TOPICS IN STATISTICAL CALIBRATION

DISSERTATION

Brandon M. Greenwell, Civilian

AFIT-ENC-DS-14-M-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

The views expressed in this dissertation are those of the author and do not reflect the official policy or position of the United States Air Force, the Department of Defense, or the United States Government.

AFIT-ENC-DS-14-M-01

TOPICS IN STATISTICAL CALIBRATION

DISSERTATION

Presented to the Faculty

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

in Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

Brandon M. Greenwell, B.S., M.S.

Civilian

March 2014

AFIT-ENC-DS-14-M-01

TOPICS IN STATISTICAL CALIBRATION

DISSERTATION

Brandon M. Greenwell, B.S., M.S.
Civilian

Approved:

Christine M. Schubert Kabban, PhD (Chairman) — 2/27/2014 Date

Raymond Hill, PhD (Member) — 2/27/2014 Date

Dursun Bulutoglu, PhD (Member) — 2/27/2014 Date

Accepted:

ADEDEJI B. BADIRU, PhD — 3/3/2014 Date
Dean, Graduate School of Engineering and Management

AFIT-ENC-DS-14-M-01

Abstract

Calibration, more generally referred to as inverse estimation, is an important and controversial topic in statistics. In this work, both semiparametric calibration and the application of calibration to grouped data is considered, both of which may be addressed through the use of the linear mixed-effects model. A method is proposed for obtaining calibration intervals that has good coverage probability when the calibration curve has been estimated semiparametrically and is biased. The traditional Bayesian approach to calibration is also expanded by allowing for a semiparametric estimate of the calibration curve. The usual methods for linear calibration are then extended to the case of grouped data, that is, where observations can be categorized into a finite set of homogeneous clusters. Observations belonging to the same cluster are often similar and cannot be considered as independent; hence, we must account for within-subject correlation when making inference. Estimation techniques begin by extending the familiar Wald-based and inversion methods using the linear mixed-effects model. Then, a simple parametric bootstrap algorithm is proposed that can be used to either obtain calibration intervals directly, or to improve the inversion interval by relaxing the normality constraint on the approximate predictive pivot. Many of these methods have been incorporated into the `R` package, `investr`, which has been developed for analyzing calibration data.

To my parents

Table of Contents

Page

List of Figures

Figure Page

List of Tables

List of Common Symbols

Symbol	Definition
$\mathbf{0}_{m\times n}$	$m \times n$ matrix of all 0's
$\boldsymbol{I}_n$	$n \times n$ identity matrix
$\mathbf{1}_n$	$n \times 1$ vector of all 1's
$\boldsymbol{J}_n$	$n \times n$ matrix of all 1's
$I(\cdot)$	the indicator function

List of Acronyms

Acronym	Definition
BLUE	best linear unbiased estimator
BLUP	best linear unbiased predictor
EBLUE	estimated (or empirical) best linear unbiased estimator
EBLUP	estimated (or empirical) best linear unbiased predictor
GLS	generalized least squares
i.i.d.	independent and identically distributed
LM	linear model
LMM	linear mixed-effects model
LS	least squares
ML	maximum likelihood
MSE	Mean squared error
NLMM	Nonlinear mixed-effects model
P-spline	penalized regression spline
PSS	penalized sum of squares
REML	restricted (or residual) maximum likelihood

TOPICS IN STATISTICAL CALIBRATION

I. Introduction

An important part of statistics is building mathematical models to help describe the relationship between variables. Often, we have a dependent variable $\mathcal{Y}$, called the response, and an independent variable $\mathcal{X}$, called the predictor. The researcher designs an experiment and procures n observed pairs $(x_1, y_1), \ldots, (x_n, y_n)$ that are then used to fit a regression model. The fitted regression model is often used to make predictions. In particular,

- predict an individual response for a given value of the predictor;
- estimate the mean response for a given value of the predictor.

Sometimes, however, the researcher is interested in the reverse problem. That is, there is a need to estimate the predictor value from an observed value of the response (*calibration*) or a specified value of the mean response (*regulation*). Both are referred to more generally as *inverse estimation*. Calibration has also been referred to in the literature as *inverse prediction*, *inverse regression*, and *discrimination*. In this paper, we discuss calibration, in particular, univariate calibration. For an overview on topics in multivariate calibration, see Brown (1982), Brown and Sundberg (1987), and Brown (1993). Point estimation is reviewed in Section 3.2, interval estimation in Section 3.3, and Bayesian calibration in Section 3.4.

A calibration experiment typically consists of two stages. In the first stage, n observations (x_i, y_i) are collected (henceforth referred to as the *standards*) and used to fit a regression model $\mathrm{E}\{\mathcal{Y}|x\} = \mu$. The fitted model, denoted $\widehat{\mu}$, is often referred

to as the *calibration curve* or *standards curve*. In the second stage, m ($m \geq 1$) values of the response are observed with unknown predictor value x_0 (henceforth referred to as the *unknowns*). The goal is to use the calibration curve $\widehat{\mu}$ to estimate x_0. The following example will help to clarify the basic idea.

Suppose a new procedure has been developed for measuring the concentration (in μg/ml) of arsenic in water samples that is cheaper, but less accurate than the existing method. An investigator has procured 32 water samples containing preselected amounts of arsenic x_i and subjected them to the new method producing measured concentrations y_i. The standards, taken from Graybill and Iyer (1994), are plotted in Figure 1.1. A new water sample is then obtained with unknown arsenic concentration x_0 and subjected to the new method, producing a measured concentration of 3.0 μg/ml. It is desired to estimate the true concentration of arsenic in the newly obtained water sample.

The goal of this work is to extend the methods of calibration to more complicated settings. In Chapter 4, we introduce *semiparametric calibration*. Here we use semiparametric regression methods to estimate the calibration curve and make inference on the unknown x_0. The benefit of this approach is that we do not have to specify the exact form of the calibration curve. The downside to this approach is that the estimated calibration curve will be biased, hence, our inference about the unknown x_0 will also be biased. We correct for this bias by taking the mixed model approach to smoothing described in, for example, Ruppert and Wand (2003). A small Monte Carlo study shows that this correction is necessary to obtain calibration intervals with coverage probability near the nominal $1 - \alpha$ level. We also extend the method of Bayesian linear calibration put forth by Hoadley (1970) by allowing for the calibration curve to be estimated semiparametrically as in Crainiceanu et al. (2005). In Chapter 5, we extend the classical methods of calibration to work with

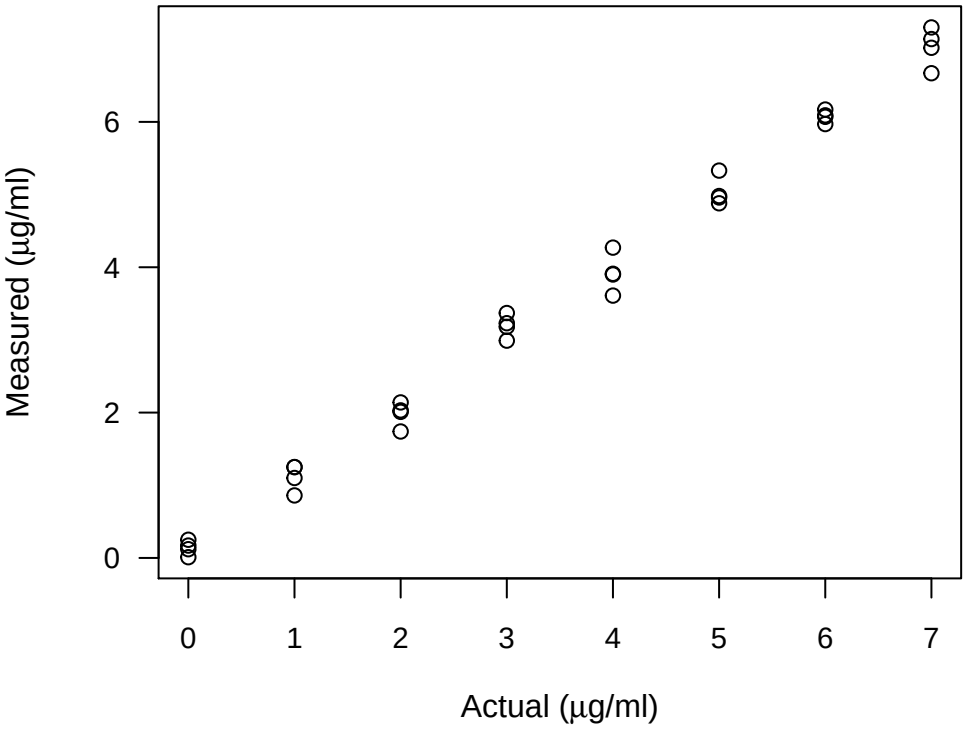


Figure 1.1: Scatterplot of the arsenic data.

linear mixed-effects models. We also propose a new parametric bootstrap algorithm that can be used to obtain an estimate of the entire sampling distribution of the estimate of x_0. We further show how this algorithm can also be used to improve upon the classical methods by removing normality assumptions that may not be satisfied in practice. We use several real datasets to demonstrate and compare our methods.

II. Statistical Background

This chapter provides an overview of some of the statistical concepts related to inverse estimation. In Section 2.1, we discuss the basic notational conventions used throughout this dissertation. Section 2.2 introduces the linear regression model. The extension to nonlinear regression models is given in Section 2.3. Section 2.4 is devoted to penalized regression splines, a type of semiparametric regression model where only part of the model is specified. Finally, in Section 2.5, we introduce the linear mixed-effects models, an extension of the linear model (LM) that allows for some of the regression coefficients to vary randomly.

2.1 Notation

For the most part, random variables will be denoted by capital letters in a calligraphic font (e.g., $\mathcal{Y}$). Vectors will be denoted by bold, lowercase letters and matrices will be denoted by bold, uppercase letters. When convenient, we will use the following notation to denote row, column, and diagonal vectors/matrices:

$$\begin{aligned}
\left\{ \underset{\text{col}}{} x_i \right\}_{i=1}^{n} &= (x_1, \ldots, x_n)' \\
\left\{ \underset{\text{row}}{} x_i \right\}_{i=1}^{n} &= (x_1, \ldots, x_n) \\
\left\{ \underset{\text{diag}}{} x_i \right\}_{i=1}^{n} &= \operatorname{diag}\{x_1, \ldots, x_n\}.
\end{aligned}$$

If $\mathcal{X}$ is a random variable, then $\mathcal{X} \sim (\mu, \sigma^2)$ simply means that $\mathcal{X}$ has some distribution with mean $\mathrm{E}\{\mathcal{X}\} = \mu$ and variance $\mathrm{Var}\{\mathcal{X}\} = \sigma^2$. Estimators and estimates will typically be denoted by the same Greek letter with a hat symbol. For example, depending on the context, $\widehat{\boldsymbol{\beta}}$ may represent a vector of estimators or their corresponding estimates.

2.2 Linear models

The linear regression model has been a mainstay of statistics for many years. It has the simple form

$$\mathcal{Y}_i = \boldsymbol{X}_i'\boldsymbol{\beta} + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.1}$$

where $\boldsymbol{X}_i = (x_{i1}, \ldots, x_{ip})'$ is a $p \times 1$ vector of predictor variables for the i-th observation, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of fixed (but unknown) regression coefficients, and $\epsilon_i \overset{iid}{\sim} (0, \sigma_\epsilon^2)$. Thus, an alternative formulation is to specify the mean response

$$\mathrm{E}\{\mathcal{Y}_i | \boldsymbol{X}\} = \boldsymbol{X}_i'\boldsymbol{\beta} = \mu_i.$$

Unless stated otherwise, $x_{i1} \equiv 1$ (i.e., the model contains an intercept). It is often convenient to work with the matrix form of (2.1), which is

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim (\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}_n), \tag{2.2}$$

where $\boldsymbol{\mathcal{Y}} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_n)'$ is an $n \times 1$ vector of response variables, $\boldsymbol{X} = (\boldsymbol{X}_1, \ldots, \boldsymbol{X}_p)'$ is an $n \times p$ matrix of predictor variables called the *design matrix*, and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_n)'$ is an $n \times 1$ vector of random errors. Equation (2.1) is special in that the response $\boldsymbol{\mathcal{Y}}$ is a linear function of the regression parameters $\boldsymbol{\beta}$.

A special case of (2.1) arises when $p = 2$ and the distribution for the errors is normal. That is,

$$\mathcal{Y}_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n, \tag{2.3}$$

where β_0 and β_1 are the intercept and slope of the regression line, respectively, and $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$. This is called the simple linear regression model and is often used for analyzing calibration data.

Another special case of the linear model (2.1) is when $x_{ij} = g_j(x_i)$, $j = 1, \ldots, p$, where each $g_j(\cdot)$ is a continuous function such as $\sqrt{\cdot}$ or $\log(\cdot)$. For example, a

polynomial model of degree p has the form

$$\mathcal{Y}_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \ldots + \beta_p x_i^p + \epsilon_i, \quad i = 1, \ldots, n.$$

Notice that a polynomial model is linear in the parameters even though it is nonlinear in the predictor variable; hence, it is a linear model.

2.2.1 Estimating the model parameters.

Estimation of $\boldsymbol{\beta}$ in the linear model (2.1) can be carried out via least squares (LS). The ordinary LS estimator of $\boldsymbol{\beta}$ minimizes the residual sum of squares,

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \ \|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|^2 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mathcal{Y}}. \tag{2.4}$$

If we make the additional assumption that the errors are normally distributed, then estimation can also be carried out by the method of maximum likelihood (ML). This has the benefit of simultaneously providing an estimator for both $\boldsymbol{\beta}$ and σ_ϵ^2. To proceed, we need to maximize the likelihood

$$\mathcal{L}\left(\boldsymbol{\beta}, \sigma_\epsilon^2 | \boldsymbol{\mathcal{Y}}\right) = \left(2\pi\sigma_\epsilon^2\right)^{-n/2} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|^2\right\},$$

or equivalently, maximize the log-likelihood

$$\mathscr{L}\left(\boldsymbol{\beta}, \sigma_\epsilon^2 | \boldsymbol{\mathcal{Y}}\right) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\left(\sigma_\epsilon^2\right) - \frac{1}{2\sigma_\epsilon^2}\|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|^2.$$

The derivatives of $\mathscr{L}(\boldsymbol{\beta}, \sigma_\epsilon^2 | \boldsymbol{\mathcal{Y}})$ with respect to the parameters $(\boldsymbol{\beta}, \sigma_\epsilon^2)$ are

$$\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \boldsymbol{\beta}} &= \frac{\boldsymbol{X}'(\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta})}{\sigma_\epsilon^2} \\
\frac{\partial \mathscr{L}}{\partial \sigma_\epsilon^2} &= \frac{\|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|^2}{2\sigma_\epsilon^2} - \frac{N}{2\sigma_\epsilon^2},
\end{aligned}$$

which, upon setting equal to zero and solving yields the ML estimators

$$\begin{aligned}
\widehat{\boldsymbol{\beta}} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mathcal{Y}} \\
\widehat{\sigma}_\epsilon^2 &= \left\|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2 / n
\end{aligned}$$

Fortunately, for the linear model (2.1) with normal errors, the ML estimator of $\boldsymbol{\beta}$ is the same as the LS estimator. It is customary to adjust the ML estimator of σ_ϵ^2 for bias by replacing it with $\widehat{\sigma}_\epsilon^2 = \left\|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}\right\|^2 /(n-p-1)$.

2.2.2 Predictions.

A common use of the fitted regression equation is to make predictions. There are two types of predictions we distinguish:

(1) estimate the mean response when $\boldsymbol{X} = \boldsymbol{X}_0$;

(2) predict a future observation corresponding to $\boldsymbol{X}_0$.

Let μ_0 and $\mathcal{Y}_0$ be the mean response and future observation of interest, respectively. We will see that the point estimators of μ_0 and $\mathcal{Y}_0$ are the same, namely the fitted value $\widehat{\mu}(x)$. Intuitively, the former should have a smaller standard error since there is less variability in estimating a fixed population parameter than in predicting a future value of a random variable. Consequently, a $100(1-\alpha)\%$ confidence interval for μ_0 at $\boldsymbol{X}_0$ will always be smaller than a $100(1-\alpha)\%$ prediction interval for $\mathcal{Y}_0$ corresponding to $\boldsymbol{X}_0$.

