Abstract

CHISWELL, KAREN ELIZABETH. Model Diagnostics for the Nonlinear Mixed Effects Model with Balanced Longitudinal Data. (Under the direction of Dr. John Monahan.)

Longitudinal data, where the response for each experimental unit is measured repeatedly on several occasions, are common in biomedical (e.g., pharmacokinetic) and agricultural (e.g., growth and yield) studies. When each unit is measured at the same time points, the data are termed balanced. The methods developed in this dissertation assume balanced longitudinal data with a normally distributed response variable.

The nonlinear mixed effects model (NLMM) provides a useful statistical framework for characterizing longitudinal data when a nonlinear regression function describes the trend for each experimental unit or subject. In a simple NLMM with two stages, a nonlinear model describes the expected trajectory for each subject at the first stage. Coefficients in the first stage model are specific to each subject. At the second stage these subject-specific coefficients are modeled as random variables (random effects) from a probability distribution. Within-subject variability (e.g., due to measurement errors or fluctuations in within-subject response) is modeled at the first stage. Variation between subjects is characterized by the second stage model for the random effects, some of which may be trivial (the coefficient value is the same for all subjects).

Our methods explore the structure of variation in the longitudinal data using traditional tools of multivariate analysis (principal component analysis, factor analysis, and multivariate regression). Assuming that preliminary estimates of the NLMM model parameters are available, a first-order linearization of the subject specific model allows us to use the multivariate results to check assumptions about the variance structure defined by the NLMM. In particular, we propose methods for identifying the non-trivial random effects, for assessing the presence of within-subject heteroscedasticity (non-constant variance), and for testing evidence of within-subject autocorrelation.

In simulation studies our methods demonstrated high power to correctly identify random effects, to detect heteroscedasticity, and to find strong AR(1) structure in within-subject residuals. We applied our methods to two examples. Data in the first example were collected from a gas uptake experiment in rats. Our methods identified significant evidence of within-subject heteroscedasticity, and suggested that between-subject variation could be explained by random effects on the two rate parameters in an open two-compartment model, or by the V_{max} metabolic parameter and rat body weight in the PBPK model. In the second example, pharmacokinetic data were collected from human subjects dosed orally with theophylline. Our analysis identified random effects on the absorption rate and clearance parameters of the open one-compartment model, and suggested heteroscedasticity in within-subject errors.

Model Diagnostics for the Nonlinear Mixed Effects Model with Balanced Longitudinal Data

by

Karen Elizabeth Chiswell

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2007

APPROVED BY:

Dr. M. Davidian	Dr. H. Zhang
Dr. H. T. Banks	Dr. J. Monahan Chair of Advisory Committee

Dedication

To my loving family, Chiswells and Faulkners.

Biography

Karen Chiswell was born in Johannesburg, South Africa, in 1972. She graduated from Roosevelt High School, Johannesburg, in 1990. In 1995 she received her Bachelor of Science in Agriculture degree, majoring in biometry, from the University of Natal (Pietermaritzburg, South Africa). After graduation, Karen worked as a forest biometrician for Mondi Forests, a forestry company with headquarters in Pietermaritzburg. In 2000, Karen moved to Raleigh, North Carolina, to pursue graduate studies in statistics at North Carolina State University. She received her Master of Statistics degree in 2002. In June of that year, Karen married Gary Faulkner.

During her graduate studies, Karen taught introductory statistics to undergraduates, participated in activities of the Statistical and Applied Mathematical Sciences Institute (SAMSI), and worked as a statistician in the Clinical Trials group at Duke Clinical Research Institute in Durham, North Carolina. In 2006 Karen began working as a preclinical statistician at the pharmaceutical company, GlaxoSmithKline, in the Research Triangle Park, North Carolina. In 2007 Karen completed the requirements for a Ph.D. in the Department of Statistics at North Carolina State University. After graduation, she will continue working at GlaxoSmithKline.

Acknowledgements

Thank you to my adviser Dr. John Monahan for his direction, assistance, and encouragement. With him as my guide, I have learned much about statistics, research, and myself. I am grateful to my committee members, Drs. Marie Davidian, H. T. Banks, and Helen Zhang for their instruction during my graduate studies, and their attention to my research. Thank you to Drs. Woody Setzer and Hugh Barton of the Environmental Protection Agency, who drew my attention to the CCl_4 gas uptake experiment and the PBPK model as an interesting application.

I appreciate the wonderful opportunities provided to me as a student in the Department of Statistics at NC State University: a rich selection of course work, trusty computing resources, the chance to instruct a statistics course, and encouragement to involve myself in the life of the department. Thank you in particular to Dr. Sastry Pantula, Adrian Blue, Janice Gaddy, and Terry Byron for greasing the wheels of the department and making these things possible. I am grateful for financial support and intellectual stimulation that I gained from involvement with the Statistical and Applied Mathematical Sciences Institute (SAMSI).

I am indebted to many statisticians and teachers who encouraged me to study and keep studying statistics when I was a student and fledgling statistician in South Africa, among them: Peter Clarke, Harvey Dicks, Yvonne Fletcher, Eric Grau, Linda Haines, Heyns Kotze and Jonathan Levin. To Peter Gardiner and Bill Esler, with whom I was lucky enough to work while at Mondi Forests: thank you for your love of science, belief in statistics and statisticians, and your wisdom about all things corporate. I am grateful for the kind and generous spirit of Karen Pieper at Duke Clinical Research Institute who gave me many opportunities to grow as a statistician. I appreciate the patience and support of the Statistical Sciences and Discovery Statistics groups at GlaxoSmithKline. In particular, I thank Mandy Bergquist and Dave Cooper for helping me focus on completing my dissertation. Special thanks are due to Steve Novick for showing me how to tackle R lattice graphics, and for his companionship and advice dispensed

on and off the jogging trail.

I am grateful for the friendship of many, and in particular for the support and encouragement of Joe Boyer, Younan Chen, Alta Dreyer, Kristen Foley, Yuliya Lokhnygina, Tena McQueen, Amy Nail, Jasmine Mathias, and Ray Webster. These acknowledgements would be incomplete if I did not mention my constant companions, Hannah and Lyra, who delight in each minute of the day, whether it is spent running, eating, or sleeping. And finally... to the people who have had the most profound influence on my life: to my parents, Caroline and Peter Chiswell, and to my brother, Matt, thank you for your unconditional love and support; and to my husband, Gary Faulkner, I am eternally grateful for your love and belief in me, and for cooking beans and potatoes.

Table of Contents

List of Tables

List of Figures

Nomenclature and List of Symbols

We use italic roman characters such as b, u and y to denote random variables, except in some cases where the context makes it clear that the realizations of these random variables are intended. Bold face italic characters such as $\boldsymbol{b}$, $\boldsymbol{u}$ and $\boldsymbol{y}$ denote random (column) vectors, while we use bold face uppercase non-italic characters such as $\mathbf{V}$ or $\mathbf{E}$ for matrices. The subscript ^{T} denotes the transpose of a vector or matrix. The letter t will generally indicate time. In most cases we assume that time (e.g., time when an observation is made) is not random, but fixed by design. When used in a context like $f(t)$, letters such as f denote a known function of t. Greek characters such as β and σ are used to denote parameters in a statistical model. When used in a context like $\mu(t)$, Greek characters such as μ denote an unknown function of t.

The following symbols are used frequently in this dissertation with the same meaning throughout.

$\mathbb{E}$	-	Expectation: $\mathbb{E}(y) = \int y \, \mathrm{d}F$ where F is the distribution of y
$\mathbb{V}$	-	Variance: $\mathbb{V}(y) = \mathbb{E}\{y - \mathbb{E}(y)\}^2$
y_{ij}	-	Random variable denoting the response on the i-th subject at time t_j
n	-	Number of subjects that are observed
p	-	Number of observations made on a subject
$\boldsymbol{y}_i$	-	Random vector containing the p responses on subject i
$\mathbf{Y}$	-	$n \times p$ data matrix with i-th row $\boldsymbol{y}_i^T$
$\boldsymbol{\mu}$	-	Vector of means, e.g., $\boldsymbol{\mu} = \mathbb{E}(\boldsymbol{y}_i)$
$\boldsymbol{\Sigma}$	-	Covariance matrix, e.g., $\boldsymbol{\Sigma} = \mathbb{V}(\boldsymbol{y}_i)$
$\boldsymbol{\Lambda}_k$	-	$p \times k$ matrix of common factor loadings
$\boldsymbol{\Psi}$	-	$p \times p$ diagonal matrix of specific variances
f_β	-	Partial derivative of function f with respect to parameter β
$\mathbf{F}_\beta$	-	$p \times q$ matrix, the columns of which are realizations of $f_{\beta^{(1)}}, \ldots, f_{\beta^{(q)}}$
H_0	-	Null hypothesis
H_1	-	Alternative hypothesis

Chapter 1

Introduction

In many biological studies, repeated measurements are made on a set of experimental units over time. For example, in a forestry experiment to compare the effect of spacing between adjacent seedlings on tree growth and yield, the diameters of trees in each plot are measured annually over a period of years. In an early-phase clinical trial to determine how a new drug is absorbed, distributed, metabolized and eliminated, several human subjects receive a dose of the drug. Samples of blood are drawn repeatedly from each subject over a period of hours, and the plasma concentration of the drug at each time point is measured. Such data are referred to as *repeated measures* data, or *longitudinal* data when the repeated measurements are made over time. In this dissertation we focus on longitudinal repeated measures data, but the methods developed could be applied to other types of repeated measures data where the same variable is measured repeatedly on an experimental unit or subject, and the sequence of measurements is determined by an ordered variable such as distance or length.

In the examples given above, the data are *balanced* by design: each experimental unit is measured the same number of times and at the same set of time points. Missing observations due, for example, to a lost blood sample, or fire damage in one corner of the forestry spacing experiment, will disrupt this balance. As the title of this dissertation implies, our methods apply to longitudinal data that are balanced. We also restrict our attention to cases where the response variable measured at each time point is a continuous variable such as concentration, blood pressure, or tree diameter, which can reasonably be assumed to follow a normal distribution.

There are several approaches to the analysis of continuous repeated measures data; for example, multivariate analysis of variance, linear or nonlinear mixed effects models, and functional

data analysis. All methods have to take account of the correlation structure that is induced in the data because the repeated measurements made on one experimental unit are likely to be correlated with one another due to latent (unobserved) subject differences. The choice of analysis method depends to a large extent on the scientific objectives of the study. For example, in the forestry experiment, a parametric growth curve may be a reasonable model for how the diameter of an individual tree changes over time. Forestry planners would be interested in estimating the parameters of the growth curve model by fitting the model to the experiment data, and in seeing how the different spacing treatments are reflected in changes in parameter values. A nonlinear mixed effects model analysis is appropriate for this purpose. For the clinical example, a nonlinear mixed effects model analysis would also be typical.

Rice (2004) compared the functional and the more traditional longitudinal data analysis approaches (e.g,. nonlinear mixed effects models) for analyzing longitudinal data. He viewed functional data analysis as primarily an exploratory approach, whereas mixed model approaches tend to be more inferential. Sometimes the same conceptual model may be implemented in different approaches. For example, one of the structural models proposed by Jennrich and Schluchter (1986) for modeling within-subject covariance structure in a linear mixed model framework, is the factor-analytic model. Multivariate factor analysis (discussed in Chapter 2) fits this same model to describe marginal covariance structure, but does not focus on modeling the expected value of responses, as is usually the primary objective in the mixed model approach. The choice of method also depends on the structure of the data. For example, multivariate methods for repeated measures analysis require balanced data, but mixed effects approaches may be able to handle unbalanced, sparsely-sampled data, depending on the choice of covariance model.

In the nonlinear mixed model analysis, the expected trajectory for an individual subject is modeled with a parametric nonlinear function. Examples of such models are given in Section 1.2 of this chapter. Coefficients in the nonlinear model typically represent quantities with biological meaning such as absorption or elimination rates. The nonlinear mixed effects model approach

provides estimates of typical values of the coefficients for the collected data, as well as estimates of how these coefficients vary between subjects. The estimated between-subject variation in coefficients provides important information for setting safe but effective doses for future clinical testing of the drug. In this dissertation we assume that fitting a nonlinear mixed effects model is the ultimate objective of the analysis. Section 1.3 of this chapter reviews the terminology and notation of a simple two-stage nonlinear mixed effects model.

In his 2002 Fisher Lecture titled "Variances are not always nuisance parameters," Carroll (2003) made the point that understanding the structure of variation in data is very important. His view was that accurately modeling variance structure can be an important goal in itself (e.g., in understanding patient variability in drug response), and may have a large impact on other parts of an analysis (e.g., in predicting the working range of an assay). He included heteroscedasticity in regression (where the variance depends on known factors such as measurement time, or the predicted response), and the selection and modeling of random effects as two important components of variance structure. These components of variation are explicitly dealt with in the two stages of the nonlinear mixed effects model.

Although there remains much interest in estimation and inference for nonlinear mixed effects model parameters, these topics are not addressed in this dissertation. Instead we assume that a preliminary fit of the nonlinear mixed effects model has been made, and that estimates of the coefficients in the nonlinear function are in hand. For example, perhaps some coefficients in the model were set at reasonable guesses, and the remaining coefficients estimated by fitting the model separately to each subject, assuming a simple homoscedastic structure for within-subject variation. Preliminary estimates of mean coefficient values might then be obtained by averaging the subject-specific estimates. We view our methods as providing a set of diagnostic tools for assessing certain assumptions of the nonlinear mixed effects model. Our methods provide guidance for formulating a realistic but parsimonious version of the nonlinear mixed effects model that could be fitted (by state-of-the art estimation methods) after the preliminary analysis. We believe that having tools to assist with formulating the model (e.g., by choosing

which coefficients should be specified as random) is valuable because state-or-the-art methods, such as those that numerically integrate out the random effects to construct a numerical approximation of the marginal likelihood, are computationally expensive, or may fail, when there is an unnecessarily large number of random effects.

The methods developed in this dissertation examine three model specification issues for the nonlinear mixed effects model: selection of random effects, identification of within-subject heteroscedasticity, and identification of within-subject serial correlation. These issues and our approaches to diagnosing them are laid out in more detail in Section 1.5 of this chapter. Methods for selecting random effects and identifying within-subject heteroscedasticity are developed in Chapter 2; these methods are tested in the simulation experiments presented in Chapter 3. Chapter 4 develops a method for detecting within-subject serial correlation, in the presence of potential bias of the nonlinear regression function. In Chapter 5 we apply our methods to two real examples. Chapter 6 concludes the dissertation with a brief discussion of the limitations of our methods, which provide motivation for possible future extensions.

The remainder of this chapter is laid out as follows. We introduce the two real examples that provide context and motivation for our work in Section 1.1. In Section 1.2 we describe the mechanistic models and resulting nonlinear regression functions that could be used to model the expected trajectory for each subject in the two examples. We review notation and terminology of the nonlinear mixed effects model in Section 1.3. In Section 1.5 we give an overview of the three model specification issues that our methods are developed to address. Our methods for exploring marginal covariance structure and within-subject autocorrelation are based on first-order linearization of the nonlinear mixed effects model; we formulate the first-order linear approximation in Section 1.4. We conclude Chapter 1 with a brief review of recent literature on the topics of random effects selection in linear and nonlinear mixed effects models, testing or modeling heteroscedasticity in nonlinear regression models, detecting autocorrelation in nonlinear models, and goodness-of-fit tests for nonlinear models. (The standard tools of multivariate analysis that are the basis for our methods are reviewed at the beginning of Chapter 2.)

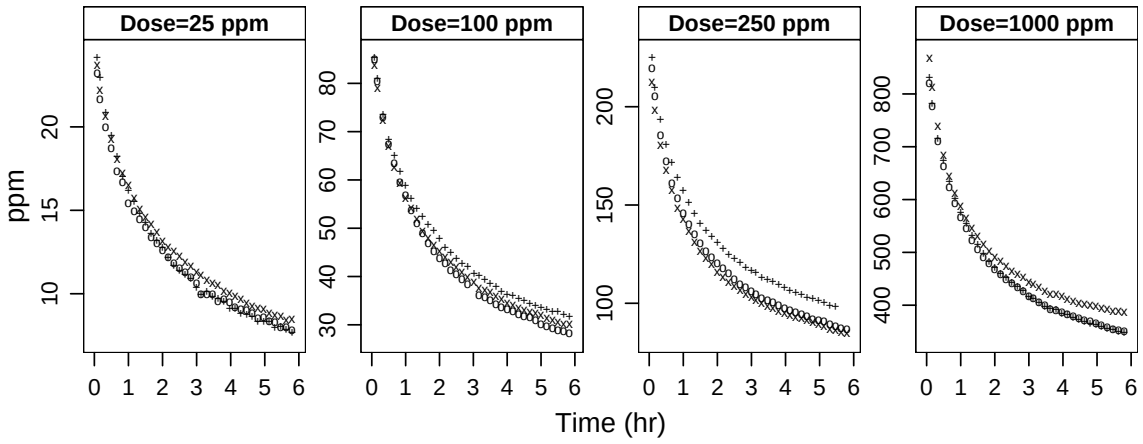

Figure 1.1: *Data from US-EPA CCl_4 gas uptake experiment. Each panel plots data for a different exposure concentration. Different plotting symbols are used for the three animals exposed at each dose.*

1.1 Motivating Examples

1.1.1 Carbon tetrachloride gas uptake experiment

Carbon tetrachloride (CCl_4) is a volatile liquid that causes chemical injury to the liver through necrosis and fat accumulation (Plaa and Charbonneau, 2000). These properties as well as the chemical's extensive use as a grain fumigant, in the production of chlorofluorocarbons, and as a solvent in dry cleaning, makes it of interest as an air pollutant. Evans, Crank, Yang, and Simmons (1994) described a gas uptake experiment in which twelve rats were individually exposed to CCl_4 in a closed chamber. Four initial exposures were used in the experiment: 1000, 250, 100 and 25 ppm. Concentration of the chemical in the chamber was measured approximately 36 times per rat over a period of 6 hours. Measurements were made at the same (or very similar) time points for all rats in the study. Figure 1.1 presents the data for the twelve rats in the gas uptake experiment.

The graphs illustrate the similarities and differences between the different subject profiles over time. While the typical pattern of change over time was similar across subjects, there

were small differences between subjects, manifest in a clear separation of a few of the subject profiles from others. The observed profile on each subject was relatively smooth. For a given subject, we might attribute the small fluctuations around a smooth curve to measurement error or small within-subject fluctuations. We will call this latter variation, *within-subject variation*. The variation between trajectories will be termed *between-subject variation*. This might be attributed to random differences between the treatment or physiology of different subjects.

A goal of the gas uptake experiment was to predict the concentration of CCl_4 over time in the liver (these concentrations were not observed directly). This concentration can be estimated using a model that predicts the concentration of CCl_4 in the liver at a given time after initial exposure to the chemical. *Mechanistic models* such a *compartment model* or a *physiologically based pharmacokinetic* (PBPK) model are typically used for this type of prediction. We describe a few examples of mechanistic models in Section 1.2. Fitting the PBPK or the compartment models to the gas uptake experiment data using the nonlinear mixed model framework would allow scientists to estimate typical model coefficients characterizing the toxicokinetics of CCl_4 in rats, and to quantify the expected rat-to-rat variation in these coefficients.

1.1.2 Pharmocokinetics of theophylline

Theophylline is a drug used to prevent wheezing and difficulty breathing caused by lung diseases such as asthma, emphysema, and chronic bronchitis. Davidian and Giltinan (1995, Chapter 5) described pharmacokinetic data collected from twelve human subjects given an oral dose of theophylline. On each subject, a measurement of the blood serum concentration (mg/L) of theophylline was taken before the dosing, and at ten subsequent time points over the next 24 hours. All subjects received the same dose (320 mg) of theophylline, therefore dose per unit of body weight (mg/kg) varied slightly from subject to subject. Figure 1.2 plots the measured serum concentration profiles for the twelve subjects.

As in the CCl_4 example, the theophylline trajectories for different subjects showed the same overall pattern, but also exhibited some between-subject variation. For example, the peak

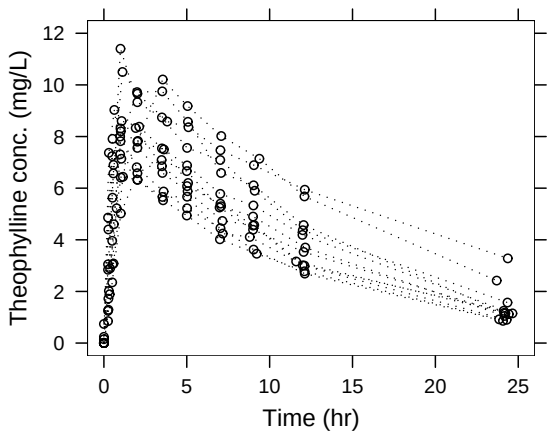

Figure 1.2: *Serum concentration of theophylline following oral dose of 320 mg in twelve subjects.*

concentration varied between subjects. This could be explained by variation in subject body weights, or by differences in the absorption rates of different subjects. Within-subject variation appeared to be small relative to the between-subject variation.

An open one-compartment (O1C) model with extravascular administration of drug is commonly used to describe the absorption and elimination of theophylline following an oral dose. We describe this model in more detail in Section 1.2. Under the O1C model, no drug is in the body at time zero. The concentration of drug increases as absorption proceeds until a peak concentration is reached. The subsequent decrease in concentration is due to elimination (excretion and metabolism) of the drug from the body. This model is typically parameterized in terms of an absorption rate k_a, and an elimination rate parameterized as Cl/V_d, where Cl is the clearance and V_d the volume of distribution. Fitting this model to the theophylline data using the nonlinear mixed model framework provides estimates of typical values of these parameters, and an estimate of subject-to-subject variation in absorption or elimination rates.

1.2 Mechanistic models

We use the term *mechanistic model* to refer to a mathematical model developed based on assumptions about the functional mechanics of a system. For example, most models used in pharmacokinetics are mechanistic—they make assumptions about how a drug is circulated, absorbed, metabolized, and eliminated. Mechanistic models may be simple (e.g., the open two-compartment model described in Section 1.2.1), or more complex (e.g., the PBPK model described in Section 1.2.2). Mechanistic models are typically parametric functions; their evaluation depends on parameters which often have a physical interpretation (e.g., a metabolic rate).

1.2.1 Compartment models

Open two-compartment model for CCl_4 gas uptake experiment

A simple model for the CCl_4 gas uptake data represents the exposure chamber with one compartment and the entire rat with a second compartment. The model is represented in Figure 1.3. In this simple model, we assume that CCl_4 is eliminated from the system (via metabolism, excretion, etc.) at a rate proportional to the concentration in the rat. Following from these

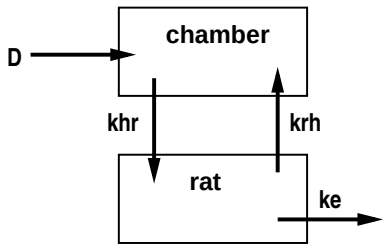

Figure 1.3: *Schematic representation of the open two-compartment (O2C) model for for CCl_4 gas uptake experiment. D is the exposure concentration, k_{hr} and k_{rh} are the rates of exchange between the chamber and rat and vice versa, and k_e is the elimination rate (due to metabolism and excretion).*

assumptions, the system of linear differential equations

$$\begin{aligned} \dot{z}_h &= -k_{hr}\, z_h + k_{rh}\, z_r \quad \texttt{[chamber]} \\ \dot{z}_r &= k_{hr}\, z_h - (k_{rh} + k_e)\, z_r \quad \texttt{[rat]} \end{aligned} \tag{1.1}$$

describes this model. The initial conditions are $z_h(0) = D$ and $z_r(0) = 0$.

The concentration of CCl_4 in the chamber at time t is given by $z_h(t)$ and the concentration in the rat is given by $z_r(t)$. The dot-notation represents a derivative with respect to time t, $\dot{z_h} \equiv \mathrm{d}z_h(t)/\mathrm{d}t$. These quantities are model predictions, not observed data, hence our choice of the notation z instead of y; y is reserved for observed data or for random variables. Initial exposure (at time 0) in the chamber was D (dose) and we assume that initial concentration of CCl_4 in the rat was 0. Transfer from the chamber to the rat due to inhalation takes place at rate k_{hr}. The rate of transfer from rat to chamber due to exhalation is k_{rh}. The rate of elimination from the rat is k_e; this combines the effects of metabolism and excretion.

The differential equations (1.1) can be solved analytically for $z_h(t)$ and $z_r(t)$ and have the following closed form solution for the concentration in the exposure chamber

$$z_h(t) = D\left\{\beta^{(1)} \exp(-\beta^{(2)} t) + \beta^{(3)} \exp(-\beta^{(4)} t)\right\}$$

where D is initial exposure concentration. The coefficients $\beta^{(1)}, \ldots, \beta^{(4)}$ are functions of the rates k_{hr}, k_{rh}, k_e and the initial exposure D. Because of the dependence of the model prediction $z_h(t)$ on $\beta^{(1)}, \ldots, \beta^{(4)}$ we will use the notation $f(t, \boldsymbol{\beta}; D)$ instead of $z_h(t)$. If we assume that a dose exactly equal to D was administered at time 0 then it is reasonable to constrain $\beta^{(1)} + \beta^{(3)} = 1$. Under this assumption we will write the model predicting the chamber concentration of CCl_4 at time t as

$$f(t, \alpha, \omega^{(1)}, \omega^{(2)}; D) = D\left\{\alpha \exp(-\omega^{(1)} t) + (1-\alpha) \exp(-\omega^{(2)} t)\right\} \tag{1.2}$$

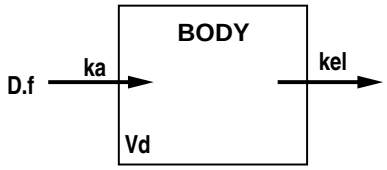

Figure 1.4: *Schematic representation of the O1C model for oral administration of theophylline. D is the administered dose, f the fraction absorbed, k_a the absorption rate constant, k_{el} the elimination rate constant, and V_d the volume of distribution.*

where $\alpha = \beta^{(1)} = 1 - \beta^{(3)}$, $\omega^{(1)} = \beta^{(2)}$, and $\omega^{(2)} = \beta^{(4)}$.

Open one-compartment model for oral dose of theophylline

The pharmacokinetics of theophylline in a human following oral dosing can be modeled with the open one-compartment (O1C) model with extravascular administration of drug (Ritschel and Kearns, 1998, Chapter 14). The O1C model represents the body with a single compartment and assumes a unidirectional input and output into and out of the system. This model assumes that drug entering the body distributes immediately between blood and other body tissues so that the entire body can be represented with one compartment. Figure 1.4 gives a schematic representation of the O1C model with extravascular (e.g., oral) administration of a dose D of drug. Sometimes only a fraction `f` of the drug may be absorbed into the body, but in the theophylline example we assume that this fraction is 100%. We assume that subjects have no previous exposure to theophylline so that theophylline concentration in the body at time zero is zero. The concentration of drug increases as absorption of drug from the gut proceeds until a peak concentration is reached. The subsequent decrease in concentration is due to elimination (excretion and metabolism) of the drug from the body.

If we assume that theophylline is absorbed at a rate proportional to the amount remaining in the gut (initially D`f`), and is eliminated from the body at a rate proportional to the

concentration in the body, then the differential equation

$$\dot{z}_b = k_a \frac{D\mathtt{f}}{V_d} \exp(-k_a t) - k_{el} z_b \tag{1.3}$$

with initial condition $z_b(0) = 0$ describes this model (Rodda, Sampson, and Smith, 1975). The concentration of theophylline in the body at time t is given by $z_b(t)$. Absorption from the gut occurs at rate k_a, and k_{el} is the rate of elimination from the body due to metabolism and excretion.

The differential equation (1.3) can be solved analytically for $z_b(t)$ (Rodda et al., 1975) and has the following closed form solution

$$z_b(t) = \frac{D\mathtt{f}}{V_d} \frac{k_a}{k_a - k_e} \left\{ \exp(-k_{el} t) - \exp(-k_a t) \right\}.$$

It is common to parameterize the elimination rate k_{el} in terms of clearance and volume of distribution, $k_{el} = Cl/V_d$. Writing the model to show explicit dependence on the parameters (k_a, Cl, V_d) gives the nonlinear function

$$f(t, k_a, Cl, V_d; D) = \frac{D\mathtt{f} k_a}{V_d(k_a - Cl/V_d)} \left\{ \exp\left(-\frac{Cl}{V_d} t\right) - \exp(-k_a t) \right\} \tag{1.4}$$

for predicting the serum concentration of theophylline at time t. We assume the dose D is known (a covariate) and that $\mathtt{f} = 1$.

1.2.2 A physiologically based pharmacokinetic (PBPK) model

In a physiologically based pharmacokinetic (PBPK) model, the body is subdivided into compartments representing organs or groups of tissues. For example, the model considered here divides the body into four compartments: the liver, fat, well-perfused tissues (central nervous system and viscera except liver), and poorly-perfused tissues (muscles and skin). A fifth compartment represents the exposure chamber. A system of differential equations describes the

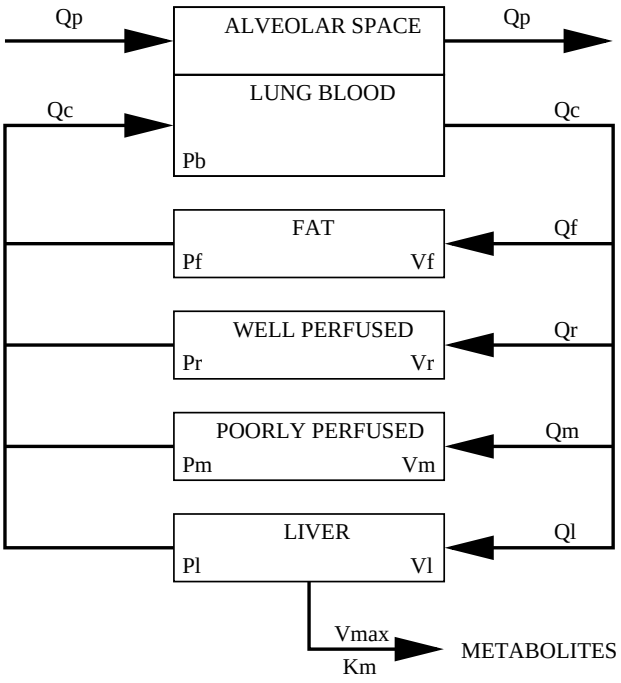

Figure 1.5: *Schematic representation of PBPK model for CCl_4 inhalation exposure in rats*

patterns of change in amounts or concentrations of the chemical over time in each compartment.

The model is parameterized by a large number of physiological, physiochemical and biochemical model parameters. PBPK models are useful for prediction of target tissue (e.g., liver) concentrations, which are often not easily observed, and are believed to provide a fairly sound basis for extrapolation to different exposure routes or to different species (Evans et al., 1994).

The PBPK model that we consider here has been established as a reasonable model in rats for describing the absorption, distribution, metabolism and elimination of CCl_4 and other volatile chemicals after exposure in a closed chamber (e.g., Evans et al., 1994; Bois, Zeise, and Tozer, 1990). A schematic representation of the PBPK model is presented in Figure 1.5. This model assumes that metabolism takes place in the liver compartment via saturable enzymatic reactions, the kinetics of which can be described by a so-called Michaelis-Menten term, $\{V_{max} z_l / V_l P_l\} / \{K_m + z_l / V_l P_l\}$. Kinetics in all other compartments are described by linear

differential equations.

$$\begin{aligned}
\dot{z}_h &= \texttt{RATS}\, Q_p \left(c_x - z_h/V_h\right) - K_L z_h && \texttt{[chamber]} \\
\dot{z}_f &= Q_f \left\{c_a - z_f/(V_f P_f)\right\} && \texttt{[fat]} \\
\dot{z}_r &= Q_r \left\{c_a - z_r/(V_r P_r)\right\} && \texttt{[rapidly perfused]} \\
\dot{z}_m &= Q_m \left\{c_a - z_m/(V_m P_m)\right\} && \texttt{[poorly perfused]} \\
\dot{z}_l &= Q_l \left\{c_a - z_l/(V_l P_l)\right\} - \frac{V_{max} z_l}{z_l + K_m V_l P_l} && \texttt{[liver]}
\end{aligned} \tag{1.5}$$

Initial conditions at $t = 0$ are given by

$$z_h(0) = z_{h0}, \quad z_f(0) = 0, \quad z_r(0) = 0, \quad z_m(0) = 0, \quad z_l(0) = 0.$$

In the differential equations, the zs denote the predicted amount of CCl_4 at a given time in a particular compartment, according to the model. Subscripts indicate model compartment: subscript h is for exposure chamber, p is for pulmonary, a is for arterial, f is for fat, r is for rapidly-perfused tissue, m is for poorly-perfused tissue, and l is for liver. The function $z_h(t)$ models the expected amount of CCl_4 in the chamber compartment for a particular rat. The Vs represent compartment volumes, Ps represent partition coefficients, and Qs represent blood flow or ventilation rates. The variable c_a is arterial concentration at time t, c_v is venous concentration at time t, c_x is concentration in exhaled breath at time t, K_L is loss rate from the empty chamber, V_h is volume of the empty exposure chamber, and B_w is rat body weight. The parameters V_{max} and K_m characterize the rate of metabolism in the liver compartment. The parameter K_m is the Michaelis-Menten constant, and V_{max} is the maximum rate of metabolism.

Table 1.1 lists typical values of the PBPK model parameters for CCl_4 metabolism in male rats. Other quantities are computed from these. A single rat was housed in each chamber in the CCl_4 experiment so that $\texttt{RATS} = 1$. The initial amount in the chamber is $z_{h0} = \texttt{CONC} \times V_h \times W_m/24450$ (mg) where $V_h = V_{CHC} - \texttt{RATS} \times B_w$ (liters) is the net chamber

Table 1.1: *Typical values of CCl_4 PBPK model input parameters (adapted from Evans et al., 1994)*

F344 Male rats body weight (B_w)	0.182 kg
Cardiac output (Q_c)	4.0 liter/hr
Ventilation rate (Q_p)	4.0 liter/hr
Fat blood flow percentage (Q_{fc})	5%
Liver blood flow percentage (Q_{lc})	15%
Slowly perfused blood flow percentage (Q_{mc})	32%
Fat volume percentage (V_{fc})	8%
Liver volume percentage (V_{lc})	4%
Slowly perfused volume percentage (V_{mc})	74%
$P_{\text{liver/air}}$ (P_{la})	16.17
$P_{\text{fat/air}}$ (P_{fa})	281.48
$P_{\text{slow/air}}$ (P_{ma})	13.3
$P_{\text{rapid/air}}$ (P_{ra})	16.17
$P_{\text{blood/air}}$ (P_b)	5.487
Molecular weight (W_m)	153.8 (g/mol)
Chamber volume (V_h)	3.8 liters
Loss rate from empty chamber (K_L)	0.02 1/hr
Michaelis-Menten constant (K_m)	1.3 mg/liter
Maximum rate of metabolism (V_{max})	0.11 mg/hr

volume. Compartment/blood partition coefficients (P_f, P_l, P_m, P_r) are computed from the compartment/air coefficients (P_{fa}, P_{la}, P_{ma}, P_{ra}) respectively, divided by the blood/air partition coefficient (P_b). Tissue volumes (V_f, V_l, V_m) equal the respective volume percentages (V_{fc}, V_{lc}, V_{mc}) times body weight (B_w) (liters). Tissue blood flow rates (Q_f, Q_l, Q_m) equal the respective blood flow percentages (Q_{fc}, Q_{lc}, Q_{mc}) times cardiac output (Q_c) (liter/hr). The

following difference equations define relationships between other quantities in the model,

$$Q_c = Q_f + Q_r + Q_m + Q_l, \quad V_f + V_r + V_m + V_l = 0.91 B_w$$
$$c_v = \frac{1}{Q_c} \sum_{*=f,r,m,l} \frac{Q_* z_*}{V_* P_*}, \quad c_a = \frac{Q_c c_v + Q_p z_h / V_h}{Q_c + Q_p / P_b}, \quad c_x = \frac{c_a}{P_b}.$$

The two metabolic parameters, K_m and V_{max}, are usually the focus of estimation based on data from a gas uptake experiment. Other parameters in the model (such as volumes, partition coefficients, blood flow rates) may be assigned values based on other studies (e.g., partition coefficients are measured in separate *in vitro* studies) or from literature. This fixing of many parameters in the model at assigned values is usually done out of necessity. In most cases it is impossible to estimate all model parameters: either there are not enough observations on each subject (sometimes fewer than the number of parameters), or parameters are not well-identified. Many authors (e.g. Gelman, Bois, and Jiang, 1996) address this problem through Bayesian analysis, placing informative prior distributions on parameters. When needed, we will distinguish between those model parameters that are fixed at assigned values (we will call these parameters $\boldsymbol{\xi}$), and those that are estimated from the data in hand (we will call these parameters $\boldsymbol{\beta}$). However, parameter estimation is not the focus of this dissertation and we will say little more about estimation.

The CCl_4 concentration in the exposure chamber and amount in the liver that are predicted by the PBPK model over a six hour period are shown in Figure 1.6 for an initial exposure of 250 ppm. Model parameters were set at the values given in Table 1.1. Because of the nonlinear differential equation in the liver compartment, there is not an analytical (closed-form) solution to the system of ODEs. To evaluate the model prediction for a given set of parameter values, we computed a numerical approximation using the `ode45` differential equation solver in Matlab (Version 7.0 (R14), The MathWorks, Inc.).

The O2C model described in Section 1.2.1 has some similarities with the PBPK model. If $K_m \approx 0$, then the differential equations describing the PBPK model is approximately linear

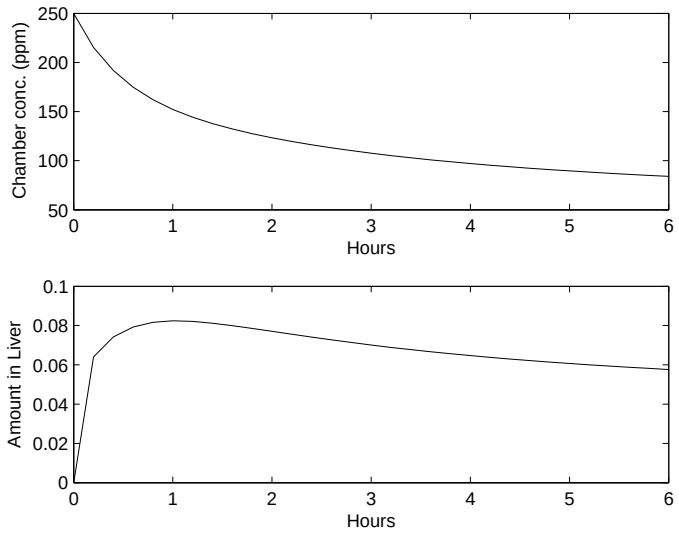

Figure 1.6: *PBPK model predictions for concentration or amount of CCl_4 in chamber and liver compartments for an initial exposure of 250 ppm. Other model parameters are given in Table 1.1.*

and has an approximate closed form solution. In this case, the solution for the concentration in the chamber is a sum of 5 exponential terms. If some of the exponential rates are very small, or rates in different compartments are similar, then the predicted chamber concentration from the PBPK model looks like that from the O2C model of equation (1.2).

1.3 Nonlinear mixed effects model (NLMM)

The CCl_4 gas uptake experiment data and the theophylline pharmacokinetic data described in Section 1.1 had a longitudinal structure: repeated measurements were made on each subject over time. This induces a correlation structure in the data: measurements made on the same subject are likely to be correlated with one another due to latent (unobserved) subject differences. The mechanistic models of Section 1.2 describe the expected trajectory for an individual subject. The *nonlinear mixed effects model* (NLMM) is a statistical approach that describes the

expected trajectory (for example with with a parametric nonlinear regression function defined by a mechanistic model), but also accounts for the correlation structure in the data by modeling between- and within-subject variability using *random effects*.

We will use the following notation for longitudinal data such as that collected in the two motivating examples. Suppose that n subjects are all measured p times at the same time points $t_1, t_2, \ldots, t_p$. Let y_{ij} be the response (e.g., CCl_4 concentration in the chamber, or serum concentration of theophylline) measured on subject i at time t_j. It will often be convenient to write the observations made on the i-th subject as a vector of length p,

$$\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})^T.$$

It is usual to assume that the random vectors, $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$, are independent and identically distributed (i.i.d.) according to some probability distribution, for example a multivariate normal distribution. The assumption of equal time points (i.e., balanced data) is necessary for the methods developed in Chapters 2 and 4.

In Section 1.2 we described mechanistic models for modeling the expected CCl_4 concentration over time in the exposure chamber for a given rat, or the expected serum theophylline concentration in a particular subject. Assuming that these models adequately describe the expected response for a subject, the observed trajectory might still differ slightly from the model prediction due to other sources of variation such as measurement error and within-subject fluctuations in metabolism rates, etc. The trajectories plotted in Figures 1.1 and 1.2 also differ between subjects. In the NLMM approach, we assume that between-subject differences are due to variation in the mechanistic model parameters over the population of subjects. The NLMM (also known as hierarchical nonlinear model) incorporates the mechanistic model, and the between- and within-subject variation into a statistical framework appropriate for describing longitudinal data.

We propose the following simple NLMM with two-stages (a subject-specific stage and a

between-subject stage) to describe longitudinal data such as those observed in the gas uptake or theophylline examples:

$$\begin{aligned} &\text{Stage-1:} \quad y_{ij} = f(t_j, \boldsymbol{\beta}_i, \boldsymbol{\xi}; \boldsymbol{x}_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \text{ independent } N(0, \psi_{ij}) \\ &\text{Stage-2:} \quad \boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{b}_i \text{ i.i.d. } N_q(0, \mathbf{D}) \end{aligned} \tag{1.6}$$

for $j = 1, \ldots, p$ and $i = 1, \ldots, n$.

In the first (subject-specific) stage of the model, y_{ij} is the observed response for subject i at measurement occasion t_j, where $\boldsymbol{\beta}_i$ is a $q \times 1$ vector of coefficients specific to subject i, $\boldsymbol{\xi}$ is a vector containing other model coefficients that are assumed known and are to be fixed at assigned values, and $\boldsymbol{x}_i$ is a vector of covariates specific to subject i (such as exposure dose). The additive term ε_{ij} is the within-subject error (due to measurement error, and other within-subject fluctuations from the mean trajectory) which is assumed to be normally distributed with variance $\psi_{ij} > 0$. The regression function f is the nonlinear function of $\boldsymbol{\beta}_i$ and $\boldsymbol{\xi}$ that is defined by the mechanistic model. In the case of the PBPK model where the differential equations cannot be solved analytically, the regression function f is defined implicitly, and f has to be evaluated numerically at $t_1, \ldots, t_p$ for given values of $\boldsymbol{\beta}_i$, $\boldsymbol{\xi}$ and $\boldsymbol{x}_i$.

In the stage-1 model we have assumed the within-subject errors to be independent, but have allowed for a different variance, ψ_{ij}, at each time point. It is common to assume a more parsimonious model for the variance of these errors. For example, we might assume the within-subject variance to be *homoscedastic*, $\psi_{ij} = \sigma^2$ for $j = 1, \ldots, p$, or to assume that ψ_{ij} is proportional to a power of the expected response, $\psi_{ij} = \sigma^2\{f(t_j, \boldsymbol{\beta}_i, \boldsymbol{\xi}; \boldsymbol{x}_i)\}^{2\gamma}$.

The regression function f may also depend on subject-level covariates $\boldsymbol{x}_i$. In the methods developed in Chapters 2 and 4, we assume that all subjects have the same covariates, or that the data can be scaled to eliminate the effect of covariates. For example, in the models for concentration of CCl_4 or theophylline, dividing by exposure or administered dose, D, eliminates this covariate from the stage-1 regression function. We will assume that this scaling also simplifies

the model for within-subject variation; i.e., makes variances approximately homoscedastic.

In the second (between-subject) stage, we assumed that the subject-specific coefficients $\boldsymbol{\beta}_i$ are independent across subjects, and normally distributed with mean $\boldsymbol{\beta}$ and between-subject covariance matrix $\mathbf{D}$. The random vectors $\boldsymbol{b}_i$ are called *random effects*. Parameter estimation for the NLMM usually focuses on the population mean parameters $\boldsymbol{\beta}$ and the covariance parameters ψ_{ij} and distinct elements in $\mathbf{D}$. There is substantial literature on methods for estimating these parameters. Most methods are based on maximizing the marginal likelihood (based on the marginal distribution of the observed data $\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n$) defined by the two-stage model in equation (1.6). For more details about the NLMM, its applications, and methods of estimation and inference, the interested reader is referred to a recent review by Davidian and Giltinan (2003).

One difference between the NLMM specification given in equation (1.6) and that found more commonly in the literature, is that we have parameterized the regression model f by two types of coefficients: $\boldsymbol{\beta}_i$ and $\boldsymbol{\xi}$. Recall that $\boldsymbol{\xi}$ contains the coefficients that are assumed known and are not to be estimated. These assigned coefficients must, by necessity, be assumed fixed across subjects, although it is possible that they too might actually vary from subject to subject. When we need to distinguish between the true and assigned values of $\boldsymbol{\xi}$, we will let $\boldsymbol{\xi}_*$ denote the true value and $\widetilde{\boldsymbol{\xi}}$ the assigned value. At some points in this dissertation where we do not concern ourselves with possible misspecification of the values of $\boldsymbol{\xi}$ (for example in Chapter 2), we have dropped $\boldsymbol{\xi}$ from our notation.

The coefficients denoted by $\boldsymbol{\beta}_i$ are potentially also of two different types: those that vary randomly between subjects, and those whose variation over subjects is comparatively small so that these coefficients could be considered fixed across subjects. We do not have separate notation for these fixed parameters. We can think of them as being part of $\boldsymbol{\beta}_i$ with their components of the random effect vector $\boldsymbol{b}_i$ equal to zero with probability one, or equivalently, their corresponding elements in the covariance matrix $\mathbf{D}$ equal to zero. If a random effect is zero (has zero variance) we will call it a trivial random effect.

1.4 First-order linearization

For now, suppose that the within-subject variances are homoscedastic, $\psi_{ij} = \sigma^2$ for all i and j. *First-order linearization* (FOL) was introduced by Beal and Sheiner (1982) as a way to approximate the marginal likelihood and compute estimates of the NLMM parameters $\boldsymbol{\beta}$, σ^2 and $\mathbf{D}$. While the FOL approximation to the marginal likelihood may be poor, and more refined methods such as Laplace's approximation (e.g., Wolfinger, 1993) are often favored for parameter estimation, we will make extensive use of the FOL approximation in the methods developed in this dissertation.

Let $\boldsymbol{\xi}_*$ be the true value of the known parameters, and $\widetilde{\boldsymbol{\xi}}$ be the assigned value. Assuming that the random effects, $\boldsymbol{b}_i$, and the differences, $(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}})$, are small, a first-order linear expansion of the true subject-specific (stage-1) regression function $f(t_j, \boldsymbol{\beta} + \boldsymbol{b}_i, \boldsymbol{\xi}_*; \boldsymbol{x}_i)$ about $\boldsymbol{b}_i = \mathbf{0}$ and $\boldsymbol{\xi}_* = \widetilde{\boldsymbol{\xi}}$ gives

$$y_{ij} \approx f(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) + \sum_{m=1}^{q} f_{\beta^{(m)}}(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) \cdot b_i^{(m)} + \sum_{h=1}^{s} f_{\xi^{(h)}}(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) \cdot (\xi_*^{(h)} - \widetilde{\xi}^{(h)}) + \varepsilon_{ij} \quad (1.7)$$

where $f_{\beta^{(m)}} = \partial f / \partial \beta^{(m)}$, $f_{\xi^{(h)}} = \partial f / \partial \xi^{(h)}$, and the partial derivatives are evaluated at $(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$. The partial derivatives are also referred to as *sensitivity functions* or *parameter sensitivities*. In Section 1.4.1 we discuss approximation of the sensitivities when f is evaluated numerically by solving a system of differential equations.

Let $\boldsymbol{f}^{(i)} = [f(t_1, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i), \ldots, f(t_p, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)]^T$, $\mathbf{F}_\beta^{(i)}$ be a $p \times q$ matrix with (j, m)-th element $f_{\beta^{(m)}}(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$, and $\mathbf{F}_\xi^{(i)}$ be a $p \times s$ matrix with (j, h)-th element $f_{\xi^{(h)}}(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$. Then (1.7) can be written in vector notation as

$$\boldsymbol{y}_i \approx \boldsymbol{f}^{(i)} + \mathbf{F}_\beta^{(i)} \boldsymbol{b}_i + \mathbf{F}_\xi^{(i)} (\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}) + \boldsymbol{\varepsilon}_i. \quad (1.8)$$

The approximate model (1.8) is linear in the random effect vector $\boldsymbol{b}_i$. Substituting the FOL approximation for the subject-specific regression function into the two-stage NLMM of equa-

tion (1.6) with $\psi_{ij} \equiv \sigma^2$, we can see that the marginal distribution of $\boldsymbol{y}_i$ is approximately normal, with marginal mean and covariance given approximately by the following expressions:

$$\mathbb{E}(\boldsymbol{y}_i) = \mathbb{E}[\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{b}_i)] \approx \mathbb{E}[\boldsymbol{f}^{(i)} + \mathbf{F}_\beta^{(i)}\boldsymbol{b}_i + \mathbf{F}_\xi^{(i)}(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}})] = \boldsymbol{f}^{(i)} + \mathbf{F}_\xi^{(i)}(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}) \equiv \boldsymbol{\mu}_i,$$

and

$$\begin{aligned}
\mathrm{Cov}(\boldsymbol{y}_i) &= \mathbb{E}[\mathbb{V}(\boldsymbol{y}_i|\boldsymbol{b}_i)] + \mathbb{V}[\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{b}_i)] \\
&\approx \mathbb{E}[\mathbb{V}(\boldsymbol{\varepsilon}_i)] + \mathbb{V}[\boldsymbol{f}^{(i)} + \mathbf{F}_\beta^{(i)}\boldsymbol{b}_i + \mathbf{F}_\xi^{(i)}(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}})] \\
&= \sigma^2\mathbf{I}_p + \mathbf{F}_\beta^{(i)}\,\mathbf{D}\,\mathbf{F}_\beta^{(i)T} \\
&\equiv \boldsymbol{\Sigma}_i.
\end{aligned}$$

The mean vector $\boldsymbol{\mu}^{(i)}$ and the covariance matrix $\boldsymbol{\Sigma}_i$ depend on the parameter values $\boldsymbol{\beta}$ and $\widetilde{\boldsymbol{\xi}}$ (this dependence is suppressed in the notation). In this dissertation, since our focus is not on estimating $\boldsymbol{\beta}$ or $\boldsymbol{\xi}$, approximate values will be substituted. For example, for the PBPK model we could use values in Table 1.1.

The dependence of these expressions on i is through the covariate vector $\boldsymbol{x}_i$. If all subjects have the same covariate values, $\boldsymbol{x}_i \equiv \boldsymbol{x}$ (e.g., they all have the same initial dose D), then the vectors $\boldsymbol{y}_i, i = 1, \ldots, n$ will be independent and approximately identically distributed according to the multivariate normal distribution with approximate marginal mean vector

$$\boldsymbol{\mu} = \boldsymbol{f} + \mathbf{F}_\xi(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}})$$

and approximate marginal covariance matrix

$$\boldsymbol{\Sigma} = \mathbf{F}_\beta\,\mathbf{D}\,\mathbf{F}_\beta^T + \sigma^2\mathbf{I}_p. \tag{1.9}$$

In this case, an approximate model for $\boldsymbol{y}_i$ is

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \, \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i. \tag{1.10}$$

If the random effects, $\boldsymbol{b}_i$, or the misspecification errors, $\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}$, are not small, then the FOL approximation of equation 1.10 will be poor. In Chapter 2, we assume that an approximation that is linear in the random effects $\boldsymbol{b}_i$ is adequate. However, if some random effects are large, then second-order terms may need to be added to the Taylor series approximation of equation (1.7). This will involve second order partial derivatives of f, and quadratic functions of the random effects in $\boldsymbol{b}_i$.

One of the model diagnostic objectives to be discussed in Section 1.5 is to discriminate between two parameterizations of a model based on the structure of the marginal covariance. The FOL approximation is also the basis of this diagnostic. Let $f^*(t, \boldsymbol{\delta}_i, \boldsymbol{\xi}; \boldsymbol{x})$ be an alternative parameterization of f with subject-specific coefficients $\boldsymbol{\delta}_i = \boldsymbol{\delta} + \boldsymbol{d}_i$. Assuming that the random effects $\boldsymbol{d}_i$ and the misspecification errors $(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}})$ are small, the first-order linear expansion of the alternative stage-1 regression function $f^*(t_j, \boldsymbol{\delta} + \boldsymbol{d}_i, \boldsymbol{\xi}_*; \boldsymbol{x}_i)$ about $\boldsymbol{d}_i = \mathbf{0}$ and $\boldsymbol{\xi}_* = \widetilde{\boldsymbol{\xi}}$ gives

$$y_{ij} \approx f^*(t_j, \boldsymbol{\delta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) + \sum_{r=1}^{q^*} f^*_{\delta^{(r)}}(t_j, \boldsymbol{\delta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) \cdot d_i^{(r)} + \sum_{h=1}^{s} f^*_{\xi^{(h)}}(t_j, \boldsymbol{\delta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) \cdot (\xi_*^{(h)} - \widetilde{\xi}^{(h)}) + \varepsilon_{ij}$$

where $f^*_{\delta^{(r)}} = \partial f^* / \partial \delta^{(r)}$ evaluated at $(t_j, \boldsymbol{\delta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$. Assuming constant covariates and, we arrive at the approximate model for $\boldsymbol{y}_i$ based on the parameterization f^*,

$$\boldsymbol{y}_i \approx \boldsymbol{\mu}^* + \mathbf{F}^*_\delta \, \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i. \tag{1.11}$$

where $\boldsymbol{\mu}^*$ and $\mathbf{F}^*_\delta$ are defined for the alternative parameterization f^* in the same manner that $\boldsymbol{\mu}$ and $\mathbf{F}_\beta$ were defined for parameterization f.

If the parameter $\boldsymbol{\beta}$ is a one-to-one function of $\boldsymbol{\delta}$, then for any value of $\boldsymbol{\delta}$ there exists a

value of $\boldsymbol{\beta}$ such that $f^*(t_j, \boldsymbol{\delta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) = f(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$. At this value of $\boldsymbol{\beta}$ it will also be the case that $f^*_{\xi^{(h)}}(t_j, \boldsymbol{\delta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) = f_{\xi^{(h)}}(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$. Thus $\boldsymbol{\mu} = \boldsymbol{\mu}^*$ and comparing the approximations in equations (1.10) and (1.11), we see that the difference between parameterizations f and f^* will potentially be manifest (to first-order) by the terms involving the random effects and sensitivity functions, $\mathbf{F}_\beta \boldsymbol{b}_i$ and $\mathbf{F}^*_\delta \boldsymbol{d}_i$. We develop this idea further in Section 2.3.3 of Chapter 2.

1.4.1 Approximating $\partial f/\partial \boldsymbol{\beta}$ when f is the solution of a system of differential equations

The partial derivative $\partial f/\partial \beta$ of a parametric model function $f(t, \beta)$ with respect to a parameter β is known as a *parameter sensitivity*. The sensitivity is a function of the variable t, and depends on the value of the parameter β_* where it is evaluated. The sensitivity reflects how rapidly the function f changes for small changes in β when β is near β_*.

The sensitivities are often used to reflect the degree of dependence of model predictions on particular parameter values. Since the true parameter values in a model are seldom known, knowing how sensitive a model's predictions are to small changes in parameter values provides useful information about the degree of uncertainty due to misspecification of a particular parameter value. The sensitivities are also used in other ways. For example in statistical applications such as nonlinear regression, the large sample estimate of the covariance matrix of the parameter estimates can be computed from the partial derivatives of the function with respect to the parameters. Numerical optimization methods will often make use of the partial derivatives of the objective function with respect to its arguments; these partial derivatives are typically functions of the parameter sensitivities.

If the function f can be written down in closed form, then the sensitivities can be derived by analytical differentiation of f with respect to elements of β. However in the PBPK model, the function describing the concentration time curve in the exposure chamber is not available in closed form. Instead it is computed as one component of the numerical solution

to a system of ordinary differential equations (ODEs). In this case the sensitivities must be approximated numerically. Suppose that β is a single model parameter under consideration and let $f(t, \beta) \equiv z(t, \beta)$ be the required model prediction obtained by solving a differential equation in z. The partial derivative of f with respect to β could be computed using a finite difference approximation, for example a central difference

$$[z(t, \beta + h) - z(t, \beta - h)]/(2h).$$

An alternative approach when z is defined by a system of differential equations is to solve a new system of differential equations called the *sensitivity equations*. The latter approach (also know as internal differentiation or variational equations) is preferable to the finite difference approximation: the computational costs being similar to the costs of finite differencing, but the accuracy being higher, especially when low accuracy methods are used to approximate z (e.g., Ostermann, 2004; Dalgaard and Larsen, 1990).

In order to explain the derivation of the sensitivity equations, suppose that the solution $z(t, \beta)$ depends on a single parameter β and is defined by a single differential equation

$$\dot{z}(t, \beta) = \mathcal{F}(z(t, \beta), \beta)$$

with initial condition $z(t_0, \beta) = z_0(\beta)$. Recall that the dot-notation denotes differentiation with respect to time, $\dot{z} = \mathrm{d}z/\mathrm{d}t$. Taking partial derivatives with respect to β on both sides of the differential equation gives

$$\frac{\partial}{\partial \beta}\dot{z}(t, \beta) = \frac{\partial}{\partial z}\mathcal{F}(z, \beta) \cdot \frac{\partial}{\partial \beta} z(t, \beta) + \frac{\partial}{\partial \beta}\mathcal{F}(z, \beta)$$

by the Chain Rule for derivatives. Let $z_{[\beta]}(t, \beta) = \partial z(t, \beta)/\partial \beta$ denote the partial derivative of $z(t, \beta)$ with respect to β. This is the sensitivity function that we seek. Assuming that we may swap the order of partial differentiation with respect to t and β, and suppressing the dependence

of z and $z_{[\beta]}$ on (t, β), we can re-write this last equation as

$$\dot{z}_{[\beta]} = \frac{\partial}{\partial z}\mathcal{F}(z, \beta) \cdot z_{[\beta]} + \frac{\partial}{\partial \beta}\mathcal{F}(z, \beta). \tag{1.12}$$

This is a differential equation for the required sensitivity function $z_{[\beta]}$. The initial condition for this equation is obtained by differentiating the original initial condition with respect to β, giving

$$z_{[\beta]}(t_0, \beta) = \frac{\partial}{\partial \beta} z_0(\beta).$$

The differential equation for $z_{[\beta]}$ is a function of z, and must be solved simultaneously with the original differential equation for z. In Appendix D we derive the sensitivity equations for several parameters of the PBPK model.

1.5 Diagnostics for the NLMM

In this dissertation we assume that a preliminary fit of the has been made, and that estimates of the coefficients in the nonlinear function are in hand. We view the methods developed in the following chapters as providing a set of diagnostic tools for assessing certain assumptions of the NLMM. Our methods provide guidance for formulating a realistic but parsimonious version of the NLMM to be fitted after the preliminary analysis. The methods developed in this dissertation examine three model specification questions for the nonlinear mixed effects model.

The first question we address is identifying how many, and which, coefficients in the NLMM should be specified as random in order to adequately characterize between-subject variation. Mechanistic models may have many parameters. For example the PBPK model described in Section 1.2.2 has approximately 20 parameters. If the model is believed to be a good description of the expected trajectory for a single subject, then the observed between-subject differences could be described by allowing each subject to have a different set of parameter values. However

it is likely that some of the parameter values will be more similar across subjects than others. Our goal is to identify the subset of parameters that require non-trivial random effects.

Our approach to identifying the non-trivial random effects is based on the FOL approximation,

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad \text{for} \quad i = 1, \ldots, n.$$

This approximation is similar to a multivariate common factor model,

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_k \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i$$

that explains the structure of the marginal covariance with k linear random effects called common factors. By fitting the common factor model to the data and looking at the relationship between the $p \times k$ loadings matrix $\boldsymbol{\Lambda}_k$ and the $p \times q$ sensitivity matrix $\mathbf{F}_\beta$ we identify which elements in $\boldsymbol{\beta}$ require non-trivial random effects. This approach is developed in Chapter 2.

The question of random effects selection can also be looked at from the viewpoint of identifying a parameterization of the nonlinear function that characterizes the between-subject variation with a few random effects. As pointed out in Section 1.4, under a different parameterization f^* of the nonlinear regression function, the FOL approximation for the data becomes

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}^*_\delta \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i.$$

By examining the relationship between $\boldsymbol{\Lambda}_k$ and $\mathbf{F}^*_\delta$, and comparing results with those for the original parameterization f, we may be able to pick one parameterization that provides a simpler explanation of between-subject variation than another.

The second model-specification issue that we address, is whether there is strong evidence that within-subject variation is heteroscedastic; i.e., does the magnitude of variation change over time? Our approach for detecting heteroscedasticity is also based on the multivariate factor model, which models within-subject variation as $\mathbb{V}(\varepsilon_{ij}) = \psi_j > 0$ for $j = 1, \ldots, p$. By

comparing the fits of factor models with different assumptions about ψ_j, we can detect the presence of within-subject heteroscedasticity, and possibly also propose an appropriate model for within-subject variation. For example, comparing two factor models, the first with $\psi_j = \sigma^2$ (homoscedasticity) and a second with $\psi_j = \sigma^2 \mu_j^{2\gamma}$ (power-of-the-mean), we can judge whether the power-of-the-mean model provides a better description of within-subject variation than the homoscedastic model.

The factor model and the simple NLMM of equation (1.6) assume that the within-subject residuals ε_{ij} in the stage-1 model are independent within subject. Examining whether this is a reasonable assumption, or whether there is evidence of within-subject serial correlation, is the third issue addressed in this dissertation. Exploring this question is complicated by potential bias in the nonlinear regression function. Even if the mechanistic model is true, it could be difficult to estimate all the coefficients from one data set, especially if there are many parameters, as in the PBPK model for the CCl_4 example. Some coefficients might be assigned known values that are erroneous, $\widetilde{\boldsymbol{\xi}} \neq \boldsymbol{\xi}_*$, leading to model misspecification and bias,

$$\mathbb{E}(y_{ij}|t_j, \boldsymbol{\beta}_i, \boldsymbol{\xi}_*; \boldsymbol{x}) = f(t_j, \boldsymbol{\beta}_i, \boldsymbol{\xi}_*; \boldsymbol{x}) \neq f(t_j, \boldsymbol{\beta}_i, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}).$$

Model bias makes it difficult to tease out the presence of time-dependence in the within-subject residuals. A method for detecting within-subject correlation in the presence of misspecified coefficient values is developed in Chapter 4. Our approach is based on the FOL of equation (1.8). For subjects i and i' with the same covariate values, the between-subject difference is approximately

$$\boldsymbol{y}_i - \boldsymbol{y}_{i'} \approx \mathbf{F}_\beta(\boldsymbol{b}_i - \boldsymbol{b}_{i'}) + (\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_{i'})$$

and the effect of misspecification of $\boldsymbol{\xi}$ cancels out, at least approximately. Diagnosis of within-subject serial correlation is then based on the residuals after regressing vectors of differences, $\boldsymbol{y}_i - \boldsymbol{y}_{i'}$ on the sensitivity matrix $\mathbf{F}_\beta$.

1.6 Literature review

The methods that we develop in Chapters 2 and 4 for constructing, or assessing the appropriateness of, the covariance structure in a NLMM, are based on standard multivariate analysis tools. We review these tools in Section 2.2 of Chapter 2.

Few diagnostic tools have been developed specifically for the NLMM. In the current section, we touch on a sampling of literature discussing approaches to modeling covariance structure in mixed models (mostly linear or generalized linear mixed models), or detecting bias, heteroscedasticity, and autocorrelation in linear and nonlinear regression. Our review of the literature is far from comprehensive, but we hope that it provides some insight into possible approaches for assessing statistical model assumptions.

There is interplay among the issues of model bias, variance heteroscedasticity, and autocorrelation. Under the assumption that residuals are homoscedastic and uncorrelated, lack-of-fit or bias in the deterministic part of the model may be assessed by examining residuals for evidence of autocorrelation or heteroscedasticity. A test of autocorrelation (Section 1.6.3) is often used as an instrument to detect model bias, for example, due to omission of important predictors. In linear regression, responses may be transformed to meet the homoscedasticity assumption. However, it is often common in nonlinear regression applications to encounter heteroscedastic variance, where a transformation of the response may not be an option.

Methods for formulating an appropriate mixed effects model, or detecting lack-of-fit in nonlinear models are reviewed in Section 1.6.1. In Section 1.6.2 we review methods that diagnose within-subject heteroscedasticity in residuals. We review methods for detecting autocorrelation in linear and nonlinear models in Section 1.6.3. The selection of random effects in mixed models, and the related topic of picking an appropriate covariance model, is covered in Section 1.6.4.

1.6.1 Model formulation and lack-of-fit of the regression function for nonlinear and mixed effects models

Molenberghs and Verbeke (2004) highlighted the importance of formulating a model that is sound from both a mathematical point of view, and adequate to answer the scientific questions posed. For mixed models, they focused on the difference between the hierarchical and marginal views of mixed models. The NLMM described in Section 1.3 is an example of a hierarchical model, with random effects defined at the subject-level inducing a correlation structure at the marginal level. Under the hierarchical model the parameters representing the variances of the random effects are constrained to be positive (because they are variances). In the marginal view, the marginal mean and covariance structure are modeled directly. In linear mixed models with normally distributed responses, the hierarchical and marginal models are equivalent (they produce the same marginal likelihood), However, even in the linear case, the choice of view has implications on appropriate tests for variance components (touched on in Section 1.6.4). For discrete responses and generalized or nonlinear mixed effects models, the differences between the hierarchical and marginal views are more fundamental.

In the motivating examples of this dissertation, the scientific question demands fitting a hierarchical NLMM. Our methods are based on linearization of the subject-level model, giving an approximate linear mixed effects model. The multivariate factor model, which is the basis of the methods developed in Chapter 2, is also a hierarchical model (the common factors are linear random effects), but modeling of the covariance structure occurs at the marginal level.

For nonlinear regression models, Diebolt and Zuber (1999) constructed a test for adequacy of the nonlinear regression function using a marked empirical process of residuals. Their test allowed for heteroscedastic variance in residuals, and they claimed that the test could be extended to a model with first-order autoregressive residuals.

Crainiceanu and Ruppert (2004a) proposed a likelihood ratio goodness-of-fit test for the adequacy of a parametric nonlinear regression function. Their approach was to embed the

parametric function into a larger family consisting of the parametric nonlinear regression function plus a nonparametric spline. Using the mixed model representation of penalized splines, the (restricted) likelihood ratio test of the adequacy of the parametric regression function essentially tests whether the variance component corresponding to the spline is zero. Their approach assumed that residuals were homoscedastic and uncorrelated. If observations are made on independent samples then the assumption of no serial correlation may be realistic. However, without a strong assumption about the structure of residuals, the method cannot assess lack-of-fit of the regression function.

Pan and Lin (2005) proposed graphical and numerical methods for checking the deterministic part of a generalized linear mixed model (GLMM). Deterministic components include the functional form of a covariate, and form of the link function. The method did not check modeling assumptions about the random effects. Their approach was based on the marginal residuals of the model. Plots of marginal residuals may be difficult to interpret because residual variability is unknown. However, the asymptotic distribution of cumulative sums of residuals can be derived under the null hypothesis that the GLMM has been correctly specified. Their method was based on the cumulative sums of residuals.

Dey, Chen, and Chang (1997) applied Bayesian methods of model selection to a nonlinear regression example, and to a nonlinear mixed effects model (the random effects in their example entered the model linearly). The Bayesian approach was used to assess adequacy of the form of the nonlinear regression function, as well as the choice of random effects. Competing models were compared using diagnostics computed from the predictive distributions of different models.

1.6.2 Within-subject heteroscedasticity

The paper by Karlsson, Beal, and Sheiner (1995) illustrated with several simulation experiments the impact that incorrect residual error models can have on parameter estimation in nonlinear mixed effects models. They showed that ignoring serially correlated errors, that might occur due to misspecification of the nonlinear regression function, led to biased estimates of random-effect

variance parameters. Misspecification of the model created bias in mean parameter estimates. Ignoring heteroscedasticity in within-subject errors created bias in all parameter estimates. These results emphasize the potential usefulness of the methods developed in this dissertation that aim to detect within-subject residual heteroscedasticity and serial correlation in repeated measures data.

Cook and Weisberg (1983) published a seminal paper on diagnostics for heteroscedastic variance in linear regression without true replicates. They developed a score test for heteroscedasticity, and suggested a graphical diagnostic procedure. The score test involves modeling the weight function for the variances as a function of predictors, or the expected response. The test for heteroscedasticity is equivalent to a test about the value of parameters in the weight function. Because the score test is sensitive to outliers and deviations from normality, the authors recommended a graphical procedure to complement their test. They suggested plots similar to those now in common practice; for example plotting standardized unweighted residuals versus predictors, versus the log of predictors, or versus fitted values. They also proposed the use of studentized residuals.

Chapter 2 of the book by Carroll and Ruppert (1988) deals with estimation of the coefficients of a nonlinear regression model when variance is heteroscedastic. These authors argued that in many applications, variability, as well as the mean, can be modeled as a smooth function of predictors, often through the mean function. They extended the graphical procedure suggested by Cook and Weisberg (1983) to nonlinear regression, and explained how graphical plots of residuals may assist with detecting and modeling heteroscedasticity. For example, plotting unweighted least-squares residuals against predicted values will expose a relationship between residual variation and mean response. Various transformations of residuals (e.g., squared, absolute value, 2/3 power) or using studentized residuals may assist with picking a model for variability.

A slightly different issue regarding heteroscedasticity in a nonlinear mixed effect model was addressed by Zeng and Davidian (1997). This issue pertains to immunoassay experiments.

In these experiments it is standard to assume that residual variance in one run of the assay increases with mean response. The variance is typically modeled as being proportional to some power of the mean function. These authors proposed a test for determining whether the parameters in the variance model change significantly from one run of the assay to another.

Lin and Wei (2003) proposed score tests for detecting different types of heteroscedasticity that might occur in nonlinear regression. Their first test was based on a proposed model for the residual variance (an extension of the score test of Cook and Weisberg (1983)). Under this model, the test for homoscedastic variance is equivalent to testing that the parameter(s) in this model are zero. The other tests are based on "randomizing" parameters in the mean or variance functions, and testing hypotheses about the variance structure of these randomized parameters. The proposed tests depend critically on the adequacy of the variance model, and postulating such a model is difficult in practice.

1.6.3 Tests for autocorrelation in linear and and nonlinear models

The Durbin-Watson test (Durbin and Watson, 1950, 1951) is commonly used to test for serial correlation in a linear regression model, $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. The null hypothesis assumes uncorrelated residuals, and the alternative hypothesis assumes the residuals follow a first-order autoregressive, AR(1), process. The test statistic is based on the residuals, e_j, from the least-squares fit of the regression model,

$$d = \frac{\sum_{j=2}^{p}(e_j - e_{j-1})^2}{\sum_{j=1}^{p} e_j^2}.$$

Rejection of the null hypothesis is usually based on upper, $d_U(\alpha)$, and lower, $d_L(\alpha)$, bounds of the level-α critical values. For example, to test for positive autocorrelation, $d < d_L(\alpha)$ indicates significant evidence that the error terms are positively autocorrelated, and $d > d_U(\alpha)$ provides significant evidence that the error terms are not positively autocorrelated. Durbin and Watson (1971) investigated methods for computing the exact null distribution of the d statistic and various approximations to the exact distribution.

The Durbin-Watson test only applies when the predictors in the regression are fixed. Bhargava, Franzini, and Narendranathan (1982) extended the Durbin-Watson test to panel data (e.g., longitudinal data), but only in the case where subject effects are modeled as fixed. The Durbin-Watson test statistic, d, can also be computed to test for autocorrelation in a nonlinear regression model, $\boldsymbol{y} = f(\mathbf{X}, \boldsymbol{\beta}) + \boldsymbol{\varepsilon}$. White (1992) approximated the null distribution of d for a nonlinear regression model using a first-order Taylor series expansion of $f(\mathbf{X}, \boldsymbol{\beta})$ around the least-squares estimates $\widehat{\boldsymbol{\beta}}$, resulting in the approximate nonlinear Durbin-Watson test.

1.6.4 Random effects selection in linear and nonlinear mixed effects models

There are two main approaches to selection of random effects in mixed models. The first approach uses model and variable selection methods. The second approach relies on formal tests about the covariance parameters of random effects, in particular, tests that the variance of a random effect is zero.

Model / variable selection methods

Since parameters in mixed models are often estimated by likelihood methods, the adequacy of an assumed mean and covariance structure can be tested using likelihood methods such as Akaike's information criterion or likelihood ratio tests comparing a sequence of nested models. Vonesh, Chinchilli, and Pu (1996) developed a pseudolikelihood ratio test for assessing the adequacy of a given covariance structure using only the model in hand. This test is based on comparing the the robust sandwich estimator of the covariance matrix of the fixed effect parameters, with the model-based version of this covariance. The disadvantage of likelihood-based methods is that they tend to be sensitive to departures from the assumed distribution of responses (often normality). Vonesh et al. (1996) also derived a measure of concordance between fitted and observed responses, similar to the R^2 measure in univariate linear models, that can be used to simultaneously assess the mean and covariance models. This measure does not require specification of a likelihood function, and so is robust to departures from an assumed

distribution.

Bayesian approaches to selection of random effects in linear and generalized linear mixed models (e.g. Chen and Dunson, 2003; Cai and Dunson, 2006; Kinney and Dunson, 2007) typically focus on model and variable selection methods, such as stochastic search variable selection. For example, these authors used variable selection-type mixture priors that place point mass on random effects with zero variance, allowing random effects to drop out of the model.

Likelihood ratio tests and score tests

Formal hypothesis tests about the covariance parameters of random effects are based on the likelihood (or restricted likelihood) function. Likelihood ratio tests and score tests are commonly used; these are discussed below. For a linear random coefficient model, Morgan (1996) suggested an F-test based on treating the model coefficients as fixed (non-random) but different across subjects. This model can be fitted with ordinary least squares. The test that a particular coefficient was equal across subjects (i.e., a fixed parameter not a random effect) is the usual full versus reduced F-test comparing the full model to a restricted model constraining the values of the parameter of interest to be equal across subjects. This test had an exact F distribution under the null hypothesis that the variance of the random effect was zero.

The seminal reference for likelihood ratio tests concerning variance components in the linear mixed effects model is the paper by Stram and Lee (1994). This paper was based on the theoretical results about asymptotic properties of likelihood ratio tests developed by Self and Liang (1987) for cases where the parameter being tested is on the boundary of the parameter space under the null hypothesis. This is the setup when testing a hypothesis about a variance component σ_b^2, that is constrained to be positive: H_0: $\sigma_b^2 = 0$ versus H_1: $\sigma_b^2 > 0$. In such boundary situations, the asymptotic null distribution of -2 times the loglikelihood is not chi-squared (as in standard testing situations), but may be approximated by a mixture of chi-squared random variables in some cases.

Crainiceanu and Ruppert (2004b) derived the finite sample distribution of the likelihood

ratio test for one variance component in a linear mixed model. These results would be useful when the number of independent subjects remains small, since the asymptotic results of Stram and Lee (1994) assume that the number of independent subjects approaches infinity while the number of observations per subject remains fixed.

Verbeke and Molenberghs (2003) discussed the use of score tests for inference on variance components. They pointed out that when variance components are constrained to be positive (as in the case of a hierarchical mixed model), the traditional two-sided score test is inappropriate; a one-sided hypothesis must be used. As with the likelihood ratio tests, since the variance component is on the boundary of the parameter space under the the null hypothesis, standard asymptotic results about the distribution of the score test cannot be leaned on. These authors showed that the likelihood ratio test and the score test are asymptotically equivalent when the variance component is constrained to be positive. (By standard theory, Wald tests, score tests and likelihood ratio tests are asymptotically equivalent in the unconstrained case.) They commented that the constrained likelihood ratio test often has practical computational advantages over the score test, despite the fact that the null and alternative models must be fitted for the likelihood ratio test, and only the null model for the score test.

Savalli, Paula, and Cysneiros (2006) presented a score-type test (with constrained alternative) for assessing the variance components in a linear mixed model when the random effects and stage-1 residuals have an elliptical distribution. The elliptical class includes all symmetric continuous distributions.

Chapter 2

Exploring the marginal covariance structure

2.1 Introduction

This chapter describes two multivariate approaches, *principal component analysis* and *maximum likelihood factor analysis*, for exploring the structure of the marginal covariance matrix $\mathbb{V}(\boldsymbol{y}_i)$ where $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})^T$. We assume that the data are *balanced*: there are p observations made at time points $t_1, t_2, \ldots, t_p$ on each of n subjects. The methods explore the structure of the observed covariance matrix of a random sample $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ in order to check the covariance modeling assumptions made at both stages of the *nonlinear mixed effects model* (NLMM).

The two-stage NLMM described in Section 1.3

$$\begin{aligned} \text{Stage 1:} \quad y_{ij} &= f(t_j, \boldsymbol{\beta}_i, \boldsymbol{\xi}; \boldsymbol{x}_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \psi_{ij}) \\ \text{Stage 2:} \quad \boldsymbol{\beta}_i &= \boldsymbol{\beta} + \boldsymbol{b}_i, \quad \boldsymbol{b}_i \overset{\text{i.i.d.}}{\sim} N_q(\mathbf{0}, \mathbf{D}) \end{aligned}$$

models the structure of the marginal covariance in the two stages: through the covariance of the *within-subject residuals* ε_{ij} in the stage-1 model, and through the covariance $\mathbf{D}$ of the *random effects* $\boldsymbol{b}_i$ in the stage-2 model. The methods of this chapter explore the covariance models at the two stages.

1. Firstly we explore how to characterize between-subject variation. We identify the number of random effects needed to capture between-subject variation, select which coefficients in the stage-1 model require non-trivial random effects, and identify an appropriate model parameterization.

2. Secondly we explore whether an assumption of homogeneous variance across time points, $\mathbb{V}(\varepsilon_{ij}) = \psi_{ij} \equiv \sigma^2$, in the stage-1 model is reasonable.

If the function f is nonlinear in the random effects $\boldsymbol{b}_i$, then the marginal covariance of $\boldsymbol{y}_i$ will be a complicated function of the variance models at the two stages,

$$\mathbb{V}(\boldsymbol{y}_i) = \mathbb{E}\left\{\mathbb{V}(\boldsymbol{y}_i|\boldsymbol{b}_i)\right\} + \mathbb{V}\left\{\mathbb{E}(\boldsymbol{y}_i|\boldsymbol{b}_i)\right\}$$

(by the law of total variance). Under the first-order linear (FOL) approximation described in Section 1.4, the stage-1 model becomes

$$\begin{aligned} \boldsymbol{y}_i &\approx \boldsymbol{f}^{(i)} + \mathbf{F}_\beta^{(i)}\boldsymbol{b}_i + \mathbf{F}_\xi^{(i)}(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}) + \boldsymbol{\varepsilon}_i \\ &= \boldsymbol{\mu}_i + \mathbf{F}_\beta^{(i)}\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \end{aligned}$$

where the (i,m)-th element of $\mathbf{F}_\beta^{(i)}$ is $\partial f/\partial\beta^{(m)}$, and the (i,h)-th element of $\mathbf{F}_\xi^{(i)}$ is $\partial f/\partial\xi^{(h)}$. These partial derivatives are sometimes called called *sensitivity functions* in applied mathematics literature (see Section 1.4 of Chapter 1). In the approximation, the function and partial derivatives are evaluated at $(t_j, \boldsymbol{\beta}, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i)$. Recall that $\widetilde{\boldsymbol{\xi}}$ is the assigned (possibly misspecified) value of parameters that are considered known.

If we assume that all n subjects have the same covariates, $\boldsymbol{x}_i \equiv \boldsymbol{x}$, then the stage-1 model is approximately

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i. \tag{2.1}$$

If in addition we assume homoscedastic within-subject variation, $\psi_{ij} \equiv \sigma^2$, then the marginal covariance is approximately

$$\mathbb{V}(\boldsymbol{y}_i) \approx \mathbf{F}_\beta \mathbf{D} \mathbf{F}_\beta^T + \sigma^2 \mathbf{I}_p \equiv \boldsymbol{\Sigma}. \tag{2.2}$$

Thus an approximate marginal model for the data is $\boldsymbol{y}_i \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

In Section 1.4 we considered an alternative parameterization $f^*(t, \boldsymbol{\delta}, \boldsymbol{\xi}; \boldsymbol{x})$ of model $f(t, \boldsymbol{\beta}, \boldsymbol{\xi}; \boldsymbol{x})$.

If $\boldsymbol{\beta}$ is a one-to-one function of $\boldsymbol{\delta}$, then under parameterization f^*, the first-order linear approximation, will give

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}^*_\delta \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i.$$

If there is strong expression of one or more random effects, then difference between parameterizations will be captured by the two sensitivity matrices $\mathbf{F}_\beta$ and $\mathbf{F}^*_\delta$.

In this chapter, we use traditional multivariate methods for analysis of a random sample of n vectors from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ to explore the structure of the approximate covariance matrix $\boldsymbol{\Sigma}$. We review these techniques in Section 2.2. We use *principal component analysis* (Section 2.3) and *factor analysis* (Section 2.4) to model the marginal covariance matrix with a factorization such as

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}$$

where $\boldsymbol{\Lambda}_k$ is a $p \times k$ matrix of rank k, and $\boldsymbol{\Psi}$ is a diagonal matrix, $\boldsymbol{\Psi} = \text{diag}\{\psi_1, \ldots, \psi_p\}$. This factorization corresponds to a common factor model for the data on the i-th subject,

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_k \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i \tag{2.3}$$

where $\boldsymbol{u}_i \overset{\text{i.i.d.}}{\sim} N_k(\mathbf{0}, \mathbf{I}_k)$, independently of $\boldsymbol{\varepsilon}_i \overset{\text{i.i.d.}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Psi})$. Note in this model that the within-subject variance structure is the same across subjects ($\boldsymbol{\Psi}$ does not depend on i).

In Section 2.3 we use principal component analysis (PCA) to determine the number, k, of independent random effects necessary to characterize between-subject variation. In Section 2.4 we fit an orthogonal common factor model using maximum likelihood factor analysis (MLFA). By fitting factor models with different assumptions about the variance $\boldsymbol{\Psi}$ of the within-subject errors $\boldsymbol{\varepsilon}_i$, we identify and model within-subject heteroscedasticity. To identify an appropriate model parameterization, and to select which parameters have non-trivial random effects, we compare the fitted common factor model (2.3) obtained by PCA or MLFA to the approximate

first-order linear approximations of the stage-1 model

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad \text{or} \quad \boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}^*_\delta \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i$$

suggested by linearization of the NLMM under parameterizations f and f^* respectively. Finally we use multivariate regression to select columns from the matrices $\mathbf{F}_\beta$ or $\mathbf{F}^*_\delta$ that best explain the columns of the estimate of $\boldsymbol{\Lambda}_k$. The selected columns correspond to the model coefficients that have non-trivial random effects.

2.2 Background: multivariate methods

This section gives a brief review of some traditional methods of multivariate analysis: factor analysis, principal component analysis, and multivariate multiple regression. Further details can be found in the books by Johnson and Wichern (2007), Mardia, Kent, and Bibby (1979), and Anderson (2003).

In the longitudinal data view, the response for each subject is measured at p time points. In multivariate analysis, the p observations on each subject may be measurements of entirely separate characteristics or variables. For balanced data with the same observation times on each subject, let $\mathbf{Y}$ be an $n \times p$ data matrix with i-th row $\boldsymbol{y}_i^T = (y_{i1}, y_{i2}, \ldots, y_{ip})$ containing the p responses on subject i. Assume that the n vectors are randomly sampled from $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\mathbf{I}_n$ be the n-dimensional identity matrix, $\mathbf{1}_n$ be an n-vector of ones, and $\mathbf{P}_1 = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$ be the $n \times n$ matrix that projects a vector in $\mathbb{R}^n$ onto the subspace spanned by $\mathbf{1}_n$. The *sample mean* is the p-vector

$$\bar{\boldsymbol{y}} = \tfrac{1}{n}\sum_{i=1}^{n} \boldsymbol{y}_i = \tfrac{1}{n}\mathbf{Y}^T\mathbf{1}_n$$

and the *sample covariance* matrix is the $p \times p$ matrix

$$\mathbf{S}_n = \tfrac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T = \tfrac{1}{n}\mathbf{Y}^T\left(\mathbf{I}_n - \mathbf{P}_1\right)\mathbf{Y}.$$

The maximum likelihood estimator (MLE) of $\boldsymbol{\mu}$ is $\bar{\boldsymbol{y}}$. When $n > p$, $\mathbf{S}_n$ is the MLE of $\boldsymbol{\Sigma}$. When $n \leq p$ the likelihood can be made arbitrarily large at singular matrices, including $\mathbf{S}_n$.

2.2.1 Factor analysis

Factor analysis seeks to describe the covariance of p variables in terms of a few unobservable latent random variables called the *common factors*. The orthogonal factor model with k independent common factors assumes

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_k \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i, \quad \text{for} \quad i = 1, \ldots, n \tag{2.4}$$

where $\boldsymbol{\mu}$ is p-vector of unknown means, $\boldsymbol{u}_i$ is k-vector of independent unobserved random variables (common factors), $\boldsymbol{\Lambda}_k$ is a $p \times k$ matrix of unknown *factor loadings*, and $\boldsymbol{\varepsilon}_i$ is p-vector of unobserved random variables called the *specific factors*. We assume that the common and specific factors are independent with distributions

$$\boldsymbol{u}_i \overset{\text{i.i.d.}}{\sim} N_k(\mathbf{0}, \mathbf{I}_k) \quad \text{and} \quad \boldsymbol{\varepsilon}_i \overset{\text{i.i.d.}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Psi})$$

where $\boldsymbol{\Psi}$ is diagonal with j-th element $\psi_j > 0$. Under model (2.4), the marginal distributions of the vectors $\boldsymbol{y}_i$ are independent and identically distributed (i.i.d.) $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}. \tag{2.5}$$

If $\mathbf{H}$ is a $k \times k$ orthogonal matrix ($\mathbf{H}^T\mathbf{H} = \mathbf{H}\mathbf{H}^T = \mathbf{I}_k$), then an alternative parameterization of model (2.4) that still satisfies the requirements of the orthogonal k-factor model is

$$\begin{aligned} \boldsymbol{y}_i &= \boldsymbol{\mu} + (\boldsymbol{\Lambda}_k \mathbf{H}^T)(\mathbf{H}\boldsymbol{u}_i) + \boldsymbol{\varepsilon}_i \\ &= \boldsymbol{\mu} + \boldsymbol{\Lambda}_k^* \boldsymbol{u}_i^* + \boldsymbol{\varepsilon}_i \end{aligned}$$

where $\boldsymbol{u}_i^* = \mathbf{H}\boldsymbol{u}_i \overset{\text{i.i.d.}}{\sim} N(\mathbf{0}, \mathbf{I}_k)$. Under the rotation $\mathbf{H}$, the marginal covariance matrix $\boldsymbol{\Sigma}$ is parameterized as

$$\boldsymbol{\Lambda}_k^* \boldsymbol{\Lambda}_k^{*T} + \boldsymbol{\Psi} = \boldsymbol{\Lambda}_k \mathbf{H}^T \mathbf{H} \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k + \boldsymbol{\Psi}.$$

Thus the factor model is unique only up to orthogonal rotations. For identifiability we choose a particular rotation: we constrain $\boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda}_k = \boldsymbol{\Delta}$ to be diagonal. For identifiability, we also require k to be small enough so that the total number of parameters in the factor model is smaller than the number of distinct elements in the unrestricted model for $\boldsymbol{\Sigma}$,

$$pk + p - \tfrac{1}{2}k(k-1) \leq \tfrac{1}{2}p(p+1).$$

In cases where $n \leq p$, k must also be small enough so that the total number of parameters estimated (including the mean $\boldsymbol{\mu}$) is smaller than the total number of observations,

$$pk + 2p - \tfrac{1}{2}k(k-1) \leq np.$$

The parameters $\boldsymbol{\mu}$, $\boldsymbol{\Lambda}_k$, and $\boldsymbol{\Psi}$ can be estimated by maximum likelihood estimation based on the marginal distribution of $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$. Standard implementations of maximum likelihood factor analysis (e.g., SAS `proc factor` and `factanal()` in R) require more observations than variables ($n > p$); in Section 2.4.1 we discuss maximum likelihood estimation when $n \leq p$.

A key question in factor analysis, and for our use of factor analysis discussed in Section 2.4, is the choice of k, the number of common factors to include in the model. When $n > p$, maximum likelihood factor analysis provides a likelihood ratio goodness-of-fit test comparing the MLE of $\boldsymbol{\Sigma}$ under the k-factor model to the unrestricted MLE, $\mathbf{S}_n$. If $\widehat{\boldsymbol{\Sigma}}_k$ is the MLE of $\boldsymbol{\Sigma}$ under the k factor model, the likelihood ratio statistic T_λ is given by

$$-2 \log T_\lambda = np(\widehat{a} - \log \widehat{g} - 1)$$

where $\widehat{a}$ and $\widehat{g}$ are the arithmetic and geometric means of the eigenvalues of $\widehat{\mathbf{\Sigma}}_k^{-1}\mathbf{S}_n$. When the k-factor model is true, the statistic $-2\log T_\lambda$ has an asymptotic chi-squared distribution with

$$s = \tfrac{1}{2}(p-k)^2 - \tfrac{1}{2}(p+k)$$

degrees of freedom. In practice it is common to use Bartlett's small-sample corrected statistic,

$$U_k = (n'/n)T_\lambda = n'p(\widehat{a} - \log\widehat{g} - 1)$$

where $n' = n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}k$. When the k-factor model is true, the approximate distribution of U_k is chi-squared with s degrees of freedom. Mardia et al. (1979, Section 9.5) recommend trusting this test when $n \geq p + 50$.

When $n \leq p$, the unrestricted model cannot be fitted ($\mathbf{S}_n$ is singular), and the likelihood ratio test cannot be used. We might attempt to use a likelihood ratio test to compare the k-factor model to a model with more factors, but this approach is not tenable. Geweke and Singleton (1980) pointed out that when the k-factor model is true, any model with more than k common factors is not identifiable. We reproduce their argument in Section 2.4.2. Since no consistent estimator for $\mathbf{\Psi}$ can then exist, standard theory for the asymptotic distribution of the likelihood ratio test cannot be invoked. Thus we cannot use a likelihood ratio test to compare the k-factor model to the $(k+1)$-factor model, or any larger common factor model.

A less formal approach to determining an appropriate number of common factors, is to measure how much variation is explained by the common factors. The k-factor model is deemed sufficient when including additional common factors in the model does not substantially increase the amount of variation explained. The contribution of the common factors to the overall variation in the data can be measured in two ways. Let $\lambda_{j,m}$ be the (j,m)-th element in $\mathbf{\Lambda}_k$.

1. The first measure looks at how the common factors explain variation of each of the individual variables. From the covariance model (2.5) we see that the variance of the j-th

variable, σ_j^2, can be decomposed as

$$\sigma_j^2 = \lambda_{j1}^2 + \lambda_{j2}^2 + \ldots + \lambda_{jk}^2 + \psi_j.$$

The sum of the squares of the elements in the j-th row of $\mathbf{\Lambda}_k$, $h_j^2 = \lambda_{j1}^2 + \lambda_{j2}^2 + \ldots + \lambda_{jk}^2$, is called the j-th *communality*. The variance ψ_j of the j-th specific factor is called the j-th *specific variance*.

2. The second measure, looks at how much of the total variation is explained by the m-th common factor. The *total variation* in the sample is measured by

$$\begin{aligned} \text{trace}(\mathbf{\Sigma}) &= \sum_{j=1}^{p} \sigma_j^2 \\ &= \sum_{j=1}^{p} \left(\lambda_{j1}^2 + \lambda_{j2}^2 + \ldots + \lambda_{jk}^2 + \psi_j\right) \\ &= \sum_{m=1}^{k} \left(\lambda_{1m}^2 + \lambda_{2m}^2 + \ldots + \lambda_{pm}^2\right) + \sum_{j=1}^{p} \psi_j. \end{aligned}$$

Thus $(\lambda_{1m}^2 + \lambda_{2m}^2 + \ldots + \lambda_{pm}^2)/\text{trace}(\mathbf{\Sigma})$ is the proportion of total variation explained by the m-th common factor, and depends on the sum of the squares of the elements in the m-th column of $\mathbf{\Lambda}_k$.

Other aspects of factor analysis that we shall not consider here are rotations of the factor loadings, and prediction of the factor scores, $\boldsymbol{u}_i$. In our method described later we shall regress the estimated factor loadings on $\mathbf{F}_\beta$ or on $\mathbf{F}_\delta^*$, and there is no advantage to doing the regression on rotated loadings. Our approach also does not have any use for the estimated factor scores.

2.2.2 Principal component analysis

Principal component analysis (PCA) seeks a few weighted sums of the p variables that explain most of the covariance of the variables. There are several motivations for PCA; we begin with

the *empirical orthogonal functions* view that is popular in meteorological applications.

Let $\mathbf{Y}^*$ be the centered $n \times p$ data matrix

$$\mathbf{Y}^* = \mathbf{Y} - \tfrac{1}{n}\mathbf{1}\bar{\boldsymbol{y}}^T = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{Y}$$

with i-th row $\boldsymbol{y}_i^{*T} = (\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T$. The sample covariance matrix

$$\mathbf{S}_n = \tfrac{1}{n}\mathbf{Y}^T(\mathbf{I} - \mathbf{P}_1)\mathbf{Y} = \tfrac{1}{n}\mathbf{Y}^{*T}\mathbf{Y}^*$$

is symmetric. By the *spectral decomposition* theorem (e.g., Mardia et al., 1979, page 469) there exist matrices $\mathbf{G}$ and $\mathbf{L}$ where $\mathbf{G}$ is orthogonal and $\mathbf{L}$ is diagonal such that

$$\mathbf{S}_n = \mathbf{G}\mathbf{L}\mathbf{G}^T = \sum_{j=1}^{p} l_j \boldsymbol{g}^{(j)}\boldsymbol{g}^{(j)T}. \tag{2.6}$$

The diagonal elements of $\mathbf{L}$ are the *eigenvalues* of $\mathbf{S}_n$, with corresponding normalized *eigenvectors* $\boldsymbol{g}^{(j)}$ given by the columns of $\mathbf{G}$. The covariance matrix $\mathbf{S}_n$ is non-negative definite, therefore its eigenvalues are non-negative. We will assume that the eigenvalues are ordered, $l_1 \geq l_2 \geq \cdots \geq l_p \geq 0$.

For $k < \min(n, p)$, the empirical orthogonal functions view of the goal of PCA is to approximate $\mathbf{Y}^*$ by a rank k matrix $\widetilde{\mathbf{Y}}^*$ that minimizes the sum of squared differences between elements of $\mathbf{Y}^*$ and the approximating matrix $\widetilde{\mathbf{Y}}^*$,

$$\text{PCASSE}_k = \sum_{i=1}^{n} \left(\boldsymbol{y}_i^* - \widetilde{\boldsymbol{y}}_i^*\right)^T \left(\boldsymbol{y}_i^* - \widetilde{\boldsymbol{y}}_i^*\right). \tag{2.7}$$

According to this criterion, PCA seeks vectors $\widetilde{\boldsymbol{y}}_i^*$ that approximate the rows $\boldsymbol{y}_i^*$ of the centered data matrix. See Johnson and Wichern (2002, Supplement 8A) for more on the geometry of

this approximation. The criterion PCASSE_k is minimized by

$$\widehat{\mathbf{Y}}^* = \mathbf{Y}^*\mathbf{G}_k\mathbf{G}_k^T = \mathbf{V}_k\mathbf{G}_k^T$$

where $\mathbf{G}_k$ contains the first k columns of $\mathbf{G}$ from decomposition (2.6); i.e., the eigenvectors corresponding to the k largest eigenvalues of $\mathbf{S}_n$. The m-th column of $\mathbf{V}_k = \mathbf{Y}^*\mathbf{G}_k$ contains the scores for the m-th *principal component* (PC),

$$\boldsymbol{v}^{(m)} = \mathbf{Y}^*\boldsymbol{g}^{(m)}.$$

The columns of $\mathbf{G}_k$ are called the *principal component loadings*. The predicted value for the centered observation vector on the i-th subject is

$$\widehat{\boldsymbol{y}}_i^* = \sum_{m=1}^{k} v_{im}\boldsymbol{g}^{(m)} \quad \text{where} \quad v_{im} = \boldsymbol{y}_i^{*T}\boldsymbol{g}^{(m)}.$$

The analysis with k principal components therefore gives the approximation

$$\boldsymbol{y}_i \approx \bar{\boldsymbol{y}} + \mathbf{G}_k\boldsymbol{v}_i, \quad \text{for} \quad i = 1, \ldots, n \tag{2.8}$$

where $\boldsymbol{v}_i^T$ is the i-th row of $\mathbf{V}_k$. Note the similarity between the principal component approximation and the k-factor model in (2.4). One difference is that the principal component approximation makes no distributional assumptions about $\boldsymbol{y}_i$.

It is worth mentioning briefly the more traditional view of PCA. In this view, the m-th principal component $\boldsymbol{v}^{(m)} = \mathbf{Y}^*\boldsymbol{g}^{(m)}$ is the weighted sum of the columns of $\mathbf{Y}^*$ that has maximum sample variance $\boldsymbol{g}^{(m)T}\mathbf{S}_n\boldsymbol{g}^{(m)}$, subject to the constraints

$$\|\boldsymbol{g}^{(m)}\| = 1 \quad \text{and} \quad \boldsymbol{g}^{(m)} \perp \boldsymbol{g}^{(r)} \quad \text{for} \quad 1 \leq r < m.$$

The normalized eigenvector $\boldsymbol{g}^{(m)}$ of $\mathbf{S}_n$ corresponding to the eigenvalue l_m, maximizes $\boldsymbol{g}^{(m)T}\mathbf{S}_n\boldsymbol{g}^{(m)}$ subject to the normalization and orthogonality constraints.

