n)=0$. (This is Assumption 3).

(iv) Any closed and bounded subset of Θ is compact.

Proof. Using the Euclidean distance on the parameter space Θ, Assumption 9 (i) and (iv) hold immediately. ■

Since all the assumptions in Wald (1949) are satisfied, the consistency of the maximum likelihood estimator follows.

4.2 Asymptotic normality

Now we establish the asymptotic normality of $\hat{\boldsymbol{\theta}}$ for $I\to\infty$ and J fixed by verifying the sufficient conditions of Huber (1967). Let $T_n=T_n(x_1,x_2,\cdots,x_n)$ be the sequence of maximum likelihood estimators. The consistency of T_n has been proved. Let $\psi(x,\theta)=(\partial/\partial\theta)\log p(x,\theta)$. Then, T_n is asymptotically normal, if the following conditions are satisfied:

(N-1) For each fixed $\theta\in\Theta$, $\psi(x,\theta)$ is a $\mathcal{U}$-measurable and $\psi(x,\theta)$ is separable.

Let $\lambda(\theta)=\mathbb{E}\psi(x,\theta)$, and $u(x,\theta,d)=\sup_{|\tau-\theta|\le d}|\psi(x,\tau)-\psi(x,\theta)|$.

(N-2) There is a $\theta_0\in\Theta$ such that $\lambda(\theta_0)=0$.

(N-3) There are strictly positive numbers a,b,c,d_0 such that

(i) $|\lambda(\theta)|\ge a|\theta-\theta_0|$ for $|\theta-\theta_0|\le d_0$,

(ii) $\mathbb{E}u(x,\theta,d)\le bd$ for $|\theta-\theta_0|+d\le d_0$, $d\ge 0$,

(iii) $\mathbb{E}[u(x,\theta,d)^2]\le cd$ for $|\theta-\theta_0|+d\le d_0$, $d\ge 0$,

(N-4) The expectation $\mathbb{E}[|\psi(x,\theta_0)|^2]$ is finite.

Proof. (N-1) is satisfied, since the log-likelihood function is continuous in $\boldsymbol{y}_{i\cdot}$ and $\boldsymbol{\theta}$. Let $M(\boldsymbol{\theta}) = \mathbb{E}[\log p(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})]$. By Lemma 1 of Wald (1976), $M(\boldsymbol{\theta})$ has a unique maximum at $\boldsymbol{\theta}_0$. Note that $\lambda(\boldsymbol{\theta})$ is the gradient of $M(\boldsymbol{\theta})$, then $\lambda(\boldsymbol{\theta}_0) = 0$ and thus (N-2) holds. Next, we verify the conditions in (N-3). Note that $\lambda(\boldsymbol{\theta})$ is continuous on $\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d_0\}$, by the mean value theorem, we have

$$\lambda(\boldsymbol{\theta}) = \lambda(\boldsymbol{\theta}_0) + H(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta} - \boldsymbol{\theta}_0),$$

where $\bar{\boldsymbol{\theta}}$ is a point between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$, and $H(\bar{\boldsymbol{\theta}})$ is the Hessian matrix of $M(\boldsymbol{\theta})$ at $\bar{\boldsymbol{\theta}}$. Note that $\lambda(\boldsymbol{\theta}_0) = 0$ and the Hessian of $M(\boldsymbol{\theta})$ is invertible at points near $\boldsymbol{\theta}_0$. Thus,

$$\|H(\bar{\boldsymbol{\theta}})^{-1}\|\|\lambda(\boldsymbol{\theta})\| \geq \|H(\bar{\boldsymbol{\theta}})^{-1}\lambda(\boldsymbol{\theta})\| = \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|,$$

and then,

$$\|\lambda(\boldsymbol{\theta})\| \geq \|H(\bar{\boldsymbol{\theta}})^{-1}\|^{-1}\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|.$$

Since $H(\boldsymbol{\theta}_0)$ is invertible, we have $\|H(\boldsymbol{\theta}_0)^{-1}\| < \infty$. Moreover, $H(\boldsymbol{\theta})$ is continuous, then there is a $c < \infty$ such that $\|H(\boldsymbol{\theta})^{-1}\| < c$ for all $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d_0$ for $d_0 > 0$. Thus, taking $a = 1/c$, we have (N-3) (i) verified.
By the differentiability of $\psi(\boldsymbol{y}_{i,\cdot}, \boldsymbol{\theta})$ and the mean value theorem,

$$\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\tau}) - \psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}) = \nabla\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta} + \eta(\boldsymbol{y}_{i\cdot}))(\boldsymbol{\tau} - \boldsymbol{\theta}),$$

for $\|\eta(\boldsymbol{y}_{i\cdot})\| < d$ and $\|\boldsymbol{\tau} - \boldsymbol{\theta}\| \leq d$. Note that $\|\mathbb{E}[\nabla\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)]\|$ is equal to the norm of the Fisher Information Matrix, and then it is finite. Moreover, since $\mathbb{E}[\nabla\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})]$ is continuous, there is a $b < \infty$ such that $\|\mathbb{E}[\nabla\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta})]\| < b$ for all $\boldsymbol{\theta}$ such that

$\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d_0 - d$ for some d_0. Since $\boldsymbol{\theta} + \eta(\boldsymbol{y}_{i\cdot}) \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq d_0 - d\}$, we have

$$\mathbb{E}[u(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}, d)] \leq b \cdot d,$$

and

$$\mathbb{E}[u(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}, d)^2] \leq c \cdot d,$$

where $c = d \cdot b^2$. Then, (N-3) (ii) and (iii) are satisfied.

Finally, note that

$$\begin{aligned}
\mathbb{E}[\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)^T \psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)] &= \mathbb{E}[\mathrm{tr}\{\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)^T \psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)\}] \\
&= \mathbb{E}[\mathrm{tr}\{\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0) \psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)^T\}] \\
&= \mathrm{tr}\{\mathbb{E}[\psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0) \psi(\boldsymbol{y}_{i\cdot}, \boldsymbol{\theta}_0)^T]\}.
\end{aligned}$$

Then, (N-4) is satisfied because the trace of the Fisher Information Matrix is finite. ■

Since all conditions of Theorem 3 in Huber (1967) are satisfied,

$$\sqrt{I}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{F}^{-1}),$$

where $\boldsymbol{F} = \mathbb{E}[\{\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{y}_{i\cdot})\}\{\frac{\partial}{\partial \boldsymbol{\theta}} \log p(\boldsymbol{y}_{i\cdot})\}^T]$ is the Fisher Information Matrix for the parameter vector $\boldsymbol{\theta}$. The elements of the partial derivatives involved can be calculated by straightforward differentiation of (4.1), which we do below.

Part I: Derivative of $\log p(\boldsymbol{y}_{i\cdot})$ w.r.t. $\boldsymbol{\mu}$

Since $\{\boldsymbol{\alpha}_i\}$ are i.i.d. from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$, we have

$$\log p(\boldsymbol{\alpha}_i) \propto -\frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_g^{-1} (\boldsymbol{\alpha}_i - \boldsymbol{\mu}).$$

Then

$$\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial \boldsymbol{\mu}} = \boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}).$$

Therefore, for $l = 1, \cdots, k$,

$$\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial \mu_l} = (\boldsymbol{\Sigma}_g^{-1})_{l\cdot}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}),$$

where $(\boldsymbol{\Sigma}_g^{-1})_{l\cdot}$ denotes the lth row of $\boldsymbol{\Sigma}_g^{-1}$. Then,

$$\begin{aligned}
\frac{\partial \log p(\boldsymbol{y}_{i\cdot})}{\partial \mu_l} =& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\frac{\partial p(\boldsymbol{y}_{i\cdot})}{\partial \mu_l} \\
=& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int \left\{\prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)\right\}\left\{\prod_{j=1}^{J} p(\boldsymbol{e}_{ij})\right\}\frac{\partial p(\boldsymbol{\alpha}_i)}{\partial \mu_l} d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i) \\
=& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int \left\{\prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)\right\}\left\{\prod_{j=1}^{J} p(\boldsymbol{e}_{ij})\right\}\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial \mu_l} p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{\alpha}_i) \\
=& \mathbb{E}\left\{\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial \mu_l}\Big|\boldsymbol{y}_{i\cdot}\right\} \\
=& \mathbb{E}\left\{(\boldsymbol{\Sigma}_g^{-1})_{l\cdot}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})|\boldsymbol{y}_{i\cdot}\right\} \\
=& (\boldsymbol{\Sigma}_g^{-1})_{l\cdot}[\mathbb{E}(\boldsymbol{\alpha}_i|\boldsymbol{y}_{i\cdot}) - \boldsymbol{\mu}].
\end{aligned}$$

Part II: Derivative of $\log p(\boldsymbol{y}_{i\cdot})$ w.r.t. $\boldsymbol{\Sigma}_g$