Let $\boldsymbol{X}_0$ be an arbitrary value of $\boldsymbol{X}$. Suppose we are interested in estimating the mean of $\mathcal{Y}$ given $\boldsymbol{X}_0$, $\mathrm{E}\{\mathcal{Y}|\boldsymbol{X}_0\} = \mu_0$. For the linear model (2.1), the best linear unbiased estimator (BLUE) of μ_0 is the fitted value $\widehat{\mu}_0 = \boldsymbol{X}_0'\widehat{\boldsymbol{\beta}}$. Furthermore, it is easy to show that

$$\mathrm{E}\{\widehat{\mu}_0\} = \boldsymbol{X}_0'\boldsymbol{\beta} = \mu_0$$

and

$$\mathrm{Var}\{\widehat{\mu}_0\} = \sigma_\epsilon^2\left[\boldsymbol{X}_0'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0\right] = S^2.$$

Assuming normal errors, it follows that

$$\frac{\widehat{\mu}_0 - \mu_0}{\widehat{\sigma}_\epsilon\sqrt{\left[\boldsymbol{X}_0'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\boldsymbol{X}_0\right]}} \sim \mathcal{T}\left(n-p\right).$$

Hence, a $100(1-\alpha)\%$ confidence interval for the mean response μ_0 is given by

$$\boldsymbol{X}_0'\widehat{\boldsymbol{\beta}} \pm t_{1-\alpha/2,n-p} \cdot \widehat{S}. \tag{2.5}$$

Let $\mathcal{Y}_0$ denote an individual or future observation at the given point $\boldsymbol{X}_0$. We assume that $\mathcal{Y}_0 = \boldsymbol{X}_0'\boldsymbol{\beta} + \epsilon_0$, where $\epsilon_0 \sim \mathcal{N}(0, \sigma_\epsilon^2)$ and is independent of $\boldsymbol{\epsilon}$. The best predictor of $\mathcal{Y}_0$ is $\widehat{y}_0 = \boldsymbol{X}_0'\widehat{\boldsymbol{\beta}}$, the same as $\widehat{\mu}_0$, however, the variance of $\widehat{\mathcal{Y}}_0$ is wider: $\text{Var}\left\{\widehat{\mathcal{Y}}_0\right\} = \sigma_\epsilon^2 + \text{Var}\left\{\widehat{\mu}_0\right\}$. This is intuitive since there is greater uncertainty in predicting the outcome of a continuous random variable than in estimating a fixed (population) parameter, such as its mean. Under the assumption of normal errors, we have that

$$\frac{\widehat{y}_0 - \mathcal{Y}_0}{\widehat{\sigma}_\epsilon \sqrt{\left[1 + \boldsymbol{X}_0' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{X}_0\right]}} \sim \mathcal{T}\left(n-p\right),$$

therefore, a $100(1-\alpha)\%$ prediction interval for $\mathcal{Y}_0$ is simply

$$\boldsymbol{X}_0'\widehat{\boldsymbol{\beta}} \pm t_{1-\alpha/2,n-p} \sqrt{\widehat{\sigma}_\epsilon^2 + \widehat{S}^2}. \tag{2.6}$$

We call (2.6) a prediction interval, as opposed to a confidence interval, since $\widehat{y}_0$ is a prediction for the outcome of the random variable $\mathcal{Y}_0$.

2.3 Nonlinear models

Often in practice there is an underlying theoretical model relating the response to the predictors, and this model may be nonlinear in the parameters, $\boldsymbol{\beta}$. Such nonlinear relationships lead us to the nonlinear regression model. For a single predictor variable, this model is

$$\mathcal{Y}_i = \mu(x_i; \boldsymbol{\beta}) + \epsilon_i, \tag{2.7}$$

where $\mu(\cdot)$ is a known expectation function that is nonlinear in at least one of the parameters in $\boldsymbol{\beta}$, and $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$.

2.3.1 Estimating the model parameters.

Borrowing from the notation in Seber and Wilde (2003), let $\mu_i(\boldsymbol{\beta}) = \mu(x_i; \boldsymbol{\beta})$,

$$\boldsymbol{\mu}(\boldsymbol{\beta}) = (\mu_1(\boldsymbol{\beta}), \ldots, \mu_N(\boldsymbol{\beta}))',$$

and

$$\boldsymbol{D}(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{\mu}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} = \left\{ \underset{\text{row}}{} \left\{ \underset{\text{col}}{} \frac{\partial \mu_i(\boldsymbol{\beta})}{\partial \beta_j} \right\}_{i=1}^{n} \right\}_{j=1}^{p}.$$

For convenience, let $\widehat{\boldsymbol{D}} = \boldsymbol{D}(\widehat{\boldsymbol{\beta}})$.

The approach to estimating $\boldsymbol{\beta}$ in the nonlinear model (2.7) is similar to the approach in the linear model (2.1). That is, we choose the value of $\boldsymbol{\beta}$ that minimizes the residual sum of squares, $\text{RSS}(\boldsymbol{\beta})$, defined by

$$\text{RSS}(\boldsymbol{\beta}) = \|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}(\boldsymbol{\beta})\|^2. \tag{2.8}$$

However, since $\mu(\cdot)$ is nonlinear in the parameters $\boldsymbol{\beta}$, minimizing Equation (2.8) requires iterative techniques such as methods of *steepest descent* or the *Gauss-Newton* algorithm, which we describe below.

Given a starting value or current guess $\boldsymbol{\beta}^{(0)}$ of the value of $\boldsymbol{\beta}$ that minimizes (2.8), we can approximate $\mu(x_i; \boldsymbol{\beta})$ with a Taylor-series approximation around $\boldsymbol{\beta}^{(0)}$. From a first-order Taylor series expansion, we get

$$\boldsymbol{\mu}(\boldsymbol{\beta}) \approx \boldsymbol{\mu}(\boldsymbol{\beta}^{(0)}) + \boldsymbol{D}(\boldsymbol{\beta}^{(0)})'(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)}), \tag{2.9}$$

The derivative matrix $\boldsymbol{D}(\boldsymbol{\beta})$ plays the same role as the design matrix $\boldsymbol{X}$ in the linear model (2.1), except that $\boldsymbol{D}(\boldsymbol{\beta})$ may depend on the unknown regression parameters $\boldsymbol{\beta}$. Substituting (2.9) into Equation (2.8), we get

$$\text{RSS}(\boldsymbol{\beta}) \approx \left\|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(0)}) - \boldsymbol{D}(\boldsymbol{\beta}^{(0)})(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})\right\|^2 \tag{2.10}$$

$$= \left\|\boldsymbol{\mathcal{Y}}^{(0)} - \boldsymbol{X}^{(0)}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(0)})\right\|^2, \tag{2.11}$$

where $\boldsymbol{\mathcal{Y}}^{(0)} = \boldsymbol{\mathcal{Y}} - \boldsymbol{\mu}(\boldsymbol{\beta}^{(0)})$, and $\boldsymbol{X}^{(0)} = \boldsymbol{D}(\boldsymbol{\beta}^{(0)})$. Equation (2.10) is of the same form as the sum of squares for the linear model. Based on this approximation, the LS estimate of $\boldsymbol{\beta}$ is then

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{\beta}^{(0)} + \left[\boldsymbol{X}^{(0)\prime}\boldsymbol{X}^{(0)}\right]^{-1} \boldsymbol{X}^{(0)\prime}\boldsymbol{\mathcal{Y}}^{(0)}.$$

Hence, given a current approximation $\boldsymbol{\beta}^{(k)}$ of $\boldsymbol{\beta}$, the updated approximation is

$$\boldsymbol{\beta}^{(k+1)} = \boldsymbol{\beta}^{(k)} + \left[\boldsymbol{X}^{(k)\prime}\boldsymbol{X}^{(k)}\right]^{-1} \boldsymbol{X}^{(k)\prime}\boldsymbol{\mathcal{Y}}^{(k)}.$$

This process is iterated until a suitable convergence criterion is met. Just as for the linear model, if we assume the errors are normally distributed, then the ML estimate of $\boldsymbol{\beta}$ is the same as the LS solution.

2.3.2 Predictions.

Let x_0 be a known value of the predictor x_0. The estimate of the mean response $\{\mathcal{Y}|x_0\} = \mu(x_0;\boldsymbol{\beta})$ is $\widehat{\mu}_0 = \mu(x_0;\widehat{\boldsymbol{\beta}})$. Furthermore, assuming normal errors, an approximate $100(1-\alpha)\%$ confidence interval for the mean response μ_0 can be obtained as in Equation (2.5) but with $\widehat{S}^2$ computed as

$$\widehat{S}^2 = \widehat{\sigma}^2_\epsilon \boldsymbol{d}_0' \left(\widehat{\boldsymbol{D}}'\widehat{\boldsymbol{D}}\right)^{-1} \boldsymbol{d}_0,$$

where

$$\boldsymbol{d}_0 = \left\{ \underset{\text{row}}{\frac{\partial \mu\,(x_0;\boldsymbol{\beta})}{\partial \beta_i}} \right\}_{i=1}^{p} \Bigg|_{\boldsymbol{\beta}=\widehat{\boldsymbol{\beta}}}.$$

An approximate $100(1-\alpha)\%$ prediction interval for a future observation is similarly obtained. These intervals, however, rely on the same linear approximation used to compute $\widehat{\boldsymbol{\beta}}$. For large samples, these intervals are often reliable. When the sample size is small or there is a lot of curvature, these intervals can be highly inaccurate. The *bootstrap* (Efron, 1979) provides an alternative method of inference under these circumstances.

2.4 Smoothing

Let $\boldsymbol{x} = (x_1, \ldots, x_n)'$ be a vector of predictor values and $\boldsymbol{\mathcal{Y}} = (\mathcal{Y}_1, \ldots, \mathcal{Y}_n)'$ be a vector of response variables. We assume that $\boldsymbol{X}$ and $\boldsymbol{\mathcal{Y}}$ are related by the regression model

$$\boldsymbol{\mathcal{Y}} = \mu(\boldsymbol{x}) + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}),$$

where $\mu(\cdot)$ is an unknown smooth function. In this section, we discuss a technique for estimating $\mu(\cdot)$ nonparametrically, often referred to as *scatterplot smoothing*, or just *smoothing*. In particular, we will focus on an important class of smoothers called *linear smoothers*. For linear smoothers, the prediction at any point x, is a linear combination of the response values: $\widehat{\mu}(x) = \boldsymbol{\omega}'\boldsymbol{\mathcal{Y}}$, where $\boldsymbol{\omega}$ is a constant that does not depend on the response vector $\boldsymbol{\mathcal{Y}}$. The vector of fitted values $\widehat{\boldsymbol{\mu}} = (\mu(x_1), \ldots, \mu(x_n))'$ can be written in matrix form as $\boldsymbol{S}\boldsymbol{\mathcal{Y}}$, where $\boldsymbol{S}$ is an $n \times n$ *smoother matrix*. Examples of linear smoothers include:

- running-mean, running-line, and running-polynomial smoothers;
- locally-weighted polynomial smoothers (i.e., LOESS);
- spline-based smoothers;
- kernel smoothers.

The remainder of this section discusses spline smoothing, specifically, the penalized regression spline (P-spline).

Regression splines represent the mean response $\mu(x)$ as a piecewise polynomial of degree p. The regions that define each piece are separated by special breakpoints $\xi_1 < \ldots < \xi_K$ called *knots*. By increasing p or the number of knots, the family of curves becomes more flexible; thus, as shown in Figure 2.1, we can easily handle any level of complexity.

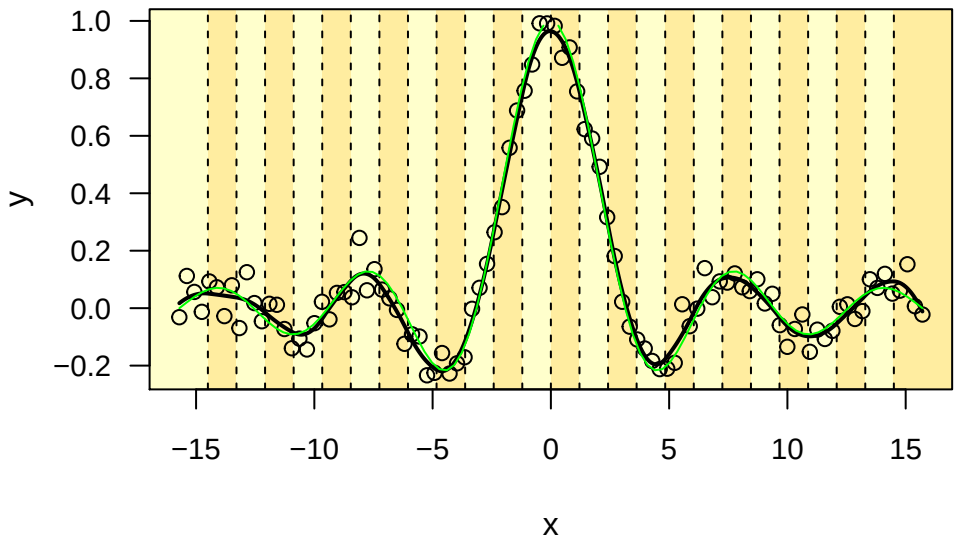


Figure 2.1: Unnormalized sinc function example. Scatterplot of 100 observations generated from a regression model with mean response $\mu(x) = \sin(x)/x$ (green line) and i.i.d. $\mathcal{N}(0, 0.05^2)$ errors. The solid line shows a quadratic P-spline fit with smoothing parameter $\lambda = 0.9986$. The dotted lines indicate the position of the knots ξ_k.

A p-th degree spline function has the form

$$\mu(x) = \beta_0 + \beta_1 x + \ldots + \beta_p x^p + \sum_{k=1}^{K} \alpha_k (x - \xi_k)_+^p, \tag{2.12}$$

where the notation a_+ denotes the positive part of a, that is, $a_+ = a \cdot I(a \geq 0)$. Many methods for choosing the number and location of the knots are given in the literature (see, for example, Ruppert (2002) and the references therein). Let n_x be the number of unique x_i. A reasonable choice for the number of knots (Ruppert and Wand, 2003, pg. 126) is $K = \min(n_x/4, 35)$ with knot locations

$$\xi_k = \left(\frac{k+1}{K+2}\right)\text{-th sample quantile of the unique } x_i, \quad k = 1, \ldots, K.$$

The idea is to choose enough knots to capture the structure of $\mu(x)$. If both p and K are too large, we run the risk of overfitting (i.e., low bias and high variance). If K is too small then the resulting fit may be too restrictive (i.e., low variance but high

bias). There is rarely the need to go beyond a cubic polynomial model and so typical choices for p are 1, 2, or 3.

In matrix form, the polynomial spline model is

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim (\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I}), \tag{2.13}$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1 & \cdots & x_1^p & (x_1 - \xi_1)_+^p & \dots & (x_1 - \xi_K)_+^p \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_n & \cdots & x_n^p & (x_n - \xi_1)_+^p & \dots & (x_n - \xi_K)_+^p \end{bmatrix}.$$

Equation (2.13) has the same form as the linear model (2.2). The ordinary least squares fit, however, will be too "wiggly." The idea behind penalized spline regression is to shrink the coefficients α_k by imposing a penalty on their size, thereby limiting their impact on the estimated response curve. The estimated coefficients minimize the penalized residual sum of squares

$$\text{PSS} = \|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda^{2p}\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta}, \tag{2.14}$$

where $\boldsymbol{D} = \text{diag}\left\{\mathbf{0}_{(p+1)\times(p+1)}, \boldsymbol{I}_{K\times K}\right\}$. The penalized least squares solution is then

$$\widehat{\boldsymbol{\beta}}_\lambda = \underset{\boldsymbol{\beta}}{\text{argmin}} \ \|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \lambda^{2p}\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta} = \left(\boldsymbol{X}'\boldsymbol{X} + \lambda^{2p}\boldsymbol{D}\right)^{-1}\boldsymbol{X}'\boldsymbol{\mathcal{Y}}. \tag{2.15}$$

Here $\lambda \geq 0$ is a *smoothing parameter* that controls the wiggliness of the fit. Small values of λ produce wiggly curves while larger values produce smoother curves. The term $\lambda^{2p}\boldsymbol{\beta}'\boldsymbol{D}\boldsymbol{\beta}$ is called the *roughness penalty*. If $\boldsymbol{D} = \boldsymbol{I}$, then the penalized least squares solution (2.15) is equivalent to the ridge regression estimate of $\boldsymbol{\beta}$. The roughness penalty can also be written as $\lambda^{2p}\|\boldsymbol{\alpha}\|^2$, where $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_K)'$ is the vector of coefficients for the spline basis functions. Hence, P-splines enforce a penalty on the ℓ^2-norm of $\boldsymbol{\alpha}$, so none of the polynomial coefficients are penalized (See Figure 2.2)!