One of the motivating problems introduced in Chapter 1 had $n > p$; the second problem had $n < p$. Let $r = \min(n, p)$. An efficient way to compute the eigenvalues of $\mathbf{S}_n$ when $n \leq p$ or $n > p$ is to use the *singular value decomposition* (SVD) of the $n \times p$ matrix,

$$n^{-1/2}\mathbf{Y}^* = \mathbf{H}\mathbf{T}\mathbf{G}_r^T \tag{2.9}$$

where $\mathbf{H}$ is $n \times r$ with $\mathbf{H}^T\mathbf{H} = \mathbf{I}_r$, $\mathbf{G}_r$ is $p \times r$ with $\mathbf{G}_r^T\mathbf{G}_r = \mathbf{I}_r$, and $\mathbf{T}$ is an $r \times r$ diagonal matrix. Implementations such as Matlab's `svd()` function have options to compute "economy" versions of the singular value decomposition where the dimension of $\mathbf{T}$ is the smaller of p and n. Using the SVD in equation (2.9) we can compute the first r terms of the spectral decomposition of $\mathbf{S}_n$

$$\mathbf{S}_n = \tfrac{1}{n}\mathbf{Y}^{*T}\mathbf{Y}^* = \mathbf{G}_r\mathbf{T}^T\mathbf{H}^T\mathbf{H}\mathbf{T}\mathbf{G}_r^T = \mathbf{G}_r\mathbf{L}_r\mathbf{G}_r^T = \sum_{j=1}^{r} l_j\boldsymbol{g}^{(j)}\boldsymbol{g}^{(j)T}$$

where $\mathbf{L}_r = \mathbf{T}^T\mathbf{T} = \mathbf{T}^2$, an $r \times r$ diagonal matrix containing the squared elements of $\mathbf{T}$ on the diagonal. The diagonal of $\mathbf{L}_r$ contains the largest r eigenvalues of $\mathbf{S}_n$, and the r columns of $\mathbf{G}_r$ contain the corresponding normalized eigenvectors. Relating this back to the spectral decomposition in (2.6), when $r < p$ we have

$$\mathbf{L} = \begin{bmatrix} \mathbf{L}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \quad \text{and} \quad \mathbf{G} = [\mathbf{G}_r, \mathbf{Q}]$$

where $\mathbf{Q}$ is a $p \times (p-r)$ matrix with orthogonal columns that are also orthogonal to the columns of $\mathbf{G}_r$. Using the SVD avoids having to compute zero eigenvalues and their corresponding eigenvectors when $\mathbf{S}_n$ is rank deficient, as in the case where $n < p$. It also ensures that the diagonal elements of $\mathbf{L}$ are real and non-negative, as they should be since $\mathbf{S}_n$ is symmetric and non-negative definite.

As in factor analysis, a key question in PCA is the choice of k, the rank of the approximating matrix $\widehat{\mathbf{Y}}^*$, or the number of principal components retained. The minimum of the PCASSE_k criterion

$$\min(\text{PCASSE}_k) = n\,(l_{k+1} + \cdots + l_p)$$

quantifies the amount of total sample variation not captured by k principal components. The total variation is $\text{trace}(\mathbf{S}_n) = \text{trace}(\mathbf{L}) = l_1 + l_2 + \cdots + l_p$ so that the proportion of total sample variation explained by the k principal components is

$$(l_1 + \cdots + l_k)\,/\text{trace}(\mathbf{S}_n).$$

The approximation with k principal components is deemed sufficient when including additional components does not substantially increase the proportion of total variation explained.

A formal likelihood ratio test for adequacy of the approximation with k principal components can be derived by assuming that $\boldsymbol{y}_i \overset{\text{i.i.d.}}{\sim} N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. If $\boldsymbol{\Sigma} = \boldsymbol{\Phi} + \sigma^2\mathbf{I}$ where $\boldsymbol{\Phi}$ is a rank k symmetric, non-negative definite matrix, then the smallest $(p-k)$ eigenvalues of $\boldsymbol{\Sigma}$ all equal σ^2. A test that the smallest $(p-k)$ eigenvalues of $\boldsymbol{\Sigma}$ are equal is used to test adequacy of the approximation with k principal components (e.g., Mardia et al. (1979, Section 8.4.3), Anderson (2003, Section 11.7.3)). Under the null hypothesis H_0: $\boldsymbol{\Sigma} = \boldsymbol{\Phi} + \sigma^2\mathbf{I}$, the smallest $(p-k)$ eigenvalues correspond to random noise which is evenly spread in all directions. The largest k eigenvalues correspond to "signal" or correlation structure among the variables.

When $n > p$, the likelihood ratio test statistic T_λ is given by

$$-2\log T_\lambda = n(p-k)\log(a_0/g_0)$$

where $a_0 = (l_{k+1} + \cdots l_p)/(p-k)$ and $g_0 = (l_{k+1} \times \cdots \times l_p)^{1/(p-k)}$ are the arithmetic and geometric means of the smallest $(p-k)$ eigenvalues of $\mathbf{S}_n$. Under the null hypothesis, the

asymptotic distribution of $-2\log T_\lambda$ is chi-squared with

$$s = \tfrac{1}{2}(p-k+2)(p-k-1)$$

degrees of freedom. In practice it is common to use Bartlett's small-sample correction

$$U_k = n'(p-k)\log(a_0/g_0)$$

where $n' = n - (2p+11)/6$. Under the hypothesis that the smallest $(p-k)$ eigenvalues of $\boldsymbol{\Sigma}$ are equal, the approximate distribution of U_k is chi-squared with s degrees of freedom.

When $n \leq p$, the sample covariance matrix $\mathbf{S}_n$ is singular and has rank $n-1$ (with probability 1). The smallest $p-n+1$ eigenvalues will be zero, $l_n = l_{n+1} = \cdots = l_p = 0$, and $g_0 = 0$. Thus the likelihood ratio test given above is not defined. In Section 2.3.1, we develop a test for adequacy of the approximation with k principal components for the case when $n \leq p$.

2.2.3 Comparison of factor analysis and PCA

Factor analysis models $\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_k \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i$, whereas PCA approximates $\boldsymbol{y}_i \approx \bar{\boldsymbol{y}} + \mathbf{G}_k \boldsymbol{v}_i$ where $\boldsymbol{v}_i^T$ is the i-th row of $\mathbf{V}_k$. Factor analysis models the marginal covariance as $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}$, whereas PCA decomposes $\mathbf{S}_n = \mathbf{G}\mathbf{L}\mathbf{G}^T \approx \mathbf{G}_k \mathbf{L}_k \mathbf{G}_k^T$.

Suppose that the population covariance matrix has the form $\boldsymbol{\Sigma} = \boldsymbol{\Phi} + \sigma^2 \mathbf{I}_p$ where $\boldsymbol{\Phi}$ is a rank k, symmetric, non-negative definite matrix. This is the form of the approximate marginal covariance matrix (2.2) under first-order linearization of the NLMM when within-subject variance is homoscedastic, $\psi_{ij} \equiv \sigma^2$. If $\theta_1^* \geq \theta_2^* \geq \cdots \geq \theta_k^* \geq 0$ are the k non-negative eigenvalues of $\boldsymbol{\Phi}$ (the remaining $(p-k)$ will be zero), then the eigenvalues of $\boldsymbol{\Sigma}$ are $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_p$ where $\theta_j = \theta_j^* + \sigma^2$ for $j = 1, \ldots, k$, and $\theta_j = \sigma^2$ for $j = k+1, \ldots, p$ (Harville, 1997, page 545). Thus the $(p-k)$ smallest eigenvalues of $\boldsymbol{\Sigma}$ all equal σ^2.

Let $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\boldsymbol{\Theta}\boldsymbol{\Gamma}^T$ be the spectral decomposition of $\boldsymbol{\Sigma}$. Partition $\boldsymbol{\Theta}$ and $\boldsymbol{\Gamma}$ into blocks of size

k and $(p-k)$ corresponding to the first k and smallest $(p-k)$ eigenvalues,

$$\mathbf{\Theta} = \begin{bmatrix} \mathbf{\Theta}_k & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_{(p-k)} \end{bmatrix} \quad \text{and} \quad \mathbf{\Gamma} = [\mathbf{\Gamma}_k, \mathbf{\Gamma}_{(p-k)}].$$

Since the matrix $\mathbf{\Gamma}$ is orthogonal, $\mathbf{\Gamma}_k \mathbf{\Gamma}_k^T + \mathbf{\Gamma}_{(p-k)} \mathbf{\Gamma}_{(p-k)}^T = \mathbf{I}_p$. Therefore we can decompose

$$\begin{aligned} \mathbf{\Sigma} &= \mathbf{\Gamma}_k \mathbf{\Theta}_k \mathbf{\Gamma}_k^T + \sigma^2 \mathbf{\Gamma}_{(p-k)} \mathbf{\Gamma}_{(p-k)}^T \\ &= \mathbf{\Gamma}_k \mathbf{\Theta}_k \mathbf{\Gamma}_k^T + \sigma^2 \left(\mathbf{I}_p - \mathbf{\Gamma}_k \mathbf{\Gamma}_k^T \right) \\ &= \mathbf{\Gamma}_k \left(\mathbf{\Theta}_k - \sigma^2 \mathbf{I}_k \right) \mathbf{\Gamma}_k^T + \sigma^2 \mathbf{I}_p. \end{aligned} \tag{2.10}$$

This is the factor decomposition with $\mathbf{\Psi} = \sigma^2 \mathbf{I}_p$ and $\mathbf{\Lambda}_k = \mathbf{\Gamma}_k (\mathbf{\Theta}_k - \sigma^2 \mathbf{I}_k)^{1/2}$. Recall that the diagonal elements of $\mathbf{\Theta}_k$ are the largest k eigenvalues of $\mathbf{\Sigma}$ and are therefore larger than σ^2. For $\mathbf{\Lambda}_k = \mathbf{\Gamma}_k (\mathbf{\Theta}_k - \sigma^2 \mathbf{I}_k)^{1/2}$,

$$\mathbf{\Lambda}_k^T \mathbf{\Psi}^{-1} \mathbf{\Lambda}_k = \sigma^{-2} (\mathbf{\Theta}_k - \sigma^2 \mathbf{I}_k)^{1/2} \mathbf{\Gamma}_k^T \mathbf{\Gamma}_k (\mathbf{\Theta}_k - \sigma^2 \mathbf{I}_k)^{1/2} = \sigma^{-2} (\mathbf{\Theta}_k - \sigma^2 \mathbf{I}_k)$$

which is diagonal so that the decomposition (2.10) also satisfies the identifiability constraint for the factor model. Thus when $\mathbf{\Psi} = \sigma^2 \mathbf{I}$, the principal component and factor decompositions of the population covariance matrix $\mathbf{\Sigma}$ are identical.

For large n, $\mathbf{S}_n \approx \mathbf{\Sigma}$ since $\mathbf{S}_n$ converges in probability to $\mathbf{\Sigma}$ as $n \to \infty$. For the distinct eigenvalues of $\mathbf{\Sigma}$, the sample eigenvalues and eigenvectors converge to the corresponding population counterparts (e.g., Mardia et al., 1979, page 229). Thus the principal component decomposition of $\mathbf{S}_n$ converges to the decomposition (2.10) in large samples. Similarly, the maximum likelihood factor model approximation $\widehat{\mathbf{\Lambda}}_k \widehat{\mathbf{\Lambda}}_k^T + \widehat{\mathbf{\Psi}}$ will converge to this same decomposition. Thus when $\mathbf{\Psi} = \sigma^2 \mathbf{I}$, the principal component and factor decompositions of the population covariance matrix are asymptotically equivalent.

In Appendix C.6.1 we derive the maximum likelihood factor estimator of $\mathbf{\Lambda}_k$ when $\mathbf{\Psi} = \sigma^2 \mathbf{I}_p$. We show that the columns of $\widehat{\mathbf{\Lambda}}_k$ are the (unnormalized) eigenvectors of $\mathbf{S}_n$ corresponding to the largest k eigenvalues. Thus in the case of homogeneous within-subject variation, the maximum likelihood factor loadings are just scaled versions of the principal component loadings.

2.2.4 Regularized or functional versions of PCA

Multivariate factor analysis and PCA ignore the fact that the p responses in vector $\boldsymbol{y}_i$ are measurements of the same variable made over time. *Functional principal component analysis* (FPCA) views the vectors $\boldsymbol{y}_i$ as realizations of a smooth function $y_i(t)$ observed at times $t_1, t_2, \ldots, t_p$. In FPCA, we seek a set of k orthonormal functions $\eta_1(t), \ldots, \eta_k(t)$ so that the approximation $\widehat{y}_i(t) = \sum_{m=1}^{k} c_{im}\eta_m(t)$ where $c_{im} = \int y_i \eta_m$, minimizes

$$\text{FPCASSE}_k = \sum_{i=1}^{n} \int \{y_i(s) - \widetilde{y}_i(s)\}^2 \, \mathrm{d}s.$$

Note the similarity between the FPCA criterion above, and the criterion (2.7) leading to the multivariate PCA approximation. In practice it is common to use *regularized* FPCA to ensure smoothness of the functions η_m. By *smoothness* we mean that $\int \{\eta''(t)\}^2 \, \mathrm{d}t$ is small. For more details on FPCA, refer to chapters 8 & 9 of Ramsay and Silverman (2005).

We did some work on penalized maximum likelihood estimation of a functional factor analysis model when $k = 1$ (our approach is sketched in Appendix A) but did not to pursue the functional or regularized approach because it did not gain us anything for our application. In particular, asymptotic theory for the functional approach (e.g., Cardot, 2000; Pezzulli and Silverman, 1993; Dauxois, Pousse, and Romain, 1982) did not help us with asymptotic approximations for the variance of the principal component loadings $\boldsymbol{g}^{(m)}$. Asymptotic theory for functional estimators has tended to focus on convergence in a functional norm, e.g., $\|\widehat{\eta}_m - \eta_m\|_2$, or mean integrated squared error, rather than the distributional properties of a vector, $\widehat{\boldsymbol{\eta}}_m$, whose components are $\widehat{\eta}_m$ evaluated at the time points, $t_1, t_2, \ldots, t_p$.

One advantage of the functional approach for our application might be the natural way in which it handles slightly unbalanced data through interpolation or smoothing, as long as observations on each curve are dense. A second gain might be that the functional estimators, which exploit smoothness, do not suffer the same asymptotic biases seen in multivariate PCA when p is large and n is small (e.g., Hall, Müller, and Wang, 2006; Johnstone, 2001), however we have not explored this in any depth.

A naive approach to smoothed PCA is to smooth the eigenvectors computed by usual multivariate PCA. This has the advantage that the smoothed loadings are a linear function of the unsmoothed loadings. This approach to smoothed PCA when $k = 1$ is discussed in Section 2.3.4.

2.2.5 Multivariate multiple regression

A tool that we shall use later in this chapter and in Chapter 4 is *multivariate multiple regression*. In multivariate multiple regression we examine the linear relationships between the p columns in the $n \times p$ data matrix $\mathbf{Y}$, with q variables in the columns of the $n \times q$ matrix $\mathbf{Z}$. The multivariate regression model is

$$\mathbf{Y} = \mathbf{Z}\mathbf{B} + \mathbf{E} \tag{2.11}$$

where the $q \times p$ matrix $\mathbf{B}$ has (h, j)-th element b_{hj}, which is the coefficient of the h-th explanatory variable $\boldsymbol{z}^{(h)}$ for the j-th response $\boldsymbol{y}^{(j)}$. If the first column of $\mathbf{Z}$ is $\mathbf{1}_n$, then the first row of $\mathbf{B}$ corresponds to an intercept for each column in $\mathbf{Y}$. The $n \times p$ matrix $\mathbf{E}$ contains the residual terms. We assume the rows of $\mathbf{E}$ are independent with zero mean and $p \times p$ covariance matrix $\boldsymbol{\Gamma}$.

Without making further distributional assumptions about the residuals $\mathbf{E}$, the coefficients may be estimated by least-squares

$$\widehat{\mathbf{B}} = \left(\mathbf{Z}^T\mathbf{Z}\right)^{-1}\mathbf{Z}^T\mathbf{Y}.$$

Sometimes it is of interest to partition $\mathbf{B}^T = \left[\mathbf{B}_{(a)}^T \mid \mathbf{B}_{(b)}^T\right]$, where $\mathbf{B}_{(a)}$ is $r \times p$ and $\mathbf{B}_{(b)}$ is $(q-r) \times p$ and to test the hypothesis

$$\mathrm{H}_0\text{: } \mathbf{B}_{(b)} = \mathbf{0}.$$

If we assume that the rows of $\mathbf{E}$ are independent and identically distributed normal random variables, $N_p(\mathbf{0}, \mathbf{\Gamma})$, then a likelihood ratio test of this hypothesis is given by Wilks' lambda statistic. Under the normality assumption, the least-squares estimator is also the maximum likelihood estimator (MLE) of $\mathbf{B}$ under the unrestricted model; call this estimator $\widehat{\mathbf{B}}_1$. The MLE of the $p \times p$ covariance matrix $\mathbf{\Gamma}$ is

$$\widehat{\mathbf{\Gamma}}_1 = \tfrac{1}{n}\left(\mathbf{Y} - \mathbf{Z}\widehat{\mathbf{B}}_1\right)^T\left(\mathbf{Y} - \mathbf{Z}\widehat{\mathbf{B}}_1\right).$$

Let $\widehat{\mathbf{B}}_0$ and $\widehat{\mathbf{\Gamma}}_0$ be the corresponding MLEs under H_0. Wilks' lambda statistic is a function of the likelihood ratio test statistic, and is given by

$$\Lambda^* = \Lambda^{2/n} = \det(\widehat{\mathbf{\Gamma}}_1)/\det(\widehat{\mathbf{\Gamma}}_0)$$

and H_0 is rejected when Λ^* is too small. When $n-q$ and $n-p$ are large, Bartlett's approximation (Mardia et al., 1979, page 84)

$$-\left[(n-q) - \tfrac{1}{2}(p-q+r+1)\right]\log\Lambda^*$$

has an approximate chi-squared distribution with $s = p(q-r)$ degrees of freedom.

Of particular interest to us will be the test comparing a model with $r-1$ variables ($r = 1, \ldots, q$) to a model with one additional variable. In this case, the exact distribution of Wilks'

lambda under the model with $r-1$ variables is given by

$$U_r = \frac{n-r-p+1}{p}(1-\Lambda^*)/\Lambda^* \sim F_{p,n-r-p+1} \tag{2.12}$$

where $F_{u,v}$ is the central F distribution with u and v degrees of freedom (Mardia et al., 1979, page 83).

To measure how well the multivariate regression model captures the variation in $\mathbf{Y}$ we propose using the *trace correlation* and *determinant correlation* coefficients defined by Mardia et al. (1979, page 171). We define these correlations for the uncentered data matrix $\mathbf{Y}$. Let $\widehat{\mathbf{E}} = \mathbf{Y} - \mathbf{Z}\widehat{\mathbf{B}}$ be the estimated residuals after fitting the multivariate regression model (2.11). Define $\mathbf{D} = (\mathbf{Y}^T\mathbf{Y})^{-1}\widehat{\mathbf{E}}^T\widehat{\mathbf{E}}$. If $\mathbf{Y}$ were a centered data matrix, then the matrix $\mathbf{D}$ would be a generalization of $1-R^2$ in the univariate multiple regression case. The matrix $\mathbf{I}_p - \mathbf{D}$ ranges between $\mathbf{0}$ and the identity matrix. If the model gives an exact prediction of $\mathbf{Y}$, $\mathbf{Y} = \mathbf{Z}\widehat{\mathbf{B}}$, then $\mathbf{I}_p - \mathbf{D}$ will be $\mathbf{I}_p$. On the other hand, if $\mathbf{Z}\widehat{\mathbf{B}} = \mathbf{0}$ so that the model explains none of the variation in $\mathbf{Y}$ then $\mathbf{I}_p - \mathbf{D} = \mathbf{0}$.

The trace correlation r_T, and the determinant correlation r_D, are defined as

$$r_T^2 = p^{-1}\mathrm{tr}(\mathbf{I}_p - \mathbf{D}) \quad \text{and} \quad r_D^2 = \det(\mathbf{I}_p - \mathbf{D}). \tag{2.13}$$

If the matrix of predictors, $\mathbf{Z}$, has column rank q that is smaller than p then some of the eigenvalues of $\mathbf{I}_p - \mathbf{D}$ will be zero and $r_D = 0$. To see this let $\mathbf{Z}$ have rank q and $\mathbf{P}_Z$ be the matrix that projects onto the column space of $\mathbf{Z}$. The rank of $\mathbf{P}_Z$ is q. The matrix $\mathbf{D}$ is

$$\mathbf{D} = (\mathbf{Y}^T\mathbf{Y})^{-1}\widehat{\mathbf{E}}^T\widehat{\mathbf{E}} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T(\mathbf{I}_n - \mathbf{P}_Z)\mathbf{Y} = \mathbf{I}_p - (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}\mathbf{P}_Z\mathbf{Y}^T$$

and $\mathbf{I}_p - \mathbf{D} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}\mathbf{P}_Z\mathbf{Y}^T$ which has rank less than or equal to $\min(p,q)$. If $q < p$ then $p-q$ of the eigenvalues of $\mathbf{I}_p - \mathbf{D}$ will be zero and the determinant correlation r_D will be zero. As Mardia et al. (1979) point out, r_D will be zero if just one eigenvalue of $\mathbf{I}_p - \mathbf{D}$ is

zero, but r_T will be zero only if all the eigenvalues of $\mathbf{I}_p - \mathbf{D}$ are zero. Both r_D and r_T will be 1 if and only if all the eigenvalues are 1. Thus if the model explains none of the variation in just one of the variables, r_D will be zero, but r_T will be positive. The determinant correlation r_D thus gives a low score to models that perform poorly on one variable, even if the model is good for other variables. If we want to ensure good predictions for all columns of $\mathbf{Y}$, then the determinant correlation r_D is the appropriate measure. However, r_D is meaningless when $\mathbf{Z}$ has fewer predictors than there are columns in $\mathbf{Y}$, so we will tend to use the trace correlation r_T.

Having introduced some of the standard multivariate tools that we will use, we move on to discussing two approaches to exploring the marginal covariance structure. The first is based on principal component analysis and is discussed in in Sections 2.3. We discuss the second approach, based on maximum likelihood factor analysis, in Section 2.4.

2.3 Exploring marginal covariance structure with principal component analysis

Assume that covariates are constant across the n subjects. Then first-order linearization of the NLMM gives the approximation of the stage-1 model,

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i.$$

The columns of $\mathbf{F}_\beta$ are the partial derivatives (sensitivities) of the nonlinear function f with respect to each coefficient in $\boldsymbol{\beta}$, evaluated at the time points $t_1, t_2, \ldots, t_p$. Under the assumption that the stage-1 residuals ε_{ij} in the NLMM are independent and identically distributed with variance σ^2, the first-order linearization gives the approximate marginal covariance structure

$$\mathbb{V}(\boldsymbol{y}_i) \approx \boldsymbol{\Sigma} = \mathbf{F}_\beta \mathbf{D} \mathbf{F}_\beta^T + \sigma^2 \mathbf{I}_p. \tag{2.14}$$

Under the assumption that $\mathbb{V}(\boldsymbol{\varepsilon}_i) = \sigma^2 \mathbf{I}_p$ it is appropriate to use *principal component analysis* (PCA) to approximate the marginal covariance matrix (see Section 2.2.3). As outlined in Section 2.2.2, PCA gives the approximation

$$\boldsymbol{y}_i \approx \bar{\boldsymbol{y}} + \mathbf{G}_k \boldsymbol{v}_i$$

where the columns of $\mathbf{G}_k$ are the eigenvectors of the sample covariance matrix $\mathbf{S}_n$ corresponding to the largest k eigenvalues. PCA approximates the sample covariance matrix by

$$\mathbf{S}_n \approx \mathbf{G}_k \mathbf{L}_k \mathbf{G}_k^T \tag{2.15}$$

where $\mathbf{L}_k$ is diagonal and contains the k largest eigenvalues of $\mathbf{S}_n$. This estimates the factorization

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \sigma^2 \mathbf{I}_p$$

of the population covariance matrix where $\boldsymbol{\Lambda}_k$ is a $p \times k$ matrix with rank k.

One question in exploring the structure of the marginal covariance matrix is to choose how many random effects are necessary to explain between-subject variation. Our first step in answering this question is to choose the number of principal components, k, required to give a good approximation to the marginal covariance matrix of the form (2.15). We address the problem of choosing k in Section 2.3.1.

The next step is to compare the PCA loadings matrix $\mathbf{G}_k$ to the matrix of sensitivity functions $\mathbf{F}_\beta$ in order to identify which elements of the coefficient vector $\boldsymbol{\beta}_i$ require non-trivial random effects. This comparison is complicated by several issues. Firstly the random effects in the NLMM may not be independent within subjects so it may not be straightforward to compare $\mathbf{G}_k$ with $\mathbf{F}_\beta$; for example, it is unlikely that a particular column of $\mathbf{G}_k$ will be proportional to a single column of $\mathbf{F}_\beta$. Secondly, although the PCA factorization may identify the need for a single random effect, a particular parameterization of the stage-1 model may require random

effects on two or more parameters in order to characterize the variation seen between subjects. Thirdly, for random effects with large magnitude, the first-order linear approximation may not be adequate, and second-order terms that are quadratic in elements of $\boldsymbol{b}_i$ may be needed to characterize between-subject variation. In this case, the PCA factorization which depends linearly on any latent variables, may identify the need for two random effects, whereas in fact a single large random effect on one parameter in the stage-1 model may suffice to capture the pattern of between-subject variation. These issues are addressed below. In Section 2.3.2 we discuss identifying which parameters in the NLMM need random effects; and in Section 2.3.3 we discuss choosing amongst several parameterizations the one that will give a simple explanation of the between-subject variation. We conclude the discussion of the principal component approach in Section 2.3.4 with an aside on smoothing the principal component loadings.

2.3.1 Selecting the number of principal components

In Section 2.2.2 we discussed informal ways to judge whether k principal components are sufficient, based on the proportion of variation explained by the k components. If the approximate covariance matrix can be factored as

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \sigma^2 \mathbf{I}$$

where $\boldsymbol{\Lambda}_k$ is $p \times k$ with rank k, then the smallest $(p-k)$ eigenvalues of $\boldsymbol{\Sigma}$ all equal σ^2. A formal test that k principal components are sufficient is equivalent to testing the hypothesis that the smallest $(p-k)$ eigenvalues of $\boldsymbol{\Sigma}$ are equal. As discussed in Section 2.2.2, when $n > p$, we can use the usual likelihood ratio test of this hypothesis. However, when $n \leq p$, this test does not exist. We propose an appropriate test for the case where $n \leq p$ below.

Case 1: $n > p$

When $n > p$, can use the usual (Bartlett corrected) likelihood ratio test for the hypothesis that $(p-k)$ eigenvalues of $\mathbf{\Sigma}$ are equal that we introduced in Section 2.2.2. The test statistic is

$$U_k = n'(p-k)\log(a_0/g_0) \tag{2.16}$$

where a_0 and g_0 are respectively the arithmetic and geometric means of the $(p-k)$ smallest eigenvalues of the sample covariance matrix $\mathbf{S}_n$, and $n' = n-(2p+11)/6$. Under the hypothesis that the $(p-k)$ smallest eigenvalues of $\mathbf{\Sigma}$ are equal, the asymptotic distribution of U_k is chi-squared with $s = \frac{1}{2}(p-k+2)(p-k-1)$ degrees of freedom.

Case 2: $n \leq p$

When $n \leq p$, the sample covariance matrix $\mathbf{S}_n$ is rank deficient and $g_0 = 0$, so the test described above does not exist. Let $\mathbf{Y}$ be the $n \times p$ data matrix. Recall that the $p \times p$ matrix $\mathbf{Y}^T\mathbf{Y}$ and the $n \times n$ matrix $\mathbf{Y}\mathbf{Y}^T$ both have rank n, and have the same positive eigenvalues. When $n \leq p$ we base our test on the eigenvalues of $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ rather than on the eigenvalues of the sample covariance matrix $\mathbf{S}_n = \frac{1}{n}\mathbf{Y}^{*T}\mathbf{Y}^*$.

The key result is that if the i-th row of $\mathbf{Y}$, $\boldsymbol{y}_i^T$, can be modeled as

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \mathbf{\Lambda}_k\boldsymbol{u}_i + \boldsymbol{\varepsilon}_i$$

where $\mathbf{\Lambda}_k$ has rank k, and $\boldsymbol{u}_i$ and $\boldsymbol{\varepsilon}_i$ are independent random vectors with $\boldsymbol{u}_i \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \mathbf{I}_k)$ and $\boldsymbol{\varepsilon}_i \overset{\text{i.i.d.}}{\sim} (\mathbf{0}, \sigma^2\mathbf{I}_p)$, then as $p \to \infty$, the $n \times n$ matrix $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ converges in probability to $\mathbf{\Sigma}^* = \mathbf{\Phi}^* + \sigma^2\mathbf{I}_n$ where $\mathbf{\Phi}^*$ is a positive definite matrix with rank $k+1$. In Theorem 2.3.1 we prove this result for $k = 1$. A similar argument extends the proof to general $k < n$.

Theorem 2.3.1. *Let* $\mathbf{Y}$ *be a* $n \times p$ *data matrix, and suppose that the* i*-th row,* $\boldsymbol{y}_i^T$ *can be*

modeled as

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\lambda} u_i + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\lambda}$ *is a* $p \times 1$ *vector,* u_i *are independent and identically distributed (i.i.d.) random variables with mean 0 and variance 1,* $\boldsymbol{\varepsilon}_i$ *are i.i.d. random vectors with mean* $\mathbf{0}$ *and covariance matrix* $\sigma^2 \mathbf{I}_p$*, and* u_i *and* $\boldsymbol{\varepsilon}_i$ *are independent. Then as* $p \to \infty$*, the* $n \times n$ *matrix* $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ *converges in probability to* $\boldsymbol{\Phi}^* + \sigma^2 \mathbf{I}_n$*, where* $\boldsymbol{\Phi}^*$ *is a positive definite* $n \times n$ *matrix with rank* 2*.*

Proof. Given in Appendix B.1.

The $n \times n$ matrix $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ is not the sample covariance matrix of a data matrix with p independent rows. Nonetheless, under the hypothesis that there are k common factors and homogeneous within-subject variation, $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ converges in probability to $\boldsymbol{\Sigma}^* = \boldsymbol{\Phi}^* + \sigma^2 \mathbf{I}_n$ where $\boldsymbol{\Phi}^*$ has rank $k+1$. The smallest $\{n - (k+1)\}$ eigenvalues of $\boldsymbol{\Sigma}^*$ equal σ^2.

The usual (i.e., when $n > p$) likelihood ratio test that the smallest $(p-k)$ eigenvalues of a $p \times p$ matrix $\boldsymbol{\Sigma}$ are equal, is equivalent to the test that $\boldsymbol{\Sigma} = \boldsymbol{\Phi} + \sigma^2 \mathbf{I}_p$ where $\boldsymbol{\Phi}$ has rank k (e.g., Anderson, 2003, Section 11.7.3). For the $n \leq p$ case, we propose using the test statistic that one would use, based on p independent observations, to test the hypothesis that the smallest $\{n - (k+1)\}$ eigenvalues of the $n \times n$ covariance matrix $\boldsymbol{\Sigma}^*$ are equal.

Following this strategy, we define a formal test of the hypothesis that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \sigma^2 \mathbf{I}_p$ when $n \leq p$, as follows. Let $l_1^* \geq l_2^* \geq \ldots \geq l_n^*$ be the eigenvalues of $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$. Let a_0^* and g_0^* be the arithmetic and geometric means of the $n-k-1$ smallest eigenvalues, $l_{k+2}^*, l_{k+3}^*, \ldots, l_n^*$. The likelihood ratio test statistic T_λ^* for the hypothesis that the smallest $\{n - (k+1)\}$ eigenvalues of $\boldsymbol{\Sigma}^*$ are equal (or equivalently that $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \sigma^2 \mathbf{I}_p$) is given by

$$-2 \log T_\lambda^* = p(n-k-1) \log(a_0^*/g_0^*).$$

Under the null hypothesis, the asymptotic distribution of $-2\log T_\lambda^*$ is chi-squared with

$$s = \tfrac{1}{2}\{n-(k+1)+2\}\{n-(k+1)-1\} = \tfrac{1}{2}(n-k+1)(n-k-2)$$

degrees of freedom.

As in the $n > p$ case, we recommend using the Bartlett small-sample corrected version of the test statistic

$$U_k^* = p'(n-k-1)\log(a_0^*/g_0^*) \tag{2.17}$$

where $p' = p - (2n+11)/6$. Under the hypothesis that there are k common factors, the approximate distribution of U_k^* is chi-squared with $s = \frac{1}{2}(n-k+1)(n-k-2)$ degrees of freedom.

Our proposed test is based on the uncentered data matrix $\mathbf{Y}$. Since we did not correct for the mean, one of the large eigenvalues of $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ represents the "signal" due to $\boldsymbol{\mu}$. We could have proposed a similar test based on the centered data matrix $\mathbf{Y}^* = (\mathbf{I}_n - \mathbf{P}_1)\mathbf{Y}$. However, it is more difficult to derive the limiting distribution of $\frac{1}{p}\mathbf{Y}^*\mathbf{Y}^{*T}$ than that of $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$. A potential disadvantage of the uncorrected version is that estimation of the mean signal from the covariance matrix is probably less precise than estimation by $\bar{\boldsymbol{y}}$, however we have not investigated the possible improvement of the test based on the centered matrix $\mathbf{Y}^*$, over that based on $\mathbf{Y}$.

2.3.2 Selection of random effects

In the previous section we proposed a test for choosing k, the number of principal components to give an approximation of the form

$$\boldsymbol{y}_i \approx \bar{\boldsymbol{y}} + \mathbf{G}_k\boldsymbol{v}_i \quad \text{or} \quad \mathbf{S}_n \approx \mathbf{G}_k\mathbf{L}_k\mathbf{G}_k^T.$$

Thus the between-subject variation is captured by a linear combination of the k orthogonal vectors $\boldsymbol{g}^{(1)}, \boldsymbol{g}^{(2)}, \ldots, \boldsymbol{g}^{(k)}$. Under the linearized model (2.1), we have

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad \text{and} \quad \mathbb{V}(\boldsymbol{y}_i) \approx \boldsymbol{\Sigma} = \mathbf{F}_\beta \mathbf{D} \mathbf{F}_\beta^T + \sigma^2 \mathbf{I}_p$$

where $\mathbf{F}_\beta$ has q columns. Under the linearization of the NLMM, the between-subject variation is characterized by the q vectors in $\mathbf{F}_\beta$ that are realizations of the sensitivity functions $\partial f / \partial \beta^{(m)} (m = 1, \ldots, q)$ at times $t_1, t_2, \ldots, t_p$. The first-order approximation will be adequate if the magnitude of random effects is small, and for now let us assume that this is the case. However, if one or more random effects are large, then a second-order approximation may be needed. In this case we may supplement $\mathbf{F}_\beta$ with several columns determined by the second partial derivatives of the function f. For example, if the variance of the random effect $b_i^{(1)}$ is expected to be large, then we may include a column which is the realization of $\partial^2 f / \partial \beta^{(1)2}$ evaluated at the time points $t_1, t_2, \ldots, t_p$.

Our approach to selecting which of the q parameters in $\boldsymbol{\beta}$ have non-trivial random effects is to investigate the relationship between the columns of $\mathbf{G}_k$ and those of $\mathbf{F}_\beta$. *Multivariate multiple regression* (reviewed in Section 2.2.5) is one tool for doing this. To do this we fit the multivariate regression model

$$\mathbf{G}_k = \mathbf{Z}\mathbf{B} + \mathbf{E} \tag{2.18}$$

where $\mathbf{Z}$ contains a subset of the columns of $\mathbf{F}_\beta$, and we assume the rows of the residual matrix $\mathbf{E}$ to be independent with zero mean and covariance matrix $\boldsymbol{\Gamma}$. We hope to identify k sensitivity functions that explain most of the variation in $\mathbf{G}_k$. To choose the most important random effects from amongst the columns of $\mathbf{F}_\beta$ we propose using *forward selection* with Wilks' lambda test statistic defined in (2.12) as the inclusion criterion. This statistic compares a model with $r - 1$ variables to a model with one additional variable.

The forward selection procedure begins by fitting the null model $\mathbf{G}_k = \mathbf{E}$. Let $\widehat{\boldsymbol{\Gamma}}_0$ denote the

maximum likelihood estimator of $\mathbf{\Gamma}$ based on this null model. At the first step of the procedure we examine all q columns of $\mathbf{F}_\beta$, including them in the model one at a time. Let $\widehat{\mathbf{\Gamma}}_1^{(m)}$ denote the maximum likelihood estimator of $\mathbf{\Gamma}$ based on the model including the intercept and the m-th predictor. For each predictor, we compute the Wilks' Lambda statistic

$$\Lambda_1^{*(m)} = \det(\widehat{\mathbf{\Gamma}}_1^{(m)})/\det(\widehat{\mathbf{\Gamma}}_0), \quad m = 1, \ldots, q.$$

Under the hypothesis that the m-th predictor has zero coefficient, the distribution of $\Lambda_1^{*(m)}$ is given by

$$U_1^{(m)} = \tfrac{n-p}{p}(1 - \Lambda_1^{*(m)})/\Lambda_1^{*(m)} \sim F_{p,n-p}.$$

Letting m^* be the index of the predictor with the smallest value of $\Lambda_1^{*(m)}$, we decide whether to include the m^*-th predictor in the model by comparing $U_1^{*(m^*)}$ with its α-level critical value obtained from the above F distribution. If the test statistic is significant, we include the m^*-th predictor in the model, and proceed to the next step where the remaining candidate predictors are tested for inclusion into the model. If the test is not significant, we stop adding predictors to the model.

At the r-th step of this stepwise procedure ($r = 1, \ldots, q$) we proceed similarly. At this step, the null model contains $r-1$ sensitivities from $\mathbf{F}_\beta$. We consider the $q - r + 1$ sensitivities still not in the model. For the m-th remaining predictor, we compute Wilks Lambda

$$\Lambda_r^{*(m)} = \det(\widehat{\mathbf{\Gamma}}_r^{(m)})/\det(\widehat{\mathbf{\Gamma}}_{r-1})$$

where $\widehat{\mathbf{\Gamma}}_{r-1}$ is the MLE of $\mathbf{\Gamma}$ based on the model including $r-1$ sensitivities, and $\widehat{\mathbf{\Gamma}}_r^{(m)}$ is the MLE of $\mathbf{\Gamma}$ based on the model that also includes the m-th sensitivity. In this case, the distribution of the Wilks' Lambda test statistic under the hypothesis that the m-th predictor

has zero coefficient is given by

$$U_r^{(m)} = \tfrac{n-r-p+1}{p}(1 - \Lambda_r^{*(m)})/\Lambda_r^{*(m)} \sim F_{p,n-r-p+1}.$$

This is the test statistic given in (2.12).

Exact test when $k = 1$

The forward selection procedure with selection criterion given by (2.12) assumes that the rows of the residual matrix $\mathbf{E}$ in model (2.18) are independent and normally distributed. Although this will not be strictly true, we expect the procedure to work adequately in practice. When $k = 1$, we have derived an exact test for strength of association between the principal eigenvector, $\boldsymbol{g}^{(1)}$, and an arbitrary p-vector $\boldsymbol{\kappa}$. When PCA has identified one significant component of between-subject variation, we may use this test to choose the most strongly associated sensitivity. Under the hypothesis that $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$ (no principal components), the distribution of $\boldsymbol{g}^{(1)}$ equals that of $\boldsymbol{z}/\|\boldsymbol{z}\|$ where $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$. This is the uniform distribution on the unit sphere in $\mathbb{R}^p$. This result is implied by Theorem 13.3.3 of Anderson (2003) and work of Stewart (1980); in Theorem 2.3.2 we prove a version of this result that suits our application.

Theorem 2.3.2. *Let $\mathbf{Y}$ be a $n \times p$ data matrix with each row independent and identically distributed according to the $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{I}_p)$ distribution. Let $\mathbf{S}_n$ be the sample covariance matrix, and let $\boldsymbol{g}^{(1)}$ be the normalized eigenvector of $\mathbf{S}_n$ corresponding to its largest eigenvalue. Then the distribution of $\boldsymbol{g}^{(1)}$ is the same as the distribution of $\boldsymbol{z}/\|\boldsymbol{z}\|$ where $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I})$.*

Proof. Given in Appendix B.2.

Let $\boldsymbol{\kappa}$ be an arbitrary p-vector. For identifiability we will usually constrain the norm of $\boldsymbol{\kappa}$, e.g., $\|\boldsymbol{\kappa}\|^2 = p$ or $\|\boldsymbol{\kappa}\| = 1$. We make use of the above result about the distribution of $\boldsymbol{g}^{(1)}$ to

test the hypothesis

$$\begin{aligned} \mathrm{H}_0\text{:} \quad & \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}, \quad \text{versus} \\ \mathrm{H}_1\text{:} \quad & \boldsymbol{\Sigma} = \sigma_1^2 \boldsymbol{\kappa}\boldsymbol{\kappa}^T + \sigma^2 \mathbf{I}_p \quad \text{where} \quad \sigma_1^2 > 0. \end{aligned}$$

We propose the following test statistic. Let $r_r = \boldsymbol{\kappa}^T \boldsymbol{g}^{(1)} / \|\boldsymbol{\kappa}\|$ and $t_r = (p-1)^{1/2} r_r (1 - r_r^2)^{-1/2}$. Reject H_0 in favor of H_1 when t_r is large. In Theorem 2.3.3 we prove that under H_0, t_r has a student's t-distribution with $p-1$ degrees of freedom.

Theorem 2.3.3. *Let $\mathbf{Y}$ be a $n \times p$ data matrix with each row independent and identically distributed according to the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{S}_n$ be the sample covariance matrix, and let $\boldsymbol{g}^{(1)}$ be the normalized eigenvector of $\mathbf{S}_n$ corresponding to its largest eigenvalue. Let $\boldsymbol{\kappa}$ be an arbitrary p-vector. Define*

$$r_r = \boldsymbol{\kappa}^T \boldsymbol{g}^{(1)} / \|\boldsymbol{\kappa}\| \quad \textit{and} \quad t_r = (p-1)^{1/2} r_r (1 - r_r^2)^{-1/2}. \tag{2.19}$$

Then $|r_r| \leq 1$ with $|r_r| = 1$ if and only if $\boldsymbol{g}^{(1)}$ and $\boldsymbol{\kappa}$ are linearly dependent. If $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$ then r_r has the distribution of a sample correlation coefficient with $\rho = 0$ and $p-1$ degrees of freedom, and t_r has a t-distribution with $p-1$ degrees of freedom.

Proof. Given in Appendix B.3.

Note that r_r is an uncentered correlation. The lack of centering accounts for the extra degree of freedom; the distribution for a centered correlation coefficient based on p data points would have $p-2$ degrees of freedom (Anderson, 2003, Theorem 4.2.1).

In Theorem 2.3.2 we derived the distribution of $\boldsymbol{g}^{(1)}$ under the assumption that $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$. When $\boldsymbol{\Sigma}$ is not $\sigma^2 \mathbf{I}_p$, the exact distribution of the eigenvectors of $n\mathbf{S}_n$ is not known. The asymptotic distributions of the eigenvalues and eigenvectors of $n\mathbf{S}_n$ as $n \to \infty$ have been derived for general $\boldsymbol{\Sigma}$ by several authors (e.g., Anderson, 2003, Theorem 13.5.1). While these

results are not particularly useful in our case where the number of subjects n is likely to be small, and sometimes smaller than the dimension p of $\mathbf{S}_n$, the asymptotic results do give some justification for using the test statistic defined in Theorem 2.3.3. When $\boldsymbol{\Sigma} = \sigma_1^2 \boldsymbol{\kappa}\boldsymbol{\kappa}^T + \sigma^2 \mathbf{I}_p$,

$$\boldsymbol{\Sigma}\boldsymbol{\kappa} = \sigma_1^2 \boldsymbol{\kappa}\|\boldsymbol{\kappa}\|^2 + \sigma^2 \boldsymbol{\kappa} = (\sigma_1^2 \|\boldsymbol{\kappa}\|^2 + \sigma^2)\boldsymbol{\kappa}$$

so that $\sigma_1^2 + \sigma^2$ is the largest eigenvalue of $\boldsymbol{\Sigma}$ with corresponding normalized principal eigenvector $\boldsymbol{\kappa}/\|\boldsymbol{\kappa}\|$. The largest eigenvalue has multiplicity one, so according to the asymptotic results referred to above, the limiting distribution of $n^{1/2}(\boldsymbol{g}^{(1)} - \boldsymbol{\kappa}/\|\boldsymbol{\kappa}\|)$ under H_1 is asymptotically normal with mean $\mathbf{0}$. Since $\boldsymbol{\kappa}/\|\boldsymbol{\kappa}\|$ is a constant, we have $\boldsymbol{g}^{(1)} \to \boldsymbol{\kappa}/\|\boldsymbol{\kappa}\|$ in probability as $n \to \infty$. Therefore, under H_1, $|r_r| \to \boldsymbol{\kappa}^T\boldsymbol{\kappa}/\|\boldsymbol{\kappa}\|^2 = 1$ in probability as $n \to \infty$. When $\boldsymbol{\Sigma} = \sigma_1^2 \boldsymbol{\kappa}\boldsymbol{\kappa}^T + \sigma^2 \mathbf{I}_p$ we can expect $|t_r|$ to be large.

2.3.3 Selection of an appropriate parameterization

Let $f(t, \boldsymbol{\beta}, \boldsymbol{\xi}; \boldsymbol{x})$ and $f^*(t, \boldsymbol{\delta}, \boldsymbol{\xi}; \boldsymbol{x})$ be alternative parameterizations of a nonlinear regression model. As discussed in Section 1.4, if $\boldsymbol{\beta}$ is a one-to-one function of $\boldsymbol{\delta}$, then the difference between parameterizations will be revealed by the difference between the terms $\mathbf{F}_\beta \boldsymbol{b}_i$ and $\mathbf{F}_\delta^* \boldsymbol{d}_i$ in the first-order linear approximations of equations (1.10) and (1.11) of Chapter 1, namely

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \quad \text{and} \quad \boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\delta^* \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i.$$

To select the parameterization that best (most parsimoniously) characterizes between-subject variation, our strategy is to compare the multivariate multiple regressions of $\mathbf{G}_k$ on $\mathbf{F}_\beta$ and $\mathbf{G}_k$ on $\mathbf{F}_\delta^*$. For each model, the multivariate forward selection procedure described in Section 2.3.2 may be used to select columns from $\mathbf{F}_\beta$ or $\mathbf{F}_\delta^*$, that best explain the variation in $\mathbf{G}_k$. We would like to answer the following questions.

1. Does one parameterization give a simpler characterization of the between-subject variation

than the other?

2. Which parameterization gives a better overall fit?

The first question may be answered by comparing the number of columns of selected from $\mathbf{F}_\beta$ or from $\mathbf{F}^*_\delta$. For example, suppose that $k = 1$, and that one column from $\mathbf{F}^*_\delta$ is selected by forward selection, but two are selected from $\mathbf{F}_\beta$. If both regressions account for a similar proportion of the variation in $\mathbf{G}_1$, then we would conclude that the parameterization $f^*(t, \boldsymbol{\delta}, \boldsymbol{\xi}; \boldsymbol{x})$ gives a simpler characterization of between-subject variation than the parameterization $f(t, \boldsymbol{\beta}, \boldsymbol{\xi}; \boldsymbol{x})$. The second question may be answered by quantifying the proportion of variation in $\mathbf{G}_k$ that is explained by the each of the two multivariate regressions. For an overall measure of how each model simultaneously explains variation in all the columns of $\mathbf{G}_k$, we propose using the trace correlation, r_T, defined in equation (2.13).

Answering these question may be complicated by the fact that our method is based on a first-order linear approximation in the random effects. If a random effect in one parameterization is large, then the linear approximation on which the PCA and the multivariate regression analysis are based, may be poor for this parameterization. In this case, we recommend supplementing the matrices $\mathbf{F}_\beta$ or $\mathbf{F}^*_\delta$ with second-order partial derivatives to account for the need for a second-order approximation to the stage-1 regression model.

Our ability to detect differences in parameterizations has several limitations. Firstly, our method which is based on linearization in the random effects, will only be able to detect differences between certain types of parameterizations. For example, consider the function describing plasma concentration in an open one-compartment model,

$$f(t, \boldsymbol{\beta}) = \beta^{(1)} \exp(-\beta^{(2)} t) - \beta^{(3)} \exp(-\beta^{(4)} t).$$

Consider the following three parameterizations of the first term of this function,

$$\begin{aligned} f^{(1)}(t, \beta^{(1)}, \beta^{(2)}) &= \beta^{(1)} \exp(-\beta^{(2)} t) \\ f^{(2)}(t, \delta^{(1)}, \delta^{(2)}) &= \mathrm{e}^{\delta^{(1)}} \exp(-\mathrm{e}^{\delta^{(2)}} t) \\ f^{(3)}(t, \gamma^{(1)}, \gamma^{(2)}) &= \frac{c}{\gamma^{(1)}(c - \gamma^{(2)}/\gamma^{(1)})} \exp(-\gamma^{(2)} t/\gamma^{(1)}). \end{aligned}$$

The parameterization $f^{(3)}$ mimics the volume-clearance parameterization of the open one-compartment model, with $\gamma^{(1)} = V$ (volume of distribution) and $\gamma^{(2)} = Cl$ (clearance). For simplicity, we assume for now that c is a known constant. The parameter vector $(\beta^{(1)}, \beta^{(2)})$ is a one-to-one function of the elements of $(\delta^{(1)}, \delta^{(2)})$ and also of the elements of $(\gamma^{(1)}, \gamma^{(2)})$.

The partial derivatives of these three functions with respect to their two parameters are

$$\begin{aligned} &\partial f^{(1)}/\partial \beta^{(1)} = \exp(-\beta^{(2)} t), \quad \partial f^{(1)}/\partial \beta^{(2)} = -\beta^{(1)} t \exp(-\beta^{(2)} t); \\ &\partial f^{(2)}/\partial \delta^{(1)} = \mathrm{e}^{\delta^{(1)}} \exp(-\mathrm{e}^{\delta^{(2)}} t), \quad \partial f^{(2)}/\partial \delta^{(2)} = -\mathrm{e}^{\delta^{(1)}} \mathrm{e}^{\delta^{(2)}} t \exp(-\mathrm{e}^{\delta^{(2)}} t); \quad \text{and} \\ &\partial f^{(3)}/\partial \gamma^{(1)} = \frac{c \exp(-\gamma^{(2)} t/\gamma^{(1)})}{\gamma^{(1)2}(c - \gamma^{(2)}/\gamma^{(1)})} \left[-1 - \gamma^{(2)} / \left\{ \gamma^{(1)}(c - \gamma^{(2)}/\gamma^{(1)}) \right\} + \gamma^{(2)} t/\gamma^{(1)} \right], \\ &\partial f^{(3)}/\partial \gamma^{(2)} = \frac{c \exp(-\gamma^{(2)} t/\gamma^{(1)})}{\gamma^{(1)2}(c - \gamma^{(2)}/\gamma^{(1)})} \left\{ -1/(c - \gamma^{(2)}/\gamma^{(1)}) - t \right\}. \end{aligned}$$

The sensitivities $\partial f^{(1)}/\partial \beta^{(1)}$ and $\partial f^{(2)}/\partial \delta^{(1)}$ differ only by a constant, as do the partial derivatives of these two functions with respect to $\beta^{(2)}$ and $\delta^{(2)}$. Since $\mathbf{F}_{\beta}^{(1)} \propto \mathbf{F}_{\delta}^{(2)}$, our method will not be able to discriminate between the parameterizations $f^{(1)}$ and $f^{(2)}$. For example we would not be able to use our method to determine whether a random effect on the scale parameter has the form $\beta^{(1)} + b_i^{(1)}$ or $\exp(\delta^{(1)} + d_i^{(1)})$. Suppose that the true parameterization is $f^{(2)}$. If $\boldsymbol{b}_i$ and $\boldsymbol{d}_i$ are small then the differences between these two parameterizations will not matter since

$$\exp(\delta^{(1)} + d_i^{(1)}) \approx \exp(\delta^{(1)}) + d_i^{(1)} \exp(\delta^{(1)}) = \beta^{(1)} + b_i^{(1)}.$$

However as the random effect $d_i^{(1)}$ gets larger, one dataset will be multivariate normal and the

other will not. In practice the parameterization $f^{(2)}$ may be selected in preference to $f^{(1)}$ on the basis that the distribution of coefficients in the population is expected to be lognormal rather than normal, or to ensure positivity of estimated coefficients.

The sensitivities $\partial f^{(3)}/\partial\gamma^{(1)}$ and $\partial f^{(3)}/\partial\gamma^{(2)}$ are not constant multiples of $\partial f^{(1)}/\partial\beta^{(1)}$ and $\partial f^{(1)}/\partial\beta^{(2)}$. If there were one random effect on say $\gamma^{(1)}$ then we could use our method to identify parameterization $f^{(3)}$ as being preferable to parameterization $f^{(1)}$. The key is that $\beta^{(1)}$ is a non-trivial function of both $\gamma^{(1)}$ and $\gamma^{(2)}$, whereas $\beta^{(1)}$ is a function of $\delta^{(1)}$ only, and $\beta^{(2)}$ is a function of $\delta^{(2)}$ only.

The second limitation to our method's ability to identify an appropriate parameterization is that if all q parameters in a particular parameterization require random effects, then our method will not be able to discriminate between alternative parameterizations. Notice that in the above example $\partial f^{(3)}/\partial\gamma^{(1)}$ and $\partial f^{(3)}/\partial\gamma^{(2)}$ can each be written as a linear combination of $\partial f^{(1)}/\partial\beta^{(1)}$ and $\partial f^{(1)}/\partial\beta^{(2)}$. Since $\mathbf{F}_\gamma^{(3)} = \mathbf{F}_\beta^{(1)}\mathbf{R}$ for some 2×2 matrix $\mathbf{R}$, our method will not be able to discriminate between parameterizations $f^{(1)}$ and $f^{(3)}$ if there are two random effects. By applying the Chain Rule for partial derivatives we see that this is true in general. If the $q\times 1$ vector $\boldsymbol{\beta} = \boldsymbol{\beta}(\boldsymbol{\delta})$ is a one-to-one function of a $q\times 1$ vector $\boldsymbol{\delta}$ then the m-th column of $\mathbf{F}_\delta^*$ is given by evaluating

$$\frac{\partial f^*}{\partial\delta^{(m)}} = \frac{\partial f}{\partial\beta^{(1)}}\cdot\frac{\partial\beta^{(1)}}{\partial\delta^{(m)}} + \frac{\partial f}{\partial\beta^{(2)}}\cdot\frac{\partial\beta^{(2)}}{\partial\delta^{(m)}} + \ldots + \frac{\partial f}{\partial\beta^{(q)}}\cdot\frac{\partial\beta^{(q)}}{\partial\delta^{(m)}}$$

at the time points $t_1, t_2, \ldots, t_p$. Let $\mathbf{J}$ be the $q\times q$ matrix with (i,j)-th element $\partial\beta^{(i)}/\partial\delta^{(j)}$. We can write $\mathbf{F}_\delta^* = \mathbf{F}_\beta\mathbf{J}$. Thus the multivariate regression of $\mathbf{G}_k$ on all q columns of $\mathbf{F}_\delta^*$ will fit as well as the multivariate regression of $\mathbf{G}_k$ on all q columns of $\mathbf{F}_\beta$ and we cannot use our method to choose between the two parameterizations.

To summarize, we expect our method to be able to discriminate between two parameterizations $f(t,\boldsymbol{\beta},\boldsymbol{\xi};\boldsymbol{x})$ and $f^*(t,\boldsymbol{\delta},\boldsymbol{\xi};\boldsymbol{x})$ only if one or more elements in $\boldsymbol{\beta}$ are each functions of several elements in $\boldsymbol{\delta}$, and the number of random effects needed in the true parameterization is

smaller than q, the dimension of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

2.3.4 Smoothing the PCA loadings

In this section we give a brief review of a naive approach to ensuring that the principal component loadings are smooth. The columns in $\mathbf{F}_\beta$ are realizations of the sensitivity functions, $f_{\beta^{(m)}}$, $m = 1, \ldots, q$, evaluated at the time points $t_1, \ldots, t_p$, and hence are expected to be smooth. It seems sensible to require columns of $\mathbf{G}_k$ to be similarly smooth. To ensure this, we could employ functional PCA (mentioned in Section 2.2.4) instead of multivariate PCA, or the functional factor approach outlined in Appendix A. A naive alternative is to smooth each column of $\mathbf{G}_k$ directly using a cubic smoothing spline.

In Appendix A we describe our approach to functional factor analysis when $k = 1$. In that discussion we show that when the smoothing parameter approaches zero (no smoothing), the estimator of the functional factor is the natural cubic spline interpolator of $\boldsymbol{g}^{(1)}$, the principal eigenvector of the sample covariance matrix. When the smoothing parameter approaches infinity (maximum smoothing), the estimator is the best linear approximation of $\boldsymbol{g}^{(1)}$. These observations motivate smoothing $\boldsymbol{g}^{(1)}$ (and subsequent eigenvectors) directly using a cubic smoothing spline. In general, this naive approach does not give the same estimator as the functional factor analysis approach. However, for zero or infinite smoothing, the naive approach will give the same fit as the functional factor estimator. For values of the smoothing parameter between zero and infinity, the naive estimator will have smoothness somewhere between exact interpolation of $\boldsymbol{g}^{(1)}$ and a linear approximation.

The naive approach has the advantage over a more rigorous approach like functional PCA, that each smoothed eigenvector is a linear function of the unsmoothed eigenvector. So for example, if $k = 1$ it is straightforward to extend the exact test defined in Theorem 2.3.3 to the smoothed eigenvector. Assume the following model for the j-th element of the eigenvector $\boldsymbol{g}^{(1)}$

$$g_j^{(1)} = \eta(t_j) + e_j \tag{2.20}$$

where η is a smooth function (e.g., twice continuously differentiable) and e_j are independent and identically distributed $N(0, \sigma_g^2)$ random variables. In vector notation, the model is

$$\boldsymbol{g}^{(1)} = \boldsymbol{\eta} + \boldsymbol{e}, \quad \text{where } \boldsymbol{e} \sim N_p(\mathbf{0}, \sigma_g^2 \mathbf{I}_p).$$

Note that this model is only a working model; it is not logically derived from the original NLMM, the first-order linear approximation, or the common factor model. In fact, the vector $\boldsymbol{g}^{(1)}$ will have a non-diagonal covariance structure because of the normalization.

To estimate η by cubic smoothing splines, we minimize the penalized least squares objective function

$$\sum_{j=1}^{p} \left\{ g_j^{(1)} - \eta(t_j) \right\}^2 + \alpha \int (\eta'')^2$$

over the space of appropriately smooth functions. The solution is the natural cubic spline interpolator of $\widehat{\boldsymbol{g}}_s^{(1)}$, a linear transformation of the original eigenvector,

$$\widehat{\boldsymbol{g}}_s^{(1)} = \mathbf{A}(\alpha) \boldsymbol{g}^{(1)}. \tag{2.21}$$

The subscript s indicates that $\widehat{\boldsymbol{g}}_s^{(1)}$ is the smoothed eigenvector. The matrix $\mathbf{A}(\alpha)$ is the so-called smoothing matrix, which depends on the measurement points t_j and the smoothing parameter α (e.g., Green and Silverman, 1994, equation (2.11)). The smoothing matrix is symmetric but not idempotent (unless $\alpha = 0$). The smoothing parameter α may be selected using cross validation. We used the Hutchinson and de Hoog algorithm described in Green and Silverman (1994, Section 3.2.2) for computation of the cross validation criterion.

The eigenvector was constrained so that $\boldsymbol{g}^{(1)T} \boldsymbol{g}^{(1)} = 1$. However, we have not constrained the smoothed estimator $\widehat{\boldsymbol{g}}_s^{(1)}$ to have norm 1. This does not cause an identification problem: the cubic spline smooth of $\boldsymbol{g}^{(1)}$ is unique if $\boldsymbol{g}^{(1)}$ is unique. Any scaling of $\widehat{\boldsymbol{g}}_s^{(1)}$ by a constant (e.g., to correct the norm), will pull the estimator $\widehat{\boldsymbol{g}}_s^{(1)}$ away from the principal eigenvector $\boldsymbol{g}^{(1)}$.

Since the smoothed estimator defined in equation (2.21) is a linear transformation of $\boldsymbol{g}^{(1)}$,

we may easily obtain its distribution from that of $\boldsymbol{g}^{(1)}$. By Theorem 2.3.2, when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$,

$$\widehat{\boldsymbol{g}}_s^{(1)} = \mathbf{A}(\alpha)\boldsymbol{g}^{(1)}$$

has the same distribution as $\mathbf{A}(\alpha)\boldsymbol{z}/\|\boldsymbol{z}\|$ where $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$. To compare the hypotheses

$$\begin{aligned} \mathrm{H}_0\text{:} \quad & \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}, \quad \text{versus} \\ \mathrm{H}_1\text{:} \quad & \boldsymbol{\Sigma} = \sigma_1^2 \boldsymbol{\kappa}\boldsymbol{\kappa}^T + \sigma^2 \mathbf{I}_p \quad \text{where} \quad \sigma_1^2 > 0. \end{aligned}$$

we can define a test statistic based on the smoothed eigenvector. Let $\tilde{\boldsymbol{\kappa}}_\alpha = \mathbf{A}(\alpha)\boldsymbol{\kappa}$ be the result of smoothing the vector $\boldsymbol{\kappa}$. Note that the smoothing parameter α is chosen (e.g., by cross-validation) in order to smooth the eigenvector $\boldsymbol{g}^{(1)}$, not the vector $\boldsymbol{\kappa}$. Recall that $\mathbf{A}(\alpha)$ is symmetric so that $\mathbf{A}(\alpha)\boldsymbol{\kappa} = \mathbf{A}^T(\alpha)\boldsymbol{\kappa}$. Consider the ratio

$$r_s = \frac{\boldsymbol{\kappa}^T \widehat{\boldsymbol{g}}_s^{(1)}}{\|\tilde{\boldsymbol{\kappa}}_\alpha\| \|\boldsymbol{g}^{(1)}\|} = \frac{\boldsymbol{\kappa}^T \{\mathbf{A}(\alpha)\boldsymbol{g}^{(1)}\}}{\|\tilde{\boldsymbol{\kappa}}_\alpha\|} = \frac{\{\boldsymbol{\kappa}^T \mathbf{A}(\alpha)\}\boldsymbol{g}^{(1)}}{\|\tilde{\boldsymbol{\kappa}}_\alpha\|} = \frac{\tilde{\boldsymbol{\kappa}}_\alpha^T \boldsymbol{g}^{(1)}}{\|\tilde{\boldsymbol{\kappa}}_\alpha\|}.$$

This the same ratio defined in (2.19), except that the smoothed vector $\tilde{\boldsymbol{\kappa}}_\alpha$ is used in place of $\boldsymbol{\kappa}$. Therefore the distribution of r_s is the same as that of r_r, namely that of a correlation coefficient with $\rho = 0$ and $p - 1$ degrees of freedom. As in the original case, we can define a t statistic,

$$t_s = (p-1)^{1/2} r_s (1 - r_s^2)^{-1/2}.$$

Under H_0, this statistic has a t distribution with $p - 1$ degrees of freedom, and should be large when H_1 is true.

For the purpose of comparing $\mathbf{G}_k$ to $\mathbf{F}_\beta$, we have not seen any advantage (e.g., in power to pick correct random effects) in using the naive smoothed eigenvector over the unsmoothed eigenvector, and have not pursued this approach (or the more sophisticated functional PCA approach) further.

This completes our discussion of the PCA approach to exploring the marginal covariance structure of the data. In the next section, we discuss how maximum likelihood factor analysis may be used to conduct a similar exploration of the marginal covariance structure.

2.4 Exploring marginal covariance with maximum likelihood factor analysis

Maximum likelihood factor analysis provides an alternative to the principal component approximation of the stage-1 model and the marginal covariance matrix. When all subjects have the same covariate values, first-order linearization of the NLMM gives the approximation

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i.$$

As outlined in Section 2.2.1, factor analysis with k common factors proposes the model

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_k \boldsymbol{u}_i + \boldsymbol{\varepsilon}_i \quad \text{for} \quad i = 1, \ldots, n$$

where the unobserved *common factors* $\boldsymbol{u}_i \overset{\text{i.i.d.}}{\sim} N_k(\mathbf{0}, \mathbf{I}_k)$ are independent of the *specific factors*, $\boldsymbol{\varepsilon}_i \overset{\text{i.i.d.}}{\sim} N_p(\mathbf{0}, \boldsymbol{\Psi})$. The columns of the $p \times k$ matrix $\boldsymbol{\Lambda}_k$ contain the common *factor loadings*. The k factor model gives the factorization of the approximate marginal covariance matrix,

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}$$

where $\boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda}_k = \boldsymbol{\Delta}$ is constrained to be diagonal.

PCA assumes that $\boldsymbol{\Psi} = \mathbb{V}(\boldsymbol{\varepsilon}_i) = \sigma^2 \mathbf{I}_p$; i.e., that within-subject variation is constant over time. However, in practice we may see heteroscedasticity in within-subject variation. Factor analysis provides a more flexible model, $\mathbb{V}(\boldsymbol{\varepsilon}_i) = \boldsymbol{\Psi}$ where $\boldsymbol{\Psi}$ is diagonal with diagonal elements $\psi_j > 0$ for $j = 1, \ldots, p$. The factor model assumes that the p specific factors (the within-subject

errors at each time point) are uncorrelated, but allows for differing magnitudes of variation across time points. Note however, that the factor model assumes that the variance at time t_j is the same across subjects, $\mathbb{V}(\varepsilon_{ij}) = \psi_j$, whereas the NLMM as defined in equation (1.6) of Chapter 1 allowed this variance to differ across subjects, $\mathbb{V}(\varepsilon_{ij}) = \psi_{ij}$. For now we assume that $\psi_{ij} \approx \psi_j$ for $i = 1, \ldots, n$.

There are several approaches to fitting the factor model; we choose maximum likelihood factor analysis (MLFA). Standard implementations of MLFA require the sample correlation matrix to be nonsingular. This requires more observations than time points, $n > p$. In the next section we show that maximum likelihood can also be used to fit the factor model when $n \leq p$. Details of the maximum likelihood estimation when $n \leq p$ are presented in Appendix C.

As in the PCA approach to exploring marginal covariance, one of the key questions is choosing the number of random effects required to characterize between-subject variation. In factor analysis, this corresponds to choosing the number of common factors k; we discuss this in Section 2.4.2. In Section 2.4.3 we propose two restricted models for the diagonal elements of $\mathbf{\Psi}$. Comparing the fits of the different models provides a strategy to detect and model within-subject heteroscedasticity. In Section 2.4.4 we end our discussion of the factor analysis approach by outlining a strategy for selecting random effects, or identifying an appropriate model parameterization. This strategy is very similar to the one used in the PCA approach, but is based on the estimated factor loadings $\widehat{\mathbf{\Lambda}}_k$ instead of on the principal component loadings $\mathbf{G}_k$.

2.4.1 Maximum likelihood factor analysis when $n \leq p$

The algorithm published by Jöreskog (1967) for maximum likelihood factor analysis is the basis of many commercial implementations; for example, `proc factor` in SAS or `factanal()` in R. This algorithm requires the sample covariance matrix $\mathbf{S}_n$ to be nonsingular. When $n \leq p$, $\mathbf{S}_n$ is singular, but the k-factor model can be fitted by maximum likelihood, with some restrictions on the number of factors, k.

For identifiability, we require k to be small enough to satisfy two constraints. Firstly, the usual constraint applies, namely that total number of parameters in the k factor representation of $\mathbf{\Sigma}$ must be smaller than the number of distinct elements in the unrestricted model for $\mathbf{\Sigma}$,

$$pk + p - \tfrac{1}{2}k(k-1) \leq \tfrac{1}{2}p(p+1).$$

Secondly, in cases where $n \leq p$, k must also be small enough so that the total number of parameters estimated (including the mean $\boldsymbol{\mu}$), is smaller than the total number of observations,

$$pk + 2p - \tfrac{1}{2}k(k-1) \leq np.$$

Our algorithm for maximum likelihood factor analysis minimizes the objective function

$$F_k(\mathbf{\Lambda}_k, \mathbf{\Psi}) = \log|\mathbf{\Sigma}| + \text{trace}\left(\mathbf{S}_n\mathbf{\Sigma}^{-1}\right) \tag{2.22}$$

where $\mathbf{\Sigma} = \mathbf{\Lambda}_k\mathbf{\Lambda}_k^T + \mathbf{\Psi}$, subject to the constraint that $\mathbf{\Lambda}_k^T\mathbf{\Psi}^{-1}\mathbf{\Lambda}_k = \mathbf{\Delta}$ is diagonal. This objective function is derived in Appendix C.1. The subscript k denotes that this is the objective function assuming k common factors. The objective function in (2.22) differs from that used in most implementations of maximum likelihood factor analysis,

$$F_k^*(\mathbf{\Lambda}_k, \mathbf{\Psi}) = \log|\mathbf{\Sigma}| + \text{trace}\left(\mathbf{S}_n\mathbf{\Sigma}^{-1}\right) - \log|\mathbf{S}_n| - p$$

which includes the additional terms $-\log|\mathbf{S}_n| - p$. When $n \leq p$, $|\mathbf{S}_n| = 0$ and the usual objective function is not defined. When $n > p$, the additional terms are free of the unknown parameters, so the objective function in (2.22) and the usual objective function F_k^* are minimized at the same value of $(\mathbf{\Lambda}_k, \mathbf{\Psi})$.

To minimize F_k, we follow the strategy of Jöreskog (1967) and Lawley and Maxwell (1971). First we find the conditional minimum of F_k for a given value of $\mathbf{\Psi}$, giving the function $f_k(\mathbf{\Psi}) =$

$F_k(\widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi}), \mathbf{\Psi})$ where $\widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi})$ is the value of $\mathbf{\Lambda}_k$ that minimizes F_k for a given $\mathbf{\Psi}$. Then we find $\mathbf{\Psi}$ (numerically) to minimize f_k. Appendix C.2 describes the details involved in minimizing F_k. We highlight the main results below. For given $\mathbf{\Psi}$, define

$$\mathbf{S}^* = \mathbf{\Psi}^{-1/2}\mathbf{S}_n\mathbf{\Psi}^{-1/2}.$$

Let $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_p$ be the eigenvalues of $\mathbf{S}^*$ and define the $k \times k$ diagonal matrix $\mathbf{\Theta} = \text{diag}\{\theta_1, \ldots, \theta_k\}$. Let $\mathbf{\Omega} = (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_k)$ be a $p \times k$ matrix whose columns are the normalized eigenvectors of $\mathbf{S}^*$ corresponding to the k largest eigenvalues. Then for fixed $\mathbf{\Psi}$, F_k is minimized, subject to the constraint that $\mathbf{\Lambda}_k^T\mathbf{\Psi}^{-1}\mathbf{\Lambda}_k = \mathbf{\Delta}$ is diagonal, at

$$\widetilde{\mathbf{\Lambda}}_k = \mathbf{\Psi}^{1/2}\mathbf{\Omega}(\mathbf{\Theta} - \mathbf{I}_k)^{1/2}. \tag{2.23}$$

Note that if any of the eigenvalues on the diagonal of $\mathbf{\Theta}$ are less than one then there will be no real solution for $\widetilde{\mathbf{\Lambda}}_k$. We will avoid this problem by choosing k small enough so that $\theta_1 \geq \theta_2 \ldots \geq \theta_k \geq 1$. This is justified in Appendix C.3 where we show that for fixed $\mathbf{\Psi}$, when $\theta_j < 1$, setting the j-th column of $\mathbf{\Lambda}_k$ to $\mathbf{0}$ minimizes the objective F_k.

Evaluating the objective function F_k at $\widetilde{\mathbf{\Lambda}}_k$ defined in (2.23) gives the objective function to be minimized in $\mathbf{\Psi}$,

$$f_k(\mathbf{\Psi}) = F_k(\widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi}), \mathbf{\Psi}) = k + \sum_{m=1}^{p} \log \psi_m + \sum_{m=1}^{k} \log \theta_m + \sum_{m=k+1}^{p} \theta_m. \tag{2.24}$$

When $n \leq p$, $\mathbf{S}_n$ and hence $\mathbf{S}^*$ will have rank $(n-1) < p$. In this case, $\theta_m = 0$ for $m = n, \ldots, p$, but the objective function f_k will still be defined as long as $k < n$.

When $k = 0$ (no common factors), the maximum likelihood estimator (MLE) of $\mathbf{\Psi}$ is just the diagonal of the sample covariance matrix $\mathbf{S}_n$, and $\widehat{\mathbf{\Lambda}}_0 = \mathbf{0}$. When $k \geq 1$, the MLE of $\mathbf{\Psi}$ is found by minimizing f_k, and then substituting this for $\mathbf{\Psi}$ in equation (2.23) to obtain the MLE of

the factor loadings, $\widehat{\mathbf{\Lambda}}_k$. Recall that θ_m is the m-th largest eigenvalue of $\mathbf{S}^* = \mathbf{\Psi}^{-1/2}\mathbf{S}_n\mathbf{\Psi}^{-1/2}$, so the eigenvalues are functions of $\mathbf{\Psi}$. The objective f_k in (2.24) is therefore a complicated function of $\mathbf{\Psi}$, but may be minimized numerically, for example using a Quasi-Newton line search method such as `fminunc()` in MATLAB. The gradient of f_k with respect to $\mathbf{\Psi}$ may be computed analytically using

$$\frac{\partial f_k}{\partial \psi_j} = \frac{1}{\psi_j}\left\{\sum_{m=1}^{k}(\theta_m - 1)\omega_{jm}^2 + 1 - s_{jj}/\psi_j\right\} \quad \text{for} \quad j = 1, \ldots, p$$

where ω_{jm} is the j-th element in the m-th eigenvector $\boldsymbol{\omega}_m$. Appendix C.4 describes further details about implementing the maximum likelihood factor analysis estimation.

In Appendix C.5 we examine briefly the behavior of the objective function f_k and the likelihood as $\mathbf{\Psi}$ gets small or as k increases. These cases are informative as they explore the boundaries of the parameter space. For example, as $\mathbf{\Psi}$ gets smaller, the likelihood approaches zero. However, for fixed $\mathbf{\Psi}$, the likelihood increases (to infinity) as the number of factors increases.

2.4.2 Selecting the number of common factors

The maximum likelihood estimation of the factor model described in the previous section assumes that the number of common factors k is known. However, a standard question in factor analysis is to determine the smallest number of common factors, k, that is necessary to provide an adequate approximation

$$\mathbf{\Sigma} = \mathbf{\Lambda}_k\mathbf{\Lambda}_k^T + \mathbf{\Psi}$$

to the marginal covariance matrix. As in the PCA approach of Section 2.3.1, different methods are needed in the cases where $n > p$ and where $n \leq p$. These are discussed below.

A note of caution is worth mentioning at this point. It may occur that no k factor model adequately describes the sample covariance matrix. For example, if the data are a nonlinear

function of unobserved latent variables, then the factor model (which is linear in the common factors) may not be able to explain the marginal covariance adequately. Since in our application, the latent variables arise due to random effects on parameters in a nonlinear model, we would expect the linear factor model to be somewhat inadequate when the magnitude of one or more random effects is large. This concern may be mediated somewhat by supposing that a second-order approximation is adequate in cases where a random effect is large. For example, a single random effect in the NLMM may be manifest as two linear latent variables, b_i and b_i^2, in the factor model approximation.

Case 1: $n > p$

When $n > p$, we may use the usual (Bartlett corrected) likelihood ratio test that was introduced in Section 2.2.1. This test compares the fit of the model with k common factors to that of an unrestricted model for $\boldsymbol{\Sigma}$. Let $\widehat{\boldsymbol{\Sigma}}_k = \widehat{\boldsymbol{\Lambda}}_k\widehat{\boldsymbol{\Lambda}}_k^T + \widehat{\boldsymbol{\Psi}}$ be the maximum likelihood estimator of $\boldsymbol{\Sigma}$ under the model with k common factors. The MLE under the unrestricted model is $\mathbf{S}_n$. The test statistic is

$$U_k = n'p(\widehat{a} - \log \widehat{g} - 1) \tag{2.25}$$

where $\widehat{a}$ and $\widehat{g}$ are the arithmetic and geometric means of the eigenvalues of $\widehat{\boldsymbol{\Sigma}}_k^{-1}\mathbf{S}_n$, and $n' = n - 1 - \frac{1}{6}(2p+5) - \frac{2}{3}k$. When the k-factor model is true, the approximate distribution of U_k is chi-squared with

$$s = \tfrac{1}{2}(p-k)^2 - \tfrac{1}{2}(p+k)$$

degrees of freedom. Recall that under the unrestricted model for $\boldsymbol{\Sigma}$ the maximum likelihood estimator of $\boldsymbol{\Sigma}$ is $\mathbf{S}_n$. Let $\widehat{F}_{\text{full}} = \log|\mathbf{S}_n| + p$ be the objective function in (2.22) evaluated at $\mathbf{S}_n$. Based on the minimized value $\widehat{F}_k$ achieved from fitting the k-factor model, the likelihood ratio test statistic defined in (2.25) may also be computed as

$$U_k = n'(\widehat{F}_k - \widehat{F}_{\text{full}}) = n'\widehat{F}_k^*.$$

This partly explains the choice of F_k^* as the usual objective function for maximum likelihood factor analysis. It is also worth noting that the minimized value $\widehat{F}_k$ equals the minimized value $\widehat{f}_k$ of the concentrated objective function defined in (2.24).

As pointed out in Section 2.2.1, it is often recommended that the test in equation (2.25) should be used with caution when $n < p + 50$. Less formal approaches based on the proportion of variation explained by the common factors may be used to complement the likelihood ratio test. These are discussed below.

Case 2: $n \leq p$

When $n \leq p$, the usual likelihood ratio test does not exist because $\mathbf{S}_n$ is singular and the likelihood evaluated at $\mathbf{S}_n$ is infinite. We might attempt to use a likelihood ratio test to compare the fit of the k-factor model to a model with more factors, but this approach is not tenable. Geweke and Singleton (1980) pointed out that when the k-factor model is true, any model with more than k factors is not identifiable. To see this suppose that $\mathbf{\Lambda}$ is a $p \times r$ matrix with rank $k < r$. Then there exists $\mathbf{Q}$, a $r \times (r-k)$ matrix for which $\mathbf{\Lambda Q} = \mathbf{0}$ and $\mathbf{Q}^T\mathbf{Q} = \mathbf{I}_{r-k}$. Let $\mathbf{M}$ be any $p \times (r-k)$ matrix with mutually orthogonal rows, where the rows have norms $\delta_1, \ldots, \delta_p$. Then $\mathbf{MM}^T = \text{diag}\left\{\delta_1^2, \ldots, \delta_p^2\right\}$. Since $\mathbf{\Lambda Q} = \mathbf{0}$ we can write

$$\begin{aligned}
\mathbf{\Lambda\Lambda}^T + \mathbf{\Psi} &= (\mathbf{\Lambda\Lambda}^T + \mathbf{MM}^T) + (\mathbf{\Psi} - \mathbf{MM}^T) \\
&= (\mathbf{\Lambda} + \mathbf{MQ}^T)(\mathbf{\Lambda} + \mathbf{MQ}^T)^T + \text{diag}\left\{\psi_1 - \delta_1^2, \ldots, \psi_p - \delta_p^2\right\} \\
&= \mathbf{\Lambda}^*\mathbf{\Lambda}^{*T} + \mathbf{\Psi}^*.
\end{aligned}$$

Thus when $\mathbf{\Lambda}$ has less than full column rank, the parameters of the factor model are not fully identified. Since no consistent estimator of $\mathbf{\Psi}$ will exist, standard theory about the asymptotic distribution of the likelihood ratio statistic does not apply. Thus we cannot use a likelihood ratio test to compare the k-factor model to a model with $k+1$ factors, or to any larger model.

Consequently, when $n \leq p$, we must rely on less formal approaches to judge adequacy of

the k-factor model. These methods are based on the proportion of variation explained by the common factors, and were outlined in Section 2.2.1. Let $\lambda_{j,m}$ be the (j,m)-th element in the maximum likelihood estimator $\widehat{\mathbf{\Lambda}}_k$. The first measure is the ratio of the j-th communality to the sample variance of the j-th variable

$$(\lambda_{j1}^2 + \lambda_{j2}^2 + \ldots + \lambda_{jk}^2)/s_{jj}^2 \tag{2.26}$$

This quantifies the proportion of variation in responses at the j-th time point that is explained by the model with k factors. The second measure

$$(\lambda_{1m}^2 + \lambda_{2m}^2 + \ldots + \lambda_{pm}^2)/\text{trace}(\mathbf{S}_n) \tag{2.27}$$

looks at how much of the total variation in the data is explained by the m-th common factor. The k-factor model is deemed sufficient (when compared with a larger model) when including additional common factors in the model does not substantially increase the amount of variation explained, as measured by (2.26) or by (2.27).

2.4.3 Identifying and modeling within-subject heteroscedasticity

The usual factor analysis model allows the *specific variance* (i.e., the within-subject variance) to be different at each time point,

$$\mathbb{V}(\boldsymbol{\varepsilon}_i) = \mathbf{\Psi} \text{ where } \mathbf{\Psi} \text{ is diagonal and } \psi_j > 0 \text{ for } j = 1, \ldots, p.$$

By fitting the factor model and estimating ψ_j, we may explore whether the assumption of the NLMM that $\mathbb{V}(\boldsymbol{\varepsilon}_i) = \sigma^2\mathbf{I}_p$ is reasonable, or whether the within-subject variation shows evidence of *heteroscedasticity*; i.e., does the magnitude of within-subject variance change across time.

Furthermore, by making some assumptions about the structure of within-subject variation, we may fit versions of the factor model that restrict the elements of $\mathbf{\Psi}$. For example, we might

assume constant variance across time,

$$\psi_j = \sigma^2 \text{ for } j = 1, \ldots, p \quad \text{or} \quad \mathbf{\Psi} = \sigma^2 \mathbf{I}_p. \tag{2.28}$$

A reasonable alternative often proposed in pharmacokinetic analysis is that the within-subject variance at time t_j is proportional to some power of the mean μ_j at that point,

$$\psi_j = \sigma^2 \mu_j^{2\gamma} \text{ for } j = 1, \ldots, p \quad \text{or} \quad \mathbf{\Psi} = \sigma^2 \text{diag}(\boldsymbol{\mu}^{2\gamma}) \tag{2.29}$$

where $\text{diag}(\boldsymbol{\mu}^{2\gamma})$ is a diagonal matrix with diagonal elements $\mu_1^{2\gamma}, \mu_2^{2\gamma}, \ldots, \mu_p^{2\gamma}$. In Appendix C.6 we give details of how these two restricted factor models may be fitted by maximum likelihood or pseudo-maximum likelihood respectively.

Note that the homoscedastic model (2.28) is nested within the power-of-the-mean model (2.29) which is nested within the unrestricted factor model. Therefore we may use a likelihood ratio test to examine the assumption of homoscedasticity. This test could compare the model with homogeneous variance against the power-of-the-mean model, or against the unrestricted factor model. Similarly, we may use a likelihood ratio test to determine whether the power-of-the-mean model is an adequate representation of within-subject variation, by comparing the fit of this model against the fit of the unrestricted model.

Let $\widehat{F}_k^{(h)}$ be the minimum value of the objective function for the factor model assuming homogeneous variance, $\psi_j = \sigma^2$ for $j = 1, \ldots, p$. Similarly, let $\widehat{F}_k^{(p)}$ and $\widehat{F}_k^{(u)}$ be the minimized values of the objective functions for the factor models with, respectively, the power-of-the-mean and unrestricted models for $\mathbf{\Psi}$. Since -2 times the loglikelihood equals nF_k, we define the

following test statistics.

$$\begin{aligned} T_{h,p} &= n(\widehat{F}_k^{(h)} - \widehat{F}_k^{(p)}) \\ T_{h,u} &= n(\widehat{F}_k^{(h)} - \widehat{F}_k^{(u)}) \\ T_{p,u} &= n(\widehat{F}_k^{(p)} - \widehat{F}_k^{(u)}) \end{aligned} \tag{2.30}$$

for comparing the homogeneous variance model to the power-of-the-mean model, the homogeneous model to the unrestricted model, and the power-of-the-mean model to the unrestricted model respectively. If $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$, then the asymptotic distribution of $T_{h,p}$ is chi-squared with 1 degree of freedom, and the asymptotic distribution of $T_{h,u}$ is chi-squared with $p-1$ degrees of freedom. If $\boldsymbol{\Psi} = \sigma^2 \boldsymbol{\mu}^{2\gamma}$ then the asymptotic distribution of $T_{p,u}$ is chi-squared with $p-2$ degrees of freedom.

2.4.4 Selection of random effects and an appropriate parameterization

Under the k-factor model

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\Lambda}_k \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_i \quad \text{and} \quad \boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}.$$

Having identified the number of common factors k (Section 2.4.2) and a model for the within-subject (or specific) variances in $\boldsymbol{\Psi}$ (Section 2.4.3), we can compute the maximum likelihood estimator of the factor loadings

$$\widehat{\boldsymbol{\Lambda}}_k = \widehat{\boldsymbol{\Psi}}^{1/2} \widehat{\boldsymbol{\Omega}} (\widehat{\boldsymbol{\Theta}} - \mathbf{I}_k)^{1/2}$$

by substituting the maximum likelihood estimator of $\boldsymbol{\Psi}$ into equation (2.23). The linearized model (2.1) with parameterizations $f(t, \boldsymbol{\beta}, \boldsymbol{\xi}; \boldsymbol{x})$ and $f^*(t, \boldsymbol{\delta}, \boldsymbol{\xi}; \boldsymbol{x})$ gives the approximations

$$\begin{aligned} \boldsymbol{y}_i &\approx \boldsymbol{\mu} + \mathbf{F}_\beta \boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \qquad \text{and} \qquad \mathbb{V}(\boldsymbol{y}_i) \approx \boldsymbol{\Sigma} = \mathbf{F}_\beta \mathbf{D} \mathbf{F}_\beta^T + \sigma^2 \mathbf{I}_p, \quad \text{or} \\ \boldsymbol{y}_i &\approx \boldsymbol{\mu} + \mathbf{F}_\delta^* \boldsymbol{d}_i + \boldsymbol{\varepsilon}_i \qquad \text{and} \qquad \mathbb{V}(\boldsymbol{y}_i) \approx \boldsymbol{\Sigma} = \mathbf{F}_\delta^* \mathbf{D}^* \mathbf{F}_\delta^{*T} + \sigma^2 \mathbf{I}_p. \end{aligned}$$

As in the PCA approach (Sections 2.3.2 and 2.3.3), we investigate the relationship between the columns of the estimated loadings matrix $\widehat{\boldsymbol{\Lambda}}_k$ and the columns of $\mathbf{F}_\beta$ or $\mathbf{F}_\delta^*$ using multivariate multiple regression (Section 2.2.5). We fit the multivariate regression model

$$\widehat{\boldsymbol{\Lambda}}_k = \mathbf{Z}\mathbf{B} + \mathbf{E} \tag{2.31}$$

where $\mathbf{Z}$ contains a subset of the columns of $\mathbf{F}_\beta$ or $\mathbf{F}_\delta^*$, and we assume the rows of the residual matrix $\mathbf{E}$ to be independent with zero mean and covariance matrix $\boldsymbol{\Gamma}$. To choose the most important random effects from amongst the columns of $\mathbf{F}_\beta$ or $\mathbf{F}_\delta^*$ we propose using forward selection. The forward selection procedure was outlined in Section 2.3.2, with the Wilks' lambda test statistic (2.12) used to test whether the m-th column of $\mathbf{F}_\beta$ should be included into the model that already contains $r-1$ other columns of $\mathbf{F}_\beta$.

In Section 2.3.2 we developed an exact test when $k = 1$ to compare the hypotheses

$$\begin{aligned} \mathrm{H}_0\text{:} \qquad & \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}, \quad \text{versus} \\ \mathrm{H}_1\text{:} \qquad & \boldsymbol{\Sigma} = \sigma_1^2 \boldsymbol{\kappa}\boldsymbol{\kappa}^T + \sigma^2 \mathbf{I}_p \quad \text{where} \quad \sigma_1^2 > 0 \end{aligned}$$

based on the first principal component loading $\boldsymbol{g}^{(1)}$. We do not have an equivalent of this test for the maximum likelihood factor estimator of $\boldsymbol{\Lambda}_1$. However, since at least in the case where $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$ we expect the principal component and factor loadings to be similar, we anticipate that the test statistic defined in equation (2.19) will be useful also for the factor

analysis approach. To apply this test to $\widehat{\mathbf{\Lambda}}_1$ define

$$r_r^M = \frac{\boldsymbol{\kappa}^T \widehat{\mathbf{\Lambda}}_1}{\|\boldsymbol{\kappa}\| \|\widehat{\mathbf{\Lambda}}_1\|} \quad \text{and} \quad t_r^M = (p-1)^{1/2} r_r (1 - r_r^2)^{-1/2}.$$

Then under H_0 we expect t_r^M to have approximately a t distribution with $p-1$ degrees of freedom.

To compare parameterizations f and f^* we follow the same strategy used with the PCA approach. The trace correlation r_T defined in (2.13) can be used to compare how well the multiple regressions on $\mathbf{F}_\beta$ or $\mathbf{F}_\delta^*$ explain variation in the columns of $\widehat{\mathbf{\Lambda}}_k$. The regression that explains a high proportion of variation in all columns of $\widehat{\mathbf{\Lambda}}_k$ based on a model that includes a small number of columns from $\mathbf{F}_\beta$ or from $\mathbf{F}_\delta^*$ will be associated with the more appropriate parameterization for the data set in question.

This concludes our development of methods to explore the marginal covariance structure of longitudinal data collected for the purpose of fitting of a nonlinear mixed effects model. In Chapter 3 we test performance of these methods in several simulation experiments. In Chapter 5 we use our proposed methods to analyze the covariance structure in the CCl_4 gas uptake experiment data, and the theophylline pharmacokinetic data introduced in Section 1.1.

Chapter 3

Simulation experiments: methods to explore the marginal covariance structure

3.1 Introduction

In this chapter we report results of simulation experiments testing the performance of the methods developed in Chapter 2. These methods explore the marginal covariance structure of repeated measures data that are collected for the purpose of fitting of a nonlinear mixed effects model (NLMM). The simulation experiments are based on four NLMM examples. We tested the ability of our methods to identify the correct number of random effects, to pick out the correct parameter sensitivities from among several candidates, and to distinguish between homoscedastic and heteroscedastic within-subject variance.

In the first example, the data were linear in the random effects. Since the principal component and factor analysis approaches assume linear common factors, we expect our methods to work well for the linear example.

In the next three examples we tested our methods on data generated with nonlinear random effects. In the nonlinear examples we also explored the ability to identify an appropriate model parameterization when there was a single random effect in one parameterization (the appropriate parameterization) that would require two random effects in a different parameterization. For nonlinear random effects the first-order linear approximation on which our methods are based may be inadequate when random effects are large. In this case a second-order approximation that includes quadratic functions of the random effects may be required.

For each example we generated data with either one or two normally distributed random

effects. In the case of two random effects, we assumed that the two random effects were uncorrelated within subjects. To simulate within-subject heteroscedasticity we used the power-of-the-mean model for residual variance at each time point

$$\mathbb{V}(\varepsilon_{ij}) = \sigma^2\{f(t_j, \boldsymbol{\beta}; \boldsymbol{x})\}^{2\gamma} \quad \text{for} \quad j = 1, \ldots, p$$

where the model f, that describes the expected trajectory for each subject, was evaluated at time t_j and at the population mean parameter $\boldsymbol{\beta}$. Note that the variance model was a function of the marginal mean, not on the conditional mean for each subject. We set σ to be small ($\sigma = 0.01$). In each simulation experiment we investigated the effects of several factors on performance of our methods:

- number of subjects (n),
- number of observations per subject (p),
- magnitude of the random effect variance, and
- amount of within-subject heteroscedasticity (γ).

The simulation experiments were run using MATLAB 7.0.0, R14 (The MathWorks, Inc.). The maximum likelihood fit of the factor model required an iterative optimization algorithm. We used MATLAB's implementation of an unconstrained optimizer, `fminunc`. This implements a Quasi-Newton line search algorithm with BFGS update of the Hessian. For the simulation experiments we set the termination tolerances at `TolX` $= 10^{-6}$ and `TolFun` $= 10^{-3}$. The algorithm terminates ("converges") if the infinity norm of the gradient of the objective function (relative to the value of this gradient at the starting value) is smaller than `TolFun`, or failing that, if the relative change in `x` (the argument of the optimization) is less than `TolX`. When analyzing results of the simulation experiments, we only used results from cases where the minimization algorithm converged.

In the next section we describe the data generating models for each example. In Section 3.3 we describe the endpoints collected from each simulation experiment and how simulation results

were analyzed. In Sections 3.4 to 3.7 we report the results for the four examples. We conclude the chapter with a summary of results in Section 3.8. Readers wishing to avoid the finer details of the simulation results are encouraged to skip directly to the summary section.

3.2 Data generating models

In this section we describe the models used to generate the simulation data for examples A, B, C, and D. In each example, the model f was used to generate the subject-specific trajectories at the first stage of the (nonlinear) mixed effects model,

$$y_{ij} = f(t_j, \boldsymbol{\beta}_i; \boldsymbol{x}) + \varepsilon_{ij} \quad \text{for} \quad j = 1, \ldots, p$$

where ε_{ij} were independent normal random variables with mean 0 and variance

$$\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \{f(t_j, \boldsymbol{\beta}; \boldsymbol{x})\}^{2\gamma}.$$

The vector $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{b}_i$ contains the model coefficients specific to the i-th subject. We assumed that the random effects $\boldsymbol{b}_i$ were independent normal random vectors with mean zero that were independent of ε_{ij}. For simplicity, we also assumed that the elements of $\boldsymbol{b}_i$ were uncorrelated within each subject. The vector $\boldsymbol{x}$ of covariates was held constant across subjects. Note that we have temporarily dropped from our notation the dependence of the model f on the known parameters $\boldsymbol{\xi}$.

In example A, the data generating model was based on a linearization of the open two-compartment model (O2C). Example B used the O2C model. This model was used to describe the trajectories for the CCl_4 gas uptake experiment described in Section 1.1. In example C we generated data using the open one-compartment (O1C) model assuming extravascular administration. This is the model used to describe the pharmacokinetics of theophylline in the second motivating example of Section 1.1. Example D is based on the three-parameter logistic

model, a common growth curve model.

In Section 3.2.5 we set out and compare relevant characteristics of the nonlinear models used in examples B, C, and D. These characteristics may help to explain differing performance of our methods in the different examples.

3.2.1 Example A: linear random effects

In this example data were generated with either one or two linear random effects. The data generating model was based on linearizing the open two-compartment (O2C) model that was described in Section 1.2.1. Under the O2C model with unit dose, the concentration of compound can be described by the function

$$f(t, \alpha, \omega^{(1)}, \omega^{(2)}) = \alpha \exp(-\omega^{(1)} t) + (1-\alpha)\exp(-\omega^{(2)} t), \quad t \in [0, 1]$$

and we chose parameter values

$$\alpha = 0.25, \quad \omega^{(1)} = 0.5, \quad \text{and} \quad \omega^{(2)} = 5$$

which produce a trajectory similar to that seen for the CCl_4 data. Assuming independent random effects on the large rate parameter $\omega_i^{(2)} = \omega^{(2)} + w_i$, and on the scale parameter $\alpha_i = \alpha + a_i$, first-order linearization gives

$$f(t, \alpha_i, \omega^{(1)}, \omega_i^{(2)}) \approx \mu(t_j) + w_i f_{\omega^{(2)}}(t_j) + a_i f_\alpha(t_j)$$

where

$$\begin{aligned}
\mu(t) &= f(t, \alpha, \omega^{(1)}, \omega^{(2)}) = \alpha \exp(-\omega^{(1)} t) + (1-\alpha)\exp(-\omega^{(2)} t) \\
f_{\omega^{(2)}}(t) &= \partial f(t)/\partial \omega^{(2)} = -(1-\alpha) t \exp(-\omega^{(2)} t), \quad \text{and} \\
f_\alpha(t) &= \partial f(t)/\partial \alpha = \exp(-\omega^{(1)} t) - \exp(-\omega^{(2)} t).
\end{aligned}$$

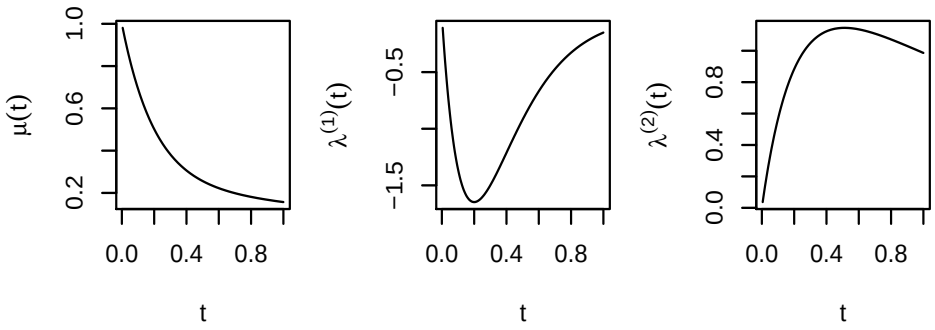

Figure 3.1: *Mean function (left panel) and factor loadings (center and right panels) for generating data with linear random effects in example A.*

The linearized model is linear in the random effects w_i and a_i.

The data generating model used for example A was

$$y_{ij} = \mu_j + \sigma_1 u_i^{(1)} \lambda_j^{(1)} + \sigma_2 u_i^{(2)} \lambda_j^{(2)} + \varepsilon_{ij} \tag{3.1}$$

where

$$\mu_j = \mu(t_j), \quad \lambda_j^{(1)} = f_{\omega^{(2)}}(t_j)/\|f_{\omega^{(2)}}\|, \quad \text{and} \quad \lambda_j^{(2)} = f_\alpha(t_j)/\|f_\alpha\|.$$

We scaled the loadings *functions* in the data generating model to have norm 1,

$$\|\lambda^{(m)}\| = \left\{ \int_0^1 \lambda^{(m)2}(t)\, \mathrm{d}t \right\}^{1/2} = 1.$$

Note that if $\boldsymbol{\lambda}^{(m)} = [\lambda^{(m)}(t_1), \ldots, \lambda^{(m)}(t_p)]^T$ is the m-th loadings *vector*, then under this scaling, $\boldsymbol{\lambda}^{(m)}$ will have norm approximately equal to $p^{1/2}$, or $\boldsymbol{\lambda}^{(m)T}\boldsymbol{\lambda}^{(m)} \approx p$. Figure 3.1 displays the mean and loadings functions.