Similarly, since $\{\boldsymbol{\alpha}_i\}$ are i.i.d. from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$,

$$\log\, p(\boldsymbol{\alpha}_i) \propto -\frac{1}{2}\log\det \boldsymbol{\Sigma}_g - \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}),$$

then,

$$d_{\boldsymbol{\Sigma}_g}\log p(\boldsymbol{\alpha}_i) = -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_g^{-1}d\boldsymbol{\Sigma}_g) + \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_g^{-1}(d\boldsymbol{\Sigma}_g)\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}).$$

For $l, r = 1, \cdots, k$, the partial derivative is

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial (\Sigma_g)_{lr}} =& -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_g^{-1}\boldsymbol{E}_{lr}) + \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{E}_{lr})\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i - \boldsymbol{\mu}) \\ =& -\frac{1}{2}(\boldsymbol{\Sigma}_g^{-1})_{lr} + \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T(\boldsymbol{\Sigma}_g^{-1})_{\cdot l}(\boldsymbol{\Sigma}_g^{-1})_{\cdot r}^T(\boldsymbol{\alpha}_i - \boldsymbol{\mu}).\end{aligned}$$

where $\boldsymbol{E}_{lr}$ the $k \times k$ matrix with 1 in the (l, r) coordinate and 0 elsewhere, and $(\boldsymbol{\Sigma}_g^{-1})_{\cdot l}$ denotes the lth column of $\boldsymbol{\Sigma}_g^{-1}$. Then

$$\begin{aligned}\frac{\partial \log p(\boldsymbol{y}_{i\cdot})}{\partial (\Sigma_g)_{lr}} =& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\frac{\partial p(\boldsymbol{y}_{i\cdot})}{\partial (\Sigma_g)_{lr}} \\ =& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int \left\{\prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)p(\boldsymbol{e}_{ij})\right\} \frac{\partial p(\boldsymbol{\alpha}_i)}{\partial (\Sigma_g)_{lr}}\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i) \\ =& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int \left\{\prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)p(\boldsymbol{e}_{ij})\right\} p(\boldsymbol{\alpha}_i)\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial (\Sigma_g)_{lr}}\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i) \\ =& \mathbb{E}\left\{\frac{\partial \log p(\boldsymbol{\alpha}_i)}{\partial (\Sigma_g)_{lr}}\Big|\boldsymbol{y}_{i\cdot}\right\} \\ =& \mathbb{E}\left\{-\frac{1}{2}(\boldsymbol{\Sigma}_g^{-1})_{lr} + \frac{1}{2}(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T(\boldsymbol{\Sigma}_g^{-1})_{\cdot l}(\boldsymbol{\Sigma}_g^{-1})_{\cdot r}^T(\boldsymbol{\alpha}_i - \boldsymbol{\mu})\Big|\boldsymbol{y}_{i\cdot}\right\} \\ =& -\frac{1}{2}(\boldsymbol{\Sigma}_g^{-1})_{lr} + \frac{1}{2}(\boldsymbol{\Sigma}_g^{-1})_{\cdot r}^T\mathbb{E}\left[(\boldsymbol{\alpha}_i - \boldsymbol{\mu})(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T\big|\boldsymbol{y}_{i\cdot}\right](\boldsymbol{\Sigma}_g^{-1})_{\cdot l}\end{aligned}$$

Part III: Derivative of $\log p(\boldsymbol{y}_{i\cdot})$ w.r.t. $\boldsymbol{\Sigma}_e$

Since $\{\boldsymbol{e}_{ij}\}$ are i.i.d. from $N(\boldsymbol{0}, \boldsymbol{\Sigma}_e)$, we have

$$\log p(\boldsymbol{e}_{ij}) \propto -\frac{1}{2}\log\det\boldsymbol{\Sigma}_e - \frac{1}{2}\boldsymbol{e}_{ij}^T\boldsymbol{\Sigma}_e^{-1}\boldsymbol{e}_{ij}.$$

Then,

$$d_{\boldsymbol{\Sigma}_e}\log p(\boldsymbol{e}_{ij}) = -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_e^{-1}d\boldsymbol{\Sigma}_e) + \frac{1}{2}\boldsymbol{e}_{ij}^T\boldsymbol{\Sigma}_e^{-1}(d\boldsymbol{\Sigma}_e)\boldsymbol{\Sigma}_e^{-1}\boldsymbol{e}_{ij}.$$

For $l, r = 1, \cdots, k$,

$$\frac{\partial \log p(\boldsymbol{e}_{ij})}{\partial (\Sigma_e)_{lr}} = -\frac{1}{2}\mathrm{tr}(\boldsymbol{\Sigma}_e^{-1}\boldsymbol{E}_{lr}) + \frac{1}{2}\boldsymbol{e}_{ij}^T\boldsymbol{\Sigma}_e^{-1}(\boldsymbol{E}_{lr})\boldsymbol{\Sigma}_e^{-1}\boldsymbol{e}_{ij}$$

$$= -\frac{1}{2}(\boldsymbol{\Sigma}_e^{-1})_{lr} + \frac{1}{2}\boldsymbol{e}_{ij}^T(\boldsymbol{\Sigma}_e^{-1})_{\cdot l}(\boldsymbol{\Sigma}_e^{-1})_{\cdot r}^T\boldsymbol{e}_{ij}.$$

Therefore,

$$\begin{aligned}
\frac{\partial \log p(\boldsymbol{y}_{i\cdot})}{\partial(\Sigma_e)_{lr}} =& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\frac{\partial p(\boldsymbol{y}_{i\cdot})}{\partial(\Sigma_e)_{lr}} \\
=& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int\left\{\prod_{i=1}^{J}p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij},\boldsymbol{\alpha}_i)\right\}\left\{\sum_{j=1}^{J}\left[\frac{\partial p(\boldsymbol{e}_{ij})}{\partial(\Sigma_e)_{lr}}\prod_{j'\neq j}p(\boldsymbol{e}_{ij'})\right]\right\} \\
& \times p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1},\cdots,\boldsymbol{e}_{iJ},\boldsymbol{\alpha}_i) \\
=& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int\left\{\prod_{i=1}^{J}p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij},\boldsymbol{\alpha}_i)\right\}\left\{\sum_{j=1}^{J}\left[\frac{\partial \log p(\boldsymbol{e}_{ij})}{\partial(\Sigma_e)_{lr}}p(\boldsymbol{e}_{ij})\prod_{j'\neq j}p(\boldsymbol{e}_{ij'})\right]\right\} \\
& \times p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1},\cdots,\boldsymbol{e}_{iJ},\boldsymbol{\alpha}_i) \\
=& \frac{1}{p(\boldsymbol{y}_{i\cdot})}\int\left\{\prod_{i=1}^{J}p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij},\boldsymbol{\alpha}_i)\right\}\left\{\sum_{j=1}^{J}\frac{\partial \log p(\boldsymbol{e}_{ij})}{\partial(\Sigma_e)_{lr}}\right\}\left\{\prod_{j=1}^{J}p(\boldsymbol{e}_{ij})\right\} \\
& \times p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1},\cdots,\boldsymbol{e}_{iJ},\boldsymbol{\alpha}_i) \\
=& \mathbb{E}\left\{\sum_{j=1}^{J}\frac{\partial \log p(\boldsymbol{e}_{ij})}{\partial(\Sigma_e)_{lr}}\Big|\boldsymbol{y}_{i\cdot}\right\} \\
=& \mathbb{E}\left\{\sum_{j=1}^{J}\left[-\frac{1}{2}(\boldsymbol{\Sigma}_e^{-1})_{lr} + \frac{1}{2}\boldsymbol{e}_{ij}^T(\boldsymbol{\Sigma}_e^{-1})_{\cdot l}(\boldsymbol{\Sigma}_e^{-1})_{\cdot r}^T\boldsymbol{e}_{ij}\right]\Big|\boldsymbol{y}_{i\cdot}\right\} \\
=& -\frac{J}{2}(\boldsymbol{\Sigma}_e^{-1})_{lr} + \frac{1}{2}(\boldsymbol{\Sigma}_e^{-1})_{\cdot r}^T\mathbb{E}\left(\sum_{j=1}^{J}\boldsymbol{e}_{ij}\boldsymbol{e}_{ij}^T\Big|\boldsymbol{y}_{i\cdot}\right)(\boldsymbol{\Sigma}_e^{-1})_{\cdot l}.
\end{aligned}$$