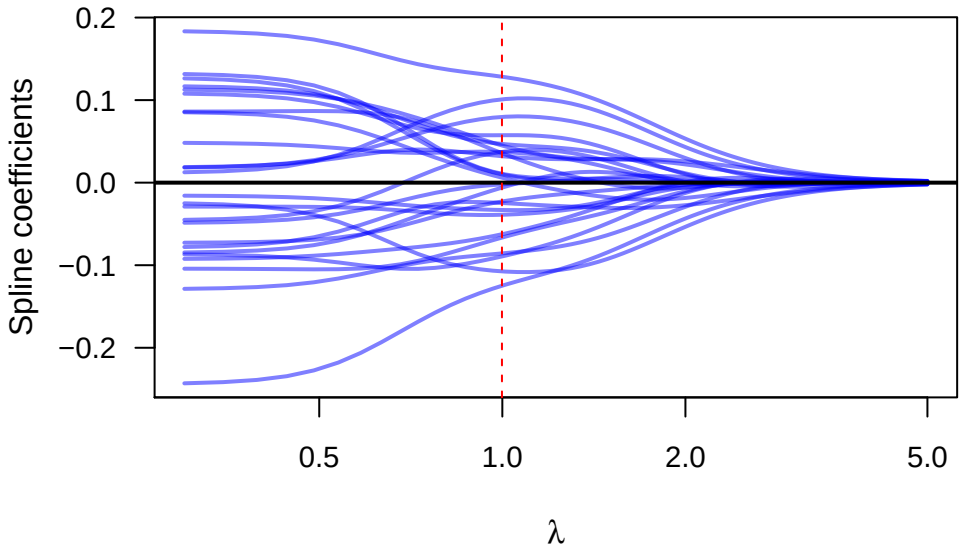


Figure 2.2: Profiles of spline coefficients as the smoothing parameter λ is varied. Coefficients are plotted versus λ. A vertical line is drawn at $\lambda = 0.9986$.

The fitted values are given by $\widehat{\boldsymbol{\mu}} = \boldsymbol{S}_\lambda \boldsymbol{\mathcal{Y}}$, where $\boldsymbol{S}_\lambda = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X} + \lambda^{2p}\boldsymbol{D}\right)^{-1}\boldsymbol{X}'$. From this point forward, we will drop the subscript λ on the smoother matrix and just write $\boldsymbol{S}$. The smoothing parameter λ is unknown but can be specified a priori. However, it is often beneficial to let the data determine the appropriate amount of smoothness. To this end, cross-validation techniques are often used to estimate λ from the given data. In Chapter 5, we discuss an alternative approach to P-splines that automatically selects an appropriate amount of smoothness.

2.4.1 Inference for linear smoothers.

Consider the general heteroscedastic error model

$$\mathcal{Y}_i = \mu(x_i) + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} (0, \sigma_\epsilon^2), \quad i = 1, \ldots n,$$

where $\mu(\cdot)$ is an unknown smooth function. Let $\widehat{\mu}(x)$ be an estimate of $\mu(\cdot)$ based on a linear smoother. The covariance matrix of the vector of fitted values $\widehat{\boldsymbol{\mu}} = \boldsymbol{S}\boldsymbol{\mathcal{Y}}$ is

$$\text{Var}\left\{\widehat{\boldsymbol{\mu}}\right\} = \boldsymbol{S}\left(\sigma_\epsilon^2 \boldsymbol{I}\right)\boldsymbol{S}' = \sigma_\epsilon^2 \boldsymbol{S}\boldsymbol{S}'. \tag{2.16}$$

For a single point, the quantity $Q = [\widehat{\mu}(x) - \mu(x)] / \operatorname{se}\{\widehat{\mu}(x)\}$ is approximately pivotal—the distribution is free of unknown parameters, at least for sufficiently large sample size n. Therefore, assuming $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \boldsymbol{I})$, Equation (2.16) can be used to form confidence intervals and prediction intervals. Note, however, that $\mathrm{E}\{\widehat{\boldsymbol{\mu}}\} = \boldsymbol{X}\left(\boldsymbol{X}'\boldsymbol{X} + \lambda^{2p}\boldsymbol{D}\right)^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\mathcal{Y}}$; hence, $\widehat{\mu}(x)$ is a biased estimator of $\mu(x)$. Unless the bias is negligible, the confidence intervals discussed here only cover $\mathrm{E}\{\widehat{\mu}(x)\}$ with $100(1-\alpha)\%$ confidence. This problem is remedied for P-splines in Chapter 4 where we discuss an alternative approach using mixed model methodology.

Let x be an arbitrary value of the explanatory variable. Recall that, for linear smoothers, the fitted value $\widehat{\mu}(x)$ can be written as $\boldsymbol{\omega}'\boldsymbol{\mathcal{Y}}$, a linear combination of the response variables. The variance of $\widehat{\mu}(x)$ is just $\operatorname{Var}\{\boldsymbol{\omega}'\boldsymbol{\mathcal{Y}}\} = \sigma_\epsilon^2\boldsymbol{\omega}'\boldsymbol{\omega}$. Given an estimate $\widehat{\sigma}_\epsilon^2$ of σ_ϵ^2, an approximate $100(1-\alpha)\%$ confidence interval for $\mu(x)$ is given by

$$\widehat{\mu}(x) \pm t_{1-\alpha/2,df}\widehat{\sigma}_\epsilon\sqrt{\boldsymbol{\omega}'\boldsymbol{\omega}},$$

where $df = n - 2\operatorname{tr}(\boldsymbol{S}) + \operatorname{tr}(\boldsymbol{S}\boldsymbol{S}')$. For sufficiently large sample size n, the quantity $t_{1-\alpha/2,df}$ can be replaced with $z_{1-\alpha/2}$, the $1-\alpha/2$ quantile of a standard normal distribution. Similarly, A $100(1-\alpha)\%$ prediction interval for a new observation is

$$\widehat{\mu}(x) \pm t_{1-\alpha/2,df}\widehat{\sigma}_\epsilon\sqrt{1 + \boldsymbol{\omega}'\boldsymbol{\omega}}.$$

2.5 Linear mixed effects models

Mixed effects models (henceforth referred to as just mixed models) represent a large and growing area of statistics. In this section, we only summarize the key aspects of a special kind of mixed model called the linear mixed-effects model (LMM). The extension to nonlinear mixed effects models and generalized linear mixed effects models is discussed in Pinheiro and Bates (2009) and McCulloch et al. (2008), respectively. Mixed models are useful for describing *grouped data* where observations belonging to the same group are correlated. One way of accounting

for such correlation is by introducing random effects into the model which induces a particular correlation structure on the response vector. Different random effects structures induce different correlation structures on the response.

The LMM extends the basic LM (Equation (2.2)) to

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{2.17}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known design matrices, $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{\alpha}$ is a vector of random effects distributed as $\boldsymbol{\alpha} \sim \mathcal{N}(0, \boldsymbol{G})$, and $\boldsymbol{\epsilon}$ is a vector of random errors distributed as $\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{R})$. Further, it is assumed that the random effects and errors are mutually independent, that is, $\boldsymbol{\alpha} \perp\!\!\!\perp \boldsymbol{\epsilon}$.

Estimating the fixed effects $\boldsymbol{\beta}$ is rather straightforward and does not require normality. Note that $\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mathcal{E}}$, where $\mathrm{E}\{\boldsymbol{\mathcal{E}}\} = \boldsymbol{0}$ and $\mathrm{Var}\{\boldsymbol{\mathcal{E}}\} = \boldsymbol{Z}\boldsymbol{G}\boldsymbol{Z}' + R = \boldsymbol{V}$. Assuming normality, the log-likelihood (ignoring constants) is given by

$$\mathscr{L}(\boldsymbol{\beta}, \boldsymbol{\theta}) = -\frac{1}{2}\log|\boldsymbol{V}| - \frac{1}{2}(\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta})'\boldsymbol{V}^{-1}(\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}), \tag{2.18}$$

where $\boldsymbol{\theta}$ is a vector containing the unique elements of $\boldsymbol{V}$. Equating the partial derivative of $\mathscr{L}(\boldsymbol{\beta}, \boldsymbol{\theta})$, with respect to the parameter $\boldsymbol{\beta}$, yields the so-called generalized least squares (GLS) estimator

$$\widetilde{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{\mathcal{Y}}. \tag{2.19}$$

Notice, however, that the GLS estimator—which happens to be the BLUE of $\boldsymbol{\beta}$—depends on the variance-covariance matrix $\boldsymbol{V}$. Since this is rarely available in practice, the usual procedure is to estimate $\boldsymbol{V}$ and then plug this into Equation (2.19). In other words, if $\widehat{\boldsymbol{V}}$ is an estimate of $\boldsymbol{V}$, then the estimated (or empirical) best linear unbiased estimator (EBLUE) of $\boldsymbol{\beta}$ is

$$\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{\mathcal{Y}}. \tag{2.20}$$

This causes some difficulties, mostly in terms of finite-sample inference, since there is no simple way to account for the variability of $\widehat{\boldsymbol{V}}$ when calculating Var $\left\{\widehat{\boldsymbol{\beta}}\right\}$. See, for example, McCulloch et al. (2008, pp. 165-167).

A technique known as *best linear unbiased prediction* is commonly used to estimate the random effects $\boldsymbol{\alpha}$. It can be shown (Henderson, 1973) that the BLUE of $\boldsymbol{\beta}$ and the best linear unbiased predictor (BLUP) of $\boldsymbol{\alpha}$, denoted $\widetilde{\boldsymbol{\alpha}}$, can be determined simultaneously as the solutions to a penalized least squares problem,

$$\begin{aligned}\begin{bmatrix}\widetilde{\boldsymbol{\beta}}\\ \widetilde{\boldsymbol{\alpha}}\end{bmatrix} &= \underset{\boldsymbol{\beta},\boldsymbol{\alpha}}{\operatorname{argmin}}\,(\boldsymbol{\mathcal{Y}}-\boldsymbol{X\beta}-\boldsymbol{Z\alpha})'\boldsymbol{R}^{-1}(\boldsymbol{\mathcal{Y}}-\boldsymbol{X\beta}-\boldsymbol{Z\alpha})+\boldsymbol{\alpha}'\boldsymbol{G}^{-1}\boldsymbol{\alpha}\\ &= \boldsymbol{\Omega}(\boldsymbol{\Omega}'\boldsymbol{R}^{-1}\boldsymbol{\Omega}+\boldsymbol{D})\boldsymbol{\Omega}'\boldsymbol{R}^{-1}\boldsymbol{\mathcal{Y}},\end{aligned} \tag{2.21}$$

where $\widetilde{\boldsymbol{\beta}}$ is as in (2.19), $\widetilde{\boldsymbol{\alpha}} = \boldsymbol{GZ}'\boldsymbol{V}^{-1}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{X}\widetilde{\boldsymbol{\beta}}\right)$, $\boldsymbol{\Omega} = (\boldsymbol{X};\boldsymbol{Z})$, and $\boldsymbol{D} = \operatorname{diag}\{\boldsymbol{0}_{p\times p}, \boldsymbol{G}^{-1}\}$. For the special case $\boldsymbol{R} = \sigma_\epsilon^2\boldsymbol{I}$ and $\boldsymbol{G} = \sigma_\alpha^2\boldsymbol{I}$, the penalized sum of squares (PSS)—the minimand in Equation (2.21)— reduces to

$$\frac{1}{\sigma_\epsilon^2}\left\|\boldsymbol{\mathcal{Y}}-\boldsymbol{X\beta}-\boldsymbol{Z\alpha}\right\|^2+\frac{1}{\sigma_\alpha^2}\left\|\boldsymbol{\alpha}\right\|^2. \tag{2.22}$$

Similar to the EBLUE of $\boldsymbol{\beta}$, the estimated (or empirical) best linear unbiased predictor (EBLUP) of $\boldsymbol{\alpha}$ is just the BLUP $\widetilde{\boldsymbol{\alpha}}$ with $\boldsymbol{G}$ and $\boldsymbol{V}$ replaced with their respective estimates, $\widehat{\boldsymbol{G}}$ and $\widehat{\boldsymbol{V}}$:

$$\widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{G}}\boldsymbol{Z}'\widehat{\boldsymbol{V}}^{-1}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{X}\widehat{\boldsymbol{\beta}}\right). \tag{2.23}$$

In a similar fashion, the BLUP and EBLUP of the mean response are given, respectively, by the equations

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{X}\widetilde{\boldsymbol{\beta}}+\boldsymbol{Z}\widetilde{\boldsymbol{\alpha}}, \tag{2.24}$$

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{X}\widehat{\boldsymbol{\beta}}+\boldsymbol{Z}\widehat{\boldsymbol{\alpha}}. \tag{2.25}$$

Note that the EBLUP $\widehat{\boldsymbol{\mu}}$ is just the fitted values.

Obviously, estimating $\widetilde{\boldsymbol{\beta}}$ and $\widetilde{\boldsymbol{\alpha}}$, requires knowledge of the variance-covariance matrices $\boldsymbol{G}$ and $\boldsymbol{R}$, which are generally unknown. In practice, we often restrict these matrices to have a simple form, usually involving only a few unknown parameters, earlier denoted by $\boldsymbol{\theta}$. These parameters can be estimated via ML estimation. ML estimators of the variance components $\boldsymbol{\theta}$, however, tend to become badly biased as the number of fixed effects in the model increases. A more effective approach, known as restricted (or residual) maximum likelihood (REML) estimation, is often used instead (see, for example, McCulloch et al. (2008, chap. 6)).

III. Overview of Statistical Calibration

In this chapter, we present a thorough overview of calibration as it pertains to our goals for this dissertation. Although we cannot cover every aspect of calibration in this chapter, we have attempted to provide an extensive bibliography for the interested reader. The main themes of this chapter are going to be point estimation and interval estimation.

3.1 Controlled Calibration vs. Natural Calibration

Calibration experiments can be classified as one of two types: *controlled calibration* experiments and *natural calibration* experiments. In controlled calibration, the predictor values are held fixed by the experimenter (i.e., x is not considered a random variable). In this case, the predictor values are chosen to cover the experimental range of interest and the response is often replicated a number of times at each design point (see, for example, the arsenic data plotted in Figure 1.1). In contrast, the n observations (x_i, y_i) in a natural calibration experiment are considered a random sample from some bivariate distribution: $(\mathcal{X}, \mathcal{Y}) \sim g(x, y)$, where g is typically assumed to be a bivariate normal distribution. In summary, the values of the independent variable are either preselected or held fixed in controlled calibration and randomly sampled from a population of values in natural calibration. Distinguishing between these two types of calibration experiments is important, especially from an inferential standpoint, as emphasized in the papers by Brown (1982) and Brown (1993).

3.2 Point estimation

Here, we discuss point estimation of x_0 for the linear calibration problem, that is, when the calibration curve has the form of the simple linear regression model.

The two most popular estimators of x_0 are the *classical estimator* and the *inverse estimator*. The classical estimator, which dates back to Eisenhart (1939), is based on inverting the calibration curve at y_0 (i.e., solving the fitted regression equation for x_0) and is easily extended to polynomial and nonlinear calibration problems. The inverse estimator, as we will see in Section 3.4, is more useful under a specific Bayesian framework.