The random effects $u_i^{(1)}$ and $u_i^{(2)}$ were generated as i.i.d. $N(0,1)$ random variables, independently of ε_{ij} which were independent $N(0, \sigma^2 \mu_j^{2\gamma})$. The parameters σ_1 and σ_2 control the amount of signal in the random effects or common factors. The parameter γ controls the amount of heteroscedasticity. To illustrate the amount of between-subject variation expressed

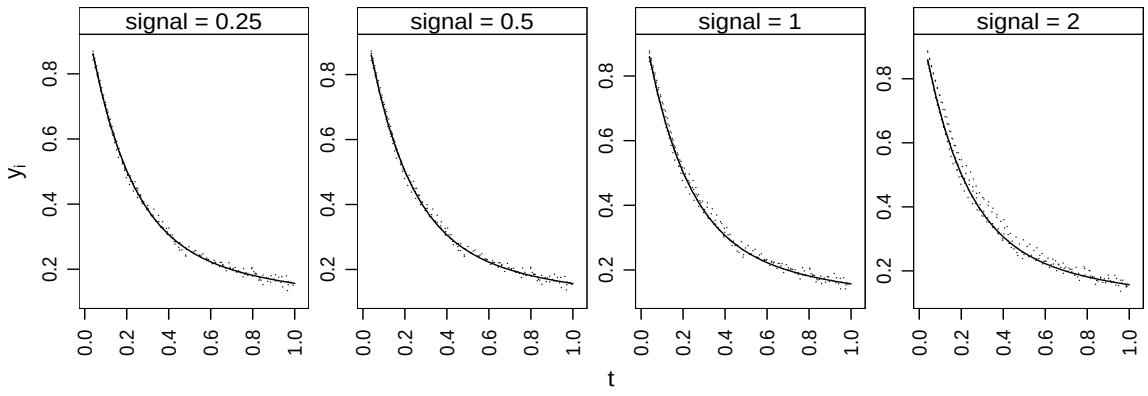

Figure 3.2: *Mean (solid curve) and subject trajectories (dotted curves) generated for example A with one random effect on $\lambda^{(1)}$. From left to right, the plots show increasing magnitude of random effect signal = σ_1/σ.*

in example A, Figure 3.2 illustrates 5 trajectories generated with one random effect $u_i^{(1)}$ with random effect signal $\sigma_1/\sigma = 0.25, 0.5, 1$, or 2. For the second random effect $\sigma_2 = 0$.

We ran two simulation experiments for example A. In the first, there was a single random effect $u_i^{(1)}$ with loadings function $\lambda^{(1)}$. In the second experiment we simulated two independent random effects: $u_i^{(1)}$ with $\sigma_1/\sigma = 2$, and a second random effect $u_i^{(2)}$ of varying magnitude.

Table 3.1: *Simulation treatment factors for example A. The first column gives the number of random effects, k.*

k	Random effects	Factor	Settings
$k = 1$	$u_i^{(1)}$	number of subjects	$n = 10, 25, 50$
		number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal for $u_i^{(1)}$	$\sigma_1/\sigma = 0, 0.25, 0.5, 1, 2$
$k = 2$	$u_i^{(1)}, u_i^{(2)}$	number of subjects	$n = 10, 25, 50$
		number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal for $u_i^{(1)}$	$\sigma_1/\sigma = 2$
		signal for $u_i^{(2)}$	$\sigma_2/\sigma = 0, 0.25, 0.5, 1, 2$

Table 3.1 gives the levels of each of the factors studied in the simulation experiments for example A. Each simulation experiment had a $3 \times 3 \times 3 \times 5$ factorial treatment design and we generated 1600 Monte Carlo replicates of each treatment combination.

3.2.2 Example B: O2C model with nonlinear random effects

In example B we generated data using the open two-compartment (O2C) model described in Section 1.2.1 that was used to model the chamber concentration of CCl_4. We assumed a dose D of 1, and used two parameterizations of the function describing concentration time course of compound in the chamber,

$$f(t, \alpha, \omega^{(1)}, \omega^{(2)}) = \alpha \exp(-\omega^{(1)} t) + (1-\alpha) \exp(-\omega^{(2)} t) \tag{3.2}$$

$$f^*(t, \alpha, \delta^{(1)}, \delta^{(2)}) = (1-\alpha) \exp(-\delta^{(2)} t) \left\{ \frac{\alpha}{1-\alpha} \exp(-\delta^{(1)} t) + 1 \right\} \tag{3.3}$$

defined for $t \in [0, 1]$. In the first parameterization we used parameter values

$$\alpha = 0.25, \quad \omega^{(1)} = 0.5, \quad \text{and} \quad \omega^{(2)} = 5$$

which produce a trajectory similar to that seen for the CCl_4 data. In the second parameterization

$$\alpha = 0.25, \quad \delta^{(1)} = \omega^{(1)} - \omega^{(2)} = -4.5, \quad \text{and} \quad \delta^{(2)} = \omega^{(2)} = 5.$$

In the first experiment with one random effect, we generated data using the parameterization f^* in equation (3.3) with a random effect on the rate parameter $\delta_i^{(2)} = \delta^{(2)} + d_i$. The data generating model was

$$y_{ij} = f^*(t, \alpha, \delta^{(1)}, \delta^{(2)} + d_i) + \varepsilon_{ij}$$

where ε_{ij} were independent $N(0, \sigma^2 \{f^*(t_j)\}^{2\gamma})$ random variables, and the random effects d_i were independent normal random variables with mean 0 and standard deviation $\upsilon_\delta |\delta^{(2)}|$. Note that

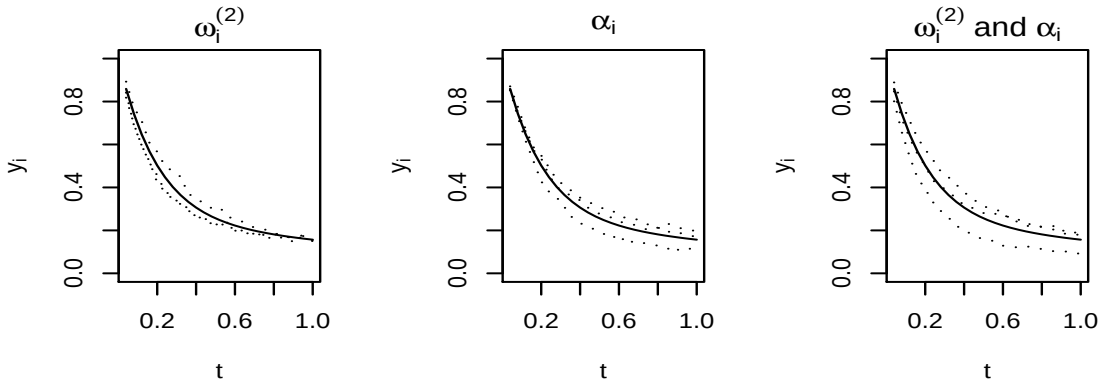

Figure 3.3: *Mean (solid curve) and random subject trajectories (dotted curves) described by parameterization* (3.2) *of the O2C model of example B with random effects on* $\omega^{(2)}$ *(left panel), on* α *(center panel), and on* $\omega^{(2)}$ *and* α *(right panel). In the figure,* $\upsilon_\omega = \upsilon_\alpha = 0.5$ *and* $\sigma = 0.01$.

the random coefficient $\delta_i^{(2)}$ induces two correlated random effects in the first parameterization f, one on each of the rate coefficients.

In the second experiment with two random effects, we generated data using the parameterization f in equation (3.2). We put random effects on the scale parameter, $\alpha_i = \alpha + a_i$, and on the large rate parameter, $\omega_i^{(2)} = \omega^{(2)} + w_i$. The data generating model was

$$y_{ij} = f(t, \alpha + a_i, \omega^{(1)}, \omega^{(2)} + w_i) + \varepsilon_{ij}$$

where ε_{ij} are independent $N(0, \sigma^2\{f(t_j)\}^{2\gamma})$ random variables, and the random effects a_i and w_i are independent normal random variables with mean 0 and standard deviations $\upsilon_\alpha|\alpha|$ and $\upsilon_\omega|\omega^{(2)}|$ respectively. We set the random effect a_i to be large ($\upsilon_\alpha = 0.4$) and varied the magnitude of the second random effect w_i. Note that the function f is a *linear* function of the rate parameter α; therefore the data generating model is linear in the random effect a_i. Figure 3.3 illustrates three trajectories generated according to parameterization (3.2) with random effects on $\omega^{(2)}$, α, and both $\omega^{(2)}$ and α.

In this example, υ_α, and υ_ω control the magnitude of the random effect "signal". However, that amount of signal is not necessarily equal to that in the linear random effect simulation of

Table 3.2: *Simulation treatment factors for example B. The first column gives the number of random effects, k.*

k	Random effects	Factor	Settings
$k = 1$	$\delta^{(2)} + d_i$	number of subjects	$n = 10, 25, 50$
		number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal (d_i)	$\upsilon_\delta = 0, 0.0125, 0.025, 0.05, 0.1$
$k = 2$	$\alpha + a_i$	number of subjects	$n = 10, 25, 50$
	$\omega^{(2)} + w_i$	number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal (a_i)	$\upsilon_\alpha = 0.4$
		signal (w_i)	$\upsilon_\omega = 0, 0.05, 0.1, 0.2, 0.4$

example A which was based on linearizing the O2C model. In example A, the signal for the common factor corresponding to the scale parameter was $\sigma_2 = \upsilon_\alpha |\alpha| \|\partial f/\partial \alpha\|$. Similarly, the signal for the common factor defined by the large rate parameter was $\sigma_1 \approx \upsilon_\omega |\omega^{(2)}| \|\partial f/\partial \omega^{(2)}\|$. In the case of a random effect on α, the relationship is exact because the function f is linear in α. The amount of signal that υ_α and υ_ω contribute to the marginal covariance depends on the magnitudes of α and $\omega^{(2)}$ as well as on the norms of their sensitivity functions.

Table 3.2 gives the levels of each of the factors studied in the two simulation experiments of example B. We had to keep υ_δ small in the first simulation based on parameterization f^* to ensure that the value of $\delta^{(1)} + \delta_i^{(2)}$ (the small rate parameter in the parameterization f) remained positive. Each simulation experiment had a $3 \times 3 \times 3 \times 5$ factorial treatment design and we generated 1600 Monte Carlo replicates of each treatment combination.

3.2.3 Example C: O1C model with nonlinear random effects

In example C we generated data using the open one-compartment (O1C) model, described previously in Section 1.2.1, and used to model the pharmacokinetics of theophylline. We assumed that 100% of the drug was absorbed (`f`=1), and used two parameterizations of the function

describing the concentration time course of compound in the body,

$$f(t, k_a, V_d, Cl; D) = \frac{Dk_a}{V_d(k_a - Cl/V_d)} \left\{ \exp\left(-\frac{Cl}{V_d} t^*\right) - \exp(-k_a t^*) \right\} \quad (3.4)$$

$$f^*(t, \delta^{(1)}, \delta^{(2)}, \delta^{(3)}; D) = D\delta^{(1)} \left\{ \exp(-\delta^{(2)} t^*) - \exp(-\delta^{(3)} t^*) \right\} \quad (3.5)$$

where $t^* = 25t$ for $t \in [0, 1]$. In the first parameterization the elimination rate has been parameterized in terms of clearance rate and the volume of distribution. We assumed that the covariate dose was $D = 5\,\text{mg/kg}$ and used the following values of model coefficients in the first parameterization f,

$$k_a = 1.5\,/\text{hr}, \quad Cl = 0.05\,\text{L/hr/kg}, \quad \text{and} \quad V_d = 0.5\,\text{L/kg}$$

which are similar to the mean coefficients estimated by Davidian and Giltinan (1995, Table 5.3) In the second parameterization f^* the corresponding coefficients were

$$\delta^{(1)} = k_a/\{V_d(k_a - Cl/V_d)\} = 2.14\,\text{kg/L}, \quad \delta^{(2)} = Cl/V_d = 0.1\,/\text{hr}, \quad \text{and} \quad \delta^{(3)} = k_a = 1.5\,/\text{hr}.$$

In the first experiment with one random effect, we generated data using the parameterization f in equation (3.4) with a random effect on the volume of distribution, $V_{di} = V_d + v_i$. The data generating model was

$$y_{ij} = f(t, k_a, V_d + v_i, Cl; D) + \varepsilon_{ij}$$

where the errors ε_{ij} were independent $N(0, \sigma^2\{f(t_j)\}^{2\gamma})$ random variables, and the random effects v_i were independent normal random variables with mean 0 and standard deviation $\upsilon_V|V_d|$. Assuming that the volume of distribution varied across subjects induced two correlated random effects in the second parameterization f^* of equation (3.5), one on the scale $\delta^{(1)}$ and one on the elimination rate $\delta^{(2)}$.

In the second experiment with two random effects, we generated data using the parameter-

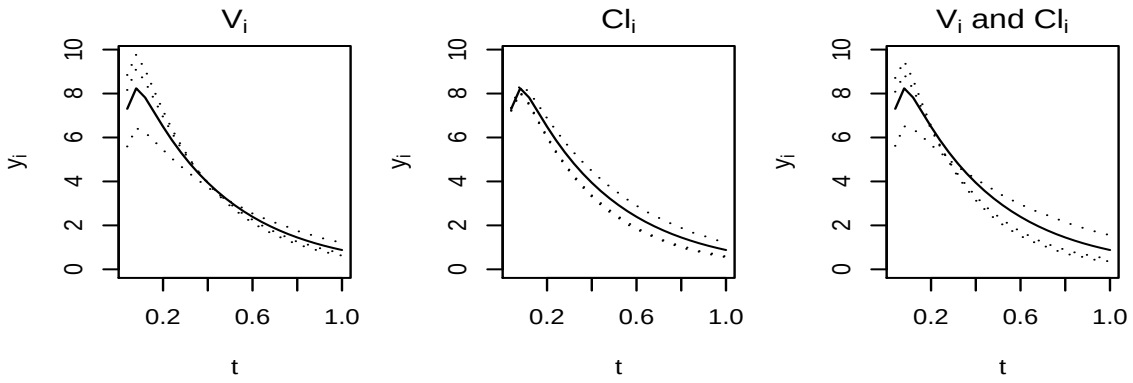

Figure 3.4: *Mean (solid curve) and random subject trajectories (dotted curves) described by parameterization* (3.4) *of the O1C model of example C with random effects on V_d (left panel), on Cl (center panel), and on V_d and Cl (right panel). In the figure, $\upsilon_V = \upsilon_C = 0.25$ and $\sigma = 0.01$.*

ization f in equation (3.4). We put random effects on the volume of distribution, $V_{di} = V_d + v_i$, and on the clearance, $Cl_i = Cl + c_i$. The data generating model was

$$y_{ij} = f(t, k_a, V_d + v_i, Cl + c_i; D) + \varepsilon_{ij}$$

where the errors ε_{ij} were independent $N(0, \sigma^2\{f(t_j)\}^{2\gamma})$ random variables, and the random effects v_i and c_i were independent normal random variables with mean 0 and standard deviations $\upsilon_V|V_d|$ and $\upsilon_C|Cl|$ respectively. We assumed that the random effect v_i was small ($\upsilon_V = 0.025$) and varied the magnitude of the second random effect c_i. Figure 3.4 illustrates three trajectories generated using parameterization (3.4) with random effects on V_d, Cl, and both V_d and Cl.

We initially ran a version of the experiment with two random effects where we set the first random effect on V_d to be large ($|\upsilon_V| = 0.2$). Unfortunately we did not take account of the possibility that the large nonlinear random effect would be detected as two or more linear common factors by our method. Consequently, we could not get a meaningful assessment of the endpoints of interest in this experiment. This experiment does emphasize that the user of our methods should be mindful of the presence of large random effects, and we discuss its results

Table 3.3: *Simulation treatment factors for example C. The first column gives the number of random effects, k.*

k	Random effects	Factor	Settings
$k = 1$	$V_d + v_i$	number of subjects	$n = 10, 25, 50$
		number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal (v_i)	$\upsilon_V = 0, 0.025, 0.05, 0.1, 0.2$
$k = 2$	$V_d + v_i$	number of subjects	$n = 10, 25, 50$
	$Cl + c_i$	number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal (v_i)	$\upsilon_V = 0.025$
		signal (c_i)	$\upsilon_C = 0, 0.025, 0.05, 0.1, 0.2$

briefly at the end of Section 3.6.

Table 3.3 gives the levels of each of the factors studied in the simulation experiments for example C. In this example, υ_V and υ_C control the magnitude of the random effect "signal". Each simulation experiment had a $3 \times 3 \times 3 \times 5$ factorial treatment design and we generated 1600 Monte Carlo replicates of each treatment combination.

3.2.4 Example D: Logistic model with nonlinear random effects

In example D we generated data using the three parameter logistic model that is commonly used to describe growth of crops. For example, Davidian and Giltinan (1995, Section 1.1.3) used this model to describe the average leaf weight of plots of soybean plants sampled over the growing season. In example D we used two parameterizations of the logistic model describing yield as a function of time,

$$f(t, \alpha, \omega^{(1)}, \omega^{(2)}) = \frac{\alpha}{1 + \exp\left\{-\omega^{(2)}(t - \omega^{(1)})\right\}}, \quad \text{and} \tag{3.6}$$

$$f^*(t, \alpha, \delta^{(1)}, \delta^{(2)}) = \frac{\alpha}{1 + \delta^{(1)} \exp\left(-\delta^{(2)} t\right)} \tag{3.7}$$

for $t \in [0, 1]$. In both parameterizations the coefficient α is the asymptotic limit of yield; as time increases, yield approaches the limit α. The parameter $\omega^{(2)}$ in parameterization f, and the parameter $\delta^{(2)}$ in parameterization f^*, quantify growth rate. Coefficient $\omega^{(1)}$ in parameterization f represents the time at which 50% of the asymptotic limit of yield is achieved; it can be thought of as the half-life of the growth curve. In the first parameterization f we set

$$\alpha = 2, \quad \omega^{(1)} = 0.5, \quad \text{and} \quad \omega^{(2)} = 10$$

for an asymptotic yield of 2 units, 50% of which is achieved at 0.5 hr, and a growth rate of 10 /hr. In the second parameterization f^* the corresponding coefficients were

$$\alpha = 2, \quad \delta^{(1)} = \mathrm{e}^{\omega^{(1)}\omega^{(2)}} = 148.4, \quad \text{and} \quad \delta^{(2)} = \omega^{(2)} = 10.$$

In the first experiment with one random effect, we generated data using the parameterization f in equation (3.6) with a random effect on the growth rate parameter, $\omega_i^{(2)} = \omega^{(2)} + w_i^{(2)}$. The data generating model was

$$y_{ij} = f(t, \alpha, \omega^{(1)}, \omega^{(2)} + w_i^{(2)}) + \varepsilon_{ij}$$

where the errors ε_{ij} were independent $N(0, \sigma^2\{f(t_j)\}^{2\gamma})$ random variables, and the random effects $w_i^{(2)}$ were independent normal random variables with mean 0 and standard deviation $\upsilon_2|\omega^{(2)}|$. Assuming that the growth rate $\omega^{(2)}$ in parameterization f varies across subjects induced two correlated random effects in the second parameterization f^* of equation (3.7), one on the growth rate $\delta^{(2)}$, and a second on the coefficient $\delta^{(1)}$.

In the second experiment with two random effects, we generated data using the parameterization f in equation (3.6). We put random effects on the the growth rate, $\omega_i^{(2)} = \omega^{(2)} + w_i^{(2)}$,

and on the half-life parameter, $\omega_i^{(1)} = \omega^{(1)} + w_i^{(1)}$. The data generating model was

$$y_{ij} = f(t, \alpha, \omega^{(1)} + w_i^{(1)}, \omega^{(2)} + w_i^{(2)}) + \varepsilon_{ij}$$

where the errors ε_{ij} were independent $N(0, \sigma^2\{f(t_j)\}^{2\gamma})$ random variables, and the random effects $w_i^{(1)}$ and $w_i^{(2)}$ were independent normal random variables with mean 0 and standard deviations $\upsilon_1|\omega^{(1)}|$ and $\upsilon_2|\omega^{(2)}|$ respectively. In the experiment with two random effects we assumed that the random effect $w_i^{(2)}$ on the growth rate was small ($\upsilon_2 = 0.05$), and we varied the magnitude of the random effect $w_i^{(1)}$ on the half-life parameter. Figure 3.5 illustrates three trajectories generated using parameterization (3.6) with random effects on $\omega^{(1)}$, $\omega^{(2)}$, and both $\omega^{(1)}$ and $\omega^{(2)}$.

We initially ran a version of the experiment with two random effects where we set the first random effect on $\omega^{(2)}$ to be large ($|\upsilon_2| = 0.4$). As with a similar experiment for example C we did not take account of the possibility that the large nonlinear random effect would be detected as two or more linear common factors by our method. Consequently, we could not get a meaningful assessment of the endpoints of interest in this experiment. This experiment again

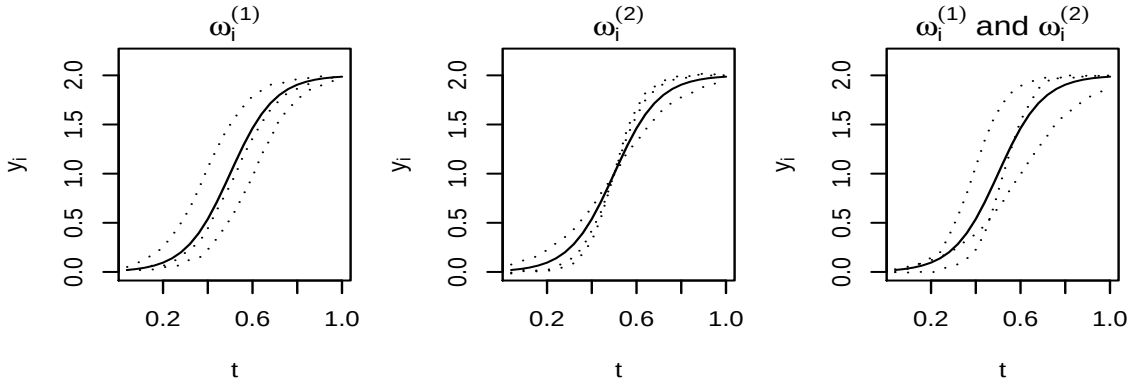

Figure 3.5: *Mean (solid curve) and random subject trajectories (dotted curves) described by parameterization* (3.6) *of the logistic model of example D with random effects on* $\omega^{(1)}$ *(left panel), on* $\omega^{(2)}$ *(center panel), and on* $\omega^{(1)}$ *and* $\omega^{(2)}$ *(right panel). In the figure,* $\upsilon_1 = 0.25$*,* $\upsilon_2 = 0.5$*, and* $\sigma = 0.01$*.*

Table 3.4: *Simulation treatment factors for example D. The first column gives the number of random effects, k.*

k	Random effects	Factor	Settings
$k = 1$	$\omega^{(2)} + w_i^{(2)}$	number of subjects	$n = 10, 25, 50$
		number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal ($w_i^{(2)}$)	$\upsilon_2 = 0, 0.05, 0.1, 0.2, 0.4$
$k = 2$	$\omega^{(2)} + w_i^{(2)}$	number of subjects	$n = 10, 25, 50$
	$\omega^{(1)} + w_i^{(1)}$	number of time points	$p = 10, 25, 50$
		heteroscedasticity	$\gamma = 0, 0.5, 1$
		signal ($w_i^{(2)}$)	$\upsilon_2 = 0.05$
		signal ($w_i^{(1)}$)	$\upsilon_1 = 0, 0.05, 0.1, 0.2, 0.4$

emphasizes that the user of our methods should be mindful of the presence of large random effects, and we briefly discuss its results at the end of Section 3.7.

Table 3.4 gives the levels of each of the factors studied in the simulation experiments for example D. In this example, υ_1 and υ_2 control the magnitude of the random effect "signal". Each simulation experiment had a $3 \times 3 \times 3 \times 5$ factorial treatment design and we generated 1600 Monte Carlo replicates of each treatment combination.

3.2.5 Comparison of the nonlinear models used in examples B, C and D

In examples B, C, and D with nonlinear random effects, we set the magnitude of the random effect standard deviation equal to a fraction of the mean parameter value. For example, if $b_i^{(m)}$ was the random effect on a parameter $\beta^{(m)}$, then we set the standard deviation of $b_i^{(m)}$ equal to $\upsilon_m|\beta^{(m)}|$. Suppose that $b_i^{(m)}$ is the only random effect, then a first-order linear approximation

of the first stage of the NLMM gives

$$\begin{aligned} y_{ij} &\approx f(t_j, \boldsymbol{\beta}) + b_i^{(m)} \partial f(t_j)/\partial \beta^{(m)} + \varepsilon_{ij} \\ &= \mu_j + \left(\upsilon_m |\beta^{(m)}| \|\partial f/\partial \beta^{(m)}\| \right) u_i^{(m)} \lambda_j^{(m)} + \varepsilon_{ij} \\ &= \mu_j + \sigma_m u_i^{(m)} \lambda_j^{(m)} + \varepsilon_{ij} \end{aligned} \tag{3.8}$$

where the common factor scores $u_i^{(m)}$ are independent $N(0,1)$ random variables and

$$\lambda_j^{(m)} = \partial f(t_j)/\partial \beta^{(m)} / \|\partial f/\partial \beta^{(m)}\|$$

is the j-th element in a vector with norm $p^{1/2}$. In the last line of equation (3.8) we have written the model in the form of a linear common factor model like the data generating model (3.1) where the common factor $u_i^{(m)}$ had unit variance. Thus the "signal" of the linear common factor associated with the random effect $b_i^{(m)}$ is approximately

$$\sigma_m \approx \upsilon_m |\beta^{(m)}| \|\partial f/\partial \beta^{(m)}\|.$$

Thus we might expect our methods, which assume a linear common factor model, to perform differently for the different nonlinear examples because the common factor "signal" is stronger in some examples than in other examples, due to the magnitude of the mean parameter $|\beta^{(m)}|$, or the norm of the sensitivity function $\|\partial f/\partial \beta^{(m)}\|$. Table 3.5 gives the numerical values of $|\beta^{(m)}| \|\partial f/\partial \beta^{(m)}\|$ for different random effect parameters for the nonlinear models used in examples B, C and D. Recall that for example B with the O2C model, we generated data for the experiment with one random effect using parameterization f^*, and using parameterization f when there were two random effects. Hence there are two sets of entries in Table 3.5 for this model.

To aid in interpretation of results, the values in the column labeled $|\beta^{(m)}| \|\partial f/\partial \beta^{(m)}\|$ may be used to multiply the values of υ_m used in each simulation experiment. For example, for a

Table 3.5: *Properties of the random effect parameters and sensitivities for the nonlinear models of examples B, C and D. The first column gives the model with parameterization in parenthesis.*

				Corr. between sensitivities	
Model	$\beta^{(m)}$	type	$\lvert\beta^{(m)}\rvert\lVert\partial f/\partial\beta^{(m)}\rVert$	centered	uncentered
O2C (f^*)	$\delta^{(2)}$	nonlinear	0.619		
O2C (f)	α	linear	0.167	0.130	−0.816
	$\omega^{(2)}$	nonlinear	0.152		
O1C (f)	V	nonlinear	2.860	0.276	−0.225
	Cl	nonlinear	3.421		
Logistic (f)	$\omega^{(1)}$	nonlinear	1.291	1.25e−3	−5.86e−5
	$\omega^{(2)}$	nonlinear	0.292		

given value of υ_m we expect the signal of the random effects in the O1C model of example C to be quite a bit larger than for random effects in the other models. The reader should also bear in mind that different levels of υ_m were used in the different nonlinear examples. For example, the maximum value of υ_m used in example C was half that of the maximum υ_m used in experiment D. Comparison of results across examples promises to be complicated.

The column "type" in Table 3.5 denotes whether a particular parameter enters the model linearly or nonlinearly. The O2C model is linear in the random effect on α. In all other cases, the models are nonlinear in the random effects that we chose to study. Because our methods are based on a linear approximation, we will pay particular attention to what happens to the performance of the methods when the random effect standard deviation becomes large. In such cases the linear approximation will be poor, and an adequate approximation may require terms that are quadratic in the random effects.

The last two columns of Table 3.5 give the centered and uncentered correlations between the sensitivities for the two random effect parameters considered in each example. Because the principal component analysis (and the factor analysis for the case of homogeneous variance) compute orthogonal factor loadings, we expect results to be easier to interpret when the sensi-

tivities of random effects used to generate data are approximately orthogonal. The uncentered correlation is of primary interest because this measures departure from orthogonality. The sensitivities for the two random effects of the logistic model used to generate data in example D were approximately orthogonal and uncorrelated. This should make it easier to interpret results for this example than in other examples where the sensitivities were more oblique or correlated.

3.3 Experiment endpoints and analysis of simulation results

To assess performance of our methods for exploring the marginal covariance structure we looked at the following measures.

Number of random effects

First we are interested in checking whether our methods were able to detect the number of random effects necessary to characterize between-subject variability. In each experiment with $k = 1$ or 2 random effects, we used the likelihood ratio tests defined in Section 2.3.1 (equation (2.16) when $n > p$ and (2.17) when $n \leq p$) for testing whether the principal component model with $k - 1$ principal components was adequate. When $n > p$, we fitted a factor model with $k - 1$ common factors and unrestricted $\mathbf{\Psi}$, and used the likelihood ratio test defined in Section 2.4.2 (equation (2.25)) to test adequacy of the approximation with $k - 1$ common factors. We estimated the power of each test to reject (at the 5% significance level) the hypothesis that a model with $k - 1$ principal components or common factors was adequate.

As mentioned in previous sections, our methods are based on an approximation of the stage-1 model that is linear in the random effects. When the stage-1 model is nonlinear in a random effect with large variance, we expect that an adequate approximation will involve quadratic functions of the large random effect. In this case, an adequate principal component or factor model will require two or more linear principal components or common factors to capture

the effect of a single large nonlinear random effect. However, in our simulation experiments we fixed the number of principal components or common factors to equal k, the number of random effects. In future work, we may allow the methods to determine an adequate number of components or common factors on the fly, based on the analysis of each simulated data set.

Test for homogeneous variance

Let MLFA-H denote the k-factor model with homogeneous within-subject variance, $\mathbf{\Psi} = \sigma^2\mathbf{I}_p$; let MLFA-P be the k-factor model with $\psi_j = \sigma^2\mu_j^{2\gamma}$ (power model); and let MLFA-U be the unrestricted k-factor model. In Section 2.4.3 we proposed likelihood ratio tests (equation (2.30)) for comparing the fit of the homogeneous k-factor model MLFA-H, with the fit of model MLFA-P or model MLFA-U. The simulated data had $\psi_j = \sigma^2\mu_i^{2\gamma}$ where $\gamma = 0$, 0.5 or 1. In each experiment we fitted the three models and computed the two likelihood ratio tests. For each test we estimated power of the tests to reject (at the 5% significance level) the hypothesis that $\mathbf{\Psi} = \sigma^2\mathbf{I}_p$.

Based on the fit of the k-factor MLFA-P model, we estimated the parameters γ and σ^2. Examining the bias in estimates of these parameters is another way to track how well the common factor model fitted the simulated data. For example, if too few common factors were included, then we would expect the estimate of σ^2 to be inflated to capture the unexplained variation in the data.

Selecting random effects

In Sections 2.3.2 and 2.4.4 we discussed our approach to identifying which parameters in the stage-1 model require non-trivial random effects. When $k = 1$, we used the exact test (equation (2.19)) to assess the strength of the relationship between the first principal component loading and the sensitivity corresponding to the true random effect parameter. We estimated the power of the exact test to reject (at the 5% significance level) the hypothesis that $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$. Although the exact test was derived based on the first principal component loading, in our sim-

ulation experiments we also computed the test statistic using the maximum likelihood estimate of the factor loading vector for a the one-factor MLFA-U and MLFA-P models.

Let $\mathbf{F}_\beta$ be the matrix with columns defined by the sensitivity functions of parameters that are candidates for random effects in the stage-1 model. The matrix may also include second-order sensitivities corresponding to a second-order approximation of the stage-1 model involving quadratic functions of nonlinear random effects with large variance. For $k = 1$ and $k = 2$, we fitted a principal component model with k components and the k-factor MLFA-U and MLFA-P models. Then we used the multivariate forward selection algorithm described in Section 2.3.2 to select columns from $\mathbf{F}_\beta$. For each simulation replicate we recorded which sensitivities were included in the selected model. We estimated the mean number of sensitivities selected, the proportion of times particular sensitivities were included, and the trace correlation r_T (equation (2.13)) of the selected model.

Comparing two parameterizations

In the nonlinear examples (B, C, and D), we generated data with a single random effect in one parameterization, which induced two correlated random effects in the alternative parameterization. Let the sensitivities for the alternative parameterization be the columns of a matrix $\mathbf{F}_\delta$. Based on the loadings for the principal component and factor models with k common factors, we used the multivariate forward selection procedure to select columns from $\mathbf{F}_\delta$. We calculated the mean number of sensitivities selected, and the trace correlation r_T for the multivariate regression based on the alternative parameterization. We would expect the mean number of sensitivities selected from $\mathbf{F}_\delta$ (the alternative parameterization) to be larger than the mean number selected from $\mathbf{F}_\beta$ (the actual parameterization), or for the trace correlation of the model selected from $\mathbf{F}_\delta$ to be smaller than that of the model selected from $\mathbf{F}_\beta$.

Analysis of the effect of simulation factors

We used analysis of variance to assess how the different simulation factors (n, p, γ, and random effect signal) affected each of the endpoints described above. Since in many cases the endpoints could not be expected to be normally distributed, we batched the MC replicates into groups of size 10, and used analysis of variance (ANOVA) based on batch means. The simulation experiments were well replicated (1600 MC reps each, or 160 batches), and it turned out that for many endpoints, all four factors as well as most of their interactions were statistically significant (e.g., at the 0.0001 level). In the following sections we have identified the factors that accounted for a large percentage of the variation in the simulation endpoints (measured by the R^2 statistic of the batched ANOVA), and have ignored other factors and higher order interactions.

When selecting random effects, results based on the principal components analysis were generally indistinguishable from results based on the k-factor MLFA-U or MLFA-P models, even when within-subject variance was heteroscedastic. For the sake of brevity, we have presented the exact test results only for the PCA analysis, and the multivariate forward selection results only for the unrestricted k-factor model, MLFA-U.

The results for the simulation experiments for examples A, B, C, and D are presented in the following sections.

3.4 Simulation results: example A

The first simulation experiment for example A had a single linear random effect with standard deviation σ_1 where σ_1/σ was varied from zero to two. The second experiment had two independent linear random effects: the first with large standard deviation ($\sigma_1/\sigma = 2$) and the second with standard deviation σ_2 where σ_2/σ was varied from zero to two. The results for these two experiments are described in Sections 3.4.1 and 3.4.2.

3.4.1 One linear random effect

The convergence rate for maximum likelihood estimation of the one-factor MLFA-U model was above 99% in all cases. The convergence rate for the one-factor MLFA-P model was 100%.

Number of random effects

In the case of linear random effects, we expect the power of the likelihood ratio tests for adequacy of zero common factors or principal components to increase with σ_1/σ, and to have size close to the nominal level of 0.05 when $\sigma_1/\sigma = 0$ (no random effect). The PCA test and the MLFA-U test for $n > p$ rely on large-n asymptotics; the PCA test for $n \leq p$ relies on large-p asymptotics. We used the small-sample corrected versions of the likelihood ratio tests to improve the approximate chi-squared distributions of the test statistics under the null hypothesis. Thus when $n > p$ or when $n < p$ we expect the tests to maintain size close to 0.05 and to readily reject the hypothesis of no common factors when $\sigma_1/\sigma > 0$. When $n \approx p$ asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. The PCA analysis assumes that $\gamma = 0$ and actually tests the hypothesis that the smallest $p - k$ eigenvalues of the marginal covariance matrix are equal. When $\gamma > 0$, the smallest eigenvalues will not be equal, even when there are no common factors. When $\gamma > 0$ we would expect the PCA test to reject the null hypothesis even when there were no common factors.

In the batched ANOVA of simulation results, the magnitude of γ, the number of subjects n, the magnitude of the random effect σ_1/σ, and interactions between these three factors explained 76% of the variation in power of the PCA test. The magnitude of σ_1/σ, γ, and the interaction between σ_1/σ and γ explained 92% of variation in the power of the MLFA-U test.

Figure 3.6 displays estimated power of the PCA test against the magnitude of the random effect signal σ_1/σ when $p = 25$. Results for $p = 10$ and $p = 50$ were similar to those presented in the figure except for a slight increase in power as p increased for cases where $\sigma_1/\sigma > 0$.

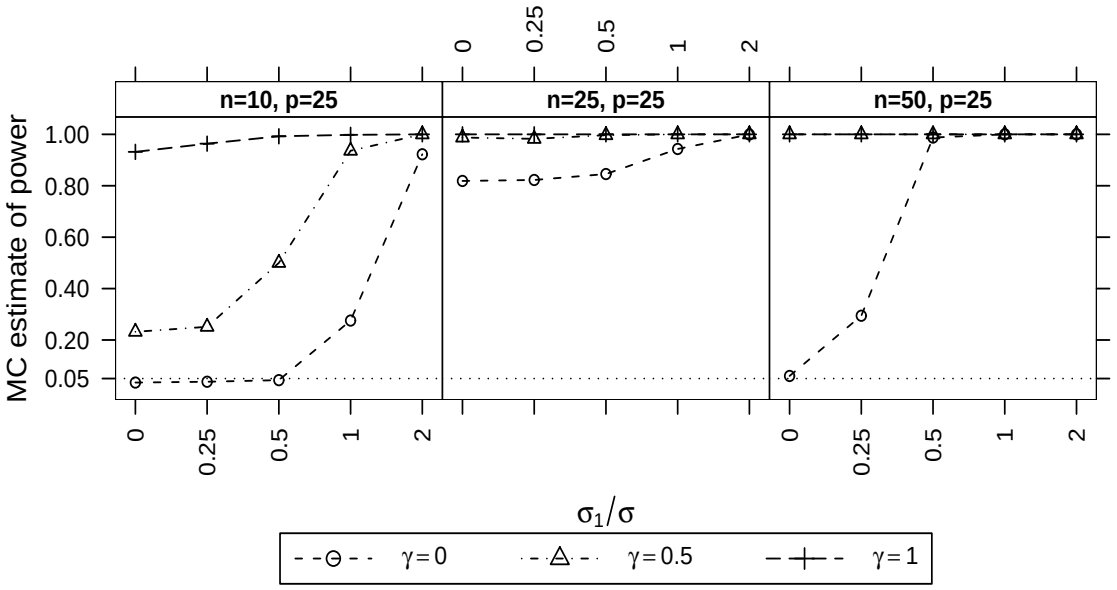

Figure 3.6: *Estimated power of the PCA test for* 0 *common factors when* $p = 25$ *and there was one linear random effect. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

The three panels in the figure show the effect of a different number of subjects n for fixed p. As n increased, power increased. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. When $\gamma = 0$ and $n < p$ or $n > p$, the test had approximately the correct size

Table 3.6: *Estimated size of the PCA test for* 0 *common factors when* $\gamma = 0$ *and there was one linear random effect with* $\sigma_1/\sigma = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	p		
n	10	25	50
10	0.161	0.034	0.038
25	0.044	0.819	0.054
50	0.054	0.060	1.000

(0.05) when $\sigma_1/\sigma = 0$. Table 3.6 gives estimated size of the PCA test at $\sigma_1/\sigma = 0$ when $\gamma = 0$ for different values of n and p. Clearly when $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on and the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there were no principal components too often. When $\gamma = 0$ and $n < p$ or $n > p$, the test had greater than 80% power to detect a random effect with $\sigma_1/\sigma = 2$.

The test of no common factors that is based on fitting the MLFA-U model only exists when $n > p$. Figure 3.7 presents the power curves for the MLFA-U test for the cases where $n > p$. The size of the test was close to the nominal 0.05 level except in the cases when $n = 50$ and $p = 25$ where the test rejected the hypothesis of no common factors in approximately 8% to 10% of replicates (see Table 3.7). In all cases where $n > p$, the test had greater than 90% power to detect a random effect with $\sigma_1/\sigma = 1$. For $\sigma_1/\sigma > 0$, power increased with increasing n

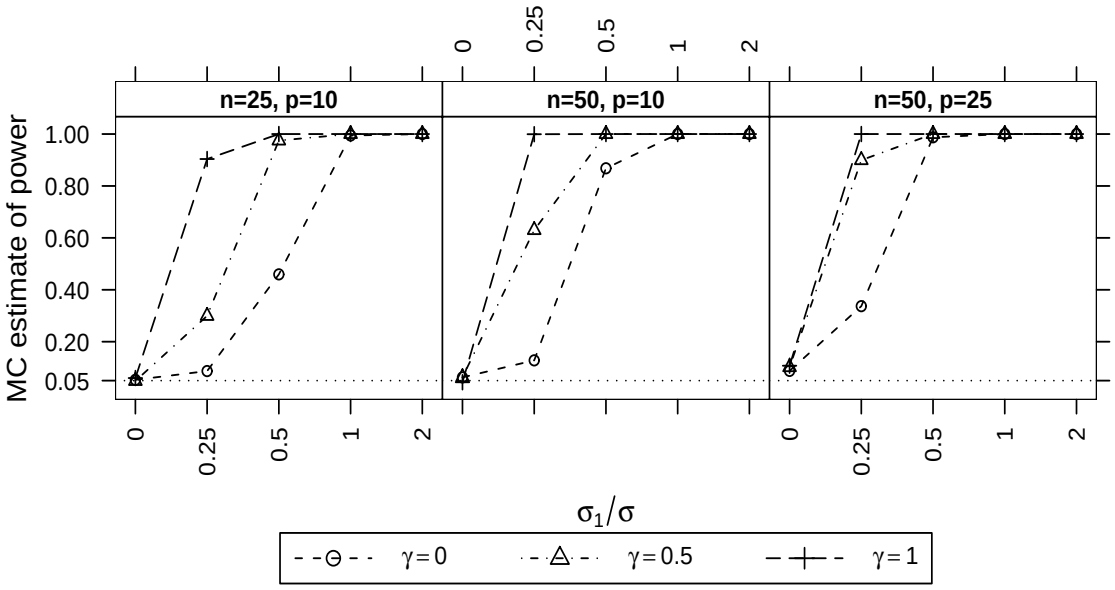

Figure 3.7: *Estimated power of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one linear random effect. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Table 3.7: *Estimated size of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one linear random effect with* $\sigma_1/\sigma = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

		γ		
n	p	0.0	0.5	1.0
25	10	0.053	0.049	0.060
50	10	0.063	0.063	0.044
50	25	0.087	0.102	0.107

and p. Interestingly, power to reject the hypothesis of no common factors also increased with increasing γ. When $n > p$ and the within-subject variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same power, although the size of the MLFA-U test was marginally higher than the PCA test when $n = 50$ and $p = 25$.

Test for homogeneous variance

For estimating p-values of the likelihood ratio tests for homogeneous variance, we relied on the asymptotic chi-squared distribution under the hypothesis that $\gamma = 0$ (we have not developed a finite-sample correction for these statistics). Therefore we expect the chi-squared approximation to be adequate when n is large relative to p. More parameters ($p - 2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ.

In the batched ANOVA of simulation results, the magnitude of γ, p, n, and interactions between these factors explained 95% of variation in the power of the likelihood ratio test comparing the MLFA-U and MLFA-H models. The magnitude of γ explained 97% of variation in power of the test comparing the MLFA-P model with the MLFA-H model. Since magnitude

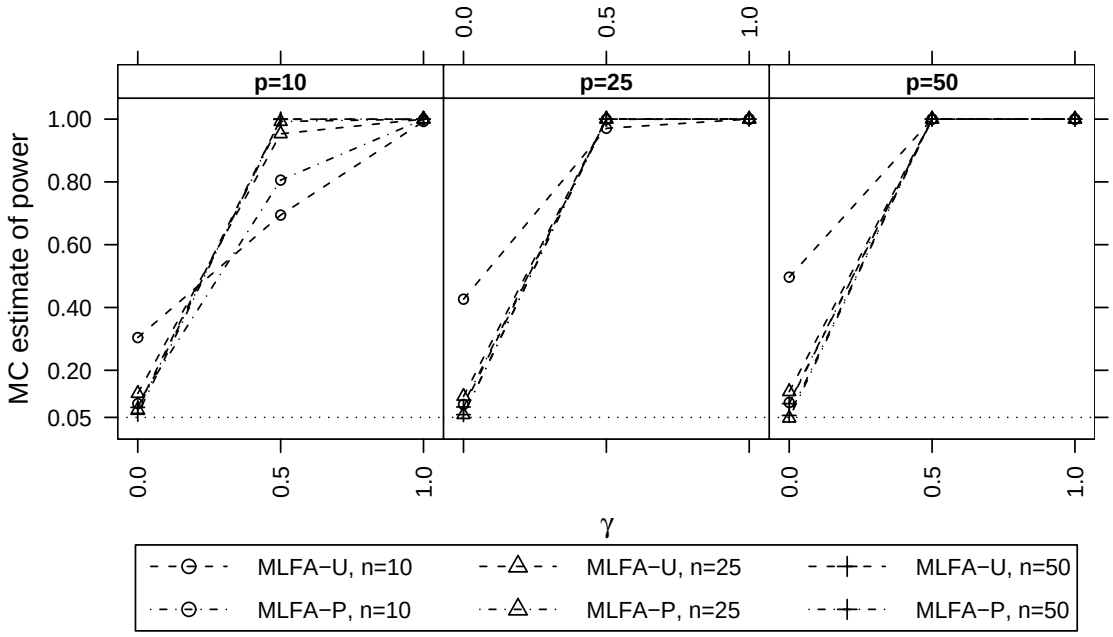

Figure 3.8: *Estimated power of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there was one linear random effect and* $\sigma_1/\sigma = 0.5$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

of the random effect signal did not have a large impact, we have presented results only for $\sigma_1/\sigma = 0.5$ in Figure 3.8.

In all cases, the size of the test comparing the homoscedastic model to the unrestricted model MLFA-U was larger than the nominal level of 0.05 (see Table 3.8). When $n = 10$, the MLFA-U test was close to useless, rejecting the hypothesis of $\gamma = 0$ in almost 50% of cases when $p = 50$. The test comparing the homoscedastic model to the power model MLFA-P performed better, although its size was larger than the nominal level when $n = 10$ or when $p = 10$. When $n > 10$ or $p > 10$, this test had greater than 90% power to detect heteroscedasticity when γ was as large as 0.5.

Table 3.8: *Estimated size of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there was one linear random effect and* $\sigma_1/\sigma = 0.5$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	MLFA-U			MLFA-P		
n	$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
10	0.304	0.426	0.497	0.094	0.091	0.098
25	0.126	0.117	0.132	0.075	0.059	0.048
50	0.082	0.083	0.094	0.060	0.059	0.056

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2\mu_j^{2\gamma}$, by fitting the one-factor power model MLFA-P. On average, the bias in $\widehat{\gamma}$ was approximately zero, but the estimate of σ^2 was smaller on average than the true value of 0.0001. In the batched ANOVA of simulation results, the simulation factors explained little of the variation in the bias of $\widehat{\gamma}$ (all factors and their interactions explained less than 4% of variation). The magnitude of the random effect, σ_1/σ, and the number of subjects, n, explained 44% of the bias in $\widehat{\sigma}^2$ (all simulation factors and their simulation factors explained 47% of the variation). Figure 3.9 illustrates the pattern in the estimate of σ^2 for $p = 10$. The pattern was similar for p larger than 10 except that there tended to be less variation between results for different settings of γ. The three panels in the figure show the effect of increasing n for fixed p; bias decreased with increasing n. On average, bias in $\widehat{\sigma}^2$ was largest when $\sigma_1/\sigma = 0$. In this case, the true model had no common factors, but one factor was forced into the model at the expense of the specific variance part of the model. This resulted in an estimate of σ^2 that was too small on average. The three curves in each panel represent the mean estimate of σ^2 for each setting of γ. The value of γ had no consistent effect on bias of the estimate of σ^2.

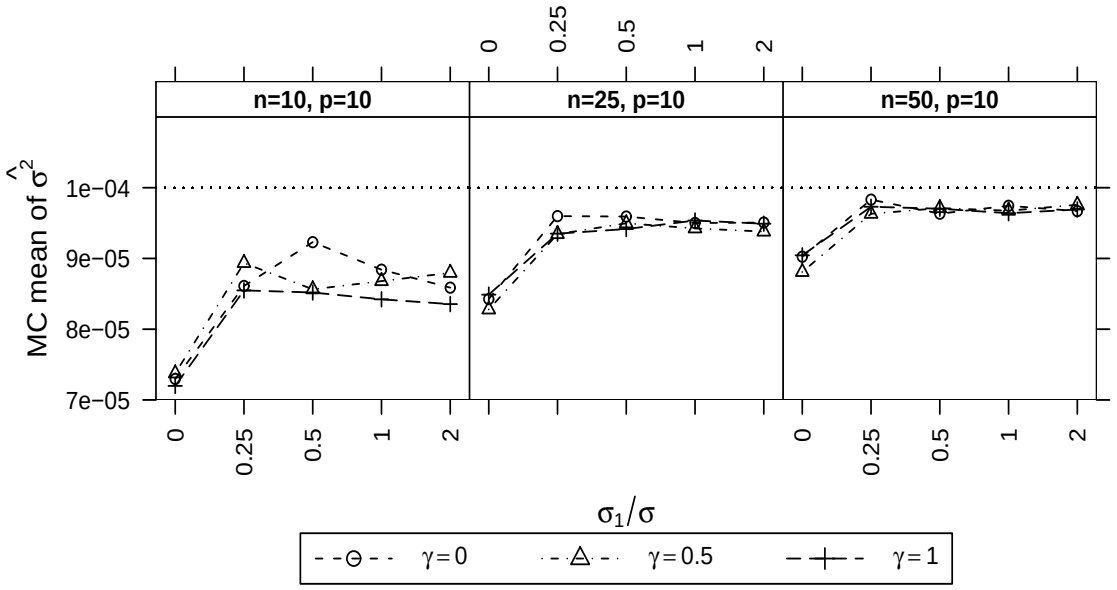

Figure 3.9: *Mean estimate of* σ^2 *from fitting the one-factor MLFA-P model when there was one linear random effect and* $p = 10$*. The true value of* σ^2 *was* $1\text{e}{-4}$*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* $2\text{e}{-6}$*.*

Selecting random effects

When $k = 1$ we used the exact test where where the j-th element of $\boldsymbol{\kappa}$ was determined by the true common factor loading,

$$\kappa_j \propto f_{\omega^{(2)}}(t_j) = -(1-\alpha)t_j \exp\left(-\omega^{(2)} t_j\right).$$

We present power of the exact test based on the PCA (results based on the one-factor MLFA-U and MLFA-P models were virtually indistinguishable from these). Since the test is based on the exact distribution of the first principal component loading under the null hypothesis, we expect the test to have the nominal size (0.05) when $\sigma_1/\sigma = 0$ and $\gamma = 0$. The power should increase for increasing signal, $\sigma_1/\sigma > 0$. Sample size, n, and number of time points, p, should have no effect on size of the test when $\gamma = 0$, but may affect power when the alternative hypothesis is

Table 3.9: *Estimated size of the PCA exact test when there was one linear random effect. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	0.048	0.147	0.171	0.049	0.105	0.109	0.054	0.088	0.062
25	0.058	0.167	0.120	0.056	0.114	0.052	0.050	0.087	0.042
50	0.054	0.159	0.064	0.050	0.086	0.027	0.051	0.072	0.024

true. When $\gamma > 0$, the principal component model that assumes homogeneous within-subject variance is incorrect, and we would expect size and power of the test to deviate somewhat from those in the case where $\gamma = 0$.

Based on the batched ANOVA, the random effect signal, σ_1/σ, γ, and interaction between these factors, explained 94% of variation in the power of the exact test. Table 3.9 gives the size of the test under the null hypothesis. When $\gamma = 0$ (homoscedastic case) the size was not significantly different from the nominal 0.05 level regardless of n and p. However when $\gamma > 0$, the size of the test was not preserved; in general when $\gamma > 0$ the size was larger than 0.05, but was smaller than 0.05 when $\gamma = 1$ and $n = 50$.

In Figure 3.10 we present estimated power of the PCA exact test to detect the true random effect. We only present power for the case where $\gamma = 0$ since size of the test was not preserved for $\gamma > 0$. Power increased for increasing random effect magnitude, σ_1/σ. The three panels in the figure show the effect of p; power increased with increasing p. In each panel, the three curves show power for sample of different size; power increased with increasing n. Except in the case where $n = p = 10$, the exact test had greater than 90% power to detect the true random effect when the magnitude of random effect variation was given by $\sigma_1/\sigma = 0.5$.

For the multivariate forward selection we used three candidate predictors proportional to

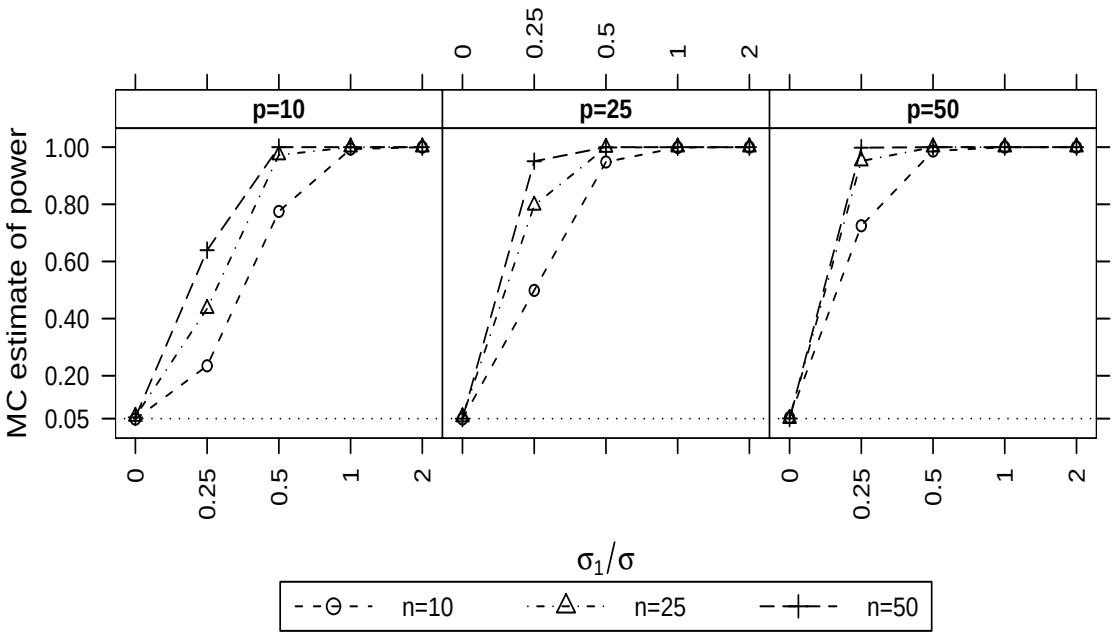

Figure 3.10: *Estimated power of the PCA exact test when* $\gamma = 0$ *and there was one linear random effect. This test rejects* H_0*:* $\mathbf{\Sigma} = \sigma^2 \mathbf{I}_p$ *in favor of one random effect with loadings function* $\lambda^{(1)}$ *. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

the sensitivity functions of the O2C model,

$$\begin{aligned} f_{\alpha}(t) &= \exp(-\omega^{(1)}t) - \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}}(t) &= -\alpha t \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}}(t) &= -(1-\alpha) t \exp(-\omega^{(2)}t). \end{aligned}$$

The true loadings function, $\lambda^{(1)}$, was proportional to $f_{\omega^{(2)}}$. When $\sigma_1/\sigma = 0$ the forward selection should select one or more sensitivities in about 5% of cases. As σ_1/σ gets larger, the forward selection should select the sensitivity $f_{\omega^{(2)}}$ in a high proportion of cases. Since the candidate predictors (parameter sensitivities) were not orthogonal, we might expect the incorrect sensitivity to be selected in more cases than if all the sensitivities were orthogonal.

We would expect the power to select the correct sensitivity to increase with p, the number of time points (also the number of rows in the multivariate regression model). Power should also increase with n as the estimator of the loadings vector should become more precise with larger n.

The multivariate regression results based on the PCA, the MLFA-U and the MLFA-P models were very similar. We present results for the unrestricted MLFA-U model. Based on the batched ANOVA of simulation results, the magnitude of the random effect σ_1/σ, γ, and the interaction between these factors explained 91% of the variation in the mean number of random effects selected, and 93% of the variation in the proportion of times that the true sensitivity, $f_{\omega^{(2)}}$, was selected. For the sake of brevity, we present results only for the cases where $\gamma = 0$ or $\gamma = 1$.

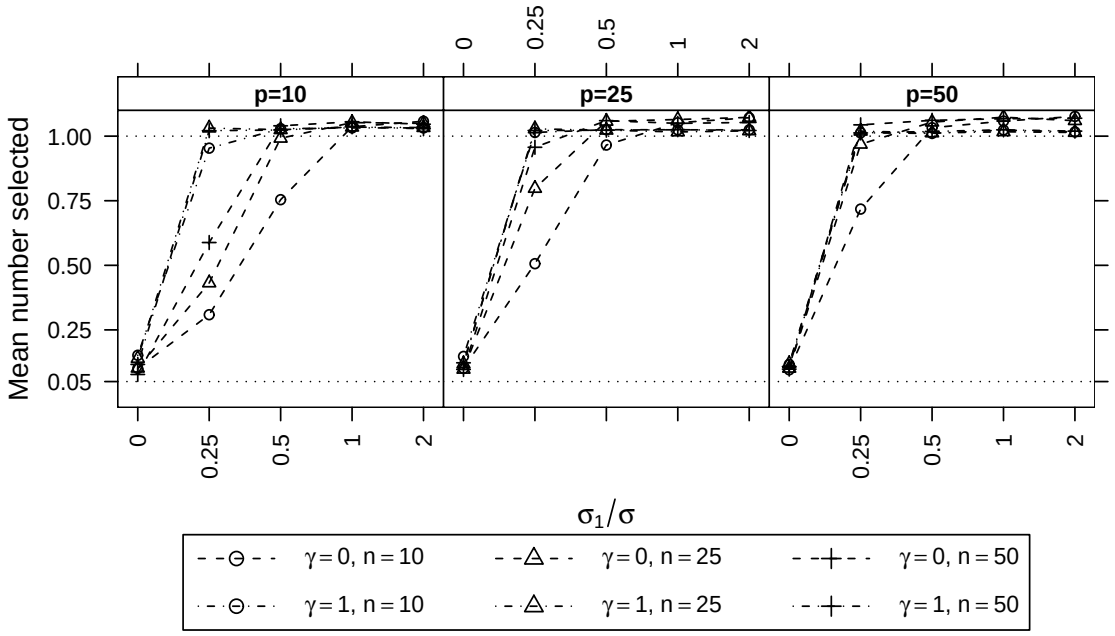

Figure 3.11: *Mean number of sensitivities selected by multivariate forward selection when there was one linear random effect. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.015*.*

Table 3.10: *Mean number of sensitivities included by forward selection for the experiment with one linear random effect with* $\sigma_1/\sigma = 0$*. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each MC mean to be less than* 0.015*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.101	0.152	0.101	0.147	0.096	0.116
25	0.102	0.140	0.098	0.118	0.102	0.119
50	0.078	0.116	0.097	0.123	0.098	0.108

Figure 3.11 displays the mean number of sensitivities selected by multivariate forward selection based on the factor loading estimated from the one-factor MLFA-U model. As the random effect signal, σ_1/σ, increased, the mean number of sensitivities selected approached 1 (the true number of random effects). When there was no random effect ($\sigma_1/\sigma = 0$), the forward selection procedure selected a model with one or more sensitivities in more cases than should be the case at the 0.05 nominal level (see Table 3.10), especially in the presence of heteroscedasticity ($\gamma = 1$). This could be due to the non-normality of the estimated loadings vector, due to multiple testing (we used the nominal 0.05 critical value at each step of the forward selection), or perhaps because the means in the table include some models that included more than 1 sensitivity. The three panels in the figure show the effect of increasing p; as expected, the mean number of sensitivities selected increased toward 1 as p increased. In each panel the dashed curves are for the case where $\gamma = 0$ and the dot-dash curves are for $\gamma = 1$. Different plotting symbols represent the different levels of n. The mean number of sensitivities selected increased toward 1 as n increased. The forward selection procedure tended to select a larger model in the presence of heteroscedasticity ($\gamma = 1$) than when within-subject variation was homoscedastic ($\gamma = 0$).

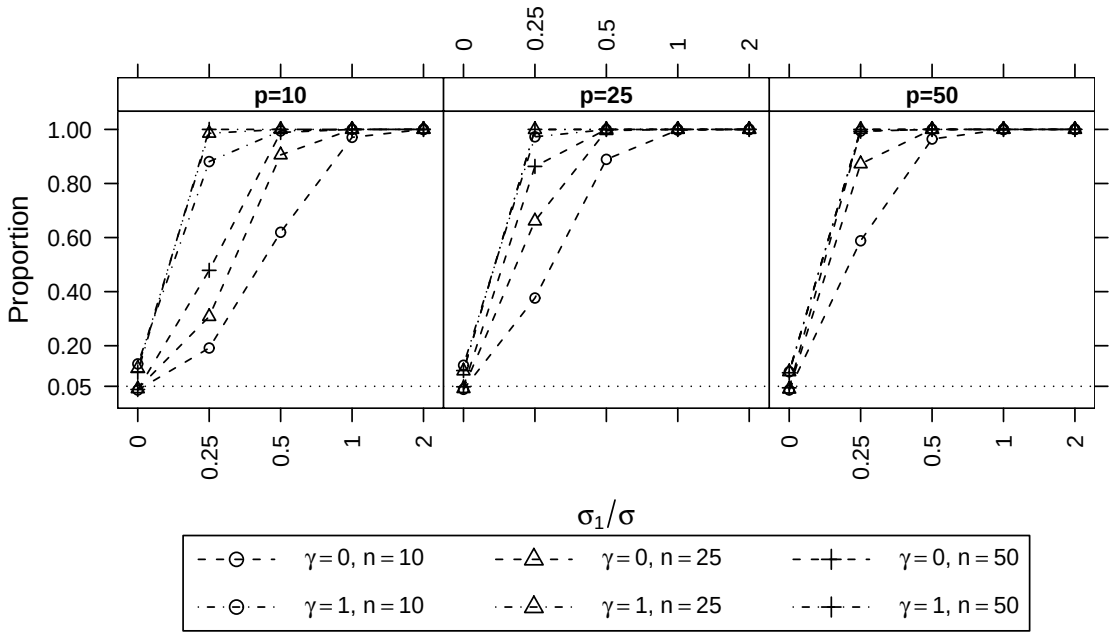

Figure 3.12: *Proportion of cases where multivariate forward selection picked the correct sensitivity* $f_{\omega^{(2)}}$ *when there was one linear random effect. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125*.*

Figure 3.12 shows the proportion of cases where multivariate forward selection picked the sensitivity $f_{\omega^{(2)}}$ corresponding to the actual random effect. When there was no random effect ($\sigma_1/\sigma = 0$), the sensitivity $f_{\omega^{(2)}}$ was selected in slightly fewer than 5% of cases when within-subject variation was homoscedastic ($\gamma = 0$) (see Table 3.11). In the presence of heteroscedasticity ($\gamma = 1$), the sensitivity $f_{\omega^{(2)}}$ was selected in approximately 10% of cases. When there was no random effect, we should not expect $f_{\omega^{(2)}}$ to be selected in preference to other sensitivities. Based on comparing the values in Table 3.10 and Table 3.11, when $\sigma_1/\sigma = 0$, the sensitivity $f_{\omega^{(2)}}$ was not included in every model that included one or more sensitivities. As shown in the figure, as the random effect signal σ_1/σ increased, the correct sensitivity was included in the model in a high proportion of cases. The three panels in the figure show the

Table 3.11: *Proportion of cases where forward selection picked* $f_{\omega^{(2)}}$ *when* $\sigma_1/\sigma = 0$ *in the experiment with one linear random effect. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.036	0.133	0.039	0.128	0.036	0.104
25	0.041	0.117	0.041	0.107	0.039	0.104
50	0.038	0.101	0.048	0.108	0.044	0.100

effect of increasing p; the correct sensitivity was picked more often when p was larger. In each panel the dashed curves represent the cases where $\gamma = 0$ and the dot-dash curves are for $\gamma = 1$. Different plotting symbols represent different levels of n. The correct sensitivity was selected in a greater proportion of cases as n increased. The sensitivity $f_{\omega^{(2)}}$ was selected more often in the presence of heteroscedasticity ($\gamma = 1$) than when within-subject variation was homoscedastic ($\gamma = 0$). To summarize briefly: when σ_1/σ, n, and p were large, the forward selection procedure tended to pick a model with one sensitivity that corresponded to the true random effect.

3.4.2 Two linear random effects

In this experiment there were two independent linear random effects. The first had large standard deviation ($\sigma_1/\sigma = 2$). The second random effect had standard deviation σ_2 where we varied σ_2/σ from zero to two. The convergence rate for maximum likelihood estimation of the two-factor MLFA-U model was above 97% in all cases. The convergence rate for the two-factor MLFA-P model was 100%. We expect results to be similar to the case of one linear random effect, except that power may tend to be lower because more parameters have to be estimated in a model with two factors than in a model with one factor.

Number of random effects

When $n > p$ or when $n < p$ (for the PCA test) we expect the likelihood ratio tests to maintain size close to 0.05 and to readily reject the hypothesis of one common factor when $\sigma_2/\sigma > 0$. When $n \approx p$ asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we would expect the PCA test to reject the null hypothesis even when there was only one common factor.

In the batched ANOVA of simulation results, the magnitude of γ, the number of subjects n, number of time points, p, the magnitude of the second random effect σ_2/σ, and two-way interactions of all four factors, explained 76% of the variation in power of the PCA test. The signal, σ_2/σ, magnitude of γ, and interaction between these two factors explained 94% of variation in the power of the MLFA-U test.

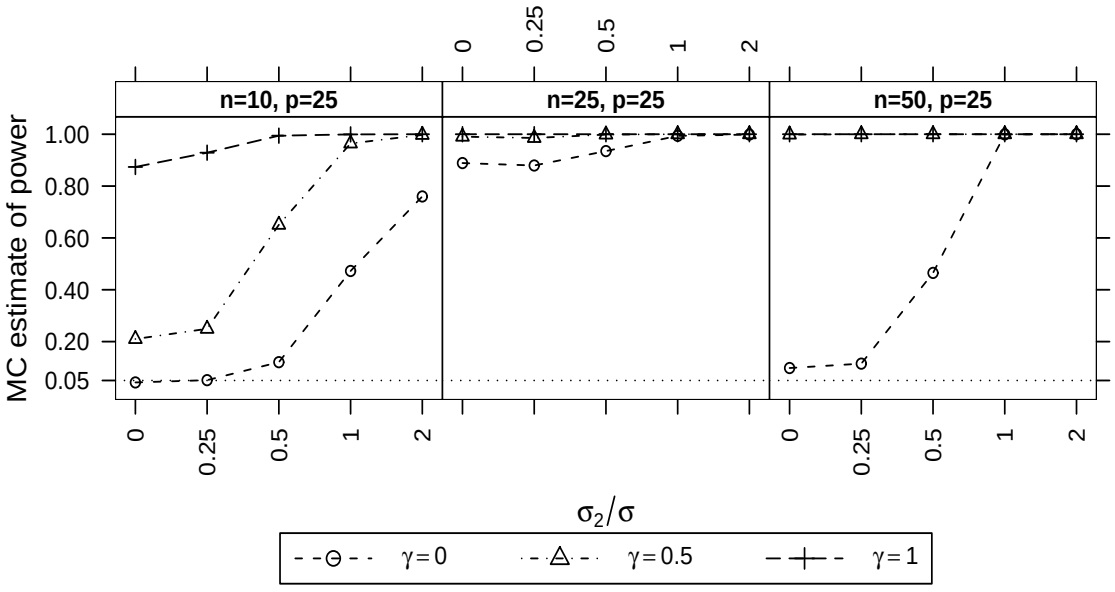

Figure 3.13: *Estimated power of the PCA test for* 1 *common factor when* $p = 25$ *and there were two linear random effects. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Figure 3.13 shows estimated power of the PCA test against the magnitude of the second random effect signal σ_2/σ when $p = 25$. Results for $p = 10$ and $p = 50$ were similar to those presented in the figure except for a slight increase in power as p increased for cases where $\upsilon_\omega > 0$. The three panels in the figure show the effect of a different number of subjects, n, for fixed p. As n increased, power increased. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$.

Table 3.23 gives estimated size of the PCA test at $\upsilon_\omega = 0$ when $\gamma = 0$ for different values of n and p. When n was much larger than p, or p was much larger than n, then test had approximately the correct size (0.05) when $\gamma = 0$. When the ratio of n to p was closer to one, the size was larger than the nominal value. Clearly when $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on and the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there was one principal component too often and would not be useful in the presence of heteroscedastic within-subject variance. In most cases, when $\gamma = 0$ and $n < p$ or $n > p$, the test had greater than 80% power to detect a random effect with $\sigma_2/\sigma = 2$, although as expected, the test was less powerful when there were two linear random effects than when there was one linear random effect.

Table 3.12: *Estimated size of the PCA test for* 1 *common factor when* $\gamma = 0$ *and there were two linear random effects, the second one with* $\sigma_2/\sigma = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

	p		
n	10	25	50
10	0.287	0.043	0.052
25	0.071	0.889	0.072
50	0.048	0.098	1.000

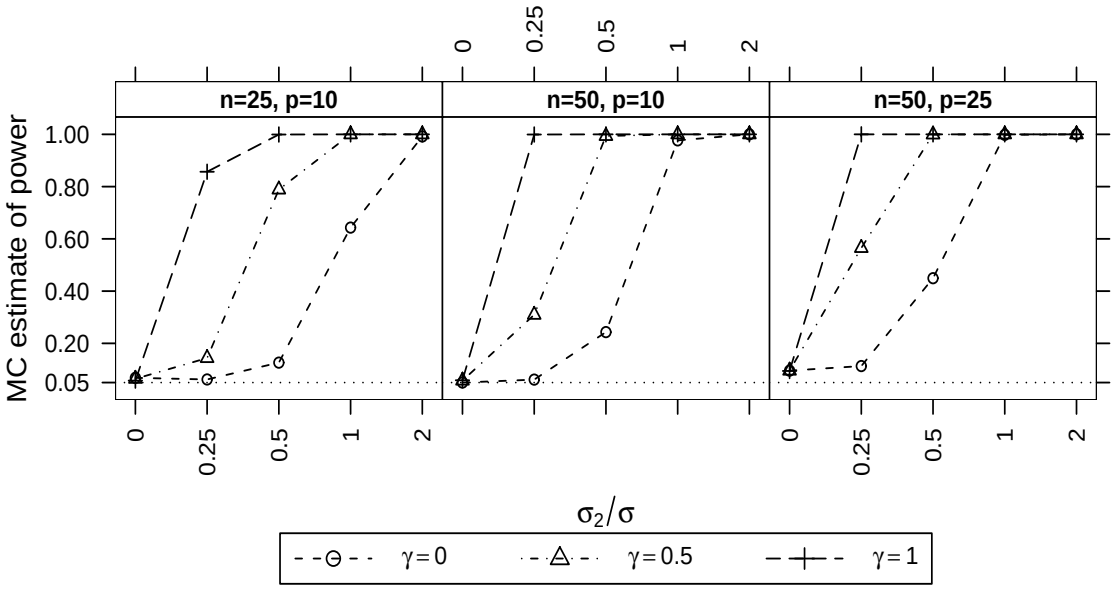

Figure 3.14: *Estimated power of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two linear random effects. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Figure 3.14 presents the power curves for the MLFA-U test for the cases where $n > p$. The size of the test was close to the nominal 0.05 level when $n = 50$ and $p = 10$. When the ratio of n to p was closer to one, the size of the test was larger than the nominal level (see Table 3.13). Except in the case where $\gamma = 0$, $n = 25$ and $p = 10$, the test had greater than 90% power to detect a second random effect with $\sigma_2/\sigma = 1$ or larger. For $\sigma_2/\sigma > 0$, power increased with increasing n and p. Interestingly, power to reject the hypothesis of one common factor increased with increasing γ. When $n > p$ and the within-subject variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same power.

Test for homogeneous variance

For the likelihood ratio tests comparing the fits of the MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation of the null distribution to be adequate when n is large relative to p. More parameters ($p - 2$ more) are estimated in fitting the MLFA-U model than

Table 3.13: *Estimated size of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two linear random effects, the second one with* $\sigma_2/\sigma = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

		γ		
n	p	0.0	0.5	1.0
25	10	0.068	0.064	0.058
50	10	0.050	0.058	0.059
50	25	0.096	0.096	0.094

in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ.

The magnitude of γ, p, n, and two-way interactions between these factors explained 94% of variation in the power to reject the hypothesis of homoscedasticity by comparing the MLFA-H model with the unrestricted MLFA-U model. The magnitude of γ explained 96% of variation in power of the test comparing the MLFA-H model with the MLFA-P model. Since magnitude of the second random effect signal did not have a large impact, we have presented results only for $\sigma_2/\sigma = 0.5$ in Figure 3.15.

In all cases, the size of the test comparing the homoscedastic model to the unrestricted MLFA-U model was larger than the nominal level of 0.05 (see Table 3.14). When $n = 10$ or $n = 25$, the MLFA-U test was close to useless, rejecting the hypothesis that $\gamma = 0$ in 85% of cases when $n = 10$ and $p = 50$. The test comparing the homoscedastic model to the MLFA-P performed better, although its size was larger than the nominal level in all cases except where $n = 50$ and $p = 50$. Compared with the results when there was a single linear random effect, when there were two random effects, the null distributions were less well approximated by the asymptotic chi-squared distribution. When $n > 10$ or $p > 10$ the test based on the MLFA-P model had greater than 90% power to detect heteroscedasticity when γ was as large as 0.5.

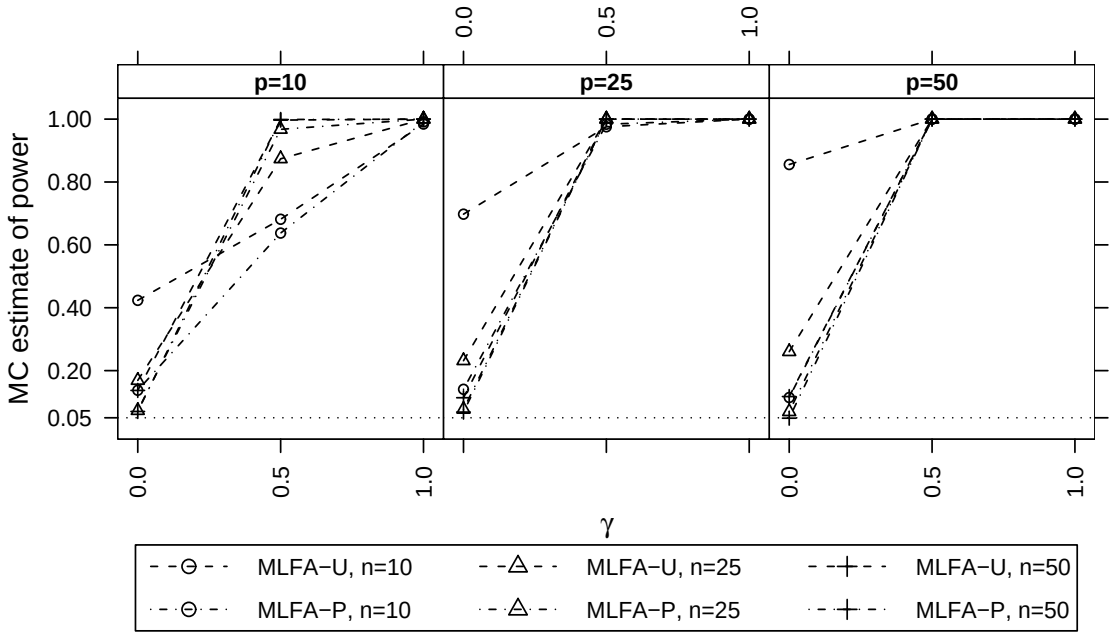

Figure 3.15: *Estimated power of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two linear random effects and* $\sigma_2/\sigma = 0.5$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the two-factor MLFA-P model. On average, the bias in $\widehat{\gamma}$ was approximately zero. In the batched ANOVA, all simulation factors and their interactions explained less than 10% of variation in the bias of $\widehat{\gamma}$. The estimate of σ^2 was smaller on average than the true value of 0.0001. Bias of $\widehat{\sigma}^2$ was larger in this experiment than when there was only one random effect. The number of subjects n, the magnitude of the random effect, σ_2/σ, and γ explained 65% of variation in bias of $\widehat{\sigma}^2$ (the ANOVA with all main effects and interactions explained 67% of variation). Figure 3.16 shows the mean estimate of σ^2 when $p = 10$ for different values of n, γ and σ_2/σ. The pattern was similar for p larger than 10. The three panels in the figure

Table 3.14: *Estimated size of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two linear random effects and* $\sigma_2/\sigma = 0.5$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	MLFA-U			MLFA-P		
n	$p = 10$	$p = 25$	$p = 50$	$p = 10$	$p = 25$	$p = 50$
10	0.424	0.698	0.856	0.138	0.141	0.114
25	0.169	0.232	0.260	0.073	0.077	0.069
50	0.137	0.114	0.117	0.070	0.068	0.049

show the effect of increasing n for fixed p; bias decreased with increasing n. The three curves in each panel represent the mean estimate of σ^2 for each setting of γ. On average, bias in $\widehat{\sigma}^2$ was smaller when the second random effect signal was large ($\sigma_2/\sigma = 2$) or when $\gamma = 1$.

Selecting random effects

We used multivariate forward selection to identify the correct random effect loading functions from amongst several possible predictors, with the first two PCA or factor loadings as multivariate response variables. The three candidate predictors were proportional to the three sensitivity functions of the O2C model,

$$\begin{aligned} f_\alpha(t) &= \exp(-\omega^{(1)}t) - \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}}(t) &= -\alpha t \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}}(t) &= -(1-\alpha)t \exp(-\omega^{(2)}t). \end{aligned}$$

The true loadings function, $\lambda^{(1)}$, of the first random effect was proportional to $f_{\omega^{(2)}}$. The true loadings function, $\lambda^{(2)}$, of the second random effect was proportional to f_α. When $\sigma_2/\sigma = 0$, the forward selection should select a model with one sensitivity corresponding to the first random

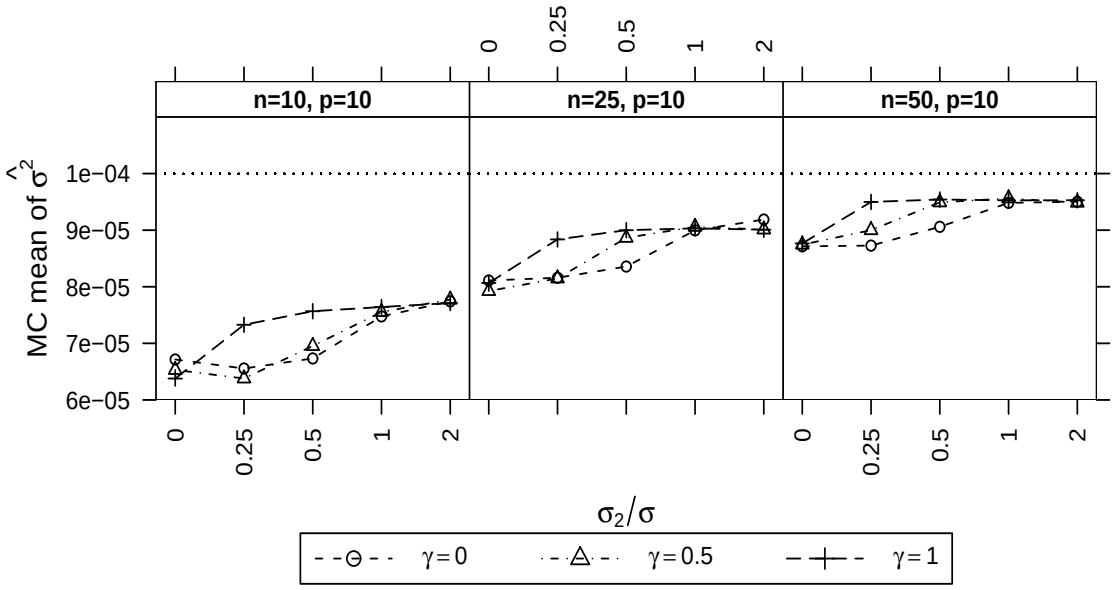

Figure 3.16: *Mean estimate of* σ^2 *from fitting the two-factor MLFA-P model when there were two linear random effects and* $p = 10$*. The true value of* σ^2 *was* $1e{-}4$*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* $2e{-}6$*.*

effect ($f_{\omega^{(2)}}$). As σ_2/σ gets larger, the forward selection should additionally select the sensitivity f_α in a high proportion of cases. Since the candidate predictors (parameter sensitivities) are not orthogonal, we might expect the incorrect sensitivity to be selected in more cases than if all the sensitivities were orthogonal. We would expect the power to select the correct sensitivity to increase with p, the number of time points (also the number of rows in the multivariate regression model). Power should also increase with n because the estimator of the loadings should become more precise with larger n.

The multivariate regression results based on the MLFA-U and the MLFA-P models were very similar. The results for the PCA analysis were similar to the MLFA results when $\gamma = 0$. When $\gamma > 0$ and the second random effect signal was small ($\sigma_2/\sigma = 0.25$), multivariate forward selection based on the PCA loadings selected a smaller model than when selection was based on the MLFA-U or MLFA-P loadings. For the sake of brevity, we present results only for the

unrestricted MLFA-U model.

The sensitivity $f_{\omega^{(2)}}$ corresponding to the first random effect was included in the multivariate regression model in almost all cases. The magnitude of the second random effect (σ_2/σ), magnitude of γ, number of time points (p), number of subjects (n), and two-way interactions of these factors explained 94% of the variation in the mean number of random effects selected, and 93% of the variation in the proportion of times that the sensitivity f_α (corresponding to the second random effect) was selected.

Figure 3.17 displays the mean number of sensitivities selected by multivariate forward selection based on the factor loading estimated from the two-factor MLFA-U model. We have only presented the results for $\gamma = 0$ and $\gamma = 1$. As the random effect signal σ_2/σ increased,

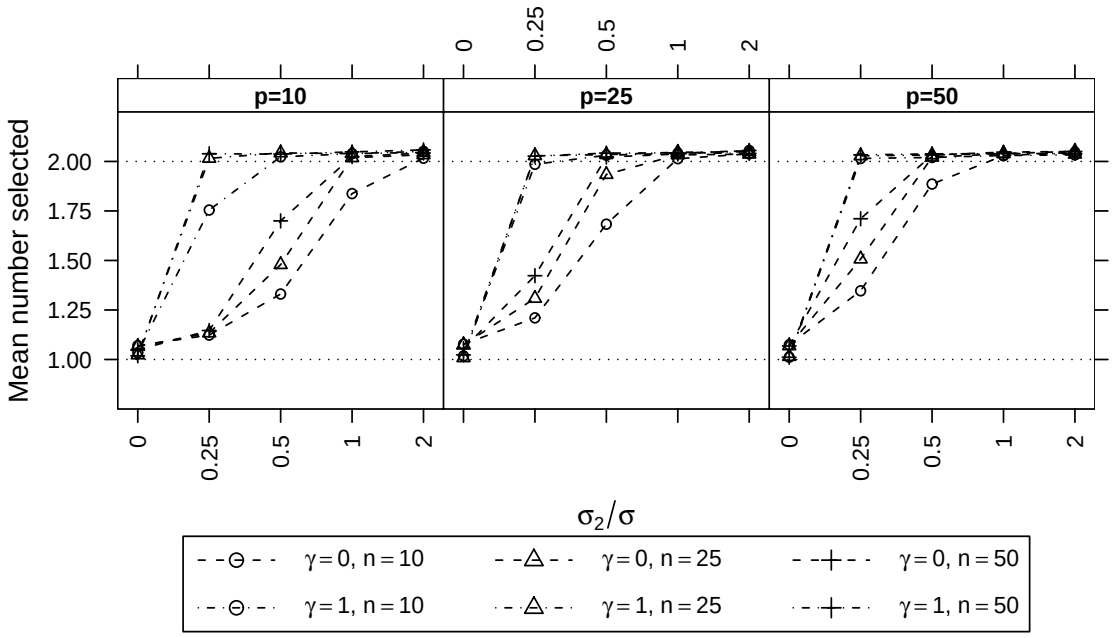

Figure 3.17: *Mean number of sensitivities selected by multivariate forward selection when there were two linear random effects. Results presented are for the regression using the factor loadings estimated from the two-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.015*.*

Table 3.15: *Mean number of sensitivities included by forward selection in the experiment with two linear random effects when* $\sigma_2/\sigma = 0$. *Results presented are for the regression using the factor loading estimated from the two-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$. *At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10, *we estimated the standard error of each MC mean to be less than* 0.015.

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	1.070	1.034	1.077	1.018	1.074	1.012
25	1.066	1.023	1.074	1.009	1.069	1.015
50	1.049	1.019	1.054	1.023	1.068	1.011

the mean number of sensitivities selected approached 2 (the true number of random effects). When there was only one random effect ($\sigma_2/\sigma = 0$), the mean model size chosen by the forward selection procedure was slightly larger than one (see Table 3.15). The three panels in the figure show the effect of increasing p; as expected, the mean number of sensitivities selected increased toward 2 as p increased. In each panel the dashed curves are for the case where $\gamma = 0$ and the dot-dash curves are for $\gamma = 1$. Different plotting symbols represent the different levels of n. The mean number of sensitivities selected increased toward 2 as n increased. Except when $\sigma_2/\sigma = 0$, the forward selection procedure selected a larger model on average in the presence of heteroscedasticity ($\gamma = 1$) than when within-subject variation was homoscedastic ($\gamma = 0$).

Figure 3.18 shows the proportion of cases where multivariate forward selection picked the sensitivity $f_{\omega^{(2)}}$ corresponding to the first random effect, and f_α corresponding to the second random effect. We have plotted results only for the case where $n = 25$. Results were similar for other values of n, with a slight increase in selection proportion when n was larger. In all cases, the sensitivity $f_{\omega^{(2)}}$ corresponding to the first random effect was selected in greater than 95% of replicates. When there was only one random effect ($\sigma_2/\sigma = 0$), the sensitivity f_α corresponding to the second random effect was selected in fewer than 5% of cases although the multivariate

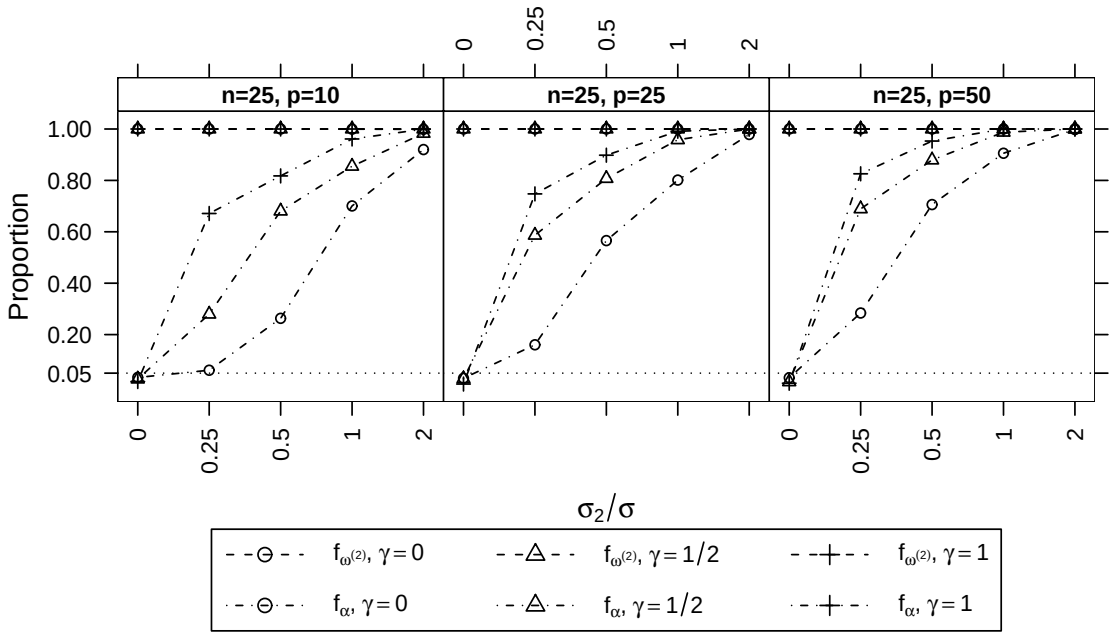

Figure 3.18: *Proportion of cases where multivariate forward selection picked the correct sensitivities, $f_{\omega^{(2)}}$ (dashed curves) and f_α (dot-dash curves), when there were two linear random effects. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $n = 25$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125*.*

forward selection used a 0.05-level critical value at each stage (see Table 3.16). When there was only one random effect with loading proportional to $f_{\omega^{(2)}}$, we should not expect f_α to be selected in preference to the other sensitivity ($f_{\omega^{(1)}}$). This may account for the smaller than 0.05 selection rate of f_α when $\sigma_2/\sigma = 0$.

As the second random effect signal σ_2/σ increased, the sensitivity f_α was included in the model in an increasing proportion of cases. The three panels in the figure show the effect of increasing p; the sensitivity corresponding to the second random effect was picked more often when p was larger. In each panel the different plotting symbols represent different levels of γ. The sensitivity f_α was selected more often when γ was larger.

It is interesting to compare Figure 3.18 for the experiment with two linear random effects,

Table 3.16: *Proportion of cases where forward selection picked f_α in the experiment with two linear random effects with $\sigma_2/\sigma = 0$. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005.

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	0.028	0.025	0.023	0.037	0.021	0.013	0.032	0.019	0.008
25	0.033	0.027	0.018	0.030	0.028	0.008	0.032	0.016	0.011
50	0.018	0.024	0.013	0.029	0.024	0.018	0.034	0.016	0.010

to Figure 3.12 for the experiment with a single linear random effect. For a given magnitude of random effect signal, the sensitivity corresponding to the second random effect was selected in fewer cases than was the sensitivity corresponding to the first random effect in the experiment with just one linear random effect.

To summarize the multivariate regression results briefly: when σ_2/σ, n, and p were large, the forward selection procedure tended to pick a model with two sensitivities that corresponded to the true random effects. Power to detect a second random effect in the presence of another large random effect was lower than power to detect a single random effect.

3.5 Simulation results: example B

The first simulation experiment of example B had a single nonlinear random effect on the rate parameter $\delta^{(2)}$ in parameterization f^* of equation (3.3). The standard deviation of this random effect was $\upsilon_\delta|\delta^{(2)}|$ where we varied υ_δ from zero to 0.1. The second experiment generated data based on the parameterization f in equation (3.2). This experiment had two independent random effects. The first was on the linear scale parameter, α, and had large standard deviation, $\upsilon_\alpha|\alpha|$, where $\upsilon_\alpha = 0.4$. The second random effect was on the rate parameter $\omega^{(2)}$, with standard

deviation $\upsilon_\omega|\omega^{(2)}|$ where υ_ω varied from zero to 0.4.

We present results for these two experiments in Sections 3.5.1 and 3.5.2 below. We expect the results to play out a little differently than was seen in example A, where random effects were linear. In the simulation experiments, when data were generated with k random effects we fitted a model with k common factors or principal components. When the magnitude of the nonlinear random effect standard deviation becomes large, the first-order linear approximation that our methods rely on will be poor. For a large nonlinear random effect, a second-order approximation that includes quadratic functions of the random effects may be required; i.e., a model with more than k common factors or principal components may be necessary to adequately capture the between-subject variation.

As a consequence, if there is one nonlinear random effect, our tests may declare a model with one principal component or one common factor to be inadequate. For nonlinear random effects, the number of common factors or principal components apparent in the data may be larger than the actual number of random effects. When a model with k factors or principal components is fitted for data with k large nonlinear random effects, estimates of the within-subject variation (specific variances) will tend to be too large, and will possibly be confounded by unexplained between-subject variation (the quadratic terms in the second-order approximation). As a result, our tests for detecting departure from within-subject homoscedasticity may reject the null hypothesis too often when it is true.

In the multivariate regression, we have included relevant second-order sensitivities as candidate predictors, and would expect these to be selected in addition to the correct first-order sensitivities when the nonlinear random effects become large.

3.5.1 One nonlinear random effect

The convergence rate for maximum likelihood estimation of the one-factor MLFA-U model was above 99% in all cases. The convergence rate of maximum likelihood estimation for the one-factor MLFA-P model was 100%.

Number of random effects

We expect the power of the likelihood ratio tests to reject the hypothesis of no common factors, to increase with the random effect signal, υ_δ, and to have size close to the nominal level of 0.05 when $\upsilon_\delta = 0$ (no random effect) and $n > p$ or $n < p$ (for the PCA test). When $n \approx p$ asymptotic behavior cannot be expected. Since the signal of the nonlinear random effect is different from the signal of the linear random effect in example A, it is difficult to compare the power for a linear versus nonlinear random effect. However, we might expect higher power for a nonlinear random effect than for a linear random effect, when second-order effects of the nonlinear random effect become larger for large υ_δ. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we would expect the PCA test to reject the null hypothesis even when there were no common factors.

In the batched ANOVA of simulation results, the magnitude of the random effect (υ_δ), the magnitude of γ, the number of subjects (n), and two-way interactions between these factors explained 74% of the variation in power of the PCA test. The magnitude of the random effect (υ_δ) explained 99% of the variation in the power of the MLFA-U test.

Figure 3.19 displays the estimated power of the PCA test as a function of the magnitude of the random effect signal, υ_δ, when $p = 25$. Results for $p = 10$ and $p = 50$ were similar to

Table 3.17: *Estimated size of the PCA test for* 0 *common factors when* $\gamma = 0$ *and there was one nonlinear random effect in the O2C model with* $\upsilon_\delta = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

	p		
n	10	25	50
10	0.176	0.031	0.048
25	0.034	0.807	0.048
50	0.041	0.054	0.999

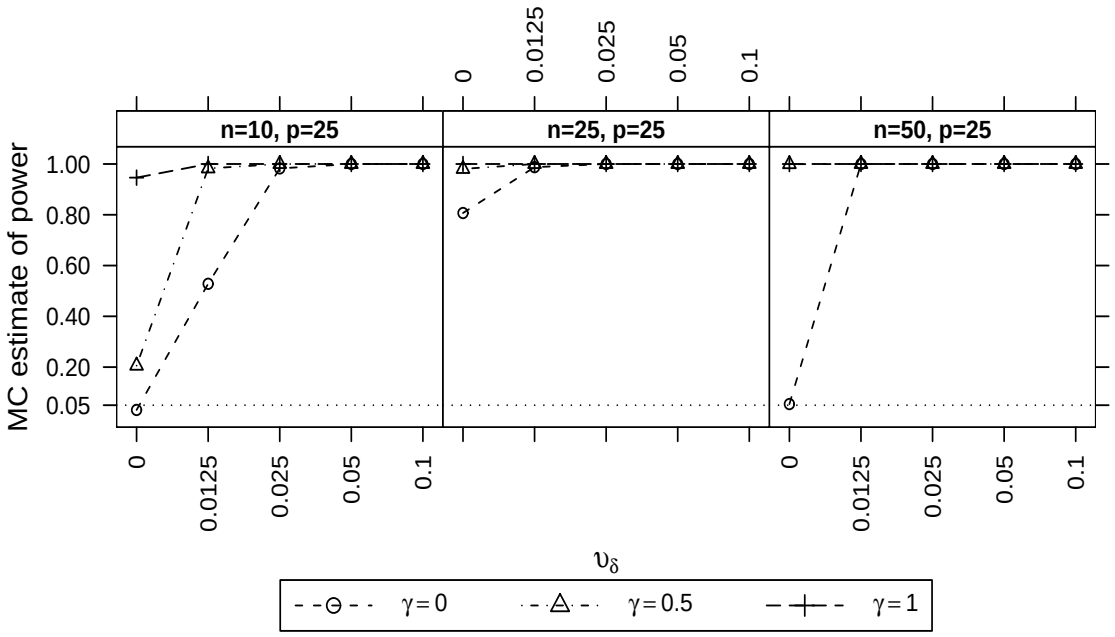

Figure 3.19: *Estimated power of the PCA test for* 0 *common factors when* $p = 25$ *and there was one nonlinear random effect in the O2C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

those presented in the figure except for a slight increase in power as p increased for cases where $\upsilon_\delta > 0$. The three panels in the figure show the effect of a different number of subjects n for fixed p. As n increased, power increased. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. When $\gamma = 0$ and $n < p$ or $n > p$, the test had approximately the correct size (0.05) when $\upsilon_\delta = 0$. Table 3.17 gives estimated size of the PCA test at $\upsilon_\delta = 0$ when $\gamma = 0$ for different values of n and p. As in the linear example, when $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on, and the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there were no principal components too often. When $\gamma = 0$ and $n < p$ or $n > p$, the test had greater than 80% power to detect a random effect with υ_δ as small as 0.0125.

Figure 3.20 presents the power curves for the MLFA-U test for the cases where $n > p$. The

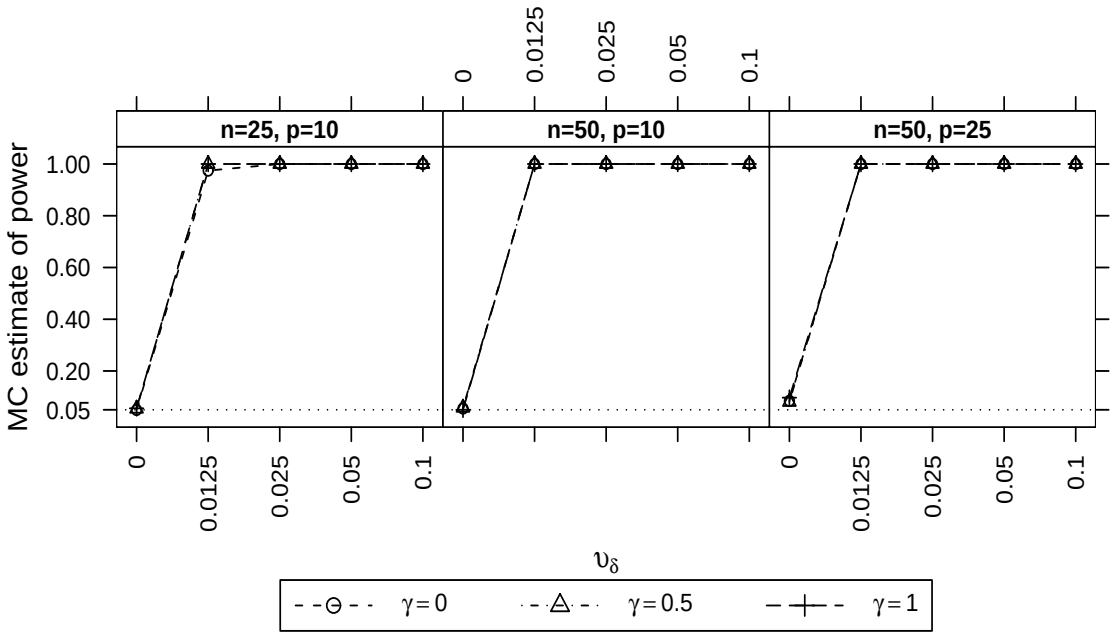

Figure 3.20: *Estimated power of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one nonlinear random effect in the O2C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

size of the test was close to the nominal 0.05 level except in the cases when $n = 50$ and $p = 25$ where the test rejected the hypothesis of no common factors in approximately 8% to 10% of replicates (see Table 3.18). In all cases where $n > p$, the test had greater than 90% power to

Table 3.18: *Estimated size of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one nonlinear random effect in the O2C model with* $\upsilon_\delta = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

		γ		
n	p	0.0	0.5	1.0
25	10	0.049	0.054	0.056
50	10	0.055	0.058	0.050
50	25	0.084	0.081	0.097

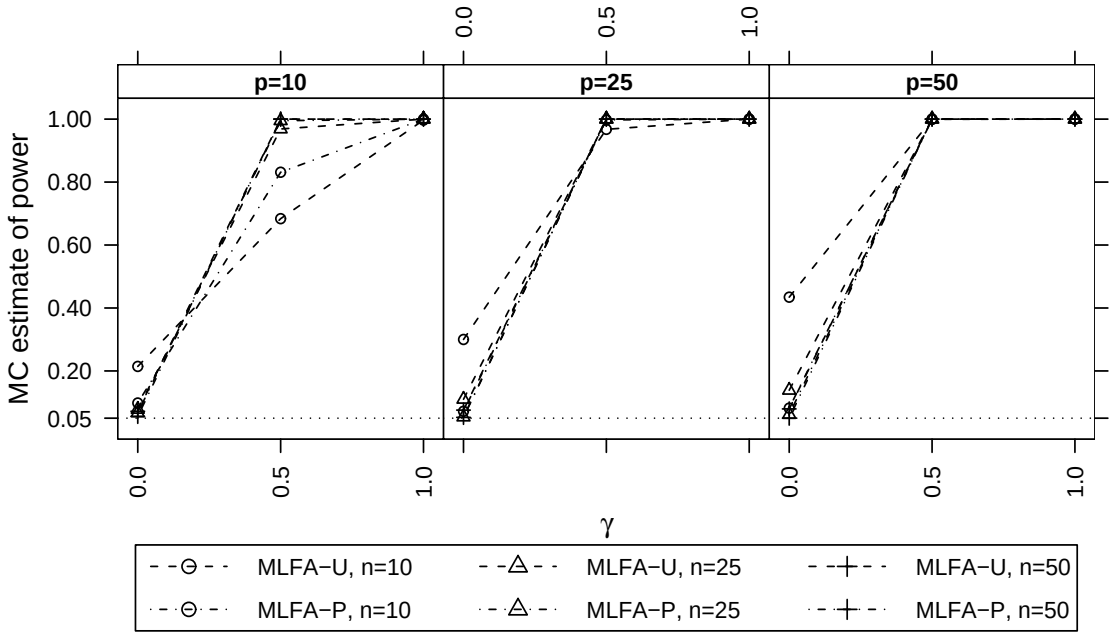

Figure 3.21: *Estimated power of the likelihood ratio tests for H_0: $\gamma = 0$ when there was one nonlinear random effect in the O2C model with $\upsilon_\delta = 0.025$. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

detect a random effect with υ_δ as small as 0.0125, and there was little effect of n, p or γ on power for this amount of random effect signal. When $n > p$ and the within-subject variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same power.