Part IV: Derivative of $\log p(\boldsymbol{y}_{i\cdot})$ w.r.t. σ^2

Since $\boldsymbol{y}_{ij}|(\boldsymbol{\alpha}_i,\boldsymbol{e}_{ij}) \sim N(\boldsymbol{f}(\boldsymbol{\phi}_{ij}),\sigma^2\boldsymbol{I}_m)$ with $\boldsymbol{\phi}_{ij} = \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij}$, we have

$$\log p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij},\boldsymbol{\alpha}_i) \propto -\frac{m}{2}\log\sigma^2 - \frac{1}{2\sigma^2}[\boldsymbol{y}_{ij}-\boldsymbol{f}(\boldsymbol{\phi}_{ij})]^T[\boldsymbol{y}_{ij}-\boldsymbol{f}(\boldsymbol{\phi}_{ij})],$$

then

$$\frac{\partial \log p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij},\boldsymbol{\alpha}_i)}{\partial\sigma^2} = -\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4}[\boldsymbol{y}_{ij}-\boldsymbol{f}(\boldsymbol{\phi}_{ij})]^T[\boldsymbol{y}_{ij}-\boldsymbol{f}(\boldsymbol{\phi}_{ij})].$$

Thus,

$$\begin{aligned}
\frac{\partial p(\boldsymbol{y}_{i\cdot})}{\partial \sigma^2} &= \int \left\{ \sum_{j=1}^{J} \left[\frac{\partial p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)}{\partial \sigma^2} \prod_{j' \neq j} p(\boldsymbol{y}_{ij'}|\boldsymbol{e}_{ij'}, \boldsymbol{\alpha}_i) \right] \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{e}_{ij}) \right\} \\
&\quad \times p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i) \\
&= \int \left\{ \sum_{j=1}^{J} \left[\frac{\partial \log p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)}{\partial \sigma^2} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i) \prod_{j' \neq j} p(\boldsymbol{y}_{ij'}|\boldsymbol{e}_{ij'}, \boldsymbol{\alpha}_i) \right] \right\} \\
&\quad \times \left\{ \prod_{j=1}^{J} p(\boldsymbol{e}_{ij}) \right\} p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i) \\
&= \int \left\{ \sum_{j=1}^{J} \frac{\partial \log p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)}{\partial \sigma^2} \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i) \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{e}_{ij}) \right\} \\
&\quad \times p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i).
\end{aligned}$$

Then,

$$\begin{aligned}
\frac{\partial \log p(\boldsymbol{y}_{i\cdot})}{\partial \sigma^2} =& \frac{1}{p(\boldsymbol{y}_{i\cdot})} \frac{\partial p(\boldsymbol{y}_{i\cdot})}{\partial \sigma^2} \\
=& \frac{1}{p(\boldsymbol{y}_{i\cdot})} \int \left\{ \sum_{j=1}^{J} \frac{\partial \log p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)}{\partial \sigma^2} \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i) \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{e}_{ij}) \right\} \\
&\quad \times p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}, \boldsymbol{\alpha}_i) \\
=& \mathbb{E} \left\{ \sum_{j=1}^{J} \frac{\partial \log p(\boldsymbol{y}_{ij}|\boldsymbol{e}_{ij}, \boldsymbol{\alpha}_i)}{\partial \sigma^2} \Big| \boldsymbol{y}_{i\cdot} \right\} \\
=& \mathbb{E} \left\{ \sum_{j=1}^{J} \left(-\frac{m}{2\sigma^2} + \frac{1}{2\sigma^4} [\boldsymbol{y}_{ij} - \boldsymbol{f}(\boldsymbol{\phi}_{ij})]^T [\boldsymbol{y}_{ij} - \boldsymbol{f}(\boldsymbol{\phi}_{ij})] \right) \Big| \boldsymbol{y}_{i\cdot} \right\} \\
=& -\frac{Jm}{2\sigma^2} + \frac{1}{2\sigma^4} \mathbb{E} \left\{ \sum_{j=1}^{J} [\boldsymbol{y}_{ij} - \boldsymbol{f}(\boldsymbol{\phi}_{ij})]^T [\boldsymbol{y}_{ij} - \boldsymbol{f}(\boldsymbol{\phi}_{ij})] \Big| \boldsymbol{y}_{i\cdot} \right\}.
\end{aligned}$$

With the formulae of these partial derivatives, we obtain the estimate of $\boldsymbol{F}$,

$$\hat{\boldsymbol{F}} = \frac{1}{I} \sum_{i=1}^{I} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log \hat{p}(\boldsymbol{y}_{i\cdot}) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\theta}} \log \hat{p}(\boldsymbol{y}_{i\cdot}) \right\}^T,$$

where $\hat{p}$ means that $\boldsymbol{\theta}$ is replaced by $\hat{\boldsymbol{\theta}}$ everywhere.

4.3 Asymptotics for the general model

In this section, we derive the asymptotic properties for the balanced one-factor model of the general shape-invariant model. We first present the model specification. Then, we show the consistency and the asymptotic normality of the maximum likelihood estimators.

Suppose that we have I fathers and each father i, $i = 1, \cdots, I$, has J descendants. Assume that the observations of all subjects are obtained at the same time grids $(t_1, \cdots, t_m)$. Thus, a total of $N = I \times J \times m$ data values available. Let y_{ijk} denote the kth observation for the jth descendant of father i and assume it follows

$$y_{ijk} = a_{ij} f(v_{ij}(t_{ijk})) + \epsilon_{ijk}, \tag{4.5}$$

where a_{ij} is the amplitude parameter, $\{v_{ij}\}$ are the inverse functions of the warping functions $\{w_i\}$, $\{\epsilon_{ijk}\}$ are i.i.d.$N(0, \sigma^2)$ random errors, and $f(t)$ is the B-spline shape function

$$f(t) = \boldsymbol{d}^T \boldsymbol{B}(t), \tag{4.6}$$

where $\boldsymbol{d}$ is a vector of the B-spline coefficients and $\boldsymbol{B}(t)$ represents the B-spline basis. Note that $v_{ij}(t)$ is a function of the Jupp-transformed knot sequence $\boldsymbol{\theta}_{ij}$, we denote $v_{ij}(t)$ as $v_{\boldsymbol{\theta}_{ij}}(t)$. Assume that the subject-specific parameters $\boldsymbol{\phi}_{ij} = (a_{ij}, \boldsymbol{\theta}_{ij})^T \in \mathbb{R}^{(p+1)}$, we propose the following one-factor design

$$\boldsymbol{\phi}_{ij} = \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij}, \tag{4.7}$$

where $\{\boldsymbol{\alpha}_i\}$ are i.i.d. $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$ representing the father effects, $\{\boldsymbol{e}_{ij}\}$ are i.i.d. $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$ indicating the environment effects and the genetic contribution of the random chosen

mother. We also assume $\{\boldsymbol{\alpha}_i\}$ are independent of $\{\boldsymbol{e}_{ij}\}$. Then

$$\boldsymbol{y}_{ij}|(\boldsymbol{\alpha}_i, \boldsymbol{e}_{ij}) \sim N(\tilde{\boldsymbol{f}}(\boldsymbol{\phi}_{ij}), \sigma^2 \boldsymbol{I}_m), \tag{4.8}$$

where

$$\boldsymbol{y}_{ij} = \begin{pmatrix} y_{ij1} \\ y_{ij2} \\ \vdots \\ y_{ijm} \end{pmatrix}, \qquad \tilde{\boldsymbol{f}}(\boldsymbol{\phi}_{ij}) = \begin{pmatrix} a_{ij} f(v_{\boldsymbol{\theta}_{ij}}(t_{ij1})) \\ a_{ij} f(v_{\boldsymbol{\theta}_{ij}}(t_{ij2})) \\ \vdots \\ a_{ij} f(v_{\boldsymbol{\theta}_{ij}}(t_{ijm}))\} \end{pmatrix}.$$

Denote $p(\boldsymbol{y}_{i\cdot}) = p(\boldsymbol{y}_{i1}, \cdots, \boldsymbol{y}_{iJ})$. Since subjects from different families are independent, the log-likelihood function is

$$l = \sum_{i=1}^{I} \log\, p(\boldsymbol{y}_{i\cdot}).$$

Under all assumptions, $\{\boldsymbol{y}_{ij}\}$ are identically distributed and they are conditionally independent given $(\boldsymbol{\alpha}_i, \boldsymbol{e}_{ij})$. Therefore, the component $p(\boldsymbol{y}_{i\cdot})$ in l is

$$p(\boldsymbol{y}_{i\cdot}) = \int \left\{ \prod_{j=1}^{J} p(\boldsymbol{y}_{ij}|\boldsymbol{\alpha}_i, \boldsymbol{e}_{ij}) \right\} \left\{ \prod_{j=1}^{J} p(\boldsymbol{e}_{ij}) \right\} p(\boldsymbol{\alpha}_i)\, d(\boldsymbol{\alpha}_i, \boldsymbol{e}_{i1}, \cdots, \boldsymbol{e}_{iJ}). \tag{4.9}$$