3.2.1 The classical estimator.

Suppose we have n observations $(x_i, \mathcal{Y}_i)$. We assume the x_i's were measured without error. (The case where there is error in both variables is discussed in Carroll and Spiegelman (1986).) Generally, the model considered is of the form

$$\mathcal{Y}_i = \mu\left(x_i; \boldsymbol{\beta}\right) + \epsilon_i, \quad i = 1, \ldots, n, \tag{3.1}$$

where $\mu = \mathrm{E}\left\{\mathcal{Y}|x\right\}$ is a known expectation function, $\boldsymbol{\beta}$ is a vector of p unknown regression parameters, and the errors are independent and identically distributed (i.i.d.) normal random variables: $\epsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\epsilon^2)$. We consider the linear calibration problem, a special case of Equation (3.1) with $\mathrm{E}\left\{\mathcal{Y}|x\right\} = \beta_0 + \beta_1 x$.

The fitted calibration line is given as

$$\widehat{\mu} = \widehat{\beta}_0 + \widehat{\beta}_1 x, \tag{3.2}$$

where $\widehat{\beta}_0$ and $\widehat{\beta}_1$ are the ML estimates of β_0 and β_1, respectively. If we observe $\mathcal{Y}_0 = y_0$, where $\mathcal{Y}_0 \sim \mathcal{N}(\beta_0 + \beta_1 x_0, \sigma_\epsilon^2)$, then the obvious estimate of x_0 is obtained by inverting the calibration line:

$$\widehat{x}_0 = \mu^{-1}(y_0; \boldsymbol{\beta}) = \frac{y_0 - \widehat{\beta}_0}{\widehat{\beta}_1} = \bar{x} + \frac{S_{xx}}{S_{xy}}(y_0 - \bar{y}), \tag{3.3}$$

where $S_{xy} = \sum(x_i - \bar{x})(y_i - \bar{y})$, and $S_{xx} = \sum(x_i - \bar{x})^2$. Under the assumption of i.i.d. normal errors, Equation (3.3) is the ML estimate of x_0. More generally, suppose in addition to the standards

$$(x_1, \mathcal{Y}_1), (x_2, \mathcal{Y}_2), \ldots, (x_n, \mathcal{Y}_n),$$

we have m unknowns

$$(x_0, \mathcal{Y}_{n+1}), (x_0, \mathcal{Y}_{n+2}), \ldots, (x_0, \mathcal{Y}_{n+m}).$$

We assume that the $\mathcal{Y}_i$'s $(i = 1, \ldots, n)$ are independent and distributed according to

$$\mathcal{Y}_i \sim \begin{cases} \mathcal{N}(\beta_0 + \beta_1 x_i, \sigma_{\text{I}}^2), & i = 1, 2, \ldots, n \\ \mathcal{N}(\beta_0 + \beta_1 x_0, \sigma_{\text{II}}^2), & i = n+1, n+2, \ldots, n+m \end{cases}.$$

In practice, assuming that $\sigma_{\text{I}}^2 = \sigma_{\text{II}}^2$ is often reasonable; however, some authors (e.g., Berkson (1969, p. 659)), have argued otherwise. In controlled calibration, one could argue that the variance in the first stage, σ_{I}^2, may be smaller than the variance from the second stage, σ_{II}^2, since the standards were likely collected under more highly controlled conditions. For example, the standards in the arsenic data may have been collected in a laboratory under tightly controlled conditions, while the measurement made on the new sample was likely made in the field (e.g., a lake) and therefore susceptible to greater measurement error.

The log-likelihood for all $n + m$ observations is

$$\mathscr{L}\left(\beta_0, \beta_1, \sigma_\epsilon^2, x_0\right) = -\frac{n+m}{2}\log(2\pi\sigma_\epsilon^2) - \frac{1}{2\sigma_\epsilon^2}\left[\sum_{i=1}^{n}(\mathcal{Y}_i - \beta_0 - \beta_1 x_i)^2 + \sum_{i=n+1}^{n+m}(\mathcal{Y}_i - \beta_0 - \beta_1 x_0)^2\right].$$

Equating the partial derivatives of $\mathscr{L}\left(\beta_0, \beta_1, \sigma_\epsilon^2, x_0\right)$ to zero and solving for the parameters produces the usual ML estimators of the slope and intercept, but also yields

$$\widehat{x}_0 = \frac{\overline{\mathcal{Y}}_0 - \widehat{\beta}_0}{\widehat{\beta}_1} \tag{3.4}$$

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{n+m-3}\left[\sum_{i=1}^{n}\left(\mathcal{Y}_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2 + \sum_{i=n+1}^{n+m}\left(\mathcal{Y}_i - \overline{\mathcal{Y}}_0\right)^2\right], \tag{3.5}$$

where $\overline{\mathcal{Y}}_0 = m^{-1}\sum_{i=n+1}^{n+m}\mathcal{Y}_i$. Equation (3.4) is known as the classical estimator of x_0. The mean of $\widehat{x}_0$ does not exist, and it has infinite variance and mean squared error

(MSE). This is not too surprising since Equation (3.4) is a ratio of jointly normal random variables (recall that a standard Cauchy distribution, which does not have any finite moments, results from the ratio of standard normal random variables). Also, note that the pooled estimate $\widehat{\sigma}^2_\epsilon$, Equation (3.5), is a weighted average of the estimates of σ^2_ϵ from the first and second stages of the calibration experiment.

The sampling distribution of $\widehat{x}_0$ is quite complicated. Fortunately, its derivation is not necessary for setting a $100(1-\alpha)\%$ confidence interval on x_0 (see Section 3.3). Nonetheless, the resulting distribution has been studied by Fieller (1932), Hinkley (1969), Buonaccorsi (1986), and Pham-Gia et al. (2006), among others. The paper by Pham-Gia et al. (2006) gives a closed-form expression for the density of the ratio of jointly normal random variables. Buonaccorsi (1986) showed that a sufficient condition for unimodality of the sampling distribution of $\widehat{x}_0$ is

$$(x_0 - \bar{x})^2 < 5.094 \left(\frac{\sigma}{\beta_1}\right)^2 \left(\frac{1}{m} + \frac{1}{n}\right).$$

A similar result also holds for the posterior of x_0 in the Hunter and Lamboy (1981) approach to Bayesian linear calibration (see Section 3.4). Unimodality here is important since some confidence intervals (e.g. the Wald interval of Section 3.3.2) are derived under the assumption that the sampling distribution of $\widehat{x}_0$ is asymptotically normal.

3.2.2 The inverse estimator.

The classical estimator involves regressing y on x and solving the fitted regression equation for the unknown x_0. The inverse estimator, however, uses the regression of x on y to obtain

$$\widetilde{x}_0 = \widehat{\gamma}_0 + \widehat{\gamma}_1 \bar{y}_0 = \bar{x} + \frac{S_{xy}}{S_{yy}}(\bar{y}_0 - \bar{y}), \tag{3.6}$$

where $\widehat{\gamma}_0$ and $\widehat{\gamma}_1$ are least squares (LS) estimates, and $S_{yy} = \sum(y_i - \bar{y})^2$. After a bit of algebra, Equation (3.6) can be re-written as

$$\widetilde{x}_0 = \left(1 - \frac{S_{xy}^2}{S_{xx}S_{yy}}\right)\bar{x} + \frac{S_{xy}^2}{S_{xx}S_{yy}}\widehat{x}_0 = \left(1 - R^2\right)\bar{x} + R^2\widehat{x}_0, \tag{3.7}$$

where R^2 is the coefficient of determination computed from (3.2). This shows $\widetilde{x}_0$ as a weighted average of the ML estimator $\widehat{x}_0$ and $\bar{x}$; hence, the inverse estimator takes into account previous information about x_0 (this is relevant to Section 3.4 where the inverse estimator is shown to be Bayes with respect to a certain prior on x_0). Also, when the error variance is zero, $R^2 = 1$ and $\widetilde{x}_0 = \widehat{x}_0$.

Inference based on this approach, at least from the frequentist perspective, is justifiable only if the observations (x_i, y_i) are sampled from a bivariate distribution (i.e., a natural calibration experiment). In controlled calibration experiments, x is held fixed and the classical estimator is mostly preferred. Nonetheless, much effort has gone into justifying the use of the inverse estimator in controlled calibration (see, for example, Krutchkoff (1967) and more recently Kannan et al. (2007)).

3.2.3 Criticisms and other estimators.

Using extensive Monte Carlo simulations, Krutchkoff (1967) argued that the inverse estimator (3.6) had a uniformly smaller Mean squared error (MSE) than the classical estimator (3.3). His experiments considered both normal and non-normal error distributions. This sparked controversy in the statistical community resulting in a renewed interest in the inverse estimator for controlled calibration. In a later paper (Krutchkoff, 1969), Krutchkoff concluded that the classical approach is superior for large sample sizes when extrapolating beyond the range of observations.

When the error distribution is normal and $n \geq 4$, the inverse estimator has finite MSE (Oman, 1985); whereas the classical estimator has infinite MSE for finite n (Williams, 1969). Furthermore, Williams (1969) argued that MSE is an inappropriate criterion for comparing the two estimators by establishing that "...no

unbiased estimator [of x_0] will have finite variance." Halperin (1970) agreed with Williams and suggested the use of the *Pitman closeness* criterion (Pitman Hobart 1937; Mood et al. 1974, pg. 290) instead. For a parameter ϕ with parameter space Φ, an estimator T_1 is said to be Pitman closer (to ϕ) than another estimator T_2 if for all $\phi \in \Phi$,

$$\Pr_\phi\left(|T_1 - \phi| < |T_2 - \phi|\right) > 0.5. \tag{3.8}$$

Unlike the MSE criterion, Pitman closeness takes into consideration the correlation between the two estimators being compared (here, they are perfectly correlated). With respect to the Pitman closeness criterion, Halperin claimed that the classical estimator is superior both outside and within the range of predictor values; though, Halperin's conclusions were based on asymptotic approximations. In addition, Halperin also compared the two estimators in terms of consistency and MSE of the relevant asymptotic distributions, showing that Krutchkoff's findings were only true for a closed interval around $\bar{x}$. The width of this interval depends on the product of the standardized slope and standard deviation of the predictor values, σ_x. Halperin also preferred the classical estimator on the basis that it yields an exact $100(1-\alpha)\%$ confidence region for x_0 (see Section 3.3.1). In response, Krutchkoff (1972) used the Pitman closeness criterion in Monte Carlo simulations and obtained the opposite results; that is, the inverse estimator was overall superior to the classical estimator. The contradiction seems to stem from the range of predictor values considered by both authors.

The classical and inverse estimators cannot be compared through exact moments; however, one can easily examine their asymptotic properties. Berkson (1969) established that the classical estimator is asymptotically unbiased while the inverse estimator is not. He derives formulas for the asymptotic bias, variance, and MSE of

both estimators. Shukla (1972) extended these formulas to an accuracy of $O(n^{-1})$:

$$
\begin{aligned}
\text{Bias}\left\{\widehat{x}_0\right\} &\approx \frac{\sigma_\epsilon^2}{S_{xx}\beta_1^2}(x_0-\bar{x}) \\
\text{Var}\left\{\widehat{x}_0\right\} &\approx \frac{\sigma_\epsilon^2}{\beta_1^2}\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}+\frac{3\sigma_\epsilon^2}{mS_{xx}\beta_1^2}\right] \\
\text{MSE}\left\{\widehat{x}_0\right\} &\approx \frac{\sigma_\epsilon^2}{\beta_1^3}\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}+\frac{3\sigma_\epsilon^2}{mS_{xx}\beta_1^2}\right] \\
\text{Bias}\left\{\widetilde{x}_0\right\} &\approx \frac{\sigma_\epsilon^2}{\beta_1^2\sigma_x^2\theta}(\bar{x}-x_0)-\frac{2\sigma_\epsilon^2(\bar{x}-x_0)}{n\beta_1^2\sigma_x^2\theta^3} \\
\text{Var}\left\{\widetilde{x}_0\right\} &\approx \frac{\sigma_\epsilon^2}{\beta_1^2\theta^2}\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}+\frac{\sigma_\epsilon^2(\theta^2-2\theta+6)}{mS_{xx}\beta_1^2\theta^2}\right] \\
&\quad -\frac{2\sigma^4(x_0-\bar{x})^2}{n\theta^4\beta_1^4\sigma_x^4} \\
\text{MSE}\left\{\widetilde{x}_0\right\} &\approx \frac{\sigma_\epsilon^2}{\beta_1^2\theta^2}\left[\frac{1}{m}+\frac{1}{n}+\frac{(x_0-\bar{x})^2}{S_{xx}}+\frac{\sigma_\epsilon^2(\theta^2-2\theta+6)}{mS_{xx}\beta_1^2\theta^2}\right] \\
&\quad -\frac{\sigma^4(x_0-\bar{x})^2}{\theta^2\beta_1^4\sigma_x^4}\left(1-\frac{6}{n\theta^2}\right),
\end{aligned}
$$

where $\sigma_x^2 = S_{xx}/(n-1)$ and $\theta = 1+\sigma_\epsilon^2/(\beta_1\sigma_x)^2$. Lwin (1981) provided a further extension by allowing the error distribution to be any member of the location-scale family of distributions. To an accuracy of $O(n^{-1})$, Lwin showed that the asymptotic MSE of the classical estimator for any error distribution from the location-scale family is the same as that obtained by Shukla (1972). The same is not true for the inverse estimator whose MSE to the same order is affected by both the skewness and kurtosis of the error distribution.

Kannan et al. (2007) obtained more accurate results supporting the use of the inverse estimator. However, they found that when $|\beta_1/\sigma|$ is moderate to large, the classical method is preferred in the sense of Pitman closeness. Ali and Ashkar (2002) discussed the impact of the coefficient of determination and proposed an estimator based on the midpoint of the *inversion interval* for x_0 (see Section 3.3.1). Lwin and Maritz (1982) take a *compound estimation* approach to the linear calibration problem. They discuss the merits of both estimators and provide further justification for each

without reference to specific distributional assumptions. Lwin and Maritz concluded that the classical estimator is preferred only if the condition of asymptotic unbiased is imposed. A likelihood analysis for the calibration problem was carried out by Minder and Whitney (1975).

A number of other estimators have been proposed in the literature; though, none have received the level of attention as the classical and inverse estimators. Ali and Singh (1981), for example, used a weighted average of the classical estimator $\widehat{x}_0$ and $\bar{x}$. The idea is to shrink the estimate toward $\bar{x}$ when x_0 is near the center of the data. Using small-disturbance asymptotic approximations, Srivastava and Singh (1989) derived an estimator that is a weighted average of the classical and inverse estimators: $\widehat{\xi} = [\widehat{x}_0 + (n-3)\widetilde{x}_0]/(n-2)$. This estimator gives more weight to the inverse estimator $\widetilde{x}_0$ when the sample size is large and vice versa. In terms of asymptotic bias, $\widehat{\xi}$ is superior to the classical and inverse estimators. Naszdi (1978) proposed an estimator that is approximately (asymptotically) unbiased, consistent, and more efficient than the classical estimator. Dahiya and McKeon (1991) obtain a confidence interval for x_0 based on the estimator in Naszdi (1978).

Many papers comparing the classical and inverse estimators have been published, none of which offer a definitive answer on which estimator is best. If point estimation is the only concern (which is rarely the case), both estimators are of value. On the other hand, if inference is to be made regarding x_0, the classical estimator is preferred for controlled calibration experiments and the inverse estimator for natural calibration experiments. Of course, if suitable prior information can be assembled, then a Bayesian approach may be the best alternative in either case, especially if the calibration curve is nonlinear (see Section 3.4).

Finally, note that most of the previous discussion was in reference to the linear calibration problem with homoscedastic normal errors. The inverse estimator is less

appropriate for nonlinear calibration curves, especially when the calibration curve has horizontal asymptotes (Jones and Rocke, 1999)

3.2.4 Arsenic example.

Here we illustrate the use of the classical and inverse estimators on the arsenic data, for which the simple linear regression model with homoscedastic normal errors seems appropriate. This is a controlled calibration experiment since the true concentrations of arsenic were preselected by the experimenter. We wish to estimate the true concentration of arsenic in a new sample based on the new observation $y_0 = 3$ μg/ml. Figure 3.1 shows a scatterplot of the standards with the fitted calibration line. The ML estimate of x_0 is 2.941 μg/ml while the inverse estimate is 2.945 μg/ml, a difference of only 0.004. In practice, the two estimators will not be much different when the coefficient of determination, R^2, is close to one. For the arsenic data, $R^2 = 0.9936$ (recall Equation 3.6).