Test for homogeneous variance

For the likelihood ratio tests comparing the fit of the MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation of the null distribution to be adequate when n is large relative to p. More parameters ($p - 2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model.

Table 3.19: *Estimated size of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there was one nonlinear random effect in the O2C model. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. The table reports results for* $\upsilon_\delta = 0.0125$ *or* 0.1*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\upsilon_\delta = 0.0125$	$\upsilon_\delta = 0.1$	$\upsilon_\delta = 0.0125$	$\upsilon_\delta = 0.1$	$\upsilon_\delta = 0.0125$	$\upsilon_\delta = 0.1$
	MLFA-U					
10	0.231	0.262	0.352	0.359	0.432	0.524
25	0.100	0.348	0.091	0.439	0.129	0.541
50	0.084	0.591	0.084	0.695	0.083	0.805
	MLFA-P					
10	0.091	0.102	0.076	0.084	0.077	0.092
25	0.056	0.092	0.049	0.079	0.054	0.071
50	0.054	0.096	0.060	0.086	0.059	0.086

Power to reject the hypothesis of homoscedasticity should increase with increasing γ. As the magnitude of the nonlinear random effect increases, we expect that the one-factor models will fail to explain between-subject variation adequately. As a result, we expect the tests to reject the null hypothesis more often than the nominal 0.05 level for $\gamma = 0$ when υ_δ is large.

The number of time points (p), and two-way interactions between magnitude of γ, number of subjects (n), and random effect signal (υ_δ), explained 91% of variation in the power of the test comparing the MLFA-H to the MLFA-U model. The magnitude of γ explained 97% of the variation in the power of the test comparing the fit of the MLFA-H model with that of the MLFA-P model.

In all cases, the size of the test comparing the homoscedastic model to the unrestricted model MLFA-U was significantly larger than the nominal level of 0.05. Size of both tests increased in the presence of large random effect signal; this was especially noticeable for the

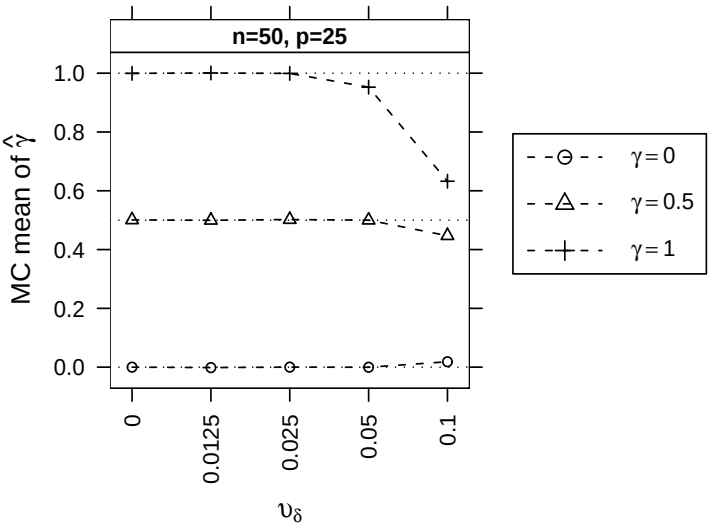

Figure 3.22: *Mean estimate of γ from fitting the one-factor MLFA-P model when there was one nonlinear random effect in the O2C model. The figure plots the case where $n = 50$ and $p = 25$. The true value of γ is given by the dotted horizontal lines. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.003*.*

MLFA-U test. Estimated size for both tests is presented in Table 3.19 for small random effect variation ($\upsilon_\delta = 0.0125$) and large random effect variation ($\upsilon_\delta = 0.1$). The test comparing the homoscedastic model to the power model MLFA-P performed better, although its size was larger than the nominal level when $n = 10$ and in the presence of a large nonlinear random effect. The magnitude of the random effect signal did not have a large impact on power (for $\gamma > 0$), so we have presented estimated power only for $\upsilon_\delta = 0.025$ in Figure 3.21. When $n > 10$ or $p > 10$ the MLFA-P test had greater than 90% power to detect heteroscedasticity when γ was as large as 0.5.

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the one-factor MLFA-P model. On average, the bias in $\widehat{\gamma}$ was approximately zero except

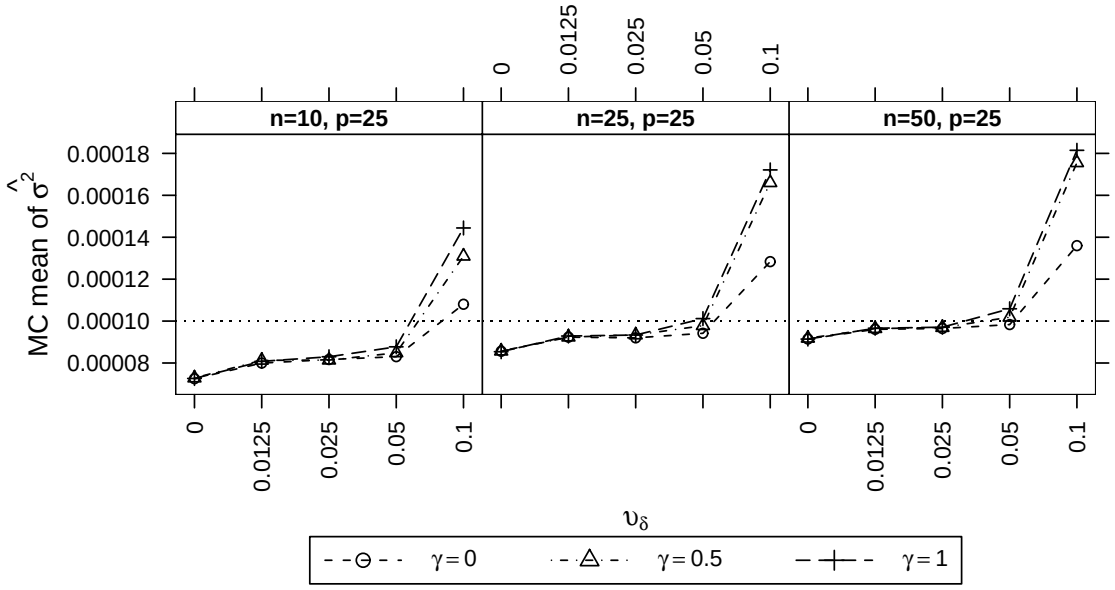

Figure 3.23: *Mean estimate of σ^2 from fitting the one-factor MLFA-P model when there was one nonlinear random effect in the O2C model and $p = 25$. The true value of σ^2 was* $1e-4$. *Based on analyzing the simulation data in batches of* 10, *we estimated the standard error of each plotted MC mean to be less than* $2e-6$.

when the nonlinear random effect was large ($\upsilon_\delta = 0.1$) and $\gamma = 1$. In this case the maximum likelihood estimator of γ was too small. The interaction between γ and random effect signal explained 82% of the variation in bias of $\widehat{\gamma}$ (all simulation factors and their interactions explained 85% of variation). Figure 3.22 displays the mean of $\widehat{\gamma}$ plotted against υ_δ for different values of γ in the case where $n = 50$ and $p = 25$.

The estimate of σ^2 was smaller on average than the true value of 0.0001 except when $\upsilon_\delta = 0.1$. In this case the estimate of σ^2 was substantially larger than the true value. When the nonlinear random effect was large, the MLFA-P model with one common factor did not adequately explain the between-subject variation. The interaction between magnitude of the random effect (υ_δ) and magnitude of γ, and the number of subjects (n) explained 82% of the variation in bias of $\widehat{\sigma}^2$ (all simulation factors and their interactions explained 88% of variation). Figure 3.23 illustrates the pattern in the estimate of σ^2 for $p = 25$. The pattern was similar for

other values of p. The three panels in the figure show the effect of increasing n for fixed p; the estimate of σ^2 increased with increasing n. The three curves in each panel represent the mean estimate of σ^2 for each setting of γ. Bias in $\widehat{\sigma}^2$ in the presence of the large nonlinear random effect was smallest when $\gamma = 0$.

Selecting random effects and an appropriate parameterization

We used the exact test where where the j-th element of $\boldsymbol{\kappa}$ was the sensitivity function of the actual random effect, evaluated at t_j,

$$\kappa_j \propto f^*_{\delta^{(2)}}(t_j) = -(1-\alpha)t_j \exp(-\delta^{(2)}t_j)\left\{\frac{\alpha}{1-\alpha}\exp(-\delta^{(1)}t_j) + 1\right\}.$$

We present results for the test based on the PCA analysis; results for the tests based on the estimated factor loading for the one-factor MLFA-U and MLFA-P models were virtually indistinguishable from those based on the first principal component loading. Since the test is based on the exact distribution of the first principal component loading under the null hypothesis, we expect the test to have the nominal size (0.05) when $\upsilon_\delta = 0$ and $\gamma = 0$. The power should increase for increasing signal $\upsilon_\delta > 0$. Sample size (n) and number of time points (p) should have no effect on size of the test when $\gamma = 0$, but may affect power when the alternative hypothesis

Table 3.20: *Estimated size of the PCA exact test for a random effect on $\delta^{(2)}$ when there was one nonlinear random effect in the O2C model with $\upsilon_\delta = 0$. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	0.049	0.016	0.002	0.051	0.007	0.001	0.049	0.013	0.001
25	0.046	0.005	0.000	0.048	0.005	0.000	0.061	0.003	0.000
50	0.051	0.002	0.000	0.043	0.003	0.000	0.058	0.001	0.000

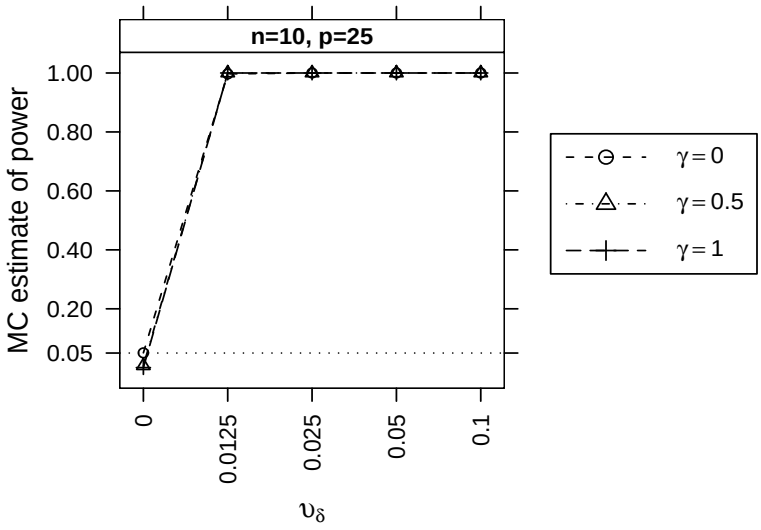

Figure 3.24: *Estimated power of the PCA exact test when there was one nonlinear random effect in the O2C model and* $n = 10, p = 25$*. This test rejects* H_0*:* $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ *in favor of one random effect with sensitivity function* $f^*_{\delta^{(2)}}$ *. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

is true. When $\gamma > 0$, the principal component model that assumes homogeneous within-subject variance is incorrect, and we would expect size and power of the test to deviate somewhat from those in the case where $\gamma = 0$.

The interaction between magnitude of the random effect (υ_δ) and magnitude of γ explained over 99% of the variation in power of the PCA exact test. Table 3.20 gives the size of the test under the null hypothesis. When $\gamma = 0$ (homoscedastic case) the size was not significantly different from the nominal 0.05 level regardless of n and p. However when $\gamma > 0$, the size was smaller than 0.05. This is different from the simulation with one linear random effect, but since $\upsilon_\delta = 0$ for cases in the table, this difference cannot be explained by the nonlinear random effect, but rather the different sensitivity functions used in the linear and nonlinear examples.

In Figure 3.24 we present estimated power of the PCA exact test to detect a random effect with sensitivity function $f^*_{\delta^{(2)}}$ (the true random effect). We only present power for the case where $n = 10$ and $p = 25$ since n and p did not have have much effect on the power of the

exact test for the levels of υ_δ examined. Power increased rapidly for increasing random effect magnitude. In all cases the exact test had greater than 90% power to detect the true random effect when the magnitude of random effect variation was given by $\upsilon_\delta = 0.0125$. The three curves in the figure show power for different amounts of within-subject heteroscedasticity, γ. As seen in Table 3.20, the effect of increasing γ was decreased size at $\upsilon_\delta = 0$.

We used multivariate forward selection, with the first PCA or factor loading as response variable, to identify the first- and second-order sensitivity functions corresponding to the correct nonlinear random effect from amongst several possible predictors. For parameterization f^* we considered the following sensitivity functions, including a second-order sensitivity for the parameter $\delta^{(2)}$,

$$\begin{aligned}
f^*_\alpha(t) &= \exp\{-(\delta^{(1)} + \delta^{(2)})t\} - \exp(-\delta^{(2)}t) \\
f^*_{\delta^{(1)}}(t) &= -\alpha t \exp\{-(\delta^{(1)} + \delta^{(2)})t\} \\
f^*_{\delta^{(2)}}(t) &= -(1-\alpha)t\exp(-\delta^{(2)}t)\left\{\frac{\alpha}{1-\alpha}\exp(-\delta^{(1)}t) + 1\right\} \\
f^*_{\delta^{(2)}\delta^{(2)}}(t) &= \alpha t^2 \exp\{-(\delta^{(1)} + \delta^{(2)})t\} + (1-\alpha)t^2\exp(-\delta^{(2)}t).
\end{aligned}$$

When $\upsilon_\delta > 0$, we would expect the forward selection to pick a model including $f^*_{\delta^{(2)}}$. As υ_δ becomes large, we would expect the second-order sensitivity $f^*_{\delta^{(2)}\delta^{(2)}}$ to be included too.

The random effect on $\delta^{(2)}$ induced two correlated random effects on parameters $\omega^{(1)}$ and $\omega^{(2)}$ in the alternative parameterization f of equation (3.2). We also used multivariate regression to select sensitivities for the parameterization f. We included the following sensitivities and

second order sensitivities as candidates for the parameterization f,

$$\begin{aligned} f_{\alpha}(t) &= \exp(-\omega^{(1)}t) - \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}}(t) &= -\alpha t \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}}(t) &= -(1-\alpha) t \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}\omega^{(1)}}(t) &= \alpha t^2 \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}\omega^{(2)}}(t) &= (1-\alpha) t^2 \exp(-\omega^{(2)}t). \end{aligned}$$

We did not include the second-order sensitivity $f_{\omega^{(1)}\omega^{(2)}}$ which is the zero function. For $\upsilon_{\delta} > 0$ we would expect the multivariate regression model to include the two sensitivities $f_{\omega^{(1)}}$ and $f_{\omega^{(2)}}$. As υ_{δ} becomes large, the second-order sensitivities $f_{\omega^{(1)}\omega^{(1)}}$ and $f_{\omega^{(2)}\omega^{(2)}}$ may also be included. We distinguished between the appropriateness of the two parameterizations based on the number of sensitivities selected for each, and the trace correlation r_T of the selected models. The more appropriate parameterization should have give a smaller multivariate regression model, and/or larger trace correlation. We expect that a smaller model should be selected for parameterization f^* than for parameterization f, indicating that the true data generating model was based on f^*.

The multivariate regression results based on the PCA, the MLFA-U and the MLFA-P models were very similar. We present results only for the unrestricted MLFA-U model. On average the trace correlation of the selected model did not differ for the parameterizations f^* and f. The main effects of the magnitude of the random effect (υ_{δ}), the magnitude of γ, number of subjects (n), and number of time points (p) explained 94% of the variation in the mean number of sensitivities selected for the true parameterization, f^*, and 95% of the variation in the mean number selected for the alternative one, f. For the sake of brevity, we present results only for the cases where $\gamma = 0$ or $\gamma = 1$. Figure 3.25 summarizes the mean number of sensitivities included in the multivariate regression models for parameterizations f^* and f. The plots on

the lower row are for $\gamma = 0$, with $\gamma = 1$ on the upper row. As was the case with one linear random effect, a larger model was selected in the presence of heteroscedasticity than when within-subject variation was homoscedastic. The effect of n, the number of subjects, is shown

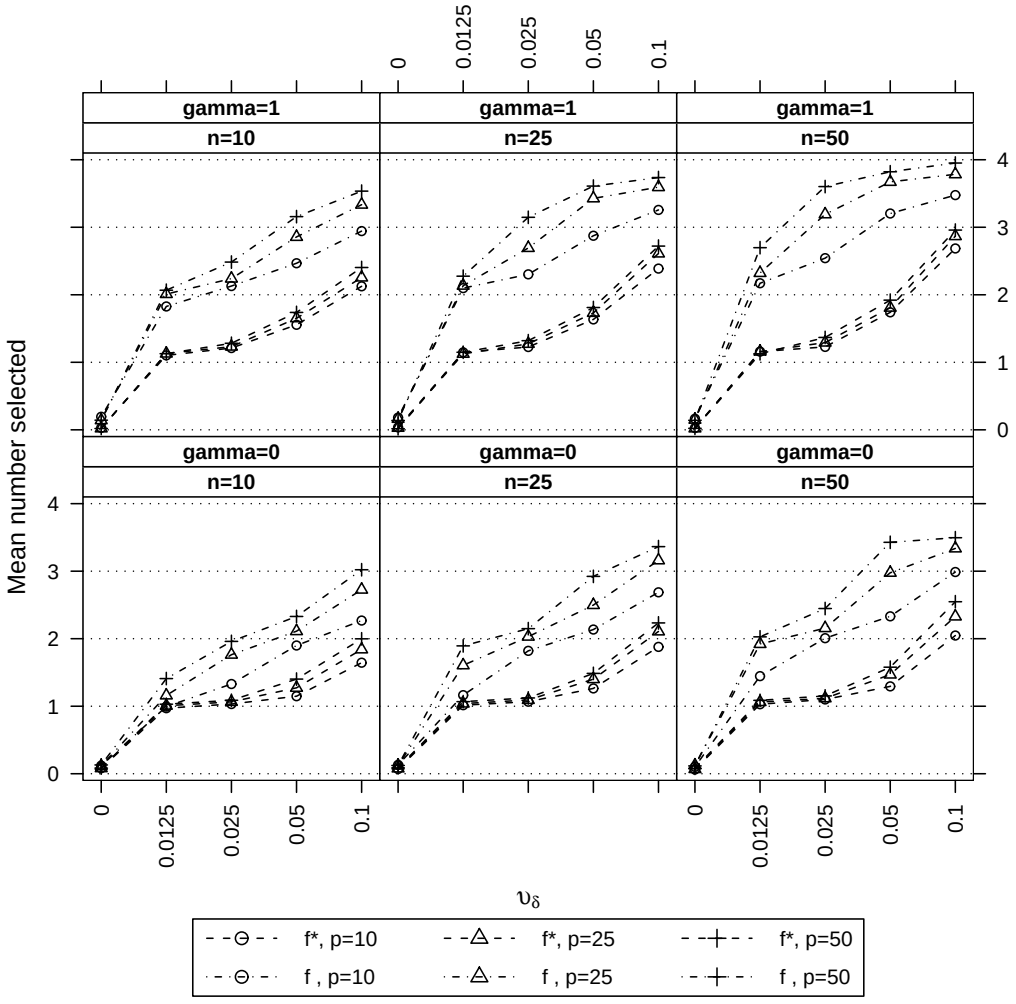

Figure 3.25: *Mean number of sensitivities selected by multivariate forward selection when there was one nonlinear random effect in the O2C model. Results presented are for the parameterizations f^* (true) and f (alternate), and the regression using the factor loading estimated from the one-factor MLFA-U model where $\gamma = 0$ or $\gamma = 1$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.02.

by the three panels in each row of the figure. As n increased, the average model size increased.

Within each panel the dashed curves show the mean size of the selected model for the true parameterization f^*. The dot-dash curves show the mean model size for the alternative parameterization. On average, for $\upsilon_\delta > 0$, fewer sensitivities were selected for the true parameterization f^* than for the alternative parameterization f. The different plotting symbols in each panel show the effect of increasing p, the number of time points. As p increased, the average size of the selected model increased. When the nonlinear random effect was small, forward selection picked a single sensitivity for the true parameterization f^*, but picked a larger model for parameterization f, especially when p, n or γ were large. As the magnitude of the random effect increased, a second or third sensitivity was included for f^*, and additional sensitivities were also included for the parameterization f.

Table 3.21 summarizes the mean number of sensitivities selected for the true parameterization f^* when there was no random effect ($\upsilon_\delta = 0$). When $\gamma = 0$, multivariate forward selection picked a model with one or more sensitivities slightly more often than might be expected by the 0.05 nominal level of the tests at each stage of the forward selection. In the presence of heteroscedasticity ($\gamma = 1$), one or more sensitivities were selected less often than we might

Table 3.21: *Mean number of sensitivities included for the true parameterization f^* by forward selection for the experiment with one nonlinear random effect in the O2C model when $\upsilon_\delta = 0$. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $\gamma = 0$ or $\gamma = 1$. At each step of the forward selection the 0.05 critical value of the test statistic was used. Based on analyzing the simulation data in batches of 10, we estimated the standard error of each MC mean to be less than 0.015.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.082	0.029	0.082	0.021	0.083	0.019
25	0.070	0.036	0.078	0.029	0.087	0.016
50	0.063	0.027	0.069	0.021	0.080	0.022

Table 3.22: *Proportion of cases where forward selection picked* $f^*_{\delta^{(2)}}$ *when* $\upsilon_\delta = 0$ *in the experiment with one nonlinear random effect in the O2C model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.012	0.001	0.009	0.003	0.012	0.003
25	0.013	0.004	0.012	0.003	0.016	0.003
50	0.006	0.006	0.012	0.002	0.011	0.004

expect.

Next we examine which sensitivities from the true parameterization f^* were selected. The main effects of the simulation factors υ_δ, γ, n and p explained 96% of variation in the proportion of times the sensitivity $f^*_{\delta^{(2)}}$ was selected, and 84% of variation in the proportion of times $f^*_{\delta^{(2)}\delta^{(2)}}$ was selected. These sensitivities correspond to the true random effect on $\delta^{(2)}$. Figure 3.26 plots the proportion of cases where multivariate forward selection picked these sensitivities. The figure only shows the case where $p = 25$. Results for other values of p were similar, but with higher proportion selected for larger values of p. When there was no random effect ($\upsilon_\delta = 0$), the sensitivity $f_{\delta^{(2)}}$ was selected in about 1% of cases when within-subject variation was homoscedastic ($\gamma = 0$) (see Table 3.22), and in less than 0.5% of cases in the presence of heteroscedasticity ($\gamma = 1$).

As the random effect signal υ_δ increased, the first-order sensitivity $f^*_{\delta^{(2)}}$ was included in the model in an increasing proportion of cases (dashed curves). The second-order sensitivity $f^*_{\delta^{(2)}\delta^{(2)}}$ (dot-dash curve) was included when the random effect magnitude became large. This confirms that the factor analysis detected a common factor related to the square of the random effect on $\delta^{(2)}$ when this random effect was large. The three panels in the figure show the effect of

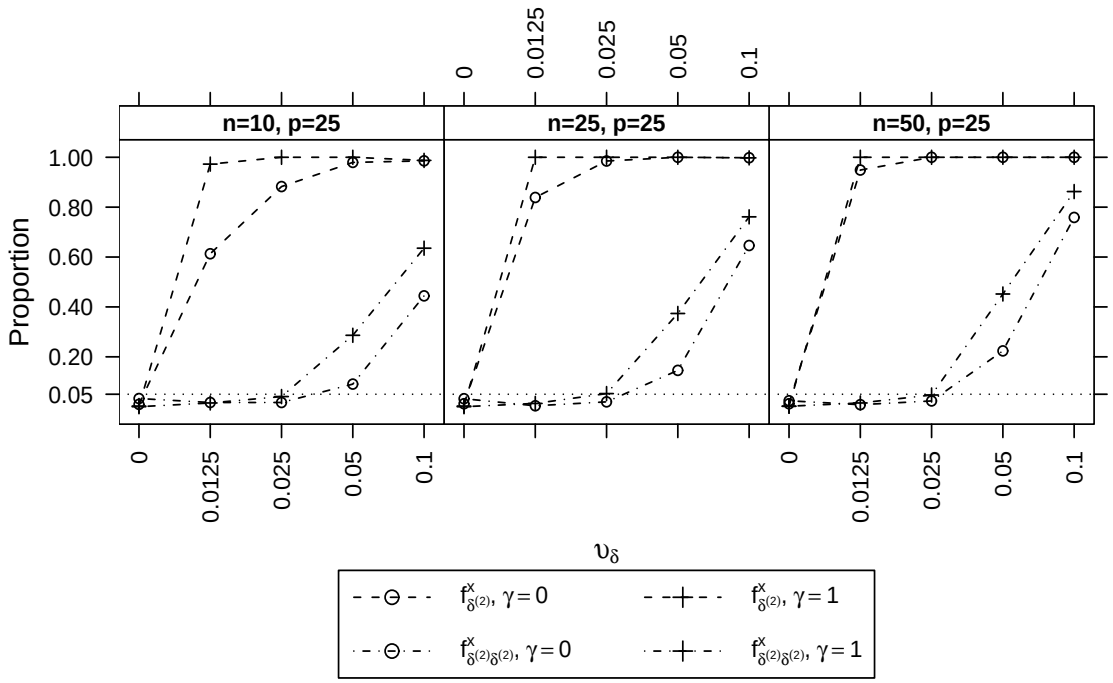

Figure 3.26: *Proportion of cases where multivariate forward selection picked sensitivities corresponding to a random effect on* $\delta^{(2)}$ *in the O2C model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0$ *or* $\gamma = 1$ *and* $p = 25$*. The dashed curves plot the proportion of times the sensitivity* $f^*_{\delta^{(2)}}$ *was selected. The dot-dash curves plot the proportion of times the second-order sensitivity* $f^*_{\delta^{(2)}\delta^{(2)}}$ *was selected. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125*.*

increasing n; the selection proportions increased when n was larger. Different plotting symbols represent different levels of γ. The sensitivity functions were selected in higher proportions in the presence of heteroscedasticity than when within-subject variation was homoscedastic.

3.5.2 Two random effects (one linear and one nonlinear)

In this experiment there were two independent random effects for parameterization f of equation (3.2). The first random effect on the linear scale parameter α had large standard deviation, $\upsilon_\alpha|\alpha|$ where $\upsilon_\alpha = 0.4$. The second random effect on the nonlinear rate parameter $\omega^{(2)}$ had stan-

dard deviation $\upsilon_\omega|\omega^{(2)}|$ where we varied υ_ω from zero to 0.4. The convergence rate for maximum likelihood estimation of the two-factor MLFA-U model was above 95% except when $\gamma = 1$ and $n = p = 10$. In this case the convergence rate was between 85% and 90% with convergence failing more often when $\upsilon_\omega = 0$. The convergence rate for the two-factor MLFA-P model was 100%.

Since the large random effect on α was linear, we expect results for this experiment to be similar to the case of two linear random effects when υ_ω is small, but to play out a little differently when υ_ω is large. As in the experiment with one nonlinear random effect, when the nonlinear random effect becomes large, the principal component and factor analyses may detect additional common factors due to the need for second order terms in the approximation to the stage-1 model. We also expect that the multivariate regression analysis will select second-order sensitivities when υ_ω becomes large.

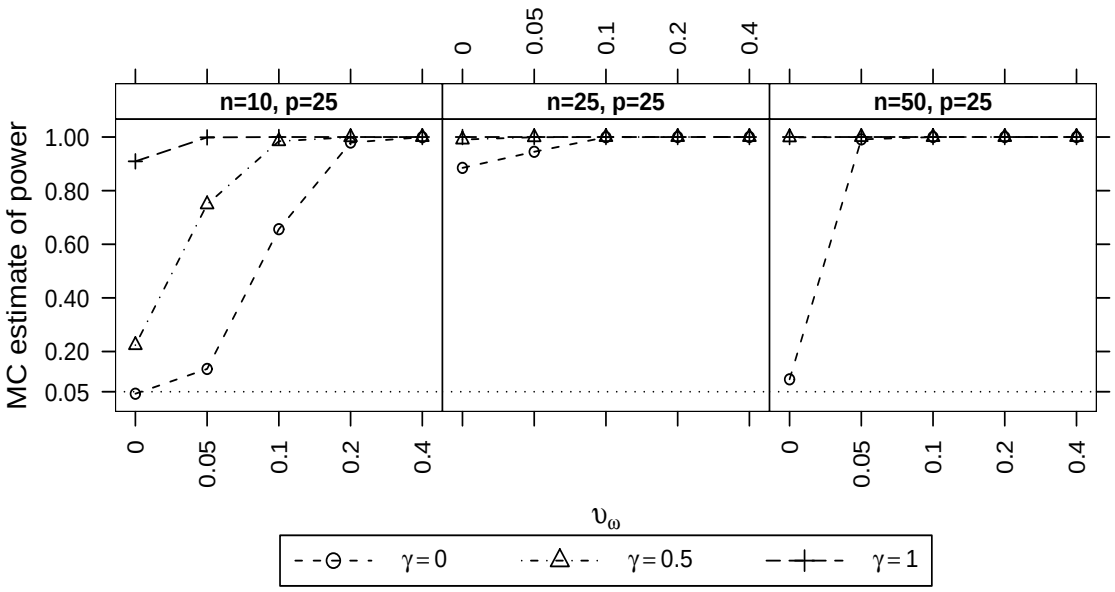

Figure 3.27: *Estimated power of the PCA test for* 1 *common factor when* $p = 25$ *and there were two random effects in the O2C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Number of random effects

When $n > p$ or when $n < p$ (for the PCA test), we expect the power of the likelihood ratio tests to reject the hypothesis of one common factor to increase with υ_ω and, since the first random effect was linear, to have size close to the nominal level of 0.05 when $\upsilon_\omega = 0$ (one random effect). When $n \approx p$, asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we would expect the PCA test to reject the null hypothesis even when there was only one common factor.

In the batched ANOVA of simulation results, the magnitude of γ, magnitude of the second random effect (υ_ω), the number of subjects (n), and two-way interactions between these factors explained 72% of the variation in power of the PCA test. The magnitude of the second random effect explained 95% of variation in power of the MLFA-U test.

Figure 3.27 plots estimated power of the PCA test against the magnitude of the second random effect signal (υ_ω) when $p = 25$. Results for $p = 10$ and $p = 50$ were similar to those presented in the figure except for a slight increase in power as p increased for cases where $\upsilon_\omega > 0$. The three panels in the figure show the effect of a different number of subjects n for fixed p. As n increased, power increased. The three curves in each panel represent power for different

Table 3.23: *Estimated size of the PCA test for* 1 *common factor when* $\gamma = 0$ *and there were two random effects in the O2C model, the second one with* $\upsilon_\omega = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

	p		
n	10	25	50
10	0.296	0.043	0.047
25	0.061	0.885	0.062
50	0.053	0.096	1.000

amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. Table 3.23 gives estimated size of the PCA test at $\upsilon_\omega = 0$ when $\gamma = 0$ for different values of n and p. When n was much larger than p, or p was much larger than n, the test had approximately the correct size (0.05) when $\gamma = 0$. When the ratio of n to p was closer to one, the size was larger than the nominal value. When $n = p$ (center panel in the figure and diagonal in the table) the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there was one principal component too often and would not be useful in the presence of heteroscedastic within-subject variance. When $\gamma = 0$ and $n < p$ or $n > p$, the PCA test had greater than 80% power to detect a random effect with $\upsilon_\omega = 0.2$.

Figure 3.28 presents the power curves for the MLFA-U test for the cases where $n > p$. The size of the test was close to the nominal 0.05 level when $p = 10$ and $n = 25$ or 50. When $n = 50$ and $p = 25$, the size of the test was larger than the nominal level (see Table 3.24). In all cases where $n > p$, the MLFA-U test had greater than 90% power to detect a second random effect with $\upsilon_\omega = 0.1$ or larger. For $\upsilon_\omega > 0$, power increased with increasing n and p. Power to reject the hypothesis of one common factor increased with increasing γ. When $n > p$ and the within-subject variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same power.

Table 3.24: *Estimated size of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two random effects in the O2C model, the second one with* $\upsilon_\omega = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

		γ		
n	p	0.0	0.5	1.0
25	10	0.054	0.053	0.061
50	10	0.048	0.049	0.051
50	25	0.093	0.104	0.086

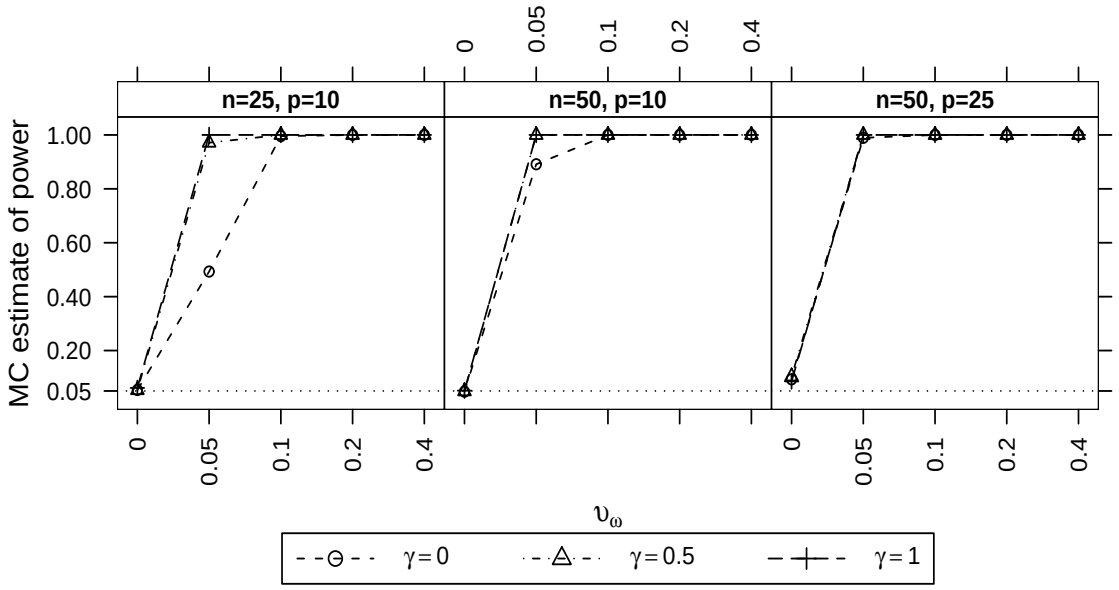

Figure 3.28: *Estimated power of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two random effects in the O2C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Test for homogeneous variance

For the likelihood ratio tests comparing the fit of the MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation of the null distribution to be adequate when n is large relative to p. More parameters ($p - 2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ.

The magnitude of γ, n, p, and the interaction between n and γ explained 87% of the variation in power to reject the hypothesis of homoscedasticity by comparing the MLFA-H model with the MLFA-U model. For the test comparing the MLFA-H and the MLFA-P models, the magnitude of γ explained 95% of variation in power. Table 3.25 presents size of the tests for small random effect ($\upsilon_\omega = 0.05$) and large random effect ($\upsilon_\omega = 0.4$).

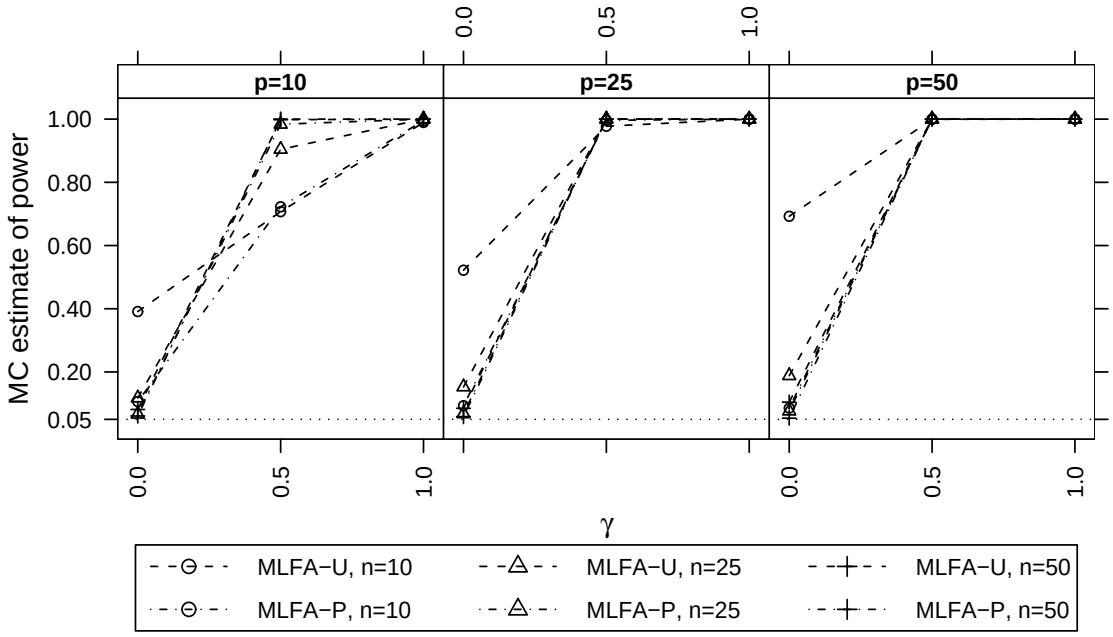

Figure 3.29: *Estimated power of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two random effects in the O2C model and* $v_\omega = 0.1$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

In all cases, the size of the test comparing the homoscedastic model to the unrestricted model MLFA-U was significantly larger than the nominal level of 0.05 (see Table 3.25), making this test close to useless. Size was increased by a large random effect signal. The test comparing the homoscedastic model to the power model MLFA-P performed better, although its size was significantly larger than the nominal level when there was a large random effect. It is possible that the size of the MLFA-P test would have been closer to the nominal level if we had fitted a factor model with more than two factors when the random effect variation was large. In the presence of a small random effect, the size of the test approached the nominal 0.05 level as n increased.

Since magnitude of the second random effect signal did not have a large impact on power

Table 3.25: *Estimated size of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two nonlinear random effects in the O2C model. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. The table reports results for* $\upsilon_\omega = 0.05$ *or* 0.4*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\upsilon_\omega = 0.05$	$\upsilon_\omega = 0.4$	$\upsilon_\omega = 0.05$	$\upsilon_\omega = 0.4$	$\upsilon_\omega = 0.05$	$\upsilon_\omega = 0.4$
	MLFA-U					
10	0.439	0.409	0.640	0.578	0.777	0.742
25	0.183	0.354	0.191	0.374	0.223	0.494
50	0.080	0.544	0.089	0.580	0.098	0.658
	MLFA-P					
10	0.141	0.166	0.109	0.159	0.102	0.159
25	0.089	0.203	0.059	0.203	0.071	0.236
50	0.059	0.342	0.058	0.339	0.060	0.403

for $\gamma > 0$, we have presented results only for $\upsilon_\omega = 0.1$ in Figure 3.29. When $n > 10$ or $p > 10$ the test based on the MLFA-P model had greater than 90% power to detect heteroscedasticity when γ was as large as 0.5.

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the two-factor MLFA-P model. On average, the bias in $\widehat{\gamma}$ was approximately zero except when the nonlinear random effect was large ($\upsilon_\omega \geq 0.2$). In these cases the maximum likelihood estimator of γ was too large. The interaction between the magnitude of γ and υ_ω explained 78% of variation in bias of $\widehat{\gamma}$ (all simulation factors and their interactions explained 84% of variation). Figure 3.30 plots the mean of $\widehat{\gamma}$ against υ_ω for values of γ in the case where $n = 50$ and $p = 25$. Bias in $\widehat{\gamma}$ was slightly larger for larger n and smaller p.

The estimate of σ^2 was smaller on average than the true value of 0.0001 except when $\upsilon_\omega = 0.4$. In this case the estimate of σ^2 was substantially larger than the true value. When the nonlinear random effect was large, two factors in the MLFA-P model did not adequately explain the between-subject variation due to the large nonlinear random effect. The magnitude of the random effect (υ_ω), γ, the number of time points (p), and two-way interactions between these factors had the largest effect, explaining 36% of variation in bias of $\widehat{\gamma}$ (all simulation factors and their interactions explained 40% of variation). Figure 3.31 illustrates the pattern in the estimate of σ^2 for $n = 25$ (note the log scale on the y-axis). The three panels in the figure show the effect of increasing p for fixed n; bias when $\upsilon_\omega = 0.4$ decreased with increasing p. The three curves in each panel represent the mean estimate of σ^2 for each setting of γ. Bias in $\widehat{\sigma}^2$ in the presence of the large nonlinear random effect was smallest when $\gamma = 0$.

The over-estimation of σ^2 and γ in the case of a large nonlinear random effect suggests that the two-factor model did not adequately explain between-subject variation when υ_ω was large.

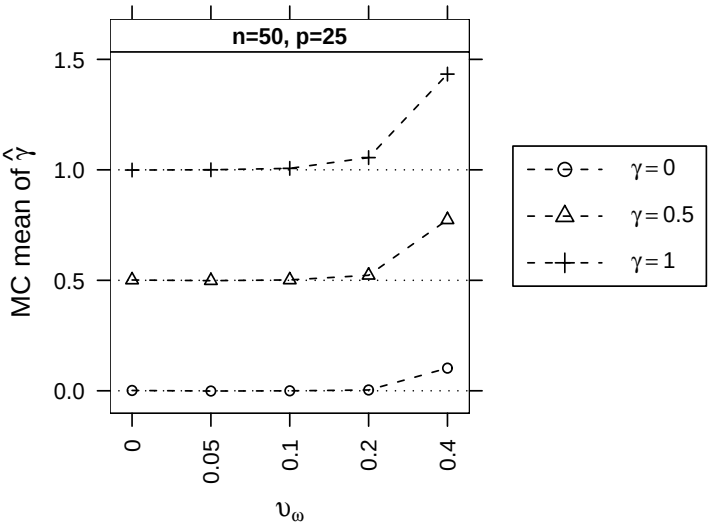

Figure 3.30: *Mean estimate of γ from fitting the two-factor MLFA-P model when there were two random effects in the O2C model. The figure plots the case where $n = 50$ and $p = 25$. The true value of γ is given by the dotted horizontal lines. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.007*.*

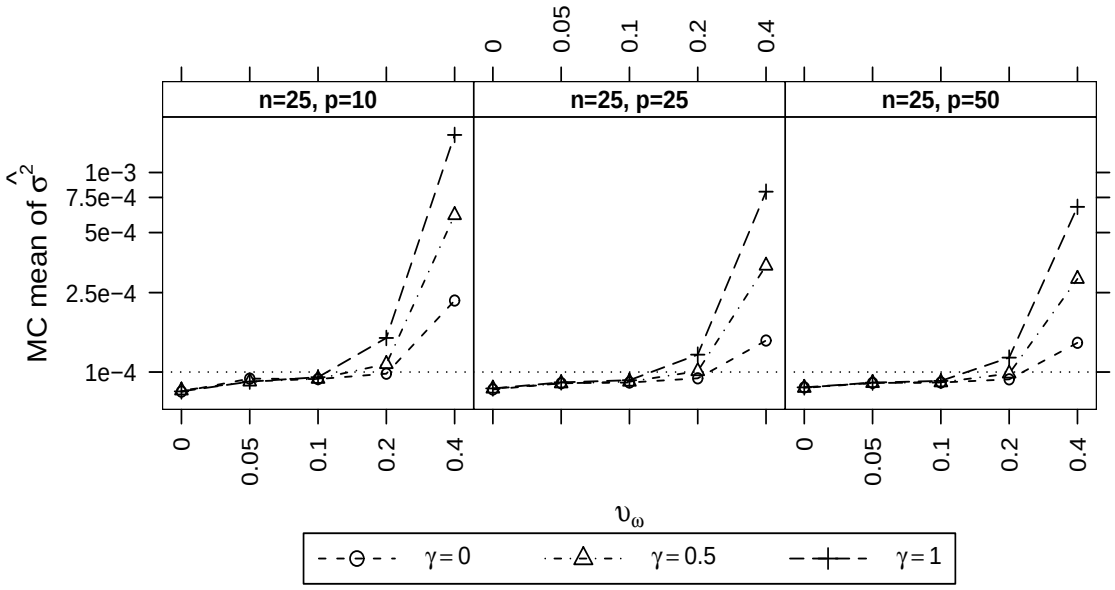

Figure 3.31: *Mean estimate of* σ^2 *from fitting the two-factor MLFA-P model when there were two random effects in the O2C model and* $n = 25$*. The true value of* σ^2 *was* $1e-4$*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* $1e-6$*.*

A model with three or more factors (to account for the second-order terms in the approximation of the stage-1 model) may have given a better fit.

Selecting random effects

We used multivariate forward selection, with the first two PCA or factor loadings as response variables, to identify the first- and second-order sensitivity functions corresponding to the correct random effects from amongst several possible predictors. For parameterization f we considered the following sensitivity functions, including a second-order sensitivity for the parameter

$\omega^{(2)}$,

$$\begin{aligned} f_{\alpha}(t) &= \exp(-\omega^{(1)}t) - \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}}(t) &= -\alpha t \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}}(t) &= -(1-\alpha)t \exp(-\omega^{(2)}t) \\ f_{\omega^{(2)}\omega^{(2)}}(t) &= (1-\alpha)t^2 \exp(-\omega^{(2)}t). \end{aligned}$$

The second-order sensitivity $f_{\alpha\omega^{(2)}}$ is proportional to $f_{\omega^{(2)}}$ and so we did not include this as a candidate predictor. We also did not include the second-order sensitivity $f_{\omega^{(1)}\omega^{(1)}}$ as a candidate because there was not a good reason to expect this to be selected, except because of non-orthogonality of the different sensitivities.

When $\upsilon_\omega = 0$ the forward selection should select a model with one sensitivity, f_α, corresponding to the linear random effect on α. As υ_ω gets larger, the forward selection should additionally select the sensitivity $f_{\omega^{(2)}}$ in a high proportion of cases. As mentioned in the previous sections, when the nonlinear random effect was large, the two-factor model may not adequately capture the between-subject variation in the data because of second-order terms in the approximation to the stage-1 model. However, we would still expect the forward selection procedure based on two common factors to pick the second-order sensitivity $f_{\omega^{(2)}\omega^{(2)}}$ in an increasing proportion of cases as υ_ω becomes large. As in the linear case, we would expect the power to select the correct sensitivity to increase with the number of time points (p), and with the number of subjects (n).

The multivariate regression results based on the PCA, MLFA-U and the MLFA-P models were very similar, and for the sake of brevity we present results only for the MLFA-U analysis. The main effects of the four simulation factors (υ_ω, γ, p, and n) explained 93% of variation in the mean number of sensitivities selected. Figure 3.32 summarizes the mean number of sensitivities included in the multivariate regression model when $n = 25$. When $\upsilon_\omega = 0$, a

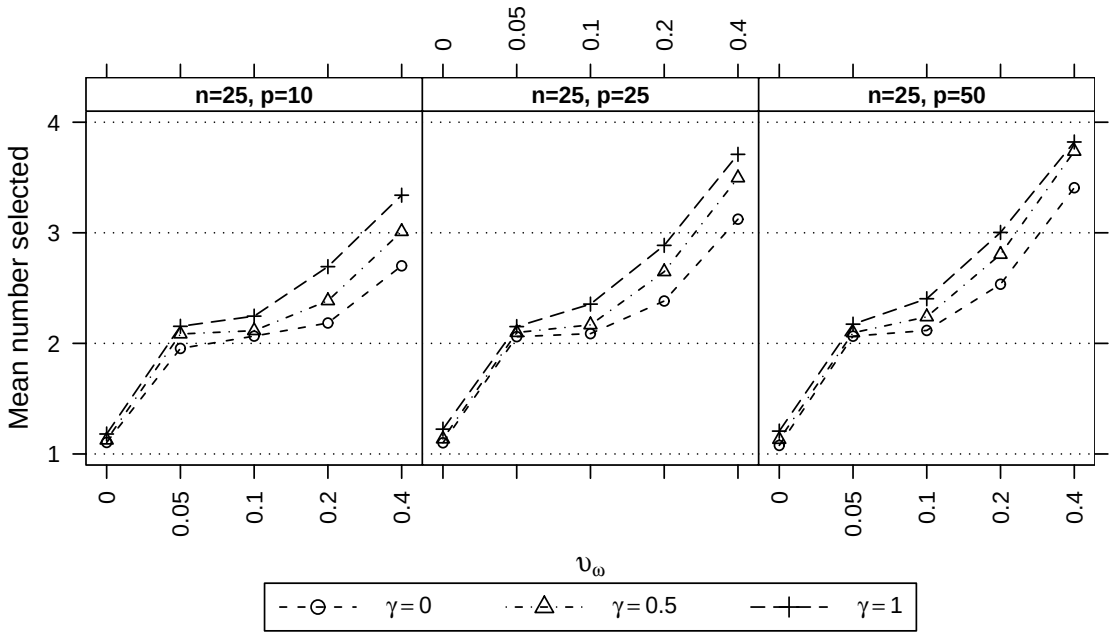

Figure 3.32: *Mean number of sensitivities selected by multivariate forward selection when there were two random effects in the O2C model. Results presented are for the regression using the factor loadings estimated from the two-factor MLFA-U model where* $n = 25$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.015*.*

model with one sensitivity was selected in most cases (see Table 3.26). As the magnitude of the random effect on $\omega^{(2)}$ increased, a larger model with two sensitivities was selected on average. When the nonlinear random effect became large ($\upsilon_\omega \geq 0.2$) the procedure selected a model with three or more sensitivities in an increasing number of cases. The three panels in the figure show the effect of increasing p for fixed n; average model size increased with increasing p. A slightly larger model was also selected for larger n. The three curves in each panel show the effect of within-subject heteroscedasticity. A larger model was selected in the presence of heteroscedasticity than when within-subject variation was homoscedastic.

The sensitivity f_α corresponding to the linear random effect on α was included in the multivariate regression model in more than 97% of cases. The main effects of υ_ω, γ, and the

Table 3.26: *Mean number of sensitivities included by forward selection when there were two random effects in the O2C model with* $\upsilon_\omega = 0$. *Results presented are for the regression using the factor loading estimated from the two-factor MLFA-U model. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each MC mean to be less than* 0.015.

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	1.118	1.140	1.223	1.108	1.133	1.219	1.100	1.114	1.183
25	1.104	1.126	1.181	1.101	1.135	1.224	1.077	1.131	1.207
50	1.093	1.126	1.179	1.094	1.124	1.194	1.086	1.123	1.205

number of time points (p) explained 96% of variation in the proportion of cases where the sensitivity $f_{\omega^{(2)}}$ was selected, and 73% of variation in the proportion of cases where $f_{\omega^{(2)}\omega^{(2)}}$ was selected. Figure 3.33 displays the proportion of cases where multivariate forward selection picked the sensitivity f_α corresponding to the first random effect, and $f_{\omega^{(2)}}$ and $f_{\omega^{(2)}\omega^{(2)}}$, the first- and second-order sensitivities corresponding to the nonlinear random effect on $\omega^{(2)}$. We have shown results only for the case where $n = 25$. Results were similar for other values of n, with a slight increase in selection proportion when n was larger. When there was only one

Table 3.27: *Proportion of cases where forward selection picked* $f_{\omega^{(2)}}$ *when there were two random effects in the O2C model with* $\upsilon_\omega = 0$. *Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model with* $\gamma = 0$ *and* $\gamma = 1$. *At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005.

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.043	0.161	0.043	0.150	0.031	0.132
25	0.043	0.136	0.036	0.143	0.026	0.149
50	0.031	0.136	0.029	0.136	0.031	0.134

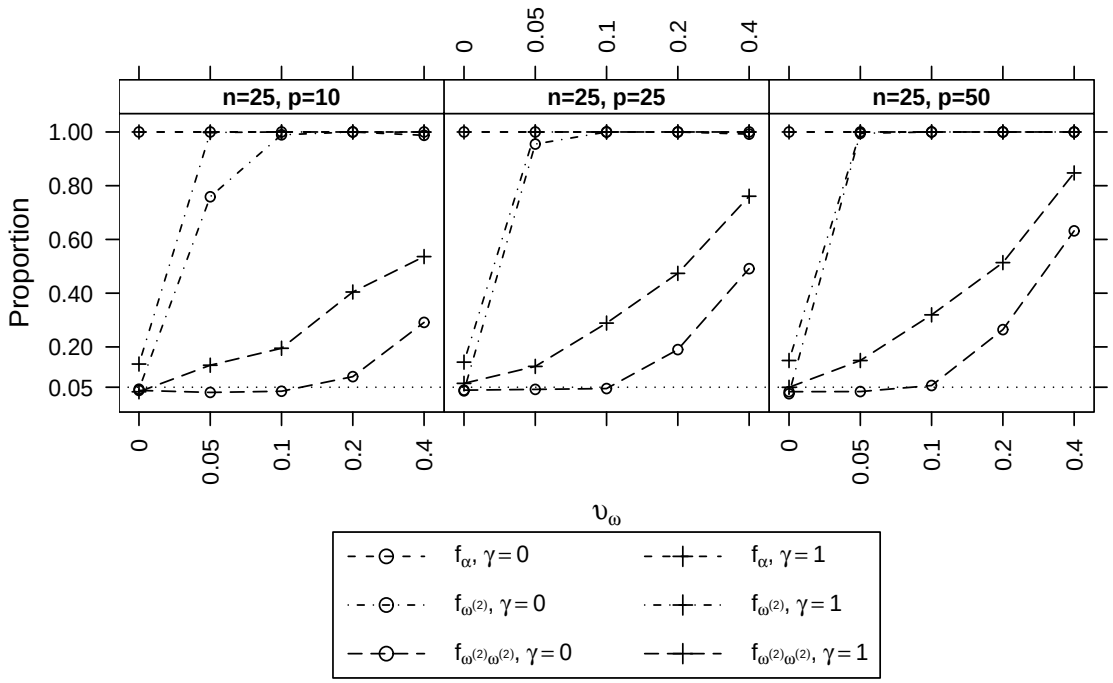

Figure 3.33: *Proportion of cases where multivariate forward selection picked the correct sensitivities, f_α (dashed curves), $f_{\omega^{(2)}}$ (dot-dashed curves), and $f_{\omega^{(2)}\omega^{(2)}}$ (long-dash curves), when there were two random effects in the O2C model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $n = 25$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125.

random effect ($\upsilon_\omega = 0$), the sensitivity $f_{\omega^{(2)}}$ was selected in less than 5% of cases when within-subject variation was homoscedastic ($\gamma = 0$), and in approximately 15% of cases in the presence of heteroscedasticity ($\gamma = 1$) — see Table 3.27.

When the second random effect signal υ_ω was positive, the sensitivity $f_{\omega^{(2)}}$ was included in the model in a high proportion of cases (dot-dash curves). As υ_ω increased, the second-order sensitivity $f_{\omega^{(2)}\omega^{(2)}}$ was included in the model with increasing frequency (long-dash curves). The three panels in the figure show the effect of increasing p; both sensitivities corresponding to the second random effect were picked more often when p was larger than when p was smaller. In each panel the different plotting symbols represent different levels of γ. The sensitivities $f_{\omega^{(2)}}$

and $f_{\omega^{(2)}\omega^{(2)}}$ were selected more often when $\gamma = 1$ than when $\gamma = 0$.

3.6 Simulation results: example C

The first simulation experiment of example C had a single nonlinear random effect on the volume parameter V_d in parameterization f of equation (3.4). The standard deviation of this random effect was $\upsilon_V|V_d|$ where we varied υ_V from zero to 0.2. The second experiment generated data with two independent nonlinear random effects, one on the volume parameter V_d, and the second on the clearance parameter, Cl. The first random effect on V_d had small standard deviation, $\upsilon_V|V_d|$, where $\upsilon_V = 0.025$. The second random effect on Cl had standard deviation $\upsilon_C|Cl|$, where υ_C varied from zero to 0.2. We present results for these two experiments in Sections 3.6.1 and 3.6.2 below.

As in example B, we expect the results for example C to play out a little differently for the nonlinear random effects than was seen in example A where random effects were linear. As the magnitude of the nonlinear random effects becomes large, we expect the models with k principal components or common factors to be inadequate, even when the true number of random effects is k. When a k-factor model is forced to fit, the estimate of within-subject variation will tend to be too large, and will be confounded with between-subject variation, making it difficult to test for homoscedastic within-subject variation. In the second experiment of example C, both random effects were nonlinear, in contrast to the two-effect experiment of example B where the first of the two random effects was linear. For this reason, in the experiment with two random effects, we kept the standard deviation of the first random effect on V_d small. Initially we conducted an experiment where the first random effect on V_d was large ($\upsilon_V = 0.2$). We discuss these results briefly in Section 3.6.3.

3.6.1 One nonlinear random effect

The convergence rate for maximum likelihood estimation of the one-factor MLFA-U model was above 99% in all cases. The convergence rate of maximum likelihood estimation for the one-factor MLFA-P model was 100%.

Number of random effects

When $n > p$ or $n < p$ (for the PCA test), we expect the power of the likelihood ratio tests to reject the hypothesis of no common factors, to increase with υ_V and to have size close to the nominal level of 0.05 when $\upsilon_V = 0$ (no random effect). When $n \approx p$ asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we

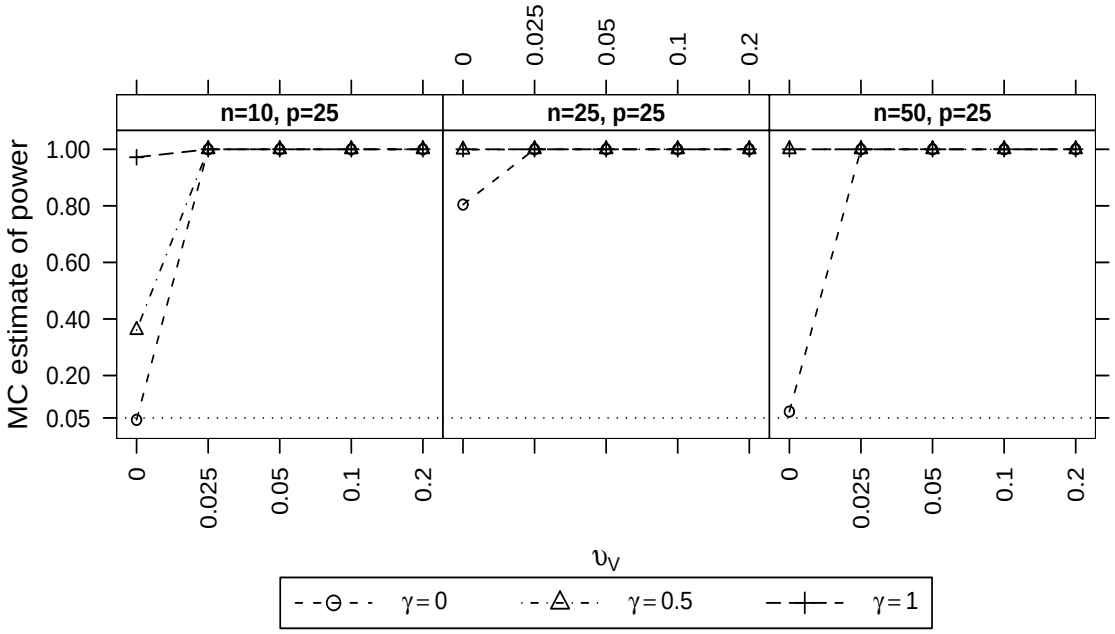

Figure 3.34: *Estimated power of the PCA test for* 0 *common factors when* $p = 25$ *and there was one nonlinear random effect in the O1C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Table 3.28: *Estimated size of the PCA test for* 0 *common factors when* $\gamma = 0$ *and there was one nonlinear random effect in the O1C model with* $\upsilon_V = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	p		
n	10	25	50
10	0.167	0.043	0.037
25	0.037	0.804	0.049
50	0.039	0.072	1.000

would expect the PCA test to reject the null hypothesis even when there were no common factors.

In the batched ANOVA of simulation results, the magnitude of the random effect (υ_V), magnitude of γ, interaction between υ_V and γ, and the number of subjects (n) explained 73% of variation in power of the PCA test to reject the hypothesis of no common factors. The random effect magnitude (υ_V) explained 99% of variation in power of the MLFA-U test.

Figure 3.34 displays estimated power of the PCA test against the magnitude of the random effect signal υ_V when $p = 25$ (results for $p = 10$ and $p = 50$ were similar to those presented in the figure). The three panels in the figure show the effect of a different number of subjects n for fixed p. Power to reject the null hypothesis was close to 100% even when the random effect magnitude was small ($\upsilon_V = 0.025$), and there was little effect of n or p on power when $\upsilon_V > 0$. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. When $\gamma = 0$ and $n < p$ or $n > p$, the size of the test was close to the nominal value (0.05) when $\upsilon_V = 0$ (see Table 3.28). When $n = 50$ and $p = 25$, the size was slightly larger than expected. When $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on and the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there were no principal components too often.

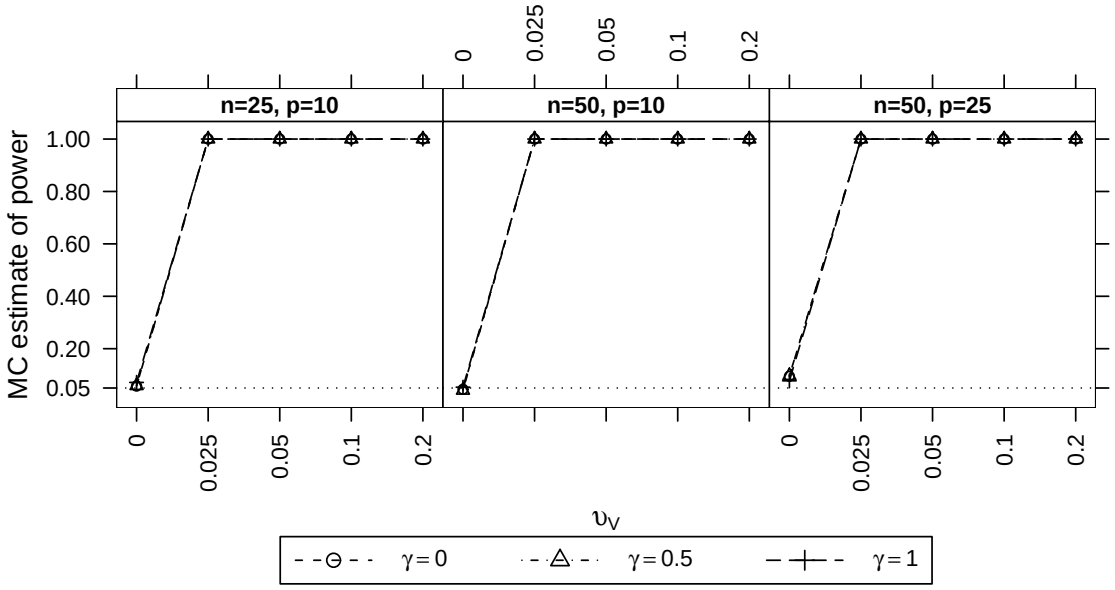

Figure 3.35: *Estimated power of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one nonlinear random effect in the O1C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Figure 3.35 presents the power curves for the MLFA-U test for the cases where $n > p$. The size of the test was close to the nominal 0.05 level when $n = 50$ and $p = 10$, and rejected the null hypothesis too often when the ratio between n and p was closer to 1 (see Table 3.29). When $n > p$, the test had almost 100% power to detect a random effect with υ_V as small as

Table 3.29: *Estimated size of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one nonlinear random effect in the O1C model with* $\upsilon_V = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

		γ		
n	p	0.0	0.5	1.0
25	10	0.058	0.059	0.071
50	10	0.045	0.041	0.053
50	25	0.095	0.097	0.077

0.025, and there was little effect of n, p or γ on power for this amount of random effect signal.

Test for homogeneous variance

For the likelihood ratio tests comparing the fits of the one-factor MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation of the null distribution to be adequate when n is large relative to p. More parameters ($p-2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ. As the magnitude of the nonlinear random effect increases, we expect that the

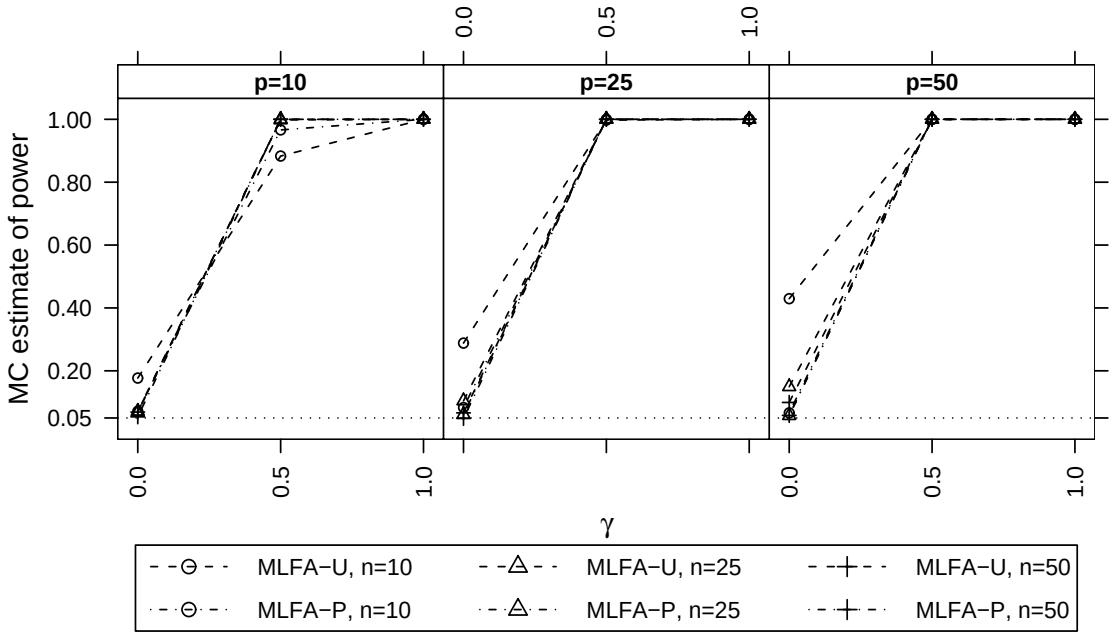

Figure 3.36: *Estimated power of the likelihood ratio tests for H_0: $\gamma = 0$ when there was one nonlinear random effect in the O1C model with $v_V = 0.025$. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Table 3.30: *Estimated size of the likelihood ratio tests for H_0: $\gamma = 0$ when there was one nonlinear random effect in the O1C model. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. The table reports results for $\upsilon_V = 0.025$ or 0.1. Nominal size of the test was 0.05. Based on 1600 MC replicates, the standard error of estimated size was approximately 0.005.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\upsilon_V = 0.025$	$\upsilon_V = 0.1$	$\upsilon_V = 0.025$	$\upsilon_V = 0.1$	$\upsilon_V = 0.025$	$\upsilon_V = 0.1$
	MLFA-U					
10	0.177	0.678	0.288	0.831	0.429	0.899
25	0.071	0.971	0.104	0.994	0.149	0.998
50	0.069	0.999	0.066	1.000	0.099	1.000
	MLFA-P					
10	0.071	0.726	0.083	0.808	0.066	0.873
25	0.068	0.984	0.061	0.995	0.058	0.998
50	0.054	1.000	0.049	1.000	0.058	1.000

one-factor models will fail to explain between-subject variation adequately. As a result, we expect the tests to reject the null hypothesis more often than the nominal 0.05 level for $\gamma = 0$ when υ_V is large.

The magnitude of γ, υ_V, and interaction of these two factors explained 90% of variation in power of the test comparing the MLFA-H and MLFA-U models, and 96% of variation in power based on comparing the MLFA-H and MLFA-P models. Figure 3.36 presents power for the two tests when the random effect signal was small ($\upsilon_V = 0.025$). When $\gamma \geq 0.5$, power to reject the hypothesis of homoscedasticity by comparing with the one-factor MLFA-U and MLFA-P models was above 90% in all cases. Size in the presence of a small random effect ($\upsilon_V = 0.025$) and a large random effect ($\upsilon_V = 0.1$) is presented in Table 3.30 for the two tests. In all cases, the size of the MLFA-U test was significantly larger than the nominal level of 0.05. Size of both tests was increased beyond usefulness in the presence of a large random effect signal. The

size of the MLFA-P test was close to the nominal value when the random effect was small ($\upsilon_V = 0.025$) and when n and p were larger than 10.

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the one-factor power model MLFA-P. When the random effect was small, mean bias in $\widehat{\gamma}$ was approximately zero. However as υ_V increased, bias in $\widehat{\gamma}$ became large, especially for $\gamma = 0$ and $\gamma = 0.5$. In these cases, the maximum likelihood estimator of γ was too large. The interaction between magnitude of γ and magnitude of the random effect (υ_V) explained 98% of variation in bias of $\widehat{\gamma}$. Figure 3.37 displays the mean of $\widehat{\gamma}$ against υ_V for values of γ in the case where $n = 50$ and $p = 25$. Bias in $\widehat{\gamma}$ was similar for other values of n and p. The large bias in $\widehat{\gamma}$ when $\gamma = 0$ for $\upsilon_V \geq 0.1$ explains why the test for homoscedasticity based on fitting the MLFA-P model rejected the null hypothesis of $\gamma = 0$ too often in the presence of a large

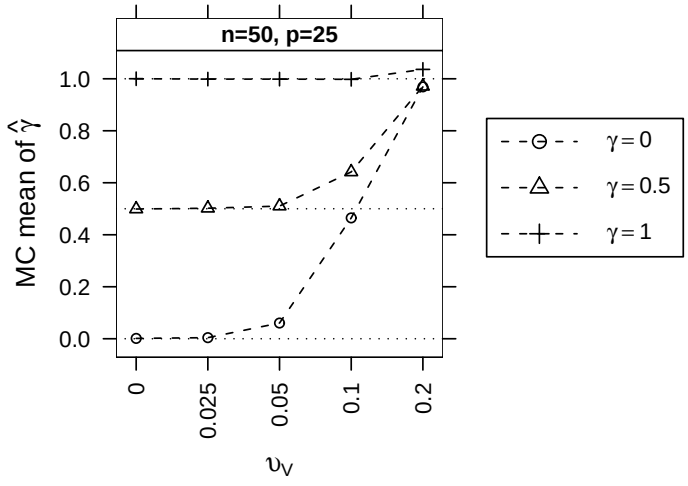

Figure 3.37: *Mean estimate of γ from fitting the one-factor MLFA-P model when there was one nonlinear random effect in the O1C model. The figure plots the case where $n = 50$ and $p = 25$. The true value of γ is given by the dotted horizontal lines. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.004*.*

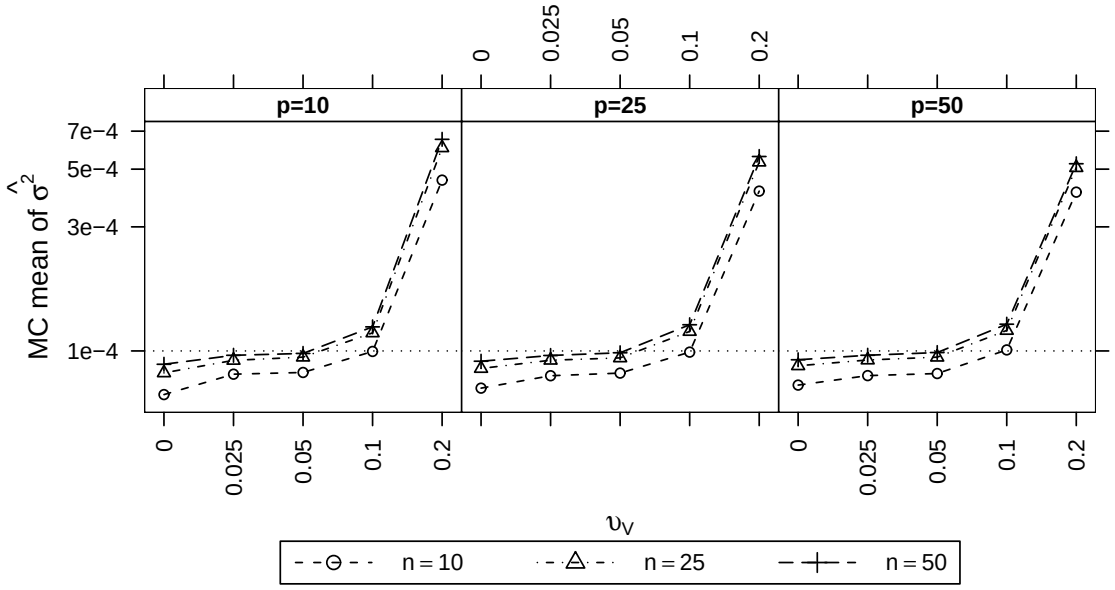

Figure 3.38: *Mean estimate of* σ^2 *from fitting the one-factor MLFA-P model when there was one nonlinear random effect in the O1C model and* $\gamma = 0.5$*. The true value of* σ^2 *was* $1e{-}4$*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* $1e{-}6$*.*

random effect.

The magnitude of the random effect (υ_V), n, and the interaction between these factors explained 85% of variation in the bias of $\widehat{\sigma}^2$ (all simulation factors and their interactions explained 86% of variation). Figure 3.38 illustrates the pattern in the estimate of σ^2 for $\gamma = 0.5$ (note the log scale on the y-axis). The pattern was similar for other values of γ. When the random effect was small, the estimate of σ^2 was smaller on average than the true value of 0.0001. When the the random effect variation was large ($\upsilon_V = 0.2$), the estimate of σ^2 was substantially larger than the true value. The three curves in each figure show the effect of increasing n. The estimate of σ^2 was smaller on average when n was smaller than when n was larger. The number of time points (p) did not have a large effect on bias of $\widehat{\sigma}^2$.

The over-estimation of σ^2 and γ in the case of a large nonlinear random effect suggests that the one-factor model did not adequately explain between-subject variation when υ_V was large.

This may explain the inflated size of the likelihood ratio tests of H_0: $\gamma = 0$ when the nonlinear random effect was large. A model with two or more factors (to account for the second-order terms in the approximation of the stage-1 model) may have given a better fit and less biased estimates of σ^2 and γ.

Selecting random effects and an appropriate parameterization

We used the exact test where the j-th element of $\boldsymbol{\kappa}$ was the first-order sensitivity function of the actual random effect, evaluated at t_j,

$$\begin{aligned}\kappa_j \propto f_V(t_j) &= \frac{Dk_a}{V_d^2(k_a - Cl/V_d)}\left[\frac{Cl}{V_d}t_j^* \exp\left(-\frac{Cl}{V_d}t_j^*\right)\right. \\ &\quad \left. - \left\{1 + \frac{Cl}{V_d(k_a - Cl/V_d)}\right\}\left\{\exp\left(-\frac{Cl}{V_d}t_j^*\right) - \exp(-k_a t_j^*)\right\}\right]\end{aligned}$$

for $t^* = 25t$ for $t \in [0, 1]$. Since the test is based on the exact distribution of the first principal component loading under the null hypothesis, we expect the test based on the PCA analysis to have the nominal size (0.05) when $\upsilon_V = 0$ and $\gamma = 0$, regardless of the magnitudes of n and p. The power should increase for increasing signal $\upsilon_V > 0$. Sample size n and number of time points p should have no effect on size of the test when $\gamma = 0$, but may affect power when the alternative hypothesis is true. When $\gamma > 0$, the principal component model that assumes homogeneous within-subject variance is incorrect, and we would expect size and power of the test to deviate somewhat from size and power in the case where $\gamma = 0$.

When $\upsilon_V > 0$, results for the tests based on the estimated factor loading for the one-factor MLFA-U and MLFA-P models were virtually indistinguishable from those based on the first principal component loading. When $\upsilon_V = 0$ and $\gamma > 0$, the exact test based on the PCA loading rejected the null hypothesis in more cases than the test based on the first MLFA loading. The random effect magnitude (υ_V), γ, number of time points (p), and two-way interaction between these factors explained 95% of variation in power of the PCA exact test. Power was close to 100% for $\upsilon_V \geq 0.025$. Figure 3.39 shows power of the exact PCA test to reject the null

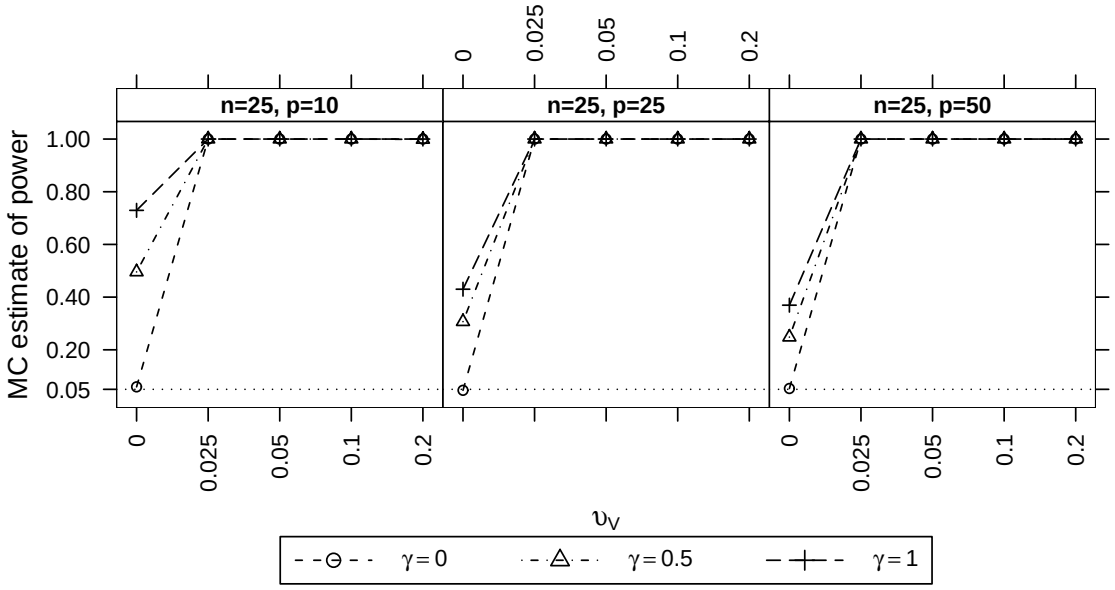

Figure 3.39: *Estimated power of the PCA exact test when there was one nonlinear random effect in the O1C model and* $n = 25$. *This test rejects* H_0*:* $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ *in favor of one random effect with sensitivity function* f_V *. Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125.

hypothesis in favor of a random effect on V_d when $n = 25$. Power and size were similar for other values of n. Table 3.31 gives the size of the exact PCA and MLFA-U tests under the null hypothesis. When $\gamma = 0$ (homoscedastic case) size for both tests was close to the nominal 0.05 level regardless of n and p. However when $\gamma > 0$, the size was much larger than 0.05. This is different from the case of a single nonlinear random effect in example B where size of the exact test tended to be smaller in the presence of heteroscedasticity. Size from fitting the MLFA-U model was similar to that from fitting the MLFA-P model. When $\gamma > 0$, size from fitting either factor model was smaller than that for the PCA test, but was still significantly higher than the nominal 0.05 level.

We used multivariate forward selection, with the first PCA or factor loading as response variable, to identify the first- and second-order sensitivity functions corresponding to the correct nonlinear random effect from amongst several possible predictors. For parameterization f

Table 3.31: *Estimated size of the PCA and MLFA-U exact tests for a random effect on V_d when there was one nonlinear random effect in the O1C model with $\upsilon_V = 0$. We present the cases where $\gamma = 0$ and $\gamma = 1$. Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
	PCA					
10	0.053	0.598	0.047	0.358	0.049	0.313
25	0.059	0.729	0.046	0.430	0.053	0.369
50	0.055	0.808	0.050	0.466	0.052	0.400
	MLFA-U					
10	0.054	0.314	0.054	0.265	0.047	0.237
25	0.063	0.270	0.053	0.254	0.041	0.223
50	0.073	0.251	0.046	0.249	0.051	0.226

we considered the following sensitivity functions, including a second-order sensitivity for the parameter V_d,

$$
\begin{aligned}
f_k(t) &\propto \frac{-Cl/V_d}{k_a - Cl/V_d}\left\{\exp\left(-\frac{Cl}{V_d}t^*\right) - \exp(-k_a t^*)\right\} + k_a t^* \exp(-k_a t^*) \\
f_V(t) &\propto \frac{Cl}{V_d}t^* \exp\left(-\frac{Cl}{V_d}t^*\right) - \left\{1 + \frac{Cl}{V_d(k_a - Cl/V_d)}\right\}\left\{\exp\left(-\frac{Cl}{V_d}t^*\right) - \exp(-k_a t^*)\right\} \\
f_C(t) &\propto \frac{1}{k_a - Cl/V_d}\left\{\exp\left(-\frac{Cl}{V_d}t^*\right) - \exp(-k_a t^*)\right\} + t^* \exp\left(-\frac{Cl}{V_d}t^*\right) \\
f_{VV}(t) &\propto \left\{2 + \frac{4Cl}{V_d(k_a - Cl/V_d)} + \frac{2Cl^2}{\{V_d(k_a - Cl/V_d)\}^2}\right\}\left\{\exp\left(-\frac{Cl}{V_d}t^*\right) - \exp(-k_a t^*)\right\} \\
&\quad + \frac{Cl}{V_d}t^* \exp\left(-\frac{Cl}{V_d}t^*\right)\left\{-4 + \frac{Cl}{V_d}t^* - \frac{2Cl}{V_d(k_a - Cl/V_d)}\right\}.
\end{aligned}
\tag{3.9}
$$

When $\upsilon_V > 0$ we would expect the forward selection to pick a model including f_V. As υ_V becomes large we would expect the second-order sensitivity f_{VV} to be included too.

The random effect on V_d induced two correlated random effects on parameters $\delta^{(1)}$ and $\delta^{(2)}$

in the alternative parameterization f^* of equation (3.5). We also used multivariate regression to select sensitivities for the parameterization f^*. We included the following sensitivities and second order sensitivities as candidates for the parameterization f^*,

$$\begin{aligned} f_1^*(t) &= D\left\{\exp(-\delta^{(2)}t^*) - \exp(-\delta^{(3)}t^*)\right\} \\ f_2^*(t) &= -D\delta^{(1)}t^*\exp(-\delta^{(2)}t^*) \\ f_3^*(t) &= -D\delta^{(1)}t\exp(-\delta^{(3)}t^*) \\ f_{22}^*(t) &= D\delta^{(1)}t^{*2}\exp(-\delta^{(2)}t^*) \end{aligned}$$

for $t^* = 25t$ for $t \in [0,1]$. We did not include the second-order sensitivity f_{11}^* which is the zero function, or the second-order sensitivity f_{12}^* which is proportional to f_2^*. For $\upsilon_V > 0$ we would expect the multivariate regression model to include the two sensitivities f_1^* and f_2^*. As υ_V becomes large, the second-order sensitivity f_{22}^* may also be included. We distinguished between the appropriateness of the two parameterizations based on the number of sensitivities selected for each, and the trace correlation r_T of the selected models. The true parameterization (f) should have give a smaller multivariate regression model, and/or larger trace correlation than the alternative parameterization (f^*).

The multivariate regression results based on the PCA, the MLFA-U and the MLFA-P models were very similar, except when $\gamma > 0$ and $\upsilon_V = 0$. In this case, the PCA analysis picked a larger model, including first- and second-order sensitivities corresponding to parameter V_d more often than would be expected at the nominal 0.05 level. We present results only for the unrestricted MLFA-U model.

The main effects of random effect magnitude (υ_V), n, and p explained 93% of variation in mean number of sensitivities selected for the true parameterization (f). The interaction between υ_V and number of time points (p) explained 98% of variation in the mean number of sensitivities selected for the alternative parameterization f^*. Since the level of γ did not have

a large effect on average model size when $\upsilon_V > 0$, we present results only for the case where $\gamma = 0.5$ in Figure 3.40.

This figure summarizes the mean number of sensitivities included in the multivariate regression models for parameterizations f (dashed curves) and f^* (dot-dash curves) of the O1C model. Contrary to expectations, when $p = 10$ a smaller model was selected for the alternate parameterization f^* than for the true parameterization f. However, when $p = 25$ or 50, the larger model was selected for the alternate parameterization except when the random effect on V_d was large ($\upsilon_V = 0.2$). The effect of n is shown by the different plotting symbols for each parameterization; as n increased, the average model size increased slightly. When the nonlinear

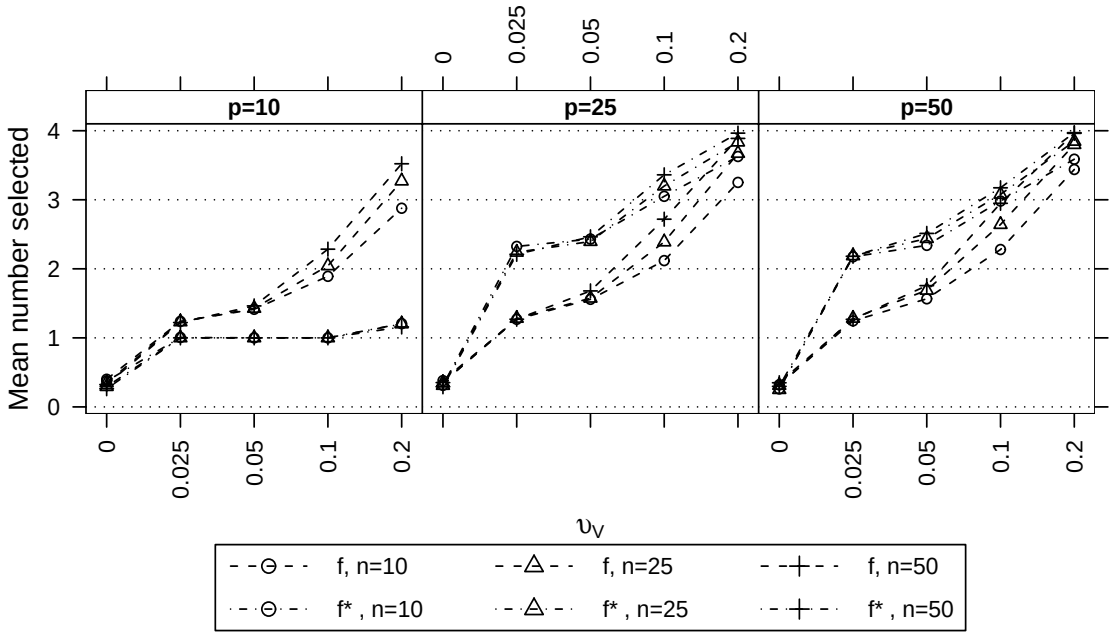

Figure 3.40: *Mean number of sensitivities selected by multivariate forward selection when there was one nonlinear random effect in the O1C model. Results presented are for the true parameterization f (dashed curves) and alternate parameterization f^* (dot-dash curves), and the regression using the factor loading estimated from the one-factor MLFA-U model when $\gamma = 0.5$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.02*.*

random effect was small, forward selection picked a single sensitivity for the true parameterization f. As the magnitude of the random effect increased, a second, third, or fourth sensitivity was included for f, and additional sensitivities were also included for the parameterization f^* when $p > 10$. When $p = 10$, the average model size selected for parameterization f^* was one, even when the random effect was large.

Table 3.32 summarizes the mean number of sensitivities selected for the true parameterization f when there was no random effect ($\upsilon_V = 0$). When $\gamma = 0$, the average model size selected by the multivariate forward selection was approximately 0.15 — slightly larger than might be expected by the 0.05 nominal level of the tests at each stage of the forward selection. In the presence of heteroscedasticity ($\gamma = 1$), the average model size was approximately 0.5.

The size of the random effect (υ_V) explained 97% of variation in the trace correlation (r_T) of the model selected for the true parameterization f. When $\upsilon_V > 0$, the mean trace correlation r_T was close to 1, and the factors n, p and γ had little effect. However, for the multivariate regression model selected for the alternate parameterization f^*, when $\upsilon_V > 0$ the trace correlation was smaller when $p = 10$ than for other values of p. The factors υ_V and p explained 95% of variation in r_T based on f^*. The lower value of r_T corresponds to the selection

Table 3.32: *Mean number of sensitivities included for the true parameterization f by forward selection for the O1C model with one nonlinear random effect with $\upsilon_V = 0$. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $\gamma = 0$ or $\gamma = 1$. At each step of the forward selection the 0.05 critical value of the test statistic was used. Based on analyzing the simulation data in batches of 10, we estimated the standard error of each MC mean to be less than 0.02.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.173	0.647	0.160	0.551	0.146	0.436
25	0.166	0.576	0.163	0.527	0.155	0.441
50	0.165	0.530	0.150	0.514	0.159	0.434

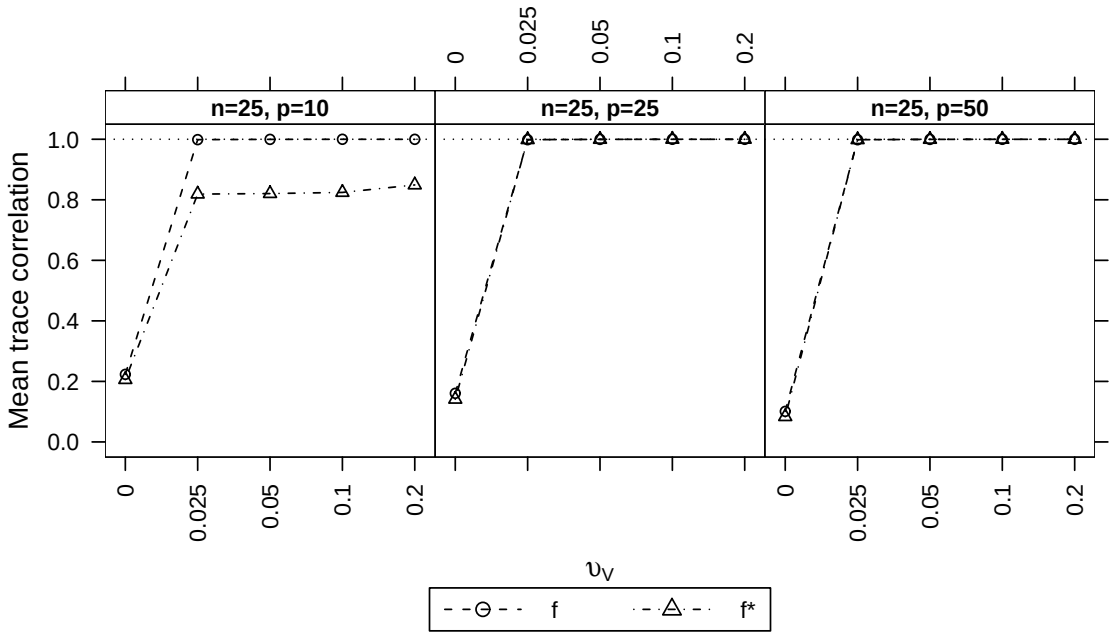

Figure 3.41: *Mean trace correlation for the selected multivariate regression models when there was one nonlinear random effect in the O1C model. Results presented are for the true parameterization f (dashed curves) and alternate parameterization f^* (dot-dash curves), and the regression using the factor loading estimated from the one-factor MLFA-U model when $n = 25$ and $\gamma = 0.5$. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.01*.*

of a smaller model for f^* when $p = 10$ than for other values of p. Figure 3.41 shows the trace correlation for the models selected for f and f^* based on the MLFA-U analysis when $n = 25$ and $\gamma = 0.5$. Results for other values of n and γ were similar. The results for the anomalous cases where $p = 10$ illustrate how discrimination between parameterizations f and f^* should be based on the size of the selected model as well as some measure (such as r_T) of how well each multivariate regression model explains variation in the principal component or common factor loadings.

Next we examine which sensitivities from the true parameterization f were selected. The simulation factors υ_V, γ, and p explained 98% of variation in the proportion of times the sensitivity f_V was selected, and 88% of variation in the proportion of cases where f_{VV} was

selected. Figure 3.42 displays the proportion of cases where multivariate forward selection picked these sensitivities. The figure only shows the case where $n = 25$. Results for other values of n were similar. When there was no random effect ($\upsilon_V = 0$), the sensitivity f_V was selected in fewer than 5% of cases when within-subject variation was homoscedastic ($\gamma = 0$), and in 15% to 25% of cases in the presence of heteroscedasticity ($\gamma = 1$) (see Table 3.33).

When $\upsilon_V > 0$, the first-order sensitivity f_V (dashed curves) was selected in close to 100% of cases, regardless of the magnitude of γ. The second-order sensitivity f_{VV} (dot-dash curves) was included when the random effect magnitude became large. This confirms that the factor analysis

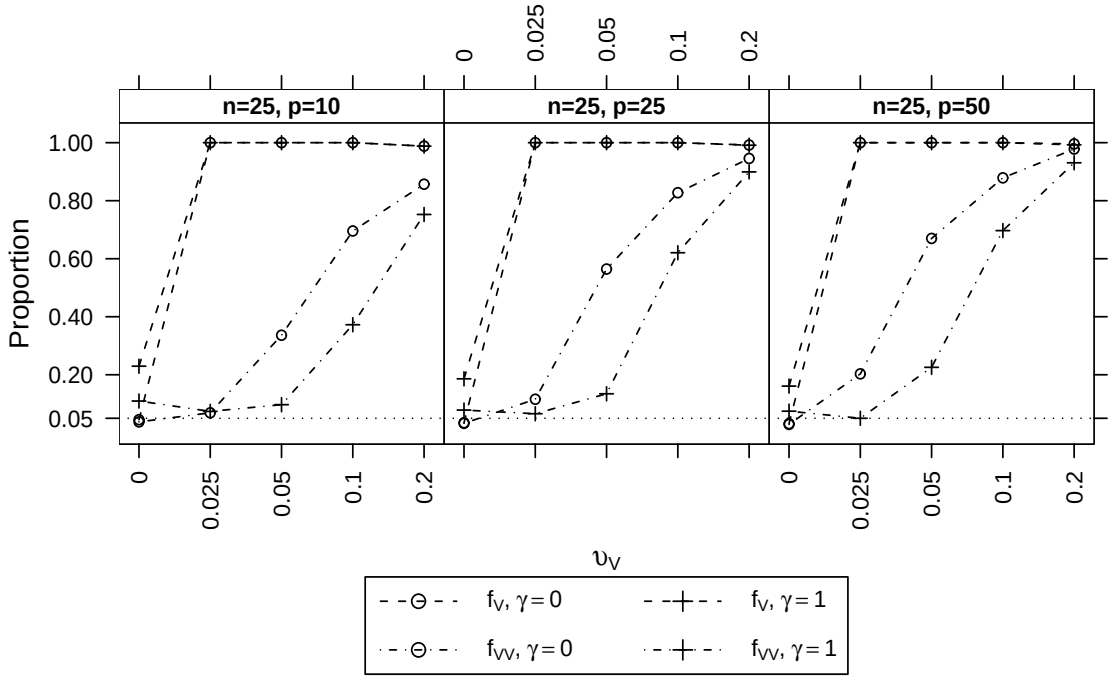

Figure 3.42: *Proportion of cases where multivariate forward selection picked sensitivities corresponding to a random effect on V_d in the O1C model with one random effect. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $\gamma = 0$ or $\gamma = 1$ and $n = 25$. The dashed curves plot the proportion of times the sensitivity f_V was selected. The dot-dash curves plot the proportion of times the second-order sensitivity f_{VV} was selected. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125.

detected a common factor related to the square of the random effect on V_d when this random effect became large, even though a model with only a single common factor was actually fitted. The three panels in the figure show the effect of increasing p; the selection of f_{VV} increased slightly when p was larger. Different plotting symbols represent different levels of γ. When $\upsilon_V > 0$, the second-order sensitivity function f_{VV} was selected in a smaller proportion of cases in the presence of heteroscedasticity than when within-subject variation was homoscedastic.

When the magnitude of the random effect became large, the average model size selected for parameterization f was three to four. When υ_V became large, the sensitivities f_C and f_k were included into the multivariate regression model in addition to the correct sensitivities f_V and f_{VV}. In most cases, in the presence of large random effect signal, the sensitivity for the clearance parameter, f_C, was included slightly more often than the sensitivity for the absorption rate, f_k.

3.6.2 Two nonlinear random effects

In this experiment there were two independent random effects for parameterization f of equation (3.4). The first random effect on the volume of distribution parameter V_d had small

Table 3.33: *Proportion of cases where forward selection picked f_V when $\upsilon_V = 0$ in the experiment with one nonlinear random effect in the O1C model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $\gamma = 0$ or $\gamma = 1$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.042	0.272	0.039	0.190	0.030	0.183
25	0.045	0.230	0.034	0.186	0.031	0.161
50	0.055	0.212	0.028	0.180	0.037	0.169

standard deviation, $\upsilon_V|V_d|$ where $\upsilon_V = 0.025$. The second random effect on the clearance parameter Cl had standard deviation $\upsilon_C|Cl|$ where we varied υ_C from zero to 0.2. The convergence rate for maximum likelihood estimation of the two-factor MLFA-U model was above 95% except when $\gamma = 0$, $\upsilon_C = 0.2$, and $n = p = 10$. In this case the convergence rate was approximately 87%. The convergence rate for the two-factor MLFA-P model was 100%.