Similar to Section 4.2, the maximum likelihood estimators of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}_g$, $\boldsymbol{\Sigma}_e$, σ^2 and $\boldsymbol{d}$ can be obtained by taking the partial derivatives of l with respect to the parameters and setting them equal to zero, from which we obtain:

$$\hat{\boldsymbol{\mu}} = \frac{1}{I} \sum_{i=1}^{I} \mathbb{E}\{\boldsymbol{\alpha}_i | \boldsymbol{y}_{i\cdot}\},$$

$$\hat{\boldsymbol{\Sigma}}_g = \frac{1}{I} \sum_{i=1}^{I} \mathbb{E}\{(\boldsymbol{\alpha}_i - \boldsymbol{\mu})(\boldsymbol{\alpha}_i - \boldsymbol{\mu})^T \big| \boldsymbol{y}_{i\cdot}\},$$

$$\hat{\boldsymbol{\Sigma}}_e = \frac{1}{n} \sum_{i=1}^{I} \mathbb{E}\left\{ \sum_{j=1}^{J} \boldsymbol{e}_{ij} \boldsymbol{e}_{ij}^T \Big| \boldsymbol{y}_{i\cdot} \right\},$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{I}\mathbb{E}\left\{\sum_{j=1}^{J}[\boldsymbol{y}_{ij}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi}_{ij})]^T[\boldsymbol{y}_{ij}-\tilde{\boldsymbol{f}}(\boldsymbol{\phi}_{ij})]\big|\boldsymbol{y}_{i\cdot}\right\},$$

$$\hat{\boldsymbol{d}} = \left[\sum_{i=1}^{I}\mathbb{E}\left\{\sum_{j=1}^{J}a_{ij}^2\sum_{k=1}^{m}\boldsymbol{B}(t_{ijk},\boldsymbol{\theta}_{ij})\boldsymbol{B}^T(t_{ijk},\boldsymbol{\theta}_{ij})\big|\boldsymbol{y}_{i\cdot}\right\}\right]^{-1} \times\left[\sum_{i=1}^{I}\mathbb{E}\left\{\sum_{j=1}^{J}a_{ij}\sum_{k=1}^{m}\boldsymbol{B}(t_{ijk},\boldsymbol{\theta}_{ij})y_{ijk}\big|\boldsymbol{y}_{i\cdot}\right\}\right].$$

Based on the model assumptions, $\{\boldsymbol{y}_{i\cdot}\}$ are independently and identically distributed from the following distribution

$$\begin{aligned}
p(\boldsymbol{y}_{i\cdot}) &= \int p(\boldsymbol{y}_{i\cdot}|\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})p(\boldsymbol{\alpha}_i)p(\boldsymbol{e}_{i\cdot})\,d(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})\\
&= \int(2\pi\sigma^2)^{-Jm/2}\exp\left\{-\frac{1}{2\sigma^2}\left[\boldsymbol{y}_{i\cdot}-\tilde{\boldsymbol{f}}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})\right]^T\left[\boldsymbol{y}_{i\cdot}-\tilde{\boldsymbol{f}}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})\right]\right\}\\
&\quad\times(2\pi)^{-1/2}|\boldsymbol{\Sigma}_g|^{-1/2}\exp\left\{-\frac{1}{2}(\boldsymbol{\alpha}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}_g^{-1}(\boldsymbol{\alpha}_i-\boldsymbol{\mu})\right\}\\
&\quad\times(2\pi)^{-J(p+1)/2}|\boldsymbol{\Sigma}_e|^{-J/2}\exp\left\{-\frac{1}{2}\boldsymbol{e}_{i\cdot}^T(\boldsymbol{I}_J\otimes\boldsymbol{\Sigma}_e)^{-1}\boldsymbol{e}_{i\cdot}\right\}d(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot}),
\end{aligned}$$

where $\boldsymbol{e}_{i\cdot}=(\boldsymbol{e}_{i1},\cdots,\boldsymbol{e}_{iJ})$ and $\tilde{\boldsymbol{f}}(\boldsymbol{\alpha}_i,\boldsymbol{e}_{i\cdot})$ is the column vector containing the responses observed at all time grids of subjects in family i. Let $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ be the row vectors containing the unique elements of the covariance matrices $\boldsymbol{\Sigma}_g$ and $\boldsymbol{\Sigma}_e$, respectively. Then, define

$$\boldsymbol{\eta}=(\boldsymbol{\mu}^T,\boldsymbol{\gamma},\boldsymbol{\lambda},\sigma^2)$$

as the unknown parameter vector of the model. Similar to Section 4.1, the consistency of the maximum likelihood estimator $\hat{\boldsymbol{\eta}}$ can be shown by verifying the assumptions in Wald (1949), which are the same for the one-factor four-parameter model and the one-factor general model except for the model identifiability. Thus, it suffices to prove the Assumption 7 for the one-factor general model.

Assumption 7. *If θ_1 is a parameter point different from the true parameter point θ_0, then $F(x,\theta_1)\neq F(x,\theta_0)$ for at least one value of x.*

Proof. Again, we show the identifiability directly on the models, which implies the identifiability of the distribution. Recall that, for the jth subject of father i, the one-factor general model is

$$y_{ij}(t) = a_{ij} f(v_{ij}(t)) + \epsilon_{ij}(t), \tag{4.10}$$

where f is the common shape function, $\{v_{ij}(t)\}$ are the inverse functions of the warping functions $\{w_{ij}(t)\}$. For identifiability, we assume that $a_{ij} \neq 0, \bar{a} = 1$, $\bar{w}(t) = t$ for all t, and f is piece-wise monotone without flat parts. Since $E\{\epsilon_{ij}(t)\} = 0$, we have that $\epsilon_{ij}(t) = y_{ij}(t) - E\{y_{ij}(t)\}$, so there is no ambiguity about the error term. Suppose that $E\{y_{ij}(t)\} = a_{ij} f\{v_{ij}(t)\} = a^*_{ij} f^*\{v^*_{ij}(t)\}$ for all i, j. Then we have

$$f(t) = \frac{a^*_{ij}}{a_{ij}} f^*[v^*_{ij}\{w_{ij}(t)\}],$$

for all t and for all i, j. Since the left hand side does not depend on i, j, we have that $a^*_{ij}/a_{ij} = c$ for all i, j and some constant c, and that $v^*_{ij}\{w_{ij}(t)\} = g(t)$ for all i, j and some function g because of the piece-wise monotonicity of f^*. Then $w_{ij}(t) = w^*_{ij}\{g(t)\}$ for all i, j, and by assumption we have $\bar{w}(t) = \bar{w}^*(t) = t$, thus $g(t) = t$. Therefore, $v_{ij} = v^*_{ij}$ for all i, j and the warping functions are identifiable. We also have $a^*_{ij} = c a_{ij}$ for all i, j, and the assumption $\bar{a^*} = \bar{a} = 1$ implies that $c = 1$; then $a_{ij} = a^*_{ij}$ for all i, j, so the scaling parameters are also identifiable. The identifiability of the scaling parameters and the warping functions implies that $f = f^*$, thus model (4.10) is identifiable. ■

The proof of the asymptotic normality of $\hat{\boldsymbol{\eta}}$ for $I \to \infty$ and J fixed is exactly the same as that of the four-parameter model by verifying the sufficient conditions of Huber (1967). Thus, we have

$$\sqrt{I}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}) \xrightarrow{D} N(\mathbf{0}, \boldsymbol{F}^{-1}),$$

where $\boldsymbol{F} = \mathbb{E}[\{\frac{\partial}{\partial \boldsymbol{\eta}} \log p(\boldsymbol{y}_{i\cdot})\}\{\frac{\partial}{\partial \boldsymbol{\eta}} \log p(\boldsymbol{y}_{i\cdot})\}^T]$ is the Fisher Information Matrix for the parameter vector $\boldsymbol{\eta}$. We can calculate the estimates of $\boldsymbol{F}$ by

$$\hat{\boldsymbol{F}} = \frac{1}{I} \sum_{i=1}^{I} \left\{ \frac{\partial}{\partial \boldsymbol{\eta}} \log \hat{p}(\boldsymbol{y}_{i\cdot}) \right\} \left\{ \frac{\partial}{\partial \boldsymbol{\eta}} \log \hat{p}(\boldsymbol{y}_{i\cdot}) \right\}^T,$$

where $\hat{p}$ means that $\boldsymbol{\eta}$ is replaced by $\hat{\boldsymbol{\eta}}$ everywhere.

Chapter 5

Simulations

In this chapter we study the finite-sample behavior of the estimators in both the one-factor and the two-factor four-parameter models via simulations. The one-factor model given by (3.8) ignores the mother factor from the genetic effects; while the two-factor model given by (3.4) and (3.5) accounts for both the father and mother effects using the additive genetic relationship matrix. In this dissertation, we only considered the relationship of half- and full-siblings in the simulated data. More complex relationship structure can be easily captured by the corresponding additive genetic relationship matrix.