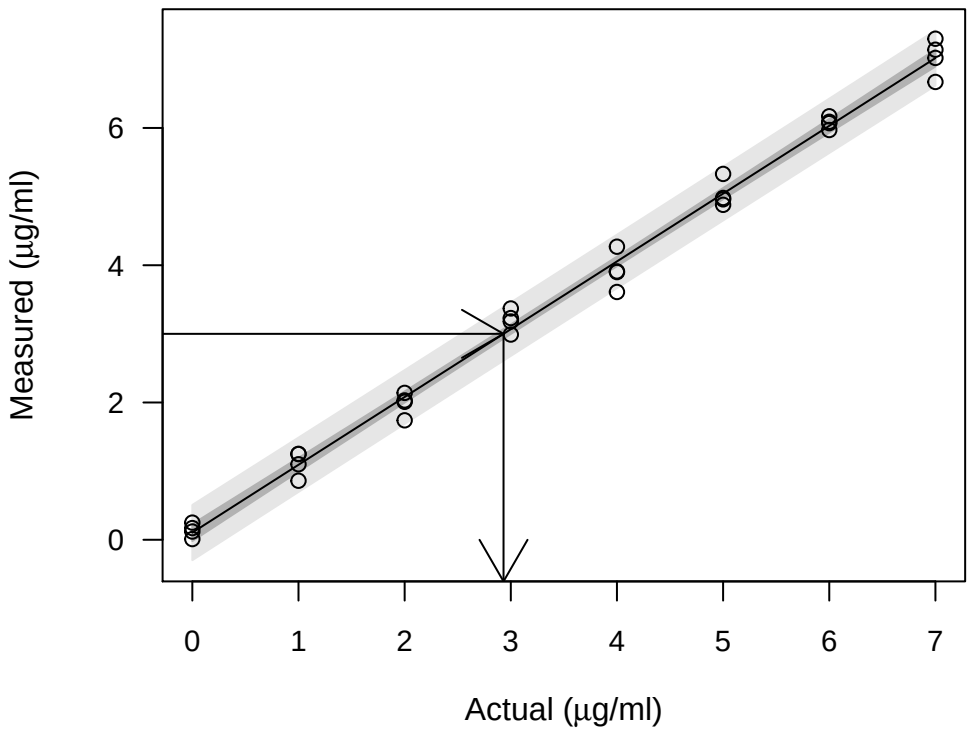


Figure 3.1: Fitted calibration line for the arsenic data. The horizontal arrow indicates the position of the observed response $y_0 = 3$ μg/ml and the vertical arrow indicates the position of the ML estimate $\widehat{x}_0 = 2.941$.

3.3 Confidence intervals

Much effort has been put into deriving and comparing point estimators for x_0. Without some measure of precision, however, a point estimate is practically useless. In this section, we discuss construction of $100(1-\alpha)\%$ confidence intervals for x_0, also known as *calibration intervals*. There are two methods commonly used for calculating calibration intervals (Zeng and Davidian, 1997b): *inversion intervals* and *Wald intervals*, additionally, we also discuss bootstrap calibration intervals. A discussion on Bayesian credible intervals is deferred until Section 3.4. For the most part, we assume the regression model (3.1) with normal errors and constant variance is appropriate and that the predictor values are fixed by design.

3.3.1 Inversion interval.

Similar to inverting the calibration curve to obtain a point estimate, a confidence interval for x_0 can be constructed by inverting a prediction interval for the response. We refer to this type of calibration interval as the inversion interval. The inversion interval relies on the distribution of the *predictive pivot*, $\mathcal{Q}$, which for the linear calibration problem is given by

$$\mathcal{Q} = \frac{\overline{\mathcal{Y}}_0 - \widehat{\beta}_0 - \widehat{\beta}_1 x_0}{\sqrt{\widehat{\sigma}_\epsilon^2 \left[\frac{1}{m} + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}\right]}}. \tag{3.9}$$

It can be shown (Graybill, 1976) that $\mathcal{Q} \sim \mathcal{T}(n+m-3)$, hence,

$$\Pr\left(t_{\alpha/2,n+m-3} < \mathcal{Q} < t_{1-\alpha/2,n+m-3}\right) = 1-\alpha.$$

Squaring both sides of the inequality and expanding, we obtain a simple quadratic in x_0:

$$\Pr\left(ax_0^2 + bx_0 + c < 0\right) = 1-\alpha, \tag{3.10}$$

where

$$\begin{aligned}
a &= \widehat{\beta}_1^2 - \widehat{\sigma}_\epsilon^2 t^2 / S_{xx} \\
b &= 2\left[\frac{\bar{x}\widehat{\sigma}_\epsilon^2 t^2}{S_{xx}} - \widehat{\beta}_1\left(\overline{\mathcal{Y}}_0 - \overline{\mathcal{Y}}\right) - \widehat{\beta}_1^2 \bar{x}\right] \\
c &= \left[\left(\overline{\mathcal{Y}}_0 - \widehat{\beta}_0\right)^2 - \widehat{\sigma}_\epsilon^2 t^2 \left(\frac{1}{m} + \frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)\right] \\
t &= t_{1-\alpha/2,n+m-3}.
\end{aligned}$$

An exact $100(1-\alpha)\%$ confidence interval for x_0 is given by the set

$$\mathcal{J}_{\text{cal}}(x) = \left\{x : ax^2 + bx + c < 0\right\}. \tag{3.11}$$

Although we use the term confidence *interval* here, it is possible that the values of x that satisfy this inequality do not form an actual interval. In particular, four possibilities exist:

(i) the set is a finite interval, $\mathcal{J}_{\text{cal}}(x) = (L, U)$;

(ii) the set is the entire real line, $\mathcal{J}_{\text{cal}}(x) = (-\infty, \infty)$;

(iii) the set consists of two semi-infinite intervals, $\mathcal{J}_{\text{cal}}(x) = (-\infty, U) \cup (L, \infty)$;

(iv) the set is empty, $\mathcal{J}_{\text{cal}}(x) = \emptyset$.

The circumstances leading to (i)-(iv) are depicted in Figure 3.2. A finite interval (i) will occur if and only if $a > 0$ and $b^2 - 4ac > 0$. Upon closer inspection of the quadratic in Equation (3.10), we see that this occurs when $a > 0$ or rather when $\widehat{\beta}_1^2/(\widehat{\sigma}_\epsilon^2/S_{xx}) > t^2$. In other words, when the slope $\widehat{\beta}_1$ is significantly different from zero at the specified α level (i.e., the regression line is not too flat), the solution to Equation (3.11) forms a $100(1-\alpha)\%$ confidence interval for x_0 and is given by

$$\bar{x} + \frac{\widehat{\beta}_1(\bar{y}_0 - \bar{y})}{a} \pm \frac{t\widehat{\sigma}}{a} \times \sqrt{a\left(\frac{1}{m} + \frac{1}{n}\right) + \frac{(\bar{y}_0 - \bar{y})^2}{S_{xx}}}. \tag{3.12}$$

or equivalently

$$\widehat{x}_0 + \frac{(\widehat{x}_0 - \bar{x})g \pm \left(t\widehat{\sigma}/\widehat{\beta}_1\right)\sqrt{(\widehat{x}_0 - \bar{x})^2/S_{xx} + (1-g)\left(\frac{1}{m} + \frac{1}{n}\right)}}{1-g},$$

where $g = \left(t^2\widehat{\sigma}_\epsilon^2\right)/\left(\widehat{\beta}_1^2 S_{xx}\right)$. The dependence on a statistically significant slope is generally not regarded as a concern since "any self-respecting calibrator will design a calibration experiment such that his or her instrument is expected to have a statistically significant slope ..." (Brown, 1993, pp. 25). Hoadley (1970) noted that the width of interval (3.12) depends on the value of the observed test statistic, $\mathcal{F}^\star = \widehat{\beta}_1^2 S_{xx}/\widehat{\sigma}_\epsilon^2 = t^2$, for testing $\mathcal{H}_0 : \beta_1 = 0$ versus $\mathcal{H}_1 : \beta_1 \neq 0$. A large value of $\mathcal{F}^\star$ is associated with a smaller interval and vice versa. An approximation can be obtained by setting $g = 0$ in Equation (3.12), which gives

$$\widehat{x}_0 \pm t_{1-\alpha/2,n+m-3}\left(\frac{\widehat{\sigma}}{\widehat{\beta}_1}\right)^2 \sqrt{\frac{1}{m} + \frac{1}{n} + \frac{(\widehat{x}_0 - \bar{x})^2}{S_{xx}}}. \tag{3.13}$$

This approximation is useful when g is small, say $g < 0.05$ (Draper and Smith, 1998). Fieller's method (Fieller, 1954), which applies to the ratio of normally distributed random variables, can also be used to derive interval (3.12) using a fiducial argument. Extending the inversion interval to the case of multiple predictors is discussed in Draper and Smith (1998, pg. 229) and in Brown (1993, chap. 3). Brown considers the special case of polynomial regression.

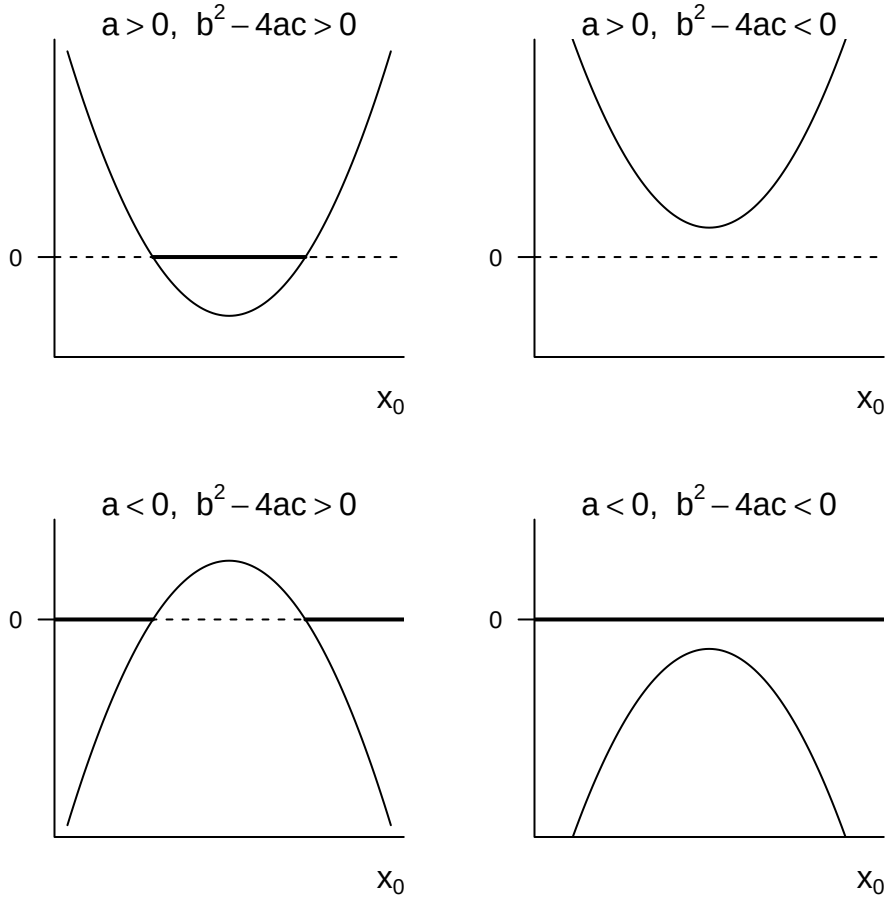


Figure 3.2: Solutions to the equation $ax^2 + bx + c < 0$. *Top left*: The set is an interval. *Top right*: The set is empty. *Bottom left*: The set consists of two semi-infinite intervals. *Bottom right*: The set is the entire real line.

For nonlinear calibration curves,

$$\mathcal{Q} = \frac{\overline{\mathcal{Y}}_0 - \mu\left(x_0; \widehat{\boldsymbol{\beta}}\right)}{\sqrt{\widehat{\sigma}_\epsilon^2/m + \widehat{\mathrm{Var}}\left\{\mu\left(x_0; \widehat{\boldsymbol{\beta}}\right)\right\}}} \tag{3.14}$$

is only an approximate pivot. We assume $\mathcal{Q} \sim \mathcal{N}(0,1)$ as n goes to ∞. An approximate $100(1-\alpha)\%$ confidence interval for x_0 based on Equation (3.14) is the set

$$\widehat{\mathcal{J}}_{\mathrm{cal}}(x) = \left\{ x : z_{\alpha/2} < \frac{\bar{y}_0 - \mu\left(x; \widehat{\boldsymbol{\beta}}\right)}{\sqrt{\widehat{\sigma}_\epsilon^2/m + \widehat{\mathrm{Var}}\left\{\mu\left(x; \widehat{\boldsymbol{\beta}}\right)\right\}}} < z_{1-\alpha/2} \right\}, \tag{3.15}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the standard normal distribution, respectively. To be more conservative, we can replace the normal quantiles with those of a $\mathcal{T}(n+m-p-1)$ distribution (p being the dimension $\boldsymbol{\beta}$). Unlike the linear calibration problem, the solution to Equation (3.15) cannot be written in closed-form and will require iterative techniques.

For the special case $m = 1$, the inversion interval is equivalent to drawing a horizontal line through the scatterplot of the standards at y_0 and finding the abscissas of its intersection with the $100(1-\alpha)\%$ (pointwise) prediction band of the calibration curve. For the straight line case, if β_1 is not significantly different from zero, the regression line is not well determined and the horizontal line drawn at y_0 will not intersect the prediction band at two points, leading to cases (ii) or (iii) outlined above; this point is illustrated in Figure 3.3. The issue in linear calibration is that prediction bands are really hyperbolas which, depending on the quality of the model/data, can bend quite severely. To circumvent this problem, Trout and Swallow (1979) proposed a similar procedure based on uniform prediction bands (i.e., prediction bands that are parallel to the calibration line). This has the advantage of always producing a (symmetric) confidence interval for x_0, without sacrificing efficiency.

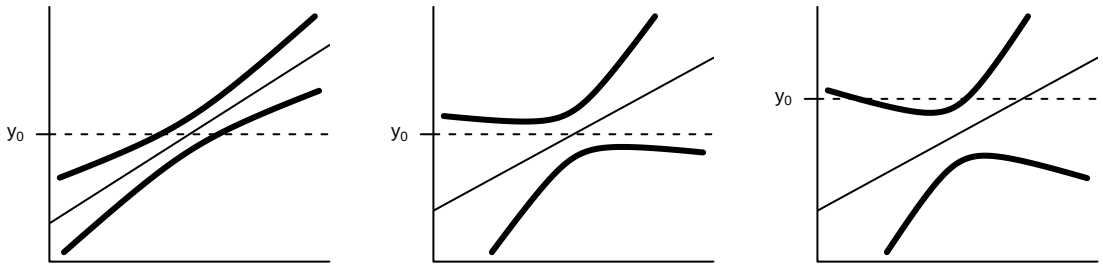


Figure 3.3: Common prediction band shapes. *Left*: Horizontal line at y_0 intersects the prediction band at two points, resulting in a finite interval. *Middle*: Horizontal line at y_0 does not intersect the prediction band at all resulting in an infinite interval. *Right*: Horizontal line at y_0 only intersects with one side of the prediction band resulting in two semi-infinite intervals.

Similarly, we can compute the inversion interval for nonlinear calibration problems by drawing the prediction band, however, inference in nonlinear regression often relies on linear approximations, large samples, and approximate normality. For nonlinear calibration curves, the bootstrap (see Section 3.3.3) may provide more accurate results. Drawing the prediction band for calibration curves also gives an idea as to what values of y_0 lead to meaningful interval estimates for x_0. For example, an observed value y_0 too close to a horizontal asymptote will produce a useless confidence interval for x_0 (if at all). Graphical methods like this are discussed by Jones and Lyons (2009) who extend the approach for longitudinal data with bivariate response.