The first random effect on V_d was nonlinear, but since we set this random effect to have small magnitude, we expect results for this experiment to be similar to the case of two linear random effects when the magnitude of the second random effect (υ_C) is small, but to play out a little differently when υ_C is large. In Section 3.6.3 we briefly discuss how results changed when the first nonlinear random effect was large. As in the experiment with one nonlinear random effect, when the second nonlinear random effect becomes large, the principal component and factor analyses may detect additional common factors due to the need for second order terms in the approximation to the stage-1 model. We also expect that the multivariate regression analysis will select second-order sensitivities when υ_C becomes large.

Table 3.34: *Estimated size of the PCA test for* 1 *common factor when* $\gamma = 0$ *and there were two random effects in the O1C model, the second one with* $\upsilon_C = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

	p		
n	10	25	50
10	0.287	0.044	0.040
25	0.058	0.880	0.073
50	0.048	0.110	1.000

Number of random effects

When $n > p$ or $n < p$ (for the PCA test), we expect the likelihood ratio tests to maintain size close to 0.05 and to readily reject the hypothesis of one common factor when $\upsilon_C > 0$. When $n \approx p$ asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we would expect the PCA test to reject the null hypothesis even when there was only one common factor.

The number of subjects (n), γ, magnitude of the second random effect υ_C, and two-way interactions explained 75% of variation in power of the PCA test. The magnitude of the second random effect (υ_C) explained 99% of variation in the power of the MLFA-U test. Figure 3.43 displays estimated power of the PCA test, plotted against the magnitude of the second random effect signal, υ_C, when $p = 25$. Results for $p = 10$ and $p = 50$ were similar to those presented

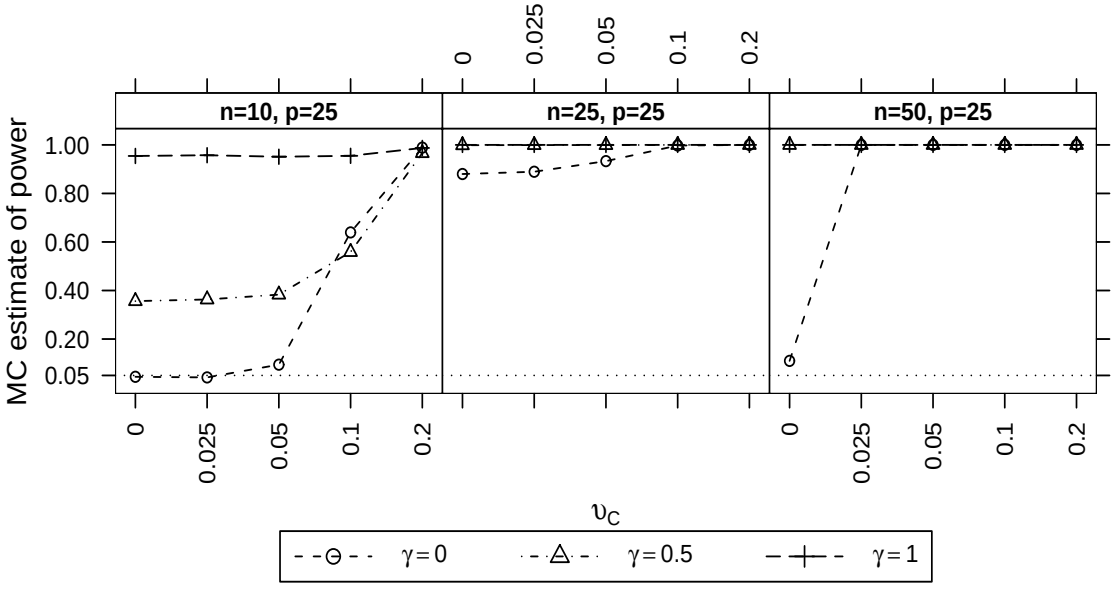

Figure 3.43: *Estimated power of the PCA test for* 1 *common factor when* $p = 25$ *and there were two random effects in the O1C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

in the figure. The three panels in the figure show the effect of a different number of subjects n for fixed p. As n increased, power increased. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. Table 3.34 gives estimated size of the PCA test at $\upsilon_C = 0$ when $\gamma = 0$ for different values of n and p. When n was much larger than p, or p was much larger than n, then test had approximately the correct size (0.05) when $\gamma = 0$. Clearly when $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on and the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there was one principal component too often and would not be useful in the presence of heteroscedastic within-subject variance. When $\gamma = 0$ and $n < p$ or $n > p$, the PCA test had greater than 90% power to detect a second random effect with $\upsilon_C = 0.2$.

Figure 3.44 presents the estimated power for the MLFA-U test for the cases where $n > p$.

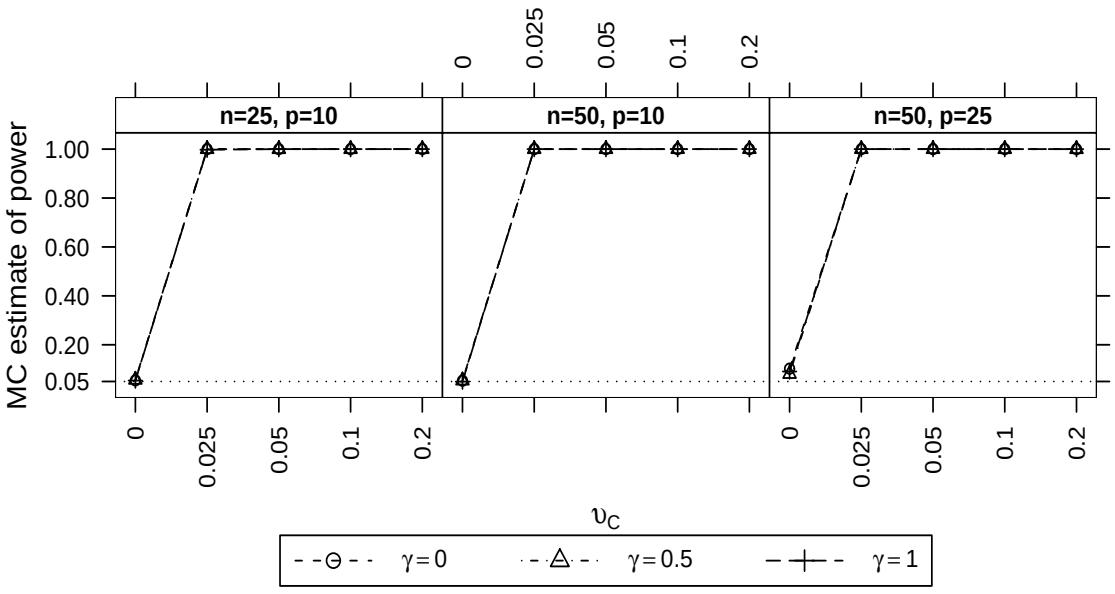

Figure 3.44: *Estimated power of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two random effects in the O1C model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Table 3.35: *Estimated size of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two random effects in the O1C model, the second one with* $\upsilon_C = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

		γ		
n	p	0.0	0.5	1.0
25	10	0.054	0.057	0.054
50	10	0.053	0.054	0.049
50	25	0.102	0.081	0.091

The size of the test was close to the nominal 0.05 level except when $n = 50$ and $p = 25$; in that case, the size was larger than the nominal level (see Table 3.35). When $n > p$, the MLFA-U test had greater than 90% power to detect a second random effect with $\upsilon_C = 0.025$ or larger. The presence of heteroscedasticity, γ, had little effect over the range of υ_C studied in this experiment. When $n > p$ and within-subject variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same size and power.

Test for homogeneous variance

For the likelihood ratio tests comparing the fits of the MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation of the null distribution to be adequate when n is large relative to p. More parameters ($p - 2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ.

The magnitude of γ, the second random effect (υ_C), n, p, and two-way interactions between these factors explained 90% of the variation in power to reject the hypothesis of homoscedasticity by comparing the MLFA-H model with the unrestricted MLFA-U model. The interaction between υ_C and γ explained 96% of variation in the power of the test comparing the MLFA-H

Table 3.36: *Estimated size of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two nonlinear random effects in the O1C model. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. The table reports results for* $\upsilon_C = 0.025$ *or* 0.1*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\upsilon_C = 0.025$	$\upsilon_C = 0.1$	$\upsilon_C = 0.025$	$\upsilon_C = 0.1$	$\upsilon_C = 0.025$	$\upsilon_C = 0.1$
	MLFA-U					
10	0.308	0.580	0.487	0.740	0.712	0.853
25	0.114	0.839	0.135	0.849	0.201	0.897
50	0.082	0.984	0.094	0.986	0.092	0.994
	MLFA-P					
10	0.120	0.168	0.105	0.170	0.109	0.217
25	0.054	0.139	0.059	0.171	0.068	0.357
50	0.065	0.140	0.050	0.229	0.059	0.598

model with the MLFA-P model. Table 3.36 presents size of the tests for small random effect ($\upsilon_C = 0.025$) and large random effect ($\upsilon_C = 0.1$).

In all cases, the size of the test comparing the homoscedastic model to the unrestricted model MLFA-U was significantly larger than the nominal level of 0.05 (see Table 3.36), making this test close to useless. Size of both tests was increased by a large random effect signal. The test comparing the homoscedastic model to the power model MLFA-P performed slightly better when the random effect on Cl was small, although its size was too large when n was small. It is possible that the size of the MLFA-P test would have been closer to the nominal level if we had fitted a factor model with more than two factors when the random effect variation was large. Since magnitude of the second random effect signal did not have a large impact on power for $\gamma > 0$, we have presented results only for $\upsilon_C = 0.05$ in Figure 3.45. When $n > 10$ or $p > 10$ the test based on the MLFA-P model had greater than 90% power to detect heteroscedasticity

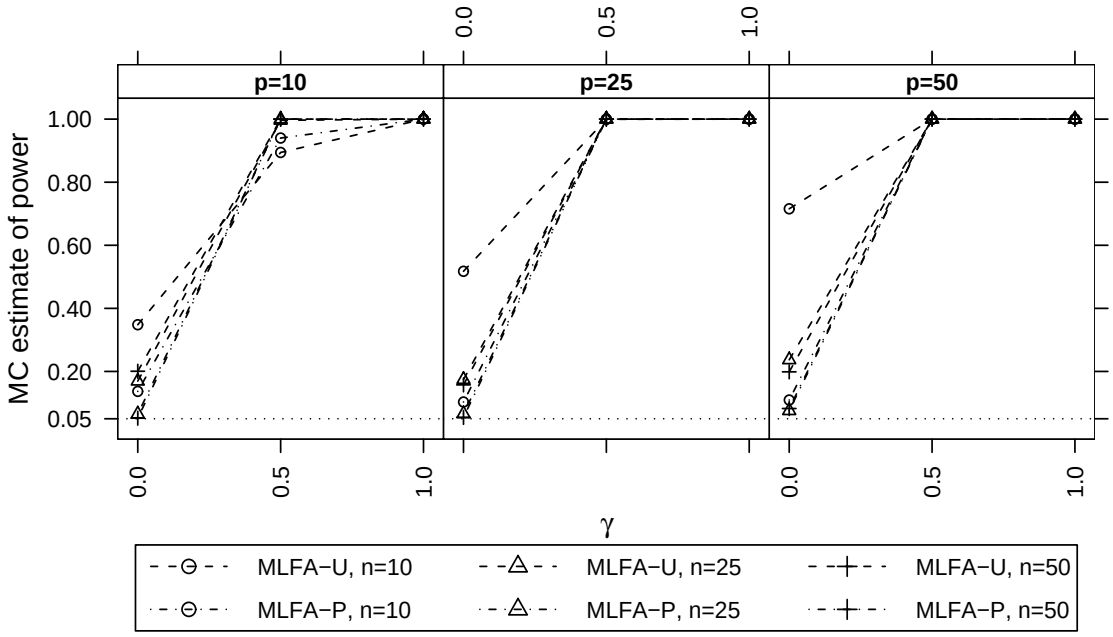

Figure 3.45: *Estimated power of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two random effects in the O1C model and* $\upsilon_C = 0.05$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U (dashed curves). The second test compared the homoscedastic model with the power model MLFA-P (dot-dashed curves). Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

when γ was as large as 0.5.

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the two-factor power model MLFA-P. On average, the bias in $\widehat{\gamma}$ was approximately zero except when the nonlinear random effect was large ($\upsilon_C = 0.2$). In this case the maximum likelihood estimator of γ was too large, especially for the case of $\gamma = 0$. The interaction between the magnitude of γ and the random effect (υ_C) explained 92% of variation in bias of $\widehat{\gamma}$. Figure 3.46 plots the mean of $\widehat{\gamma}$ against υ_C for values of γ in the case where $n = 50$ and $p = 25$.

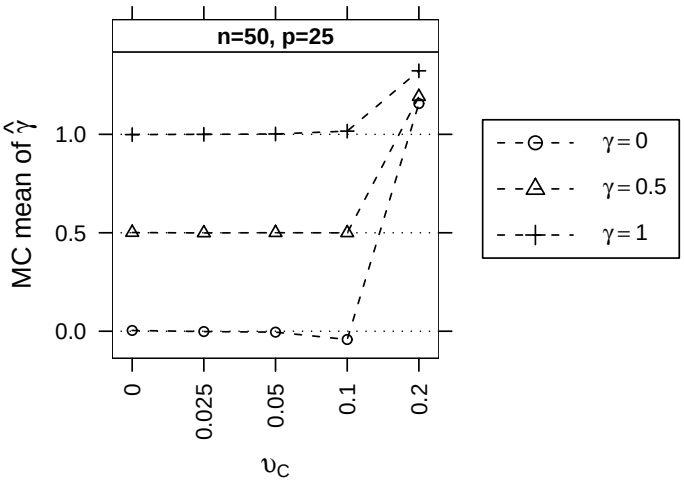

Figure 3.46: *Mean estimate of γ from fitting the two-factor MLFA-P model when there were two random effects in the O1C model. The figure plots the case where $n = 50$ and $p = 25$. The true value of γ is given by the dotted horizontal lines. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.005*.*

The magnitude of γ, the random effect υ_C, number of observations (n), number of time points (p), and two-way interactions between these factors explained 59% of variation in bias of $\widehat{\sigma}^2$ (the model including all interactions explained 67% of variation). Figure 3.47 illustrates the pattern in the estimate of σ^2 for $p = 25$. The estimate of σ^2 was smaller on average than the true value of 0.0001 except in some cases when $\upsilon_C = 0.1$ or 0.2. In these cases bias in the estimate of σ^2 was large, but sometimes $\widehat{\sigma}^2$ was larger than the true value of σ^2, and in other cases it was smaller than the true value. Since the pseudolikelihood estimator of σ^2 for the MLFA-P model depends on the estimator of γ (equation (C.13)), the behavior of $\widehat{\sigma}^2$ for $\upsilon_C = 0.2$ can perhaps be explained by the fact that $\widehat{\gamma}$ was significantly larger than γ when $\upsilon_C = 0.2$, especially when the true value of γ was zero. Clearly the two-factor MLFA-P model did not adequately explain the between-subject variation due to the large nonlinear random effect, but in this experiment, the estimate of γ, not σ^2, was increased to take account of the unexplained within-subject variation, in contrast to what we saw in example B.

The three panels in Figure 3.47 show the effect of increasing n for fixed p; the estimate of

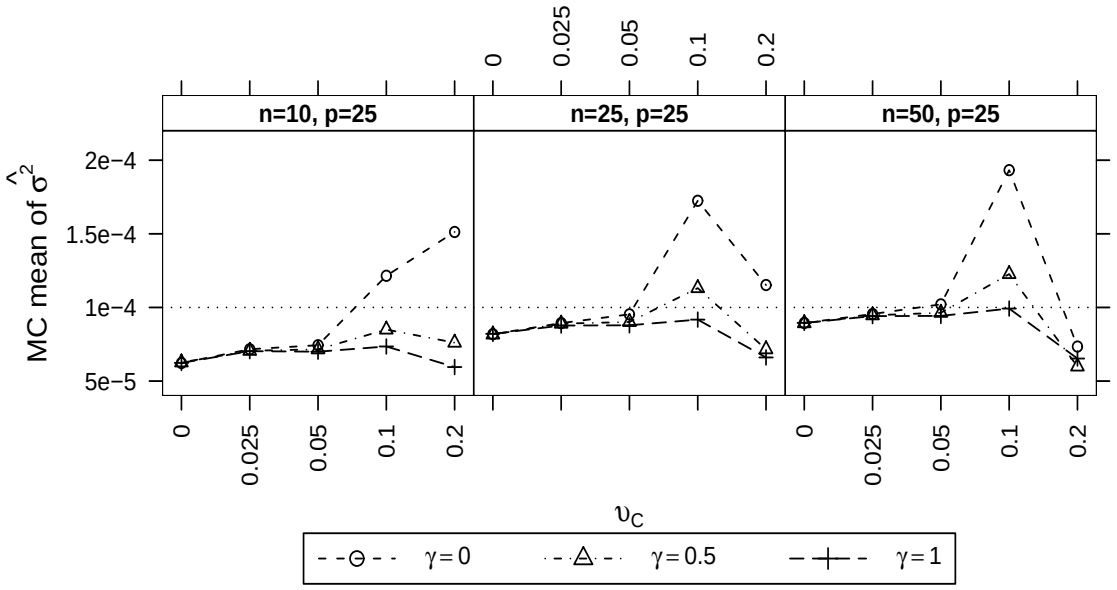

Figure 3.47: *Mean estimate of* σ^2 *from fitting the two-factor MLFA-P model when there were two random effects in the O1C model and* $p = 25$*. The true value of* σ^2 *was* $1e{-}4$*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* $1e{-}6$*.*

σ^2 tended to be larger for larger n, except when $\upsilon_C = 0.2$. The three curves in each panel represent the mean estimate of σ^2 for each setting of γ. Bias in $\widehat{\sigma}^2$ in the presence of the large nonlinear random effect was largest when $\gamma = 0$.

The bias in the estimates of σ^2 and γ in the case of a large nonlinear random effect suggests that the two-factor model did not adequately explain between-subject variation when υ_C was large. We expect that a model with three or more factors (to account for the second-order terms in the approximation of the stage-1 model) would have given a better fit.

Selecting random effects

We used multivariate forward selection, with the first two PCA or factor loadings as response variables, to identify the first- and second-order sensitivity functions corresponding to the correct nonlinear random effects from amongst several possible predictors. For parameterization f

we considered the sensitivity functions f_k, f_V, f_C, and f_{VV} that were given in equation (3.9). We did not consider the second-order sensitivity f_{VC} for parameters V_d and Cl because it is is a linear function of f_V and f_{VV}. We also did not include the second-order sensitivity f_{CC} for parameter Cl because there was a linear dependency among f_V, f_C, f_{VV} and f_{CC}. In retrospect it would have made more sense to consider f_{CC} in the place of f_{VV}, since the magnitude of the random effect on Cl was larger than that on V_d for most of the simulation treatments. However, our "incorrect" choice of candidate predictors does give some insight into how the multivariate forward selection behaves when an incomplete set of candidate second-order sensitivities is supplied. Also, since f_{CC} is a linear function of the candidate sensitivities, we can make some inference on selection of f_{CC} by the frequency in which f_{VV} was selected in the presence of f_V and f_C.

When $\upsilon_C = 0$ the forward selection should select a model with one sensitivity, f_V, corresponding to the random effect on V_d. As υ_C gets larger, the forward selection should additionally select the sensitivity f_C in a high proportion of cases. As mentioned in the previous sections, when the nonlinear random effect is large, the two-factor model may not adequately capture the between-subject variation in the data because of second-order terms in the approximation to the stage-1 model. However, we would still expect the forward selection procedure based on two common factors to pick the second-order sensitivity for parameter Cl when υ_C is large. Although the sensitivity f_{CC} was not included as a candidate, this sensitivity is a linear function of f_V, f_{VV}, and f_C, so we would expect f_{VV} to be included in an increasing proportion of cases as υ_C becomes large.

The multivariate regression results based on the PCA, MLFA-U and the MLFA-P models were very similar when within-subject variation was homoscedastic. When $\gamma > 0$, and $p = 10$, average model size based on the PCA analysis was larger than for the MLFA-U and MLFA-P analyses. Results were almost indistinguishable in all other cases. For the sake of brevity we present results only for the MLFA-U analysis. The magnitude of the random effect (υ_C), number of subjects (n), and interaction between p and γ explained 93% of variation in the number of

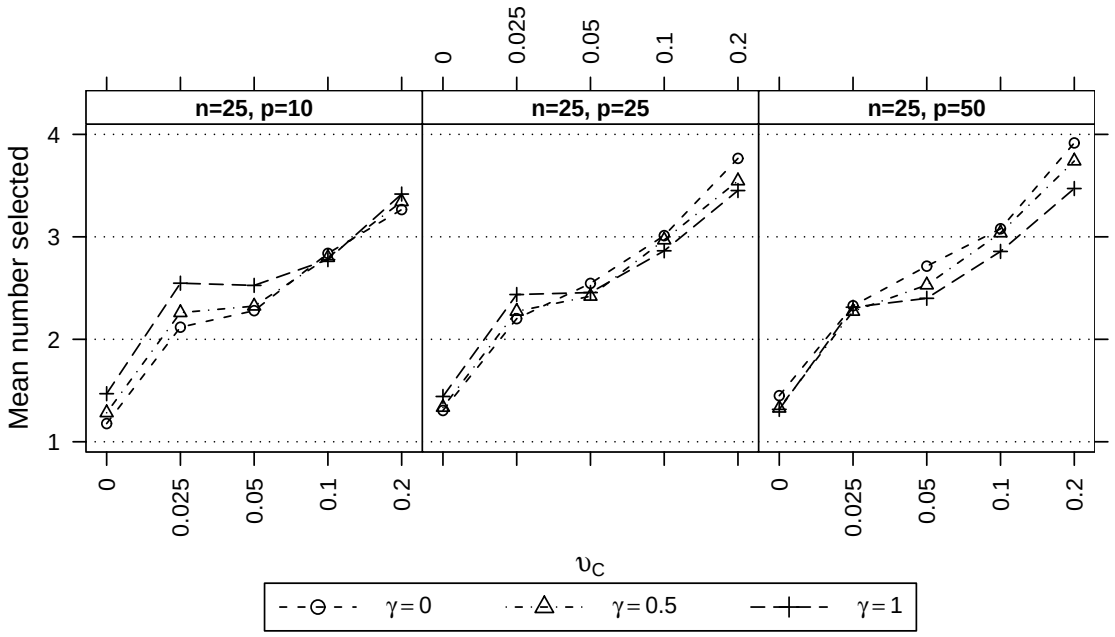

Figure 3.48: *Mean number of sensitivities selected by multivariate forward selection when there were two random effects in the O1C model. Results presented are for the regression using the factor loadings estimated from the two-factor MLFA-U model where* $n = 25$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.02*.*

sensitivities selected. Figure 3.48 summarizes the mean number of sensitivities included in the multivariate regression model when $n = 25$.

When $\upsilon_C = 0$, a model with one sensitivity was selected in most cases (see Table 3.37), but as p or γ increased, a model with more than one random effect was selected in an increasing proportion of cases. When the magnitude of the random effect on Cl was greater than zero but small, the average model size selected was between two and three. When the nonlinear random effect became large ($\upsilon_C \geq 0.1$), the procedure selected a model with three or more sensitivities in an increasing number of cases. The three panels in the figure show the effect of increasing p for fixed n; average model size increased slightly with increasing p. The three curves in each panel show the effect of within-subject heteroscedasticity. There was no consistent effect of γ

Table 3.37: *Mean number of sensitivities included by forward selection when there were two random effects in the O1C model with* $\upsilon_C = 0$. *Results presented are for the regression using the factor loading estimated from the two-factor MLFA-U model. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each MC mean to be less than* 0.02.

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	1.144	1.290	1.456	1.245	1.331	1.442	1.361	1.295	1.308
25	1.177	1.283	1.469	1.305	1.336	1.442	1.450	1.328	1.316
50	1.177	1.263	1.456	1.341	1.344	1.465	1.528	1.351	1.323

on the mean number of sensitivities included in the multivariate regression model.

The sensitivity f_V corresponding to the small nonlinear random effect on V_d was included in the multivariate regression model in greater than 95% of cases. The magnitude of the random effect (υ_C), γ, and the number of time points (p) explained 99% of the variation in the proportion of cases where the sensitivity f_C was selected. These factors explained 87% of variation in the proportion of cases in which f_{VV} was selected.

Figure 3.49 displays the proportion of cases where multivariate forward selection picked the sensitivity f_V corresponding to the first random effect, and f_C and f_{VV}, the first- and second-order sensitivities corresponding to the nonlinear random effects on Cl and V_d respectively. Since the second-order sensitivity f_{CC} was a linear function of f_V, f_C and f_{VV}, inclusion of f_{VV} in the model could also be interpreted as evidence of second-order behavior of the random effect on Cl. We plotted results only for the case where $n = 25$. Results were similar for other values of n, with a slight increase in selection of f_{VV} when n was larger.

When the second random effect signal υ_C was positive, the sensitivity f_C was included in the model in greater than 99% of cases (dot-dash curves). As υ_C increased, the second-order sensitivity f_{VV} was included in the model with increasing frequency (long-dash curves). The three panels in the figure show the effect of increasing p; the sensitivity f_{VV} was picked more

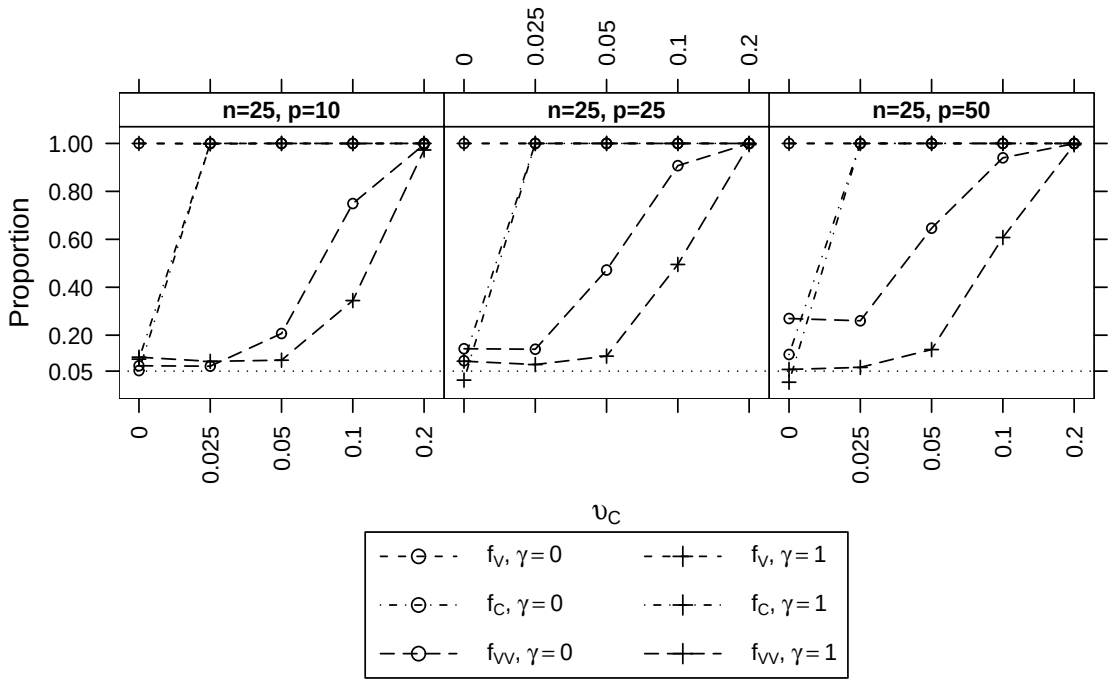

Figure 3.49: *Proportion of cases where multivariate forward selection picked the sensitivities, f_V (dashed curves), f_C (dot-dashed curves), and f_{VV} (long-dash curves), when there were two random effects in the O1C model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $n = 25$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125*.*

often when p was larger. In each panel the different plotting symbols represent different levels of γ. The second-order sensitivity f_{VV} was selected more often when $\gamma = 0$ than when $\gamma = 1$.

When there was only one random effect on V_d ($\upsilon_C = 0$), the sensitivity f_C was selected in 5% to 10% of cases when $\gamma = 0$, with increased selection as p increased — see Table 3.38. In contrast, when $\gamma = 1$, selection of f_C decreased from 10% to 1% of cases with increasing p. The second order sensitivity f_{VV} was selected in about 30% of replicates when $\upsilon_C = 0$ for the case where $p = 50$ and $\gamma = 0$. This selection of f_{VV} when $\upsilon_C = 0$ was most likely due to the presence of the first random effect on V_d. This is consistent with results of the experiment with a single random effect on V_d that had standard deviation proportional to $\upsilon_V = 0.25$ — see the

Table 3.38: *Proportion of cases where forward selection picked f_C when there were two random effects in the O1C model with $v_C = 0$. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model with $\gamma = 0$ and $\gamma = 1$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005.

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.050	0.094	0.073	0.012	0.106	0.006
25	0.052	0.100	0.093	0.013	0.119	0.004
50	0.046	0.109	0.096	0.015	0.119	0.008

last panel of Figure 3.42 in Section 3.6.1.

This experiment exemplifies how the interpretation of the multivariate regression results may be complicated for large nonlinear random effects. It may be the case that not all parameter sensitivities may be included as candidates for the multivariate forward selection because of linear dependencies between them. The selection of first-order sensitivities seems to be fairly straight-forward: in the experiments of examples B and C, the correct first-order sensitivities were included in the model in a high proportion of cases even when the random effect variation was small. However, interpreting the inclusion of second-order sensitivities may be more complicated. For example, in the most recent experiment, sensitivity f_{VV} could have been included because of slight second-order behavior of the small random effect on V_d, or because of the large nonlinear random effect on Cl and the linear dependency between the candidates f_V, f_C, f_{VV} and f_{CC}. Having said this, perhaps clear identification of second-order sensitivities is not important if the selection of first-order sensitivities has already clearly identified the need for random effects on V_d and Cl. A "hierarchy of predictors" rule could be adopted in which second-order sensitivities are added to the multivariate regression only if the first-order sensitivities have been selected.

In addition, sensitivities are likely to be highly correlated or non-orthogonal, even if not

linearly dependent. When p was large, a model including all four candidate sensitivities was picked in a good number of cases when $\upsilon_C = 0.2$ (Figure 3.48), although there was no random effect on the absorption rate, k_a. In the linear examples of Section 3.4, selection of incorrect sensitivities was rare. We conclude that the inclusion of f_k was due to the presence of a large nonlinear random effect on parameter Cl, and non-orthogonality of the candidate sensitivities. As a remedy, we would recommend that the order in which sensitivities are included by the forward selection procedure should be taken into account when deciding which parameters truly require non-trivial random effects. Admittedly, we have not presented any evidence here to suggest that this will be a fool-proof strategy.

3.6.3 Two nonlinear random effects, the first one being large

Before running the experiment described in the previous section, we ran a similar experiment, with random effects on V_d and Cl, except that the magnitude of the first random effect on V_d was large ($\upsilon_V = 0.2$). When applying our methods to data generated in this experiment, we fitted models with two principal components or common factors. However the large nonlinear random effect on V_d mostly likely required second-order terms in the Taylor series approximation of the stage-1 model. The results of this experiment did not give much information about the performance of our methods, except that a two-factor model was inadequate to capture the between-subject variation induced by two large nonlinear random effects.

In the experiment with a large random effect on V_d, the likelihood ratio tests for number of random effects rejected the null hypothesis that there was one linear latent variable in 100% of cases even when the magnitude of the second random effect, υ_C, was zero. The tests for homoscedastic within-subject variance had size much larger than the nominal 0.05 level due to unexplained between-subject variation after fitting two common factors. The estimates of γ and σ^2 were much larger (sometimes 100-fold larger) than truth for most values of υ_C. In most cases, more than three sensitivities were included in the multivariate models chosen by the forward selection procedure, due to the second-order behavior of the large random effect on

V_d. All four of the candidate sensitivities were included in most cases, especially when n and p were larger than 25.

If we were to redo such an experiment, we would ensure that sufficient principal components or common factors were included in the multivariate models to adequately describe the between-subject variation, before proceeding with the various diagnostic tests based on the fit of these models.

3.7 Simulation results: example D

The first simulation experiment for example D had a single nonlinear random effect on the growth rate parameter $\omega^{(2)}$ in parameterization f of equation (3.6). The standard deviation of this random effect was $\upsilon_2|\omega^{(2)}|$ where we varied υ_2 from zero to 0.4. The second experiment generated data with two independent nonlinear random effects, one on the growth rate parameter $\omega^{(2)}$, and the second on the half-life parameter, $\omega^{(1)}$. The first random effect on $\omega^{(2)}$ had small standard deviation, $\upsilon_2|\omega^{(2)}|$, where $\upsilon_2 = 0.05$. The second random effect on $\omega^{(1)}$ had standard deviation $\upsilon_1|\omega^{(1)}|$, where υ_1 varied from zero to 0.4. We present results for these two experiments in Sections 3.7.1 and 3.7.2 below.

As the magnitude of the nonlinear random effects becomes large, we expect the models with k principal components or common factors to be inadequate, even when the true number of random effects is k. When a k-factor model is forced to fit, the estimate of within-subject variation will tend to be too large, and will be confounded with between-subject variation, making it difficult to test for homoscedastic within-subject variation. In the experiment with two nonlinear random effects, we kept the standard deviation of the first random effect on $\omega^{(2)}$ small. In the course of our work we also conducted an experiment where the first random effect on $\omega^{(2)}$ was large ($\upsilon_2 = 0.4$). We discuss these results briefly in Section 3.7.3.

3.7.1 One nonlinear random effect

The convergence rate for maximum likelihood estimation of the one-factor MLFA-U model was above 95% in all cases. The convergence rate of maximum likelihood estimation for the one-factor MLFA-P model was 100%.

Number of random effects

When $n > p$ or $n < p$ (for the PCA test), we expect power of the likelihood ratio tests to reject the hypothesis of no common factors to increase with υ_2, and to have size close to the nominal level of 0.05 when $\upsilon_2 = 0$ (no random effect). When $n \approx p$ asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we expect the PCA test to reject the null hypothesis even when there were no common factors.

In the batched ANOVA of simulation results, the number of subjects (n) and interaction between magnitude of the random effect (υ_2) and γ explained 78% of the variation in power of the PCA test. The magnitude of the random effect (υ_2) explained 99% of variation in power of the MLFA-U test.

Figure 3.50 displays estimated power of the PCA test plotted against the magnitude of the random effect signal (υ_2) when $p = 25$. Results for $p = 10$ and $p = 50$ were similar to those

Table 3.39: *Estimated size of the PCA test for* 0 *common factors when* $\gamma = 0$ *and there was one nonlinear random effect in the logistic model with* $\upsilon_2 = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	p		
n	10	25	50
10	0.191	0.039	0.039
25	0.042	0.806	0.051
50	0.039	0.067	1.000

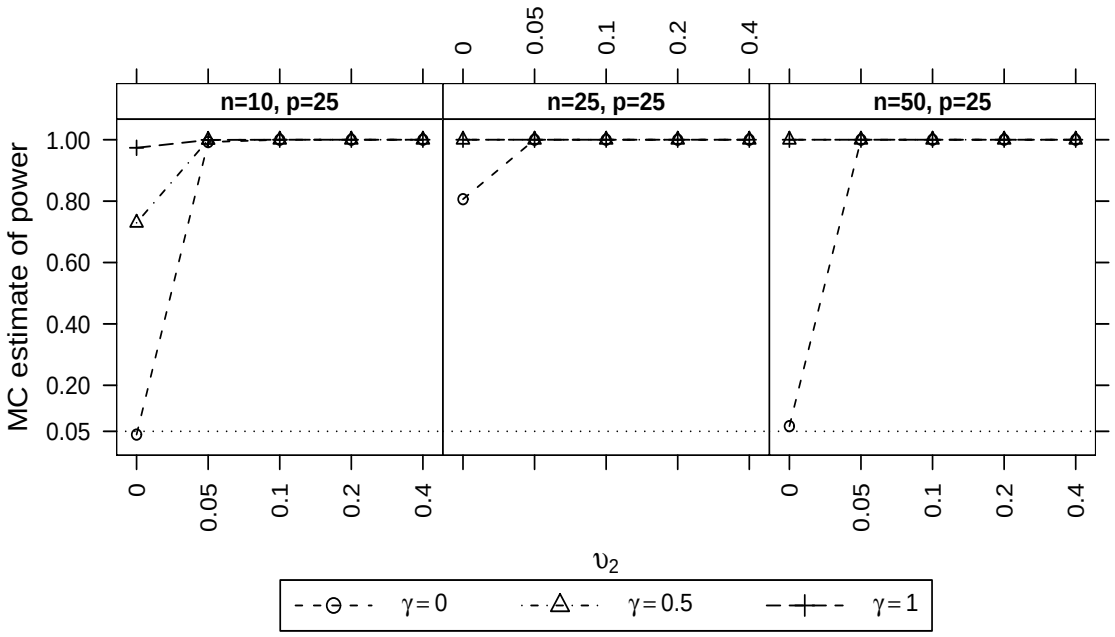

Figure 3.50: *Estimated power of the PCA test for* 0 *common factors when* $p = 25$ *and there was one nonlinear random effect in the logistic model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

presented in the figure. The three panels in the figure show the effect of a different number of subjects n for fixed p. Power to reject the null hypothesis was close to 100% even when the random effect magnitude was small ($\upsilon_2 = 0.05$), and there was little effect of n or p on power when $\upsilon_2 > 0$. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. When $\gamma = 0$ and $n < p$ or $n > p$, the test had size close to the nominal value (0.05) when $\upsilon_2 = 0$ (see Table 3.39). When $n = 50$ and $p = 25$, the size was slightly larger than expected. Clearly when $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on and the PCA test was not useful. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there were no principal components too often.

Figure 3.51 presents the power curves for the MLFA-U test for the cases where $n > p$. The size of the test was close to the nominal 0.05 level except when $n = 50$ and $p = 25$. In this case

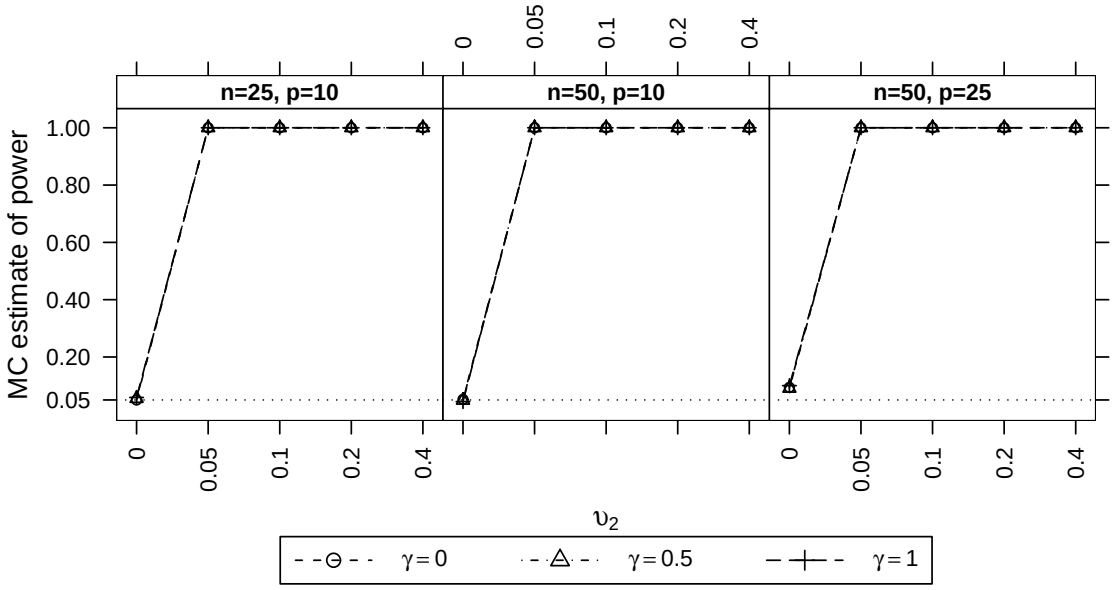

Figure 3.51: *Estimated power of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one nonlinear random effect in the logistic model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

the test rejected the null hypothesis too often (see Table 3.40). The MLFA test had almost 100% power to detect a random effect with υ_2 as small as 0.05, and there was little effect of n, p or γ on power for this amount of random effect signal. When $n > p$ and the within-subject

Table 3.40: *Estimated size of the MLFA-U test for* 0 *common factors when* $n > p$ *and there was one nonlinear random effect in the logistic model with* $\upsilon_2 = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

		γ		
n	p	0.0	0.5	1.0
25	10	0.050	0.055	0.059
50	10	0.052	0.049	0.044
50	25	0.094	0.091	0.100

variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same power.

Test for homogeneous variance

For the likelihood ratio tests comparing the fits of the MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation of the null distribution to be adequate when n is large relative to p. More parameters ($p - 2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ. As the

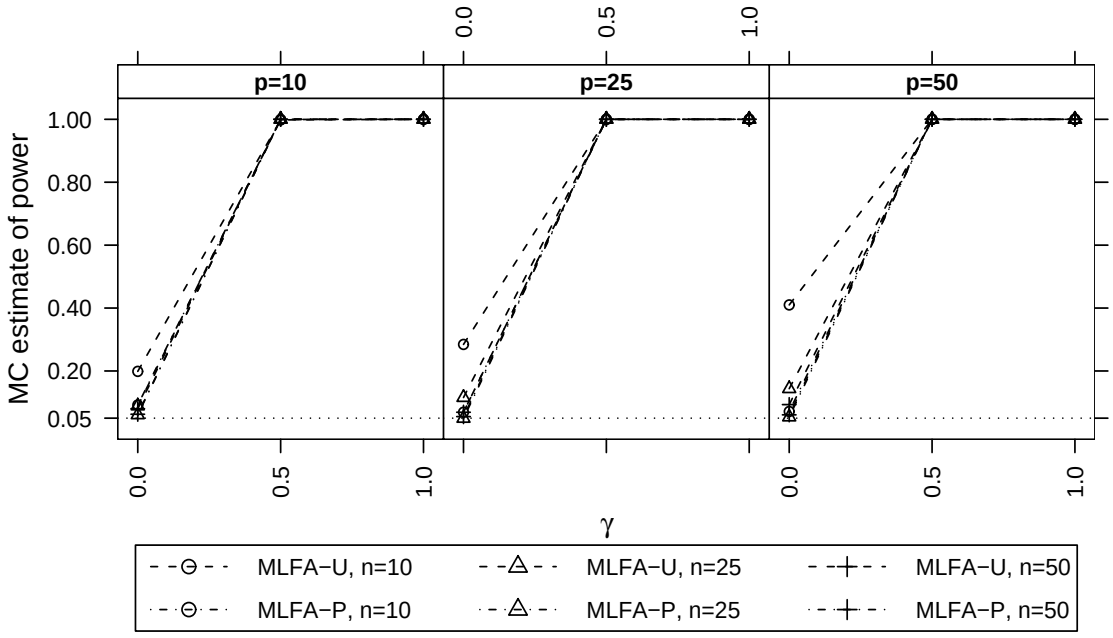

Figure 3.52: *Estimated power of the likelihood ratio tests for H_0: $\gamma = 0$ when there was one nonlinear random effect in the logistic model with $\upsilon_2 = 0.1$. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

magnitude of the nonlinear random effect increases, we expect that the one-factor models will fail to explain between-subject variation adequately. As a result, when υ_2 is large and $\gamma = 0$, we expect the tests to reject the null hypothesis more often than the nominal 0.05 level.

The interaction between γ and the magnitude of the random effect (υ_2) explained 86% of variation in the power of the test comparing the MLFA-H and MLFA-U models. This interaction explained 94% of the variation in power of the test comparing the MLFA-H model with the MLFA-P model. Figure 3.52 presents power for the two tests when the random effect signal was moderate ($\upsilon_2 = 0.1$). When $\gamma \geq 0.5$, power to reject the hypothesis of homoscedasticity by comparing with the one-factor MLFA-U and MLFA-P models was close to 100% in the presence of no random effect or a moderate random effect.

Size in the presence of a small random effect ($\upsilon_2 = 0.05$) and a large random effect ($\upsilon_2 = 0.2$) is presented in Table 3.41 for the two tests. In all cases, the size of the MLFA-U test was

Table 3.41: *Estimated size of the likelihood ratio tests for H_0: $\gamma = 0$ when there was one nonlinear random effect in the logistic model. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. The table reports results for $\upsilon_2 = 0.05$ or 0.2. Nominal size of the test was 0.05. Based on 1600 MC replicates, the standard error of estimated size was approximately 0.005.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\upsilon_2 = 0.05$	$\upsilon_2 = 0.2$	$\upsilon_2 = 0.05$	$\upsilon_2 = 0.2$	$\upsilon_2 = 0.05$	$\upsilon_2 = 0.2$
	MLFA-U					
10	0.204	0.334	0.279	0.401	0.424	0.542
25	0.076	0.563	0.111	0.549	0.147	0.667
50	0.074	0.858	0.068	0.837	0.099	0.927
	MLFA-P					
10	0.094	0.079	0.083	0.074	0.077	0.076
25	0.054	0.036	0.054	0.049	0.062	0.076
50	0.047	0.036	0.059	0.052	0.057	0.121

significantly larger than the nominal level of 0.05, and increased beyond usefulness in the presence of a large random effect signal. The size of the MLFA-P test was close to the nominal value when the random effect was small ($\upsilon_2 = 0.025$) and when $n \geq 25$. For $p = 10$ and 25, the size decreased slightly in the presence of a large random effect, but increased when $p = 50$.

When the random effect was very large ($\upsilon_2 = 0.4$ — these results are not shown in the figure or table), the size of the MLFA-U test was close to 100%, and the power of the MLFA-P test was very much reduced. Although the likelihood for the fitted MLFA-P model should be larger than that for the model that assumes homogeneous variance, there were some cases when $\upsilon_2 = 0.4$ where the likelihood ratio test statistic was negative. Neither test was useful for testing the hypothesis of homoscedasticity in the presence of a large random effect on $\omega^{(2)}$.

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the one-factor power model MLFA-P. The interaction between γ and magnitude of the random effect (υ_2) explained 99% of variation in the bias of $\widehat{\gamma}$. Figure 3.53 shows the mean of $\widehat{\gamma}$ plotted against υ_2 for values of γ in the case where $n = 50$ and $p = 25$. Bias in $\widehat{\gamma}$ was similar for other values of n and p. When the random effect was small, mean bias in $\widehat{\gamma}$ was approximately zero, except when $\gamma = 1$ where $\widehat{\gamma}$ was smaller than the true γ. However as υ_2 increased, bias in $\widehat{\gamma}$ became large for $\gamma = 0.5$ and $\gamma = 1$. In these cases, the maximum likelihood estimator of γ was too small, with the mean of $\widehat{\gamma}$ approaching 0 as υ_2 increased. The underestimation of $\widehat{\gamma}$ when $\gamma = 1$ for large random effect explains why the test for homoscedasticity based on fitting the MLFA-P model failed to reject the hypothesis of homoscedasticity in some cases when $\gamma = 1$.

The magnitude of the random effect υ_2, n, and the two-way interaction between these factors explained 94% of variation in the bias of $\widehat{\sigma}^2$. Figure 3.54 illustrates the pattern in the estimate of σ^2 for $p = 25$ (note the log scale on the y-axis). The pattern was similar for other values of p. When the random effect was small, the estimate of σ^2 was smaller on average than the true

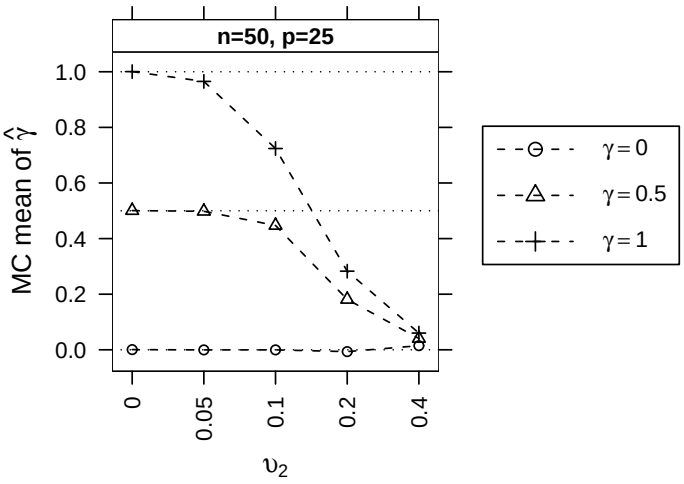

Figure 3.53: *Mean estimate of γ from fitting the one-factor MLFA-P model when there was one nonlinear random effect in the logistic model. The figure plots the case where $n = 50$ and $p = 25$. The true value of γ is given by the dotted horizontal lines. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.002*.*

value of 0.0001. When the the random effect variation was large ($\upsilon_2 \geq 0.2$), the estimate of σ^2 was substantially larger (up to 10-fold larger) than the true value. The three panels in the figure show the effect of increasing n, and the three curves in each figure show the effect of γ. The estimate of σ^2 increased slightly with increasing n and γ.

The over-estimation of σ^2 in the case of a large nonlinear random effect suggests that the one-factor model did not adequately explain between-subject variation when υ_2 was large. A model with two or more factors (to account for the second-order terms in the approximation of the stage-1 model) may have given a better fit and less biased estimates of σ^2 and γ. It is a little puzzling that the estimate of γ tended to be biased toward toward zero when υ_2 was large and $\gamma > 0$, but that the estimate of σ^2 did not appear to be affected to a great extent by the level of γ in the simulation experiment. Since the pseudolikelihood estimator of σ^2 for the MLFA-P model depends on the estimator of γ (equation (C.13)), we would have expected much larger bias in σ^2 when the estimate of γ was significantly smaller than truth. This contrasts

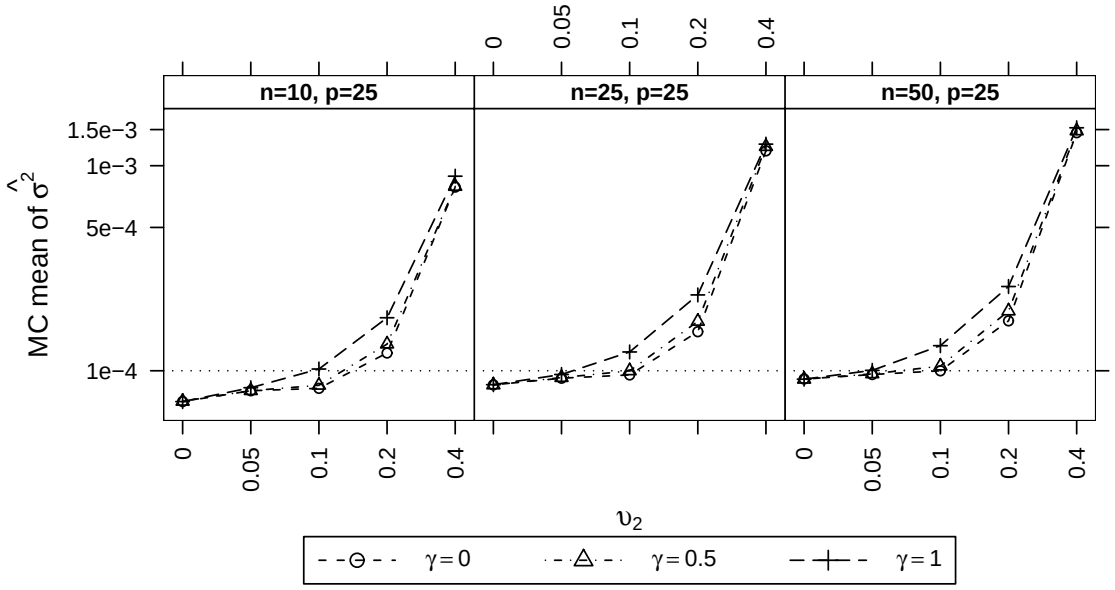

Figure 3.54: *Mean estimate of* σ^2 *from fitting the one-factor MLFA-P model when there was one nonlinear random effect in the logistic model and* $p = 25$*. The true value of* σ^2 *was* $1\text{e}{-4}$*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* $5\text{e}{-7}$*.*

with results for example C where one or both of σ^2 and γ tended to be too large when the nonlinear random effect was large.

Selecting random effects and an appropriate parameterization

We used the exact test where the j-th element of $\boldsymbol{\kappa}$ was the first-order sensitivity function of the actual random effect, evaluated at t_j,

$$\kappa_j \propto f_{\omega^{(2)}}(t_j) \quad = \quad \frac{\alpha(t_j - \omega^{(1)}) \exp\{-\omega^{(2)}(t_j - \omega^{(1)})\}}{[1 + \exp\{-\omega^{(2)}(t_j - \omega^{(1)})\}]^2}.$$

Results for the tests based on the estimated factor loading for the one-factor MLFA-U and MLFA-P models were virtually indistinguishable from those based on the first principal component loading. Since the test is based on the exact distribution of the first principal component

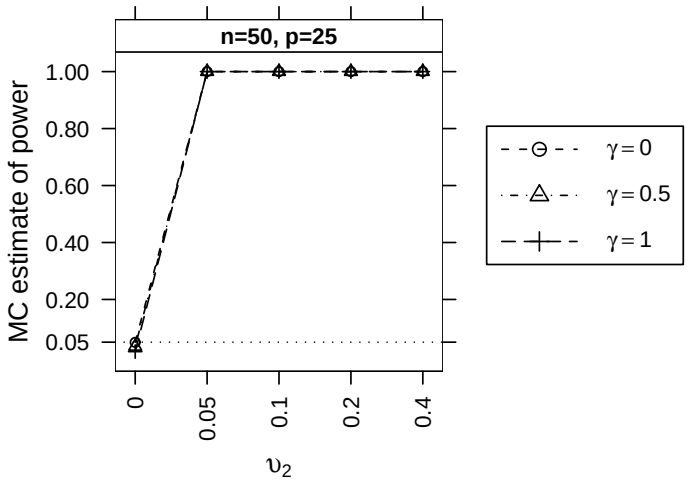

Figure 3.55: *Estimated power of the PCA exact test when there was one nonlinear random effect in the logistic model and* $n = 50$, $p = 25$. *This test rejects* H_0: $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ *in favor of one random effect with sensitivity function* $f_{\omega^{(2)}}$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125.

loading under the null hypothesis, we expect the test to have the nominal size (0.05) when $\upsilon_2 = 0$ and $\gamma = 0$, regardless of the magnitudes of n and p. The power should increase for increasing signal $\upsilon_2 > 0$. Sample size n and number of time points p should have no effect on size of the test when $\gamma = 0$, but may affect power when the alternative hypothesis is true. When $\gamma > 0$, the principal component model that assumes homogeneous within-subject variance is incorrect, and we would expect size and power of the test to deviate somewhat from those in the case where $\gamma = 0$.

The magnitude of the random effect (υ_2) explained over 99% of variation in power of the PCA exact test. Figure 3.55 displays power to reject the null hypothesis in favor of a random effect on $\omega^{(2)}$ when $n = 50$ and $p = 25$. Power was close to 100% for $\upsilon_2 \geq 0.05$. Power and size were similar for other values of n and p. Table 3.42 gives the size of the exact PCA test under the null hypothesis. When $\gamma = 0$ (homoscedastic case) size was close to the nominal 0.05 level regardless of n and p, and size decreased when $\gamma > 0$. This is similar to the case of a single nonlinear random effect in example B where size of the exact test tended to be smaller

Table 3.42: *Estimated size of the PCA exact test for a random effect on* $\omega^{(2)}$ *when there was one nonlinear random effect in the logistic model with* $\upsilon_2 = 0$*. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005*.*

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	0.052	0.026	0.012	0.049	0.037	0.025	0.048	0.046	0.031
25	0.036	0.014	0.006	0.056	0.048	0.021	0.049	0.045	0.026
50	0.048	0.015	0.003	0.049	0.035	0.017	0.054	0.036	0.019

in the presence of heteroscedasticity, and different from example C where size increased in the presence of heteroscedasticity.

We used multivariate forward selection, with the first PCA or factor loading as response variable, to identify the first- and second-order sensitivity functions corresponding to the correct nonlinear random effect from amongst several possible predictors. For parameterization f we considered the following sensitivity functions, including a second-order sensitivity for the parameter $\omega^{(2)}$,

$$\begin{aligned}
f_{\alpha}(t) &= [1 + \exp\{-\omega^{(2)}(t - \omega^{(1)})\}]^{-1} \\
f_{\omega^{(1)}}(t) &= \frac{-\alpha\omega^{(2)} \exp\{-\omega^{(2)}(t - \omega^{(1)})\}}{[1 + \exp\{-\omega^{(2)}(t - \omega^{(1)})\}]^2} \\
f_{\omega^{(2)}}(t) &= \frac{\alpha(t - \omega^{(1)}) \exp\{-\omega^{(2)}(t - \omega^{(1)})\}}{[1 + \exp\{-\omega^{(2)}(t - \omega^{(1)})\}]^2} \\
f_{\omega^{(2)}\omega^{(2)}}(t) &= \frac{-\alpha(t - \omega^{(1)})^2 \exp\{-\omega^{(2)}(t - \omega^{(1)})\}}{[1 + \exp\{-\omega^{(2)}(t - \omega^{(1)})\}]^2} \\
&\quad + \frac{2\alpha(t - \omega^{(1)})^2 \exp\{-2\omega^{(2)}(t - \omega^{(1)})\}}{[1 + \exp\{-\omega^{(2)}(t - \omega^{(1)})\}]^3}.
\end{aligned} \tag{3.10}$$

When $\upsilon_2 > 0$ we would expect the forward selection to pick a model including $f_{\omega^{(2)}}$. As υ_2 becomes large we would expect the second-order sensitivity $f_{\omega^{(2)}\omega^{(2)}}$ to be included too.

The random effect on $\omega^{(2)}$ induced two correlated random effects on parameters $\delta^{(1)}$ and $\delta^{(2)}$

in the alternative parameterization f^* of equation (3.7). We also used multivariate regression to select sensitivities for the parameterization f^*. We included the following sensitivities and second order sensitivities as candidates for the parameterization f^*,

$$\begin{aligned}
f^*_{\alpha} &= \{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^{-1} \\
f^*_{\delta^{(1)}} &= \frac{-\alpha\exp(-\delta^{(2)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^2} \\
f^*_{\delta^{(2)}} &= \frac{\alpha\delta^{(1)}t\exp(-\delta^{(2)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^2} \\
f^*_{\delta^{(1)}\delta^{(1)}} &= \frac{2\alpha\exp(-2\delta^{(2)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^3} \\
f^*_{\delta^{(2)}\delta^{(2)}} &= \frac{2\alpha\delta^{(1)2}t^2\exp(-2\delta^{(2)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^3} - \frac{\alpha\delta^{(1)}t^2\exp(-\delta^{(2)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^3} \\
f^*_{\delta^{(1)}\delta^{(2)}} &= \frac{-2\alpha\delta^{(1)}t\exp(-2\delta^{(3)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^3} + \frac{\alpha t\exp(-\delta^{(2)}t)}{\{1+\delta^{(1)}\exp(-\delta^{(2)}t)\}^2}.
\end{aligned}$$

For $\upsilon_2 > 0$ we would expect the multivariate regression model to include the two sensitivities $f^*_{\delta^{(1)}}$ and $f^*_{\delta^{(2)}}$. As υ_2 becomes large, the second-order sensitivities $f^*_{\delta^{(1)}\delta^{(1)}}$, $f^*_{\delta^{(2)}\delta^{(2)}}$, and $f^*_{\delta^{(1)}\delta^{(2)}}$ may also be included. We distinguished between the appropriateness of the two parameterizations based on the number of sensitivities selected for each, and the trace correlation r_T of the selected models. The more appropriate parameterization should have give a smaller multivariate regression model, and/or larger trace correlation. We expect that a smaller model, or a model with larger trace correlation, should be selected for the true parameterization f than for parameterization f^*, indicating that the true data generating model was based on f.

The multivariate regression results based on the PCA, the MLFA-U and the MLFA-P models were very similar. We present results only for the unrestricted MLFA-U model. The magnitude of the random effect υ_2, the interaction between n and υ_2, and the interaction between p and υ_2 explained 95% of variation in the mean number of sensitivities selected for the true parameterization f. The interaction between the magnitude of the random effect (υ_2) and number of time points (p) explained 96% of variation in the mean number selected for the

alternative parameterization f^*. Since the size of γ did not have a large effect on average model size when $\upsilon_2 > 0$, we present results only for the case where $\gamma = 0.5$ in Figure 3.56.

This figure summarizes the mean number of sensitivities included in the multivariate regression models for parameterizations f (dashed curves) and f^* (dot-dash curves) of the logistic model. As in example C, when $p = 10$ and υ_2 was large, a smaller model was selected for the alternate parameterization f^* than for the true parameterization f (this is contrary to expectations). However, when $p = 25$ or 50, the larger model was generally selected for the alternate parameterization. The effect of n is shown by the different plotting symbols for each parameterization; n did not have a consistent effect on average model size. When the nonlinear

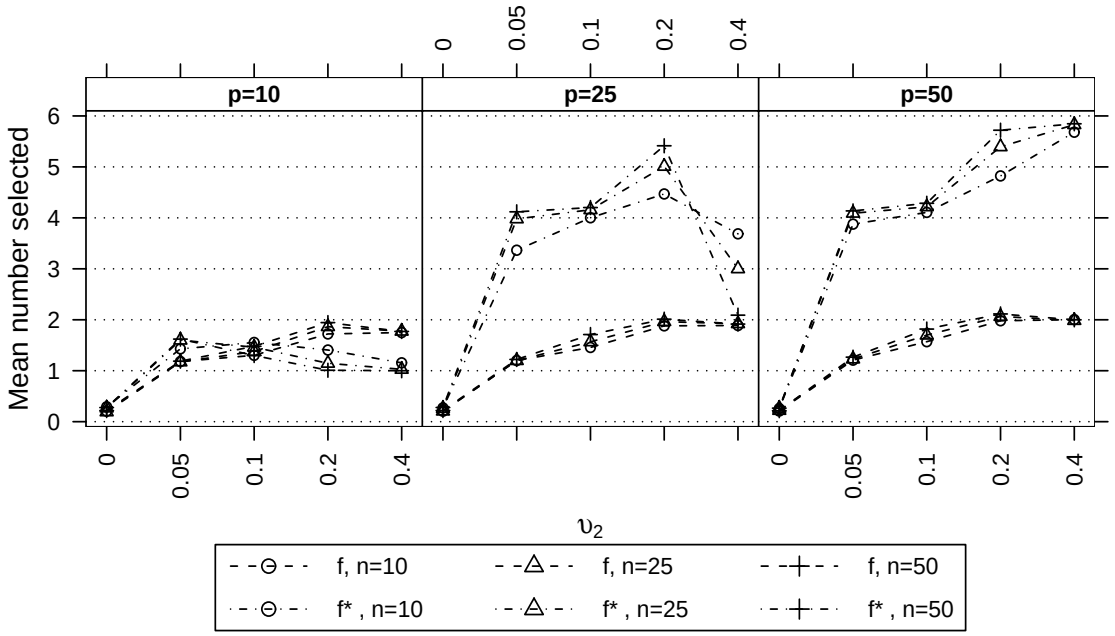

Figure 3.56: *Mean number of sensitivities selected by multivariate forward selection when there was one nonlinear random effect in the logistic model. Results presented are for the true parameterization f (dashed curves) and alternate parameterization f^* (dot-dash curves), and the regression using the factor loading estimated from the one-factor MLFA-U model when $\gamma = 0.5$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.015*.*

Table 3.43: *Mean number of sensitivities included for the true parameterization f by forward selection for the logistic model with one nonlinear random effect with $\upsilon_2 = 0$. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where $\gamma = 0$ or $\gamma = 1$. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each MC mean to be less than* 0.015*.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.187	0.241	0.167	0.201	0.164	0.200
25	0.159	0.196	0.174	0.198	0.153	0.218
50	0.166	0.186	0.145	0.207	0.167	0.210

random effect was small ($\upsilon_2 = 0.05$), in the majority of cases, forward selection picked a single sensitivity for the true parameterization f, and three or more sensitivities for the alternate parameterization f^* when $p = 25$ or 50. As the magnitude of the random effect increased, a second sensitivity was included for f, and additional sensitivities were also included for the parameterization f^* when $p > 10$. When $p = 10$, the average model size selected for parameterization f^* approached one as the random effect became large.

Table 3.43 summarizes the mean number of sensitivities selected for the true parameterization f when there was no random effect ($\upsilon_2 = 0$). When $\gamma = 0$, multivariate forward selection picked a model with one or more sensitivities in about 15% of cases — slightly more often than might be expected by the 0.05 nominal level of the tests at each stage of the forward selection. In the presence of heteroscedasticity ($\gamma = 1$), one or more sensitivity was selected in about 20% of cases.

The random effect magnitude, υ_2, explained 99% of the variation in the trace correlation, r_T, for the multivariate regression model selected for the true parameterization, f. For the alternative parameterization, f^*, the interaction between υ_2 and number of time points (p) explained 99% of the variation in r_T. Figure 3.57 shows the trace correlation for the models

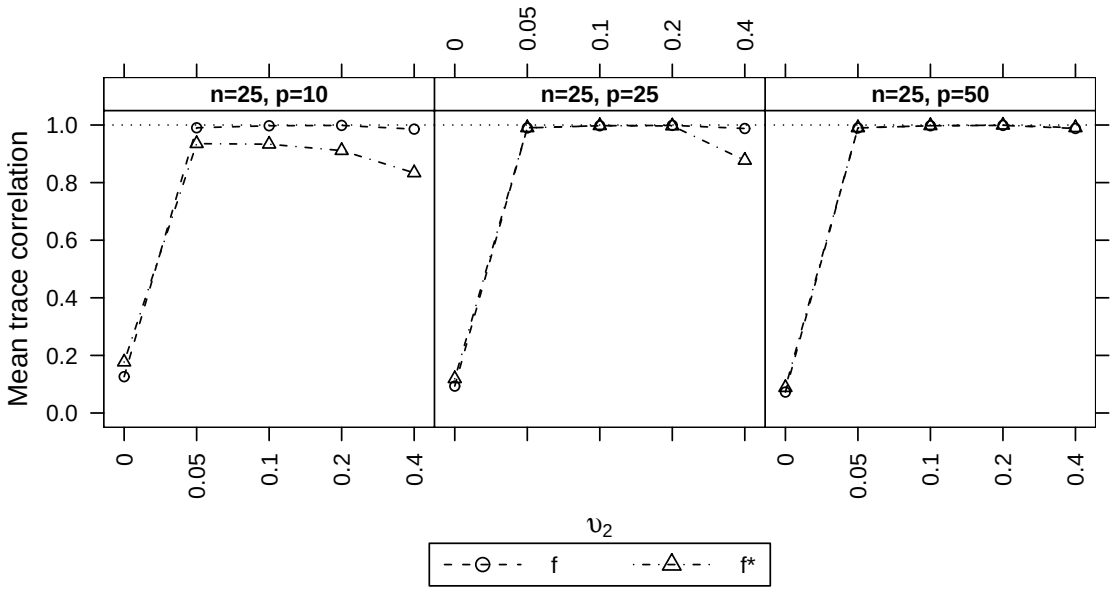

Figure 3.57: *Mean trace correlation for the selected multivariate regression models when there was one nonlinear random effect in the logistic model. Results presented are for the true parameterization f (dashed curves) and alternate parameterization f^* (dot-dash curves), and the regression using the factor loading estimated from the one-factor MLFA-U model when $n = 25$ and $\gamma = 0.5$. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.01*.*

selected for f and f^* based on the MLFA-U analysis when $n = 25$ and $\gamma = 0.5$. Results for other values of n and γ were similar. When $\upsilon_2 > 0$, the mean trace correlation r_T was close to 1 for f. For the alternate parameterization, f^*, the mean trace correlation was smaller than that for the true parameterization when $p = 10$ and $\upsilon_2 > 0$ or when $p = 25$ and $\upsilon_2 = 0.4$. The lower values of r_T correspond to the cases where a smaller model was selected for f^* than for f. As in example C, the results for the anomalous cases where $p = 10$ illustrate how discrimination between parameterizations f and f^* should be based on the size of the selected model as well as some measure (such as r_T) of how well the multivariate regression model explains variation in the principal component or common factor loadings.

Next we examine which sensitivities from the true parameterization f were selected. The interaction between the number of time points (p) and magnitude of the random effect (υ_2)

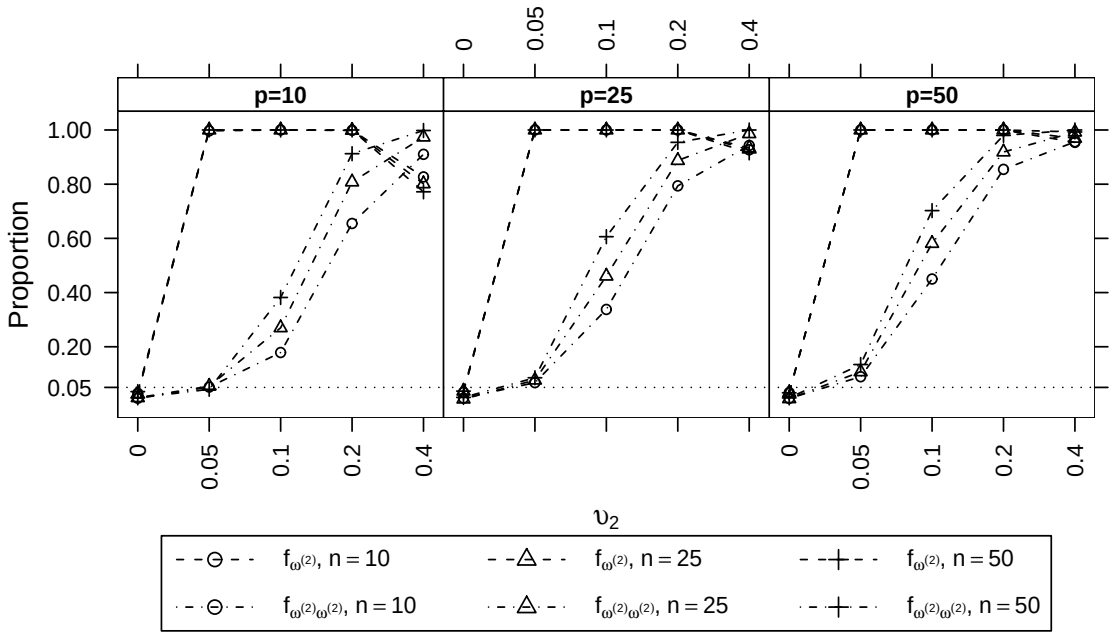

Figure 3.58: *Proportion of cases where multivariate forward selection picked sensitivities corresponding to a random effect on* $\omega^{(2)}$ *in the logistic model with one random effect. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $\gamma = 0.5$*. The dashed curves plot the proportion of times the sensitivity* $f_{\omega^{(2)}}$ *was selected. The dot-dash curves plot the proportion of times the second-order sensitivity* $f_{\omega^{(2)}\omega^{(2)}}$ *was selected. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125*.*

explained 98% of variation in the proportion of times that the sensitivity $f_{\omega^{(2)}}$ was selected, and 92% of variation in the proportion of times the second-order sensitivity $f_{\omega^{(2)}\omega^{(2)}}$ was selected. These sensitivities correspond to the actual random effect on $\omega^{(2)}$. Figure 3.58 displays the proportion of cases where multivariate forward selection picked these sensitivities. The figure only shows the case where $\gamma = 0.5$. Results for other values of γ were similar. When there was no random effect ($\upsilon_2 = 0$), the sensitivity $f_{\omega^{(2)}}$ was selected in less than 5% of cases when within-subject variation was homoscedastic ($\gamma = 0$), and in fewer cases in the presence of heteroscedasticity ($\gamma > 0$) — see Table 3.44.

As shown in Figure 3.58, when $0 < \upsilon_2 < 0.4$, the first-order sensitivity $f_{\omega^{(2)}}$ (dashed curves)

Table 3.44: *Proportion of cases where forward selection picked* $f_{\omega^{(2)}}$ *when* $\upsilon_2 = 0$ *in the experiment with one nonlinear random effect in the logistic model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of the estimated proportion was approximately* 0.005.

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	0.047	0.028	0.018	0.038	0.032	0.019	0.042	0.031	0.016
25	0.024	0.026	0.015	0.051	0.036	0.015	0.038	0.030	0.018
50	0.033	0.034	0.013	0.032	0.036	0.014	0.044	0.030	0.010

was selected in close to 100% of cases, regardless of the magnitudes of n or p. Curiously, when $\upsilon_2 = 0.4$, this sensitivity was selected in fewer cases, but the selection rate increased with p. The second-order sensitivity $f_{\omega^{(2)}\omega^{(2)}}$ (dot-dash curves) was included with increasing frequency as the random effect magnitude became large. This confirms that the factor analysis detected a common factor related to the square of the random effect on $\omega^{(2)}$ when this random effect became large, even though a model with only a single common factor was actually fitted. The three panels in the figure show the effect of increasing p; the selection proportions increased slightly as p increased. Different plotting symbols represent different levels of n. When $\upsilon_2 > 0$, the proportion of times that the second-order sensitivity function was included, increased slightly with increasing n.

The multivariate forward selection procedure worked well for this example. As the random effect became large, the average size of the selected model approached two (Figure 3.56), and the two selected sensitivities were $f_{\omega^{(2)}}$ and $f_{\omega^{(2)}\omega^{(2)}}$, corresponding to the actual random effect on the growth rate parameter $\omega^{(2)}$ (Figure 3.58). One reason that the model selection worked better in example D than in examples B or C, may be that the sensitivities corresponding to the parameters $\omega^{(1)}$ and $\omega^{(2)}$ in the logistic model were almost orthogonal — see Table 3.5 in Section 3.2.5.

3.7.2 Two nonlinear random effects

In this experiment there were two independent random effects for parameterization f of equation (3.6). The first random effect on the growth rate parameter $\omega^{(2)}$ had small standard deviation, $\upsilon_2|\omega^{(2)}|$ where $\upsilon_2 = 0.05$. The second random effect on the half-life parameter $\omega^{(1)}$ had standard deviation $\upsilon_1|\omega^{(1)}|$ where we varied υ_1 from zero to 0.4. The convergence rate for maximum likelihood estimation of the two-factor MLFA-U model was above 95% except when $\upsilon_1 \geq 0.2$, and $n \leq p$. In these cases the convergence rate was sometimes as low as 80%. The convergence rate for the two-factor MLFA-P model was 100%.

The first random effect on $\omega^{(2)}$ was nonlinear, but since we set this random effect to have small magnitude, we expect results for this experiment to be similar to the case of two linear random effects when the magnitude of the second random effect (υ_1) is small, but to play out a little differently when υ_1 is large. In Section 3.7.3 we briefly discuss how results changed when the first nonlinear random effect was large. As in the experiment with one nonlinear random effect, when the second nonlinear random effect becomes large, the principal component and factor analyses may require additional common factors due to the need for second order terms in the approximation to the stage-1 model. We also expect that the multivariate regression analysis will select second-order sensitivities when υ_1 becomes large.

Table 3.45: *Estimated size of the PCA test for* 1 *common factor when* $\gamma = 0$ *and there were two random effects in the logistic model, the second one with* $\upsilon_1 = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

	p		
n	10	25	50
10	0.273	0.042	0.046
25	0.051	0.882	0.083
50	0.051	0.094	1.000

Number of random effects

When $n > p$ or $n < p$ (for the PCA test), we expect the power of the likelihood ratio tests to reject the hypothesis of one common factor, to increase with υ_1 and, since the first non-linear random effect was small, to have size close to the nominal level of 0.05 when $\upsilon_1 = 0$ (one random effect). When $n \approx p$ asymptotic behavior cannot be expected. When $\gamma = 0$ (homoscedastic within-subject variance), the principal component and maximum likelihood factor analysis methods should give similar results. When $\gamma > 0$ we would expect the PCA test to reject the null hypothesis even when there was only one common factor.

The interaction between magnitude of the second random effect (υ_1) and γ explained 74% of variation in the power of the PCA test. The random effect signal υ_1 explained 97% of variation in power of the MLFA-U test. Figure 3.59 displays estimated power of the PCA test plotted against the magnitude of the second random effect signal, υ_1, when $p = 25$. Results for $p = 10$

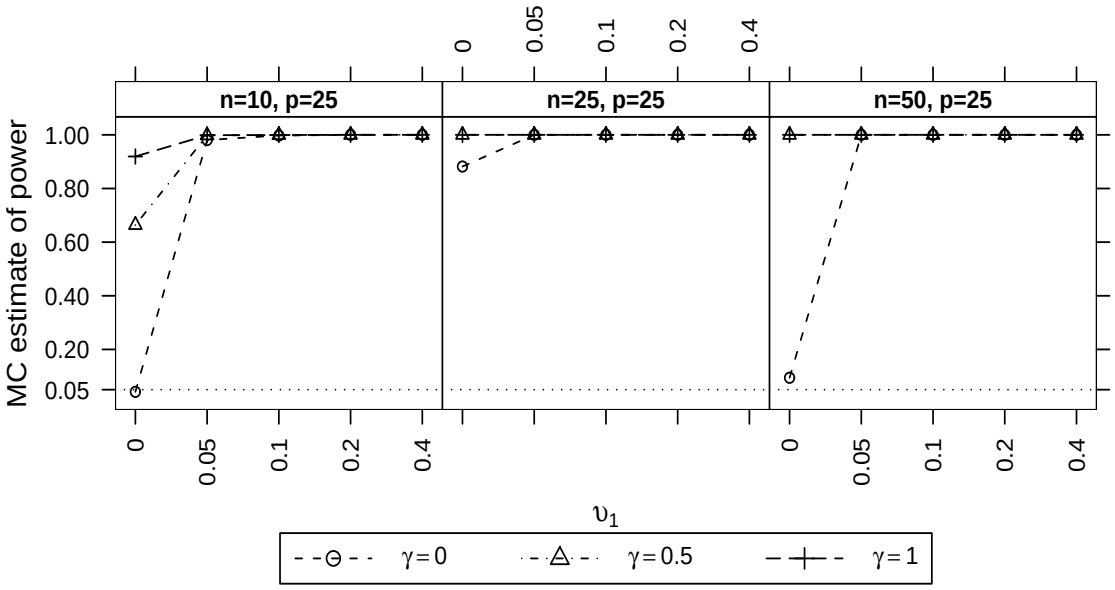

Figure 3.59: *Estimated power of the PCA test for* 1 *common factor when* $p = 25$ *and there were two random effects in the logistic model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

and $p = 50$ were similar to those presented in the figure. The three panels in the figure show the effect of a different number of subjects n for fixed p. The three curves in each panel represent power for different amounts of within-subject heteroscedasticity: $\gamma = 0$ (homoscedastic within-subject variance), $\gamma = 0.5$, or $\gamma = 1$. As anticipated, when $\gamma > 0$, the PCA test rejected the hypothesis that there was one principal component too often and would not be useful in the presence of heteroscedastic within-subject variance.

Table 3.45 gives estimated size of the PCA test at $\upsilon_1 = 0$ when $\gamma = 0$ for different values of n and p. When n was much larger than p, or p was much larger than n, then test had approximately the correct size (0.05) when $\gamma = 0$. When the ratio of n to p was closer to one, the size was larger than the nominal value. Clearly when $n = p$ (center panel in the figure and diagonal in the table), the large-p asymptotics cannot be leaned on and the PCA test was not useful. When $\gamma = 0$ and $n < p$ or $n > p$, the PCA test had close to 100% power to detect a second random effect with $\upsilon_1 = 0.05$.

Figure 3.60 presents the power curves for the MLFA-U test for the cases where $n > p$. The size of the test was close to the nominal 0.05 level except when $n = 50$ and $p = 25$ or $\gamma = 1$, where the size was larger than the nominal level (see Table 3.46). When $n > p$, the MLFA-U test had close to 100% power to detect a second random effect with $\upsilon_1 = 0.05$ or larger. When

Table 3.46: *Estimated size of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two random effects in the logistic model, the second one with* $\upsilon_C = 0$. *Nominal size of the test was* 0.05. *Based on* 1600 *MC replicates, the standard error of estimated size was approximately* 0.005.

		γ		
n	p	0.0	0.5	1.0
25	10	0.051	0.063	0.061
50	10	0.051	0.058	0.080
50	25	0.084	0.099	0.439

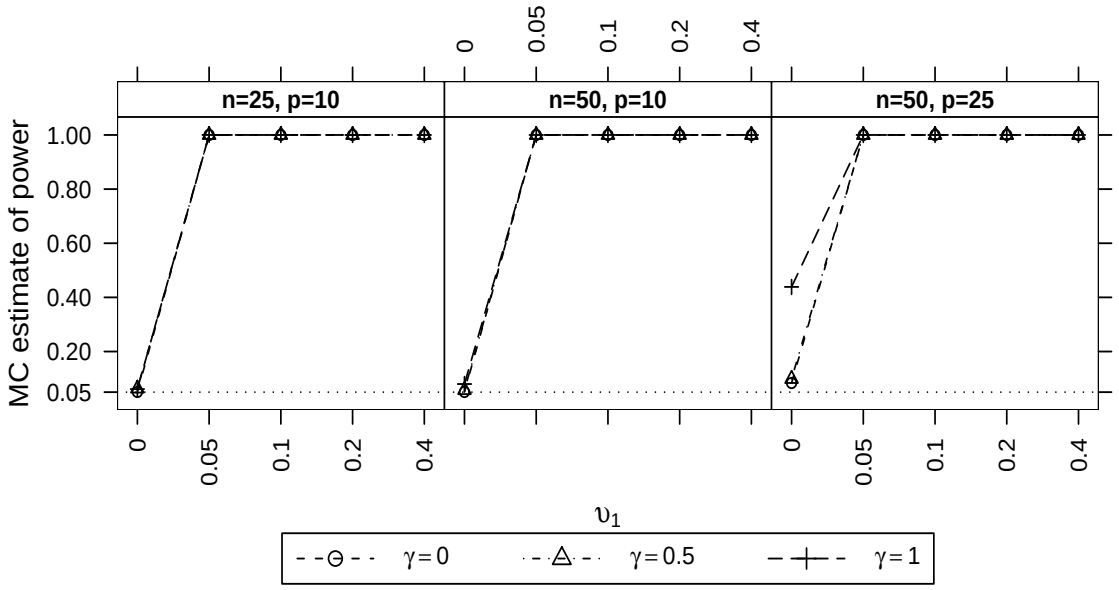

Figure 3.60: *Estimated power of the MLFA-U test for* 1 *common factor when* $n > p$ *and there were two random effects in the logistic model. Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

$n > p$ and within-subject variation was homoscedastic ($\gamma = 0$), the PCA and MLFA-U tests had approximately the same size and power.

Test for homogeneous variance

For the likelihood ratio tests comparing the fits of the MLFA-H, MLFA-U, and MLFA-P models, we expect the chi-squared approximation to be adequate when n is large relative to p. More parameters ($p-2$ more) are estimated in fitting the MLFA-U model than in fitting the MLFA-P model, and so we expect that the test based on the MLFA-P model may have size closer to the nominal level of 0.05 than the test based on the MLFA-U model. Power to reject the hypothesis of homoscedasticity should increase with increasing γ.

The interaction between the magnitude of γ and the random effect (υ_1) explained 80% of the variation in power of the test comparing the fits of the MLFA-H and MLFA-U models. The $\gamma \times \upsilon_1$ interaction, and the number of time points (p) explained 93% of the variation in the

power of the test comparing the MLFA-H model with the MLFA-P model. Figure 3.52 presents power for the two tests when the random effect signal was small ($\upsilon_1 = 0.05$). When the random effect was small and $\gamma \geq 0.5$, power to reject the hypothesis of homoscedasticity using either test was close to 100%.

Table 3.47 presents size of the tests for small random effect ($\upsilon_1 = 0.05$) and moderate random effect ($\upsilon_1 = 0.1$). In all cases, the size of the test comparing the homoscedastic model to the unrestricted model MLFA-U was significantly larger than the nominal level of 0.05, making this test close to useless. Size of both tests was increased by a moderate random effect signal. The test comparing the homoscedastic model to the power model MLFA-P performed slightly better when the random effect on $\omega^{(1)}$ was small, although its size was still a little higher than the nominal value, except when n and p were 25 or larger.

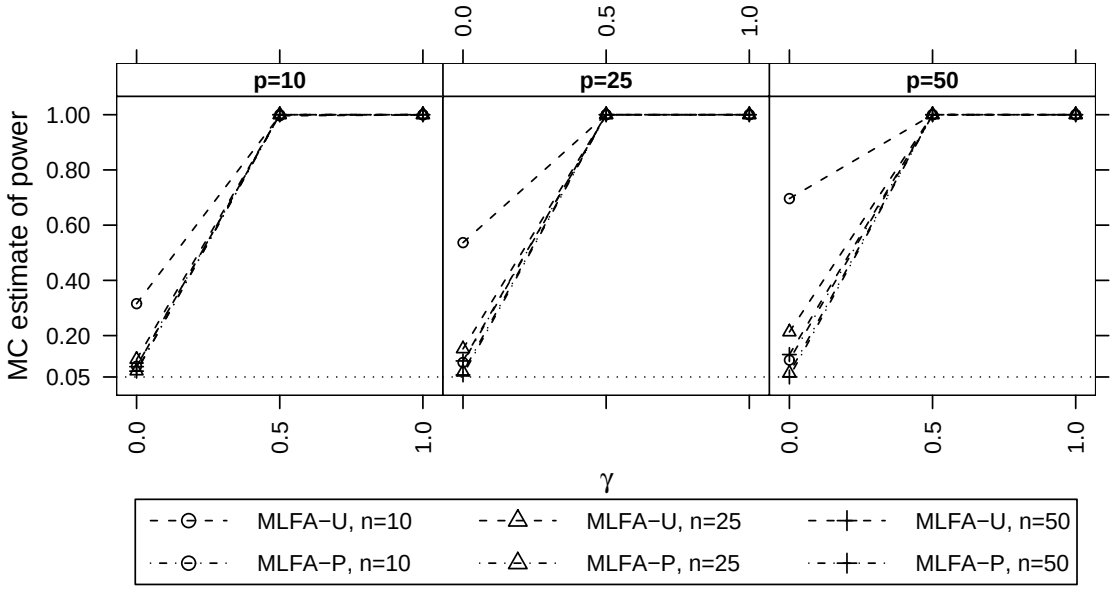

Figure 3.61: *Estimated power of the likelihood ratio tests for* H_0*:* $\gamma = 0$ *when there were two random effects in the logistic model and* $\upsilon_1 = 0.05$*. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U (dashed curves). The second test compared the homoscedastic model with the power model MLFA-P (dot-dashed curves). Nominal size of the test was* 0.05*. Based on* 1600 *MC replicates, the standard error of estimated power was less than* 0.0125*.*

Table 3.47: *Estimated size of the likelihood ratio tests for H_0: $\gamma = 0$ when there were two nonlinear random effects in the logistic model. The first test compared the homoscedastic factor model with the unrestricted model MLFA-U. The second test compared the homoscedastic model with the power model MLFA-P. The table reports results for $\upsilon_1 = 0.05$ or 0.1. Nominal size of the test was 0.05. Based on 1600 MC replicates, the standard error of estimated size was approximately 0.005.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\upsilon_1 = 0.05$	$\upsilon_1 = 0.1$	$\upsilon_1 = 0.05$	$\upsilon_1 = 0.1$	$\upsilon_1 = 0.05$	$\upsilon_1 = 0.1$
	MLFA-U					
10	0.316	0.350	0.537	0.534	0.696	0.700
25	0.114	0.239	0.152	0.297	0.213	0.386
50	0.087	0.419	0.108	0.416	0.131	0.529
	MLFA-P					
10	0.089	0.119	0.102	0.128	0.111	0.139
25	0.072	0.077	0.068	0.102	0.064	0.150
50	0.071	0.082	0.058	0.121	0.050	0.165

As was the case in the experiment with a single random effect, the tests for homoscedasticity were not well-behaved in the presence of a large random effect. When the random effect was large ($\upsilon_1 = 0.4$), the size of the MLFA-U test was close to 100%, size of the MLFA-P test was greater than 50%, and power of the MLFA-P test was very much reduced (these results are not shown in the figure or table). Although the likelihood for the fitted MLFA-P model should be larger than that for the model that assumes homogeneous variance, there were some cases when $\upsilon_1 = 0.4$ where the likelihood ratio test statistic was negative. Neither test was useful for testing the hypothesis of homoscedasticity in the presence of a large random effect on $\omega^{(1)}$. We expect that the size of the MLFA-P test would have been closer to the nominal level if we had fitted a factor model with more than two factors when the random effect variation was moderate or large.

Estimating γ and σ^2

We estimated the parameters γ and σ^2 of the within-subject variance model $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ by fitting the two-factor power model MLFA-P. Patterns in the bias of $\widehat{\gamma}$ and $\widehat{\sigma}^2$ for this experiment were similar to results for the experiment with a single random effect on parameter $\omega^{(2)}$. The interaction between γ and random effect magnitude (υ_1), the interaction between p and υ_1, and the number of subjects (n) explained 79% of variation in bias of $\widehat{\gamma}$ (all simulation factors and their interactions explained 85% of variation). Figure 3.62 shows the mean of $\widehat{\gamma}$ plotted against υ_1 for different values of γ in the case where $n = 50$ and $p = 25$. Bias in $\widehat{\gamma}$ was similar for other values of n and p. When the random effect was small, mean bias in $\widehat{\gamma}$ was approximately zero. However as υ_1 increased, bias in $\widehat{\gamma}$ became large, especially for $\gamma = 0.5$ and $\gamma = 1$. The maximum likelihood estimator of γ was too small in the presence of a large random effect, with the mean of $\widehat{\gamma}$ approaching 0 or negative values as υ_1 increased. The underestimation of $\widehat{\gamma}$ when $\gamma = 1$ for $\upsilon_1 = 0.4$ explains why the test for homoscedasticity based on comparing the MLFA-H

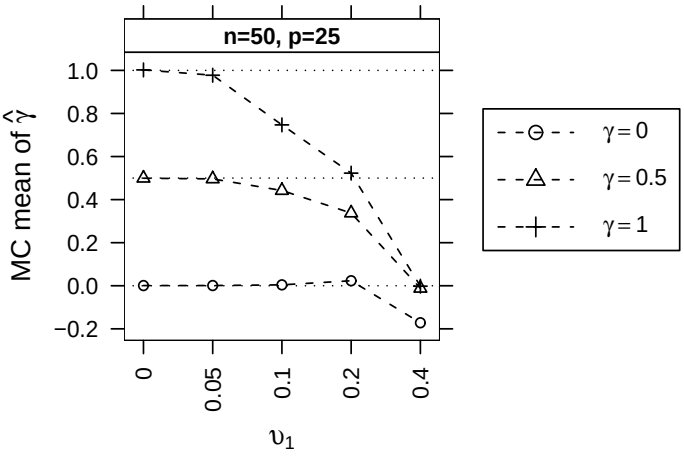

Figure 3.62: *Mean estimate of γ from fitting the two-factor MLFA-P model when there were two random effects in the logistic model. The figure plots the case where $n = 50$ and $p = 25$. The true value of γ is given by the dotted horizontal lines. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.005*.*

and MLFA-P models failed to reject the hypothesis of homoscedasticity in some cases when $\gamma = 1$ and the random effect variation was large.

The magnitude of γ, and the interaction between the number of subjects (n) and magnitude of the random effect (υ_1), explained 90% of variation in bias of the estimate of σ^2. Figure 3.63 illustrates the pattern in the estimate of σ^2 for $p = 25$ (note the log scale on the y-axis). The estimate of σ^2 was smaller on average than the true value of 0.0001 when υ_1 was small, but rapidly became larger on average than the true value as υ_1 increased. The three panels in Figure 3.63 show the effect of increasing n for fixed p; the estimate of σ^2 tended to be larger for larger n. The three curves in each panel represent the mean estimate of σ^2 for each setting of γ. The estimate of σ^2 increased slightly with γ when $\upsilon_1 > 0$.

The over-estimation of σ^2 in the case of a large nonlinear random effect suggests that the two-factor model did not adequately explain between-subject variation when υ_1 was large. A

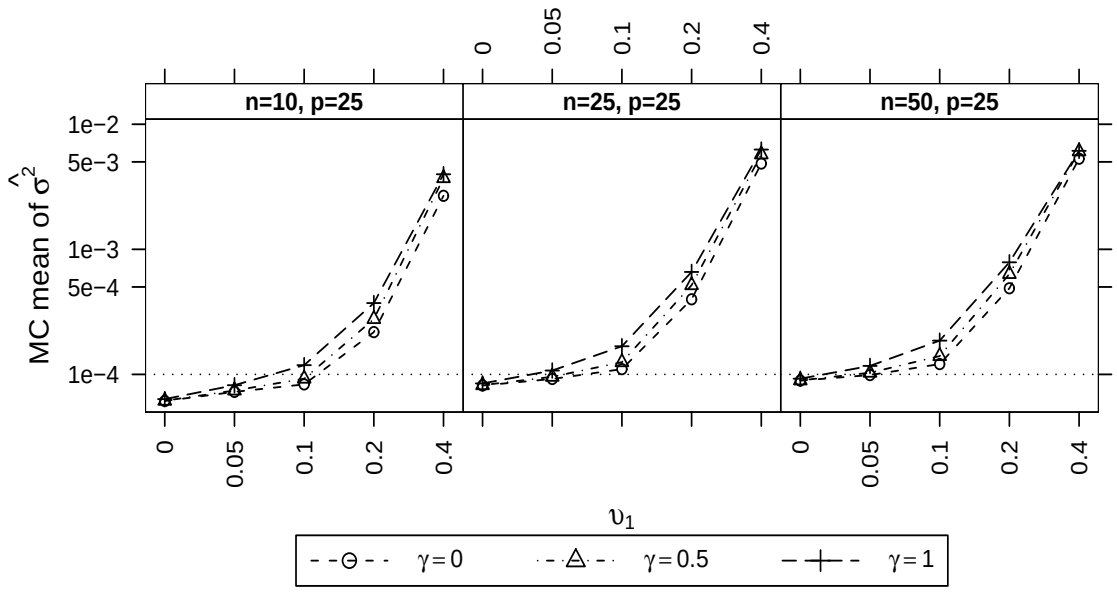

Figure 3.63: *Mean estimate of σ^2 from fitting the two-factor MLFA-P model when there were two random effects in the logistic model and $p = 25$. The true value of σ^2 was* 1e−4*. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 5e−7*.*

model with three or more factors (to account for the second-order terms in the approximation of the stage-1 model) may have given a better fit and less biased estimates of σ^2 and γ. As in the previous experiment where there was a single nonlinear random effect, it is a little puzzling that the estimate of γ had such large negative bias when υ_1 was large, but the estimate of σ^2 did not appear to be affected to a great extent by the level of γ.

Selecting random effects

We used multivariate forward selection, with the first two PCA or factor loadings as multivariate response variables, to identify the first- and second-order sensitivity functions corresponding to the correct nonlinear random effects from amongst several possible predictors. For parameterization f we considered the sensitivity functions f_α, $f_{\omega^{(1)}}$, $f_{\omega^{(2)}}$, $f_{\omega^{(1)}\omega^{(1)}}$, $f_{\omega^{(2)}\omega^{(2)}}$, and $f_{\omega^{(1)}\omega^{(2)}}$ that were given in equation (3.10).

When $\upsilon_1 = 0$ the forward selection should select a model with one sensitivity, $f_{\omega^{(2)}}$, corresponding to the random effect on $\omega^{(2)}$. As υ_1 gets larger, the forward selection should additionally select the sensitivity $f_{\omega^{(1)}}$ in a high proportion of cases. When the nonlinear random effect is large, the two-factor model may not adequately capture the between-subject variation in the data because of second-order terms in the approximation to the stage-1 model. However, we would still expect the forward selection procedure based on two common factors to pick the second-order sensitivity $f_{\omega^{(1)}\omega^{(1)}}$ and possibly $f_{\omega^{(1)}\omega^{(2)}}$ when υ_1 is large.

The multivariate regression results based on the PCA, MLFA-U and the MLFA-P models were almost indistinguishable from each other across all the simulation scenarios. We present results only for the MLFA-U analysis. The number of subjects (n), and interaction between the magnitude of the random effect (υ_1) and number of time points (p) explained 97% of variation in the mean number of sensitivities selected. Figure 3.64 summarizes the mean number of sensitivities included in the multivariate regression model when $\gamma = 0.5$. When $\upsilon_1 = 0$, the mean model size was approximately 1.25 (see Table 3.48), increasing slightly with p or in the presence of within-subject heteroscedasticity. As the magnitude of υ_1 increased, additional

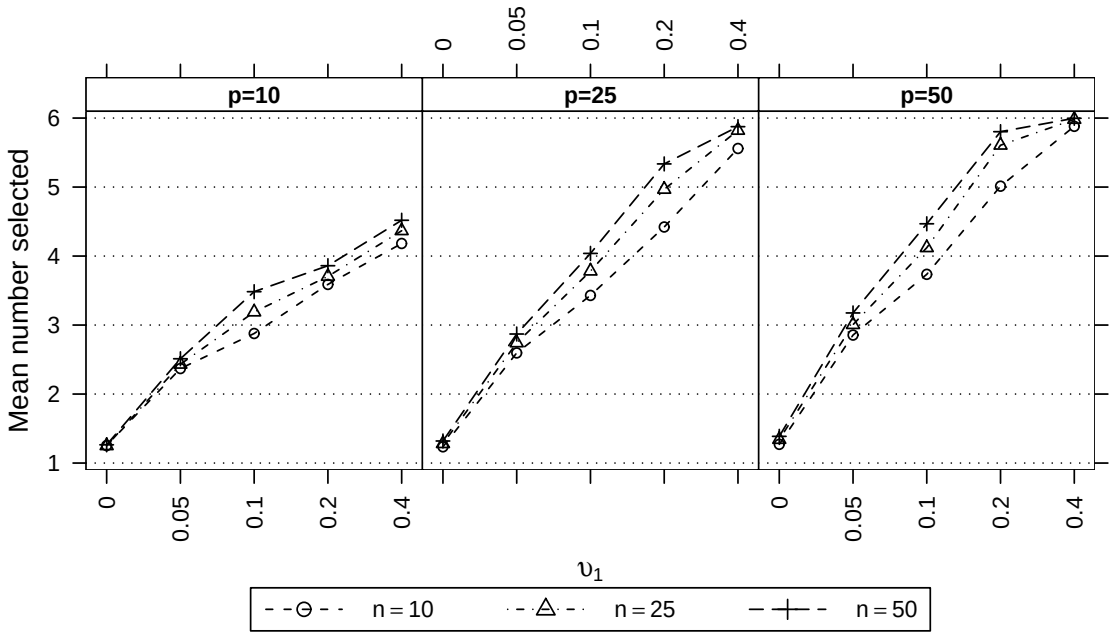

Figure 3.64: *Mean number of sensitivities selected by multivariate forward selection when there were two random effects in the logistic model. Results presented are for the regression using the factor loadings estimated from the two-factor MLFA-U model where* $\gamma = 0.5$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.02*.*

sensitivities were included into the model. The three panels in the figure show the effect of increasing number of time points, p; as p increased, the average model size increased. The three curves in each panel show the effect of increasing number of subjects, n; the number of sensitivities selected increased with increasing n. When $\upsilon_1 = 0.4$, all six candidate sensitivities were included in the model in a high proportion of cases when n and p were large.