5.1 One-factor four-parameter model

We generated balanced data for the one-factor four-parameter model given by (3.8) with various number of fathers $I = 20, 50, 100$ and a fixed number of descendants $J = 10$ per father. The shape function was $f(t) = [16 + 2(t-7)]\mathbb{1}(t \leq 7) + [16 - 0.2(t-7)]\mathbb{1}(t > 7)$. The curve increases with a slope equal to 2 in the early phase, and then decreases with speed 0.2 in the late phase. The father effects $\boldsymbol{\alpha}_i$'s were i.i.d. from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$ with $\boldsymbol{\mu} = (2, 16, -0.2, 7)^T$ and $\boldsymbol{\Sigma}_g = \text{diag}(0.5^3, 1.5, 0.1^2, 0.5)$. The $\boldsymbol{e}_{ij}$'s were i.i.d. from $N(\mathbf{0}, \boldsymbol{\Sigma}_e)$ with $\boldsymbol{\Sigma}_e = \text{diag}(0.3^3, 0.3^3, 0.3^3, 0.3^3)$. The variance of the

random errors σ^2 was set to be 0.1^2.

For each descendant j of the ith father, we computed its responses with the simulated parameter vector $\boldsymbol{\phi}_{ij} = \boldsymbol{\alpha}_i + \boldsymbol{e}_{ij}$ on equally spaced time grid of $m = 6$ and $m = 15$ points in $[2, 12]$ respectively. Each scenario was replicated 300 times. For each simulated data set, we computed the maximum likelihood estimators of the proposed four-parameter model derived in Section 3.3. In the estimation procedure, we used $c = 2$ in the transition function $\text{trn}(x) = e^{cx}/(1 + e^{cx})$. As measures of performance we used the bias, the standard deviation and the root mean square error, defined as follows: for $\boldsymbol{\xi} \in \mathbb{R}^p$, and $\hat{\boldsymbol{\xi}}$ is the estimator, then $\text{bias}(\hat{\boldsymbol{\xi}}) = \|E(\hat{\boldsymbol{\xi}}) - \boldsymbol{\xi}\|$, $\text{sd}(\hat{\boldsymbol{\xi}}) = E\{\|\hat{\boldsymbol{\xi}} - E(\hat{\boldsymbol{\xi}})\|^2\}^{1/2}$, and $\text{rmse}(\hat{\boldsymbol{\xi}}) = \{\text{bias}^2(\hat{\boldsymbol{\xi}}) + \text{sd}^2(\hat{\boldsymbol{\xi}})\}^{1/2}$. The estimation errors for each scenario are reported in Table 5.1.

Table 5.1: Simulation Results. Bias, standard deviation and root mean square error for estimators of one-factor four-parameter model.

	m=6								
	I=20			I=50			I=100		
	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse
$\hat{\boldsymbol{\mu}}$	0.0090	0.3349	0.3439	0.0157	0.2066	0.2223	0.0147	0.1514	0.1661
$\hat{\boldsymbol{\Sigma}}_g$	0.0594	0.5036	0.5630	0.0266	0.3296	0.3562	0.0317	0.2176	0.2493
$\hat{\boldsymbol{\Sigma}}_e$	0.0063	0.0078	0.0141	0.0051	0.0046	0.0097	0.0050	0.0023	0.0072
	m=15								
	I=20			I=50			I=100		
	bias	sd	rmse	bias	sd	rmse	bias	sd	rmse
$\hat{\boldsymbol{\mu}}$	0.0064	0.3359	0.3422	0.0058	0.2142	0.2200	0.0043	0.1556	0.1599
$\hat{\boldsymbol{\Sigma}}_g$	0.0676	0.4766	0.5441	0.0594	0.3065	0.3659	0.0307	0.2208	0.2515
$\hat{\boldsymbol{\Sigma}}_e$	0.0047	0.0058	0.0104	0.0049	0.0033	0.0082	0.0050	0.0023	0.0074

As we see from Table 5.1, the standard deviation and the root mean square error decrease as I becomes larger, which is to be expected. Besides, the estimation errors from data sampled on 6 points are almost identical as those from 15 time points. This implies, the proposed model can provide good estimation even with few observations on each subject.

5.2 Two-factor four-parameter model

In this section, we study the finite-sample behavior of the estimators in the two-factor four-parameter model defined by (3.4) and (3.5). The structure of each pedigree was designed as follows: there were J different mothers nested within each father i ($i = 1, \cdots, I$), and for each pair of father and mother they had K descendants. Moreover, all mothers associated with one father were different from those of another. Therefore, there were $n = I \times J \times K$ individuals in each simulated data set, and three types of relationship exist between subjects: independent individuals, half-siblings, and full-siblings. We then generated balanced data with various combinations of I, J, and K.

The shape function was the same as that in the one-factor four-parameter model, $f(t) = [16+2(t-7)]\mathbb{1}(t \leq 7)+[16-0.2(t-7)]\mathbb{1}(t > 7)$. In the two-factor model, both the father and mother factors contribute to the genetic effects. The genetic effects $\{\boldsymbol{g}_i\}$ were normally distributed from $N(\boldsymbol{\mu}, \boldsymbol{\Sigma}_g)$ with $\boldsymbol{\mu} = (2, 16, -0.2, 7)^T$ and $\boldsymbol{\Sigma}_g = \text{diag}(0.5^3, 1.5, 0.1^2, 0.5)$. Since $\{\boldsymbol{g}_i\}$ were genetically correlated, they were generated from $\boldsymbol{g} \sim N(\tilde{\boldsymbol{\mu}}, \boldsymbol{A} \otimes \boldsymbol{\Sigma}_g)$ with $\boldsymbol{g} = (\boldsymbol{g}_1^T, \cdots, \boldsymbol{g}_n^T)^T$ and $\tilde{\boldsymbol{\mu}} = (\boldsymbol{\mu}^T, \cdots, \boldsymbol{\mu}^T)^T$. The additive genetic relationship matrix $\boldsymbol{A}$ was calculated by the method discussed in Section 2.5 using the identity information of all subjects. The environmental effects $\{\boldsymbol{e}_i\}$ were i.i.d. from $N(\boldsymbol{0}, \boldsymbol{\Sigma}_e)$ with $\boldsymbol{\Sigma}_e = \text{diag}(0.3^3, 0.3^3, 0.3^3, 0.3^3)$. The variance of the random errors σ^2 was set to be 0.1^2.

We used $I = 5, 10$, $J = 3, 5$ and $K = 5, 10$, which produced eight possible combinations of I, J and K in total. Each scenario was replicated 300 times. For each subject i, we computed its responses with the simulated parameter vector $\boldsymbol{\phi}_i = \boldsymbol{g}_i + \boldsymbol{e}_i$ on an equally spaced time grid of 6 points in $[2, 12]$. For each simulated data set, we calculated the maximum likelihood estimators derived in Section 3.2, and the estimation errors for each scenario are given in Table 5.2.

We observe from Table 5.2 that the standard deviation and root mean square error

Table 5.2: Simulation Results. Bias, standard deviation and root mean square error for estimators of two-factor four-parameter model.

	I=5, J=3, K=5			I=5, J=3, K=10		
	bias	sd	rmse	bias	sd	rmse
$\hat{\boldsymbol{\mu}}$	0.0078	0.4060	0.4137	0.0184	0.3850	0.4034
$\hat{\boldsymbol{\Sigma}}_g$	0.0445	0.1315	0.1760	0.0421	0.0438	0.0859
$\hat{\boldsymbol{\Sigma}}_e$	0.3061	0.4386	0.7447	0.2677	0.4203	0.6880
	I=5, J=5, K=5			I=5, J=5, K=10		
$\hat{\boldsymbol{\mu}}$	0.0240	0.3701	0.3941	0.0141	0.3575	0.3716
$\hat{\boldsymbol{\Sigma}}_g$	0.0324	0.0863	0.1187	0.0434	0.0428	0.0862
$\hat{\boldsymbol{\Sigma}}_e$	0.2395	0.4095	0.6490	0.2568	0.3710	0.6279
	I=10, J=3, K=5			I=10, J=3, K=10		
$\hat{\boldsymbol{\mu}}$	0.0065	0.2797	0.2862	0.0248	0.2688	0.2936
$\hat{\boldsymbol{\Sigma}}_g$	0.0349	0.0710	0.1059	0.0455	0.0370	0.0825
$\hat{\boldsymbol{\Sigma}}_e$	0.1923	0.3687	0.5609	0.2077	0.3555	0.5632
	I=10, J=5, K=5			I=10, J=5, K=10		
$\hat{\boldsymbol{\mu}}$	0.0090	0.2518	0.2608	0.0109	0.2656	0.2765
$\hat{\boldsymbol{\Sigma}}_g$	0.0410	0.0491	0.0901	0.0493	0.0253	0.0746
$\hat{\boldsymbol{\Sigma}}_e$	0.1595	0.3303	0.4898	0.1031	0.2919	0.3950

decrease as I, J and K increase. For the estimates of $\boldsymbol{\mu}$, the contribution of I to the improvement is relatively larger than those of J and K. This is sensible because the overall mean is largely driven by the genetic factor of different families. For relatively small values of I and J ($I = 5, J = 3$), the improvement of the root mean square error of the estimates of $\boldsymbol{\Sigma}_g$ is more from the increase of K; while the improvement of using large K is getting smaller when I and J become larger. This implies it would be better to have more information that can reflect the genetic correlation between subjects by using more fathers and more mothers. For the estimates of $\boldsymbol{\Sigma}_e$, increasing the number of I has more effects in the improvement of the standard error and the root mean square error than those given by increasing J and K. Therefore, taking all into account, we would recommend to use a relatively large number of families I to obtain better estimates of the parameters.