3.3.2 Wald interval.

Another common approach for obtaining a confidence interval for the unknown x_0 is to use the delta method (VerHoef, 2012; Dorfman, 1938), also see Casella and Berger (2002). Let the variance-covariance matrix of $\left(\overline{\mathcal{Y}}_0, \widehat{\boldsymbol{\beta}}\right)'$ be given by $\boldsymbol{\Sigma}$ where

$$\boldsymbol{\Sigma} = \begin{bmatrix} \text{Var}\left\{\overline{\mathcal{Y}}_0\right\} & \text{Cov}\left\{\overline{\mathcal{Y}}_0, \widehat{\boldsymbol{\beta}}\right\} \\ \text{Cov}\left\{\overline{\mathcal{Y}}_0, \widehat{\boldsymbol{\beta}}\right\} & \text{Var}\left\{\widehat{\boldsymbol{\beta}}\right\} \end{bmatrix}.$$

Recall that $\widehat{\boldsymbol{\beta}}$ is a linear function of the observations $\boldsymbol{\mathcal{Y}}$. Since $\overline{\mathcal{Y}}_0$ is independent of $\boldsymbol{\mathcal{Y}}$, it follows that $\overline{\mathcal{Y}}_0$ and $\widehat{\boldsymbol{\beta}}$ are also independent; hence, $\text{Cov}\left\{\overline{\mathcal{Y}}_0, \widehat{\boldsymbol{\beta}}\right\} = \mathbf{0}_{p\times p}$. Therefore, $\boldsymbol{\Sigma}$ simplifies to

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\epsilon^2/m & \mathbf{0}_{p\times p} \\ \mathbf{0}_{p\times p} & \sigma_\epsilon^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \end{bmatrix}.$$

The classical estimator, $\widehat{x}_0$, has the form $x = \mu^{-1}\left(y; \boldsymbol{\beta}\right)$. Let $\mu_1^{-1}\left(y; \boldsymbol{\beta}\right) = \frac{\partial}{\partial y}\mu^{-1}\left(y; \boldsymbol{\beta}\right)$ and $\mu_2^{-1}\left(y; \boldsymbol{\beta}\right) = \frac{\partial}{\partial \boldsymbol{\beta}}\mu^{-1}\left(y; \boldsymbol{\beta}\right)$. Note that if p (the dimension of $\boldsymbol{\beta}$) is greater than one, then $\mu_2^{-1}\left(y; \boldsymbol{\beta}\right)$ will be a vector valued function. The delta method estimate of the variance of $\widehat{x}_0 = \mu^{-1}\left(\overline{\mathcal{Y}}_0; \widehat{\boldsymbol{\beta}}\right)$, based on a first-order Taylor series expansion, is given by

$$\widehat{\text{Var}}\left\{\widehat{x}_0\right\} = \frac{\widehat{\sigma}_\epsilon^2}{m}\left[\mu_1^{-1}\left(\overline{\mathcal{Y}}_0; \widehat{\boldsymbol{\beta}}\right)\right]^2 + \widehat{\sigma}_\epsilon^2\left[\mu_2^{-1}\left(\overline{\mathcal{Y}}_0; \widehat{\boldsymbol{\beta}}\right)\right]'\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left[\mu_2^{-1}\left(\overline{\mathcal{Y}}_0; \widehat{\boldsymbol{\beta}}\right)\right]. \tag{3.16}$$

For the simple linear calibration problem, Equation (3.16) reduces to

$$\widehat{\text{Var}}\left\{\widehat{x}_0\right\} = \frac{\widehat{\sigma}_\epsilon^2}{\widehat{\beta}_1^2}\left[\frac{1}{m} + \frac{1}{n} + \frac{(\widehat{x}_0 - \bar{x})^2}{S_{xx}}\right].$$

The estimated standard error (se) of $\widehat{x}_0$ is just $\widehat{\text{se}}\left\{\widehat{x}_0\right\} = \sqrt{\widehat{\text{Var}}\left\{\widehat{x}_0\right\}}$.

Assuming that

$$\mathcal{W} = \frac{\widehat{x}_0 - x_0}{\widehat{\text{se}}\left\{\widehat{x}_0\right\}} \dot{\sim} \mathcal{N}(0,1) \tag{3.17}$$

for "large" n leads to an approximate $100(1-\alpha)\%$ Wald-based confidence interval for x_0 of

$$\widehat{x}_0 \pm t_{1-\alpha/2, n+m-p-1} \cdot \widehat{\text{se}}\left\{\widehat{x}_0\right\}, \tag{3.18}$$

where $t_{1-\alpha/2,n+m-p-1}$ is the $1-\alpha/2$ quantile of a Student's t-distribution with $n+m-p-1$ degrees of freedom. If the sample size is large enough, we could replace $t1-\alpha/2n+m-p-1$ with $z_{1-\alpha/2}$, the $1-\alpha/2$ quantile of a standard

normal distribution. Unlike the inversion interval, Equation (3.15), the Wald-based interval always exists and is symmetric about the point estimate $\widehat{x}_0$. This approach is useful when it is difficult (or impossible) to invert a corresponding prediction interval. The symmetry of the Wald interval, however, may be unrealistic in nonlinear calibration problems when, for example, $\bar{y}_0$ is near a horizontal asymptote (Schwenke and Millikem, 1991). Perhaps the biggest drawback is the approximate normality assumption for $\mathcal{W}$, which is not always reasonable in practice. For example, in simple linear calibration, the distribution of $\widehat{x}_0$ may not even be unimodal (Buonaccorsi, 1986). On the other hand, for the simple linear calibration problem, the Wald-based interval is equivalent to the approximate inversion interval given in Equation (3.13). Thus, provided g is "small," the Wald interval may perform well even when $\mathcal{W}$ is not asymptotically normal.

Schwenke and Millikem (1991) discuss the inversion and Wald confidence intervals for nonlinear calibration. Using simulation, they showed that the inversion and Wald intervals can attain the desired confidence level in samples of size 20 for a nonlinear exponential decay model. They also discuss testing the equality of two calibration points. However, Schwenke and Millikem treat $\mathcal{Y}_0$ as a fixed constant. In regulation-type problems, this is the correct approach. However, in calibration, the confidence interval for x_0 will be too narrow. To illustrate, consider the conventional treatment group of the postmortem data analyzed in Schwenke and Millikem (1991). In essence, Schwenke and Millikem computed the time corresponding to a mean pH level of 6.0, not an observed pH level of 6.0. They list a 95% Wald-based confidence interval for x_0 of $(3.315, 4.317)$. Accounting for the correct variation, however, this interval should actually be $(2.615, 5.016)$. The standard error they computed for $\widehat{x}_0$ is based on $\boldsymbol{T} = (\widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)'$ whereas here it is based on $\boldsymbol{T} = (\mathcal{Y}_0, \widehat{\beta}_0, \widehat{\beta}_1, \widehat{\beta}_2)'$. In short, Schwenke and Millikem only account for the variance of the estimated regression parameters

whereas in a true calibration problem, the variance of $\mathcal{Y}_0$ needs to be taken into account.

There is an interesting relationship in the linear calibration problem (with $m = 1$) between the Wald interval for x_0, interval (3.18), and a prediction interval for $\mathcal{Y}_0$. Let L_w and U_w denote the lower and upper bounds, respectively, from a $100(1-\alpha)\%$ Wald confidence interval for x_0 corresponding to an observed y_0. If $\widehat{x}_0$ is assumed given, then it is easy to show that a $100(1-\alpha)\%$ prediction interval for $\mathcal{Y}_0$ is $(\widehat{\beta}_0+\widehat{\beta}_1 L_w, \widehat{\beta}_0+\widehat{\beta}_1 U_w)$ if $\widehat{\beta}_1 > 0$ and $(\widehat{\beta}_0 + \widehat{\beta}_1 U_w, \widehat{\beta}_0 + \widehat{\beta}_1 L_w)$ if $\widehat{\beta}_1 < 0$. This relationship is illustrated in Figure 3.4 for the arsenic example.

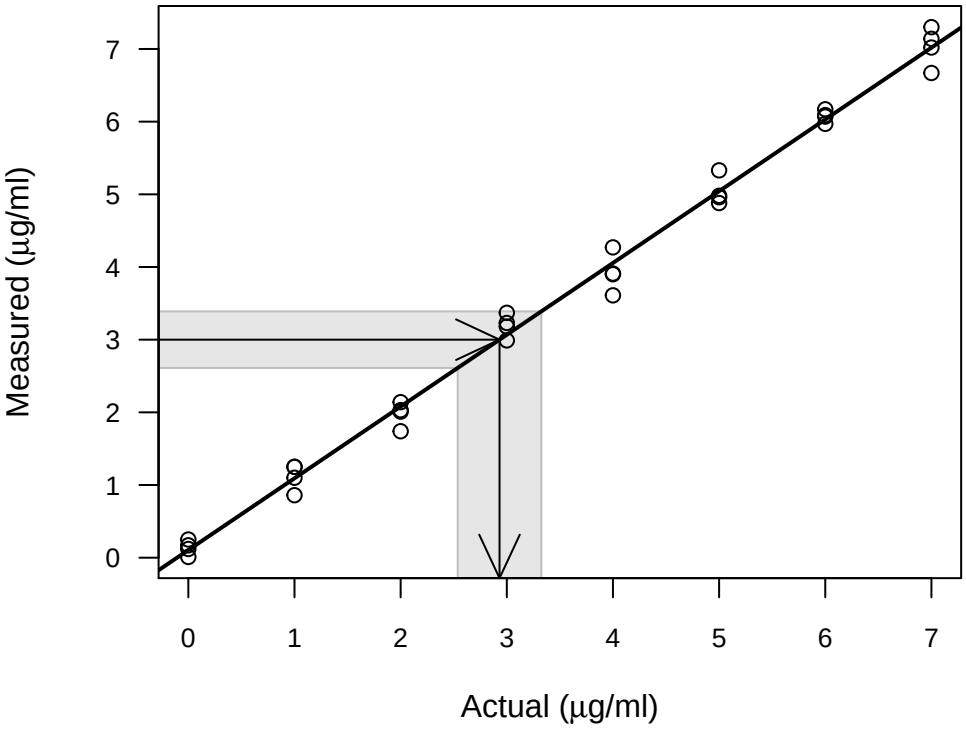


Figure 3.4: Scatterplot of the arsenic data with fitted calibration line. The horizontal arrow represents the response measurement for the new sample and the vertical arrow represents the ML estimate of the unknown concentration. The vertical gray band represents a 95% Wald confidence interval for x_0 corresponding to $y_0 = 3$ and the horizontal gray band represents a 95% prediction interval for $\mathcal{Y}_0$ corresponding to $x_0 = \widehat{x}_0$.

3.3.3 Bootstrap intervals.

In many applications, the calibration curve is usually nonlinear (e.g., dose-response curves and other types of assay data). Although the inversion and Wald procedures can still be applied, they can be highly inaccurate when the sample size is small. A useful alternative in these situations is to use the nonparametric bootstrap (Efron, 1979), which is a computer intensive technique based on sampling with replacement from the observed data. As such, the bootstrap provides an alternative means to computing bias, standard errors, and confidence intervals. Unlike the delta method, however, which is only first-order accurate, the bootstrap can often have "second-order" accuracy (Casella and Berger, 2002, pg. 517). The bootstrap is not without assumptions (e.g., independent observations are still required), but it does allow us to relax the usual assumptions of large sample size and normality. The supporting mathematics can be quite involved, but the interested reader is pointed to Efron and Tibshirani (1994) and Hall (1992). A detailed and practical guide to the bootstrap is given by Davison and Hinkley (1997).

Let $\widehat{\mu}_i = \mu\left(x_i; \widehat{\boldsymbol{\beta}}\right)$ be the *fitted values* and $\widehat{\boldsymbol{\beta}}$ be the least squares estimate of $\boldsymbol{\beta}$. The two (nonparametric) bootstrap resampling schemes for regression are:

case resampling: pairs of data are sampled with replacement to produce

$(x_1, y_1)^\star, \ldots, (x_n, y_n)^\star$;

model-based resampling: the residuals $e_1, \ldots, e_n$, or a modified version thereof, are sampled with replacement and added to the fitted values to produce

$(x_1, \widehat{\mu}_1 + e_1^\star), \ldots, (x_n, \widehat{\mu}_n + e_n^\star)$.

Efron and Tibshirani (1994) and Davison and Hinkley (1997) discuss the merits of both schemes. Model-based resampling is more appropriate when the values of the independent variable are fixed by design (e.g., controlled calibration experiments) and the errors have constant variance. Resampling cases, on the other hand,

is less accurate but more robust to violations of the model assumptions such as homoscedastistic errors.

Various bootstrap approaches to calibration have been suggested in the literature. Rosen and Cohen (1995b) discuss controlled calibration based on a parametric bootstrap where, instead of sampling directly from the residuals, samples are drawn from a normal distribution. This procedure will work even when the calibration curve is estimated nonparametrically (e.g., *cubic smoothing splines*). Zeng and Davidian (1997a) commented that the procedure proposed by Rosen and Cohen requires a large number of resamples to achieve the desired accuracy. They suggested a bootstrap adjustment to the inversion and Wald intervals that requires far fewer bootstrap resamples. Given the speed and multicore functionality of modern computers, a large number of resamples, say 10,000 or more, is less of a concern.

In this section, we outline a general bootstrap approach to controlled calibration in Algorithm 1, but first we discuss a rather naive approach. For an observed $\bar{y}_0$, suppose we compute R bootstrap replicates of $\widehat{x}_0$, and then use, for example, the sample $\alpha/2$ and $1-\alpha/2$ quantiles as a $100(1-\alpha)\%$ confidence interval for x_0. This results in $\bar{y}_0$ being treated as a fixed parameter in the bootstrap simulation, but in fact, $\bar{y}_0$ is an observed value of the random variable $\overline{\mathcal{Y}}_0$ which has variance σ^2_ϵ/m. In other words, some variation is not getting accounted for in this approach. This is akin to ignoring Var $\left\{\overline{\mathcal{Y}}_0\right\}$ in the delta method discussed previously. One solution is to simulate the correct variance by adding a small amount of noise to the observed responses from the second stage of the calibration experiment.

A few issues regarding the residuals in Algorithm 1 are worth considering. For the linear case, it is preferable to scale the residuals before centering (Davison and Hinkley, 1997). In particular, compute $r_i = e_i/\sqrt{1-h_{ii}}$, where h_{ii} are the diagonal elements of the hat matrix, $\boldsymbol{H} = \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}$. We then sample $\epsilon^\star_j$ from the modified

Algorithm 1: Model-based resampling for controlled calibration.

for $r = 1$ **to** R **do**

(1) **for** $i = 1$ **to** n **do**

(a) set $x_i^\star = x_i$;

(b) randomly sample $\epsilon_i^\star$ from the centered residuals $e'_1, \ldots, e'_n$;

(c) set $y_i^\star = \mu\left(x_i; \widehat{\boldsymbol{\beta}}\right) + \epsilon_i^\star$;

end

(2) Fit model to $(x_1^\star, y_1^\star), \ldots, (x_n^\star, y_n^\star)$, giving estimates $\widehat{\boldsymbol{\beta}}_r^\star$, $\widehat{\sigma}_r^{2\star}$;

(3) **for** $j = 1$ **to** m **do**

(a) randomly sample $\epsilon_j^\star$ from the centered residuals $e'_1, \ldots, e'_n$;

(b) set $y_{0j}^\star = \bar{y}_0 + \epsilon_j^\star$;

end

(4) Set $\bar{y}_{0r}^\star = \sum_{k=1}^m y_{0k}^\star / m$;

(5) Set $\widehat{x}_{0r}^\star = \mu^{-1}\left(\bar{y}_{0r}^\star; \widehat{\boldsymbol{\beta}}_r^\star\right)$;

if calculating an interval based on the studentized bootstrap, then additionally:

(6) Compute $\widehat{\text{se}}\left\{\widehat{x}_{0r}^\star\right\}$;

(7) Set $\mathcal{W}_r^\star = (\widehat{x}_{0r}^\star - \widehat{x}_0)/\widehat{\text{se}}\left\{\widehat{x}_{0r}^\star\right\}$;

end

residuals $r_1 - \bar{r}, \ldots, r_n - \bar{r}$. This transforms the residuals so that they are centered and have constant variance σ_ϵ^2. For nonlinear calibration curves, the residuals should be corrected for bias in addition to centering them (Davison and Hinkley, 1997). When there are outliers in the residuals, the bootstrap distribution of $\widehat{x}_0$ can become skewed or multimodal. One suggestion is to replace the residuals in step (3) with

something smoother, say random variates from a normal distribution. For the case $m > 1$, Jones and Rocke (1999) suggest using a "residual pool" formed by $e_1, \ldots, e_n$ plus residuals computed from the unknowns: $e_{n+j} = y_{n+j} - \bar{y}_0$ $(j = 1, 2, \ldots, m)$. In Section 3.2, we mentioned that it is often reasonable to assume $\sigma_{\mathrm{I}}^2 = \sigma_{\mathrm{II}}^2$. However, if this assumption is not valid, then we can still proceed by resampling separately from the two sets of residuals $\left\{e_i\right\}_{i=1}^{n}$ and $\left\{e_j\right\}_{j=n+1}^{n+m}$ (Gruet and Jolivet, 1993; Jones and Rocke, 1999). For instance, we can resample from $\left\{e_i\right\}_{i=1}^{n}$ in step (1) and from $\left\{e_j\right\}_{j=n+1}^{n+m}$ in step (3) of Algorithm 1. We can also accommodate nonconstant variance (i.e., heteroscedasticity) by applying the *wild bootstrap* (Davison and Hinkley, 1997, pp. 272). This approach is discussed in Huet et al. (2004, pp. 142).