The interactions $n \times \upsilon_1$ and $p \times \upsilon_1$ explained most of the variation in the proportion of cases where the sensitivity $f_{\omega^{(2)}}$ (corresponding to the first random effect on $\omega^{(2)}$), and the sensitivities $f_{\omega^{(1)}}$, $f_{\omega^{(1)}\omega^{(1)}}$, and $f_{\omega^{(1)}\omega^{(2)}}$ (the first- and second-order sensitivities corresponding to the random effect on $\omega^{(1)}$) were included in the multivariate regression model. For $f_{\omega^{(2)}}$, these two interactions explained 52% of variation (all simulation factors explained 56% of the

Table 3.48: *Mean number of sensitivities included by forward selection when there were two random effects in the logistic model with* $\upsilon_1 = 0$*. Results presented are for the regression using the factor loading estimated from the two-factor MLFA-U model. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each MC mean to be less than* 0.02*.*

	$p = 10$			$p = 25$			$p = 50$		
n	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$	$\gamma = 0$	$\gamma = \frac{1}{2}$	$\gamma = 1$
10	1.201	1.253	1.239	1.204	1.236	1.232	1.229	1.271	1.266
25	1.196	1.251	1.260	1.231	1.281	1.270	1.252	1.342	1.343
50	1.212	1.265	1.239	1.239	1.320	1.326	1.295	1.387	1.499

variation). For $f_{\omega^{(1)}}$, $f_{\omega^{(1)}\omega^{(1)}}$, and $f_{\omega^{(1)}\omega^{(2)}}$, the two interactions explained, respectively, 99%, 88%, and 95% of variation in the proportion of cases where the sensitivities were selected.

Figure 3.65 displays the proportion of cases where multivariate forward selection picked the sensitivities corresponding to random effects on $\omega^{(2)}$ and $\omega^{(1)}$. We plotted results for the case where $n = 25$ and $\gamma = 0.5$. Selection rates were slightly higher when n was larger, and did not differ much with γ. The sensitivity $f_{\omega^{(2)}}$ corresponding to the small nonlinear random effect on $\omega^{(2)}$ was included in the multivariate regression model in greater than 65% of cases (dashed curves with open circles). The proportion of times this sensitivity was included increased with increasing p and n, but varied with υ_1, depending on which other sensitivities were included in the model. The first-order sensitivity $f_{\omega^{(1)}}$ corresponding to the random effect on $\omega^{(1)}$ was included in almost 100% of cases when $\upsilon_1 > 0$ (dot-dash curves with open triangles). The second-order sensitivity $f_{\omega^{(1)}\omega^{(2)}}$ was included in an increasing proportion of cases as υ_1 increased (long-dash curves with crosses). The sensitivity $f_{\omega^{(1)}\omega^{(1)}}$ was picked less often, but its selection rate increased with increasing υ_1 and p (solid curves with plus signs). When there was only one random effect on $\omega^{(2)}$ ($\upsilon_1 = 0$), the sensitivity $f_{\omega^{(1)}}$ was selected in about 5% of cases — see Table 3.49.

This experiment also exemplifies how the interpretation of the multivariate regression results

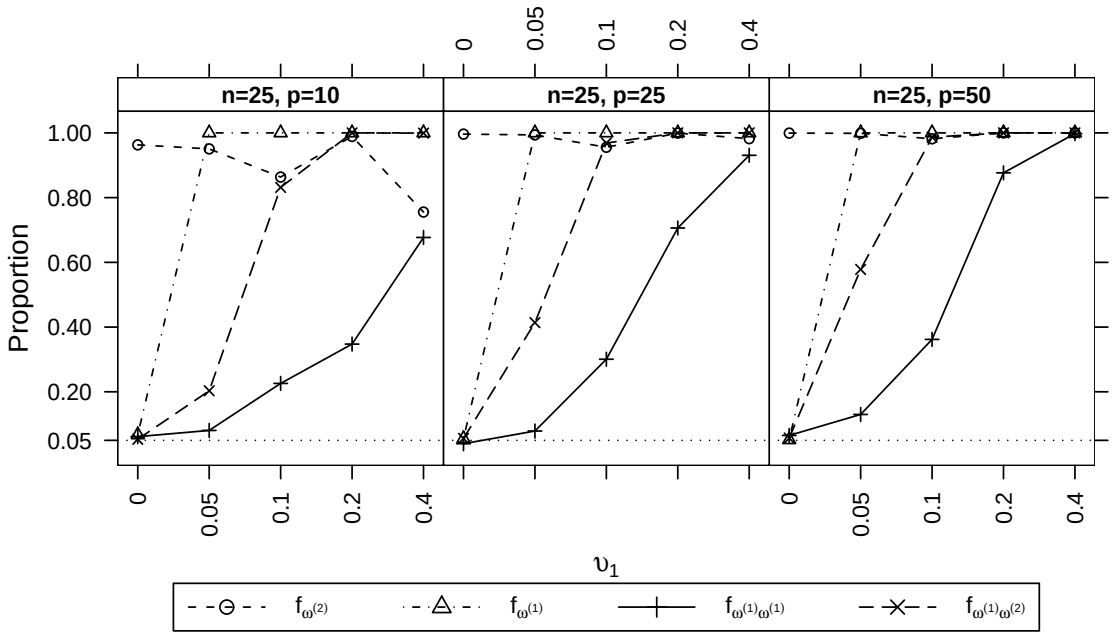

Figure 3.65: *Proportion of cases where multivariate forward selection picked the sensitivities,* $f_{\omega^{(2)}}$ *(dashed curve),* $f_{\omega^{(1)}}$ *(dot-dashed curves),* $f_{\omega^{(1)}\omega^{(1)}}$ *(solid curves), and* $f_{\omega^{(1)}\omega^{(2)}}$ *(long-dash curves) when there were two random effects in the logistic model. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model where* $n = 25$ *and* $\gamma = 0.5$*. At each step of the forward selection the* 0.05 *critical value of the test statistic was used. Based on* 1600 *MC replicates, the standard error of each plotted proportion was less than* 0.0125*.*

may be complicated when there is more than one nonlinear random effect. As discussed in example C, the candidate sensitivities (especially when second-order sensitivities are included) are likely to be highly correlated or non-orthogonal. In this experiment, when p was large, a model including all six candidate sensitivities was picked in a good number of cases when $\upsilon_1 = 0.4$ (Figure 3.64), although there was no random effect on the asymptote parameter α. This contrasts with the one-effect experiment for example D, where only the two correct sensitivities were included, even when random effect signal was large. We conclude that the inclusion of f_α was due to the presence of a large nonlinear random effect on parameter $\omega^{(1)}$, and non-orthogonality of the candidate sensitivities. As a remedy, we would recommend that

Table 3.49: *Proportion of cases where forward selection picked $f_{\omega^{(1)}}$ when there were two random effects in the logistic model with $\upsilon_1 = 0$. Results presented are for the regression using the factor loading estimated from the one-factor MLFA-U model with $\gamma = 0$ and $\gamma = 1$. At each step of the forward selection the 0.05 critical value of the test statistic was used. Based on 1600 MC replicates, the standard error of the estimated proportion was approximately 0.005.*

	$p = 10$		$p = 25$		$p = 50$	
n	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$	$\gamma = 0$	$\gamma = 1$
10	0.048	0.059	0.039	0.046	0.039	0.044
25	0.045	0.052	0.039	0.038	0.031	0.035
50	0.048	0.044	0.044	0.034	0.036	0.036

the order in which sensitivities are included by the forward selection procedure should be taken into account when deciding which parameters truly require non-trivial random effects.

3.7.3 Two nonlinear random effects, the first one being large

Before running the experiment described in the previous section, we ran a similar experiment, with random effects on the growth rate parameter $\omega^{(2)}$ and the half-life parameter $\omega^{(1)}$, except that the magnitude of the first random effect on $\omega^{(2)}$ was large ($\upsilon_2 = 0.4$). When applying our methods to data generated in this experiment, we fitted models with two principal components or common factors. However the large nonlinear random effect on $\omega^{(2)}$ mostly likely required second-order terms in the Taylor series approximation of the stage-1 model. As was the case in example C, the results of this experiment did not give much information about the performance of our methods, except that a two-factor model was inadequate to capture the between-subject variation induced by two large nonlinear random effects.

In the experiment with a large random effect on $\omega^{(2)}$ and random effect on $\omega^{(1)}$ with varying magnitude, the likelihood ratio tests for number of random effects rejected the null hypothesis that there was one linear latent variable in close to 100% of cases even when the magnitude of the second random effect, υ_1, was zero. The tests for homoscedastic within-subject variance

had size much larger than the nominal 0.05 level in the presence of a moderate random effect on $\omega^{(1)}$ ($\upsilon_1 = 0.1$) due to unexplained between-subject variation after fitting two common factors. The estimates of γ and σ^2 were much larger (between 10- and 100-fold larger) than truth for most values of υ_1. More than three sensitivities were included in the multivariate models chosen by the forward selection procedure even when $\upsilon_1 = 0$, due to the second-order behavior of the large random effect on $\omega^{(2)}$. All six of the candidate sensitivities were included when $\upsilon_1 > 0$ and n and p were larger than 25.

If we were to redo such an experiment, we would ensure that sufficient principal components or common factors were included in the multivariate models to adequately describe the between-subject variation, before proceeding with the various diagnostic tests based on the fit of these models.

3.8 Summary of results

In example A where the random effects were linear, our methods performed as anticipated. Results for the experiment with two linear random effects were similar to those for the experiment with one linear random effect, with slightly smaller power when there were two random effects. When there were k linear random effects, the MLFA likelihood ratio test of the hypothesis that $k-1$ common factors are adequate, had approximately the correct size, and had high power to detect a k-th random effect with moderate magnitude when $n > p$. The PCA likelihood ratio test had acceptable size and power when within-subject variation was homoscedastic and $n > p$ or $n < p$. The PCA test was not useful (its size was too large) when $n \approx p$, or in the presence of within-subject heteroscedasticity. For general practice we recommend the (finite sample corrected) MLFA likelihood ratio test when $n > p$. We would only recommend the PCA test if within-subject variation was homoscedastic, or if the test could be applied on a transformed scale that would ensure homoscedastic within-subject variation.

For testing the hypothesis of within-subject homoscedasticity, the likelihood ratio test com-

paring the homoscedastic factor model (MLFA-H) to the power model (MLFA-P) had acceptable size when n and p were large, and had high power to discriminate between the variance models $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$ and $\mathbb{V}(\varepsilon_{ij}) = \sigma^2\mu_j^{2\gamma}$ when $\gamma \geq 0.5$. A finite sample correction for this test to improve the size under the null hypothesis would be useful. The test comparing the MLFA-H model to the unrestricted factor model (MLFA-U) was not practical, rejecting the null hypothesis far too often when $\gamma = 0$. When a k-factor MLFA-P model was fitted to data generated with k linear factors, the estimate of γ was unbiased on average, and the estimate of σ^2 was too small, with bias diminishing with increasing n. We recommend use of the test of homoscedasticity based on fitting the MLFA-P model (or some other model that describes within-subject variation with a small number of parameters), but caution that size of the test may be larger than the nominal level unless n and p are large.

In the experiment with a single linear random effect, the exact test of the hypothesis that $\mathbf{\Sigma} = \sigma^2\mathbf{I}_p$ had the correct size when within-subject variation was homoscedastic, and had high power to detect the true random effect. In the presence of heteroscedasticity, the size of the test deviated somewhat from the nominal value (sometimes it was smaller than 0.05 and sometimes larger), making this test unpredictable if within-subject variation was heteroscedastic. Although the exact test was derived based on the first principal component loading, it behaved similarly when the first factor loading was estimated from fitting the MLFA-U or MLFA-P models, to how it performed when the test statistic was based on the first PCA loading. The multivariate forward selection procedure worked as anticipated, picking the sensitivity functions corresponding to the true random effects in a high proportion of cases when random effect signal was moderate to large. We recommend the multivariate forward selection procedure for identifying non-trivial random effects (random effects with non-zero variance).

It is worth cautioning that the range over which random effect standard deviation was varied in example A resulted in a very modest amount of between-subject variation (see Figure 3.2), but our methods were very sensitive to the presence of these random effects. In practice, we would advise being mindful that when n and p are large, the methods may identify random

effects that are statistically significant but lack much practical importance.

Examples B, C and D tested the performance of our methods when random effects were nonlinear. The main differences from results in example A can be explained by the need for higher order (e.g., quadratic) terms in the Taylor series approximation of the stage-1 model in the presence of a large nonlinear random effect. What this means in practice, is that when there are k nonlinear random effects, a factor model or principal component model with more than k common factors will need to be fitted in order to adequately capture between-subject variation induced by the nonlinear random effects.

This obviously has an impact on interpreting results of the likelihood ratio tests for adequacy of the k-factor model. Recall that these tests assume that the common factors are linear. When a k-factor model appears to be adequate, this does not necessarily mean that there are k nonlinear random effects. The results of these tests should be interpreted in conjunction with the multivariate forward selection results that identify which first- and second-order sensitivities best explain the patterns in between-subject variation. For example, in the experiment with one nonlinear random effect on the growth-rate parameter $\omega^{(2)}$ in example D, a model with two sensitivities was selected, but these were the first- and second-order sensitivities corresponding to a single random effect on $\omega^{(2)}$.

When the nonlinear random effects were large, the test for homoscedasticity was greatly affected by fitting a model with too few common factors. The resulting unexplained between-subject variation caused large bias in the estimates of γ and σ^2 in the MLFA-P model, and clouded the ability of the test based on the fit of the MLFA-P model or the MLFA-U model to discriminate between homoscedastic and heteroscedastic within-subject variation.

An aspect that was not explored in the linear random effect cases of example A, was the ability to distinguish between two parameterizations of a model. When a single random effect on a parameter in one parameterization induced two random effects in an alternate parameterization, our methods were able to distinguish between the two parameterizations. In most cases the multivariate forward selection procedure picked a smaller model for the true param-

eterization than for the alternate parameterization. In a few cases, the procedure selected a smaller model for the alternate parameterization. However, in these cases, the smaller model had smaller trace correlation r_T than the model selected for the true parameterization.

There were several marked differences between performance of our methods for the different nonlinear examples. For example, when there was one nonlinear random effect in example D, multivariate forward selection picked a model with two sensitivities, even when the random effect magnitude was large. However a model with three or more sensitivities was picked with one large random effect in example C. There could be several reasons for these differences. In Table 3.5 we compared the multiplying factors $|\beta^{(m)}|\|\partial f/\partial\beta^{(m)}\|$ that translate the relative standard deviation υ_m of the random effect on parameter $\beta^{(m)}$ into a linearized latent variable "signal". For example, the parameter V_d in the O1C model (example C) induces approximately 10 times larger signal in the linearized model than parameter $\omega^{(2)}$ of the logistic model (example D), for a given value of υ. A second reason for the differences between the different nonlinear examples could be due to how correlated or non-orthogonal the first- and second-order candidate sensitivities were for each model. We would expect the multivariate selection procedure to pick models that were too large in examples where the sensitivities were highly correlated with one another.

In conclusion, we would recommend our methods for application in examples with nonlinear random effects, with a few cautions or safety checks. The user should ensure that sufficient principal components or common factors are included in the multivariate models to adequately describe the between-subject variation, before proceeding with the various diagnostic tests based on the fit of these models. To to avoid selecting too many random effects, the order in which the multivariate forward selection picks sensitivities should be noted, and used to assist in identifying practically important random effects. The "hierarchy of predictors" principle, according to which higher-order sensitivities would be included in the model only if the corresponding first-order sensitivities were included, could also assist in identifying practically significant random effects.

Chapter 4

Detecting within-subject autocorrelation in the presence of model misspecification

4.1 Introduction

A common assumption when fitting a nonlinear regression model is that the nonlinear function f correctly specifies the expected trajectory being described. However, it is quite common in particular applications (e.g., toxicology, pharmacokinetics) to work with mechanistic models that suffer from some degree of misspecification, or model bias. In mechanistic models with many parameters, model bias may arise when some regression coefficients are set at "known" values, rather than being estimated by fitting the model to the data in hand. In the nonlinear mixed effects model (NLMM) notation developed in Section 1.3 of Chapter 1, we denoted these coefficients as $\boldsymbol{\xi}$, with $\boldsymbol{\xi}_*$ representing the true value, and $\widetilde{\boldsymbol{\xi}}$ the assigned value.

An important implication of model bias, especially where the data being modeled have a longitudinal structure, is the resulting difficulty in modeling the within-subject covariance structure. The following sentences quoted from Davidian and Giltinan (1995, Chapter 5) summarize the difficulty.

> "It is likely that there will be some interplay between the mean response function and an assumed correlation structure. Systematic deviations from the assumed form for the mean may be attributable to misspecification of mean function or to autocorrelation; without additional information or assumptions, this may not be possible to determine."

In the presence of model misspecification, the stage-1 regression model of the NLMM introduced in Chapter 1 becomes

$$\text{Stage-1:} \quad y_{ij} = f(t_j, \boldsymbol{\beta}_i, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) + \text{ misspecification error } + \varepsilon_{ij}$$

In this chapter we assume, for simplicity, that the within-subject variances are homoscedastic, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$. When the coefficients in $\boldsymbol{\xi}$ are assigned incorrect values, the misspecification error will be a function of time. This induces autocorrelation in the conditional residuals of the misspecified model

$$y_{ij} - f(t_j, \beta_i, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i) = \text{ misspecification error } + \varepsilon_{ij}$$

even when the true residuals ε_{ij} are uncorrelated.

Consider, as an example, the PBPK model for the CCl_4 gas uptake experiment. We used ordinary least-squares to estimate for each rat separately, the metabolic parameters, K_m and

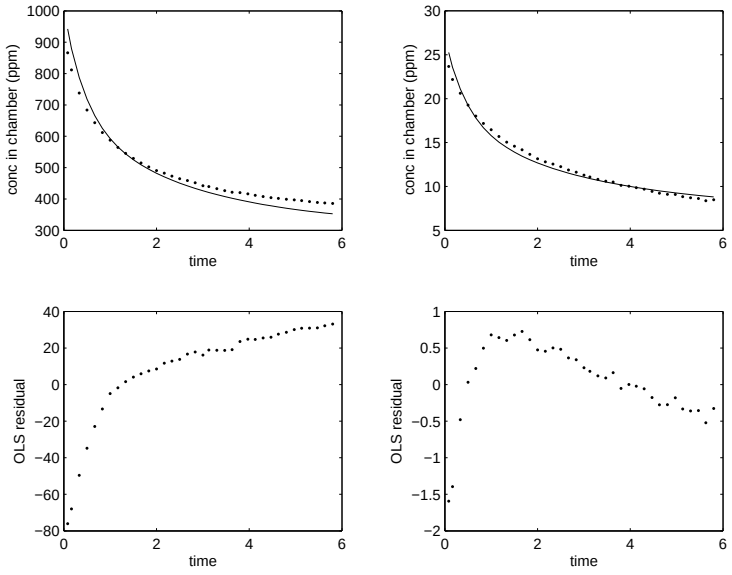

Figure 4.1: *OLS PBPK model fits and residuals for rat 3 (left) and rat 12 (right) from the CCl_4 gas uptake experiment*

V_{max}, and the true initial exposure in the chamber, y_{h0}. The other coefficients in the model were set at values in Table 1.1 of Chapter 1. Examples of the fitted model and residual plots are displayed in Figure 4.1. The inadequacy of the fitted model is clearly apparent in the smooth pattern in the residual plots. Although experience might suggest that the cause of the inadequacy is primarily due to model bias (possibly due to parameters such as body weight, B_w, being set at incorrect values), other model assumptions may also be incorrect. For example, the OLS fit assumes that observations made at different times on the same rat are independent and have constant measurement error variance. However based on the residual plots we cannot rule out the possibility of some time-dependent structure in measurements on each subject, over and above that accounted for by the PBPK model bias.

This chapter describes a statistical approach to disentangle the effects of model bias from genuine AR(1) structure of within-subject variation. The remainder of this chapter is laid out as follows. In Section 4.2 we describe our method for disentangling the effect of model bias from inherent serial correlation in the data. In Section 4.3 we present results of three simulation experiments designed to test our method. Application of the methods of this chapter to the CCl_4 gas uptake experiment and the theophylline example is postponed until Chapter 5.

4.2 Statistical Method

We begin with the two-stage NLMM described in equation (1.6) of Chapter 1, but assume homoscedastic within-subject variance, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$ for all i and j. Assume that, given the true values of subject-specific coefficients, $\boldsymbol{\beta}_i$, and known coefficients, $\boldsymbol{\xi}$, the regression function f accurately describes the expected trajectory for subject i in the stage-1 model, i.e.,

$$\mathbb{E}\left(y_{ij}|\boldsymbol{\beta}_i\right) = f(t_{ij}, \boldsymbol{\beta}_i, \boldsymbol{\xi}_*; \boldsymbol{x}_i)$$

where $\boldsymbol{\xi}_*$ is the true value of $\boldsymbol{\xi}$. If the assigned value $\widetilde{\boldsymbol{\xi}} \neq \boldsymbol{\xi}_*$, then model bias occurs due to the misspecification of $\boldsymbol{\xi}$,

$$\mathbb{E}\,(y_{ij}|\boldsymbol{\beta}_i) \neq f(t_{ij}, \boldsymbol{\beta}_i, \widetilde{\boldsymbol{\xi}}; \boldsymbol{x}_i).$$

We restrict our attention to model bias that occurs through the misspecification of the parameter $\boldsymbol{\xi}$, although other sources of model bias, for example, due to an inadequate choice of the regression function f, may be encountered in practice.

In earlier work (Chiswell and Monahan, 2004), we looked at estimation of the parameters of the NLMM $\boldsymbol{\beta}$, $\mathbf{D}$ and σ^2 when $\widetilde{\boldsymbol{\xi}} \neq \boldsymbol{\xi}_*$. We do not revisit this issue here. Instead, as in previous chapters, we assume that we have estimates of $\boldsymbol{\beta}$ (population mean values of the regression coefficients) in hand, perhaps from a preliminary analysis at the subject level. We focus on the problem of teasing apart the effects of model bias from genuine AR(1) structure of within-subject variation.

Our approach exploits the balanced structure of the data, namely that each subject is measured at the same, or very similar, time points. This allows us to compute pairwise differences between the trajectories for subjects with the same covariates. Multivariate regression analysis of these between-subject differences allows us to analyze the structure of within-subject variance, (approximately) independently of model bias. This approach is discussed in the following subsections.

4.2.1 Multivariate regression of between-subject differences

Our approach is based on the first-order linear (FOL) approximation introduced in Section 1.4 of Chapter 1. For two subjects, i and i', with the same observations times and the same covariates, we have the approximation given in equation (1.10) for each subject,

$$\begin{aligned} \boldsymbol{y}_i &\approx \boldsymbol{f} + \mathbf{F}_\xi\,(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}) + \mathbf{F}_\beta\,\boldsymbol{b}_i + \boldsymbol{\varepsilon}_i \\ \boldsymbol{y}_{i'} &\approx \boldsymbol{f} + \mathbf{F}_\xi\,(\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}) + \mathbf{F}_\beta\,\boldsymbol{b}_{i'} + \boldsymbol{\varepsilon}_{i'}. \end{aligned}$$

Taking the difference in response vectors for the two subjects gives

$$\boldsymbol{w}_{(i,i')} \equiv \boldsymbol{y}_i - \boldsymbol{y}_{i'} \approx \mathbf{F}_\beta(\boldsymbol{b}_i - \boldsymbol{b}_{i'}) + (\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_{i'}). \tag{4.1}$$

If the misspecification errors, $\boldsymbol{\xi}_* - \widetilde{\boldsymbol{\xi}}$ are small, then the bias due to misspecification of $\widetilde{\boldsymbol{\xi}}$ cancels out, at least approximately.

The matrix $\mathbf{F}_\beta$ is evaluated at the assigned value $\widetilde{\boldsymbol{\xi}}$; our method assumes that misspecification of $\boldsymbol{\xi}$ will not have a large effect on the sensitivity functions that make up the columns of $\mathbf{F}_\beta$. We also assume that the correct random effects have been identified, and that a first-order approximation is adequate. If non-trivial random effects are missing from the model, or are large (so that the FOL approximation fails), then these effects will also be absorbed into the conditional stage-1 residuals, contributing to additional autocorrelation in these residuals. It is not necessary to include all q sensitivities in the matrix $\mathbf{F}_\beta$. For example, using the methods of Chapter 2 we may have identified a subset of non-trivial random effects to include. Similarly, if these methods identified that a random effect was large enough to warrant inclusion of a second-order effect, then the corresponding second-order sensitivity should be included in $\mathbf{F}_\beta$. However, for simplicity of notation we will assume that $\mathbf{F}_\beta$ has q columns.

There are $(n-1)$ linearly independent pairs of differences, $\boldsymbol{y}_i - \boldsymbol{y}_{i'}$. If $n-1 > p-q-1$, then we cannot use all of the pairs, but must choose $r = \min(n-1, p-q-1)$ pairs. Unless there is some reason to choose a particular set, these could be picked at random. We collect the chosen pairwise differences as the columns of a $p \times r$ matrix, for example,

$$\mathbf{W} = \left[\boldsymbol{w}_{(1,2)}, \boldsymbol{w}_{(1,3)}, \ldots, \boldsymbol{w}_{(1,r+1)}\right].$$

Based on the approximate relationship in equation (4.1), we fit a multivariate multiple regression model

$$\mathbf{W} = \mathbf{F}_\beta \mathbf{B} + \mathbf{E} \tag{4.2}$$

where $\mathbf{B}$ is a $q \times r$ matrix of regression coefficients, and $\mathbf{E}$ is a $p \times r$ matrix containing the differences of the stage-1 residuals $\boldsymbol{\varepsilon}_i - \boldsymbol{\varepsilon}_{i'}$. Let $\boldsymbol{\Gamma}_0$ be the $r \times r$ covariance matrix of the j-th row of $\mathbf{E}$. The elements of $\boldsymbol{\Gamma}_0$ depend on the choice of pairwise differences included in $\mathbf{W}$. Since $\mathbb{V}(\varepsilon_{ij} - \varepsilon_{i'j}) = 2\sigma^2$, the diagonal elements of $\boldsymbol{\Gamma}_0$ equal $2\sigma^2$. The off-diagonal elements of $\boldsymbol{\Gamma}_0$ will either equal zero (if the two pairwise differences involve no common subject) or $\pm\sigma^2$ (when the two pairwise differences involve one common subject). For example, if $\mathbf{W} = \left[\boldsymbol{w}_{(1,2)}, \boldsymbol{w}_{(1,3)}, \ldots, \boldsymbol{w}_{(1,r+1)}\right]$, then the off-diagonal elements of $\boldsymbol{\Gamma}_0$ all equal σ^2 so that $\boldsymbol{\Gamma}_0 = \sigma^2\mathbf{I}_r + \sigma^2\mathbf{1}\mathbf{1}^T$.

Note that although the coefficients in $\mathbf{B}$ correspond to the random effect differences $\boldsymbol{b}_i - \boldsymbol{b}_{i'}$, we treat these coefficients as fixed, not random, when fitting the multivariate regression model. Statistical aspects of fitting a multivariate regression model, like the one in equation (4.2), were discussed in Section 2.2.5 of Chapter 2. These methods assume that the rows of $\mathbf{E}$ are independent, which will be the case if the within-subject residuals are uncorrelated.

If the proposed multivariate regression model (4.2) is adequate, then this model should fit well, for example have trace correlation r_T close to one. The trace correlation was defined in equation (2.13) of Chapter 2. If the stage-1 residuals ε_{ij} are uncorrelated within a subject, then we should see no obvious time-dependent pattern in rows of the residual matrix $\mathbf{E}$ of the fitted multivariate regression. In the next section we derive the expected multivariate autocorrelation structure amongst the rows of $\mathbf{E}$ that would arise if the stage-1 residuals follow a first-order autoregressive, AR(1), process.

4.2.2 Testing for AR(1) structure in within-subject variation

The primary goal of fitting the multivariate regression model of equation (4.2) is to attempt to disentangle model bias from possible underlying time-dependent structure in the errors ε_{ij}. To do this we assume a first-order autoregressive, AR(1), model for within-subject time dependence in the original data,

$$\varepsilon_{ij} = \rho\varepsilon_{i,j-1} + a_{ij} \quad \text{with} \quad a_{ij} \text{ i.i.d. } N(0, \sigma_a^2) \tag{4.3}$$

for $j = 2, \ldots, p$ and $i = 1, \ldots, n$. Under model (4.3),

$$\mathbb{V}(\varepsilon_{ij}) = \sigma_a^2/(1-\rho^2) \equiv \sigma^2 \quad \text{and} \quad \text{Cov}(\varepsilon_{ij}, \varepsilon_{i,j-1}) = \rho\sigma_a^2/(1-\rho^2) = \rho\sigma^2$$

for $j = 2, \ldots, p$. We assume the residual for the first time point, ε_{i1}, is i.i.d. $N(0, \sigma^2)$. The residuals from different subjects are uncorrelated.

Under the assumptions of model (4.3), time dependence amongst the rows of $\mathbf{E}$ follows a first-order vector autoregressive, VAR(1), process. To see this, let $\mathbf{E}_j^T$ be the j-th row of $\mathbf{E}$. The r elements of the vector $\mathbf{E}_j$ are of the form $\varepsilon_{ij} - \varepsilon_{i'j}$, and can be written as follows, based on the AR(1) process in equation (4.3),

$$\varepsilon_{ij} - \varepsilon_{i'j} = \rho(\varepsilon_{i,j-1} - \varepsilon_{i',j-1}) + (a_{ij} - a_{i'j}).$$

The same relation holds for other elements of $\mathbf{E}_j$. Let $\boldsymbol{c}_j$ be a vector with r elements of the form $c_{(i,i'),j} = a_{ij} - a_{i'j}$. Since the a_{ij} are i.i.d. $N(0, \sigma_a^2)$, the elements of $\boldsymbol{c}_j$ have mean 0 and variance $2\sigma_a^2$. Elements of $\boldsymbol{c}_j$ may be correlated (with covariance $\pm\sigma_a^2$), but the $\boldsymbol{c}_j$ vectors are uncorrelated across time points. Therefore we can write $\mathbf{E}_j$ as a first-order vector autoregressive, VAR(1), process,

$$\mathbf{E}_j = \rho\mathbf{E}_{j-1} + \boldsymbol{c}_j \quad \text{for} \quad j = 2, \ldots, p. \tag{4.4}$$

The lag-1 covariance matrix $\boldsymbol{\Gamma}_1 = \text{Cov}(\mathbf{E}_j, \mathbf{E}_{j-1})$ for the j-th and $(j-1)$-th rows of $\mathbf{E}$, has elements of three types. On the diagonal, the elements involve the same two subjects, so this element is $2\rho\sigma_a^2/(1-\rho^2) = 2\rho\sigma^2$. The off-diagonal elements are either zero (if the two pairwise differences involve no common subjects), or $\pm\rho\sigma_a^2/(1-\rho^2) = \pm\rho\sigma^2$ (if the two pairwise differences involve one common subject). For example, if $\mathbf{W} = \left[\boldsymbol{w}_{(1,2)}, \boldsymbol{w}_{(1,3)}, \ldots, \boldsymbol{w}_{(1,r+1)}\right]$, then the off-diagonal elements of $\boldsymbol{\Gamma}_1$ all equal $\rho\sigma^2$ so that

$$\boldsymbol{\Gamma}_1 = \rho\sigma^2\mathbf{I}_r + \rho\sigma^2\mathbf{1}\mathbf{1}^T = \rho\boldsymbol{\Gamma}_0.$$

For any choice of pairwise differences, the diagonal elements of $\mathbf{\Gamma}_1$ are proportional to the diagonal elements of $\mathbf{\Gamma}_0$, with constant of proportionality ρ. If $\rho \equiv 0$ for all subjects, then there will be no time-dependence amongst rows of $\mathbf{E}$.

To test whether $\rho = 0$, we propose the following (somewhat ad-hoc) approach. Let

$$\widehat{\mathbf{B}} = (\mathbf{F}_\beta^T \mathbf{F}_\beta)^{-1} \mathbf{F}_\beta^T \mathbf{W}$$

the the least-squares estimator of $\mathbf{B}$ in the multivariate regression model (4.2). Let $\widehat{\mathbf{E}} = \mathbf{W} - \mathbf{F}_\beta \widehat{\mathbf{B}}$ be the estimated residual matrix. Since the multivariate regression model does not necessarily contain an intercept, we center the residuals, $\widehat{\mathbf{E}}^* = (\mathbf{I}_p - \mathbf{P}_1)\widehat{\mathbf{E}}$ where $\mathbf{P}_1 = p^{-1}\mathbf{1}\mathbf{1}^T$. An estimator of $\mathbf{\Gamma}_0$ is

$$\mathbf{G}_0 = (p-q)^{-1} \sum_{j=1}^{p} \widehat{\mathbf{E}}_j^* \widehat{\mathbf{E}}_j^{*T}$$

where $\widehat{\mathbf{E}}_j^{*T}$ is the j-th row of $\widehat{\mathbf{E}}^*$. An estimator of the lag-1 covariance $\mathbf{\Gamma}_1$ is

$$\mathbf{G}_1 = (p-q-1)^{-1} \sum_{j=2}^{p} \widehat{\mathbf{E}}_{j-1}^* \widehat{\mathbf{E}}_j^{*T}.$$

As mentioned earlier, for any set of pairwise differences, the diagonal elements of $\mathbf{\Gamma}_1$ equal ρ times the diagonal elements of $\mathbf{\Gamma}_0$. Let $r_{(i,i')} = (\mathbf{G}_1)_{(i,i')}/(\mathbf{G}_0)_{(i,i')}$ be the ratio of the diagonal elements of $\mathbf{G}_1$ and $\mathbf{G}_0$ corresponding to the difference $\boldsymbol{w}_{(i,i')}$. To test whether there is significant evidence of AR(1) autocorrelation, we compute Fisher's Z transformation for each of the r ratios,

$$Z'_{(i,i')} = .5 \log\left(\frac{1 + r_{(i,i')}}{1 - r_{(i,i')}}\right).$$

Each diagonal element is based on $p-1$ cross-product terms, so using standard results (e.g., Steel and Torrie, 1980, Section 11.4), $Z'_{(i,i')}$ has an approximate normal distribution with mean $.5\log\{(1+\rho)/(1-\rho)\}$ and variance $(p-1-3)^{-1} = (p-4)^{-1}$. Let $\bar{Z}'$ be the mean of the r values of $Z'_{(i,i')}$. Then $\bar{Z}'$ has an approximate normal distribution with mean $.5\log\{(1+\rho)/(1-\rho)\}$.

An estimator of ρ is

$$\widehat{\rho} = \frac{\mathrm{e}^{2\bar{Z}'} - 1}{\mathrm{e}^{2\bar{Z}'} + 1}. \tag{4.5}$$

Because the elements $Z'_{(i,i')}$ are most likely correlated, the variance of $\bar{Z}'$ could be larger or smaller than $\{r(p-4)\}^{-1}$. To test the hypothesis H_0: $\rho = 0$, we suggest using the test statistic

$$T_{\rho(i,i')} = (p-4)^{1/2} Z'_{(i,i')} \tag{4.6}$$

for each of the r ratios of the diagonal elements of $\mathbf{G}_1$ and $\mathbf{G}_0$. Under the hypothesis H_0, $T_{\rho(i,i')}$ will have an approximately standard normal distribution. We need to correct for the r multiple comparisons. We suggest a simple Bonferroni-correction, i.e., to use the $\alpha/(2r)$ critical value for each of the r two-sided comparisons. The hypothesis is rejected if in one or more tests, $|T_{\rho(i,i')}|$, exceeds the $\alpha/(2r)$ critical value from the standard normal distribution, $z_{\alpha/(2r)}$. In the simulation experiments of the next section, we test the size and power of the test procedure, and the estimation of ρ using $\widehat{\rho}$ defined in (4.5).

4.3 Simulation experiments

In this section we present results of three simulation experiments designed to test the ability of our method to detect AR(1) structure in the first stage of the NLMM. In the first experiment, the data generating model had a single linear random effect, and homoscedastic within-subject variation, and we investigated power to reject H_0: $\rho = 0$ over a range of ρ. The second experiment was similar to the first, but we investigated the effect of heteroscedastic within-subject variance. In the third experiment, we generated data according to the open-two compartment (O2C) model, with a random effect on the large rate parameter $\omega^{(2)}$ (a nonlinear random effect). We looked at the effect of possible misspecification of the rate parameter, on which the sensitivity used as a predictor in the multivariate regression for between-subject differences depends. In this last experiment we assumed that within-subject variance was homoscedastic.

4.3.1 Experiment I: one linear random effect and homoscedastic within-subject variance

In this experiment we generated data according to the following model where an AR(1) process defined the correlation structure in the stage-1 residuals

$$\begin{aligned}
&\text{Stage-1:} \qquad y_{ij} = w_i f_{\omega^{(2)}}(t_j) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2),\ \mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i,k}) = \sigma^2 \rho^{|j-k|} \\
&\text{Stage-2:} \qquad w_i \text{ i.i.d. } N(0, \sigma_w^2), \quad \sigma_w = \upsilon_w |\omega^{(2)}|
\end{aligned}$$

for $i = 1, \ldots, n$ subjects, and $j = 1, \ldots, p$ time points, t_j, that were equally spaced in $[0, 1]$. We assumed that the residuals ε_{ij} were independent across subjects, and independent of the random effects w_i. The function $f_{\omega^{(2)}}$ was the sensitivity for $\omega^{(2)}$ in the O2C model,

$$f_{\omega^{(2)}}(t) = -(1-\alpha) t \exp(-\omega^{(2)} t)$$

scaled to have norm 1 (i.e., $\sum_{j=1}^{p} \{f_{\omega^{(2)}}(t_j)\}^2 = p$). We set parameter values at $\alpha = 0.25$, $\omega^{(2)} = 5$, $\sigma = 0.01$, and $\upsilon_w = 0.1$ (random effect standard deviation was 10% of $\omega^{(2)}$).

We did not consider the effect of misspecification of a linear parameter, because this effect cancels out completely in the analysis. For example, if the true stage-1 model was

$$y_{ij} = w_i f_{\omega^{(2)}}(t_j) + \xi f_2(t_j) + \varepsilon_{ij}$$

where $f_2(t)$ was a function free of $\omega^{(2)}$ and ξ, then the difference between responses of subjects i and i' is free of ξ,

$$y_{ij} - y_{i'j} = (w_i - w_{i'}) f_{\omega^{(2)}}(t_j) + (\varepsilon_{ij} - \varepsilon_{i'j})$$

and therefore indifferent to the effects of misspecification in the value of ξ.

Table 4.1 lists the factors and levels considered in this experiment. The number of pairwise differences, r, used in the multivariate regression approach was limited by the number of time

Table 4.1: *Simulation treatment factors for experiment I*

Factor	Levels
AR(1) correlation, ρ	0, 0.05, 0.1, 0.25, 0.5, 0.95
No. time points, p	10, 25, 50
Relative no. subjects, n/p	1/2, 3/4

points, p. Therefore we only considered values of n smaller than p. The value of n used was $(n/p)p$ rounded up to the nearest integer. We generated 1,000 Monte Carlo replicates of each treatment combination. We implemented the estimation and hypothesis testing approach outlined earlier in this chapter, and generated the simulation data, using Matlab (Version 7.0 (R14)).

We used the test defined in equation (4.6) to test for evidence of AR(1) correlation at the 5% significance level. We estimated ρ using the estimator of equation (4.5). The ability to detect AR(1) structure should increase as ρ increases. We would expect the power of the test to increase with number of time points p. The estimate of ρ should be more accurate and more precise with increasing p, and should also improve with larger n. Because the Bonferroni correction depends on the number of between-subject differences, which is a function of n, we expect that the test may be more conservative for larger n.

Table 4.2: *Estimated size of test for H_0: $\rho = 0$ when $\rho = 0$ for experiment I. In this experiment, there was one linear random effect, and within-subject variance was homoscedastic. Nominal level of the test was* 0.05*. Based on* 1000 *MC replicates, the standard error of estimated size was less than* 0.007*.*

	n/p	
p	1/2	3/4
10	0.037	0.040
25	0.030	0.046
50	0.045	0.045

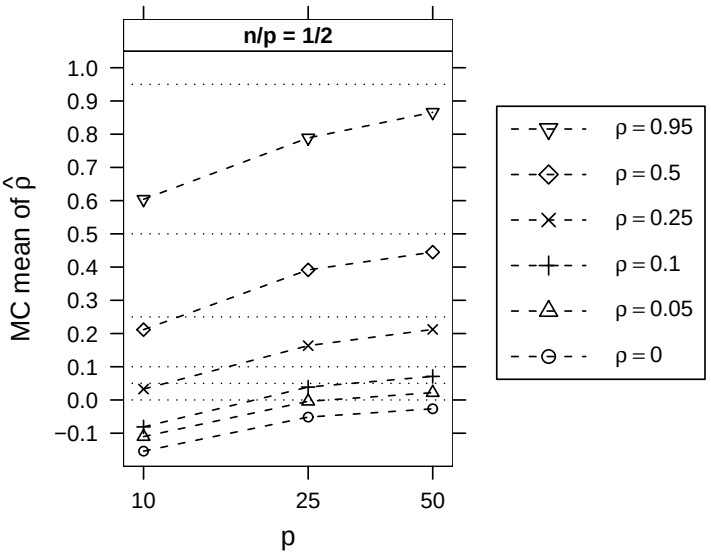

Figure 4.2: *MC means of $\widehat{\rho}$ plotted against p when there was one linear random effect, and within-subject variance was homoscedastic. The six curves show the estimates for different values of AR(1) correlation, ρ. The ratio of the number of subjects to number of time points, n/p had little effect on power; we plot results only for $n/p = 1/2$. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.007 *when $p = 10$, less than* 0.004 *when $p = 25$, and less than* 0.003 *when $p = 50$.*

We used analysis of variance to assess how the different simulation factors (p, n/p, and ρ) affected the power of the test and the estimate of ρ. Since these endpoints cannot be expected to be normally distributed, we batched the MC replicates into groups of size 10, and calculated the ANOVA on batch means. The ratio of number of subjects to number of time points had little effect on the estimation of ρ (p-value = 0.3701), and only marginal effect on power of the test (p-value = 0.0101). Power of the test, and bias in $\widehat{\rho}$ were primarily affected by the true value of ρ and by p (p-values < 0.0001).

The estimate of ρ was too small on average. The absolute magnitude of bias was larger for larger ρ, and bias decreased with increasing p — see Figure 4.2. The Monte Carlo standard error of the estimates of ρ also decreased with increasing p.

Figure 4.3 plots the Monte Carlo estimate of power of the test for H_0: $\rho = 0$, as a function

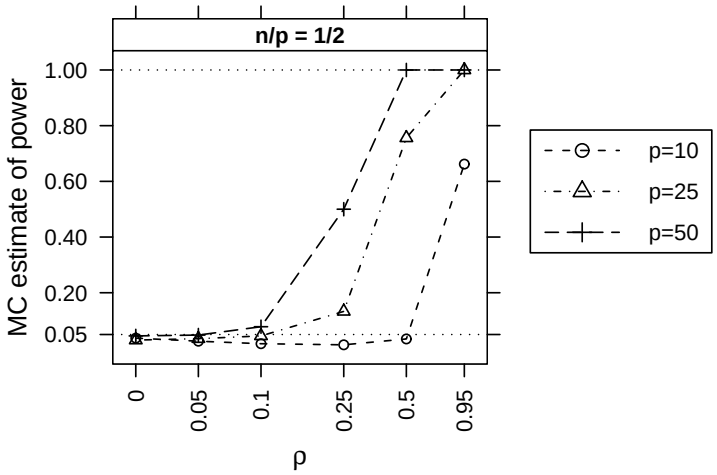

Figure 4.3: *Estimated power to reject* H_0*:* $\rho = 0$ *as a function of* ρ *when there was one linear random effect, and within-subject variance was homoscedastic. The three curves show the effect of increasing number of time points,* p*. The ratio of the number of subjects to number of time points,* n/p *had little effect on power; we plot results only for* $n/p = 1/2$*. Nominal size of the test was* 0.05*. Based on* 1000 *MC replicates, the standard error of estimated power was less than* 0.016*.*

of ρ. Since the ratio n/p had little effect, we have plotted results only for $n/p = 1/2$. The three curves in each plot show that, as expected, power increased with increasing number of time points, p. When $p = 10$, power to detect even strong autocorrelation ($\rho = 0.95$) was low (less than 65%). However, when $p = 50$, a correlation of $\rho = 0.5$ was detected in 100% of cases. When $\rho = 0$, the size of the test was close to, although slightly smaller than the nominal 0.05 level (see Table 4.2). This is not surprising since the Bonferroni correction will be conservative (reject a true null hypothesis less often than suggested by the 0.05 nominal level) when the r hypotheses are not independent.

4.3.2 Experiment II: one linear random effect and heteroscedastic within-subject variance

The data generating model for this experiment was similar to that used in Experiment I, except that here we also examined the effect of within-subject heteroscedasticity on the detection and

estimation of AR(1) correlation. We generated data according to the following model where a first-order heteroscedastic autoregressive, ARH(1), process defined the correlation structure in the stage-1 residuals

$$\begin{aligned}
&\text{Stage-1:} \qquad y_{ij} = w_i f_{\omega^{(2)}}(t_j) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2\{\mu(t_j)\}^{2\gamma}),\\
&\qquad\qquad\quad \mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i,k}) = \sigma^2\{\mu(t_j)\}^{\gamma}\{\mu(t_k)\}^{\gamma}\rho^{|j-k|}\\
&\text{Stage-2:} \qquad w_i \text{ i.i.d. } N(0, \sigma_w^2), \quad \sigma_w = \upsilon_w|\omega^{(2)}|
\end{aligned}$$

for $i = 1, \ldots, n$ subjects, and $j = 1, \ldots, p$ time points, t_j, that were equally spaced in $[0, 1]$. We assumed that the residuals ε_{ij} were independent across subjects, and independent of the random effects w_i.

As in Experiment I, the function $f_{\omega^{(2)}}$ was the sensitivity for $\omega^{(2)}$ in the O2C model. To simulate heteroscedasticity, we set the variance of the residuals at time t_j to be proportional to $\{\mu(t_j)\}^{2\gamma}$ where μ was defined by the O2C model,

$$\mu(t) = \alpha \exp(-\omega^{(1)}t) + (1 - \alpha)\exp(-\omega^{(2)}t)$$

scaled to have norm 1. We did not include a fixed mean in the data generating model because this would cancel out in computing the between-subject residuals. We set parameter values at $\alpha = 0.25$, $\omega^{(1)} = 0.5$, $\omega^{(2)} = 5$, $\sigma = 0.01$, and $\upsilon_w = 0.1$ (10% of $\omega^{(2)}$).

Since the number of subjects, n, had little effect in the first experiment, we set $n = p/2$ (rounded up to the nearest integer) in Experiment II. Table 4.3 lists the factors and levels

Table 4.3: *Simulation treatment factors for experiment II*

Factor	Levels
AR(1) correlation, ρ	0, 0.05, 0.1, 0.25, 0.5, 0.95
Heteroscedasticity, γ	0, 0.5, 1
No. time points, p	10, 25, 50

considered in this experiment. We generated 1,000 Monte Carlo replicates of each treatment combination.

We expect results to the same as those in Experiment I when within-subject variation is homoscedastic, $\gamma = 0$. When $\gamma > 0$, the bias in the estimate of ρ should increase. We expect that the power and size of the test for H_0: $\rho = 0$ will be affected by heteroscedasticity, but we could not predict before-hand whether it would decrease or increase.

We used analysis of variance (based on means of batches of size 10) to assess whether the different simulation factors (p, γ, and ρ) affected the power of the test and the estimator of ρ. All three factors affected the power and size of the test for H_0: $\rho = 0$, and bias in the estimator of ρ (p-values < 0.001). The number of time points, p, and the true value of ρ, had the largest

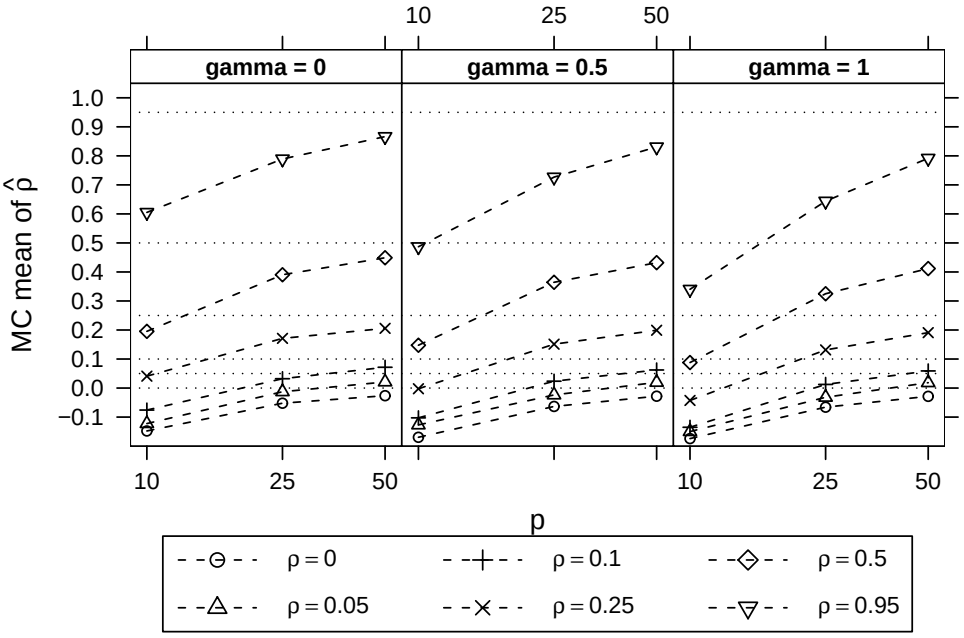

Figure 4.4: *MC means of $\widehat{\rho}$ plotted against p when there was one linear random effect, and within-subject variance was heteroscedastic. The three panels show the effect of heteroscedasticity (γ). The six curves show the estimates for different values of AR(1) correlation, ρ. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.008 *when* $p = 10$*, less than* 0.005 *when* $p = 25$*, and less than* 0.004 *when* $p = 50$*.*

affect on results, but results were also significantly affected by the presence of heteroscedasticity ($\gamma > 0$).

As in Experiment I, the estimate of ρ was too small on average. The absolute magnitude of bias was larger for larger ρ, and bias decreased with increasing p — see Figure 4.4. The Monte Carlo standard error of the estimates of ρ also decreased with increasing p. Comparing the results in the three panels, we see that the presence of heteroscedasticity increased bias in $\widehat{\rho}$, especially when p was small.

Figure 4.5 plots the Monte Carlo estimate of power of the test for H_0: $\rho = 0$, as a function of ρ. The three panels show the effect of heteroscedasticity on the power of the test. Heteroscedasticity increased the size of the test when $p = 25$ or $p = 50$ (see Table 4.4). When $p = 10$, power to detect AR(1) structure with ρ as large as 0.95 was reduced to less than 40% when $\gamma = 1/2$, and to less than 15% when $\gamma = 1$. As in Experiment I, power of the test

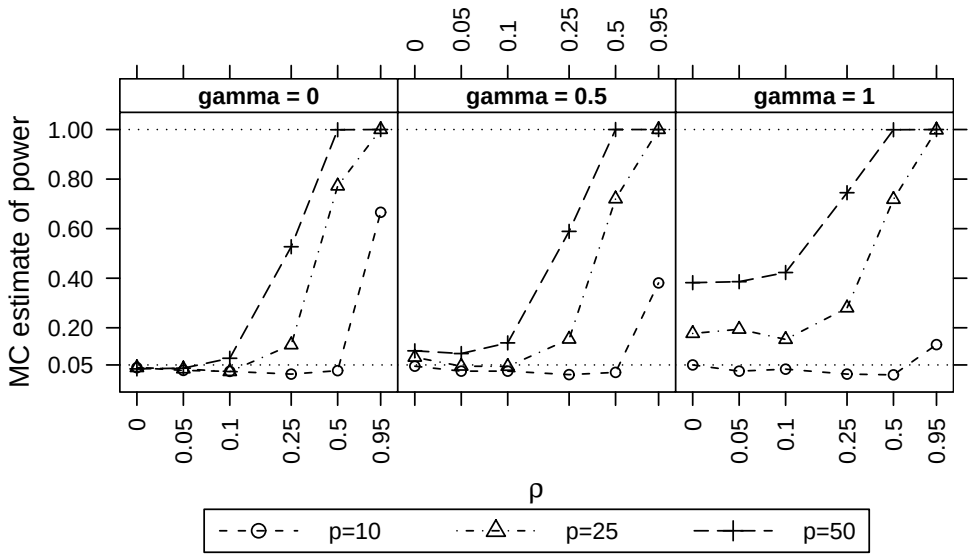

Figure 4.5: *Estimated power to reject H_0: $\rho = 0$ as a function of ρ when there was one linear random effect, and within-subject variance was heteroscedastic. The three panels show the effect of heteroscedasticity (γ). In each panel, the three curves show the effect of increasing number of time points, p. Nominal size of the test was* 0.05*. Based on* 1000 *MC replicates, the standard error of estimated power was less than* 0.016*.*

Table 4.4: *Estimated size of test for H_0: $\rho = 0$ when $\rho = 0$ for experiment II. In this experiment, there was one linear random effect, and within-subject variance was heteroscedastic. Nominal level of the test was* 0.05*. Based on* 1000 *MC replicates, the standard error of estimated size was less than* 0.007*.*

	γ		
p	0	1/2	1
10	0.039	0.045	0.050
25	0.039	0.081	0.177
50	0.034	0.107	0.382

increased with increasing p.

4.3.3 Experiment III: one nonlinear random effect and misspecification of one nonlinear parameter

In the third experiment we generated data according to the O2C model with a small random effect on the large rate parameter, $\omega^{(2)}$, and possible misspecification of this rate parameter. An AR(1) process (with homoscedastic variance) defined the correlation structure in the stage-1 residuals

$$\begin{aligned}
&\text{Stage-1:} \qquad y_{ij} = \alpha \exp(-\omega^{(1)} t_j) + (1-\alpha)\exp\{-(\omega^{(2)} + w_i)t_j\} + \varepsilon_{ij}, \\
&\qquad\qquad\quad \varepsilon_{ij} \sim N(0, \sigma^2),\ \mathrm{Cov}(\varepsilon_{ij}, \varepsilon_{i,k}) = \sigma^2 \rho^{|j-k|} \\
&\text{Stage-2:} \qquad w_i \text{ i.i.d. } N(0, \sigma_w^2), \quad \sigma_w = \upsilon_w |\omega^{(2)}|
\end{aligned}$$

for $i = 1, \ldots, n$ subjects, and $j = 1, \ldots, p$ time points, t_j, that were equally spaced in $[0, 1]$. We assumed that the residuals ε_{ij} were independent across subjects, and independent of the random effects w_i. We set parameter values at $\alpha = 0.25$, $\omega^{(1)} = 0.5$, $\omega^{(2)} = 5$, $\sigma = 0.01$, and $\upsilon_w = 0.1$ (random effect standard deviation was 10% of $\omega^{(2)}$).

Let $\widetilde{\xi} = \delta\omega^{(2)}$ be a possibly misspecified value of the rate parameter, where δ takes values

Table 4.5: *Simulation treatment factors for experiment III*

Factor	Levels
AR(1) correlation, ρ	0, 0.05, 0.1, 0.25, 0.5, 0.95
Misspecification, δ	0.75, 0.9, 1, 1.1, 1.25
No. time points, p	10, 25, 50

$0.75, 0.9, 1, 1.1, 1.25$. We used the true value, $\xi_* = \omega^{(2)}$, to generate the data. In the regression of the multivariate-between subject sensitivities, we assumed that $\omega^{(2)}$ was misspecified at the value $\widetilde{\xi}$, and evaluated the sensitivity $f_{\omega^{(2)}}$ at $\widetilde{\xi}$ instead of at $\omega^{(2)}$,

$$\widetilde{f}_{\omega^{(2)}}(t) = -(1-\alpha)t\exp(-\delta\omega^{(2)}t).$$

This simulates how our method might work if one parameter in the model was misspecified. We chose to misspecify $\omega^{(2)}$ because misspecification of the other rate parameter, $\omega^{(1)}$, would have had no effect ($f_{\omega^{(2)}}$ is free of $\omega^{(1)}$), and misspecification of the scale parameter α gives a function proportional to the true sensitivity function.

Table 4.5 lists the factors and levels considered in this experiment. As in Experiment II, we set $n = p/2$ (rounded up to the nearest integer). We generated 1,000 Monte Carlo replicates of

Table 4.6: *Estimated size of test for H_0: $\rho = 0$ when $\rho = 0$ for experiment III. In this experiment, there was a random effect on $\omega^{(2)}$, misspecification of $\omega^{(2)}$, and within-subject variance was homoscedastic. The factor δ controls the amount of misspecification in $\omega^{(2)}$, $\widetilde{\xi} = \delta\omega^{(2)}$. Nominal level of the test was* 0.05*. Based on* 1000 *MC replicates, the standard error of estimated size was less than* 0.007*.*

	δ		
p	1	1.1	1.25
10	0.024	0.034	0.027
25	0.031	0.071	0.129
50	0.051	0.123	0.394

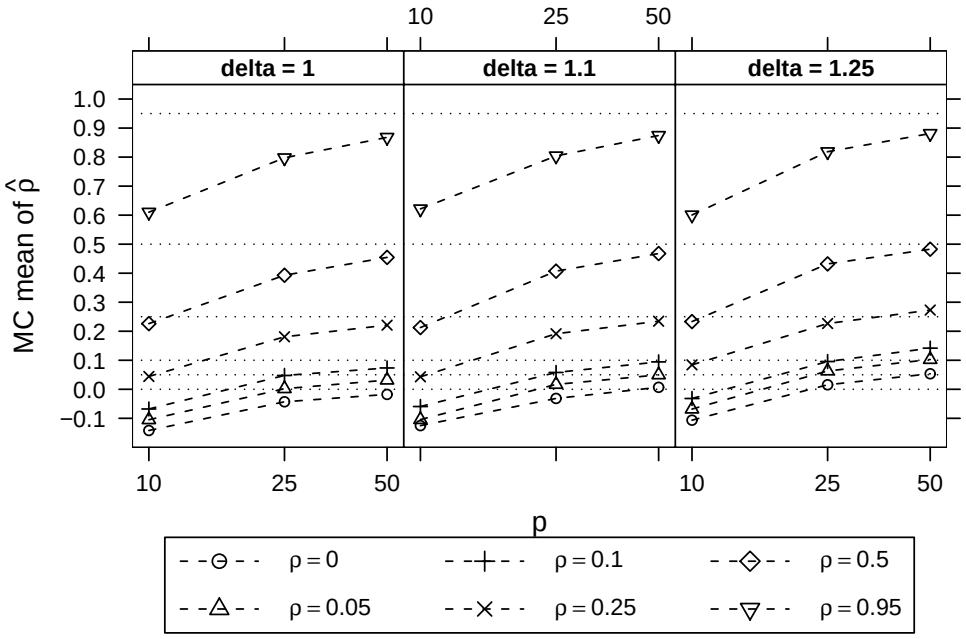

Figure 4.6: *MC means of $\widehat{\rho}$ plotted against p when there was a random effect on $\omega^{(2)}$, misspecification of $\omega^{(2)}$, and within-subject variance was homoscedastic. The three panels show the effect of different amounts of misspecification, $\widetilde{\xi} = \delta\omega^{(2)}$. The six curves in each panel show the estimates for different values of AR(1) correlation, ρ. Based on analyzing the simulation data in batches of* 10*, we estimated the standard error of each plotted MC mean to be less than* 0.008 *when* $p = 10$*, less than* 0.005 *when* $p = 25$*, and less than* 0.004 *when* $p = 50$*.*

each treatment combination.

When $\delta = 1$ (no misspecification) we expect results to be similar to those of Experiment I, because the magnitude of the nonlinear random effect is small. However, when $|\delta - 1|$ is large, our method may detect AR(1) structure when there is none; therefore the size of the test may be inflated.

We used analysis of variance (based on means of batches of 10 replicates) to assess whether the different simulation factors (p, δ, and ρ) affected the power of the test and the estimate of ρ. All three factors, as well as their interactions had a significant effect on bias in $\widehat{\rho}$ and the power of the test for AR(1) structure (p-values < 0.0001). Results did not depend on the direction of misspecification of $\omega^{(2)}$; for example, results when $\delta = 0.75$ were similar to those

when $\delta = 1.25$. For the sake of brevity we have presented results only for the cases where $\delta \geq 1$.

Figure 4.6 shows the average values of the estimates of ρ, plotted against p. The three panels show the effect of misspecification of parameter $\omega^{(2)}$. Within each panel, six curves plot the estimates for different values of the truth (ρ). As in the previous two experiments, the estimate of ρ was generally too small on average. However, when there was large misspecification, large number of time points ($p = 50$), and the true correlation was small ($\rho \leq 0.25$), the estimator $\widehat{\rho}$ was larger on average than the true value. That is, when ρ is small and p is large, our method may over-estimate AR(1) correlation in the presence of large misspecification error.

Figure 4.7 plots the Monte Carlo estimate of power of the test for H_0: $\rho = 0$, as a function of ρ. The three panels show the effect of misspecification of parameter $\omega^{(2)}$. When $\rho = 0$ and $\delta = 1$ (no misspecification), the size of the test was close to, or smaller than, the nominal 0.05

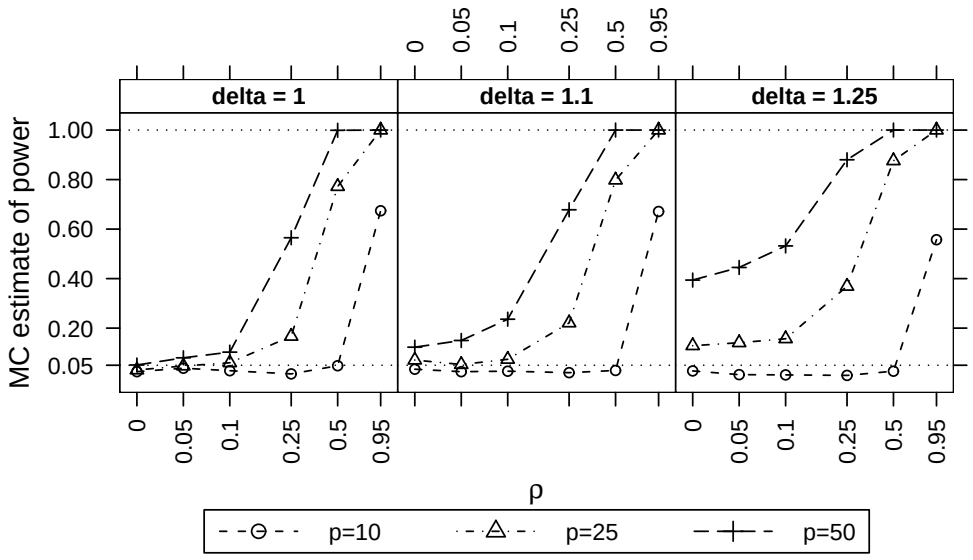

Figure 4.7: *Estimated power to reject H_0: $\rho = 0$ as a function of ρ when there was a random effect on $\omega^{(2)}$, misspecification of $\omega^{(2)}$, and within-subject variance was homoscedastic. The three panels show the effect of different amounts of misspecification, $\widetilde{\xi} = \delta\omega^{(2)}$. The three curves in each panel show the effect of increasing number of time points, p. Nominal size of the test was* 0.05*. Based on* 1000 *MC replicates, the standard error of estimated power was less than* 0.016*.*

level (see Table 4.6). The small nonlinear random effect did not inflate the size of the test. When the number of time points was few ($p = 10$), misspecification did not have much effect on the size or power of the test. However, when $p = 25$ or $p = 50$, misspecification error increased the size of the test (last two columns of Table 4.6). As in Experiments I and II, power of the test increased with increasing p.

4.3.4 Summary of simulation results

In this chapter, we developed an approach to disentangle AR(1) structure in within-subject residuals from correlations caused by misspecified regression coefficients. Based on regressing between-subject differences on random effect parameter sensitivities, we proposed an estimator for AR(1) correlation ρ, and a test for the hypothesis H_0: $\rho = 0$.

The estimator of the AR(1) correlation coefficient, ρ, was generally too small on average. Bias in $\widehat{\rho}$ decreased with increasing number of time points p. The number of subjects, n did not have much effect on estimating ρ, or on testing for evidence of AR(1) structure.

The first simulation experiment showed that the test maintained the correct level (0.05) when $\rho = 0$ and within-subject variance was homoscedastic. Power to detect within-subject correlation when ρ was as large as 0.5 was acceptable when the number of time points was 25 or larger. We would not recommend the test when p is small, but the performance was acceptable for $p \geq 25$.

Our method is based on the assumption of homoscedastic within-subject variance. The second simulation experiment showed that the size of the test was notably inflated when the assumption of homoscedasticity was strongly violated ($\gamma = 1$). This was most evident when sample size was large ($p = 25$ or $p = 50$).

The main purpose of our method was to attempt to disentangle the effects of model bias caused by misspecification of coefficient values from true within-subject autocorrelation. The ability of our test to do this was tested in experiment III. This experiment showed that our test worked as expected when the nonlinear random effect was small. When misspecification

error was small (10%), our test had slightly inflated size, but provided a useful test for within-subject autocorrelation. When misspecification error was as large as 25%, our method could not distinguish between true AR(1) correlation, and dependence caused by the misspecification.

The reader might question the utility of an approach that assumes or accepts bias in the function that describes the stage-1 trajectories in the NLMM. We agree that parameter estimation and modeling using a biased model may not make sense; for example, estimates of regression coefficients in a biased model may not be meaningful. However, it may be the case that the modeler cannot decide whether a model is really biased, or whether the observed pattern in residuals is due to true AR(1) structure. Our method provides one way of attempting to tease apart these two issues. If, according to our method, there appears to be no AR(1) structure, then any lack of fit may be attributed to model bias. The analyst then has to decide how to address the issue of model bias.

Chapter 5

Application: motivating examples

In this chapter we apply the methods developed in Chapters 2 and 4 to the two motivating examples that were introduced in Section 1.1. In the first example (discussed in Section 5.1), the open two-compartment (O2C) model and a physiologically-based pharmacokinetic (PBPK) model are alternative functions used to describe longitudinal measurements of the chamber concentration of carbon tetrachloride recorded in a gas uptake experiment. In the second example (discussed in Section 5.2), the open one-compartment (O1C) model is used to describe repeated measurements of serum concentrations of the drug theophylline recorded in 12 human subjects.

We apply the methods of Chapter 2 to explore the marginal covariance structure of the two data sets. We attempt to select the number and identity of random effect parameters in the nonlinear models that are needed to characterize between-subject variation. For the O1C and O2C models we investigate which of two parameterizations is best suited to describe between-subject variation. We investigate whether there is evidence that within-subject variation is heteroscedastic, or whether we can assume this variation to be constant over time. Because nonlinear models do not always provide an unbiased prediction of subject-level trajectories, it may be difficult to determine whether observed correlation in within-subject residuals is due to model bias or true serial correlation in measurements. We use the methods of Chapter 4 to attempt to disentangle serial correlation in within-subject measurement errors from the effects of model bias.

We conclude the chapter with a few comments and additional insights about our methods, gained through applying them to the two examples.

5.1 Carbon tetrachloride gas uptake experiment

We described the CCl_4 gas uptake experiment in Section 1.1 of Chapter 1. Twelve rats were individually exposed to CCl_4 in a closed chamber. Four initial exposures were used in the experiment: 1000, 250, 100 and 25 ppm. Concentration of CCl_4 in the chamber was measured approximately 36 times per rat over a period of 6 hours. Measurements were made at the same (or very similar) time points for all rats in the study.

Because our methods require balanced data (the same observation time points for each subject), we used linear interpolation to impute the few missing values in the data set, and assumed that measurement times were identical across subjects. Figure 5.1 displays the observed chamber concentrations (including interpolated values) for the 12 rats in the gas uptake experiment. We divided the plotted concentrations by the initial exposure level to eliminate the scaling effect due to dose. By undertaking this scaling, we are implicitly assuming that measurement error variance, on the original scale, is heterogeneous (a function of initial exposure), with larger

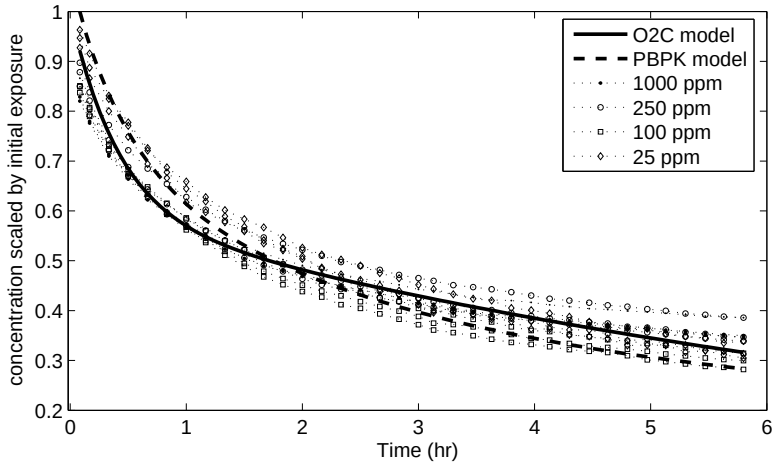

Figure 5.1: *Chamber concentration of CCl_4 following inhalation exposure in 12 rats. The plotted concentrations are scaled by the initial exposure (ppm). Dotted curves represent the observed trajectories with different plotting symbols distinguishing the four exposure groups. The solid curve is the predicted concentration profile based on the O2C model with $D = 1$, $\alpha = 0.595$, $\omega^{(1)} = 0.109$, and $\omega^{(2)} = 2.41$. The dashed curve is the prediction based on the PBPK model with parameters given in Table 1.1.*

variance for larger response.

In this example we consider two models to describe the chamber concentration for each animal. The first is the open two-compartment (O2C) model

$$f(t, \alpha, \omega^{(1)}, \omega^{(2)}; D) = D\{\alpha \exp(-\omega^{(1)} t) + (1 - \alpha) \exp(-\omega^{(2)} t)\}$$

for $t \in [0, 6]$, which predicts the expected concentration at a given time point for each subject. This model was described in more detail in Section 1.2. Because we scaled the concentration by initial dose, we will assume that $D = 1$ ppm for all subjects. To obtain preliminary estimates of the parameters α, $\omega^{(1)}$, and $\omega^{(2)}$, we fitted the O2C model to the data for each subject separately using ordinary least squares (SAS `proc nlin`), and averaged the coefficient estimates across the 12 subjects. The mean coefficient values were $\alpha = 0.595$, $\omega^{(1)} = 0.109$ 1/hr, and $\omega^{(2)} = 2.41$ 1/hr. The solid curve in Figure 5.1 plots the predicted concentration trajectory for these coefficient values.

The second model considered is the PBPK model that was described in Section 1.1.1. This model divides the body into four compartments: the liver, fat, well-perfused tissues (central nervous system and viscera except liver), and poorly-perfused tissues (muscles and skin). A fifth compartment represents the exposure chamber. A system of nonlinear differential equations describes the patterns of change in amounts or concentrations of the CCl_4 over time in each compartment. We used Matlab's `ode45` differential equation solver (Version 7 (R14), The MathWorks, Inc.) to approximate the solution to the system of differential equations, and to give a prediction of CCl_4 concentration in the chamber. For this analysis, we used the parameter values listed in Table 1.1 of Chapter 1. The dashed curve in Figure 5.1 plots the predicted concentration trajectory for the PBPK model.

In the Section 5.1.1 we apply the methods developed in Chapter 2 for exploring the marginal covariance of the dose-scaled CCl_4 data, and use this analysis to select random effects for the O2C and PBPK models. In Section 5.1.2 we apply the approach of Chapter 4 to test whether

there is significant evidence of AR(1) structure in the within-subject residuals when the O2C or PBPK models describe the expected trajectory for each rat. For the interpolated data there were $p = 36$ observations on $n = 12$ subjects. For the purpose of this analysis, we ignore the exposure covariate, and treat all 12 subjects as though they had the same covariates.

5.1.1 Exploring the marginal covariance structure

Choosing the number of common factors

Figure 5.2 shows the eigenvalues of the the sample covariance matrix $\mathbf{S}_n$ (note the log scale on the y-axis). Since $n < p$ in this example, the covariance matrix has rank 11, and the smallest $p - n + 1 = 25$ eigenvalues of $\mathbf{S}_n$ are zero. Recall that the eigenvalues are the variances of the sample principal components. The first two eigenvalues are more than ten-fold larger than the other eigenvalues. This suggests that two principal components should be sufficient to capture the between-subject variation in the data.

We used the Bartlett-corrected likelihood ratio statistic defined in equation (2.17) for $n < p$ to test whether a model with $k = 1, 2, 3$ or 4 principal components is adequate. Results are presented in Table 5.1. The PCA test results for this example suggest that more than four

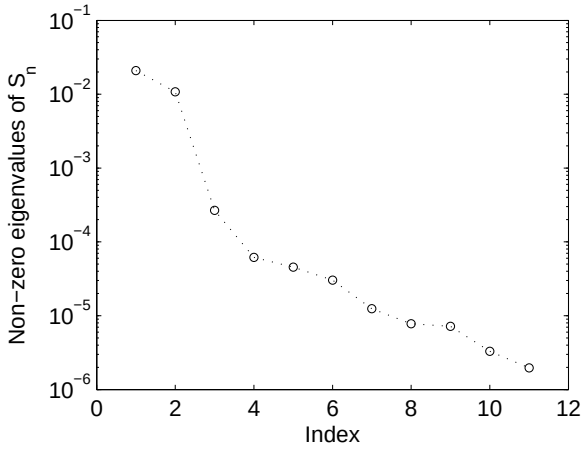

Figure 5.2: *Non-zero eigenvalues of the sample covariance matrix, $\mathbf{S}_n$, for the CCl_4 example.*

Table 5.1: *PCA test results for the hypothesis that k principal components are adequate for the CCl_4 data. The column LRT_k gives the value of the likelihood ratio test statistic based on the PCA model, and df_k gives the degrees of freedom of the test.*

k	LRT_k	df_k	p-value
1	399.1	54	$<.0001$
2	171.7	44	$<.0001$
3	135.6	35	$<.0001$
4	99.3	27	$<.0001$

principal components are needed. The simulation results of Chapter 3 showed that if within-subject variation is homoscedastic, then since $p = 3n$ in this example, the size of the PCA test should be close to the nominal level. However, if there is within-subject heteroscedasticity, then the smallest eigenvalues of $\mathbf{\Sigma}$ will not be equal, and the PCA test will tend to reject the hypothesis of k principal components too often when it is true.

Factor analysis allows a more flexible model for the within-subject variation than the principal component model, which assumes that $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$ for all i and j. The unrestricted factor model (MLFA-U) allows for a distinct variance at each time point, the only restriction being that the variances must be positive, $\mathbb{V}(\varepsilon_{ij}) = \psi_j > 0$. In Section 2.4.3 we also introduced a restricted factor model, MLFA-P, that models the within-subject variance as a power of the mean, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ where μ_j is the marginal mean at time t_j.

For the CCl_4 example, we fitted the MLFA-U and MLFA-P models with one, two, and three common factors. The iterative algorithm for fitting the MLFA-U model with four common factors had trouble converging, so we did not attempt to fit a model with more than three common factors. Since $n < p$ in this example, we could not use the Bartlett-corrected likelihood ratio tests to judge the adequacy of the k-factor models. An informal way to judge the adequacy of the k-factor models is to compute the percentage of total sample variation ($\text{tr}(\mathbf{S}_n)$) accounted for by the k common factors. For this example, we expect within-subject variation to be small relative to between-subject variation (see Figure 5.1), so an adequate factor model should

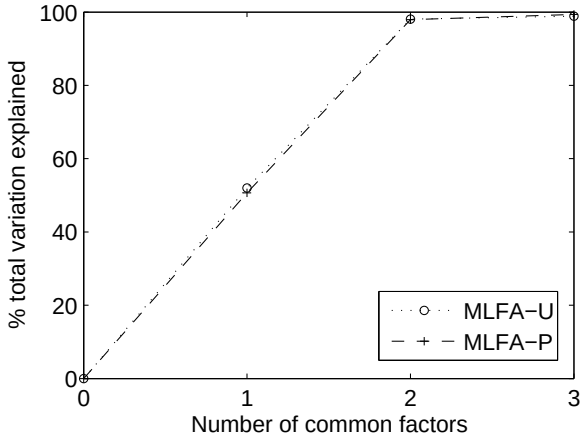

Figure 5.3: *Percentage of total sample variation explained by the k-factor MLFA-U and MLFA-P models for the CCl_4 example. The percentage of variation explained is plotted against the number of common factors included in the MLFA-U model (dotted curve) or in the MLFA-P model (dashed curve).*

explain a high percentage of the total sample variance. Figure 5.3 displays the percentage of variation explained plotted against the number of common factors included in the MLFA-U and MLFA-P models. The two-factor MLFA-U and MLFA-P models explained approximately 98% of the total sample variation. Including a third factor increased the percentage of variation explained to 99%. Based on the large magnitude of the first two principal components (relative to smaller eigenvalues) and the high percentage of variation explained by the two-factor models, we based the next stages of this analysis on models including two common factors.

Within-subject heteroscedasticity

As is common in much biological data, we expect to see within-subject heteroscedasticity for the CCl_4 data, perhaps with increased variability for larger responses. In Section 2.4.3 we described an approach to testing the hypothesis of homoscedastic within-subject variance by comparing the fit of the factor model (MLFA-H) that assumes homogeneous variance, $\mathbb{V}(\varepsilon_{ij}) = \psi_j = \sigma^2$ for all i and j, with the fit of the MLFA-P or MLFA-U models that specify more flexible models for the specific variances ψ_j.

Table 5.2: *Likelihood ratio test results for structure of the specific variances in the CCl_4 example. The results are based on fitting the MLFA-H, MLFA-P and MLFA-U models with 2 common factors. The first column denotes the models that are compared*

Comparison	LRT	*df*	p-value
MLFA-H v. MLFA-P	159.3	1	$<.0001$
MLFA-H v. MLFA-U	276.6	35	$<.0001$
MLFA-P v. MLFA-U	117.3	34	$<.0001$

In the simulation experiments of Chapter 3 we saw that the test comparing the MLFA-H model to the MLFA-P model had approximately the correct size when n and p were small (e.g., 10), but that the size of the test comparing the MLFA-H model with the MLFA-U model could be greatly inflated when n and p were small. Both tests were also affected by the presence of a large random effect, but primarily because we did not include sufficient common factors in the models used in the simulation experiments. This resulted in the estimates of within-subject variation being contaminated with unexplained between-subject variation. We would not expect this to be the case for the CCl_4 example, where two common factors explained approximately 98% of the total sample variation. We did not examine the behavior of the test comparing the MLFA-P model with the MLFA-U model in the simulation experiments, but expect that this test would behave similarly to the test comparing the MLFA-H and MLFA-U models.

Table 5.2 reports the results of the likelihood ratio tests comparing the fit of the MLFA-H, MLFA-P and MLFA-U models when two common factors were included. The fit of the MLFA-P and MLFA-U models were significantly better than that of the MLFA-H model. These tests provide strong evidence against the hypothesis of within-subject homoscedasticity. However, the fit of the MLFA-P model appears to be significantly worse than that of the unrestricted MLFA-U model, suggesting that the power model, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2\mu_j^{2\gamma}$, has some shortcomings. Based on the simulation experiences reported in Chapter 3, we interpret the tests based on comparison with the MLFA-U model with some degree of caution since n was small relative to

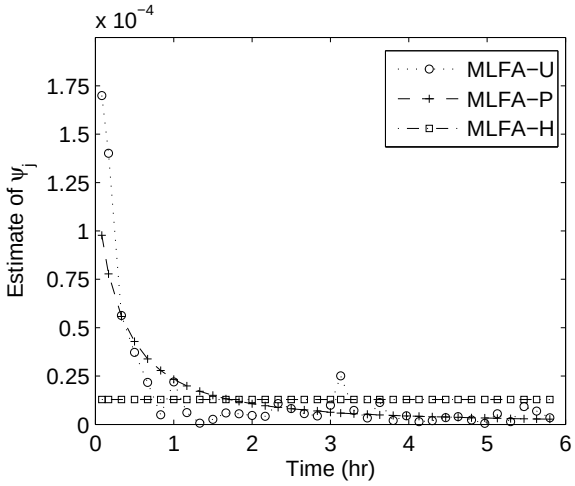

Figure 5.4: *Estimates of specific variances ψ_j based on fitting the 2-factor models. In each panel the estimates are shown for the MLFA-U model (open circles), MLFA-P model (plus signs), and MLFA-H model (open squares).*

p in this example.

Figure 5.4 shows the estimates of the specific variances ψ_j obtained by fitting the MLFA-U, MLFA-P and MLFA-H models with two common factors against time. The specific variances estimated by fitting the MLFA-U model suggest that within-subject variation was large early in the time course when response was highest, and decreased with time. The specific variances estimated by the MLFA-P model follow the same pattern as the MLFA-U estimates, although the power model was not able to capture the steep decline in specific variance during the first hour. This shortcoming most likely explains the lack of fit of the MLFA-P model compared to the MLFA-U model. The estimate of (γ, σ^2) obtained from fitting the two-factor MLFA-P model was $(1.86, 1.6\text{e}{-4})$. The MLFA-H model clearly does not capture the change in within-subject variation over time. The MLFA-H factor model is equivalent to the PCA model that assumes homogeneous within-subject variation. After examining evidence of heteroscedasticity, we conclude that PCA analysis is not appropriate for the CCl_4 example.

Because the MLFA-H and PCA models did not describe within-subject variation adequately, we have based further analysis on the loadings estimated from fitting the two-factor MLFA-U

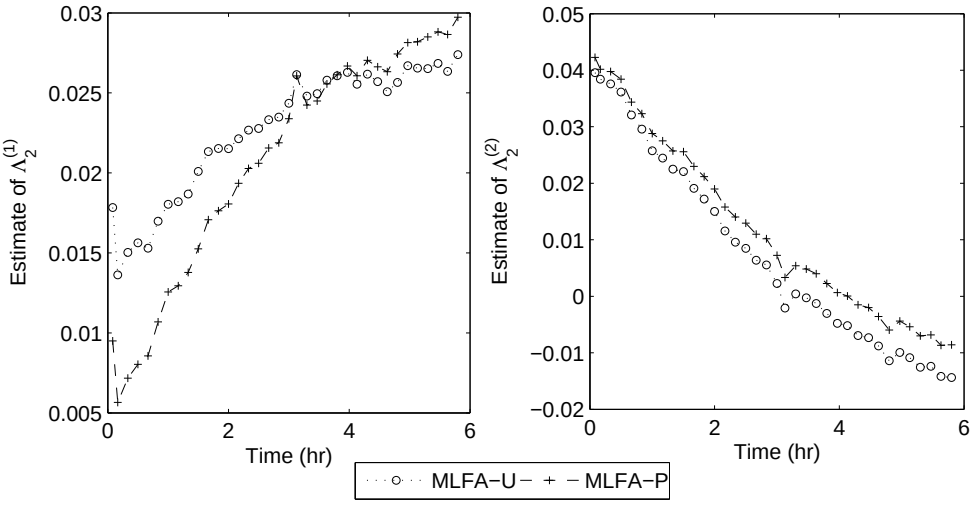

Figure 5.5: *Estimated factor loadings for the CCl_4 example based on the two-factor MLFA-U model (open circles) and two-factor MLFA-P model (plus signs). Estimates of the first loading are plotted against time in the left hand panel. Those for the second loading are plotted in the right hand panel.*

and MLFA-P models. Figure 5.5 displays the estimated factor loadings from the two-factor MLFA-U and MLFA-P models against time. The two models produced similar estimated loadings. The differences between the MLFA-U and MLFA-P estimates of the first loading (left panel in the figure) corresponded to the difference between the estimates of specific variance for the two models.

Selection of random effects and a parameterization: O2C model

We applied the multivariate forward selection procedure to pick sensitivity functions for the parameterization

$$f(t, \alpha, \omega^{(1)}, \omega^{(2)}; D) = D\{\alpha \exp(-\omega^{(1)} t) + (1 - \alpha) \exp(-\omega^{(2)} t)\}$$

of the O2C model. The selected sensitivities explain the variation in the factor loadings from the two-factor MLFA-U and MLFA-P models. At each step of the forward selection the 0.05

critical value of the test statistic was used. We considered the following first- and second-order sensitivities as candidates for forward selection,

$$\begin{aligned} f_{\alpha}(t) &\propto \exp(-\omega^{(1)}t) - \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}}(t) &\propto -\alpha t \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}}(t) &\propto -(1-\alpha) t \exp(-\omega^{(2)}t) \\ f_{\omega^{(1)}\omega^{(1)}}(t) &\propto \alpha t^2 \exp(-\omega^{(1)}t) \\ f_{\omega^{(2)}\omega^{(2)}}(t) &\propto (1-\alpha) t^2 \exp(-\omega^{(2)}t). \end{aligned}$$

We did not include the second-order sensitivities $f_{\alpha\alpha}$ and $f_{\omega^{(1)}\omega^{(2)}}$ which equal the zero function. The sensitivities $f_{\alpha\omega^{(1)}}$ and $f_{\alpha\omega^{(2)}}$ are proportional to $f_{\omega^{(1)}}$ and $f_{\omega^{(2)}}$, respectively and were also not included as candidate predictors.

Table 5.3 presents the forward selection results for the five candidate sensitivities corresponding to the parameters α, $\omega^{(1)}$ and $\omega^{(2)}$ of the O2C model. The analyses based on the two-factor MLFA-U and MLFA-P models agreed. Our analysis suggests that all three parameters need non-trivial random effects. However, since the model with two common factors explained most of the total sample variation, we might consider picking only the first two sensitivities that were selected, $f_{\omega^{(1)}}$ and $f_{\omega^{(2)}}$. These correspond to random effects on the two rate parameters, $\omega^{(1)}$ and $\omega^{(2)}$.

Table 5.3: *Multivariate forward selection results for the CCl_4 example and parameterization f of the O2C model. Selection was based on loadings of the 2-factor MLFA-U and MLFA-P models. Selection at each step was based on the 0.05 critical value of the test statistic. The p-values are sequential; i.e., calculated at the stage when the sensitivity entered the model.*

	#		Order selected (p-value)				
Model	selected	r_T	f_{α}	$f_{\omega^{(1)}}$	$f_{\omega^{(2)}}$	$f_{\omega^{(1)}\omega^{(1)}}$	$f_{\omega^{(2)}\omega^{(2)}}$
MLFA-U	5	0.991	3 (<.0001)	1 (<.0001)	2 (<.0001)	5 (0.0082)	4 (0.0001)
MLFA-P	5	0.991	3 (<.0001)	1 (<.0001)	2 (<.0001)	5 (0.0089)	4 (0.0001)

We also considered the alternative parameterization of the O2C model

$$f^*(t, \alpha, \delta^{(1)}, \delta^{(2)}; D) = D(1-\alpha)\exp(-\delta^{(2)}t)\left\{\frac{\alpha}{1-\alpha}\exp(-\delta^{(1)}t) + 1\right\}.$$

The coefficients in the alternative parameterization, corresponding to those used in parameterization f, were $D = 1$, $\alpha = 0.5950$, $\delta^{(1)} = \omega^{(1)} - \omega^{(2)} = -2.30$, $\delta^{(2)} = \omega^{(2)} = 2.41$. For the alternative parameterization we considered the following sensitivity functions as candidates for the multivariate regression model,

$$\begin{aligned}
f^*_{\alpha}(t) &\propto \exp\{-(\delta^{(1)} + \delta^{(2)})t\} - \exp(-\delta^{(2)}t) \\
f^*_{\delta^{(1)}}(t) &\propto -\alpha t \exp\{-(\delta^{(1)} + \delta^{(2)})t\} \\
f^*_{\delta^{(2)}}(t) &\propto -(1-\alpha)t\exp(-\delta^{(2)}t)\left\{\frac{\alpha}{1-\alpha}\exp(-\delta^{(1)}t) + 1\right\} \\
f^*_{\delta^{(1)}\delta^{(1)}}(t) &\propto \alpha t^2 \exp\{-(\delta^{(1)} + \delta^{(2)})t\} \\
f^*_{\delta^{(2)}\delta^{(2)}}(t) &\propto \alpha t^2 \exp\{-(\delta^{(1)} + \delta^{(2)})t\} + (1-\alpha)t^2\exp(-\delta^{(2)}t).
\end{aligned}$$

We did not include the second-order sensitivities $f^*_{\alpha\alpha}$ which equals the zero function, or the second-order sensitivities $f^*_{\alpha\delta^{(1)}}$, $f^*_{\alpha\delta^{(2)}}$, and $f^*_{\delta^{(1)}\delta^{(2)}}$ which can each be written as linear combinations of one or more of the sensitivities listed above.

Table 5.4 presents the forward selection results for the five candidate sensitivities corre-

Table 5.4: *Multivariate forward selection results for the CCl_4 example and the alternative parameterization f^* of the O2C model. Selection was based on loadings of the 2-factor MLFA-U and MLFA-P models. Selection at each step was based on the 0.05 critical value of the test statistic. The p-values are sequential; i.e., calculated at the stage when the sensitivity entered the model.*

	#		Order selected (p-value)				
Model	selected	r_T	f^*_{α}	$f^*_{\delta^{(1)}}$	$f^*_{\delta^{(2)}}$	$f^*_{\delta^{(1)}\delta^{(1)}}$	$f^*_{\delta^{(2)}\delta^{(2)}}$
MLFA-U	5	0.991	3 (<.0001)	1 (<.0001)	2 (<.0001)	5 (<.0001)	4 (0.0459)
MLFA-P	5	0.991	3 (<.0001)	1 (<.0001)	2 (<.0001)	5 (<.0001)	4 (0.0486)

sponding to parameters α, $\delta^{(1)}$ and $\delta^{(2)}$ in parameterization f^*. As in the analysis for the parameterization f, the sensitivities corresponding to the two rate parameters were picked first and second for the MLFA-U and MLFA-P analysis, suggesting that random effects on the rate parameters are largely responsible for between-subject variation in the CCl_4 example.

For this parameterization, it is interesting to note that if the 0.01 critical value had been chosen instead of the 0.05 value, then the forward selection algorithm would have stopped after choosing the sensitivity f_α; the second-order sensitivities would not have been added to the model. However, it is not clear that this indicates a preference for the parameterization f^* as all three parameters would still have been identified as requiring random effects. The trace correlation r_T for the multivariate regression selected for f^* is the same as that for the multivariate regression models selected for f. As discussed in Section 2.3.3, our method cannot distinguish between parameterizations (at least based on first-order sensitivities) when all parameters have random effects.

Selection of random effects: PBPK model

The physiologically based pharmacokinetic (PBPK) model has many physiological, physiochemical and biochemical model parameters, some of which are listed in Table 1.1, and others are calculated based on these parameters. Several of the parameters may be considered fixed or unlikely to vary from subject to subject. For example, the molecular weight (W_m) of CCl_4 is 153.8 g/mol in all cases. We will also assume that the tissue/air partition coefficients (and therefore the compartment/blood partition coefficients) do not vary from animal to animal. Although blood flow rates and tissue volumes are likely to vary from animal to animal, we assume that this variation can be explained by variation in individual animal cardiac output, Q_c, and body weight, B_w, respectively, and that the blood flow percentages and volume percentages are relatively constant across animals. Evans et al. (1994) explain that cardiac output Q_c was calculated using an allometric function of body weight, and the pulmonary ventilation rate Q_p was set equal to Q_c. However, for the purpose of our analysis we will assume that these

parameters, although probably related to body weight, potentially make their own contribution to between-subject variation in chamber concentration. Therefore, in order to characterize between-subject variation we focus on just a few of the PBPK model parameters as candidates for random effects. These are body weight, B_w, for each animal, the loss rate from an empty chamber K_L, the metabolic parameters V_{max} and K_m, total cardiac output, Q_c, and pulmonary ventilation rate, Q_p.

In order to apply the multivariate forward selection, we need to compute the sensitivity functions for each of the parameters identified in the paragraph above. In Section 1.4.1 we discussed numerical approximation of the parameter sensitivities when the model is defined by a system of differential equations. In Appendix D we have derived the sensitivity equations for the parameters B_w, K_L, V_{max}, K_m, Q_c, and Q_p. Using Matlab's `ode45` differential equation solver (Matlab 7.0 (R14), The MatWorks, Inc.), we numerically approximated the sensitivity functions for these parameters, evaluating them at values of the parameters given in Table 1.1 of Chapter 1 and at the $p = 36$ time points $t_1, t_2, \ldots, t_p$. We applied the multivariate forward selection procedure to pick a subset of these sensitivities that explain the variation in the factor loadings from the two-factor MLFA-U and MLFA-P models. We did not consider second-order sensitivities for the PBPK model.

Table 5.5 presents the forward selection results for the six PBPK model candidate sensitivities corresponding to body weight ($z_{h[B_w]}$), loss rate from the empty chamber ($z_{h[K_L]}$), the maximum rate of metabolism ($z_{h[V_{max}]}$), the Michaelis-Menten constant ($z_{h[K_m]}$), cardiac output ($z_{h[Q_c]}$), and ventilation rate ($z_{h[Q_p]}$). Five out of the six candidate sensitivities were included in the multivariate regression model fitted to the loadings of the two-factor MLFA-U and MLFA-P models. Figure 5.6 shows the first three PBPK sensitivities selected, those corresponding to V_{max}, B_w, and K_L.

The gas uptake experiment was primarily designed to allow estimation of the parameters describing metabolism of CCl_4 in the liver, V_{max} and K_m. The multivariate regression results indicated strong evidence that the maximum rate of metabolism V_{max} varied from animal to

Table 5.5: *Multivariate forward selection results for the CCl_4 example and the PBPK model. Selection was based on loadings of the 2-factor MLFA-U and MLFA-P models. Selection at each step was based on the 0.05 critical value of the test statistic. The p-values are sequential; i.e., calculated at the stage when the sensitivity entered the model.*

			Order selected (p-value)					
Model	# selected	r_T	$z_{h[B_w]}$	$z_{h[K_L]}$	$z_{h[V_{max}]}$	$z_{h[K_m]}$	$z_{h[Q_c]}$	$z_{h[Q_p]}$
MLFA-U	5	0.999	2	3	1	-	5	4
			(<.0001)	(<.0001)	(<.0001)	-	(<.0001)	(<.0001)
MLFA-P	5	0.999	2	3	1	-	5	4
			(<.0001)	(<.0001)	(<.0001)	-	(<.0001)	(<.0001)

animal, but that the Michaelis-Menten constant K_m could be treated as fixed. As might have been anticipated, there was strong evidence that variation in body weight B_w also contributed to between-animal variation in chamber concentration. Evans et al. (1994) reported that the body weight for the twelve subjects at time of exposure ranged from 0.165 to 0.205 kg. A reasonable approach would be to treat body weight as an observed animal-level covariate when fitting the PBPK model.

A multivariate regression model containing only the sensitivity $z_{h[V_{max}]}$ had trace correlation $r_T = 0.706$; including the sensitivity $z_{h[B_w]}$ increased r_T to 0.958. Since a factor model with two common factors explained 98% of total sample variation, putting a random effect on V_{max} and substituting the true values of each animals body weight, would explain most of the observed between-subject variation.

The third sensitivity to be selected, $z_{h[K_L]}$, corresponds to the loss rate from the empty chamber, K_L. According to the experimental protocol reported in Evans et al. (1994), the loss rate from the empty chamber was determined before each exposure. These authors reported that typical values for K_L ranged between 0.02 and 0.025 /hr, and were independent of the initial concentration in the chamber, `CONC`. These estimates of K_L could not be used directly as covariates, because after measuring K_L, the chamber was opened and aired before placing an

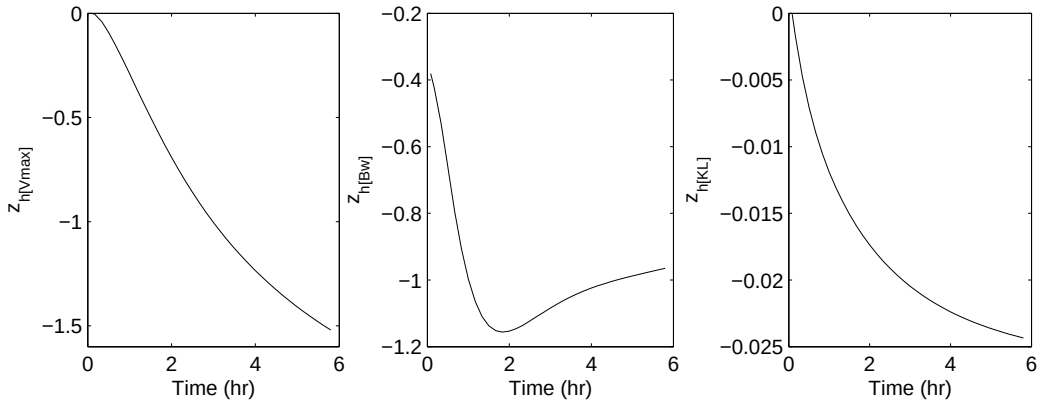

Figure 5.6: *Sensitivities for parameter V_{max} (left panel), B_w (center panel), and K_L (right panel) for the PBPK model. The sensitivities were evaluated at the parameters given in Table 1.1 of Chapter 1 and scaled to have norm $p^{1/2}$.*

animal in the chamber and resealing it. It is likely that the loss rate depended on the quality of the seal (a stainless steel lid clamped to a jar, with a thin layer of vacuum grease used as a sealant), which changed each time the chamber was opened.

The fourth and fifth sensitivities picked, correspond to the cardiac output, Q_c, and ventilation rate, Q_p, parameters. As mentioned earlier, these parameters are assumed to be strongly related to individual body weight. Their inclusion in the multivariate regression model further supports the inclusion of B_w as a measured covariate when fitting the PBPK model. However, the effect of body weight on between-subject variation in chamber concentrations appears to be mediated through the cardiac output, ventilation rate and derived tissue blood flow rates, not only through the compartment volumes.

5.1.2 Within-subject serial correlation

We applied the method developed in Chapter 4 to test for evidence of first-order autoregressive, AR(1), structure in the within-subject residuals for the CCl_4 data. As in the previous sections, we scaled the observed chamber concentrations by exposure dose, and pooled the data from the four exposure groups. In this example, $n = 12$ and $p = 36$. A set of $r = n - 1 = 11$

linearly independent between-subject differences, $\{\boldsymbol{y}_1-\boldsymbol{y}_2, \boldsymbol{y}_1-\boldsymbol{y}_3, \ldots, \boldsymbol{y}_1-\boldsymbol{y}_{12}\}$, were selected for the columns of the response matrix $\mathbf{W}$ for the multivariate regression of equation (4.2) in Chapter 4.