Chapter 6

Example: Flour-beetle growth data

In this chapter we apply the two-factor four-parameter model to the flour-beetle growth data. The whole data set consists of 1124 insects, representing 134 full-sib families nested with 29 half-sib families. The number of mother nested within each father is between 3 and 6. The number of descendants of each pair of father and mother varies between 1 and 10, with a median of 8. The data structure is illustrated in Table 6.1.

Father ID	Mother ID	# of Descendants
1	1	5
1	2	3
1	3	5
1	4	9
1	5	9
1	6	8
2	7	3
2	8	7
2	9	5
2	10	1
⋮	⋮	⋮

Table 6.1: Beetle-growth data structure.

In general, there are four phases in the life cycle of flour-beetles: egg, larva, pupa, and adult. The flour-beetle growth data are collected from the day of egg hatching

up to the day of eclosion from pupa phase. For each flour-beetle, its body mass was measured about every 3 days in the early phase of the growth curve and more frequently, usually every day, in the late phase. Thus, there are 5 to 13 measurements per beetle. For better visualization, we only plotted part of the raw data and the log-transformed data and they are displayed in Figure 6.1.

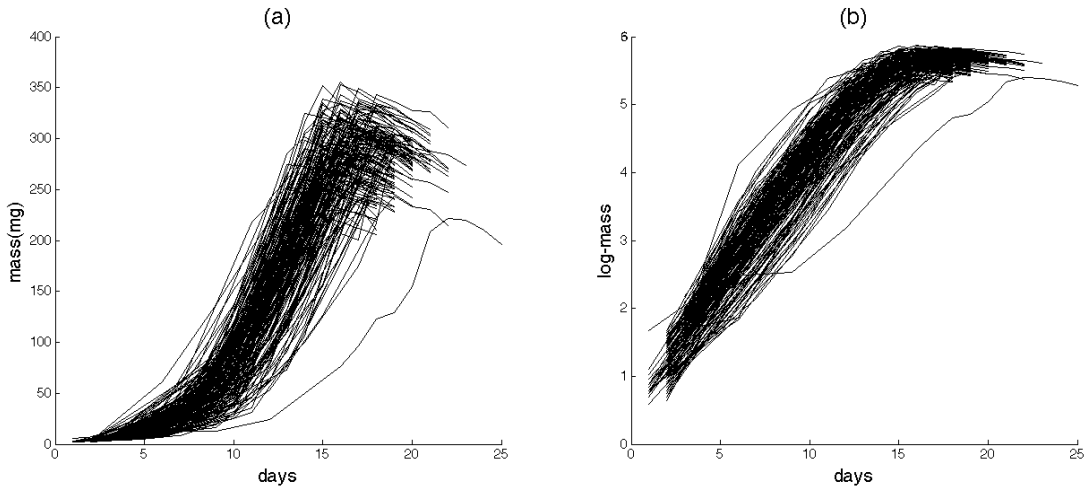

Figure 6.1: Beetle-growth curves. (a) Raw mass trajectories; (b) log-mass trajectories.

As we can see from Figure 6.1, both the starting and end points of curves are irregular and the timing of beetles reaching pupation period varies from one to another. The loss of body mass from larva to pupa is because the beetle stops eating and starts searching for a place to pupate. In addition, different beetles have different birth masses and growth rates. All these physical traits have a high impact on the individual fitness, including mating probability, mating success, ovariole number and fecundity (Irwin and Carter, 2013), and the genetic factor contained in these physical traits represents plenty of useful information in the development and evolution of the flour-beetles. Figure 6.2 exhibits the variation of these traits between different families: the main difference is on the birth mass, the maximum mass, and the turning point; while the growth rates before and after the turning points seem to vary less.

We apply the four-parameter shape-invariant model (3.3) to the flour-beetle growth

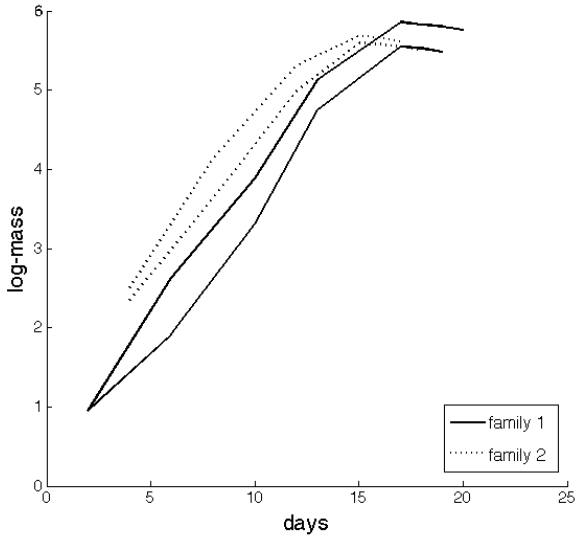

Figure 6.2: Beetle-growth trajectories (log-mass) from two different families.

data. The approximate maximum likelihood estimates are:

$$\hat{\boldsymbol{\mu}} = (0.3569, 5.6444, -0.0465, 14.2167),$$

$$\hat{\boldsymbol{\Sigma}}_g = \begin{pmatrix} 0.0001 & 0.0009 & -0.0004 & -0.0048 \\ & 0.0135 & -0.0048 & -0.0513 \\ & & 0.0024 & 0.0247 \\ & & & 0.2756 \end{pmatrix},$$

$$\hat{\boldsymbol{\Sigma}}_e = \begin{pmatrix} 0.0000 & 0.0000 & 0.0001 & -0.0020 \\ & 0.0001 & 0.0001 & -0.0030 \\ & & 0.0002 & -0.0046 \\ & & & 0.1349 \end{pmatrix},$$

$$\hat{\sigma}^2 = 0.0161,$$

All the parameters in $\boldsymbol{\phi}_i$ are highly correlated, this is because they uniquely define

the piece-wise linear structure of the two growth modes. As expected, the genetic effect accounts more in the variances of the four parameters than the environmental effect. The location of the turning point has much larger variance than the peck body mass and the growth rates before and after the turning point, which is consistent as we saw from Figures 6.1 (b). It implies that the starting time of the pupation period is affected more by the genetic effect, and the larvae tend to have a target peak body mass before starting to pupate. This result may help the biologists perform selection for further investigation. The fitted curves $\hat{f}_{ij}(t)$ of the random chosen

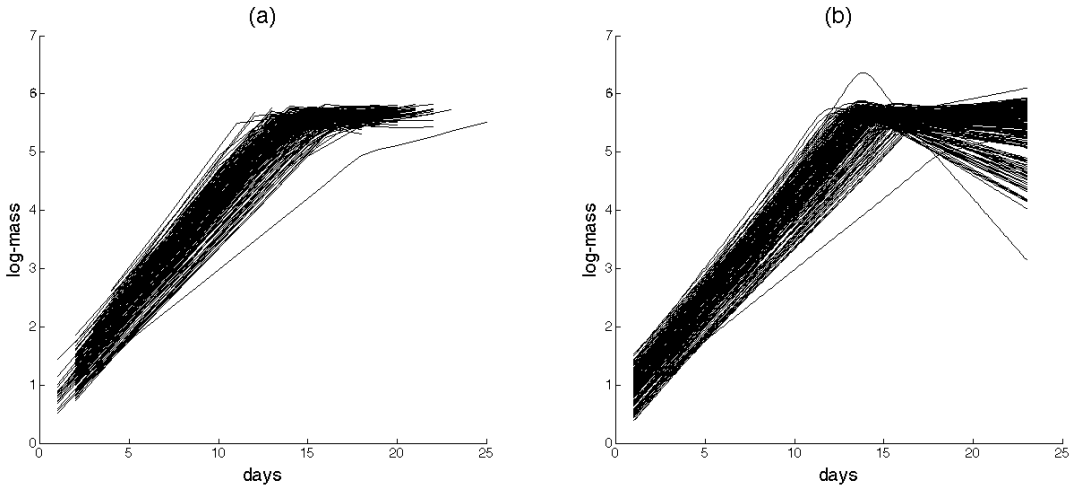

Figure 6.3: Fitted flour-beetle growth trajectories. (a) Fitted curves using time grids of the raw data; (b) Fitted curves using common and finer time grids.

sample are shown in Figure 6.3 which demonstrates the overall performance of this model. Moreover, as illustrated in Figure 6.4, the estimate of the turning point by our four-parameter model is relatively smaller than what the curves present. The reason is that we use a piece-wise linear shape function and the nature of the piece-wise linear functions tend to locate a earlier time of the turning point.