Once we have our bootstrap replicates, a number of different bootstrap confidence intervals can be constructed. For a studentized interval, we compute R bootstrap replicates of $\mathcal{W}$ (Equation (3.17)): $\mathcal{W}_1^\star, \mathcal{W}_2^\star, \ldots, \mathcal{W}_R^\star$. This is outlined in steps (6)-(7) of Algorithm 1. Instead of relying on normal theory assumptions about $\mathcal{W}$, as when computing the Wald interval, the bootstrap estimates the distribution of $\mathcal{W}$ directly from the data. Thus, instead of approximating the quantiles of $\mathcal{W}$ using a standard table, a "table is built for the data at hand" (Efron and Tibshirani, 1994). An approximate $100(1-\alpha)\%$ confidence interval for x_0 based on the studentized bootstrap is given by

$$\left(\widehat{x}_0 - \gamma_{1-\alpha/2}^\star \cdot \widehat{\mathrm{se}}\left\{\widehat{x}_0\right\}, \widehat{x}_0 - \gamma_{\alpha/2}^\star \cdot \widehat{\mathrm{se}}\left\{\widehat{x}_0\right\}\right), \tag{3.19}$$

where $\gamma_{\alpha/2}^\star$ and $\gamma_{1-\alpha/2}^\star$ are the sample quantiles from the bootstrap distribution of $\mathcal{W}$. One drawback of this approach is that it requires an estimate of the standard error, $\widehat{\mathrm{se}}\left\{\widehat{x}_{0r}^\star\right\}$. We could use a Taylor series approximation, as in the delta method, but this is known to be unreliable for small samples. Another alternative is to use the bootstrap to estimate the standard error of each $\widehat{x}_{0r}^\star$. This implies a second level of bootstrapping nested within the first and can be computationally expensive. The

studentized interval for controlled calibration using the large-sample formula for the standard error (see Equation (3.16)) is discussed by Zeng and Davidian (1997a) and Jones and Rocke (1999). This interval can perform poorly in practice and is easily influenced by outliers in the data.

Yet another technique (Jones and Rocke, 1999; Huet et al., 2004) is to bootstrap the predictive pivot in Equation (3.14):

$$\mathcal{Q}^\star = \frac{\bar{y}_0^\star - \mu\left(\widehat{x}_0; \widehat{\boldsymbol{\beta}}^\star\right)}{\widehat{\text{se}}^\star\left\{\bar{y}_0^\star - \mu\left(\widehat{x}_0; \widehat{\boldsymbol{\beta}}^\star\right)\right\}} = \frac{\bar{y}_0^\star - \mu\left(\widehat{x}_0; \widehat{\boldsymbol{\beta}}^\star\right)}{\sqrt{\widehat{\sigma}_\epsilon^{2\star}/m + \widehat{\text{Var}}^\star\left\{\mu\left(\widehat{x}_0; \widehat{\boldsymbol{\beta}}^\star\right)\right\}}}.$$

It is is easy to see that, for the linear calibration problem, $\mathcal{Q}^\star = \mathcal{W}^\star$; therefore, the resulting interval is the same. For the nonlinear calibration problem, however, a $100(1-\alpha)\%$ confidence interval for x_0 based on the bootstrapping the predictive pivot is given by the set

$$\mathcal{J}_{\text{cal}}^\star(x) = \left\{x : q_{\alpha/2}^\star \leq \mathcal{Q} \leq q_{1-\alpha/2}^\star\right\},$$

where $q_{\alpha/2}^\star$ and $q_{1-\alpha/2}^\star$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the bootstrap distribution of $\mathcal{Q}$, respectively. This mimics the inversion method discussed in Section 3.3.1 but usually performs better since we no longer have to rely on the approximate normality of $\mathcal{Q}$. Therefore, we might think of this as a bootstrap adjusted inversion interval. According to Huet et al. (2004), this procedure performs reasonably well in practice, even with small sample sizes and R as small as 200.

On the other hand, we can avoid computing a pivot altogether by working directly with the bootstrap replicates of $\widehat{x}_0$ from step (5) of Algorithm 1. A simple and often satisfactory procedure, known as the percentile bootstrap, uses the sample $\alpha/2$ and $1-\alpha/2$ quantiles of $\widehat{x}_{01}^\star, \ldots, \widehat{x}_{0R}^\star$ as a confidence interval for x_0. A more accurate technique is the *bias-corrected and accelerated* (BC_a) interval, which is essentially a modification of the percentile interval; for details see Efron and Tibshirani (1994,

chap. 14, sec. 3). The BC_a method tends to work well, but usually requires R to be large. It is both second-order accurate (Efron and Tibshirani, 1994, pp. 187) and transformation respecting. For example, if we want a BC_a confidence interval for $\log(x_0)$, we can simply log the endpoints of the corresponding BC_a interval for x_0. The studentized interval is also second-order accurate, but not transformation respecting, while the percentile interval is transformation respecting but only first-order accurate. For an in-depth discussion regarding the different bootstrap confidence interval procedures, see Davison and Hinkley (1997, chap. 5).

A related method for computing calibration intervals, based on the *jackknife*, was given by Miller (1974). Let $\beta = h(\boldsymbol{\beta})$ be a function of the regression parameters. Miller showed that, under reasonable conditions, the jackknife estimate

$$\widehat{\beta}_{\text{jack}} = n\widehat{\beta} - \frac{n-1}{n}\sum_{i=1}^{n}\widehat{\beta}_{(-i)},$$

where $\widehat{\beta}_{(-i)}$ denotes the estimate of β with the i-th observation removed from the data, is asymptotically normally distributed (even if the errors ϵ_i are not normally distributed). Hence, an approximate $100(1-\alpha)\%$ confidence interval for x_0 can be developed that does not require a normality assumption for the errors of the model. However, x_0 is not just a function of $\boldsymbol{\beta}$, but also of $\overline{\mathcal{Y}}_0$. This is not a problem as long as $m > 1$ and $m/n \to c$, $0 < c < \infty$. For regulation (Graybill and Iyer, 1994, pp. 431-432), x_0 is a function of the regression parameters only; hence, the jackknife-based interval is easily applied.

3.4 Bayesian calibration

Although calibration has been the subject of much discussion and debate from a frequentist point of view, it is just as intriguing from a Bayesian perspective. Let $\boldsymbol{y}$, $\boldsymbol{y}_0$, and **data** represent the data from the standards, unknowns, and both, respectively. Also, if x has a nonstandardized t-distribution with mean μ, precision

τ (i.e., recipricol of the variance), and k degrees of freedom—denoted $x \sim \mathcal{T}_k(\mu, \tau)$—then the probability density function of x is

$$f(x; \mu, \tau, k) = \frac{\Gamma(k/2 + 1/2)}{\Gamma(k/2)\sqrt{\tau k \pi}} \left[1 + \frac{(x-\mu)^2}{\tau k}\right]^{-(k+1)/2},$$

for $-\infty < x < \infty$, $-\infty < \mu < \infty$, $\tau > 0$, and $k \geq 1$.

Two influential papers on Bayesian calibration are Hoadley (1970) and Hunter and Lamboy (1981). Aitchison and Dunsmore (1980, chap. 10) discuss Bayesian solutions to the linear calibration problem for both natural and controlled calibration experiments. They also briefly discuss calibration under a general utility structure. Dunsmore (1967) derived the inverse estimator (3.6) as the conditional mean of $x_0|\mathbf{data}$ assuming independent observations from a bivariate normal distribution (i.e., Gaussian data from a natural calibration experiment). For controlled calibration experiments, Hoadley (1970) argued that the classical estimator is unsatisfactory, since the width of the inversion interval (3.12) depends on the magnitude of the F-statistic for testing the significance of the slope; in other words, the data contain information about the precision of $\widehat{x}_0$. Because of this, Hoadley argues that less weight should be given to this estimator when it is known to be unreliable, which is what a Bayes estimator does. He proposed a class of Bayesian solutions assuming, *a priori*, that x_0 is independent of $(\beta_0, \beta_1, \sigma_\epsilon^2)$. His results are based on a general prior of the form

$$\pi\left(\beta_0, \beta_1, \sigma_\epsilon^2, x_0\right) = \pi\left(\beta_0, \beta_1, \sigma_\epsilon^2\right) \pi\left(x_0\right),$$

but he obtained explicit results for the diffuse prior $\pi\left(\beta_0, \beta_1, \sigma_\epsilon^2\right) \propto 1/\sigma_\epsilon^2$. For the case $m = 1$, he was able to derive the inverse estimator (3.6) as a Bayes estimator under squared error loss with respect to a nonstandardized Student's t prior distribution for x_0:

$$x_0 \sim \mathcal{T}_{n-3}\left\{\bar{x}, \left(1 + \frac{1}{n}\right)\frac{S_{xx}}{n-3}\right\}. \tag{3.20}$$

This leads to the very simple result

$$x_0|\mathbf{data} \sim \mathcal{T}_{n-2}\left\{\widetilde{x}_0, \frac{\widehat{\sigma}_\epsilon^2 S_{xx}}{S_{yy}}\left(1+\frac{1}{n}+\frac{(y_0-\bar{y})^2}{S_{yy}}\right)\right\}. \tag{3.21}$$

A $100(1-\alpha)\%$ shortest credible interval for x_0 can be obtained from (3.21) and is given by

$$\widetilde{x}_0 \pm \sqrt{\frac{\widehat{\sigma}_\epsilon^2 S_{xx}}{S_{yy}}\left(1+\frac{1}{n}+\frac{(y_0-\bar{y})^2}{S_{yy}}\right) F_{1-\alpha,1,n-2}}. \tag{3.22}$$

Although this provides some theoretical justification for using the inverse estimator in controlled calibration, Hoadley is aware that it is merely a by-product of the Bayesian approach and recommends a careful elicitation of prior information. Aitchison and Dunsmore (1980, pp. 198, 204) give extensions of distributions (3.20)-(3.21) and interval (3.22) for the case $m > 1$. As they point out, this extension is somewhat nonsensical since the prior variance of x_0 depends on m, the number of future replicates. This is rather unrealistic, but "... tractability in these calibration problems can lead to an unnecessary departure from reality ..." (Aitchison and Dunsmore, 1980, pp. 198). Though many authors have criticized the classical estimator on the grounds that it has infinite MSE, Hoadley also contested its use based on the inherent problems of the inversion interval (3.10). These difficulties, however, do not arise with the shortest credible interval (3.22). Figure 3.5 shows the prior and posterior distributions of x_0 for the arsenic example (Section 3.2.4) using Hoadley's approach. The mean of the posterior is 2.945 μg/ml, the same as the inverse estimator, and the corresponding 95% shortest credible interval for x_0 is $(2.547, 3.342)$.

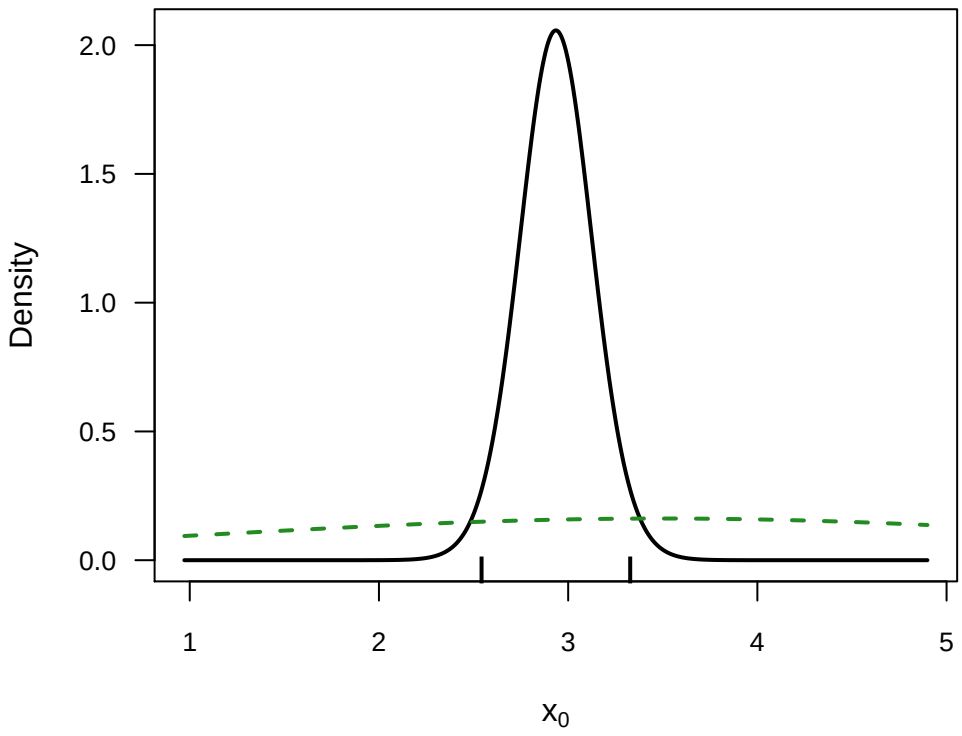


Figure 3.5: Bayesian calibration for the arsenic example. The black and green lines represent the posterior and prior for x_0, respectively. The posterior is symmetric and centered at $\widetilde{x}_0 = 2.945$. The tick marks indicate the endpoints of a 95% HPD interval for x_0.

Hunter and Lamboy (1981) also considered the linear calibration problem, but proposed a slightly different approach. Let $\eta = \beta_0 + \beta_1 x_0$ and assume, *a priori*, that η and $(\beta_0, \beta_1, \sigma_\epsilon^2)$ are independent. Assuming normal errors, a prior for the unknown $x_0 = (\eta - \beta_0)/\beta_1$ is induced by specifying improper reference priors for η, β_0, β_1, and σ_ϵ^2. That is, assuming

$$\pi\left(\eta, \beta_0, \beta_1, \sigma_\epsilon^2\right) = \pi\left(\beta_0, \beta_1, \sigma_\epsilon^2\right) \pi\left(\eta\right) \propto 1/\sigma_\epsilon^2.$$

This is quite different from the approach put forth by Hoadley (1970) who does not treat x_0 as an explicit function of η, β_0, and β_1. Hunter and Lamboy also discuss the case where σ_ϵ^2 is known, or at least, assumed known. The posterior they obtain for x_0 is equivalent to the posterior density for the ratio of bivariate normal

random variables (σ_ϵ^2 known) or bivariate t random variables (σ_ϵ^2 unknown), both of which have infinite variance. The latter is actually a generalization of the structural distribution for x_0 obtained by Kalotay (1971). A more thorough analysis of these posteriors is given by Hunter and Lamboy (1979a) and Hunter and Lamboy (1979b). Hunter and Lamboy claim that, under reasonable conditions, the inversion and Wald intervals (Section 3.3.1 and Section 3.3.2) provide accurate approximations to the highest posterior density (HPD) region for x_0. That said, Hill (1981), Orban (1981), Lwin (1981), and Brown (1982) all noted that Hunter and Lamboy's method offer some Bayesian justification for the classical estimator and confidence intervals.