For the analysis of the O2C model, we set the parameters at the average of the least-squares estimates for each subject, $\alpha = 0.595$, $\omega^{(1)} = 0.109$ 1/hr, and $\omega^{(2)} = 2.41$ 1/hr. We set the exposure dose D equal to one, which could be a source of misspecification error. In the previous section, we identified that all three parameters, α, $\omega^{(1)}$, and $\omega^{(2)}$, required random effects. There was also some evidence that there were second-order effects on the two rate parameters. For the O2C model, the multivariate regression predictor matrix, $\mathbf{F}_\beta$, contained $q = 5$ sensitivities,

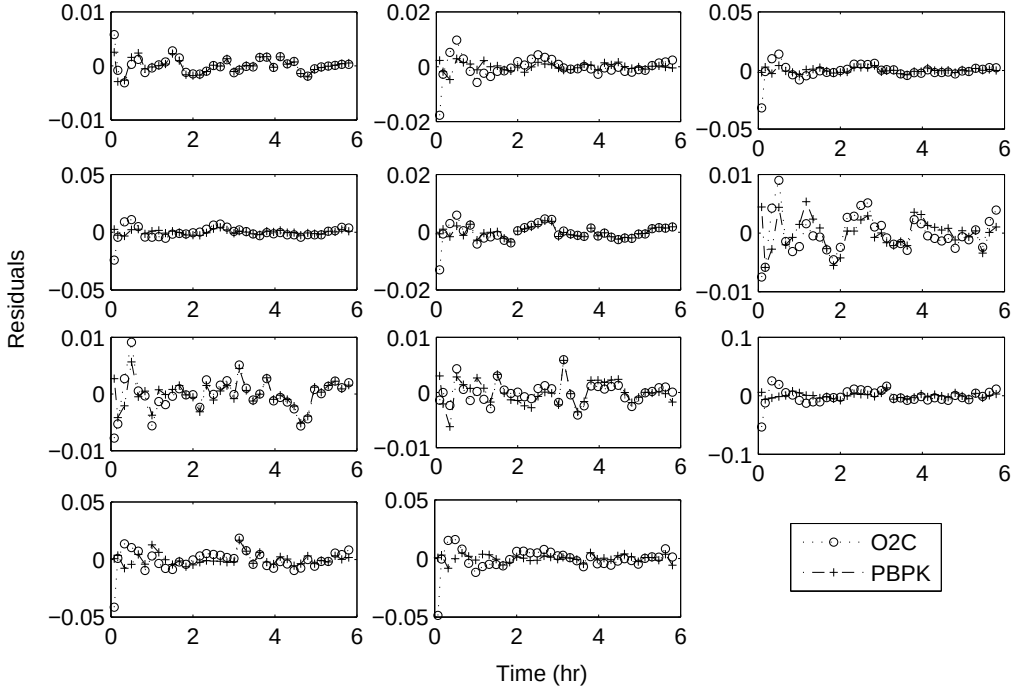

Figure 5.7: *Uncentered residuals from the multivariate regression of between-subject differences for the CCl_4 example. Each subplot graphs the residuals from one between-subject difference for the regression on the O2C sensitivities (open circles) and for the regression on the PBPK sensitivities (plus signs). The order of between subject differences is, row-wise from top left, $\boldsymbol{y}_1 - \boldsymbol{y}_2$, $\boldsymbol{y}_1 - \boldsymbol{y}_3$, $\boldsymbol{y}_1 - \boldsymbol{y}_4$, ...*

f_α, $f_{\omega^{(1)}}$, $f_{\omega^{(2)}}$ and the second-order sensitivities $f_{\omega^{(1)}\omega^{(1)}}$ and $f_{\omega^{(2)}\omega^{(2)}}$.

In analyzing the PBPK model, we fixed all parameters at the values given in Table 1.1 of Chapter 1. There may be misspecification in several of these parameters. In the previous section, we identified five coefficients as requiring random effects, B_w, K_L, V_{max}, Q_c, and Q_p. The predictor matrix, $\mathbf{F}_\beta$, for the PBPK model contained $q = 5$ columns corresponding to the sensitivities of these parameters.

Figure 5.7 displays the uncentered residuals plotted against time for the set of $r = 11$ between-subject differences, $\boldsymbol{y}_i - \boldsymbol{y}_{i'}$. Each panel shows the residuals for a particular between-subject difference. Subjects 1 to 3 were exposed at 1000 ppm, subjects 4 to 6 at 250 ppm, subjects 7 to 9 at 100 ppm, and subjects 10 to 12 at 25 ppm. The residuals were largest (last three plots) when the between subject differences were between animal 1 (exposed at 1000 ppm) and animals exposed at 25 ppm. The plots do not show strong evidence of within-subject autocorrelation, nor strong evidence of heteroscedasticity. For the O2C analysis, the residual at the first time point tends to be larger than at other time points.

The fitted multivariate regression model for between-subject differences had a trace correlation of $r_T = 0.64$ for both the the O2C and the PBPK analyses. The values of $r_{(i,i')}$ for the 11 between-subject differences ranged from -0.06 to 0.39 for the O2C analysis, and from -0.13 to 0.30 for the PBPK analysis. Based on equation (4.5), the estimated AR(1) correlation, ρ, was 0.27 for the O2C analysis and 0.12 for the PBPK analysis. The hypothesis of no AR(1) correlation structure, H_0: $\rho = 0$, was not rejected at the 0.05 significance level for either model. There does not appear to be strong evidence of within-subject serial correlation in the CCl_4 gas uptake example.

5.2 Pharmocokinetics of theophylline

We described the theophylline pharmacokinetic data in Section 1.1 of Chapter 1. On each of twelve subjects, a measurement of the blood serum concentration (mg/L) of theophylline was

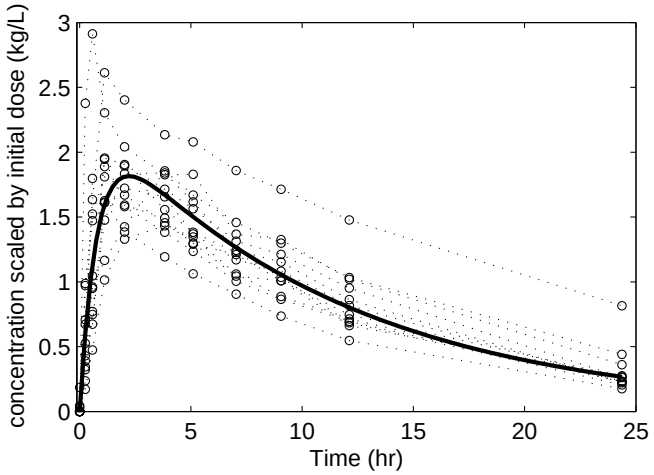

Figure 5.8: *Serum concentration of theophylline following oral dose in twelve subjects. The plotted concentrations are scaled by the initial dose (per body weight). Dotted curves represent the observed trajectories. The solid curve is the predicted concentration profile based on the O1C model with* $D = 1$*,* $k_a = 1.303$*,* $Cl = 0.0405$*, and* $V_d = 0.452$*.*

taken before dosing with theophylline, and at 10 subsequent time points over the next 24 hours. Figure 5.8 displays the measured serum concentration profiles for the twelve subjects. We scaled the concentrations by the initial dose (mg/kg of body weight). Scaling by dose implies a belief that that, on the original scale, measurement error variance is heterogeneous (a function of initial dose), with larger variance for larger response.

The open one-compartment (O1C) model

$$f(t, k_a, V_d, Cl; D) = \frac{Dk_a}{V_d(k_a - Cl/V_d)} \left\{ \exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t) \right\}$$

for $t \in [0, 25]$ predicts the expected concentration at a given time point for each subject. This model was described in more detail in Section 1.2. Because we scaled the concentration by initial dose, we will assume that $D = 1$ kg/L for all subjects. Based on estimates obtained by Davidian and Giltinan (e.g., 1995, Table 5.3) we chose the following mean values for the model coefficients: $k_a = 1.303$ 1/hr, $Cl = 0.0405$ L/hr/kg, and $V_d = 0.452$ L/kg. The solid curve in Figure 5.8 graphs the predicted concentration trajectory for these coefficient values. These

authors allowed all three parameters to vary across the population of subjects, and identified the presence of within-subject heteroscedasticity, with larger residual variation associated with larger response. They modeled within-subject variation using

$$\mathbb{V}(\varepsilon_{ij}) = \sigma^2\{f(t_j, k_{ai}, Cl_i, V_{di}; D_i)\}^{2\gamma}$$

and estimated $\gamma \approx 0.6$ and $\sigma^2 \approx 0.1$.

In the Section 5.2.1 we apply the methods developed in Chapter 2 to explore the structure of the marginal covariance of the theophylline data, and use the results to select random effects for the O1C model. In Section 5.2.2 we use the approach of Chapter 4 to determine whether there is significant evidence of AR(1) structure in the within-subject residuals when the O1C model is used to describe the subject trajectories. Several of the concentration values measured at baseline were zero (these subjects most likely had no previous exposure to theophylline). The zero values make it difficult to model residual variation at time zero. We excluded the baseline measurement for all subjects from our analysis; as a result there were $p = 10$ observations on $n = 12$ subjects.

5.2.1 Exploring the marginal covariance structure

Choosing the number of common factors

Figure 5.9 shows the eigenvalues of the the sample covariance matrix $\mathbf{S}_n$ for the theophylline example. Recall that the eigenvalues are the variances of the sample principal components. The plot suggests that three principal components should be sufficient to capture the between-subject variation in the data. Because we anticipate within-subject heteroscedasticity in the theophylline data set, we have focused on the maximum likelihood analysis for this data set, rather than the principal component analysis, which assumes homogeneous within-subject variation. The simulation results presented in Chapter 3 illustrated that the principal component analysis could be misleading in the presence of heteroscedasticity.

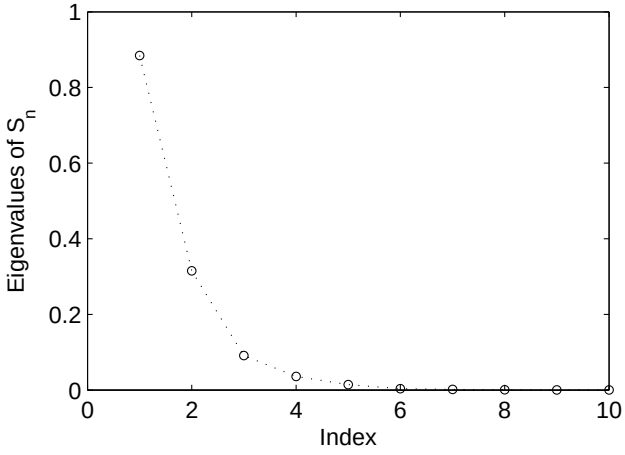

Figure 5.9: *Eigenvalues of the sample covariance matrix,* $\mathbf{S}_n$*, for the theophylline example.*

The factor analysis model allows a more flexible model for the within-subject variation than the principal component model which assumes that $\mathbb{V}(\varepsilon_{ij}) = \sigma^2$ for all i and j. The unrestricted factor model (MLFA-U) allows for a distinct variance at each time point, the only restriction being that the variances must be positive, $\mathbb{V}(\varepsilon_{ij}) = \psi_j > 0$. We also fitted the MLFA-P model, that models the within-subject variance as a power of the mean, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2 \mu_j^{2\gamma}$ where μ_j is the marginal mean at time t_j.

For the theophylline example, we fitted the MLFA-U and MLFA-P models with one, two, three, four and five common factors. The MLFA-U model with six common factors has the same number of parameters ($pk + p - \frac{1}{2}k(k-1) = 55$) as there are unique elements in $\mathbf{S}_n$ ($\frac{1}{2}p(p+1) = 55$); the six-factor model offers no simplification of the covariance structure. Therefore we did not attempt to fit a model with six common factors. An informal way to judge the adequacy of the k-factor models is to compute the percentage of total sample variation ($\text{tr}(\mathbf{S}_n)$) accounted for by the k common factors. For this example, we expect within-subject variation to be small relative to between-subject variation (see Figure 5.8), so an adequate factor model should explain a high percentage of the total sample variance. Figure 5.10 shows the percentage of variation explained against the number of common factors included in the MLFA-

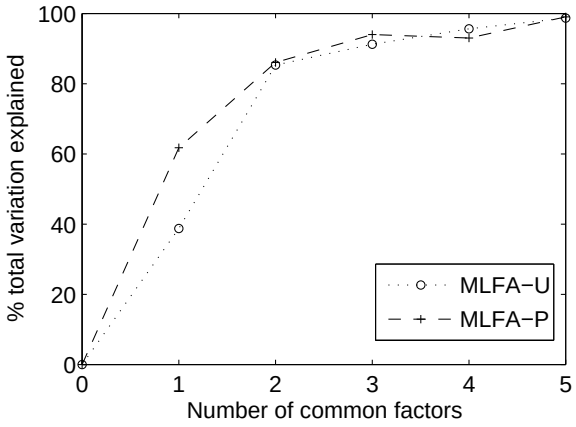

Figure 5.10: *Percentage of total sample variation explained by the k-factor MLFA-U and MLFA-P models for the theophylline example. The percentage of variation explained is plotted against the number of common factors included in the MLFA-U model (dotted curve) or in the MLFA-P model (dashed curve).*

U and MLFA-P models. The two-factor MLFA-U and MLFA-P models explained greater than 85% of the total sample variation. Including a third factor increased the percentage of variation explained to over 90%. A fourth factor increased the percentage of variation explained to 95% for the MLFA-U model, but did not improve the fit of the MLFA-P model. The five-factor MLFA-U and MLFA-P models explained 99% of total sample variation.

Since $n > p$ in this example, we were able to employ the Bartlett-corrected likelihood ratio test described in equation (2.25) of Section 2.4.2, to test whether k common factors were sufficient to adequately models the marginal covariance matrix $\mathbf{\Sigma}$. We treat the results of this test with a degree of caution here, since the number of subjects, n, is not much larger than the number of time points, p. The simulation results of Chapter 3 suggest that the size of the test may be larger than the nominal level when $n \approx p$.

The MLFA-P model has fewer parameters than the MLFA-U model, and it is possible that the smaller model may provide a more reliable test that a particular factor model adequately models the marginal covariance matrix $\mathbf{\Sigma}$. We did not discuss this test in Chapter 2, and develop it briefly here. Let $\widehat{F}_k^{ps}$ be the minimized value of the objective function for the MLFA-

P model with k common factors, and let $\widehat{F}_{\text{full}} = \log|\mathbf{S}_n| + p$ be the objective function evaluated at $\mathbf{S}_n$, which is the MLE of $\boldsymbol{\Sigma}$ under the unrestricted model. The likelihood ratio test statistic T_λ for comparing the MLFA-P model with k common factors to an unrestricted model for $\boldsymbol{\Sigma}$ is given by

$$-2\log T_\lambda = n(\widehat{F}_k^{ps} - \widehat{F}_{\text{full}}) \equiv U_k^{ps}.$$

By the usual arguments about likelihood ratio tests, when the k-factor MLFA-P model is true, the statistic U_k^{ps} has an asymptotic chi-squared distribution with

$$df_k = \tfrac{1}{2}p(p+1) - \{2 + pk - \tfrac{1}{2}k(k-1)\} = \tfrac{1}{2}(p-k)^2 + \tfrac{1}{2}(p-k-4)$$

degrees of freedom. We use the statistic U_k^{ps} to test the hypothesis that the k-factor MLFA-P model gives an adequate description of $\boldsymbol{\Sigma}$, compared to the unrestricted estimator, $\mathbf{S}_n$. Note that the test based on U_k^{ps} may reject the null hypothesis if k common factors are insufficient to adequately describe the between-subject variation, or if the power model for $\boldsymbol{\Psi}$ is not appropriate for modeling within-subject variances. We have not checked the small-sample performance of the likelihood ratio test based on the MLFA-P model in simulation experiments, but for interest sake, we apply it in this analysis.

Table 5.6: *Factor analysis test results for the hypothesis that k common factors are adequate for the theophylline data. The column LRT_k gives the value of the likelihood ratio test statistic based on the MLFA-U or MLFA-P model, and df_k gives the degrees of freedom of the test.*

	MLFA-U			MLFA-P		
k	LRT_k	df_k	p-value	LRT_k	df_k	p-value
1	71.6	35	0.0003	275.9	43	$<.0001$
2	51.6	26	0.0020	195.2	34	$<.0001$
3	38.6	18	0.0032	149.9	26	$<.0001$
4	25.6	11	0.0073	117.0	19	$<.0001$
5	11.8	5	0.0383	63.8	13	$<.0001$

Table 5.6 summarizes the results of the likelihood ratio tests based on the MLFA-U and MLFA-P models. In all cases, the MLFA-P model showed significant lack of fit. As we shall see in the next section, the MLFA-P model is most likely not appropriate for modeling within-subject variation in this experiment, and so the very small p-values based on the fit of the MLFA-P model could be due to inappropriateness of the model for $\mathbf{\Psi}$, rather than a need for more common factors. The MLFA-U models all showed some evidence of lack of fit. Comparing the p-values to the 0.01 significance level, the five-factor model appears to provide an adequate fit. This model explained 99% of total sample variation (Figure 5.10). However the three- or four-factor models also explained a high percentage of total sample variation. For the next stages of this analysis we compare the results based on models with three, four or five common factors.

Within-subject heteroscedasticity

As is common in pharmacokinetic data, we expect to see within-subject heteroscedasticity here, perhaps with increased variability for larger responses. In Section 2.4.3 we described an approach to testing the hypothesis of homoscedastic within-subject variance by comparing the fit of the MLFA-H model with the fit of the MLFA-P or MLFA-U models.

Based on the simulation results of Chapter 3, we expect the the test comparing the MLFA-H model to the MLFA-P model to have approximately the correct size in this example. Because of the small sample size, the size of the test comparing the MLFA-H model with the MLFA-U model could be greatly inflated.

Table 5.7 reports the results of the likelihood ratio tests comparing the fit of the MLFA-H, MLFA-P and MLFA-U models when two, three, four, or five common factors were included. Except in the two-factor model, the specific variances estimated by fitting the MLFA-U model suggest that within-subject variation was larger at the half and one hour time points than at other measurements occasions. Peak serum theophylline concentration was observed at these times, supporting the view that larger measurement error occurred when the response was large.

Table 5.7: *Likelihood ratio test results for structure of the specific variances in the theophylline example. The results are based on fitting the MLFA-H, MLFA-P and MLFA-U models with* $k = 2$, 3, 4 *and* 5 *common factors.*

	Model comparison								
	MLFA-H v. MLFA-P			MLFA-H v. MLFA-U			MLFA-P v. MLFA-U		
k	LRT	*df*	p-value	LRT	*df*	p-value	LRT	*df*	p-value
2	4.68	1	0.0306	87.4	9	$<.0001$	82.7	8	$<.0001$
3	0.147	1	0.7017	54.3	9	$<.0001$	54.1	8	$<.0001$
4	-9.99	1	1.0000	33.2	9	0.0001	43.2	8	$<.0001$
5	0.893	1	0.3447	24.4	9	0.0037	23.5	8	0.0028

In all cases, the MLFA-P model did not fit significantly better than the MLFA-H model that assumes homogeneous variance. In fact, in the four factor case, the fit of the MLFA-P model was slightly worse (negative value of the LRT statistic). Comparing the MLFA-H or MLFA-P model to the unrestricted MLFA-U model suggests that there is significant heteroscedasticity in the within-subject variation, but that it is not well described by the power model, $\mathbb{V}(\varepsilon_{ij}) = \sigma^2\mu_j^{2\gamma}$.

Figure 5.11 displays the estimates of the specific variances ψ_j obtained by fitting the MLFA-U, MLFA-P and MLFA-H models with two, three, four, and five common factors against time. Note the log scale on the x-axis and the different y-axis scales in the upper and lower panels. The estimates of ψ_j decreased in magnitude as the number of common factors included increased.

In the three and five factor models, the estimates of ψ_j from the MLFA-P were similar to those from fitting the MLFA-H model, explaining why the likelihood ratio tests did not reject the hypothesis of homoscedasticity by comparing the fit of the MLFA-H model to the MLFA-P model. In the four-factor model, the MLFA-P estimates are different from the MLFA-H estimates, but because this model was not able to model the specific variances well (perhaps because of the within-subject variation remains large as time approaches 24 hours), the fit of this model was poor compared with that of the simpler MLFA-H model. In the two-factor model, the MLFA-P model fitted slightly better than the MLFA-H model, but did not adequately

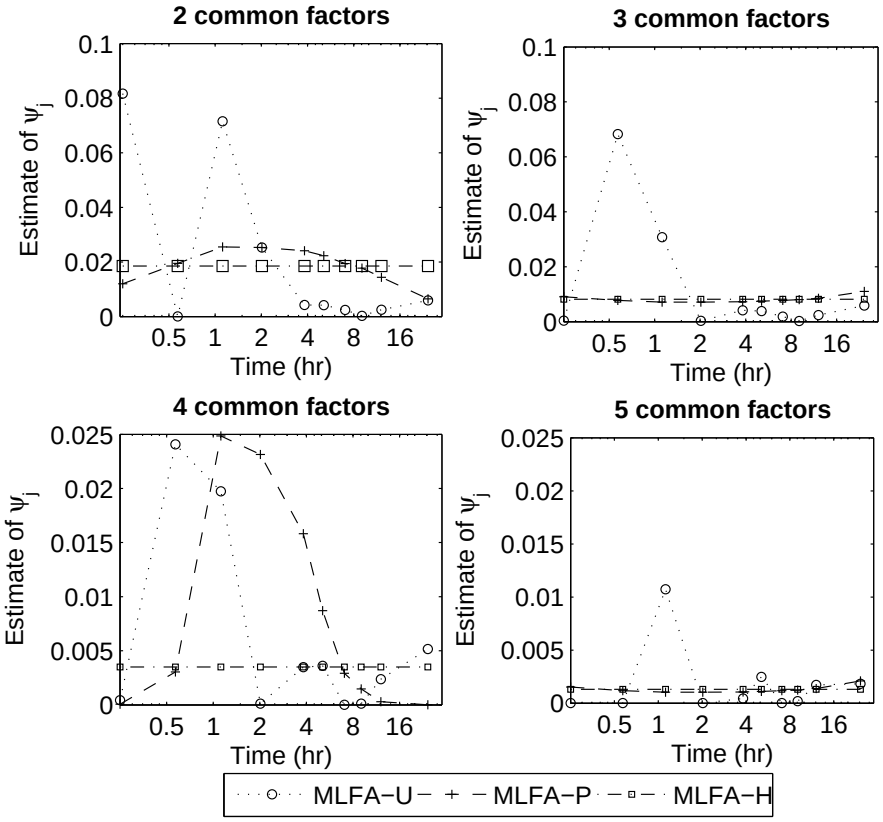

Figure 5.11: *Estimates of specific variances ψ_j based on fitting the 2-factor, 3-factor, 4-factor, and 5-factor models. In each panel the estimates are shown for the MLFA-U model (open circles), MLFA-P model (plus signs), and MLFA-H model (open squares).*

capture the pattern in the unrestricted estimates of ψ_j.

The estimate of (γ, σ^2) obtained from fitting the MLFA-P model was $(0.397, 0.0163)$ with two factors, $(-0.125, 0.0082)$ when three factors were included, $(3.09, 0.0008)$ with four factors, and $(-0.206, 0.0013)$ for the five-factor model. The negative estimates of γ in the three- and five-factor MLFA-P model were unexpected. In the panels in the right hand column of Figure 5.11, the MLFA-P estimates of ψ_j increased slightly at the end of the time course, suggesting that within-subject variance increased with decreasing expected response. This surprising result is probably caused by the slightly larger specific variance observed at the end of the time course and the leverage of the final time point (there was approximately a 12 hour time difference

between the penultimate and final measurement times).

Davidian and Giltinan (1995, Section 5.5) modeled the within-subject variance for the theophylline data with the power-of-the-mean model

$$\mathbb{V}(\varepsilon_{ij}) = \sigma^2 f^{2\gamma}(t_j, k_{ai}, Cl_i, V_{di}; D_i)$$

at the subject-level (i.e., conditional on the random effects). If measurement error variance is proportional to some power of the expected response, then it makes sense to specify this model at the subject level; the residual variance for a measurement on a subject would be related to that subject's expected response (the conditional mean), not the average response, at that time point. Our factor analysis approach models the specific variances as a function of the marginal mean μ_j.

We mention this distinction here to clarify that although our analysis did not support the power-of-the-mean variance model used by Davidian & Giltinan, perhaps more investigation would be in order before rejecting the power-of-the-mean model for this example. The difference between modeling variance as a function of the marginal mean versus the conditional mean may be significant in the theophylline example, where the expected response could differ by a factor of two from subject to subject. In the simulation experiments of Chapter 3, we generated data where the residual variance at time point t_j was proportional to a power of the marginal mean at that time point. We have not investigated how different results would be if variance was actually a function of the subject-specific mean. We expect that if random effects are small then differences should not be too great.

Selection of random effects and a parameterization

Because the MLFA-P and MLFA-H models did not describe within-subject variation adequately, we have based further analysis on the loadings estimated from fitting the two, three, four and five factor MLFA-U models. We applied the multivariate forward selection procedure to pick

sensitivity functions for the volume-clearance parameterization of the O1C model that explain the variation in the factor loadings from the fitted MLFA-U models. At each step of the forward selection we used the 0.05 critical value of the test statistic. We considered the following first- and second-order sensitivities as candidates for forward selection,

$$
\begin{aligned}
f_k(t) &\propto \frac{-Cl/V_d}{k_a - Cl/V_d}\left\{\exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t)\right\} + k_a t\exp(-k_a t) \\
f_V(t) &\propto \frac{Cl}{V_d}t\exp\left(-\frac{Cl}{V_d}t\right) - \left\{1 + \frac{Cl}{V_d(k_a - Cl/V_d)}\right\}\left\{\exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t)\right\} \\
f_C(t) &\propto \frac{1}{k_a - Cl/V_d}\left\{\exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t)\right\} + t\exp\left(-\frac{Cl}{V_d}t\right) \\
f_{kk}(t) &\propto \frac{-2}{k_a - Cl/V_d}\left\{\exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t)\right\} + 2t\exp(-k_a t) \\
&\quad -\frac{2k_a}{ka - Cl/V_d}t\exp(-k_a t) - k_a t^2\exp(-k_a t) \\
&\quad -\frac{2k_a}{(k_a - Cl/V_d)^2}\left\{\exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t)\right\} \\
f_{CC}(t) &\propto t^2\exp\left(-\frac{Cl}{V_d}t\right) - \frac{2}{k_a - Cl/V_d}t\exp\left(-\frac{Cl}{V_d}t\right) \\
&\quad +\frac{2}{(k_a - Cl/V_d)^2}\left\{\exp\left(-\frac{Cl}{V_d}t\right) - \exp(-k_a t)\right\}
\end{aligned}
$$

for $t \in [0, 25]$. We did not consider the second-order sensitivity f_{VC} for parameters V_d and Cl because it is is a linear function of other sensitivities. There was a linear dependency among f_V, f_C, f_{VV} and f_{CC}; we chose to include f_{CC} instead of f_{VV} because in preliminary analysis we found that the sensitivity f_C was picked but not f_V ("hierarchy of predictors" principle).

Table 5.8 presents the forward selection results for the five candidate sensitivities corresponding to the parameters k_a, Cl and V_d of the O1C model. To a large extent, there was agreement between the analyses based on two, three, four or five common factors. Our analysis suggests that the absorption rate coefficient k_a and the clearance parameter Cl require non-trivial, but small, random effects. The analysis based on the two-factor model suggested that the random effect on the absorption rate k_a may be large (the second-order sensitivity was

Table 5.8: *Multivariate forward selection results for the theophylline example and the volume-clearance parameterization of the O1C model. Selection was based on loadings of the k-factor MLFA-U model including $k = 2, 3, 4, 5$ common factors. Selection at each step was based on the 0.05 critical value of the test statistic. The p-values are sequential; i.e., calculated at the stage when the sensitivity entered the model.*

			Order selected (p-value)				
k	# selected	r_T	f_k	f_V	f_C	f_{kk}	f_{CC}
2	4	0.990	3 (0.0020)	4 (0.0183)	1 (<.0001)	2 (<.0001)	-
3	3	0.848	1 (<.0001)	3 (0.0160)	2 (<.0001)	-	-
4	3	0.781	1 (<.0001)	-	2 (0.0001)	3 (0.0174)	-
5	2	0.630	1 (<.0001)	-	2 (0.0002)	-	-

selected).

We also considered the alternative parameterization of the O1C model

$$f^*(t, \delta^{(1)}, \delta^{(2)}, \delta^{(3)}; D) = D\delta^{(1)} \left\{ \exp(-\delta^{(2)}t) - \exp(-\delta^{(3)}t) \right\}$$

where $t \in [0, 25]$. The coefficients in the alternative parameterization corresponding to those used for the volume-clearance parameterization were $D = 1$, $\delta^{(1)} = k_a/\{V_d(k_a - Cl/V_d)\} = 2.38\,\text{kg/L}$, $\delta^{(2)} = Cl/V_d = 0.0896\,\text{/hr}$, and $\delta^{(3)} = k_a = 1.303\,\text{/hr}$. For the alternative parameterization we considered the following sensitivity functions as candidates for the multivariate regression model,

$$\begin{aligned}
f_1^*(t) &= D \left\{ \exp(-\delta^{(2)}t) - \exp(-\delta^{(3)}t) \right\} \\
f_2^*(t) &= -D\delta^{(1)} t \exp(-\delta^{(2)}t) \\
f_3^*(t) &= -D\delta^{(1)} t \exp(-\delta^{(3)}t) \\
f_{22}^*(t) &= D\delta^{(1)} t^{*2} \exp(-\delta^{(2)}t) \\
f_{33}^*(t) &= D\delta^{(1)} t^{*2} \exp(-\delta^{(3)}t)
\end{aligned}$$

Table 5.9: *Multivariate forward selection results for the theophylline example and the alternative parameterization of the O1C model. Selection was based on loadings of the k-factor MLFA-U model including $k = 2, 3, 4, 5$ common factors. Selection at each step was based on the 0.05 critical value of the test statistic. The p-values are sequential; i.e., calculated at the stage when the sensitivity entered the model.*

			Order selected (p-value)				
k	# selected	r_T	f_1^*	f_2^*	f_3^*	f_{22}^*	f_{33}^*
2	4	0.990	1 (<.0001)	2 (<.0001)	3 (0.0009)	-	4 (0.0178)
3	3	0.848	3 (0.0089)	2 (<.0001)	1 (<.0001)	-	-
4	3	0.844	-	2 (0.0002)	1 (<.0001)	-	3 (0.0144)
5	3	0.764	-	2 (0.0002)	1 (<.0001)	-	3 (0.0233)

for $t \in [0, 25]$. We did not include the second-order sensitivities f_{11}^* and f_{23}^* which equal the zero function, or the second-order sensitivities f_{12}^* and f_{13}^* which are proportional to f_2^* and f_3^* respectively.

Table 5.9 presents the forward selection results for the five candidate sensitivities corresponding to coefficients $\delta^{(1)}$, $\delta^{(2)}$, and $\delta^{(3)}$ in parameterization f^*. The analysis based on three, four, or five common factors suggested that the two rate coefficients $\delta^{(3)}$ (corresponding to k_a) and $\delta^{(2)}$ (corresponding to Cl/V_d) require non-trivial random effects. Recall that for the volume-clearance parameterization f, the two parameters identified for random effects were k_a and Cl. The three-, four-, and five-factor analyses for the two parameterizations both identify that random variation in the rate coefficients is largely responsible for observed between-subject variation.

The analysis based on the two-factor and three-factor models also identified a random effect on the scale coefficient, $\delta^{(1)}$. In the analysis for parameterization f, the first-order sensitivities for all three parameters were selected based on three common factors. The trace correlation r_T was the same ($r_T = 0.990$ and 0.848 respectively) for the two- and three-factor analyses for parameterizations f and f^*. As discussed in Section 2.4.4, when there are random effects on all parameters in a model, our method cannot distinguish between alternative parameterizations.

For the analyses based on four or five common factors, the regression models selected for f^* had slightly higher trace correlation than those selected for f. This suggests that the alternative parameterization f^*, with a scale and two rate coefficients, is better able to characterize the between-subject variation with two random rate coefficients, than the volume-clearance parameterization, where random effects on k_a and Cl induce random variation in the scale, $Dk_a/\{V_d(k_a - Cl/V_d)\}$, which is not evident in the observed data.

5.2.2 Within-subject serial correlation

We applied the method developed in Chapter 4 to test whether there was evidence of first-order autoregressive, AR(1), structure in the within-subject residuals for the theophylline data. We scaled the serum concentrations by the administered dose (on the mg/kg scale), so that the dose covariate could be ignored in comparing trajectories. When fitting the O1C model, none of the parameters were fixed at pre-assigned values. However, we did assume that the dose, D, and fraction of drug absorbed, `f`, were one. This could be viewed as a potential source of model misspecification.

In the previous section, we identified that the parameters k_a and Cl required non-trivial random effects. The matrix of predictors, $\mathbf{F}_\beta$, for the multivariate regression of between-subject differences (equation (4.2) of Chapter 4) had $q = 2$ columns corresponding to the sensitivities f_k and f_C. In this example, $n = 12$ and $p = 10$ (the time zero observation was excluded), so $r = p - q - 1 = 7$ sets of linearly independent between-subject differences were randomly selected from the 11 possible between-subject differences. The resulting columns of $\mathbf{W}$ were

$$\mathbf{W} = (\boldsymbol{y}_1 - \boldsymbol{y}_3, \boldsymbol{y}_1 - \boldsymbol{y}_5, \boldsymbol{y}_1 - \boldsymbol{y}_4, \boldsymbol{y}_1 - \boldsymbol{y}_9, \boldsymbol{y}_1 - \boldsymbol{y}_2, \boldsymbol{y}_1 - \boldsymbol{y}_6, \boldsymbol{y}_1 - \boldsymbol{y}_{11}).$$

Figure 5.12 shows the uncentered residuals plotted against time for the seven between-subject differences (note the log scale on the time axis). With only a few time points ($p = 10$), it is difficult to detect any evidence of serial correlation in the plot. The multivariate residuals

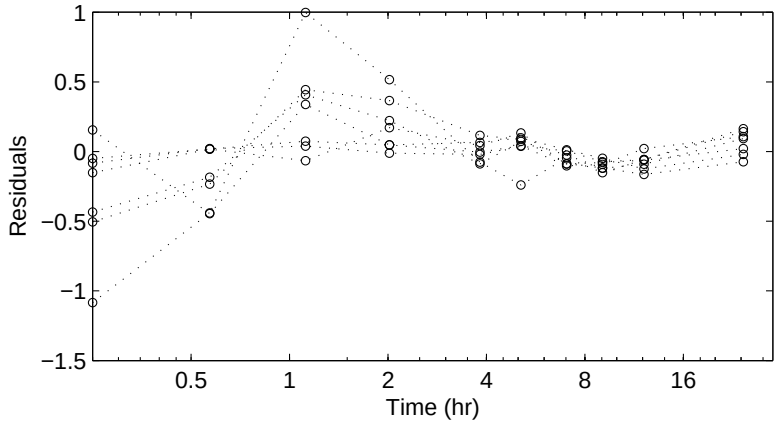

Figure 5.12: *Uncentered residuals from the multivariate regression of between-subject differences for the theophylline example.*

showed larger variation early in the time course, and less variation as time progressed. This confirms the evidence of within-subject heteroscedasticity detected by the factor analysis results presented in the previous section.

The fitted multivariate regression model had a trace correlation of $r_T = 0.51$. The values of $r_{(i,i')}$ for the seven between-subject differences ranged from -0.61 to 0.43. Based on equation (4.5), the estimated AR(1) correlation, ρ, was 0.053, and the hypothesis H_0: $\rho = 0$ was not rejected at the 0.05 significance level. There is little evidence of within-subject serial correlation in the theophylline example. Admittedly, the simulation results of Chapter 4 showed that power to detect AR(1) structure was low when there were few time points (e.g., $p = 10$), as in this example.

5.3 Summary of applications

In this chapter we applied our methods for exploring the marginal covariance structure, and examining within-subject autocorrelation, to the CCl_4 gas uptake experiment, and to data collected to characterize the pharmacokinetics of the drug theophylline. In the CCl_4 example, $n = 12$ and $p = 36$. For the theophylline example $n = 12$ and $p = 10$.

In the CCl_4 example, we concluded that two common factors were needed to capture between-subject variation. A model with two common factors captured 98% of total sample variation. Since $n < p$, we could not use the likelihood ratio test based on the fit of the factor model to test the adequacy of this model. The PCA test, that rejected adequacy of the two-factor model, proved to be inappropriate for this example because of strong evidence of within-subject heteroscedasticity. The power model (MLFA-P) did a fairly good job of describing the within-subject specific variances, although it did not capture the steep decrease in specific variance observed during the early part of the time course. Despite the apparent inadequacy of the fit of the MLFA-P model, results for the rest of the analysis based on the MLFA-P model were almost identical to those based on the MLFA-U model.

For both parameterizations of the O2C model, random effects on the two rate parameters were identified as accounting for most of the between-subject variation in the CCl_4 chamber concentrations. Our method could not distinguish between the two parameterizations of the O2C model. For the PBPK model, we identified the metabolic parameter, V_{max}, and individual rat body weight, B_w, as accounting for most of the between-subject variation.

Our analysis found little evidence of AR(1) structure in the within-subject residuals of the CCl_4 data. With 36 time points in this example, we are confident in the conclusion of no significant AR(1) correlation. The first-stage of the NLMM can safely assume that the residuals are independent.

For the theophylline example, we compared results using factor models with three, four and five common factors. In retrospect, a two-factor model may have been adequate, but the likelihood ratio tests, as well as the smaller percentage of variation accounted for by a two-factor model, suggested that at least three factors should be included. There was clear evidence of within-subject heteroscedasticity for this example, but the power-of-the-mean (MLFA-P) model did not adequately capture the much larger variation observed when serum concentration of theophylline was high.

In the volume-clearance parameterization, f, of the O1C model, the multivariate forward

selection picked the absorption parameter, k_a, and the clearance parameter, Cl, as requiring random effects. For the alternate parameterization, f^*, the procedure picked the two rate parameters as explaining most of the between-subject variation. Our method gave a slight advantage to the parameterization f^*, that had a separate scale coefficient, over the volume-clearance parameterization, where the random effects on k_a and Cl induced random variation in the scale. This implied that there was little evidence of subject-to-subject variation in scale in the theophylline data set.

Our analysis found little evidence of AR(1) structure in the within-subject residuals for the theophylline data. However with only 10 time points, our method had little power to detect AR(1) structure.

The likelihood ratio tests for identifying an adequate number of common factors did not prove useful in either example, rejecting the adequacy of factor models that accounted for close to 100% of total sample variation. The results of these tests should be interpreted in conjunction with the percentage of variation explained by the k-factor model.

Modeling within-subject heteroscedasticity was a difficult task. The ability to compare estimates of the specific variances from the MLFA-U model to those obtained by assuming a particular model (e.g., the power-of-the-mean model) is a nice feature of our approach. For example, a graphical comparison of estimates of the specific variances ψ_j obtained by fitting the unrestricted MLFA-U model and an alternative model may be useful in understanding the structure of within-subject variation. We suspect that the likelihood ratio tests may tend to reject the simpler specific variance models in cases where they provide a useful approximation in practice (as was the case in the CCl_4 example).

In both examples, our method for assessing AR(1) structure of within-subject variation gave a clear answer that there was none. However, this method was developed under the assumption of homoscedastic within-subject variance, and there was strong evidence that within-subject variation was heteroscedastic in both examples. The simulation results of Section 4.3.2 (Chapter 4) showed that the size of the test was inflated when within-subject variance increased with

mean response and p was relatively large (e.g., 25 or larger), but that power to detect even strong AR(1) correlation was low when p was as small as 10. The simulation results also showed that our method was not able to separate the effect of large misspecification errors from AR(1) structure. For the theophylline example where $p = 10$, there is a risk that our test did not detect AR(1) correlation structure in the residuals due to small sample size. However, for the CCl_4 example where $p = 36$, we believe that the fact that the test did not detect significant AR(1) correlation is a reliable assurance that there was none; misspecification or heteroscedasticity would only have increased the chance of falsely detecting AR(1) structure.

Chapter 6

Summary and future work

In this dissertation we developed methods for exploring between- and within-subject variation in balanced longitudinal data. The goal of the exploration was to guide the modeling of covariance structure in a nonlinear mixed effects model (NLMM). In particular, we developed methods for selecting random effects, choosing an appropriate parameterization of the nonlinear regression function, identifying within-subject heteroscedasticity, and determining if there is strong evidence of AR(1) structure in within-subject residuals.

In simulation experiments, our methods were able to select the correct random effects (linear or nonlinear), even when the random effect variance was relatively small. We could identify the simpler parameterization of a regression function, when one random effect in the true parameterization induced two random effects in an alternative parameterization. Our methods were able to identify and model the presence of within-subject heteroscedasticity. We could detect large AR(1) correlation in within-subject residuals, and were able to distinguish between correlation structure due to model bias caused by a small misspecification error in a coefficient value, and true AR(1) structure.

Our methods were based on a first-order linear approximation of the first stage of the NLMM. When random effects or parameter misspecification was large, simulation experiments showed that the first-order approximation was inadequate. For a large random effect, the multivariate methods picked up additional common factors due to second-order terms involving the large random effects. Including second-order sensitivities in the multivariate regression models helped to explain these second-order effects.

We applied our methods to two real examples. For the CCl_4 gas uptake experiment, we

concluded that two random effects characterized between-subject variation. The two rate parameters of the open-two compartment model explained most of the between-subject variation. For the PBPK model, between-subject variation was explained by the metabolic parameter, V_{max}, and by individual rat body weight, B_w. There was strong evidence that within-subject variation was heteroscedastic, but no evidence of AR(1) structure in within-subject residuals.

For the theophylline example, between-subject variation was explained by random effects on the two rate parameters in the O1C model. In the volume-clearance parameterization, we identified random effects on the absorption rate, k_a, and the clearance parameter, Cl. There was strong evidence of within-subject heteroscedasticity, but little evidence of AR(1) correlation (although in this example with only 10 time points, the test for AR(1) structure had little power).

There are several outstanding issues which we did not address in this dissertation. These provide motivation for possible future extensions of our methods.

- Our methods for exploring structure of between- and within-subject variation rely on multivariate data analysis methods that require balanced data. However in many longitudinal applications, data are sparsely sampled at irregular times, or are incomplete due to missing observations.
- Subject-level covariates may explain much between-subject variation. For example, in the CCl_4 gas uptake experiment, our methods identified that body weight (a measured covariate and one of the input parameters in the PBPK model) explained a large portion of between-subject variation. In developing our methods we assumed that covariates were constant across subjects, or that they could be removed from consideration, for example, by scaling.
- We evaluated the sensitivity functions at the values of the parameters that were in hand, for example, from a preliminary analysis of the data. We did not evaluate how robust our methods are to small changes in values of the parameters at which sensitivities are evaluated.

- For the likelihood ratio tests for heteroscedasticity, we used the asymptotic chi-squared approximation of the null distributions. Finite sample corrections would make these tests more reliable (i.e., preserve the nominal level) for the moderate sample sizes expected in NLMM applications.
- In the NLMM with heteroscedastic within-subject variance, it is common to assume that variance is a function of the expected mean response for that subject, e.g.,

  $$\mathbb{V}(\varepsilon_{ij}) = \sigma^2\{f(t_j, \beta + b_i)\}^{2\gamma}.$$

  Thus the variance for the i-th subject depends on the random effect b_i. The factor analysis model assumed that within-subject variance, although possibly different at each time point, was the same for all individuals at that time point,

  $$\mathbb{V}(\varepsilon_{ij}) = \psi_j, \quad \text{for all } i.$$

  In our simulation experiments we simulated heteroscedasticity according to

  $$\mathbb{V}(\varepsilon_{ij}) = \sigma^2\{f(t_j, \beta)\}^{2\gamma}$$

  and did not investigate how a random effect in the variance model would change the results of the factor analysis.
- In examining the effect of model bias on ability to detect AR(1) structure in residuals, we only considered model bias arising from misspecification of parameters in the model. We did not explore other types of bias that might arise due to structural misspecification of the regression function.
- When misspecification error was small, our method could distinguish between AR(1) correlation and model bias. However, when misspecification error was large, evaluating the sensitivity at the misspecified parameter value inflated the size of the test

for H_0: $\rho = 0$.

- In most cases, our estimator for AR(1) correlation was too small on average. Our approach to estimating ρ and testing the hypothesis H_0: $\rho = 0$ could perhaps be improved, for example by deriving the maximum likelihood estimator for ρ, and developing a likelihood ratio test for the hypothesis of interest.
- Heteroscedasticity is commonly encountered in nonlinear regression applications, however, our test for H_0: $\rho = 0$ had inflated size in the presence of heteroscedasticity. Our approach for testing for within-subject correlation could perhaps be extended to a first-order heterogeneous autoregressive ARH(1) model.
- Our methods assume that all random variables are normally distributed. We did not investigate the robustness of our approach to departures from this assumption.

Bibliography

T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley Interscience, 3rd edition, 2003.

S. L. Beal and L. B. Sheiner. Estimating population kinetics. *CRC Critical Reviews in Biomedical Engineering*, 8:195–222, 1982.

A. Bhargava, L. Franzini, and W. Narendranathan. Serial correlation and the fixed effects model. *The Review of Economic Studies*, 49:533–549, 1982.

F. Y. Bois, L. Zeise, and T. N. Tozer. Precision and sensitivity of pharmacokinetic models for cancer risk assessment: Tetrachloroethylene in mice, rats and humans. *Toxicology and Applied Pharmacology*, 102:300–315, 1990.

B. Cai and D. B. Dunson. Bayesian covariance selection in generalized linear mixed models. *Biometrics*, 62:446–457, 2006.

H. Cardot. Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Journal of Nonparametric Statistics*, 12(4):503–538, 2000.

R. J. Carroll. Variances are not always nuisance parameters. *Biometrics*, 59:211–220, 2003.

R. J. Carroll and D. Ruppert. *Transformation and Weighting in Regression*. Chapman and Hall, New York, 1988.

Z. Chen and D. B. Dunson. Random effects selection in linear mixed models. *Biometrics*, 59: 762–769, 2003.

K. Chiswell and J. F. Monahan. Dealing with model bias in a nonlinear mixed model. In *2004 Proceedings of the American Statistical Association, Statistical Computing Section [CD-ROM], Alexandria, VA: American Statistical Association*, pages 2419–2426, 2004.

K. Chiswell and J. F. Monahan. Random effects selection in a nonlinear mixed model with many parameters. In *2005 Proceedings of the American Statistical Association, Section on Nonparametric Statistics [CD-ROM], Alexandria, VA: American Statistical Association*, pages 1628–1635, 2005.

R. D. Cook and S. Weisberg. Diagnostics for heteroscedasticity in regression. *Biometrika*, 70: 1–10, 1983.

C. M. Crainiceanu and D. Ruppert. Likelihood ratio tests for goodness-of-fit of a nonlinear regression model. *Journal of Multivariate Analysis*, 91:35–52, 2004a.

C. M. Crainiceanu and D. Ruppert. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B*, 66(1):165–185, 2004b.

P. Dalgaard and M. Larsen. Fitting numerical solutions of differential equations to experimental data: A case study and some general remarks. *Biometrics*, 46:1097–1109, 1990.

J. Dauxois, A. Pousse, and Y. Romain. Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12:136–154, 1982.

M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall / CRC, Boca Raton, FL, 1995.

M. Davidian and D. M. Giltinan. Nonlinear models for repeated measurement data: an overview and update. *Journal of Agricultural, Biological and Environmental Statistics*, 8:387–419, 2003.

D. K. Dey, M. H. Chen, and H. Chang. Bayesian approach for nonlinear random effects models. *Biometrics*, 53:1239–1252, 1997.

J. Diebolt and J. Zuber. Goodness-of-fit tests for nonlinear heteroscedastic regression models. *Statistics & Probability Letters*, 42:53–60, 1999.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression, i. *Biometrika*, 37:409–428, 1950.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression, ii. *Biometrika*, 38:159–178, 1951.

J. Durbin and G. S. Watson. Testing for serial correlation in least squares regression, iii. *Biometrika*, 58:1–19, 1971.

M. V. Evans, W. D. Crank, H. Yang, and J. E. Simmons. Applications of sensitivity analysis to a physiologically based pharmacokinetic model for carbon tetrachloride in rats. *Toxicology and Applied Pharmacology*, 128:36–44, 1994.

A. Gelman, F. Bois, and J. Jiang. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association*, 91:1400–1412, 1996.

J. F. Geweke and K. J. Singleton. Interpreting the likelihood ratio statistic in factor models when sample size is small. *Journal of the American Statistical Association*, 75(369):133–137, 1980.

P. J. Green and B. W. Silverman. *Nonparametric Regression and Generalized Linear Models: a Roughness Penalty Approach*. Chapman & Hall, London, 1994.

P. Hall, H.-G. Müller, and J.-L. Wang. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*, 34(3):1493–1517, 2006.

D. A. Harville. *Matrix Algebra from a Statistician's Perspective*. Springer, New York, 1997.

R. I. Jennrich and M. D. Schluchter. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics*, 42:805–820, 1986.

R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, New Jersey, 5th edition, 2002.

R. A. Johnson and D. W. Wichern. *Applied Multivariate Statistical Analysis*. Prentice Hall, Upper Saddle River, New Jersey, 6th edition, 2007.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *The Annals of Statistics*, 29(2):295–327, 2001.

K. G. Jöreskog. Some contributions to maximum likelihood factor analysis. *Psychometrika*, 32 (4):443–482, 1967.

M. O. Karlsson, S. L. Beal, and L. B. Sheiner. Three new residual error models for populationpk/pd analyses. *Journal of Pharmacokinetics and Biopharmaceutics*, 23:651–672, 1995.

S. K. Kinney and D. B. Dunson. Fixed and random effects selection in linear and logistic models. *Biometrics (OnlineEarly Articles)*, 2007.

D. N. Lawley and A. E. Maxwell. *Factor Analysis as a Statistical Method*. London Butterworths, London, second edition, 1971.

J.-G. Lin and B.-C. Wei. Testing for heteroscedasticity in nonlinear regression models. *Communications in Statistics: Theory and Methods*, 32(1):171–192, 2003.

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

J. E. Marsden and M. J. Hoffman. *Elementary Classical Analysis*. W.F. Freeman and Company, New York, 2nd edition, 1993.

G. Molenberghs and G. Verbeke. Meaningful statistical model formulations for repeated measures. *Statistica Sinica*, 14:989–1020, 2004.

J. E. Morgan. *Methods for testing variance parameters equal to zero in a random coefficient regression model*. PhD thesis, North Carolina State University, 1996.

A. Ostermann. Sensitivity analysis. Privately published: http://techmath.uibk.ac.at/mathematik/publikationen, 2004.

Z. Pan and D. Y. Lin. Goodness-of-fit methods for generalized linear mixed models. *Biometrics*, 61:1000–1009, 2005.

S. Pezzulli and B. W. Silverman. Some properties of smoothed principal components analysis for functional data. *Computational Statistics*, 8:1–16, 1993.

G. L. Plaa and M. Charbonneau. *Principles and Methods of Toxicology*, chapter 24. Taylor & Francsis, New York, fourth edition, 2000.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, New York, 2nd edition, 2005.

J. A. Rice. Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica*, 14:631–647, 2004.

J. A. Rice and B. W. Silverman. Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B: Methodological*, 53:233–243, 1991.

W. A. Ritschel and G. L. Kearns. *Handbook of Basic Pharmacokinetics—including Clinical Applications*. The American Pharmaceutical Association, Washington, DC, 1998.

B. E. Rodda, C. B. Sampson, and D. W. Smith. The one-compartment open model: some statistical aspects of parameter estimation. *Applied Statistics*, 24:309–318, 1975.

H. L. Royden. *Real Analysis*. Macmillan Publishing Company, New York, 3rd edition, 1988.

C. Savalli, G. A. Paula, and F. J. A. Cysneiros. Assessment of variance components in elliptical linear mixed models. *Statistical Modelling*, 6:59–76, 2006.

S. G. Self and K. Liang. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, 82(398):605–610, 1987.

R. G. D. Steel and J. H. Torrie. *Principles and Procedures of Statistics: a Biometrical Approach*. McGraw-Hill, Singapore, 2nd edition, 1980.

G. W. Stewart. The efficient generation of random orthogonal matrices with an application to condition estimators. *SIAM Journal on Numerical Analysis*, 17:403–409, 1980.

D. O. Stram and J. W. Lee. Variance components testing in the longitudinal mixed effects model. *Biometrics*, 50:1171–1177, 1994.

G. Verbeke and G. Molenberghs. The use of score tests for inference on variance components. *Biometrics*, 59:254–262, 2003.

E. F. Vonesh, V. M. Chinchilli, and K. Pu. Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, 52(2):572–587, 1996.

K. J. White. The Durbin-Watson test for autocorrelation in nonlinear models. *The Review of Economics and Statistics*, 74(2):370–373, 1992.

R. Wolfinger. Laplace's approximation for nonlinear mixed models. *Biometrika*, 80(4):791–795, 1993.

Q. Zeng and M. Davidian. Testing homgeneity of intra-run variance parameters in immunoassay. *Statistics in Medicine*, 16:1756–1776, 1997.

Appendices

Appendix A

Functional factor analysis when $k = 1$

In this appendix we describe an approach to smoothed PCA that we referred to in Section 2.2.4. We call this method *functional factor analysis*, and use *penalized maximum likelihood* to fit the functional factor model. We have only developed the approach for one common factor. For more factors, the functional PCA approach of Ramsay and Silverman (2005, Chapters 8 & 9) is probably easier to implement than an extension of our method would be; for example, our method would require a different smoothing parameter for each functional factor.

Our approach is similar to the method proposed by Rice and Silverman (1991), an early paper in the area of functional principal component analysis (FPCA). In that paper, the model included k common factors, but did not include a residual error term. The mean function was estimated separately from estimation of the common factor functions. In our method, we allow for a residual term, and estimation of the common factor function depends on the smoothed estimate of the mean function.

A.1 Functional factor model

Suppose that the number of parameters to be considered as candidates for random effects in the nonlinear mixed effects model (NLMM) is only one (i.e., $q = 1$). Then the FOL approximation for subjects with common covariates is of the form

$$y_{ij} \approx \mu(t_j) + w_i\,\eta(t_j) + \varepsilon_{ij}, \quad \text{for} \quad i = 1, \ldots, n; \quad j = 1, \ldots, p \tag{A.1}$$

where μ and η are functions common to all subjects with $\mu(t_j) \approx f(t_j, \beta, \boldsymbol{\xi}^*; \boldsymbol{x}_i)$, $\eta(t_j) \approx f_\beta^{(i)}/\gamma$, $w_i \approx \gamma b_i$ independent and identically distributed (i.i.d.) $N(0, \gamma^2\sigma_i^2)$, and ε_{ij} are i.i.d. $N(0, \sigma^2)$ independently of w_i. The number $\gamma \in \mathbb{R}$ is a scale factor resulting from the need to make the model in (A.1) identifiable. In the case of identical observation times on each subject we can make the model identifiable by constraining $\sum_{j=1}^p \eta^2(t_j) = p$. In this case, $\gamma = \left\{\sum_{j=1}^p \left(f_\beta^{(i)}\right)^2 / p\right\}^{1/2}$. The model in (A.1) is similar to a multivariate factor analysis model with one factor, except that the common factor is assumed to be a smooth function (hence the term *functional factor analysis*), and the specific factors ε_{ij} are all assumed to have common variance σ^2. In Section A.2 we describe a nonparametric method for estimating the functions μ and η.

At first glance it might seem more natural to impose the constraint that $\sum_{j=1}^p \eta^2(t_j) = 1$, however we prefer the constraint $\sum_{j=1}^p \eta^2(t_j) = p$ for the following reason. Suppose that the domain of η is $[0, 1]$. Then the L_2-norm of η is $\|\eta\| = \{\int_0^1 \eta^2(t)\mathrm{dt}\}^{1/2}$. If the observation times $t_1, t_2, \ldots, t_p$ are distributed evenly over the interval $[0, 1]$ with $t_1 = 1/p, t_2 = 2/p, \ldots, t_p = 1$, then

$$\|\eta\| = \left\{\int_0^1 \eta^2(t)\mathrm{dt}\right\}^{1/2} \approx \left\{\frac{1}{p}\sum_{j=1}^p \eta^2(t_j)\right\}^{1/2}.$$

In the functional factor model, it makes sense to think of trying to estimate a function η that has a norm that does not depend on the number of observations, p. The constraint $\|\eta\| = 1$ in the functional space implies

$$\sum_{j=1}^p \eta^2(t_j) \approx p.$$

A.2 Penalized maximum likelihood estimation of μ and η

Because of the clustered (longitudinal) structure of the data, it is convenient to use vector notation for the observed data on each subject. Let $\boldsymbol{y}_i = (y_{i1}, y_{i2}, \ldots, y_{ip})^T$, $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \varepsilon_{i2}, \ldots, \varepsilon_{ip})^T$, $\boldsymbol{\mu} = (\mu(t_1), \mu(t_2), \ldots, \mu(t_p))^T$, and $\boldsymbol{\eta} = (\eta(t_1), \eta(t_2), \ldots, \eta(t_p))^T$. Then model (A.1) can be

written

$$\boldsymbol{y}_i \approx \boldsymbol{\mu} + w_i\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i \quad \text{for} \quad i = 1, \ldots, n. \tag{A.2}$$

To avoid identification problems due to scaling of $\boldsymbol{\eta}$ and w_i we will assume that $\boldsymbol{\eta}$ has norm p, or $\boldsymbol{\eta}^T\boldsymbol{\eta} = p$. Under the assumption that $\boldsymbol{\varepsilon}_i$ are independently distributed $N(\mathbf{0}, \sigma^2\mathbf{I}_p)$ and w_i are i.i.d. $N(0, \sigma_w^2)$, independent of $\boldsymbol{\varepsilon}_i$, the $\boldsymbol{y}_i$ vectors have (to first order in $|w_i|$) a multivariate normal distribution with marginal mean

$$\mathrm{E}(\boldsymbol{y}_i) \approx \mathrm{E}(\boldsymbol{\mu} + w_i\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i) = \boldsymbol{\mu}$$

and marginal covariance matrix

$$\mathrm{Var}(\boldsymbol{y}_i) \approx \mathrm{Var}(\boldsymbol{\mu} + w_i\boldsymbol{\eta} + \boldsymbol{\varepsilon}_i) = \sigma_w^2\boldsymbol{\eta}\boldsymbol{\eta}^T + \sigma^2\mathbf{I}_p \equiv \boldsymbol{\Sigma}.$$

Since

$$\begin{aligned}
\det(\boldsymbol{\Sigma}) &= \det(\sigma_w^2\boldsymbol{\eta}\boldsymbol{\eta}^T + \sigma^2\mathbf{I}_p) \\
&= (\sigma^2)^p \det\left\{(\sigma_w^2/\sigma^2)\boldsymbol{\eta}\boldsymbol{\eta}^T + \mathbf{I}_p\right\} \\
&= (\sigma^2)^p \det\left\{(\sigma_w^2/\sigma^2)\boldsymbol{\eta}^T\boldsymbol{\eta} + 1\right\} \\
&= (\sigma^2)^p \det\left(1 + p\sigma_w^2/\sigma^2\right)
\end{aligned}$$

the probability density (pdf) of $\boldsymbol{y}_i$ is

$$\begin{aligned}
p\left(\boldsymbol{y}_i\right) &= (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\mathrm{e}^{-(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})/2} \\
&= (2\pi)^{-p/2}(\sigma^2)^{-p/2}\left(1 + p\sigma_w^2/\sigma^2\right)^{-1/2}\mathrm{e}^{-(\boldsymbol{y}_i-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{y}_i-\boldsymbol{\mu})/2}
\end{aligned}$$

where

$$\boldsymbol{\Sigma}^{-1} = \left(\sigma_w^2\boldsymbol{\eta}\boldsymbol{\eta}^T + \sigma^2\mathbf{I}_p\right)^{-1} = [\mathbf{I}_p - \{\sigma_w^2/(\sigma^2 + p\sigma_w^2)\}\boldsymbol{\eta}\boldsymbol{\eta}^T]/\sigma^2.$$

Appendix A. Functional factor analysis when $k = 1$

By independence of subjects, the marginal likelihood function is the product of $p(\boldsymbol{y}_1)$, ..., $p(\boldsymbol{y}_n)$.

The -2-loglikelihood function can be written as

$$\begin{aligned} -2\log L = np\log(2\pi) + np\log(\sigma^2) + n\log(1 + p\sigma_w^2/\sigma^2) \\ + n(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}) + \text{trace}\left(\boldsymbol{\Sigma}^{-1}\mathbf{E}\right) \end{aligned} \tag{A.3}$$

where $\bar{\boldsymbol{y}} = \sum_{i=1}^n \boldsymbol{y}_i/n$ and $\mathbf{E} = \sum_{i=1}^n (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T$.

If we assume that μ and η are smooth functions of a continuous variable $t \in [a, b]$; i.e., they do not fluctuate rapidly, then we can define μ and η belonging to the model space $\mathcal{S}[a, b]$, the set of all twice-differentiable functions on $[a, b]$. A natural way to quantify the roughness of a function in $\mathcal{S}$ is to compute its integrated squared second derivative, e.g., $\int_a^b [\mu''(s)]^2 \mathrm{d}s$. Under the constraint of smoothness, we estimate μ and η by minimizing a penalized version of the objective function in (A.3), namely

$$S(\mu, \eta, \sigma^2, \sigma_w^2) = -2\log L + \alpha_\mu \int [\mu'']^2 + \alpha_\eta \int [\eta'']^2$$

for $\mu, \eta \in \mathcal{S}[a, b]$ and $\sigma^2, \sigma_w^2 > 0$. The parameters α_μ and α_η are so-called smoothing parameters. They control the trade-off between goodness-of-fit and smoothness.

Using standard results from smoothing spline literature (e.g., Green and Silverman, 1994), the above objective function can be written as

$$\begin{aligned} S^* = np\log(2\pi) + np\log(\sigma^2) + n\log(1 + p\sigma_w^2/\sigma^2) + n(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}) \\ + \text{trace}\left(\boldsymbol{\Sigma}^{-1}\mathbf{E}\right) + \alpha_\mu \boldsymbol{\mu}^T\mathbf{K}\boldsymbol{\mu} + \alpha_\eta \boldsymbol{\eta}^T\mathbf{K}\boldsymbol{\eta} \end{aligned} \tag{A.4}$$

where $\mathbf{K}$ is a $p \times p$ matrix depending on the time points $t_1, \ldots, t_p$ and the form of the penalty function. For fixed σ^2, σ_w^2, α_μ and α_η, let $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_p)^T$ and $\widehat{\boldsymbol{\eta}} = (\widehat{\eta}_1, \ldots, \widehat{\eta}_p)^T$ be the minimizers of (A.4) over $\mathbb{R}^p$ subject to the constraint that $\boldsymbol{\eta}^T\boldsymbol{\eta} = p$. The minimizers $\widehat{\mu}$ and $\widehat{\eta}$

of $S(\mu, \eta, \sigma^2, \sigma_w^2)$ over $\mathcal{S}[a, b]$ are the unique natural cubic spline interpolators of the elements of $\widehat{\boldsymbol{\mu}}$ and $\widehat{\boldsymbol{\eta}}$ with knots at $t_1, \ldots, t_p$ (see Green and Silverman, 1994, Chapter 2). Therefore the penalized likelihood estimators of the functions μ and η can be computed by taking derivatives of S^* with respect to the vectors $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ and setting the derivatives equal to zero. Our strategy is to treat $\boldsymbol{\eta}$ as fixed and estimate $\boldsymbol{\mu}$, then treat $\boldsymbol{\mu}$ as fixed and estimate $\boldsymbol{\eta}$. We derive results in Section A.5.

The resulting estimator for $\boldsymbol{\mu}$ solves

$$\left\{\mathbf{I}_p + (\alpha_\mu/n)\widehat{\boldsymbol{\Sigma}}\mathbf{K}\right\}\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}} \tag{A.5}$$

where $\widehat{\boldsymbol{\Sigma}} = \sigma^2\mathbf{I}_p + \sigma_w^2\widehat{\boldsymbol{\eta}}\widehat{\boldsymbol{\eta}}^T$. So $\widehat{\boldsymbol{\mu}}$ is a weighted version of $\bar{\boldsymbol{y}}$, the subject means at each time point. The estimator for $\boldsymbol{\eta}$ maximizes $\boldsymbol{\eta}^T\mathbf{A}\boldsymbol{\eta}$ where

$$\mathbf{A} = n\sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)](\bar{\boldsymbol{y}} - \widehat{\boldsymbol{\mu}})(\bar{\boldsymbol{y}} - \widehat{\boldsymbol{\mu}})^T + \sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\mathbf{E} - \alpha_\eta\mathbf{K} \tag{A.6}$$

subject to the constraint that $\widehat{\boldsymbol{\eta}}^T\widehat{\boldsymbol{\eta}} = p$. Thus, $\widehat{\boldsymbol{\eta}}$ is the eigenvector of $\mathbf{A}$ associated with its largest eigenvalue, that is scaled to have norm $p^{1/2}$. The estimation of $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ are coupled, and the two problems should be solved simultaneously (e.g., to ensure convergence to the true maximum likelihood estimator). However, in our preliminary attack of this problem, we use an easier approach which is to solve the two problems separately. We propose an iterative algorithm:

(a) Initial iterates: let $\boldsymbol{\mu}_c = \boldsymbol{\mu}_0$ and $\boldsymbol{\eta}_c = \boldsymbol{\eta}_0$ for some starting values $\boldsymbol{\mu}_0$ and $\boldsymbol{\eta}_0$.

(b) Solve (A.5) for $\widehat{\boldsymbol{\mu}}$ with $\boldsymbol{\eta}$ set at $\boldsymbol{\eta}_c$.

(c) Evaluate $\mathbf{A}$ at $\boldsymbol{\mu}_c$. Find the largest eigenvalue of $\mathbf{A}$ and its corresponding eigenvector $\widehat{\boldsymbol{\eta}}$, scaled to have norm $p^{1/2}$.

(d) Set $\boldsymbol{\mu}_c = \widehat{\boldsymbol{\mu}}$ and $\boldsymbol{\eta}_c = \widehat{\boldsymbol{\eta}}$ and return to step (b).

Repeat steps (b) – (d) a fixed number of times.

Results from a simulation experiment showing performance of our functional factor analysis method method, and application of the method to the CCl_4 gas uptake experiment were reported in Chiswell and Monahan (2005). In that work, we found that estimation of μ was better when between-subject variation was smaller than when it was larger. This was to be expected, because the FOL approximation is only valid when the magnitude of the nonlinear random effect is small. On the other hand, the estimation of η improved when the between-subject variation was larger. We believe this is because the larger random effect created a stronger "signal" in the covariance structure so that it was easier to identify the largest eigenvalue of the matrix $\mathbf{A}$ in (A.6).

A.3 Behavior under maximum or no smoothing

We briefly consider what happens to our estimators $\widehat{\mu}$ and $\widehat{\eta}$ when the smoothing parameters α_μ and α_η approach zero or infinity.

When $\alpha_\mu = 0$, equation (A.5) reduces to $\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$ so that $\widehat{\mu}$ is just the natural cubic spline interpolator of the subject mean scores $\bar{\boldsymbol{y}}$, with knots at $t_1 < t_2 < \ldots < t_p$. When $\alpha_\eta = 0$, the matrix $\mathbf{A}$ in (A.6) becomes

$$\mathbf{A} = \sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)] \sum_{i=1}^{n} (\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}})(\boldsymbol{y}_i - \widehat{\boldsymbol{\mu}})^T \tag{A.7}$$

i.e., a multiple of the residual sample covariance matrix after fitting the mean trend. Thus in this case $\widehat{\eta}$ is the natural cubic spline interpolator of the eigenvector corresponding to the largest eigenvalue of the residual covariance matrix.

Our choice of the integrated squared second derivative roughness penalty (e.g., $\int [\mu'']^2$) dictates that making α_μ or α_η large, forces $\widehat{\mu}$ or $\widehat{\eta}$ respectively to be linear. In this case, the estimator $\widehat{\mu}$ will be the best linear regression fit to subject means at each point, i.e., the best

linear fit to the points $(t_j, \bar{y}_j)$ for $j = 1, \ldots, p$. The estimator $\widehat{\eta}$ will be the best linear regression fit to the eigenvector corresponding to the largest eigenvalue of the residual covariance matrix given in (A.7).

A.4 Choice of smoothing parameters

In implementing our method, we chose the smoothing parameters α_μ and α_η subjectively by examining the data. More objective (data-based) methods such as cross-validation would be preferable. For example, Rice and Silverman (1991) proposed "leave-out-one-curve" cross-validation.

The variance components σ^2 and σ_w^2 also need to be estimated from the data. We propose maximum likelihood or residual maximum likelihood (REML) estimation, but have not pursued details of the estimation of the variance components.

A.5 Derivation of the estimators of $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$

In this section we derive equations (A.5) and (A.6) for estimating $\boldsymbol{\mu}$ and $\boldsymbol{\eta}$ by minimizing the objective function S^* defined in formula (A.4). For $\boldsymbol{\mu}$, we fix $\boldsymbol{\eta}$ and set the derivative of S^* with respect to $\boldsymbol{\mu}$ equal to $\mathbf{0}$ and solve for $\boldsymbol{\mu}$. Since $\boldsymbol{\Sigma}$ and $\mathbf{K}$ are symmetric we can write

$$\begin{aligned} \partial S^*/\partial \boldsymbol{\mu} &= -2n\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu}) + 2\alpha_\mu \mathbf{K}\boldsymbol{\mu} \\ &= -2n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{y}} + 2(n\boldsymbol{\Sigma}^{-1} + \alpha_\mu \mathbf{K})\boldsymbol{\mu}. \end{aligned}$$

Setting the vector of partial derivatives equal to $\mathbf{0}$ gives the following system of linear equations in $\widehat{\boldsymbol{\mu}}$:

$$\left(n\boldsymbol{\Sigma}^{-1} + \alpha_\mu \mathbf{K}\right)\widehat{\boldsymbol{\mu}} = n\boldsymbol{\Sigma}^{-1}\bar{\boldsymbol{y}} \quad \Leftrightarrow \quad \left\{\mathbf{I}_p + (\alpha_\mu/n)\boldsymbol{\Sigma}\mathbf{K}\right\}\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$$

where $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p + \sigma_w^2 \boldsymbol{\eta}\boldsymbol{\eta}^T$. Notice that the solution is a function of $\boldsymbol{\eta}$.

Appendix A. Functional factor analysis when $k = 1$

To find $\widehat{\boldsymbol{\eta}}$ we first show that for fixed $\boldsymbol{\mu}$, σ_w^2 and σ^2 (and smoothing parameters), the objective function S^* in (A.4) can be written as a quadratic form in $\boldsymbol{\eta}$, i.e., $S^* = C^* - \boldsymbol{\eta}^T \mathbf{A} \boldsymbol{\eta}$ where $\mathbf{A}$ is a $p \times p$ symmetric matrix free of $\boldsymbol{\eta}$, and C^* is a constant. We then use the result (Rayleigh-Ritz Theorem) that the vector $\widehat{\boldsymbol{\eta}}$ that maximizes $\boldsymbol{\eta}^T \mathbf{A} \boldsymbol{\eta}$ subject to the constraint that $\boldsymbol{\eta}^T \boldsymbol{\eta} = p$ is the eigenvector of $\mathbf{A}$ associated with its largest eigenvalue, that is scaled to have norm $p^{1/2}$. This eigenvector $\widehat{\boldsymbol{\eta}}$ therefore minimizes S^* subject to the constraint that $\boldsymbol{\eta}^T \boldsymbol{\eta} = p$.

For fixed $\boldsymbol{\mu}$, σ_w^2 and σ^2, the objective function S^* can be written as the following function of $\boldsymbol{\eta}$:

$$S^* = C + n(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu}) + \text{trace}\left(\boldsymbol{\Sigma}^{-1} \mathbf{E}\right) + \alpha_\eta \boldsymbol{\eta}^T \mathbf{K} \boldsymbol{\eta}$$

where C is free of $\boldsymbol{\eta}$. Substituting $\boldsymbol{\Sigma}^{-1} = [\mathbf{I}_p - \{\sigma_w^2/(\sigma^2 + p\sigma_w^2)\}\boldsymbol{\eta}\boldsymbol{\eta}^T]/\sigma^2$ in the second term gives

$$\begin{aligned} n(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu}) &= n(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T [\mathbf{I}_p - \sigma_w^2/(\sigma^2 + p\sigma_w^2)\boldsymbol{\eta}\boldsymbol{\eta}^T](\bar{\boldsymbol{y}} - \boldsymbol{\mu})/\sigma^2 \\ &= C_1 - n\sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)](\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\eta}\boldsymbol{\eta}^T (\bar{\boldsymbol{y}} - \boldsymbol{\mu}) \\ &= C_1 - n\sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\boldsymbol{\eta}^T (\bar{\boldsymbol{y}} - \boldsymbol{\mu})(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\eta} \end{aligned}$$

where C_1 is another constant free of $\boldsymbol{\eta}$. Substituting $\boldsymbol{\Sigma}^{-1} = [\mathbf{I}_p - \sigma_w^2/(\sigma^2 + p\sigma_w^2)\boldsymbol{\eta}\boldsymbol{\eta}^T]/\sigma^2$ in the third term of S^* gives

$$\begin{aligned} \text{trace}\left(\boldsymbol{\Sigma}^{-1} \mathbf{E}\right) &= \text{trace}\left\{[\mathbf{I}_p - \sigma_w^2/(\sigma^2 + p\sigma_w^2)\boldsymbol{\eta}\boldsymbol{\eta}^T]\mathbf{E}/\sigma^2\right\} \\ &= C_2 - \sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\ \text{trace}(\boldsymbol{\eta}\boldsymbol{\eta}^T \mathbf{E}) \\ &= C_2 - \sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\ \text{trace}(\boldsymbol{\eta}^T \mathbf{E}\boldsymbol{\eta}) \\ &= C_2 - \sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\boldsymbol{\eta}^T \mathbf{E}\boldsymbol{\eta} \end{aligned}$$

where C_2 is another constant free of $\boldsymbol{\eta}$. Substituting back into S^* gives

$$\begin{aligned} S^* &= C^* - n\sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\boldsymbol{\eta}^T(\bar{\boldsymbol{y}} - \boldsymbol{\mu})(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T\boldsymbol{\eta} \\ &\quad -\sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\boldsymbol{\eta}^T\mathbf{E}\boldsymbol{\eta} + \alpha_\eta\boldsymbol{\eta}^T\mathbf{K}\boldsymbol{\eta} \\ &= C^* - \boldsymbol{\eta}^T\mathbf{A}\boldsymbol{\eta} \end{aligned}$$

where

$$\mathbf{A} = n\sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)](\bar{\boldsymbol{y}} - \boldsymbol{\mu})(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T + \sigma_w^2/[\sigma^2(\sigma^2 + p\sigma_w^2)]\mathbf{E} - \alpha_\eta\mathbf{K}$$

and the terms free of $\boldsymbol{\eta}$ have been accumulated in C^*. Symmetry of $\mathbf{A}$ follows from symmetry of $(\bar{\boldsymbol{y}} - \boldsymbol{\mu})(\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T$, $\mathbf{E}$, and $\mathbf{K}$.

Appendix B

Proofs of theorems for principal component analysis approach

B.1 Proof of Theorem 2.3.1

Limiting matrix of $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$

Theorem. *Let* $\mathbf{Y}$ *be a* $n \times p$ *data matrix, and suppose that the* i*-th row,* $\boldsymbol{y}_i^T$ *can be modeled as*

$$\boldsymbol{y}_i = \boldsymbol{\mu} + \boldsymbol{\lambda} u_i + \boldsymbol{\varepsilon}_i$$

where $\boldsymbol{\lambda}$ *is a* $p \times 1$ *vector,* u_i *are independent and identically distributed (i.i.d.) random variables with mean 0 and variance 1,* $\boldsymbol{\varepsilon}_i$ *are i.i.d. random vectors with mean* $\mathbf{0}$ *and covariance matrix* $\sigma^2\mathbf{I}_p$*, and* u_i *and* $\boldsymbol{\varepsilon}_i$ *are independent. Then as* $p \to \infty$*, the* $n \times n$ *matrix* $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ *converges in probability to* $\boldsymbol{\Phi}^* + \sigma^2\mathbf{I}_n$*, where* $\boldsymbol{\Phi}^*$ *is a positive definite* $n \times n$ *matrix with rank* 2*.*

Proof. Write the $p \times n$ matrix $\mathbf{Y}^T$ as

$$\mathbf{Y}^T = [\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n] = \boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T + \mathbf{E}$$

where $\boldsymbol{u} = (u_1, u_2, \ldots, u_n)^T$ and $\mathbf{E} = [\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_n]$. Consider the $n \times n$ matrix

$$\begin{aligned}
\tfrac{1}{p}\mathbf{Y}\mathbf{Y}^T &= \tfrac{1}{p}\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T + \mathbf{E}\right)^T\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T + \mathbf{E}\right) \\
&= \tfrac{1}{p}\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T\right)^T\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T\right) + \tfrac{1}{p}\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T\right)^T\mathbf{E} \\
&\quad + \tfrac{1}{p}\mathbf{E}^T\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T\right) + \tfrac{1}{p}\mathbf{E}^T\mathbf{E}.
\end{aligned}$$

Appendix B. Proofs of theorems for principal component analysis approach

Let the first term be $\mathbf{\Phi}_p^* = \frac{1}{p}\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T\right)^T\left(\boldsymbol{\mu}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T\right)$. If $\boldsymbol{\lambda}$ is not a scalar multiple of $\boldsymbol{\mu}$, and $\boldsymbol{u} \neq \mathbf{0}$, then the matrix $\boldsymbol{u}\mathbf{1}^T + \boldsymbol{\lambda}\boldsymbol{u}^T$ has rank two, and so $\mathbf{\Phi}_p^*$ has rank two and is positive definite. In fact,

$$\begin{aligned}
\mathbf{\Phi}_p^* &= \tfrac{1}{p}\left\{(\boldsymbol{\mu}^T\boldsymbol{\mu})\mathbf{1}\mathbf{1}^T + (\boldsymbol{\lambda}^T\boldsymbol{\mu})\left(\boldsymbol{u}\mathbf{1}^T + \mathbf{1}\boldsymbol{u}^T\right) + (\boldsymbol{\lambda}^T\boldsymbol{\lambda})\boldsymbol{u}\boldsymbol{u}^T\right\} \\
&= \tfrac{1}{p}\left\{\|\boldsymbol{\mu}\|^2\mathbf{1}\mathbf{1}^T + (\boldsymbol{\lambda}^T\boldsymbol{\mu})\left(\boldsymbol{u}\mathbf{1}^T + \mathbf{1}\boldsymbol{u}^T\right) + \|\boldsymbol{\lambda}\|^2\boldsymbol{u}\boldsymbol{u}^T\right\} \\
&= \tfrac{1}{p}\begin{pmatrix} a\mathbf{1} + b\boldsymbol{u}, & c\boldsymbol{u} \end{pmatrix}\begin{pmatrix} a\mathbf{1}^T + b\boldsymbol{u}^T \\ c\boldsymbol{u}^T \end{pmatrix}
\end{aligned}$$

where $a = \|\boldsymbol{\mu}\|$, $b = \boldsymbol{\mu}^T\boldsymbol{\lambda}/\|\boldsymbol{\mu}\|$, and $c = \left\{\|\boldsymbol{\lambda}\|^2\|\boldsymbol{\mu}\|^2 - (\boldsymbol{\mu}^T\boldsymbol{\lambda})^2\right\}^{1/2}/\|\boldsymbol{\mu}\|$. The constant c is real by the Cauchy-Schwartz inequality: $\|\boldsymbol{\lambda}\|^2\|\boldsymbol{\mu}\|^2 \geq (\boldsymbol{\mu}^T\boldsymbol{\lambda})^2$.

Now we investigate the limiting matrix of $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ as $p \to \infty$. We will assume that the elements of $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are realizations of smooth functions μ and λ defined on $[0,1]$ (i.e., second derivatives of these functions are integrable). Suppose $\boldsymbol{\mu}_j = \mu(t_j)$ and $\boldsymbol{\lambda}_j = \lambda(t_j)$ for $t_j \in [0,1]$, and that as p increases, the density of observation times in $[0,1]$ increases. Then as $p \to \infty$,

$$\begin{aligned}
\tfrac{1}{p}\|\boldsymbol{\mu}\|^2 &= \tfrac{1}{p}\sum_{j=1}^{n}\mu^2(t_j) \to \int_0^1 \mu^2(t)\mathrm{d}t = \|\mu\|^2 \\
\tfrac{1}{p}\|\boldsymbol{\lambda}\|^2 &= \tfrac{1}{p}\sum_{j=1}^{n}\lambda^2(t_j) \to \int_0^1 \lambda^2(t)\mathrm{d}t = \|\lambda\|^2, \quad \text{and} \\
\tfrac{1}{p}\boldsymbol{\lambda}^T\boldsymbol{\mu} &= \tfrac{1}{p}\sum_{j=1}^{n}\lambda(t_j)\mu(t_j) \to \int_0^1 \lambda(t)\mu(t)\mathrm{d}t = \langle\mu,\lambda\rangle
\end{aligned}$$

Thus, conditional on $\boldsymbol{u}$,

$$\mathbf{\Phi}_p^* \to \|\mu\|^2\mathbf{1}\mathbf{1}^T + \langle\mu,\lambda\rangle\left(\boldsymbol{u}\mathbf{1}^T + \mathbf{1}\boldsymbol{u}^T\right) + \|\lambda\|^2\boldsymbol{u}\boldsymbol{u}^T = \mathbf{\Phi}^*$$

which is also a rank 2 positive definite matrix (by a similar argument to the one used above for

$\boldsymbol{\Phi}_p^*$).

The elements of the matrix $\mathbf{E}$ are all i.i.d. random variables with mean 0 and variance σ^2, so by the weak law of large numbers, $\frac{1}{p}\mathbf{E}^T\mathbf{E}$ converges in probability to $\sigma^2\mathbf{I}_n$ as $p \to \infty$. The remaining terms in $\frac{1}{p}\mathbf{Y}\mathbf{Y}^T$ all involve cross-products of the form $\boldsymbol{\mu}^T\mathbf{E}$ or $\boldsymbol{\lambda}^T\mathbf{E}$. Since these vectors are sums of p i.i.d. random variables with mean 0 and variance σ^2, by the weak law of large numbers these terms converge in probability to zero as $p \to \infty$. □

B.2 Proof of Theorem 2.3.2

Distribution of $\boldsymbol{g}^{(1)}$ when $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_p$

Theorem. *Let* $\mathbf{Y}$ *be a* $n \times p$ *data matrix with each row independent and identically distributed according to the* $N_p(\boldsymbol{\mu}, \sigma^2\mathbf{I}_p)$ *distribution. Let* $\mathbf{S}_n$ *be the sample covariance matrix, and let* $\boldsymbol{g}^{(1)}$ *be the normalized eigenvector of* $\mathbf{S}_n$ *corresponding to its largest eigenvalue. Then the distribution of* $\boldsymbol{g}^{(1)}$ *is the same as the distribution of* $\boldsymbol{z}/\|\boldsymbol{z}\|$ *where* $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I})$.

Proof. Let $\mathcal{O}_p$ be the space of orthogonal $p \times p$ matrices. A measure ν on $\mathcal{O}_p$ is left translation-invariant if for any subset $\mathcal{F} \subset \mathcal{O}_p$, $\nu(\mathbf{H}\mathcal{F}) = \nu(\mathcal{F})$ for any fixed $\mathbf{H} \in \mathcal{O}_p$. Similarly, the measure ν is right translation-invariant if $\nu(\mathcal{F}\mathbf{H}) = \nu(\mathcal{F})$ for any $\mathcal{F} \subset \mathcal{O}_p$ and a fixed orthogonal translation $\mathbf{H} \in \mathcal{O}_p$.

It can be proved that $\mathcal{O}_p$ is a compact topological group. As a result, there exists a unique regular left translation-invariant Borel measure ν on $\mathcal{O}_p$ that satisfies $\nu(\mathcal{O}_n) = 1$. Because of compactness, this measure is also the unique regular right translation-invariant Borel measure satisfying $\nu(\mathcal{O}_n) = 1$ (see for example, Royden, 1988, Corollary 20, page 380). This measure ν is known as the Haar distribution. It is the only probability distribution that has the property of left and right translation-invariance.

If $\mathbf{H}$ is a fixed orthogonal translation and $\mathbf{U}$ is a random matrix in $\mathcal{O}_p$ with the Haar distribution, then $\mathbf{HU}$ has the same distribution as $\mathbf{U}$, and similarly, $\mathbf{UH}$ has the same distribution

as $\mathbf{U}$. We will write $\mathbf{HU} \stackrel{\mathrm{d}}{=} \mathbf{U} \stackrel{\mathrm{d}}{=} \mathbf{UH}$ to indicate that the matrices $\mathbf{HU}$, $\mathbf{U}$, and $\mathbf{UH}$ are equal in distribution.

Let $\mathbf{S}_n = \mathbf{GLG}^T$ be the spectral decomposition of the sample covariance matrix $\mathbf{S}_n$. The matrix $\mathbf{G}$ is orthogonal. Note that since $n\mathbf{S}_n = \mathbf{G}(n\mathbf{L})\mathbf{G}^T$, $\mathbf{G}$ also contains the eigenvectors of the sum of squares and crossproducts matrix, $n\mathbf{S}_n$. When the rows of $\mathbf{Y}$ are independent and identically distributed $N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$, the distribution of $n\mathbf{S}_n$ is Wishart, $W_p(\sigma^2\mathbf{I}_p, n-1)$. Next we argue that the joint distribution of the eigenvectors $n\mathbf{S}_n$ is the Haar distribution on $\mathcal{O}_p$.

Since $n\mathbf{S}_n \sim W_p(\sigma^2\mathbf{I}_p, n-1)$, we can write $n\mathbf{S}_n = \sum_{i=1}^{n-1} \boldsymbol{z}_i\boldsymbol{z}_i^T$ where $\boldsymbol{z}_i$ are independently and identically distributed according to $N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$ (Anderson, 2003, Section 3.3). Let $\mathbf{H}$ be a fixed orthogonal matrix chosen arbitrarily from $\mathcal{O}_p$. Then since uncorrelated normal random variables are invariant under orthogonal transformations, $\mathbf{H}\boldsymbol{z}_i \stackrel{\mathrm{d}}{=} \boldsymbol{z}_i$ for $i = 1, \ldots, n-1$. As a result, the matrices $n\mathbf{S}_n$ and $\mathbf{H}(n\mathbf{S}_n)\mathbf{H}^T$ are equal in distribution since

$$n\mathbf{S}_n = \sum_{i=1}^{n-1} \boldsymbol{z}_i\boldsymbol{z}_i^T \stackrel{\mathrm{d}}{=} \sum_{i=1}^{n-1} \mathbf{H}\boldsymbol{z}_i(\mathbf{H}\boldsymbol{z}_i)^T = \mathbf{H}\left(\sum_{i=1}^{n-1} \boldsymbol{z}_i\boldsymbol{z}_i^T\right)\mathbf{H}^T = \mathbf{H}(n\mathbf{S}_n)\mathbf{H}^T.$$

Therefore

$$\mathbf{G}(n\mathbf{L})\mathbf{G}^T = n\mathbf{S}_n \stackrel{\mathrm{d}}{=} \mathbf{H}(n\mathbf{S}_n)\mathbf{H}^T = \mathbf{H}\{\mathbf{G}(n\mathbf{L})\mathbf{G}^T\}\mathbf{H}^T = (\mathbf{HG})n\mathbf{L}(\mathbf{HG})^T$$

so that $\mathbf{HG} \stackrel{\mathrm{d}}{=} \mathbf{G}$. Thus the distribution of $\mathbf{G}$, the matrix of norm-1 eigenvectors of $n\mathbf{S}_n$, is invariant under left translations. Hence $\mathbf{G}$ is distributed according to the Haar distribution on $\mathcal{O}_p$. Next we derive the distribution of the first column of $\mathbf{G}$, $\boldsymbol{g}^{(1)}$.

Let $\boldsymbol{g}^{(1)}$ be the normalized eigenvector of $\mathbf{S}_n$ (or $n\mathbf{S}_n$) corresponding to the largest eigenvalue. We call $\boldsymbol{g}^{(1)}$ the the principal eigenvector of $\mathbf{S}_n$. Stewart (1980) presents the following interesting observation about random matrices in $\mathcal{O}_p$ that are distributed according to the Haar distribution. An easy characterization of the Haar distribution follows from this observation. Let $\mathbf{X}$ be a $p \times p$ matrix whose elements are all independently and identically distributed as

$N(0, \sigma^2)$ random variables. Let $\mathbf{X} = \mathbf{QR}$ be the QR decomposition of $\mathbf{X}$, where $\mathbf{Q}$ is a $p \times p$ orthogonal matrix in $\mathcal{O}_p$, and $\mathbf{R}$ is an upper triangular matrix normalized so that its diagonal elements are positive. Let $\mathbf{H} \in \mathcal{O}_p$ be a fixed orthogonal transformation. Because the elements of $\mathbf{X}$ are uncorrelated $N(0, \sigma^2)$ random variables, the distribution of $\mathbf{X}$ is invariant under fixed orthogonal transformations, i.e., $\mathbf{HX} \stackrel{\mathrm{d}}{=} \mathbf{X}$. Substituting $\mathbf{X} = \mathbf{QR}$ gives that $\mathbf{QR} \stackrel{\mathrm{d}}{=} (\mathbf{HQ})\mathbf{R}$, and hence since $\mathbf{R}$ is invertible, $\mathbf{Q} \stackrel{\mathrm{d}}{=} \mathbf{HQ}$. Therefore the distribution of $\mathbf{Q}$ on $\mathcal{O}_p$ is invariant under left translations. Hence $\mathbf{Q}$ is distributed according to the Haar distribution.

We are interested in the distribution of the first column of $\mathbf{G} \in \mathcal{O}_p$ that is distributed according to the Haar distribution. Stewart's argument implies that there exists a matrix $\mathbf{X}$ whose elements are independently distributed according to $N(0, \sigma^2)$ and a matrix $\mathbf{R}$ which is upper triangular normalized to have positive diagonal elements, such that $\mathbf{X} = \mathbf{GR}$ is the QR decomposition (with $\mathbf{Q} = \mathbf{G}$). One way of computing the QR decomposition is by using the Gram-Schmidt algorithm. From this it is easy to see that if $\mathbf{X} = \mathbf{GR}$, then the first column of $\mathbf{G}$ is just the first column of $\mathbf{X}$ (call this $\boldsymbol{x}_1$), divided by its length $\|\boldsymbol{x}_1\|$. Since the elements of $\mathbf{X}$ are independent and identically distributed according to the $N(0, \sigma^2)$ distribution, the distribution of the first column of $\mathbf{G}$ is that of a $N_p(\mathbf{0}, \sigma^2\mathbf{I}_p)$ random vector $\boldsymbol{x}_1$ divided by its length, i.e., $\boldsymbol{g}^{(1)} \stackrel{\mathrm{d}}{=} \boldsymbol{x}_1/\|\boldsymbol{x}_1\|$. Finally, since $\boldsymbol{x}_1 \stackrel{\mathrm{d}}{=} \sigma\boldsymbol{z}$ where $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$, and $\|\boldsymbol{x}_1\| = \sigma\|\boldsymbol{z}\|$,

$$\boldsymbol{g}^{(1)} \stackrel{\mathrm{d}}{=} \boldsymbol{z}/\|\boldsymbol{z}\|, \quad \text{where } \boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I}_p). \quad \square$$

B.3 Proof of Theorem 2.3.3

Distribution of t_r when $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$

Theorem. *Let $\mathbf{Y}$ be a $n \times p$ data matrix with each row independent and identically distributed according to the $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution. Let $\mathbf{S}_n$ be the sample covariance matrix, and let $\boldsymbol{g}^{(1)}$ be the normalized eigenvector of $\mathbf{S}_n$ corresponding to its largest eigenvalue. Let $\boldsymbol{\kappa}$ be an arbitrary p-vector. Define*

$$r_r = \boldsymbol{\kappa}^T \boldsymbol{g}^{(1)} / \|\boldsymbol{\kappa}\| \quad \textit{and} \quad t_r = (p-1)^{1/2} r_r (1 - r_r^2)^{-1/2}.$$

Then $|r_r| \leq 1$ with $|r_r| = 1$ if and only if $\boldsymbol{g}^{(1)}$ and $\boldsymbol{\kappa}$ are linearly dependent. If $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$ then r_r has the distribution of a sample correlation coefficient with $\rho = 0$ and $p - 1$ degrees of freedom, and t_r has a t-distribution with $p - 1$ degrees of freedom.

Proof. For any $\boldsymbol{\Sigma}$, by the Cauchy-Schwarz inequality,

$$|\boldsymbol{\kappa}^T \boldsymbol{g}^{(1)}| \leq \|\boldsymbol{\kappa}\| \|\boldsymbol{g}^{(1)}\| = \|\boldsymbol{\kappa}\|$$

since $\boldsymbol{g}^{(1)}$ has norm 1, so that $|r_r| \leq 1$. It also follows that $|\boldsymbol{\kappa}^T \boldsymbol{g}^{(1)}| = \|\boldsymbol{\kappa}\|$ (and thus $|r_r| = 1$) if and only if $\boldsymbol{\kappa}$ and $\boldsymbol{g}^{(1)}$ are linearly dependent (see for example the proof of the Cauchy-Schwarz inequality in Marsden and Hoffman (1993, pages 61–62)).

When $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p$, the distribution of $\boldsymbol{g}^{(1)}$ is known exactly, and is equal to the distribution of $\boldsymbol{z}/\|\boldsymbol{z}\|$ where $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ (Theorem 2.3.2). Thus we can write

$$r_r \stackrel{\mathrm{d}}{=} \frac{\boldsymbol{\kappa}^T \boldsymbol{z}}{\|\boldsymbol{\kappa}\| \|\boldsymbol{z}\|}.$$

If $\boldsymbol{z} \sim N_p(\mathbf{0}, \mathbf{I}_p)$ then $\boldsymbol{\kappa}^T \boldsymbol{z} \sim N(0, \|\boldsymbol{\kappa}\|^2)$ and $\boldsymbol{\kappa}^T \boldsymbol{z}/\|\boldsymbol{\kappa}\| \sim N(0, 1)$. Let

$$\mathbf{P}_\kappa = \boldsymbol{\kappa}(\boldsymbol{\kappa}^T \boldsymbol{\kappa})^{-1} \boldsymbol{\kappa}^T = \boldsymbol{\kappa}\boldsymbol{\kappa}^T / \|\boldsymbol{\kappa}\|^2,$$

a projection matrix with rank 1. Then $\boldsymbol{z}^T(\mathbf{I}_p - \mathbf{P}_\kappa)\boldsymbol{z}$ has a χ^2 distribution with $p-1$ degrees of freedom. Since $\mathbf{P}_\kappa\boldsymbol{\kappa} = \boldsymbol{\kappa}$ we have

$$(\boldsymbol{\kappa}^T/\|\boldsymbol{\kappa}\|)\mathrm{Var}(\boldsymbol{z})(\mathbf{I} - \mathbf{P}_\kappa) = \boldsymbol{\kappa}^T\,\mathbf{I}\,(\mathbf{I} - \mathbf{P}_\kappa)/\|\boldsymbol{\kappa}\| = (\boldsymbol{\kappa}^T - \boldsymbol{\kappa}^T)/\|\boldsymbol{\kappa}\| = \mathbf{0}^T.$$

Therefore the quadratic form is independent of $\boldsymbol{\kappa}^T\boldsymbol{z}/\|\boldsymbol{\kappa}\|$, and the ratio

$$\frac{\boldsymbol{\kappa}^T\boldsymbol{z}/\|\boldsymbol{\kappa}\|}{\{\boldsymbol{z}^T(\mathbf{I} - \mathbf{P}_\kappa)\boldsymbol{z}/(p-1)\}^{1/2}} \tag{B.1}$$

follows a Student's t distribution with $p-1$ degrees of freedom. We claim that the ratio in (B.1) has the same distribution as t_r of the theorem.

Simplifying the ratio we get

$$\begin{aligned}
(p-1)^{1/2}\frac{\boldsymbol{\kappa}^T\boldsymbol{z}/\|\boldsymbol{\kappa}\|}{\{\boldsymbol{z}^T\boldsymbol{z} - \boldsymbol{z}^T\boldsymbol{\kappa}\boldsymbol{\kappa}^T\boldsymbol{z}/\|\boldsymbol{\kappa}\|^2\}^{1/2}} &= (p-1)^{1/2}\frac{\boldsymbol{\kappa}^T\boldsymbol{z}/\|\boldsymbol{\kappa}\|}{\{\|\boldsymbol{z}\|^2 - (\boldsymbol{\kappa}^T\boldsymbol{z})^2/\|\boldsymbol{\kappa}\|^2\}^{1/2}} \\
&= (p-1)^{1/2}\frac{\boldsymbol{\kappa}^T\boldsymbol{z}/\|\boldsymbol{\kappa}\|\|\boldsymbol{z}\|}{\{1 - (\boldsymbol{\kappa}^T\boldsymbol{z})^2/\|\boldsymbol{\kappa}\|^2\|\boldsymbol{z}\|^2\}^{1/2}} \\
&\stackrel{\mathrm{d}}{=} (p-1)^{1/2}\frac{r_r}{(1-r_r^2)^{1/2}} \\
&= t_r.
\end{aligned}$$

Hence t_r has the same distribution as the ratio in formula (B.1), namely a student's t distribution with $p-1$ degrees of freedom.