For comparison purpose, we also applied the four-parameter model using $c = 4$ in the transition function. Compared to the results from $c = 2$, the changes of estimates for parameters in $\boldsymbol{\mu}$ and σ^2 is relatively small with the most relative change 8.53% given by $\hat{\phi}_{i3}$ (the estimates of the growth rate after the turning point). However, the

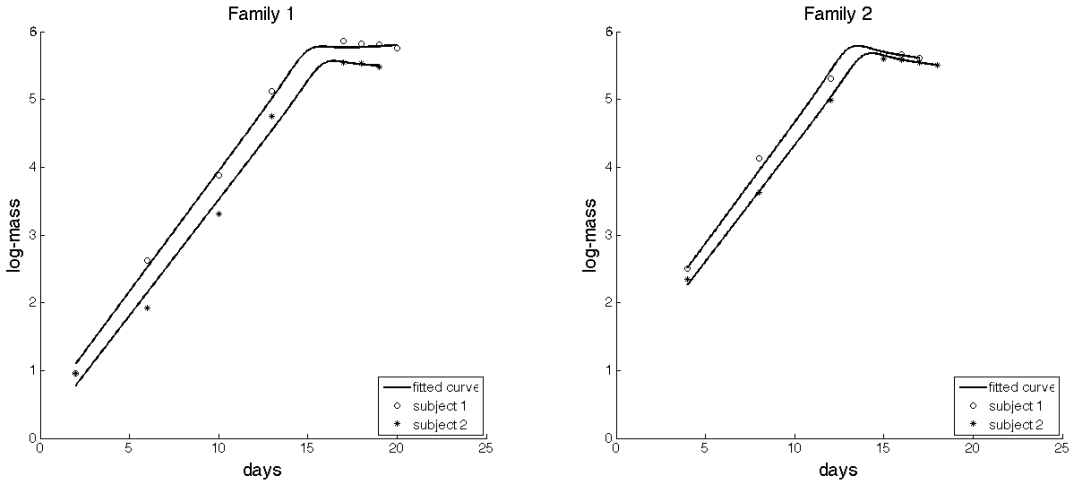

Figure 6.4: Fitted trajectories of beetles of two different families: fitted growth trajectories (solid line), growth trajectories of raw data (circle and star).

relative changes in the estimates of $\mathbf{\Sigma}_g$ and $\mathbf{\Sigma}_e$ is larger with the largest change of 23%. The reason is that, for the parameters that are not significantly different from zero, using different values of c in the transition function always causes larger change in the estimates (Morrell *et al.*, 1995).

Chapter 7

Conclusion

In this dissertation, we proposed shape-invariant models that can fit various kinds of nonlinear relationship. They explicitly address the amplitude and phase variation and make the unobservable genetic and environmental effects estimable. These models are very flexible especially when the data structure is complex and the knowledge of modeling building is limited. The four-parameter shape-invariant model is based on the flour-beetle growth data, but it can be generalized to other curve patterns. For the transition function, it is helpful to plot the sample data to see if the change around the turning points is sharp or relative slow to adjust the value of c. From our experience, $c = 2$ is usually a good choice. The general self modeling regression provides a more sophisticated way for the common shape function. To overcome the differentiability problem, one can use differentiable families for the time-warping functions. In the simulation, we studied the correlation of full and half siblings. One can simply modify the additive genetic relationship matrix to accommodate other complex correlation structures.

Estimation of these shape-invariant models can be conducted using Laplace's approximation. We only implemented the four-parameter model in Matlab, and the implementation of the general model is of future interest. One issue of this procedure

is the long computational time when the sample size becomes large. Possible solutions are using the basic programming languages, like C/C++ or Fortran, or developing other efficient approximation techniques. For the model diagnostics, one can perform the residual plots for the random errors and the environmental random effects. As mentioned in Lindstrom (1995), to check the assumption that all the individuals are generated from the common shape function, one can plot the shifted/scaled response variable versus the shifted/scaled covariates. If the model fits the data, the plot for all individuals will lie approximately along a single curve indicative of the common shape function.

Bibliography

[1] Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*. New York: Wiley.

[2] Bürger, R. (2000). *The Mathematical Theory of Selection, Recombination, and Mutation*. New York: Wiley.

[3] Davidian, M. and Gallant, A. R. (1992). *Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine*. Journal of Pharmacokinetics and Biopharmaceutics, Vol. 20, No. 5, pp. 529-556.

[4] Davidian, M. and Giltinan, D. M. (1995). *Nonlinear Models for Repeated Measurement Data*. London: Chapman & Hall.

[5] de Boor, C. (1978). *A Practical Guide to Splines*. New York: Springer-Verlag.

[6] Eubank, R. L. (1999). *Nonparametric Regression and Spline Smoothing*. New York: Marcel Dekker.

[7] Falconer, D. S. (1981). *Introduction to Quantitative Genetics*. London: Longman Press.

[8] Fritsch, F. N. and Carlson, R. E. (1980). *Monotone piecewise cubic interpolation*. SIAM Journal on Numerical Analysis, Vol. 17, No. 2, pp. 238-246.

[9] Gervini, D. and Gasser, T. (2003). *Alignment of functional data via self-modeling warping functions*.

[10] Gervini, D. and Gasser, T. (2004). *Selfmodelling warping functions*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 66, No. 4, pp. 959-971.

[11] Gervini, D. and Carter, P. A. (2014). *Warped functional analysis of variance*. Biometrics, Vol. 70, No. 3, pp. 526-535.

[12] Geweke, J. (1989). *Bayesian inference in econometric models using Monte Carlo integration*. Econometrica: Journal of the Econometric Society, 1317-1339.

[13] Heckman, N. E. (2003). *Functional data analysis in evolutionary biology*. Recent Advances and Trends in Nonparametric Statistics, 49-60.

[14] Henderson, C. R. (1975). *Use of relationships among sires to increase accuracy of sire evaluation*. Journal of Dairy Science Vol. 58, No. 11, pp. 1731-1738.

[15] Henderson, C. R. (1976). *A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values*. Biometrics, Vol. 32, No. 1, pp. 69-83.

[16] Huber, P. J. (1967). *The behavior of maximum likelihood estimates under nonstandard conditions*. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, No. 1.

[17] Hürtgen, H. and Gervini, D. (2009). *Semiparametric shape-invariant models for periodic data*. Journal of Applied Statistics, Vol. 36, No. 10, pp. 1055-1065.

[18] Irwin, K. K. and Carter, P. A. (2013). *Constraints on the evolution of function-valued traits: a study of growth in Tribolium castaneum*. Journal of Evolutionary Biology, Vol. 26, No. 12, pp. 2633-2643.

[19] Jupp, D. L. (1978). *Approximation to data by splines with free knots*. SIAM Journal on Numerical Analysis, Vol. 15, No. 2, pp. 328-343.

[20] Kneip, A. and Gasser, T. (1988). *Convergence and consistency results for self-modeling nonlinear regression*. The Annals of Statistics, Vol. 16, No. 1, pp. 82-112.

[21] Kirkpatrick, M. and Heckman, N. E. (1989). *A quantitative genetic model for growth, shape, reaction norms, and other infinite dimensional characters*. Journal of mathematical biology, Vol. 27, No. 4, pp. 429-450.

[22] Laird, N. M. and Ware, J. H. (1982). *Random-effects models for longitudinal data*. Biometrics, Vol. 38, No. 4, pp. 963-974.

[23] Lawton, W. H., Sylvestre, E. A. and Maggio, M. S. (1972). *Self-modeling nonlinear regression*. Technometrics, Vol. 14, No. 3, pp. 513-532.

[24] Lindstrom, M. J. and Bates, B. M. (1990). *Nonlinear mixed effects models for repeated-measures data*. Biometrics, Vol. 46, No. 3, pp. 673-687

[25] Lindstrom, M. J. (1995). *Self-modeling with random shift and scale parameters and a free-knot spline shape function*. Statistics in Medicine, 14, 2009-2021.

[26] Marubini, E., Resele, L. F., Tanner, J. M. and Whitehouse, R. H. (1972). *The fit of Gompertz and logistic curves to longitudinal data during adolescence on height, sitting height and biacromial diameter in boys and girls of the Harpenden Growth Study*. Human Biology, 511-523.

[27] Morrell, C. H., Pearson, J. D., Carter, H. B. and Brant, L. J. (1995). *Estimating unknown transition times using a piecewise nonlinear mixed-effects model in men with prostate cancer*. Journal of the American Statistical Association, Vol. 90, No. 429, pp. 45-53.

[28] Mrode, R. A. (2014). *Linear Models for the Prediction of Animal Breeding Values*. Cambridge: CABI.

[29] Pinheiro, J. C. and Bates, D. M. (1995). *Approximations to the log-likelihood function in the nonlinear mixed-effects model*. Journal of computational and Graphical Statistics, Vol. 4, No. 1, pp. 12-35.

[30] Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. New York: Springer.

[31] Ramsay, J. O., and Li, X. (1998). *Curve registration*. Journal of the Royal Statistical Society: Series B (Statistical Methodology), Vol. 60, No. 2, pp. 351-363.

[32] Telesca, D. and Inoue, L.Y.T. (2008). *Bayesian hierarchical curve registration*. Journal of the American Statistical Association, Vol. 103, No.481, pp. 328-339.

[33] Tierney, L. and Kadane, J. B. (1986). *Accurate approximations for posterior moments and marginal densities*. Journal of the American Statistical Association, Vol. 81, No. 393, pp. 82-86.

[34] Wald, A. (1949). *Note on the consistency of the maximum likelihood estimate*. The Annals of Mathematical Statistics, Vol. 20, No. 4, pp. 595-601.

[35] Wang, K. and Gasser, T. (1999). *Synchronizing sample curves nonparametrically*. Annals of Statistics, Vol. 27, No. 2, pp. 439-460.

[36] Wright, S. (1922). *Coefficients of inbreeding and relationship*. American Naturalist, 330-338.

Appendix

This appendix covers the MATLAB routines used in simulations and case study of the four-parameter shape invariant model of this dissertation.

Main function:

```
function out = FourParaModel(indivID, t, y, invA, para0, rd0)
% indivID  (N X 1)  vector of IDs of all subjects
% t        (N X 1)  vector of time grids of all subjects
% y        (N X 1)  vector of masses of all subjects
% invA     (n X n)  inverse of the genetic relationship matrix
% para0     struct  a struct consists of the initial values of the
%                   parameters
% rd0      (8n X 1)  initial values of genetic and environment effects
ID = unique(indivID);
n = length(ID);
N = length(indivID);
cmu = 1/(ones(1,n)*invA*ones(n,1));
cnt = 1;itermax = 20;
err = 1;tol = 1e-3;
convflag = 0;
while err>=tol && cnt<=itermax
    invGam = inv(para0.Gam);
```

```
invLam = inv(para0.Lam);
[rd,exitflag,fv] = minObj(invA, invGam, invLam, para0.mu,...
        para0.s2, indivID, t, y, rd0);
G = reshape(rd(1:4*n),4,n);
V = bsxfun(@minus,G,para0.mu);
E = reshape(rd(4*n+1:end),4,n);
para.Lam = zeros(4);
para.s2 = 0;
for k = 1:n
    idx1 = 4*(k-1)+1:4*k;
    idx2 = 4*(n+k-1)+1:4*(n+k);
    difmass = y(indivID==ID(k))-h(t(indivID==ID(k)),...
       rd(idx1)+rd(idx2));
    para.s2 = para.s2 + (difmass'*difmass);
end
para.s2 = para.s2/N;
para.Lam = (E*E')/n;
para.Gam = (V*invA*V')/n;
para.mu = cmu*(G*invA*ones(n,1));
erro.s2 = para.s2-para0.s2;
erro.mu = para.mu-para0.mu;
erro.Gam = para.Gam-para0.Gam;
erro.Lam = para.Lam-para0.Lam;
para0 = para;
rd0 = rd;
err = max([max(abs(erro.mu)./abs(para.mu)),...
  abs(erro.s2)/abs(para.s2),...
```

```
        max(abs(erro.Gam)./abs(para.Gam)),...
        max(abs(erro.Lam)./abs(para.Lam))]);
    fprintf('\nLaplace Iteration %f with Error %f ', cnt, err);
    cnt = cnt + 1;
end
if err<tol
    convflag = 1;
end
out.para = para0;
out.rd = rd0;
out.convflag = convflag;
end
```

Auxiliary functions:

```
function [para,exitflag,fval] = minObj(invA, invGam, invLam, mu,...
 s2, indivID, t, y, para0)
% find the minimum point of the objective function
maxI = 1000*length(para0);
options = optimoptions(@fminunc,'Algorithm','quasi-newton','GradObj',...
 'on','TolFun',0.01,'MaxIter',maxI,'display','off');
[para,fval,exitflag] = fminunc(@(para)gradObj(invA, invGam, invLam,...
 mu, s2, indivID, t, y, para),para0,options);
end

function [fval,grad,hess] = gradObj(invA, invGam, invLam, mu, s2,...
 indivID, t, y, rd)
% find the gradient of the objective function
% D  (4n x 8n)   Design matrix
```

```
% t  (N  x 1)    Vector of time grids
% rd (8n x 1)    Vector of random effects
ID = unique(indivID);
n = length(ID);
N = length(t);
D = [eye(4*n),eye(4*n)];
rdnew = D*rd;
bm = zeros(size(rd));
bm(1:4*n) = kron(ones(n,1),mu);
invSig = [kron(invA,invGam), zeros(4*n);
          zeros(4*n),kron(eye(n),invLam)];
hy = zeros(size(y));
grdh = zeros(8*n,N);
cnt = 0;
for k = 1:n
    tk = t(indivID==ID(k));
    m = length(tk);
    idx1 = 4*(k-1)+1:4*k;
    idx2 = cnt+1:cnt+m;
    grdh(idx1,idx2) = parh(tk,rdnew(idx1));
    hy(idx2) = h(tk,rdnew(idx1));
    cnt = cnt + m;
end
grdh(4*n+1:end,:) = grdh(1:4*n,:);
grad = -2*(grdh*(y-hy))/s2 + 2*invSig*(rd-bm);
hess = 2*(grad*grad')/s2 + 2*invSig;
fval = (y-hy)'*(y-hy)/s2 + (rd-bm)'*invSig*(rd-bm);
```

```
end

function ph = parh(t,p)
% calculate the partial derivative of h
% t is a scalar or vector
% p is a parameter vector
ph = zeros(4,length(t));
ph(1,:) = (t-p(4)).*trn(p(4)-t);
ph(2,:) = trn(p(4)-t)+trn(t-p(4));
ph(3,:) = (t-p(4)).*trn(t-p(4));
ph(4,:) = -p(1)*trn(p(4)-t)+ptrn(p(4)-t).*h1(t,p)-...
    p(3)*trn(t-p(4))-ptrn(t-p(4)).*h2(t,p);
end

function y = h(t,p)
% calculate the response of the curve
% t is a scalar or vector
% p is a parameter vector
y = h1(t,p).*trn(p(4)-t) + h2(t,p).*trn(t-p(4));
end

function y = h1(t,p)
% the first piece of the curve
% t is a scalar or vector
% p is a parameter vector
y = p(2) + p(1)*(t-p(4));
end
```

```
function y = h2(t,p)
% the second piece of the curve
% t is a scalar or vector
% p is a parameter vector
y = p(2) + p(3)*(t-p(4));
end

function y = trn(x)
% calculate the transition function
% x is a scalar or vector
c = 2;
y = exp(c*x)./(1+exp(c*x));
end

function py = ptrn(x)
% find the partial derivative of the transition function
% x is a scalar or vector
c = 2;
py = c*exp(c*x)./(1+exp(c*x)).^2;
end
```

CURRICULUM VITAE

WEN YANG

2015 Ph.D. in Mathematics, University of Wisconsin, Milwaukee, WI.

Dissertation: Shape-Invariant Models for Non-Independent Functional Data.

Advisor: Daniel Gervini.

2008 B.S. in Applied Mathematics, Nankai University, Tianjin, China.

WORK EXPERIENCE

- Statistics Software Development Intern, May 2014-Aug 2014, The MathWorks Inc., Natick, MA.
- Research Assistant, Jan 2014-May 2014, , Department of Mathematical Sciences, University of Wisconsin, Milwaukee, WI.
- Teaching Assistant, Sep 2009-May 2015, Department of Mathematical Sciences, University of Wisconsin, Milwaukee, WI.

SKILLS AND CERTIFICATES

- Knowledge in Matlab, R, SAS, and C++ programming.
- SAS Certified Base/Advanced Programmer and Business Analyst for SAS9.
- Passed Society of Actuaries Exam P.