Since t_r is a one-to-one function of r_r, the distribution of r_r can be derived using a change of variables approach (e.g., Anderson, 2003, page 120). The density of t_r is

$$f_p(t) = \frac{\Gamma(p/2)}{(p-1)^{1/2}\,\Gamma[(p-1)/2](\pi)^{1/2}}\left(1 + \frac{t^2}{p-1}\right)^{-p/2}.$$

Appendix B. Proofs of theorems for principal component analysis approach

Substituting $t = (p-1)^{1/2} r (1-r^2)^{-1/2}$ with Jacobian $(p-1)^{1/2}(1-r^2)^{-3/2}$ gives

$$\begin{aligned} f_p(r) &= \frac{\Gamma(p/2)}{\Gamma[(p-1)/2](\pi)^{1/2}} \left(1 + \frac{r^2}{1-r^2}\right)^{-p/2} (1-r^2)^{-3/2} \\ &= \frac{\Gamma(p/2)}{\Gamma[(p-1)/2](\pi)^{1/2}} \left(1-r^2\right)^{p/2} (1-r^2)^{-3/2} \\ &= \frac{\Gamma(p/2)}{\Gamma[(p-1)/2](\pi)^{1/2}} \left(1-r^2\right)^{(p-3)/2} \end{aligned}$$

which is the density of a correlation coefficient with true correlation $\rho = 0$ and $p-1$ degrees of freedom. □

Appendix C

Details about maximum likelihood factor analysis

C.1 Derivation of the MLFA objective function, F_k, when $n \leq p$

In this section we derive objective function F_k in (2.22) that is minimized in order to obtain the maximum likelihood estimators of the parameters $\mathbf{\Lambda}_k$ and $\mathbf{\Psi}$ of the k-factor model. Consider the factor model described in equation (2.4) of Section 2.2.1. For now, assume that the number of common factors is known to equal k. Under the k-factor model, the marginal distribution of $\boldsymbol{y}_i$ is $N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$, with density

$$f(\boldsymbol{y}; \boldsymbol{\mu}, \mathbf{\Sigma}) = (2\pi)^{-p/2} |\mathbf{\Sigma}|^{-1/2} \exp\left\{-\tfrac{1}{2}(\boldsymbol{y} - \boldsymbol{\mu})^T \mathbf{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu})\right\}.$$

Appendix C. Details about maximum likelihood factor analysis

Based on the joint distribution of $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ we define a likelihood function for the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$,

$$\begin{aligned}
L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) &= \prod_{i=1}^{n} (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\tfrac{1}{2}(\boldsymbol{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu})\right\} \\
&= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n} (\boldsymbol{y}_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \boldsymbol{\mu})\right\} \\
&= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\tfrac{1}{2} \sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})\right\} \\
&\quad \times \exp\left\{-\tfrac{n}{2} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\right\} \\
&= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\tfrac{1}{2} \text{trace}\left[\sum_{i=1}^{n} (\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T \boldsymbol{\Sigma}^{-1}\right]\right\} \\
&\quad \times \exp\left\{-\tfrac{n}{2} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\right\} \\
&= (2\pi)^{-np/2} |\boldsymbol{\Sigma}|^{-n/2} \exp\left\{-\tfrac{n}{2} \text{trace}\left(\mathbf{S}_n \boldsymbol{\Sigma}^{-1}\right)\right\} \\
&\quad \times \exp\left\{-\tfrac{n}{2} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\right\}.
\end{aligned}$$

The loglikelihood function is

$$\log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = c - \tfrac{n}{2}\left[\log|\boldsymbol{\Sigma}| + \text{trace}\left(\mathbf{S}_n \boldsymbol{\Sigma}^{-1}\right) + (\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\bar{\boldsymbol{y}} - \boldsymbol{\mu})\right]$$

for constant c free of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. Since $\boldsymbol{\Sigma}$ (and therefore $\boldsymbol{\Sigma}^{-1}$) is positive definite, the quadratic form is always non-negative, and is minimized (with a value of 0) when $\boldsymbol{\mu} = \bar{\boldsymbol{y}}$. Hence the maximum likelihood estimator (MLE) of $\boldsymbol{\mu}$ is $\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$ for any positive definite $\boldsymbol{\Sigma}$. The maximum likelihood estimator of $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}$ thus minimizes the objective function

$$F_k(\boldsymbol{\Lambda}_k, \boldsymbol{\Psi}) = \log|\boldsymbol{\Sigma}| + \text{trace}\left(\mathbf{S}_n \boldsymbol{\Sigma}^{-1}\right)$$

subject to the constraint that $\boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda}_k = \boldsymbol{\Delta}$ is diagonal.

C.2 Minimizing the objective function F_k

To minimize F_k we start by finding the partial derivatives of F_k with respect to elements of $\mathbf{\Lambda}_k$ and $\mathbf{\Psi}$. Let σ_{ij} denote the ij-th element of $\mathbf{\Sigma}$. From Lawley and Maxwell (1971) we know that

$$\begin{aligned}
\frac{\partial \log|\mathbf{\Sigma}|}{\partial \sigma_{jj}} &= (\mathbf{\Sigma}^{-1})_{jj} \\
\frac{\partial \log|\mathbf{\Sigma}|}{\partial \sigma_{ji}} &= 2(\mathbf{\Sigma}^{-1})_{ji} \quad (i \neq j) \\
\frac{\partial \text{trace}(\mathbf{S}_n\mathbf{\Sigma}^{-1})}{\partial \sigma_{jj}} &= -(\mathbf{\Sigma}^{-1}\mathbf{S}_n\mathbf{\Sigma}^{-1})_{jj} \\
\frac{\partial \text{trace}(\mathbf{S}_n\mathbf{\Sigma}^{-1})}{\partial \sigma_{ji}} &= -2(\mathbf{\Sigma}^{-1}\mathbf{S}_n\mathbf{\Sigma}^{-1})_{ji} \quad (i \neq j)
\end{aligned}$$

where the subscript notation $(\mathbf{A})_{ji}$ denotes the element in the j-th row and the i-th column of the matrix $\mathbf{A}$. Therefore

$$\begin{aligned}
\frac{\partial F_k}{\partial \sigma_{jj}} &= (\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{S}_n\mathbf{\Sigma}^{-1})_{jj} \\
\frac{\partial F_k}{\partial \sigma_{ji}} &= 2(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1}\mathbf{S}_n\mathbf{\Sigma}^{-1})_{ji} \quad (i \neq j).
\end{aligned}$$

These partial derivatives for our objective function F_k are the same as those for the objective function of Jöreskog (1967). Let λ_{is} denote the is-th element of $\mathbf{\Lambda}_k$, and ψ_j denote the j-th diagonal element of $\mathbf{\Psi}$. Since $\mathbf{\Sigma} = \mathbf{\Lambda}_k\mathbf{\Lambda}_k^T + \mathbf{\Psi}$, the individual elements of $\mathbf{\Sigma}$ can be written as $\sigma_{jj} = \sum_{s=1}^{k} \lambda_{js}^2 + \psi_j$ and $\sigma_{ij} = \sigma_{ji} = \sum_{s=1}^{k} \lambda_{is}\lambda_{js}$ for $i \neq j$. Therefore

$$\frac{\partial \sigma_{jj}}{\partial \lambda_{jr}} = 2\lambda_{jr}, \quad \frac{\partial \sigma_{ji}}{\partial \lambda_{jr}} = \lambda_{ir} \ (i \neq j), \quad \frac{\partial \sigma_{jj}}{\partial \psi_j} = 1, \quad \text{and} \quad \frac{\partial \sigma_{ji}}{\partial \psi_j} = 0 \ (i \neq j).$$

Appendix C. Details about maximum likelihood factor analysis

By the chain rule,

$$\begin{aligned}
\frac{\partial F_k}{\partial \lambda_{jr}} &= \sum_{i=1}^{p} \left[\left(\partial F_k / \partial \sigma_{ji} \right) \left(\partial \sigma_{ji} / \partial \lambda_{jr} \right) \right] \\
&= 2 \sum_{i=1}^{p} \left(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1} \mathbf{S}_n \mathbf{\Sigma}^{-1} \right)_{ji} \lambda_{ir} \\
&= 2 \left[\mathbf{\Sigma}^{-1} \left(\mathbf{\Sigma} - \mathbf{S}_n \right) \mathbf{\Sigma}^{-1} \mathbf{\Lambda} \right]_{jr}, \quad \text{and} \\
\frac{\partial F_k}{\partial \psi_j} &= \left(\partial F_k / \partial \sigma_{jj} \right) \left(\partial \sigma_{jj} / \partial \psi_j \right) \\
&= \left(\mathbf{\Sigma}^{-1} - \mathbf{\Sigma}^{-1} \mathbf{S}_n \mathbf{\Sigma}^{-1} \right)_{jj} \\
&= \left[\mathbf{\Sigma}^{-1} \left(\mathbf{\Sigma} - \mathbf{S}_n \right) \mathbf{\Sigma}^{-1} \right]_{jj}.
\end{aligned}$$

In matrix notation,

$$\begin{aligned}
\frac{\partial F_k}{\partial \mathbf{\Lambda}_k} &= 2 \mathbf{\Sigma}^{-1} \left(\mathbf{\Sigma} - \mathbf{S}_n \right) \mathbf{\Sigma}^{-1} \mathbf{\Lambda}_k, \quad \text{and} \\
\frac{\partial F_k}{\partial \mathbf{\Psi}} &= \operatorname{diag} \left[\mathbf{\Sigma}^{-1} \left(\mathbf{\Sigma} - \mathbf{S}_n \right) \mathbf{\Sigma}^{-1} \right].
\end{aligned} \tag{C.1}$$

Our strategy is to first find the conditional minimum of F_k for a given value of $\mathbf{\Psi}$, giving the function $f_k(\mathbf{\Psi}) = F_k(\widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi}), \mathbf{\Psi})$ where $\widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi})$ is the value of $\mathbf{\Lambda}_k$ that minimizes F_k for a given $\mathbf{\Psi}$. Then we find $\mathbf{\Psi}$ to minimize f_k. For a fixed $\mathbf{\Psi}$, the minimum of F_k occurs when $\partial F_k / \partial \mathbf{\Lambda}_k = \mathbf{0}$:

$$\begin{aligned}
\frac{\partial F_k}{\partial \mathbf{\Lambda}_k} = \mathbf{0} &\iff \mathbf{\Sigma}^{-1} (\mathbf{\Sigma} - \mathbf{S}_n) \mathbf{\Sigma}^{-1} \mathbf{\Lambda}_k = \mathbf{0} \\
&\iff (\mathbf{\Sigma} - \mathbf{S}_n) \mathbf{\Sigma}^{-1} \mathbf{\Lambda}_k = \mathbf{0}.
\end{aligned} \tag{C.2}$$

We have the following identities,

$$\mathbf{\Sigma}^{-1} = (\mathbf{\Lambda}_k \mathbf{\Lambda}_k^T + \mathbf{\Psi})^{-1} = \mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1} \mathbf{\Lambda}_k (\mathbf{I}_k + \mathbf{\Delta})^{-1} \mathbf{\Lambda}_k^T \mathbf{\Psi}^{-1} \tag{C.3}$$

where $\boldsymbol{\Delta} = \boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda}_k$, and

$$\begin{aligned}
\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda}_k &= \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k - \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}_k^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k \\
&= \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k - \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta} \\
&= \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k(\mathbf{I}_k - (\mathbf{I}_k + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}) \\
&= \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1}
\end{aligned} \tag{C.4}$$

since $(\mathbf{I}+\mathbf{PQ})^{-1} = \mathbf{I} - \mathbf{P}(\mathbf{I}+\mathbf{QP})^{-1}\mathbf{Q}$ with $\mathbf{P} = \mathbf{I}_k$ and $\mathbf{Q} = \boldsymbol{\Delta}$. Substituting (C.4) into (C.2) gives

$$\begin{aligned}
& \mathbf{0} = (\boldsymbol{\Sigma} - \mathbf{S}_n)\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1} \\
\iff \quad & \mathbf{0} = (\boldsymbol{\Sigma} - \mathbf{S}_n)\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k \\
\iff \quad & \mathbf{0} = (\boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi} - \mathbf{S}_n)\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k \\
\iff \quad & \mathbf{S}_n\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k + \boldsymbol{\Lambda}_k \\
\iff \quad & \mathbf{S}_n\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_k\boldsymbol{\Delta} + \boldsymbol{\Lambda}_k \\
\iff \quad & \mathbf{S}_n\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k = \boldsymbol{\Lambda}_k(\mathbf{I}_k + \boldsymbol{\Delta}).
\end{aligned}$$

Premultiplying by $\boldsymbol{\Psi}^{-1/2}$ gives

$$(\boldsymbol{\Psi}^{-1/2}\mathbf{S}_n\boldsymbol{\Psi}^{-1/2})(\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Lambda}_k) = (\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Lambda}_k)(\mathbf{I}_k + \boldsymbol{\Delta}). \tag{C.5}$$

Since $\mathbf{S}^* = \boldsymbol{\Psi}^{-1/2}\mathbf{S}_n\boldsymbol{\Psi}^{-1/2}$ is symmetric, under the constraint that $\boldsymbol{\Delta} = \boldsymbol{\Lambda}_k^T\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k$ is diagonal, equation (C.5) is satisfied when the columns of $\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Lambda}_k$ are a set of k (unnormalized) eigenvectors of $\mathbf{S}^*$ and the diagonal elements of $(\mathbf{I}_k + \boldsymbol{\Delta})$ are the corresponding eigenvalues.

The $p \times p$ matrix $\mathbf{S}^* = \boldsymbol{\Psi}^{-1/2}\mathbf{S}_n\boldsymbol{\Psi}^{-1/2}$ has p eigenvalues, $\theta_1 \geq \theta_2 \geq \ldots \geq \theta_p$. Let $\boldsymbol{\Omega} = (\boldsymbol{\omega}_1, \boldsymbol{\omega}_2, \ldots, \boldsymbol{\omega}_k)$ be a set of unit length eigenvectors corresponding to the k largest eigenvalues. The columns of $\boldsymbol{\Omega}$ are orthogonal so that $\boldsymbol{\Omega}^T\boldsymbol{\Omega} = \mathbf{I}_k$. Let $\boldsymbol{\Theta} = \text{diag}\{\theta_1, \theta_2, \ldots, \theta_k\}$. It turns out

that the absolute minimum of F_k over $\mathbb{R}^{p\times k}$ (conditional on a given $\mathbf{\Psi}$) is attained by choosing the k columns of $\mathbf{\Omega}$ (appropriately normalized) as the columns of $\mathbf{\Psi}^{-1/2}\mathbf{\Lambda}_k$. Since

$$\mathbf{\Theta} = \mathbf{I}_k + \mathbf{\Delta} = \mathbf{I}_k + (\mathbf{\Psi}^{-1/2}\mathbf{\Lambda}_k)^T(\mathbf{\Psi}^{-1/2}\mathbf{\Lambda}_k),$$

the eigenvectors (i.e., columns of $\mathbf{\Psi}^{-1/2}\mathbf{\Lambda}_k$) should be scaled so that the sum of squares of the j-th column equals $\theta_j - 1$. Thus, for a given $\mathbf{\Psi}$, let $\widetilde{\mathbf{\Lambda}}_k$ be defined by

$$\mathbf{\Psi}^{-1/2}\widetilde{\mathbf{\Lambda}}_k = \mathbf{\Omega}(\mathbf{\Theta} - \mathbf{I}_k)^{1/2}, \quad \text{or}$$

$$\widetilde{\mathbf{\Lambda}}_k = \mathbf{\Psi}^{1/2}\mathbf{\Omega}(\mathbf{\Theta} - \mathbf{I}_k)^{1/2} \tag{C.6}$$

Note that if any of the eigenvalues on the diagonal of $\mathbf{\Theta}$ are less than one then there will be no real solution for $\widetilde{\mathbf{\Lambda}}_k$. Jöreskog (1967) notes that this happens seldom in practice unless k is very large. We will avoid this problem by choosing k small enough so that $\theta_1 \geq \theta_2 \ldots \geq \theta_k \geq 1$. This is justified in Section C.3.

Next we evaluate

$$F_k(\mathbf{\Lambda}_k, \mathbf{\Psi}) = \log|\mathbf{\Sigma}| + \text{trace}\left(\mathbf{S}_n\mathbf{\Sigma}^{-1}\right)$$

at $\widetilde{\mathbf{\Lambda}}_k$ defined in (C.6). Let $\widetilde{\mathbf{\Sigma}} = \widetilde{\mathbf{\Lambda}}_k\widetilde{\mathbf{\Lambda}}_k^T + \mathbf{\Psi}$. Then

$$\begin{aligned}
|\mathbf{\Psi}^{-1/2}\widetilde{\mathbf{\Sigma}}\mathbf{\Psi}^{-1/2}| &= |\mathbf{\Psi}^{-1/2}(\widetilde{\mathbf{\Lambda}}_k\widetilde{\mathbf{\Lambda}}_k^T + \mathbf{\Psi})\mathbf{\Psi}^{-1/2}| \\
&= |\mathbf{I}_p + \mathbf{\Psi}^{-1/2}\widetilde{\mathbf{\Lambda}}_k\widetilde{\mathbf{\Lambda}}_k^T\mathbf{\Psi}^{-1/2}| \\
&= |\mathbf{I}_k + \widetilde{\mathbf{\Delta}}| \\
&= |\mathbf{\Theta}|
\end{aligned}$$

where $\widetilde{\mathbf{\Delta}} = \widetilde{\mathbf{\Lambda}}_k^T \mathbf{\Psi}^{-1} \widetilde{\mathbf{\Lambda}}_k$. Thus

$$|\mathbf{\Psi}|^{-1/2} |\widetilde{\mathbf{\Sigma}}| |\mathbf{\Psi}|^{-1/2} = |\mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Sigma}} \mathbf{\Psi}^{-1/2}| = \theta_1 \theta_2 \dots \theta_k$$

or $|\widetilde{\mathbf{\Sigma}}| = (\theta_1 \theta_2 \dots \theta_k)(\psi_1 \psi_2 \dots \psi_p)$. From (C.3) we get

$$\begin{aligned}
(\mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Sigma}} \mathbf{\Psi}^{-1/2})^{-1} &= \mathbf{\Psi}^{1/2} \widetilde{\mathbf{\Sigma}}^{-1} \mathbf{\Psi}^{1/2} \\
&= \mathbf{\Psi}^{1/2} [\mathbf{\Psi}^{-1} - \mathbf{\Psi}^{-1} \widetilde{\mathbf{\Lambda}}_k (\mathbf{I}_k + \widetilde{\mathbf{\Delta}})^{-1} \widetilde{\mathbf{\Lambda}}_k^T \mathbf{\Psi}^{-1}] \mathbf{\Psi}^{1/2} \\
&= \mathbf{I}_p - \mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Lambda}}_k \mathbf{\Theta}^{-1} \widetilde{\mathbf{\Lambda}}_k^T \mathbf{\Psi}^{-1/2}.
\end{aligned}$$

Using equation (C.6) and substituting the above we get

$$\begin{aligned}
\operatorname{trace}(\mathbf{S}_n \widetilde{\mathbf{\Sigma}}^{-1}) &= \operatorname{trace}(\mathbf{S}_n \mathbf{\Psi}^{-1/2} \mathbf{\Psi}^{1/2} \widetilde{\mathbf{\Sigma}}^{-1} \mathbf{\Psi}^{1/2} \mathbf{\Psi}^{-1/2}) \\
&= \operatorname{trace}[(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2})(\mathbf{\Psi}^{1/2} \widetilde{\mathbf{\Sigma}}^{-1} \mathbf{\Psi}^{1/2})] \\
&= \operatorname{trace}[(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2})(\mathbf{I}_p - \mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Lambda}}_k \mathbf{\Theta}^{-1} \widetilde{\mathbf{\Lambda}}_k^T \mathbf{\Psi}^{-1/2})] \\
&= \operatorname{trace}(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2}) - \operatorname{trace}([(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2})(\mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Lambda}}_k) \mathbf{\Theta}^{-1} \widetilde{\mathbf{\Lambda}}_k^T \mathbf{\Psi}^{-1/2}] \\
&= \operatorname{trace}(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2}) - \operatorname{trace}[(\mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Lambda}}_k) \mathbf{\Theta} \mathbf{\Theta}^{-1} \widetilde{\mathbf{\Lambda}_k}^T \mathbf{\Psi}^{-1/2}] \\
&= \operatorname{trace}(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2}) - \operatorname{trace}(\mathbf{\Psi}^{-1/2} \widetilde{\mathbf{\Lambda}_k} \widetilde{\mathbf{\Lambda}_k}^T \mathbf{\Psi}^{-1/2}) \\
&= \operatorname{trace}(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2}) - \operatorname{trace}(\widetilde{\mathbf{\Lambda}}_k^T \mathbf{\Psi}^{-1} \widetilde{\mathbf{\Lambda}}_k) \\
&= \operatorname{trace}(\mathbf{\Psi}^{-1/2} \mathbf{S}_n \mathbf{\Psi}^{-1/2}) - \operatorname{trace}(\mathbf{\Theta} - \mathbf{I}_k) \\
&= \sum_{m=1}^{p} \theta_m - \sum_{m=1}^{k} (\theta_m - 1) \\
&= \theta_{k+1} + \ldots + \theta_p + k.
\end{aligned}$$

Thus, F_k evaluated at $\widetilde{\mathbf{\Lambda}}_k = \widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi})$ defined in (C.6) gives

$$\begin{aligned} f_k(\mathbf{\Psi}) &= F_k(\widetilde{\mathbf{\Lambda}}_k(\mathbf{\Psi}), \mathbf{\Psi}) \\ &= \log|\widetilde{\mathbf{\Sigma}}| + \operatorname{trace}\left(\mathbf{S}_n\widetilde{\mathbf{\Sigma}}^{-1}\right) \\ &= \log[(\theta_1\theta_2\cdots\theta_k)(\psi_1\psi_2\cdots\psi_p)] + \sum_{m=k+1}^{p}\theta_m + k \\ &= k + \sum_{m=1}^{p}\log\psi_m + \sum_{m=1}^{k}\log\theta_m + \sum_{m=k+1}^{p}\theta_m. \end{aligned} \tag{C.7}$$

Note that the function f_k is defined as long as $\psi_m > 0$ for $m = 1, \ldots, p$, and $\theta_m > 0$ for $m = 1, \ldots, k$. The second condition is automatically satisfied since we have chosen k so that $\theta_m \geq 1$ for $m = 1, \ldots, k$.

Any set of k eigenvectors of $\mathbf{S}^* = \Psi^{-1/2}\mathbf{S}_n\mathbf{\Psi}^{-1/2}$ will satisfy the equation (C.5). Now we show that choosing the eigenvectors that correspond to the k largest eigenvalues as the columns of $\mathbf{\Psi}^{-1/2}\widetilde{\mathbf{\Lambda}}_k$ minimizes F_k for a given value of $\mathbf{\Psi}$. First note that for $x > 1$, the function $x - \log x$ is strictly increasing since $\frac{\mathrm{d}}{\mathrm{d}x}(x - \log x) = 1 - 1/x > 0$ for $x > 1$. Now suppose that θ_{k+b} in (C.7) is exchanged with θ_a where $a \leq k$ and $b \geq 1$. We assume that θ_a and θ_{k+b} are both greater than or equal to 1 (otherwise $\mathbf{\Lambda}_k$ will be imaginary). Interchanging these roots will lead to the following value of the objective function f_k

$$\bar{f}_k(\mathbf{\Psi}) = k + \sum_{m=1}^{p}\log\psi_m + \sum_{m\in\mathcal{M}}\log\theta_m + \sum_{m\in\bar{\mathcal{M}}}\theta_m$$

where $\mathcal{M} = \{\theta_1, \theta_2, \ldots, \theta_{a-1}, \theta_{k+b}, \theta_{a+1}, \ldots, \theta_k\}$ and $\bar{\mathcal{M}} = \{\theta_1, \theta_2, \ldots, \theta_p\} \setminus \mathcal{M}$. Subtracting this value of the objective function from (C.7) gives

$$\bar{f}_k(\mathbf{\Psi}) - f_k(\mathbf{\Psi}) = \log\theta_{k+b} + \theta_a - \log\theta_a - \theta_{k+b} = (\theta_a - \log\theta_a) - (\theta_{k+b} - \log\theta_{k+b}).$$

Since $x - \log x$ is increasing for $x > 1$ and $\theta_a \geq \theta_{k+b}$ we have $\bar{f}_k(\mathbf{\Psi}) - f_k(\mathbf{\Psi}) \geq 0$. Hence

choosing the k largest eigenvalues and the corresponding eigenvectors as the solution to (C.5) gives the minimum value of F_k for a given value of $\mathbf{\Psi}$.

C.3 Minimizing F_k when $\theta_j < 1$

For fixed $\mathbf{\Psi}$, the minimizer $\widetilde{\mathbf{\Lambda}}_k$ of F_k defined in (C.6) requires that $\theta_j \geq 1$ for $j = 1, \ldots, k$. In this section we show that when $\theta_j < 1$, the objective F_k is minimized when the j-th column of $\widetilde{\mathbf{\Lambda}}_k$ is zero.

Let $\mathbf{B} = \mathbf{\Psi}^{-1/2}\mathbf{\Lambda}_k$. Then for fixed $\mathbf{\Psi}$, as a minimizer of the objective function F_k defined in (2.22), $\mathbf{B} = \mathbf{\Psi}^{-1/2}\mathbf{\Lambda}_k$ satisfies equation (C.5) so that

$$\mathbf{S}^*\mathbf{B} = \mathbf{B}(\mathbf{I}_k + \mathbf{\Delta}) \quad \text{where} \quad \mathbf{\Delta} = \mathbf{B}^T\mathbf{B}. \tag{C.8}$$

Let $\mathbf{B} = \mathbf{U}\mathbf{C}\mathbf{Q}^T$ be the singular value decomposition of $\mathbf{B}$ where $\mathbf{U}$ is $p \times k$ with $\mathbf{U}^T\mathbf{U} = \mathbf{I}_k$, $\mathbf{C}$ is a $k \times k$ diagonal matrix, and $\mathbf{Q}$ is a $k \times k$ orthogonal matrix. Let $\mathbf{B}^* = \mathbf{B}\mathbf{Q} = \mathbf{U}\mathbf{C}$. Notice that $\mathbf{B}^{*T}\mathbf{B}^* = \mathbf{C}\mathbf{U}^T\mathbf{U}\mathbf{C} = \mathbf{C}^2$, a diagonal matrix with j-th element c_j^2 equal to the squared norm of the j-th column of $\mathbf{B}$. The rotated matrix $\mathbf{B}^*$ also satisfies equation (C.8) since post-multiplying by $\mathbf{Q}$ gives

$$\begin{aligned}
& \mathbf{S}^*\mathbf{B} = \mathbf{B}(\mathbf{I}_k + \mathbf{\Delta}) \\
\iff \quad & \mathbf{S}^*\mathbf{B}\mathbf{Q} = \mathbf{B}(\mathbf{I}_k + \mathbf{\Delta})\mathbf{Q} \\
\iff \quad & \mathbf{S}^*\mathbf{B}\mathbf{Q} = \mathbf{B}\mathbf{Q}\mathbf{Q}^T(\mathbf{I}_k + \mathbf{B}^T\mathbf{B})\mathbf{Q} \\
\iff \quad & \mathbf{S}^*\mathbf{B}^* = \mathbf{B}^*(\mathbf{I}_k + \mathbf{B}^{*T}\mathbf{B}^*).
\end{aligned}$$

Also, this choice of $\mathbf{Q}$ satisfies the constraint that $\mathbf{\Delta} = \mathbf{B}^{*T}\mathbf{B}^* = \mathbf{C}^2$ is diagonal. Since $(\mathbf{I}_k + \mathbf{C}^2)$ is diagonal, the equation

$$\mathbf{S}^*\mathbf{B}^* = \mathbf{B}^*(\mathbf{I}_k + \mathbf{C}^2)$$

is solved if $\mathbf{B}^*$ are the unnormalized eigenvectors of $\mathbf{S}^*$, with the eigenvalues being the diagonal elements of $\mathbf{C}^2$.

We showed in Section C.2 that for fixed $\mathbf{\Psi}$, F_k is minimized when the columns of $\mathbf{B}^*$ are the (unnormalized) eigenvectors of $\mathbf{S}^*$ that correspond to the largest k eigenvalues of $\mathbf{S}^*$, $\{\theta_1 \geq \theta_2 \geq \ldots \geq \theta_k\}$. Let $\mathbf{\Theta} = \text{diag}(\theta_1, \theta_2, \ldots, \theta_k)$. Since $\mathbf{\Theta} = \mathbf{I}_k + \mathbf{C}^2$, the norm of the j-th column of $\mathbf{B}^*$ equals $c_j = (\theta_j - 1)^{1/2}$. Thus, we only get a real solution (i.e., interior point solution) for $\mathbf{B}^*$ when $c_j^2 \geq 0$ or when $\theta_j \geq 1$ for $j = 1, \ldots, k$.

We now show that when $\theta_j < 1$, F_k is minimized for fixed $\mathbf{\Psi}$ when the j-th column of $\mathbf{B}^*$ is zero (i.e., on the boundary of the parameter space). To see this we write the objective function $F_k = \log|\mathbf{\Sigma}| + \text{trace}(\mathbf{S}_n\mathbf{\Sigma}^{-1})$ in terms of $\mathbf{B}^*$. Since we can write $\mathbf{\Lambda}_k = \mathbf{\Psi}^{-1/2}\mathbf{B}^*$,

$$\begin{aligned}
\mathbf{\Sigma} &= \mathbf{\Lambda}_k\mathbf{\Lambda}_k^T + \mathbf{\Psi} = \mathbf{\Psi}^{1/2}\mathbf{B}^*\mathbf{B}^{*T}\mathbf{\Psi}^{1/2} + \mathbf{\Psi}, \\
\log|\mathbf{\Sigma}| &= \log\left(|\mathbf{\Psi}^{1/2}| \cdot |\mathbf{I}_p + \mathbf{B}^*\mathbf{B}^{*T}| \cdot |\mathbf{\Psi}^{1/2}|\right) \\
&= \log|\mathbf{\Psi}| + \log|\mathbf{I}_k + \mathbf{C}^2|, \quad \text{and} \\
\text{trace}(\mathbf{S}_n\mathbf{\Sigma}^{-1}) &= \text{trace}\left\{\mathbf{S}_n^{-1}\mathbf{\Psi}^{-1/2}(\mathbf{I}_p + \mathbf{B}^*\mathbf{B}^{*T})^{-1}\mathbf{\Psi}^{-1/2}\right\} \\
&= \text{trace}\left[\mathbf{\Psi}^{-1/2}\mathbf{S}_n^{-1}\mathbf{\Psi}^{-1/2}\left\{\mathbf{I}_p - \mathbf{B}^*(\mathbf{I}_k + \mathbf{C}^2)^{-1}\mathbf{B}^{*T}\right\}\right] \\
&= \text{trace}(\mathbf{S}^*) - \text{trace}\left\{\mathbf{S}^*\mathbf{B}^*(\mathbf{I}_k + \mathbf{C}^2)^{-1}\mathbf{B}^{*T}\right\} \\
&= \text{trace}(\mathbf{S}^*) - \text{trace}\left\{\mathbf{B}^{*T}\mathbf{S}^*\mathbf{B}^*(\mathbf{I}_k + \mathbf{C}^2)^{-1}\right\}
\end{aligned}$$

where $\mathbf{S}^* = \mathbf{\Psi}^{-1/2}\mathbf{S}_n^{-1}\mathbf{\Psi}^{-1/2}$. Thus for fixed $\mathbf{\Psi}$, the objective function is

$$F_k(\mathbf{B}^*) = \text{constant} + \log|\mathbf{I}_k + \mathbf{C}^2| - \text{trace}\left\{\mathbf{B}^{*T}\mathbf{S}^*\mathbf{B}^*(\mathbf{I}_k + \mathbf{C}^2)^{-1}\right\}. \tag{C.9}$$

Appendix C. Details about maximum likelihood factor analysis

Since $\mathbf{B}^* = \mathbf{UC}$ and $\mathbf{B}^{*T}\mathbf{B}^* = \mathbf{C}^2$ is diagonal with j-th element c_j^2 we can write

$$\log|\mathbf{I}_k + \mathbf{C}^2| = \sum_{j=1}^{k} \log(1 + c_j^2)$$

and

$$\begin{aligned}
\operatorname{trace}\left\{\mathbf{B}^{*T}\mathbf{S}^*\mathbf{B}^*(\mathbf{I}_k + \mathbf{B}^{*T}\mathbf{B}^*)^{-1}\right\} &= \operatorname{trace}\left\{\mathbf{C}\mathbf{U}^T\mathbf{S}^*\mathbf{U}\mathbf{C}(\mathbf{I}_k + \mathbf{C}^2)^{-1}\right\} \\
&= \operatorname{trace}\left\{\mathbf{U}^T\mathbf{S}^*\mathbf{U}\mathbf{C}(\mathbf{I}_k + \mathbf{C}^2)^{-1}\mathbf{C}\right\} \\
&= \operatorname{trace}\left\{\mathbf{U}^T\mathbf{S}^*\mathbf{U}\mathbf{C}(\mathbf{I}_k + \mathbf{C}^2)^{-1}\mathbf{C}\right\} \\
&= \sum_{j=1}^{k} \boldsymbol{u}_j^T\mathbf{S}^*\boldsymbol{u}_j \frac{c_j^2}{1 + c_j^2}
\end{aligned}$$

where $\boldsymbol{u}_j$ is the j-th column of $\mathbf{U}$. The objective function F_k in (C.9) can now be written as

$$F_k(\mathbf{B}^*) = \mathbf{F}_k(\mathbf{U}, \mathbf{C}) = \text{constant} + \sum_{j=1}^{k} \log(1 + c_j^2) - \sum_{j=1}^{k} \boldsymbol{u}_j^T\mathbf{S}^*\boldsymbol{u}_j \frac{c_j^2}{1 + c_j^2}.$$

Let $\delta_j = c_j^2 \geq 0$ for $j = 1, \ldots, k$ (the squared norm of the j-the column of $\mathbf{B}^*$). To minimize F_k over $\boldsymbol{u}_j$ and δ_j note that since $\|\boldsymbol{u}_j\| = 1$ and the $\boldsymbol{u}_j$'s are orthogonal,

$$\sum_{j=1}^{k} \boldsymbol{u}_j^T\mathbf{S}^*\boldsymbol{u}_j \frac{\delta_j}{1 + \delta_j}$$

is maximized with respect to the $\boldsymbol{u}_j$'s when each $\boldsymbol{u}_j$ is the normalized eigenvector of $\mathbf{S}^*$ corresponding to the j-th largest eigenvalue θ_j. Profiling out the $\boldsymbol{u}_j$'s by substituting the maximized value of the above sum (i.e., $\sum_{j=1}^{k} \theta_j \frac{\delta_j}{1+\delta_j}$), the objective function then becomes

$$F_k^*(\delta_1, \ldots, \delta_k) = \text{constant} + \sum_{j=1}^{k} \left\{\log(1 + \delta_j) - \theta_j \frac{\delta_j}{1 + \delta_j}\right\}.$$

To minimize F_k^* over the region where $\delta_j \geq 0, j = 1, \ldots, k$, take the partial derivatives with

respect to δ_j and set equal to zero. The partial derivative of F_k^* with respect to δ_j, $j = 1, \ldots, k$ equals the partial derivative of the function

$$g(\delta_j) = \log(1 + \delta_j) - \theta_j \frac{\delta_j}{1 + \delta_j}, \quad \text{where } \delta_j \geq 0.$$

The derivative of g with respect to δ_j is

$$g'(\delta_j) = \frac{1}{1 + \delta_j} - \theta_j \frac{1}{(1 + \delta_j)^2}.$$

Note that $g'(0) = 1 - \theta_j$. Consider three cases.

$\theta_j > 1$: In this case, $g'(0) < 0$ and $g'(\delta_j) = 0 \Rightarrow \delta_j = \theta_j - 1 > 0$. Thus g is minimized at an interior point. This is the usual solution for j-th column of $\mathbf{B}^*$ described in (C.6) when $\theta_j > 1$.

$\theta_j = 1$: In this case, $g'(0) = 0$ and $g'(\delta_j) = 0 \Rightarrow \delta_j = 0$. So g is minimized on the boundary, and the j-th column of $\mathbf{B}^*$ therefore has norm $c_j = \delta_j^{1/2} = 0$ (i.e., these factor loadings are zero).

$\theta_j < 1$: In this case, $g'(0) > 0$ and $g'(\delta_j) = 0 \Rightarrow \delta_j < 0$ (outside the parameter space) so that g is minimized on the boundary, $\delta_j = 0$. Here again, the j-th column of $\mathbf{B}^*$ has norm $c_j = \delta_j^{1/2} = 0$ (i.e., these factor loadings are zero).

Thus when $\theta_j < 1$, the objective function F_k is minimized for fixed $\mathbf{\Psi}$ when the j-th column of $\mathbf{B}^*$ is zero, or equivalently when the j-th column of $\widetilde{\mathbf{\Lambda}}_k$ is zero. In practice, the number of factors k should be decreased until $\theta_j > 1$ for $j = 1, \ldots, k$.

C.4 Implementing maximum likelihood factor analysis

For $k = 0$ (i.e., no common factors) the approximate marginal covariance is $\mathbf{\Sigma} = \mathbf{\Psi}$, a diagonal matrix. Under this model the loglikelihood is

$$\log L_0 = -\tfrac{n}{2}\left[\log|\mathbf{\Psi}| + \operatorname{trace}(\mathbf{S}_n\mathbf{\Psi}^{-1})\right] = -\tfrac{n}{2}\sum_{j=1}^{p}(\log\psi_j + s_{jj}/\psi_j)$$

where s_{jj} is the j-th diagonal element of $\mathbf{S}_n$. Under this model, the maximum likelihood estimator of $\mathbf{\Psi}$ solves

$$0 = \frac{\partial \log\psi_j}{\partial\psi_j} + s_{jj}\frac{\partial\psi_j^{-1}}{\partial\psi_j} = \frac{1}{\psi_j} - \frac{s_{jj}}{\psi_j^2}.$$

Thus the the MLE of $\mathbf{\Psi}$ is $\widehat{\mathbf{\Psi}} = \operatorname{diag}(\mathbf{S}_n)$, the diagonal elements of the sample covariance matrix.

For $k \geq 1$, an iterative algorithm is used to minimize the concentrated objective function $f_k(\mathbf{\Psi})$ defined in (C.7) over the space where $\psi_j > 0$. To handle the positivity constraint we reparameterize in terms of $\psi_j = \mathrm{e}^{\alpha_j}$ where $\alpha_j \in \mathbb{R}$. To minimize f_k we use a Quasi-Newton line search method such as the `fminunc()` optimizer in MATLAB. Let $\widehat{\mathbf{\Psi}}$ be the minimizer of f_k. The maximum likelihood estimates of the factor loadings are calculated by substituting $\widehat{\mathbf{\Psi}}$ back into the equation (C.6) to get

$$\widehat{\mathbf{\Lambda}}_k = \widehat{\mathbf{\Psi}}^{1/2}\widehat{\mathbf{\Omega}}(\widehat{\mathbf{\Theta}} - \mathbf{I}_k)^{1/2}$$

where $\widehat{\mathbf{\Theta}}$ and $\widehat{\mathbf{\Omega}}$ are the k largest eigenvalues and corresponding normalized eigenvectors of $\widehat{\mathbf{S}}^* = \widehat{\mathbf{\Psi}}^{-1/2}\mathbf{S}_n\widehat{\mathbf{\Psi}}^{-1/2}$.

If the user does not supply a starting value $\mathbf{\Psi}_0$ for the iterative algorithm to minimize f_k, then we used the following defaults. If $n > p$ we used the starting value recommended by Jöreskog (1967)

$$\mathbf{\Psi}_0 = \left(1 - \tfrac{k}{2p}\right)\left\{\operatorname{diag}\left(\mathbf{S}_n^{-1}\right)\right\}^{-1}$$

where diag($\mathbf{A}$) is a diagonal matrix containing the diagonal of $\mathbf{A}$. If $n \leq p$ we used

$$\boldsymbol{\Psi}_0 = \text{diag}(\mathbf{S}_n)/(k+1).$$

C.4.1 Gradient of f_k with respect to diagonal elements of $\boldsymbol{\Psi}$

The Quasi-Newton line search algorithm works best when an analytical gradient (vector of partial derivatives of f_k with respect to ψ_j, $j = 1, \ldots, p$) is supplied. Following the argument in Jöreskog (1967), the gradient of f_k at a given value of $\boldsymbol{\Psi}$ is equal to $\partial F_k / \partial \boldsymbol{\Psi}$ evaluated at a point where F_k is minimum with respect to $\boldsymbol{\Lambda}_k$ (i.e., evaluated at $\widetilde{\boldsymbol{\Lambda}}_k$ defined in (C.6)). In equation (C.1) we derived

$$\frac{\partial F_k}{\partial \boldsymbol{\Psi}} = \text{diag}\left[\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{\Sigma} - \mathbf{S}_n\right)\boldsymbol{\Sigma}^{-1}\right].$$

From Section C.2 we have the identities

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1}\boldsymbol{\Lambda}_k^T\boldsymbol{\Psi}^{-1}, \quad \text{and}$$

$$(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n)\boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{\Lambda}}_k = 0$$

so that substituting $\widetilde{\boldsymbol{\Sigma}} = \widetilde{\boldsymbol{\Lambda}}_k\widetilde{\boldsymbol{\Lambda}}_k^T + \boldsymbol{\Psi}$ where $\widetilde{\boldsymbol{\Lambda}}_k$ is defined by (C.6), gives

$$\begin{aligned}
\widetilde{\boldsymbol{\Sigma}}^{-1}(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n) &= \left\{\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{\Lambda}}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1}\widetilde{\boldsymbol{\Lambda}}_k^T\boldsymbol{\Psi}^{-1}\right\}(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n) \\
&= \boldsymbol{\Psi}^{-1}(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n) + \mathbf{0}, \quad \text{and} \\
\boldsymbol{\Psi}^{-1}(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n)\widetilde{\boldsymbol{\Sigma}}^{-1} &= \boldsymbol{\Psi}^{-1}(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n)\left\{\boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\widetilde{\boldsymbol{\Lambda}}_k(\mathbf{I}_k + \boldsymbol{\Delta})^{-1}\widetilde{\boldsymbol{\Lambda}}_k^T\boldsymbol{\Psi}^{-1}\right\} \\
&= \boldsymbol{\Psi}^{-1}(\widetilde{\boldsymbol{\Sigma}} - \mathbf{S}_n)(\boldsymbol{\Psi}^{-1} + \mathbf{0}).
\end{aligned}$$

Thus, when $\partial F_k/\partial \mathbf{\Psi}$ is evaluated at $\widetilde{\mathbf{\Lambda}}_k$ we get

$$\frac{\partial f_k}{\partial \mathbf{\Psi}} = \text{diag}\left[\mathbf{\Psi}^{-1}(\widetilde{\mathbf{\Sigma}} - \mathbf{S}_n)\mathbf{\Psi}^{-1}\right] = \text{diag}\left[\mathbf{\Psi}^{-1}\left(\widetilde{\mathbf{\Lambda}}_k\widetilde{\mathbf{\Lambda}}_k^T + \mathbf{\Psi} - \mathbf{S}_n\right)\mathbf{\Psi}^{-1}\right].$$

Recall from (C.6) that

$$\mathbf{\Psi}^{-1/2}\widetilde{\mathbf{\Lambda}}_k = \mathbf{\Omega}(\mathbf{\Theta} - \mathbf{I}_k)^{1/2}, \quad \text{so that}$$

$$\mathbf{\Psi}^{-1}\widetilde{\mathbf{\Lambda}}_k\widetilde{\mathbf{\Lambda}}_k^T\mathbf{\Psi}^{-1} = \mathbf{\Psi}^{-1/2}\mathbf{\Omega}(\mathbf{\Theta} - \mathbf{I}_k)\mathbf{\Omega}^T\mathbf{\Psi}^{-1/2}.$$

The j-th diagonal element of this matrix is

$$\sum_{m=1}^{k}(\theta_m - 1)w_{jm}^2/\psi_j.$$

Thus the partial derivative of f_k with respect to ψ_j can be written as

$$\frac{\partial f_k}{\partial \psi_j} = \frac{1}{\psi_j}\left\{\sum_{m=1}^{k}(\theta_m - 1)\omega_{jm}^2 + 1 - s_{jj}/\psi_j\right\}, \quad \text{for} \quad j = 1, \ldots, p$$

where ω_{jm} is the j-th element in the m-th eigenvector $\boldsymbol{\omega}_m$. Because of the reparameterization $\psi_j = \mathrm{e}^{\alpha_j}$ we actually use the partial derivatives with respect to α_j,

$$\frac{\partial f_k}{\partial \alpha_j} = \frac{\partial f_k}{\partial \psi_j}\cdot\frac{\partial \psi_j}{\partial \alpha_j} = \frac{\partial f_k}{\partial \psi_j}\cdot \mathrm{e}^{\alpha_j} = \sum_{m=1}^{k}(\theta_m - 1)\omega_{jm}^2 + 1 - s_{jj}/\psi_j$$

since $\mathrm{e}^{\alpha_j} = \psi_j$.

C.4.2 Computing the eigenvalues of $\mathbf{S}^* = \mathbf{\Psi}^{-1/2}\mathbf{S}_n\mathbf{\Psi}^{-1/2}$.

The sample covariance matrix is $\mathbf{S}_n = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T$. Let $\mathbf{Y}^*$ be the $n \times p$ centered data matrix with i-th row $(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^T$. Then we can write $\mathbf{S}_n = \frac{1}{n}\mathbf{Y}^{*T}\mathbf{Y}^*$ and

$$\mathbf{S}^* = \mathbf{\Psi}^{-1/2}\mathbf{S}_n\mathbf{\Psi}^{-1/2} = \tfrac{1}{n}\mathbf{\Psi}^{-1/2}\mathbf{Y}^{*T}\mathbf{Y}^*\mathbf{\Psi}^{-1/2}.$$

An efficient way to find eigenvalues and the corresponding eigenvectors of $\mathbf{S}^*$ is to use the singular value decomposition. This method of calculation also ensures that the computed eigenvalues of $\mathbf{S}^*$ will be real and non-negative — as they should be since $\mathbf{S}^*$ is symmetric and non-negative definite. Let the $p \times n$ matrix $\mathbf{S}^{*1/2} = n^{-1/2}\boldsymbol{\Psi}^{-1/2}\mathbf{Y}^{*T}$ have the singular value decomposition

$$\mathbf{S}^{*1/2} = \mathbf{U}\mathbf{D}\mathbf{V}^T$$

where $\mathbf{D}$ is diagonal with dimension $r = \min(n, p)$, $\mathbf{U}$ is a $p \times r$ matrix with $\mathbf{U}^T\mathbf{U} = \mathbf{I}_r$, and $\mathbf{V}$ is a $n \times r$ matrix with $\mathbf{V}^T\mathbf{V} = \mathbf{I}_r$. Since

$$\mathbf{S}^* = \mathbf{S}^{*1/2}(\mathbf{S}^{*1/2})^T = \mathbf{U}\mathbf{D}^2\mathbf{U}^T$$

the squares of the singular values are the eigenvalues of $\mathbf{S}^*$ (i.e., $\theta_m = d_m^2$ for $m = 1, \ldots, r$), and the columns of $\mathbf{U}$ are the normalized eigenvectors of $\mathbf{S}^*$. The singular value decomposition can be computed in MATLAB using the function `svd(·,'econ')`.

C.5 Behavior of the likelihood as $\boldsymbol{\Psi}$ decreases or k increases

The loglikelihood for the k factor model equals

$$\log L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = c - \tfrac{n}{2}\left\{\log|\boldsymbol{\Sigma}| + \operatorname{trace}(\mathbf{S}_n\boldsymbol{\Sigma}^{-1}) + (\bar{\boldsymbol{y}} - \boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\bar{\boldsymbol{y}} - \boldsymbol{\mu})\right\}$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k\boldsymbol{\Lambda}_k^T + \boldsymbol{\Psi}$. When $\widehat{\boldsymbol{\mu}} = \bar{\boldsymbol{y}}$ the loglikelihood equals a constant minus the objective function F_k defined in (2.22). In Section C.2 we showed that for fixed $\boldsymbol{\Psi}$, F_k takes a minimum value of

$$f_k(\boldsymbol{\Psi}) = k + \sum_{m=1}^{p}\log\psi_m + \sum_{m=1}^{k}\log\theta_m + \sum_{m=k+1}^{p}\theta_m.$$

Thus one way to explore the behavior of the likelihood is to examine the behavior of f_k as the elements of $\boldsymbol{\Psi}$ get small, or as the number of factors increases. These limits are of interest

because the maximum likelihood estimation may fail in these cases, and it is of interest to understand what these failures imply about the behavior of the likelihood in these regions.

To illustrate possible behavior of the likelihood as $\mathbf{\Psi}$ gets small, suppose that $\mathbf{\Psi}_\delta = \delta \mathbf{I}_p$. Then $\mathbf{S}^* = \delta^{-1}\mathbf{S}_n$. Suppose that $d_1 \geq \ldots \geq d_p$ are the p eigenvalues of $\mathbf{S}_n$ (some of these may be zero if $n \leq p$). Then the eigenvalues of $\mathbf{S}^*$ are $\theta_m = \delta^{-1} d_m$ and the objective function f_k becomes

$$\begin{aligned} f_k(\mathbf{\Psi}_\delta) &= k + p\log\delta + \sum_{m=1}^{k} \log(\delta^{-1} d_m) + \sum_{m=k+1}^{p} \delta^{-1} d_m \\ &= k + (p-k)\log\delta + \sum_{m=1}^{k} \log d_m + \delta^{-1} \sum_{m=k+1}^{p} d_m. \end{aligned}$$

Since $\log z + \frac{1}{z} \to \infty$ as $z \to 0$, $f_k(\mathbf{\Psi}_\delta) \to \infty$ as $\delta \to 0$. Thus as $\mathbf{\Psi}_\delta$ gets smaller, the objective function f_k increases. Hence the likelihood goes to zero as $\mathbf{\Psi}_\delta$ approaches zero.

Now consider what happens to f_k if $\mathbf{\Psi}$ is fixed and small, and the number of factors k increases. First recall that $\theta > \log\theta$ for any $\theta > 0$ and that the function $\theta - \log\theta$ has a minimum value of 1 at $\theta = 1$. Thus $\log\theta - \theta \leq -1$ for any $\theta > 0$. Consider the difference between f_{k+1} and f_k,

$$\begin{aligned} f_{k+1}(\mathbf{\Psi}) - f_k(\mathbf{\Psi}) &= (k+1) - k + \left(\sum_{m=1}^{k+1} \log\theta_m - \sum_{m=1}^{k} \log\theta_m \right) + \left(\sum_{m=k+2}^{p} \theta_m - \sum_{m=k+1}^{p} \theta_m \right) \\ &= 1 + \log\theta_{k+1} - \theta_{k+1} \\ &\leq 1 - 1 = 0. \end{aligned}$$

Thus $f_{k+1}(\mathbf{\Psi}) \leq f_k(\mathbf{\Psi})$ for fixed $\mathbf{\Psi}$, so increasing k decreases the objective function f_k and increases the likelihood. It is reasonable to expect the likelihood to increase with increasing k. Increasing k gives a sequence of nested models with increasing number of parameters. A model with a larger number of parameters must fit better and hence have larger likelihood. Although there are some constraints on k to ensure identifiability, what this result implies is that as k

increases toward or across that identifiability boundary, the likelihood increases. In particular, if $n \leq p$ then some of the eigenvalues of $\mathbf{S}^*$ will be zero. In this case, if $k \geq n$, then $\log \theta_k = -\infty$ so that $f_k = -\infty$ (infinite likelihood).

C.6 Maximum likelihood estimation of factor models with restricted $\boldsymbol{\Psi}$

In this section we consider two restricted cases of the factor model (2.4). The first case assumes that $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$, a constant diagonal matrix. The second case assumes that each element of $\boldsymbol{\Psi}$ is a power of the the mean μ_j at that time point, $\psi_j = \sigma^2 \mu_j^{2\gamma}$.

C.6.1 Case 1: $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$

When $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$, we have

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \sigma^2 \mathbf{I}_p$$

with the constraint that $\sigma^{-2} \boldsymbol{\Lambda}_k^T \boldsymbol{\Lambda}_k = \boldsymbol{\Delta}$ is diagonal. When $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$,

$$\mathbf{S}^* = \boldsymbol{\Psi}^{-1/2} \mathbf{S}_n \boldsymbol{\Psi}^{-1/2} = \sigma^{-2} \mathbf{S}_n$$

and

$$\boldsymbol{\Delta} = \boldsymbol{\Lambda}_k^T \boldsymbol{\Psi}^{-1} \boldsymbol{\Lambda}_k = \sigma^{-2} \boldsymbol{\Lambda}_k^T \boldsymbol{\Lambda}_k \quad \text{diagonal}$$

so that in this case the columns of $\boldsymbol{\Lambda}_k$ are constrained to be orthogonal. For fixed σ^2, substituting $\boldsymbol{\Psi} = \sigma^2 \mathbf{I}_p$ into (C.5) gives the eigenvalue-vector equation

$$(\sigma^{-2} \mathbf{S}_n)(\sigma^{-1} \boldsymbol{\Lambda}_k) = (\sigma^{-1} \boldsymbol{\Lambda}_k)(\mathbf{I}_k + \sigma^{-2} \boldsymbol{\Lambda}_k^T \boldsymbol{\Lambda}_k).$$

Appendix C. Details about maximum likelihood factor analysis

Let $\mathbf{\Theta} = \text{diag}(\theta_1, \ldots, \theta_k)$ be a $k \times k$ diagonal matrix having the k largest eigenvalues of $\sigma^{-2}\mathbf{S}_n$ on the diagonal, and let $\mathbf{\Omega}$ be the corresponding normalized eigenvectors. Then for fixed σ^2, the maximum likelihood estimator of the loadings matrix $\mathbf{\Lambda}_k$ is

$$\widetilde{\mathbf{\Lambda}}_k = \sigma\mathbf{\Omega}(\mathbf{\Theta} - \mathbf{I}_k)^{1/2}.$$

The factor loadings are thus just scaled versions of the eigenvectors of $\mathbf{S}_n$ that correspond to the largest k eigenvectors. The factor loadings are scaled versions of the principal component loadings.

Substituting $\mathbf{\Lambda}_k = \widetilde{\mathbf{\Lambda}}_k(\sigma^2)$ into F_k gives an objective function in σ^2,

$$f_k(\sigma^2) = F_k(\widetilde{\mathbf{\Lambda}}_k(\sigma^2), \sigma^2) = \log|\widetilde{\mathbf{\Sigma}}| + \text{trace}(\mathbf{S}_n\widetilde{\mathbf{\Sigma}}^{-1})$$

where $\widetilde{\mathbf{\Sigma}} = \widetilde{\mathbf{\Lambda}}_k\widetilde{\mathbf{\Lambda}}_k + \sigma^2\mathbf{I}$. Since $\mathbf{S}^* = \sigma^{-2}\mathbf{S}_n$, the m-th eigenvalue of $\mathbf{S}^*$ is $\theta_m = \sigma^{-2}d_m$, where d_m is the m-th largest eigenvalue of the sample covariance matrix $\mathbf{S}_n$. Using the identities derived in Section C.2, we see that

$$|\widetilde{\mathbf{\Sigma}}| = \sigma^{2p}\theta_1 \ldots \theta_k = (\sigma^2)^{(p-k)}d_1 \ldots d_k, \quad \text{and}$$

$$\text{trace}(\mathbf{S}_n\widetilde{\mathbf{\Sigma}}^{-1}) = \text{trace}(\mathbf{S}^*) - \text{trace}(\mathbf{\Theta} - \mathbf{I}_k) = k + \sigma^{-2}\sum_{m=k+1}^{p} d_m.$$

Therefore the objective function to be minimized over σ^2 is

$$f_k(\sigma^2) = k + \sum_{m=1}^{k}\log d_m + (p-k)\log\sigma^2 + \sigma^{-2}\sum_{m=k+1}^{p} d_m$$

where the eigenvalues d_m are free of σ^2 for $m = 1, \ldots, p$.

Taking the partial derivative of f_k with respect to σ^2 and setting it equal to 0 gives the

equation

$$0 = \partial f_k/\partial\sigma^2 = (p-k)\sigma^{-2} - \sigma^{-4} \sum_{m=k+1}^{p} d_m$$

so that the MLE of σ^2 is

$$\widehat{\sigma}_k^2 = \sum_{m=k+1}^{p} d_m/(p-k).$$

The MLE of $\mathbf{\Lambda}_k$ for the k-factor model is

$$\widehat{\mathbf{\Lambda}}_k = \widehat{\sigma}_k \mathbf{\Omega}(\widehat{\mathbf{\Theta}} - \mathbf{I}_k)^{1/2}$$

where $\widehat{\mathbf{\Theta}} = \widehat{\sigma}_k^{-2}\text{diag}(d_1, \ldots, d_k)$, $d_1 \geq \ldots \geq d_k$ are the k largest eigenvalues of $\mathbf{S}_n$, and the columns of $\mathbf{\Omega}$ are the corresponding normalized eigenvectors.

The objective function F_k is minimized at $(\widehat{\mathbf{\Lambda}}_k, \widehat{\sigma}_k^2)$. Let the minimized value be $\widehat{F}_k$, which equals the minimized value of f_k, $f_k(\widehat{\sigma}_k^2)$. Thus

$$\widehat{F}_k = f_k(\widehat{\sigma}_k^2) = p + (p-k)\log\widehat{\sigma}_k^2 + \sum_{m=1}^{k} \log d_m.$$

Note that if $k = 0$, then the maximum likelihood estimator of σ^2 minimizes

$$F_0(\sigma^2) = p\log\sigma^2 + \text{trace}(\mathbf{S}_n)/\sigma^2$$

so that $\widehat{\sigma}_0^2 = \text{trace}(\mathbf{S}_n)/p$ and $f_0(\widehat{\sigma}_0^2) = p\log\widehat{\sigma}_0^2 + p$.

Assuming that the true number of factors is at least k, we can use a likelihood ratio test to compare the fit of the k-factor model with $\mathbf{\Psi} = \sigma^2\mathbf{I}$ to the fit of the k-factor model with unrestricted positive definite diagonal $\mathbf{\Psi}$. We would expect the distribution of this test to have an asymptotic χ^2 distribution with $p-1$ degrees of freedom under the model with $\mathbf{\Psi} = \sigma^2\mathbf{I}$. We might also compare the fit of the k-factor model with $\mathbf{\Psi} = \sigma^2\mathbf{I}$ to the fit of the k-factor model with $\mathbf{\Psi} = \sigma^2\text{diag}(\boldsymbol{\mu}^{2\gamma})$ (fitting this model is discussed in the next section). We expect

the (pseudo)likelihood ratio test for comparing these two models would have an asymptotic χ^2 distribution with 1 degree of freedom when the true model has $\mathbf{\Psi} = \sigma^2\mathbf{I}$. These likelihood ratio test statistics are given in equation (2.30).

C.6.2 Case 2: $\psi_j = \sigma^2\mu_j^{2\gamma}$

Assume that the diagonal elements of $\mathbf{\Psi}$ are a function of the mean vector $\boldsymbol{\mu}$, $\psi_j = \sigma^2\mu_j^{2\gamma}$. Then the k-factor model for the covariance is

$$\mathbf{\Sigma} = \mathbf{\Lambda}_k\mathbf{\Lambda}_k^T + \sigma^2\text{diag}(\boldsymbol{\mu}^{2\gamma})$$

where the exponent 2γ is applied elementwise to $\boldsymbol{\mu}$ and $\text{diag}(\boldsymbol{v})$ is a diagonal matrix with $\boldsymbol{v}$ on the diagonal. The usual identifiability constraint becomes $\sigma^{-2}\mathbf{\Lambda}_k^T\text{diag}(\boldsymbol{\mu}^{-2\gamma})\mathbf{\Lambda}_k = \mathbf{\Delta}$ diagonal. The number of parameters in the covariance model with k factors is

$$2 + pk - k(k-1)/2.$$

The difference between the number of parameters in the unstructured $\mathbf{\Sigma}$ and the model above is

$$\begin{aligned} s &= \{p(p+1) - 4 - 2pk + k(k-1)\}/2 \\ &= \{p^2 - 2pk + k^2 + p - k - 4\}/2 \\ &= \{(p-k)^2 + (p-k-4)\}/2 \end{aligned}$$

and for identifiability we require $s > 0$. When $n \leq p$ we also need to check that the number of parameters (also counting the p parameters in $\boldsymbol{\mu}$) is no larger than the total number of observations. We require

$$p + 2 + pk - k(k-1)/2 \leq np \iff k + 1 + 2/p - k(k-1)/2p \leq n$$

which will be satisfied if

$$k \leq n - 1 - 2/p.$$

Although the covariance now contains some information about the mean vector $\boldsymbol{\mu}$, our approach will be to assume that $\boldsymbol{\mu}$ is a nuisance parameter that can be consistently estimated by $\bar{\boldsymbol{y}}$. To estimate $\boldsymbol{\Lambda}, \sigma^2$ and γ we maximize the pseudolikelihood function

$$L_{ps}(\boldsymbol{\Sigma}; \boldsymbol{\mu} = \bar{\boldsymbol{y}}) = (2\pi)^{-np/2} |\boldsymbol{\Sigma}_{ps}|^{-n/2} \exp\left\{-\tfrac{n}{2} \text{trace}(\mathbf{S}_n \boldsymbol{\Sigma}_{ps}^{-1})\right\}$$

where $\boldsymbol{\Sigma}_{ps} = \boldsymbol{\Lambda}_k \boldsymbol{\Lambda}_k^T + \sigma^2 \text{diag}(\bar{\boldsymbol{y}}^{2\gamma})$ and $\sigma^{-2} \boldsymbol{\Lambda}_k^T \text{diag}(\bar{\boldsymbol{y}}^{-2\gamma}) \boldsymbol{\Lambda}_k$ is constrained to be diagonal. To compute the pseudolikelihood estimator we will minimize the objective function

$$F_k^{ps}(\boldsymbol{\Lambda}_k, \sigma^2, \gamma) = \log |\boldsymbol{\Sigma}_{ps}| + \text{trace}(\mathbf{S}_n \boldsymbol{\Sigma}_{ps}^{-1})$$

subject to the identifiability constraint.

When $k = 0$, the pseudolikelihood estimators of σ^2 and γ minimize

$$F_0^{ps}(\sigma^2, \gamma) = p \log \sigma^2 + 2\gamma \sum_{i=1}^{p} \log \bar{y}_i + \sigma^{-2} \sum_{j=1}^{p} s_{jj} \bar{y}_i^{-2\gamma}.$$

Taking partial derivatives and setting equal to zero gives the equations

$$\begin{aligned} 0 &= p\sigma^{-2} - \sigma^{-4} \sum_{j=1}^{p} s_{jj} \bar{y}_j^{-2\gamma} \quad \Rightarrow \quad \sigma^2 = \sum_{j=1}^{p} s_{jj} \bar{y}_j^{-2\gamma}/p \\ 0 &= 2 \sum_{j=1}^{p} \log \bar{y}_j - 2\sigma^{-2} \sum_{j=1}^{p} s_{jj} \bar{y}_j^{-2\gamma} \log \bar{y}_j \end{aligned}$$

Substituting the first equation into the second gives the following equation in γ,

$$0 = \sum_{j=1}^{p} \log \bar{y}_j - p \left(\sum_{j=1}^{p} s_{jj} \bar{y}_j^{-2\gamma} \right)^{-1} \left(\sum_{j=1}^{p} s_{jj} \bar{y}_j^{-2\gamma} \log \bar{y}_j \right). \tag{C.10}$$

So when $k = 0$, we solve equation (C.10) for γ and then substitute $\widehat{\gamma}$ in

$$\sigma^2 = \sum_{j=1}^{p} s_{jj}\bar{y}_j^{-2\gamma}/p$$

to get $\widehat{\sigma}_0^{2,ps}$. The minimum of F_0^{ps} is

$$p\log\widehat{\sigma}_0^{2,ps} + 2\widehat{\gamma}_0^{ps}\sum_{j=1}^{p}\log\bar{y}_j + p.$$

When $k \geq 1$ the strategy to minimize F_k^{ps} mimics that in Section C.2. For fixed σ^2 and γ, the pseudolikelihood estimator of $\mathbf{\Lambda}_k$ solves the usual eigenvalue-vector equation

$$\mathbf{S}_{ps}^{*}(\mathbf{\Psi}_{ps}^{-1}\mathbf{\Lambda}_k) = (\mathbf{\Psi}_{ps}^{-1}\mathbf{\Lambda}_k)(\mathbf{I}_k + \mathbf{\Delta}_{ps}) \tag{C.11}$$

where $\mathbf{\Psi}_{ps} = \sigma^2\text{diag}(\bar{\boldsymbol{y}}^{-2\gamma})$,

$$\mathbf{S}_{ps}^{*} = \mathbf{\Psi}_{ps}^{-1/2}\mathbf{S}_n\mathbf{\Psi}_{ps}^{-1/2} = \sigma^{-2}\,\text{diag}(\bar{\boldsymbol{y}}^{-\gamma})\mathbf{S}_n\,\text{diag}(\bar{\boldsymbol{y}}^{-\gamma})$$

and $\mathbf{\Delta}_{ps} = \mathbf{\Lambda}_k^T\mathbf{\Psi}_{ps}^{-1}\mathbf{\Lambda}_k$. For fixed k, σ^2 and γ, the pseudolikelihood estimator of the loadings matrix $\mathbf{\Lambda}_k$ is

$$\widetilde{\mathbf{\Lambda}}_k^{ps}(\sigma^2,\gamma) = \mathbf{\Psi}_{ps}^{1/2}\mathbf{\Omega}_{ps}(\mathbf{\Theta}_{ps} - \mathbf{I}_k)^{1/2}$$

where $\mathbf{\Theta}_{ps} = \text{diag}(\theta_1,\ldots,\theta_k)$ contains the k largest eigenvalues of $\mathbf{S}_{ps}^{*}$ on the diagonal, and $\mathbf{\Omega}_{ps}$ contains the corresponding normalized eigenvectors.

Substituting $\mathbf{\Lambda} = \widetilde{\mathbf{\Lambda}}_k^{ps}(\sigma^2,\gamma)$ gives the objective function in σ^2 and γ,

$$f_k^{ps}(\sigma^2,\gamma) = F_k(\widetilde{\mathbf{\Lambda}}_k^{ps}(\sigma^2,\gamma),\sigma^2,\gamma) = \log|\widetilde{\mathbf{\Sigma}}_{ps}| + \text{trace}(\mathbf{S}_n\widetilde{\mathbf{\Sigma}}_{ps}^{-1})$$

where $\widetilde{\mathbf{\Sigma}}_{ps} = \widetilde{\mathbf{\Lambda}}_{ps}\widetilde{\mathbf{\Lambda}}_{ps}^T + \mathbf{\Psi}_{ps}$. Let $\mathbf{S}_{ps}^{**} = \text{diag}(\bar{\boldsymbol{y}}^{-\gamma})\mathbf{S}_n\text{diag}(\bar{\boldsymbol{y}}^{-\gamma})$, and let $\theta_1^{**} \geq \ldots \geq \theta_p^{**}$ be the ordered eigenvalues of $\mathbf{S}_{ps}^{**}$. These eigenvalues are functions of γ, but are free of σ^2. Then since

$\mathbf{S}_{ps}^{*} = \sigma^{-2}\mathbf{S}_{ps}^{**}$, the eigenvalues of $\mathbf{S}_{ps}^{*}$ are $\theta_m = \sigma^{-2}\theta_m^{**}$ and $\mathbf{\Theta} = \sigma^{-2}\mathbf{\Theta}^{**}$. Using the identities from Section C.2, we get

$$\begin{aligned} |\widetilde{\mathbf{\Sigma}}_{ps}| &= |\mathbf{\Psi}_{ps}|\theta_1 \ldots \theta_k \\ &= |\sigma^2 \mathrm{diag}(\bar{\boldsymbol{y}}^{2\gamma})| \times (\sigma^{-2})^k \theta_1^{**} \ldots \theta_k^{**} \\ &= (\sigma^2)^{p-k} \prod_{i=1}^{p} \bar{y}_i^{2\gamma} \prod_{m=1}^{k} \theta_m^{**}, \quad \text{and} \end{aligned}$$

$$\begin{aligned} \mathrm{trace}(\mathbf{S}_n \widetilde{\mathbf{\Sigma}}_{ps}^{-1}) &= \mathrm{trace}(\mathbf{S}_{ps}^{*}) - \mathrm{trace}(\mathbf{\Theta}_{ps} - \mathbf{I}_k) \\ &= k + \sigma^{-2}\mathrm{trace}(\mathbf{S}_{ps}^{**}) - \sigma^{-2}\mathrm{trace}(\mathbf{\Theta}_{ps}^{**}) \\ &= k + \sigma^{-2} \sum_{m=k+1}^{p} \theta_m^{**}. \end{aligned}$$

Substituting into the objective function f_k^{ps} gives

$$f_k^{ps}(\sigma^2, \gamma) = k + \sigma^{-2} \sum_{m=k+1}^{p} \theta_m^{**} + (p-k)\log\sigma^2 + 2\gamma \sum_{i=1}^{p} \log \bar{y}_i + \sum_{m=1}^{k} \log \theta_m^{**}. \tag{C.12}$$

The eigenvalues θ_m^{**} are functions of γ but are free of σ^2. Holding γ fixed, and differentiating (C.12) with respect to σ^2 and setting equal to zero, gives the pseudolikelihood estimator of σ^2 as a function of γ,

$$\widetilde{\sigma}_k^{2,ps} = \sum_{m=k+1}^{p} \theta_m^{**}/(p-k). \tag{C.13}$$

Substituting $\widetilde{\sigma}_k^{2,ps}$ into the objective function (C.12) gives an objective function in γ alone,

$$\phi_k^{ps}(\gamma) = f_k(\widetilde{\sigma}_k^{2,ps}(\gamma), \gamma) = p + (p-k)\log\widetilde{\sigma}_k^{2,ps} + 2\gamma \sum_{i=1}^{p} \log \bar{y}_i + \sum_{m=1}^{k} \log \theta_m^{**}. \tag{C.14}$$

To find the pseudolikelihood estimator of γ, we minimize ϕ_k^{ps} over γ using a quasi-Newton line search algorithm such as MATLAB's `fminunc` to give $\widehat{\gamma}_{ps}$. If the user did not supply a starting

guess for γ we used $\gamma = 1$. Substituting the resulting solution for γ into (C.13) gives the pseudolikelihood estimator

$$\widehat{\sigma}_k^{2,ps} = \sum_{m=k+1}^{p} \theta_m^{**}/(p-k).$$

Substituting for σ^2 and γ in (C.11) gives the pseudolikelihood estimator of $\mathbf{\Lambda}$,

$$\widehat{\mathbf{\Lambda}}_k^{ps} = \widehat{\mathbf{\Psi}}_{ps}^{1/2}\widehat{\mathbf{\Omega}}_{ps}(\widehat{\mathbf{\Theta}}_{ps} - \mathbf{I}_k)^{1/2}$$

where $\widehat{\mathbf{\Psi}}_{ps} = \widehat{\sigma}_k^{ps,2}\text{diag}(\bar{\boldsymbol{y}}^{2\widehat{\gamma}_{ps}})$, $\widehat{\mathbf{\Theta}}_{ps}$ contains the k largest eigenvalues of $\widehat{\mathbf{\Psi}}_{ps}^{-1}\mathbf{S}_n\widehat{\mathbf{\Psi}}_{ps}^{-1}$, and $\widehat{\mathbf{\Omega}}_{ps}$ contains the corresponding normalized eigenvectors.

Following a similar argument to the one in Section C.4.1, we see that $\partial\phi_k^{ps}/\partial\gamma$ at a given value of γ is equal to $\partial F_k^{ps}/\partial\gamma$, evaluated at the point where F_k^{ps} is minimum with respect to $\mathbf{\Lambda}_k$ and σ^2 for a given γ. For $\mathbf{\Psi}_{ps} = \sigma^2\text{diag}(\bar{\boldsymbol{y}}^{2\gamma})$ we have

$$\begin{aligned}
\partial\mathbf{\Sigma}_{ps}/\partial\gamma &= \partial\mathbf{\Psi}_{ps}/\partial\gamma \\
&= \sigma^2\,\text{diag}(2\bar{y}_1^{2\gamma}\log\bar{y}_1, \ldots, 2\bar{y}_p^{2\gamma}\log\bar{y}_p) \\
&= 2\sigma^2\,\text{diag}(\bar{y}_1^{2\gamma}, \ldots, \bar{y}_p^{2\gamma})\,\text{diag}(\log\bar{y}_1, \ldots, \log\bar{y}_p) \\
&= 2\mathbf{\Psi}_{ps}\,\text{diag}(\log\bar{y}_1, \ldots, \log\bar{y}_p).
\end{aligned}$$

Since

$$\partial\log|\mathbf{\Sigma}_{ps}|/\partial\gamma = \text{trace}\left\{\mathbf{\Sigma}_{ps}^{-1}\left(\partial\mathbf{\Sigma}_{ps}/\partial\gamma\right)\right\}, \quad \text{and}$$

$$\partial\text{trace}(\mathbf{S}_n\mathbf{\Sigma}_{ps}^{-1})/\partial\gamma = \text{trace}\left\{\mathbf{S}_n\left(\partial\mathbf{\Sigma}_{ps}^{-1}/\partial\gamma\right)\right\} = -\text{trace}\left\{\mathbf{S}_n\mathbf{\Sigma}_{ps}^{-1}\left(\partial\mathbf{\Sigma}_{ps}/\partial\gamma\right)\mathbf{\Sigma}_{ps}^{-1}\right\}$$

we have

$$\begin{aligned}\partial F_k^{ps}/\partial\gamma &= \partial\left\{\log|\boldsymbol{\Sigma}_{ps}| + \operatorname{trace}(\mathbf{S}_n\boldsymbol{\Sigma}_{ps}^{-1})\right\}/\partial\gamma \\ &= \operatorname{trace}\left\{\boldsymbol{\Sigma}_{ps}^{-1}\left(\partial\boldsymbol{\Sigma}_{ps}/\partial\gamma\right)\right\} - \operatorname{trace}\left\{\mathbf{S}_n\boldsymbol{\Sigma}_{ps}^{-1}\left(\partial\boldsymbol{\Sigma}_{ps}/\partial\gamma\right)\boldsymbol{\Sigma}_{ps}^{-1}\right\} \\ &= \operatorname{trace}\left\{\boldsymbol{\Sigma}_{ps}^{-1}(\boldsymbol{\Sigma}_{ps}-\mathbf{S}_n)\boldsymbol{\Sigma}_{ps}^{-1}\left(\partial\boldsymbol{\Sigma}_{ps}/\partial\gamma\right)\right\}.\end{aligned}$$

As in Section C.4.1, for a given value of γ and σ^2, when $\widetilde{\boldsymbol{\Lambda}}_k^{ps}$ solves (C.11),

$$\widetilde{\boldsymbol{\Sigma}}_{ps}^{-1}(\widetilde{\boldsymbol{\Sigma}}_{ps}-\mathbf{S}_n)\widetilde{\boldsymbol{\Sigma}}_{ps}^{-1} = \boldsymbol{\Psi}_{ps}^{-1}(\widetilde{\boldsymbol{\Sigma}}_{ps}-\mathbf{S}_n)\boldsymbol{\Psi}_{ps}^{-1}$$

where $\widetilde{\boldsymbol{\Sigma}}_{ps} = \widetilde{\boldsymbol{\Lambda}}_k^{ps}(\widetilde{\boldsymbol{\Lambda}}_k^{ps})^T + \boldsymbol{\Psi}_{ps}$. The objective function F_k^{ps} evaluated at $\widetilde{\boldsymbol{\Lambda}}_k^{ps}$ is f_k^{ps}. Thus

$$\partial f_k^{ps}/\partial\gamma = \operatorname{trace}\left\{\boldsymbol{\Psi}_{ps}^{-1}(\widetilde{\boldsymbol{\Sigma}}_{ps}-\mathbf{S}_n)\boldsymbol{\Psi}_{ps}^{-1}\left(\partial\boldsymbol{\Sigma}_{ps}/\partial\gamma\right)\right\}.$$

Substituting for $\partial\boldsymbol{\Sigma}_{ps}/\partial\gamma$ and $\widetilde{\boldsymbol{\Sigma}}_{ps} = \widetilde{\boldsymbol{\Lambda}}_k^{ps}(\widetilde{\boldsymbol{\Lambda}}_k^{ps})^T + \boldsymbol{\Psi}_{ps}$ gives

$$\begin{aligned}\partial f_k^{ps}/\partial\gamma &= \operatorname{trace}\left[\boldsymbol{\Psi}_{ps}^{-1}(\widetilde{\boldsymbol{\Sigma}}_{ps}-\mathbf{S}_n)\boldsymbol{\Psi}_{ps}^{-1}\left\{2\boldsymbol{\Psi}_{ps}\operatorname{diag}(\log\bar{y}_1,\ldots,\log\bar{y}_p)\right\}\right] \\ &= 2\operatorname{trace}\left\{\boldsymbol{\Psi}_{ps}^{-1}(\widetilde{\boldsymbol{\Sigma}}_{ps}-\mathbf{S}_n)\operatorname{diag}(\log\bar{y}_1,\ldots,\log\bar{y}_p)\right\} \\ &= 2\operatorname{trace}\left[\left\{\boldsymbol{\Psi}_{ps}^{-1}\widetilde{\boldsymbol{\Lambda}}_k^{ps}(\widetilde{\boldsymbol{\Lambda}}_k^{ps})^T + \mathbf{I}_p - \boldsymbol{\Psi}_{ps}^{-1}\mathbf{S}_n\right\}\operatorname{diag}(\log\bar{y}_1,\ldots,\log\bar{y}_p)\right].\end{aligned}$$

Recall that $\widetilde{\boldsymbol{\Lambda}}_k^{ps} = \boldsymbol{\Psi}_{ps}^{1/2}\boldsymbol{\Omega}_{ps}(\boldsymbol{\Theta}_{ps}-\mathbf{I}_k)^{1/2}$ so that

$$\widetilde{\boldsymbol{\Lambda}}_k^{ps}(\widetilde{\boldsymbol{\Lambda}}_k^{ps})^T = \boldsymbol{\Psi}_{ps}^{1/2}\boldsymbol{\Omega}_{ps}(\boldsymbol{\Theta}_{ps}-\mathbf{I}_k)\boldsymbol{\Omega}_{ps}^T\boldsymbol{\Psi}_{ps}^{1/2}$$

where $\boldsymbol{\Theta}_{ps}$ contains the k-largest eigenvalues of $\mathbf{S}_{ps}^* = \boldsymbol{\Psi}_{ps}^{-1/2}\mathbf{S}_n\boldsymbol{\Psi}_{ps}^{-1/2}$ and $\boldsymbol{\Omega}_{ps}$ contains the respective eigenvectors. Therefore the j-th diagonal element of $\boldsymbol{\Psi}_{ps}^{-1}\widetilde{\boldsymbol{\Lambda}}_k^{ps}(\widetilde{\boldsymbol{\Lambda}}_k^{ps})^T$ is

$$\sum_{m=1}^{k}(\theta_m-1)\omega_{jm}^2.$$

Appendix C. Details about maximum likelihood factor analysis

The derivative of f_k^{ps} with respect to γ is

$$\partial f_k^{ps}/\partial\gamma = 2\sum_{j=1}^{p}\left\{\sum_{m=1}^{k}(\theta_m - 1)\omega_{jm}^2 + 1 - s_{jj}/\psi_j\right\}\log\bar{y}_i$$

where s_{jj} is the j-th diagonal element of $\mathbf{S}_n$ and $\psi_j = \sigma^2\bar{y}_j^{2\gamma}$. For a given value of γ, evaluating $\partial f_k^{ps}/\partial\gamma$ at $\widetilde{\sigma}_k^{2,ps}(\gamma)$ defined in (C.13) gives the gradient of the objective function ϕ_k^{ps} defined in (C.14),

$$\partial\phi_k^{ps}(\gamma)/\partial\gamma = \partial f_k^{ps}/\partial\gamma\,\Big|_{\sigma^2=\widetilde{\sigma}_k^{2,ps}}$$

Recall that $\theta_m = \sigma^{-2}\theta_m^{**}$ where θ_m^{**} is the m-th largest eigenvalue of $\mathbf{S}_{ps}^{**} = \mathrm{diag}(\bar{\boldsymbol{y}}^{-\gamma})\,\mathbf{S}_n\,\mathrm{diag}(\bar{\boldsymbol{y}}^{-\gamma})$.

Assuming that the true number of factors is at least k, we could use a (pseudo)likelihood ratio test to compare the k-factor model with $\boldsymbol{\Psi} = \sigma^2\mathrm{diag}(\boldsymbol{\mu}^{2\gamma})$ to a k-factor model with unrestricted positive definite diagonal $\boldsymbol{\Psi}$. We would expect the distribution of this test to have an asymptotic χ^2 distribution with $p-2$ degrees of freedom under the model with $\boldsymbol{\Psi} = \sigma^2\mathrm{diag}(\boldsymbol{\mu}^{2\gamma})$. The test statistic is given in equation (2.30).

Appendix D

Sensitivity equations for the PBPK model

In this appendix we derive the sensitivity equations used to numerically approximate the sensitivity functions (partial derivatives) for several parameters in the physiologically based pharmacokinetic (PBPK) model. We introduced the PBPK model in Section 1.2.2 of Chapter 1; it is described by the following system of differential equations,

$$\begin{aligned}
\dot{z}_h &= \texttt{RATS}\, Q_p \,(c_x - z_h/V_h) - K_L z_h \\
\dot{z}_f &= Q_f \,\{c_a - z_f/(V_f P_f)\} \\
\dot{z}_r &= Q_r \,\{c_a - z_r/(V_r P_r)\} \\
\dot{z}_m &= Q_m \,\{c_a - z_m/(V_m P_m)\} \\
\dot{z}_l &= Q_l \,\{c_a - z_l/(V_l P_l)\} - \frac{V_{max} z_l}{z_l + K_m V_l P_l}
\end{aligned} \tag{D.1}$$

Initial conditions at $t = 0$ are given by

$$z_h(0) = z_{h0}, \quad z_f(0) = 0, \quad z_r(0) = 0, \quad z_m(0) = 0, \quad z_l(0) = 0.$$

We described the meaning of the symbols and parameters previously in Section 1.2.2.

Table 1.1 of Chapter 1 gives values of the basic input parameters. Other quantities are computed from these. A single rat was housed in each chamber in the CCl_4 experiment so that $\texttt{RATS} = 1$. The initial amount in the chamber is $z_{h0} = \texttt{CONC} \times V_h \times W_m/24450$ (mg) where $V_h = V_{CHC} - \texttt{RATS} \times B_w$ (liters) is the net chamber volume. Compartment/blood partition coefficients (P_f, P_l, P_m, P_r) are computed from the compartment/air coefficients (P_{fa}, P_{la}, P_{ma}, P_{ra}) respectively, divided by the blood/air partition coefficient (P_b). Tissue volumes

(V_f, V_l, V_m) equal the respective volume percentages (V_{fc}, V_{lc}, V_{mc}) times body weight (B_w) (liters). Tissue blood flow rates (Q_f, Q_l, Q_m) equal the respective blood flow percentages (Q_{fc}, Q_{lc}, Q_{mc}) times cardiac output (Q_c) (liter/hr). The following difference equations define relationships between other quantities in the model,

$$Q_c = Q_f + Q_r + Q_m + Q_l, \quad V_f + V_r + V_m + V_l = 0.91 B_w$$
$$c_v = \frac{1}{Q_c} \sum_{*=f,r,m,l} \frac{Q_* z_*}{V_* P_*}, \quad c_a = \frac{Q_c c_v + Q_p z_h / V_h}{Q_c + Q_p / P_b}, \quad c_x = \frac{c_a}{P_b}.$$

In equation (1.12) of Chapter 1 we derived the basic form of the supplemental differential equation (the *sensitivity equation*) that must be solved simultaneously with the original system to provide a numerical approximation of the sensitivity function. In the sections below we derive the sensitivity equations for the following PBPK model parameters: body weight (B_w), loss rate from an empty chamber (K_L), the metabolic parameters (V_{max}) and (K_m), total cardiac output, (Q_c), and ventilation rate, (Q_p). The system of differential equations models the amount (mg) in the chamber, although the collected data measured concentration (ppm). For a fixed chamber volume, the sensitivity on the concentration scale is proportional to the sensitivity on the amount scale, where the constant of proportionality is free of time, t. We derive the sensitivities below on the amount (mg) scale.

D.1 Sensitivity equations for body weight, B_w

The volume parameters V_* are all functions of rat body weight, B_w,

$$\begin{aligned} V_h = V_{CHC} - \texttt{RATS} B_w &\Rightarrow \partial V_h/\partial B_w = -\texttt{RATS} \\ V_f = V_{fc} B_w &\Rightarrow \partial V_f/\partial B_w = V_{fc} \\ V_l = V_{lc} B_w &\Rightarrow \partial V_l/\partial B_w = V_{lc} \\ V_m = V_{mc} B_w &\Rightarrow \partial V_m/\partial B_w = V_{mc} \\ V_r = V_{rc} B_w &\Rightarrow \partial V_r/\partial B_w = V_{rc} \end{aligned}$$

where $V_{rc} = 0.91 - V_{fc} - V_{lc} - V_{mc}$.

To approximate the sensitivity, $z_{h[B_w]} = \partial z_h/\partial B_w$, for the body weight, B_w, of a rat, the following system of differential equations can be solved simultaneously with the original system in equation (D.1). In the equations below, the subscript $[B_w]$ indicates a partial derivative with respect to B_w.

$$\begin{aligned} \dot{z}_{h[B_w]} &= \texttt{RATS}\, Q_p \left(c_{x[B_w]} - z_{h[B_w]}/V_h - \texttt{RATS}\, z_h/V_h^2\right) - K_L z_{h[B_w]} \\ \dot{z}_{f[B_w]} &= Q_f \left\{c_{a[B_w]} - z_{f[B_w]}/(V_f P_f) + V_{fc} z_f/(V_f^2 P_f)\right\} \\ \dot{z}_{r[B_w]} &= Q_r \left\{c_{a[B_w]} - z_{r[B_w]}/(V_r P_r) + V_{rc} z_r/(V_r^2 P_r)\right\} \\ \dot{z}_{m[B_w]} &= Q_m \left\{c_{a[B_w]} - z_{m[B_w]}/(V_m P_m) + V_{mc} z_m/(V_m^2 P_m)\right\} \\ \dot{z}_{l[B_w]} &= Q_l \left\{c_{a[B_w]} - z_{l[B_w]}/(V_l P_l) + V_{lc} z_l/(V_l^2 P_l)\right\} - \frac{V_{max} z_{l[B_w]}}{z_l + K_m V_l P_l} + \frac{V_{max} z_l z_{l[B_w]}}{(z_l + K_m V_l P_l)^2} \\ &\quad + \frac{V_{max} K_m V_{lc} P_l z_l}{(z_l + K_m V_l P_l)^2} \end{aligned}$$

where

$$c_{v[B_w]} = Q_c^{-1} \sum_{*=f,r,m,l} \{Q_* z_{*[B_w]}/(V_* P_*) - Q_* V_{*c} z_*/(V_*^2 P_*)\},$$

$$c_{a[B_w]} = \{Q_c c_{v[B_w]} + Q_p z_{h[B_w]}/V_h + \texttt{RATS}\, Q_p z_h/V_h^2\}/(Q_c + Q_p/P_b),$$

and $c_{x[B_w]} = c_{a[B_w]}/P_b$. The initial condition $z_{h0} = \texttt{CONC}\, V_h W_m/24450$ is a function of B_w since $V_h = V_{CHC} - \texttt{RATS}\, B_w$ so that the initial conditions for the supplemental system are $z_{h[B_w]}(0) = -\texttt{CONC RATS}\, W_m/24450$ and

$$z_{f[B_w]}(0) = 0,\ z_{r[B_w]}(0) = 0,\ z_{m[B_w]}(0) = 0,\ z_{l[B_w]}(0) = 0.$$

D.2 Sensitivity equations for loss rate, K_L

The following system of differential equations can be solved simultaneously with the original ODE system to approximate the sensitivity, $z_{h[K_L]} = \partial z_h/\partial K_L$, for the parameter describing loss rate from the empty chamber, K_L. In the equations below, the subscript $[K_L]$ indicates a partial derivative with respect to K_L.

$$\begin{aligned}
\dot{z}_{h[K_L]} &= \texttt{RATS}\, Q_p \left(c_{x[K_L]} - z_{h[K_L]}/V_h\right) - K_L z_{h[K_L]} - z_h \\
\dot{z}_{f[K_L]} &= Q_f \left\{c_{a[K_L]} - z_{f[K_L]}/(V_f P_f)\right\} \\
\dot{z}_{r[K_L]} &= Q_r \left\{c_{a[K_L]} - z_{r[K_L]}/(V_r P_r)\right\} \\
\dot{z}_{m[K_L]} &= Q_m \left\{c_{a[K_L]} - z_{m[K_L]}/(V_m P_m)\right\} \\
\dot{z}_{l[K_L]} &= Q_l \left\{c_{a[K_L]} - z_{l[K_L]}/(V_l P_l)\right\} - \frac{V_{max} z_{l[K_L]}}{z_l + K_m V_l P_l} + \frac{V_{max} z_l z_{l[K_L]}}{(z_l + K_m V_l P_l)^2}
\end{aligned}$$

where

$$c_{v[K_L]} = Q_c^{-1} \sum_{*=f,r,m,l} Q_* z_{*[K_L]}/(V_* P_*),$$

$$c_{a[K_L]} = \{Q_c c_{v[K_L]} + Q_p z_{h[K_L]}/V_h\}/(Q_c + Q_p/P_b),$$

and $c_{x[K_L]} = c_{a[K_L]}/P_b$. Since the initial conditions of the original system were free of K_L, the initial conditions for the supplemental equations are

$$z_{h[K_L]}(0) = 0,\ z_{f[K_L]}(0) = 0,\ z_{r[K_L]}(0) = 0,\ z_{m[K_L]}(0) = 0,\ z_{l[K_L]}(0) = 0.$$

D.3 Sensitivity equations for V_{max}

The parameter V_{max} is the maximum rate of metabolism. The following system of differential equations can be solved simultaneously with the original ODE system to approximate the sensitivity, $z_{h[V_{max}]} = \partial z_h / \partial V_{max}$, for V_{max}. In the equations below, the subscript $[V_{max}]$ indicates a partial derivative with respect to V_{max}.

$$
\begin{aligned}
\dot{z}_{h[V_{max}]} &= \texttt{RATS}\, Q_p \left(c_{x[V_{max}]} - z_{h[V_{max}]}/V_h\right) - K_L z_{h[V_{max}]} \\
\dot{z}_{f[V_{max}]} &= Q_f \left\{c_{a[V_{max}]} - z_{f[V_{max}]}/(V_f P_f)\right\} \\
\dot{z}_{r[V_{max}]} &= Q_r \left\{c_{a[V_{max}]} - z_{r[V_{max}]}/(V_r P_r)\right\} \\
\dot{z}_{m[V_{max}]} &= Q_m \left\{c_{a[V_{max}]} - z_{m[V_{max}]}/(V_m P_m)\right\} \\
\dot{z}_{l[V_{max}]} &= Q_l \left\{c_{a[V_{max}]} - z_{l[V_{max}]}/(V_l P_l)\right\} - \frac{V_{max} z_{l[V_{max}]}}{z_l + K_m V_l P_l} + \frac{V_{max} z_l z_{l[V_{max}]}}{(z_l + K_m V_l P_l)^2} - \frac{z_l}{z_l + K_m V_l P_l}
\end{aligned}
$$

where

$$c_{v[V_{max}]} = Q_c^{-1} \sum_{*=f,r,m,l} Q_* z_{*[V_{max}]}/(V_* P_*),$$

$$c_{a[V_{max}]} = \{Q_c c_{v[V_{max}]} + Q_p z_{h[V_{max}]}/V_h\}/(Q_c + Q_p/P_b),$$

and $c_{x[V_{max}]} = c_{a[V_{max}]}/P_b$. Since the initial conditions of the original system were free of V_{max}, the initial conditions for the supplemental equations are

$$z_{h[V_{max}]}(0) = 0,\ z_{f[V_{max}]}(0) = 0,\ z_{r[V_{max}]}(0) = 0,\ z_{m[V_{max}]}(0) = 0,\ z_{l[V_{max}]}(0) = 0.$$

D.4 Sensitivity equations for Michaelis-Menten constant, K_m

To approximate the sensitivity, $z_{h[K_m]} = \partial z_h / \partial K_m$, for the Michaelis-Menten constant, K_m, the following system of differential equations can be solved simultaneously with the original system. In the equations below, the subscript $[K_m]$ indicates a partial derivative with respect to K_m.

$$\begin{aligned}
\dot{z}_{h[K_m]} &= \texttt{RATS}\, Q_p \left(c_{x[K_m]} - z_{h[K_m]}/V_h\right) - K_L z_{h[K_m]} \\
\dot{z}_{f[K_m]} &= Q_f \left\{c_{a[K_m]} - z_{f[K_m]}/(V_f P_f)\right\} \\
\dot{z}_{r[K_m]} &= Q_r \left\{c_{a[K_m]} - z_{r[K_m]}/(V_r P_r)\right\} \\
\dot{z}_{m[K_m]} &= Q_m \left\{c_{a[K_m]} - z_{m[K_m]}/(V_m P_m)\right\} \\
\dot{z}_{l[K_m]} &= Q_l \left\{c_{a[K_m]} - z_{l[K_m]}/(V_l P_l)\right\} - \frac{V_{max} z_{l[Km]}}{z_l + K_m V_l P_l} + \frac{V_{max} z_l z_{l[K_m]}}{(z_l + K_m V_l P_l)^2} + \frac{V_{max} V_l P_l z_l}{(z_l + K_m V_l P_l)^2}
\end{aligned}$$

where

$$c_{v[K_m]} = Q_c^{-1} \sum_{*=f,r,m,l} Q_* z_{*[K_m]}/(V_* P_*),$$

$$c_{a[K_m]} = \{Q_c c_{v[K_m]} + Q_p z_{h[K_m]}/V_h\}/(Q_c + Q_p/P_b),$$

and $c_{x[K_m]} = c_{a[K_m]}/P_b$. Since the initial conditions of the original system were free of K_m, the initial conditions for the supplemental equations are

$$z_{h[K_m]}(0) = 0,\ z_{f[K_m]}(0) = 0,\ z_{r[K_m]}(0) = 0,\ z_{m[K_m]}(0) = 0,\ z_{l[K_m]}(0) = 0.$$

D.5 Sensitivity equations for total cardiac output, Q_c

The tissue blood flow rates Q_* are all functions of the cardiac output, Q_c, of a rat

$$\begin{aligned} Q_f = Q_{fc}Q_c &\Rightarrow \partial Q_f/\partial Q_c = Q_{fc} \\ Q_l = Q_{lc}Q_c &\Rightarrow \partial Q_l/\partial Q_c = Q_{lc} \\ Q_m = Q_{mc}Q_c &\Rightarrow \partial Q_m/\partial Q_c = Q_{mc} \\ Q_r = Q_{rc}Q_c &\Rightarrow \partial Q_r/\partial Q_c = Q_{rc} \end{aligned}$$

where $Q_{rc} = 1 - Q_{fc} - Q_{lc} - Q_{mc}$. The following differential equations can be solved simultaneously with the original system to approximate the sensitivity, $z_{h[Q_c]} = \partial z_h/\partial Q_c$, for Q_c. In the equations below, the subscript $[Q_c]$ indicates a partial derivative with respect to Q_c.

$$\begin{aligned} \dot{z}_{h[Q_c]} &= \texttt{RATS}\, Q_p \left(c_{x[Q_c]} - z_{h[Q_c]}/V_h\right) - K_L z_{h[Q_c]} \\ \dot{z}_{f[Q_c]} &= Q_f \left\{c_{a[Q_c]} - z_{f[Q_c]}/(V_f P_f)\right\} + Q_{fc}\left\{c_a - z_f/(V_f P_f)\right\} \\ \dot{z}_{r[Q_c]} &= Q_r \left\{c_{a[Q_c]} - z_{r[Q_c]}/(V_r P_r)\right\} + Q_{rc}\left\{c_a - z_r/(V_r P_r)\right\} \\ \dot{z}_{m[Q_c]} &= Q_m \left\{c_{a[Q_c]} - z_{m[Q_c]}/(V_m P_m)\right\} + Q_{mc}\left\{c_a - z_m/(V_m P_m)\right\} \\ \dot{z}_{l[Q_c]} &= Q_l \left\{c_{a[Q_c]} - z_{l[Q_c]}/(V_l P_l)\right\} + Q_{lc}\left\{c_a - z_l/(V_l P_l)\right\} - \frac{V_{max} z_{l[Q_c]}}{z_l + K_m V_l P_l} + \frac{V_{max} z_l z_{l[Q_c]}}{(z_l + K_m V_l P_l)^2} \end{aligned}$$

where

$$c_{v[Q_c]} = \sum_{*=f,r,m,l} Q_{*c} z_{*[Q_c]}/(V_* P_*),$$

$$c_{a[Q_c]} = \{Q_c c_{v[Q_c]} + Q_p z_{h[Q_c]}/V_h + c_v\}/(Q_c + Q_p/P_b) - \{Q_c c_v + Q_p z_h/V_h\}/(Q_c + Q_p/P_b)^2,$$

and $c_{x[Q_c]} = c_{a[Q_c]}/P_b$. Since the initial conditions of the original system were free of Q_c, the initial conditions for the supplemental equations are

$$z_{h[Q_c]}(0) = 0,\ z_{f[Q_c]}(0) = 0,\ z_{r[Q_c]}(0) = 0,\ z_{m[Q_c]}(0) = 0,\ z_{l[Q_c]}(0) = 0.$$

D.6 Sensitivity equations for ventilation rate, Q_p

To approximate the sensitivity, $z_{h[Q_p]} = \partial z_h / \partial Q_p$, for the ventilation rate parameter, Q_p, the following differential equations can be solved simultaneously with the original system. In the equations below, the subscript $[Q_p]$ indicates a partial derivative with respect to Q_p.

$$\begin{aligned}
\dot{z}_{h[Q_p]} &= \texttt{RATS}\, Q_p \left(c_{x[Q_p]} - z_{h[Q_p]}/V_h\right) + \texttt{RATS}\, \left(c_x - z_h/V_h\right) - K_L z_{h[Q_p]} \\
\dot{z}_{f[Q_p]} &= Q_f \left\{c_{a[Q_p]} - z_{f[Q_p]}/(V_f P_f)\right\} \\
\dot{z}_{r[Q_p]} &= Q_r \left\{c_{a[Q_p]} - z_{r[Q_p]}/(V_r P_r)\right\} \\
\dot{z}_{m[Q_p]} &= Q_m \left\{c_{a[Q_p]} - z_{m[Q_p]}/(V_m P_m)\right\} \\
\dot{z}_{l[Q_p]} &= Q_l \left\{c_{a[Q_p]} - z_{l[Q_p]}/(V_l P_l)\right\} - \frac{V_{max} z_{l[Q_p]}}{z_l + K_m V_l P_l} + \frac{V_{max} z_l z_{l[Q_p]}}{(z_l + K_m V_l P_l)^2}
\end{aligned}$$

where

$$c_{v[Q_p]} = Q_c^{-1} \sum_{*=f,r,m,l} Q_* z_{*[Q_p]}/(V_* P_*),$$

$$c_{a[Q_p]} = \{Q_c c_{v[Q_p]} + (Q_p z_{h[Q_p]} + z_h)/V_h\}/(Q_c + Q_p/P_b) + \{Q_c c_v + Q_p z_h/V_h\}/\{P_b (Q_c + Q_p/P_b)^2\},$$

and $c_{x[Q_p]} = c_{a[Q_p]}/P_b$. Since the initial conditions of the original system were free of Q_p, the initial conditions for the supplemental equations are

$$z_{h[Q_p]}(0) = 0,\ z_{f[Q_p]}(0) = 0,\ z_{r[Q_p]}(0) = 0,\ z_{m[Q_p]}(0) = 0,\ z_{l[Q_p]}(0) = 0.$$