Lawless (1981) criticized the authors for not using a more flexible family of priors and described the sole reliance on improper priors as "... merely attempting to dress classical frequency procedures in Bayesian clothes." Hill (1981) made the same criticism and argued that a carefully selected gamma prior for β_1 would have been more useful since, in practice, it is often reasonable to assume that β_1 is positive and not too close to zero. Lawless and Hill also pointed out that it is more realistic to assume, *a priori*, independence of $(\beta_0, \beta_1, \sigma_\epsilon^2)$ and x_0 (as did Hoadley), rather than independence of $(\beta_0, \beta_1, \sigma_\epsilon^2)$ and $\eta = \beta_0 + \beta_1 x_0$. Hill also pointed out that the approach used by Hunter and Lamboy (1981) is just a special case of that proposed by Hoadley (1970) with

$$\pi\left(\beta_0, \beta_1, \sigma_\epsilon^2, x_0\right) \propto \begin{cases} |\beta_1|/\sigma_\epsilon^2, & \sigma_\epsilon^2 \text{ unknown} \\ |\beta_1|, & \sigma_\epsilon^2 \text{ known} \end{cases}.$$

On the other hand, Lwin (1981) considered their approach somewhat attractive since it "... leads to an analysis parallel to the well-known 'ratio of means problem' ...". Lwin also objects to the use of non-informative priors and, in agreement with Hill, argues that the posterior of x_0 should be conditional on $\beta_1 > 0$. Another major criticism was that Hunter and Lamboy provide no justification for their choice of

locally uniform priors, nor did they seem interested in exploring any properties of the resulting posterior distribution and HPD intervals.

Bayesian methods can easily be extended to handle more complex situations. For example, Racine-Poon (1988) discusses the essential role of calibration in assay-type problems where relationships are inherently nonlinear. Following Hoadley (1970), Poon assumes that, *a priori*, x_0 and $(\boldsymbol{\beta}, \sigma_\epsilon^2)$ are independent, that is,

$$\pi\left(\boldsymbol{\beta}, \sigma_\epsilon^2, x_0\right) = \pi\left(\boldsymbol{\beta}, \sigma_\epsilon^2\right) \pi\left(x_0\right).$$

The resulting posterior is then given by

$$\pi\left(\boldsymbol{\beta}, \sigma_\epsilon^2, x_0 | \mathbf{data}\right) \propto \pi\left(\boldsymbol{y}_0 | x_0, \boldsymbol{\beta}, \sigma_\epsilon^2\right) \pi\left(\boldsymbol{y} | \boldsymbol{\beta}, \sigma_\epsilon^2\right) \pi\left(\boldsymbol{\beta}, \sigma_\epsilon^2\right) \pi\left(x_0\right).$$

Poon noted the substantial amount of numerical integration involved in obtaining the posterior of x_0 and proposed instead an approximation method. Given the current state of statistical software such as R (R Development Core Team, 2011) and OpenBUGS (Lunn et al., 2009), this is less of an issue. A good discussion on nonlinear calibration problems using proper priors is given by Hamada et al. (2003). Du Plessis and Van Der Merwe (1996) use Bayesian calibration to estimate the age of rhinoceros—a multivariate, nonlinear calibration problem.

3.4.1 Nasturtium example.

We end our discussion of Bayesian calibration with a nonlinear calibration example. Consider the nasturtium data from Racine-Poon (1988). The objective was to determine the concentrations of an agrochemical (e.g., pesticide) present in soil samples. Bioassays were performed on a type of garden cress called nasturtium. The response is weight of the plant in milligrams (mg) after three weeks of growth, and the predictor is the concentration of the agrochemical in the soil. In the first stage of the experiment, six replicates of the response $\mathcal{Y}_i$ were measured at each of seven preselected concentrations x_i (g/ha). Figure 3.6 shows a scatterplot of the standards

with concentration on the natural log scale (we added 0.01 to the zero concentrations before taking the logarithm).

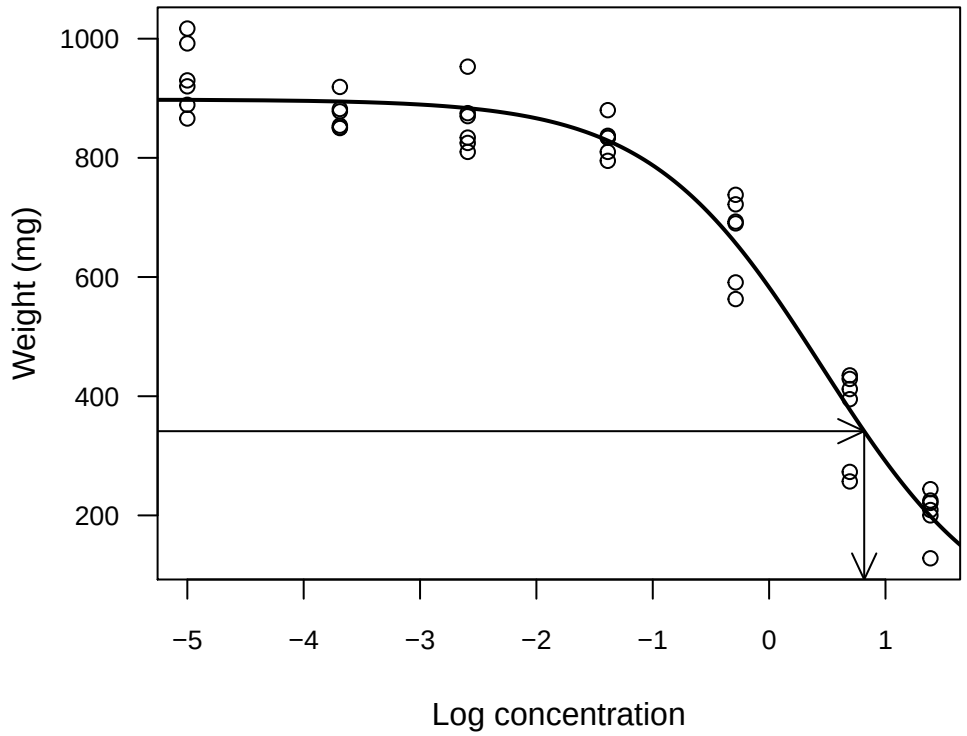


Figure 3.6: Scatterplot of the nasturtium data with fitted logit-log model. The horizontal arrow corresponds to the observed $\bar{y}_0 = 341.333$ mg and the vertical arrow corresponds to the logarithm of the estimated concentration $\log(\widehat{x}_0) = 0.817$.

A logit-log regression function is used to describe the data:

$$\mu(x; \beta_1, \beta_2, \beta_3) = \begin{cases} \beta_1, & x = 0 \\ \beta_1 / \left[1 + \exp\left\{\beta_2 + \beta_3 \ln(x)\right\}\right], & x > 0, \end{cases}.$$

The errors are assumed to be normally distributed with mean zero and constant variance σ^2_ϵ. Normality was checked using a normal Q-Q plot of the residuals and the constant variance assumption appears reasonable from the scatterplot. In the second stage of the experiment, the observed weights corresponding to three new soil samples all sharing the same concentration x_0 were observed to be 309, 296, and 419 mg. An

estimate of the unknown concentration is obtained by inverting the fitted calibration curve and found to be 2.2639. We follow Poon in assuming, *a priori*, that x_0 is independent of $(\beta_1, \beta_2, \beta_3, \sigma_\epsilon^2)$. Improper uniform priors were given to the parameters $(\beta_1, \beta_2, \beta_3, \sigma_\epsilon^2)$, and the prior for x_0 was chosen to be uniform over the experimental range: $x_0 \sim \mathcal{U}(0, 4)$. The posterior for x_0 is shown in Figure 3.7 and is nearly identical to the approximation obtained by Racine-Poon (1988, fig. 9). A histogram of the bootstrap distribution of $\widehat{x}_0$ is also shown in Figure 3.7 for comparison. The mean of the posterior density is 2.3434, and the mode is 2.3124. A corresponding 95% HPD interval for the unknown x_0 is (1.7566, 3.0011). For comparison, we also provide the inversion, Wald, and BC_a intervals for x_0 in Table 3.1. If our prior information accurately reflects the truth, then the HPD interval is preferred. Otherwise, the inversion or bootstrap interval are both reasonable. The bootstrap distribution and posterior density in Figure 3.7 are clearly skewed to the right; thus, a symmetric interval, such as the Wald interval, may not be appropriate here.

Table 3.1: Comparison of 95% calibration intervals for the nasturtium example.

Inversion interval	(1.772, 2.969)
Wald interval	(1.689, 2.839)
BC_a interval	(1.805, 2.903)
HPD interval	(1.7566, 3.0011)

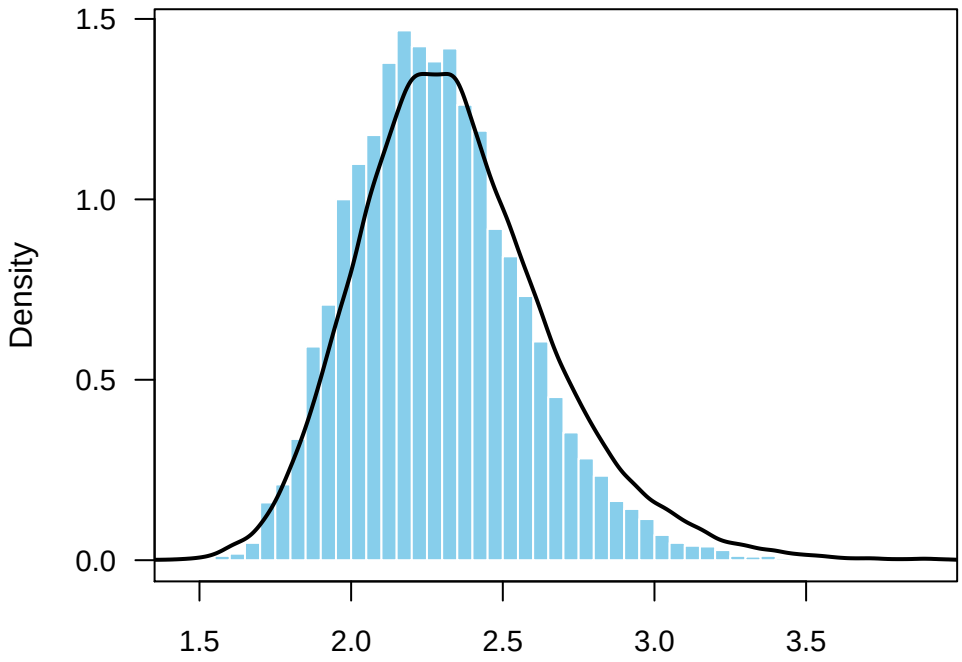


Figure 3.7: Posterior of x_0 (solid curve) together with the bootstrap distribution of $\widehat{x}_0$ (histogram).

IV. Semiparametric Calibration

So far, we have considered calibration curves in which the mean response $\mu(x) = \mathrm{E}\{\mathcal{Y}|x\}$ has a known form that depends on a small number of unknown parameters $\boldsymbol{\beta}$. Finding a good parametric model, however, can be time consuming and require a great deal of expertise. Therefore, it is sometimes useful to assess the effects of the explanatory variable x without completely specifying the structural form of $\mu(x)$. In this chapter, we propose a simple and fast *semiparametric* approach to computing calibration curves. By "semiparametric", we mean that only part of the model is specified. Our treatment of semiparametric calibration curves follows the work of Brumback et al. (1999), Ruppert (2002), and Ruppert and Wand (2003), and Crainiceanu et al. (2005). Nonparametric calibration has also been discussed in the literature by, for example, Clark (1979), Clark (1980), and Rosen and Cohen (1995b). Our approach to calibration here is similar to that of Clark (1980) in that we are inverting bias-adjusted prediction intervals. However, the LMM representation of P-splines (Ruppert and Wand, 2003) we use here yields a rather simple method for making this adjustment. Rosen and Cohen used a nonparametric bootstrap to obtain calibration intervals from a cubic smoothing spline. Their approach, however, made no attempt to correct for bias in the smoothed calibration curve.

In Section 4.1, we discuss the linear mixed-effects model representation of P-splines. Section 4.2 proposes a method for obtaining bias-adjusted calibration intervals based on the mixed model representation. A small Monte Carlo study demonstrates that the these intervals have coverage probability close to the nominal $1-\alpha$ level. A Bayesian analog of this procedure is proposed in Section 4.3. Section 4.4 concludes this chapter with a discussion on ideas for future research.

4.1 Mixed model representation of P-splines

Recall the polynomial spline model from Section 2.4:

$$\mathcal{Y}_i = \sum_{j=0}^{p} \beta_j x_i^j + \sum_{k=1}^{K} \alpha_k \left(x_i - \xi_k\right)_+^p + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \left(0, \sigma_\epsilon^2\right), \quad i = 1, \ldots, n. \tag{4.1}$$

We can easily write this in matrix form, such as in Equation (2.13), but a more useful form is obtained by separating the polynomial and spline terms as in

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \left(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}\right), \tag{4.2}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$ are the coefficients of the polynomial basis functions, $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)'$ are the coefficients of the spline basis functions, and $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known design matrices with i-th rows equal to

$$\boldsymbol{X}_i = \left(1, x_i, x_i^2, \ldots, x_i^p\right)' \quad \text{and} \quad \boldsymbol{Z}_i = \left(\left(x_i - \xi_1\right)_+^p, \ldots, \left(x_i - \xi_K\right)_+^p\right)',$$

respectively. Although we choose $\boldsymbol{Z}$ to be the truncated polynomial spline basis matrix, any other basis will do (e.g., radial basis, B-spline basis, etc.). Based on the matrix Equation (4.2), the penalized spline fitting criterion, Equation (2.14), becomes

$$\left\|\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}\right\|^2 + \lambda^{2p} \left\|\boldsymbol{\alpha}\right\|^2,$$

which is proportional to Equation (2.22), the PSS for an LMM with hierarchical structure $\boldsymbol{\mathcal{Y}}|\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha}, \sigma_\epsilon^2 \boldsymbol{I})$ and $\boldsymbol{\alpha} \sim \mathcal{N}(0, \sigma_\alpha^2 \boldsymbol{I})$. Dividing through by the error variance, σ_ϵ^2, gives the smoothing parameter as $\lambda^{2p} = \sigma_\epsilon^2 / \sigma_\alpha^2$, a simple ratio of the variance components. This leads to the following mixed model representation of P-splines (Brumback et al., 1999):

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\epsilon} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{0} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha^2 \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \sigma_\epsilon^2 \boldsymbol{I} \end{bmatrix}\right). \tag{4.3}$$

The parameters $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, σ_α^2, and σ_ϵ^2 can all be estimated using standard mixed model methodology and software. Under this framework, a P-spline is really just the BLUP

from a special LMM! The amount of smoothing is determined automatically by $\widehat{\lambda} = \widehat{\sigma}^2_\epsilon / \widehat{\sigma}^2_\alpha$, where $\widehat{\sigma}^2_\epsilon$ and $\widehat{\sigma}^2_\alpha$ are the REML estimates of σ^2_ϵ and σ^2_α, respectively. Although λ could be estimated via ordinary ML estimation or cross-validation, Krivobokova and Kauermann (2007) showed that the REML-based estimate is less affected by the presence of different correlation structures for the errors, $\boldsymbol{\epsilon}$. The BLUP of $\boldsymbol{\mu}$, denoted $\widetilde{\boldsymbol{\mu}}$, was given in Equation (2.24); however, an equivalent expression for $\widetilde{\boldsymbol{\mu}}$ is given by

$$\widetilde{\boldsymbol{\mu}} = \boldsymbol{\Omega} \left(\boldsymbol{\Omega}'\boldsymbol{\Omega} + \frac{\sigma^2_\epsilon}{\sigma^2_\alpha} \boldsymbol{D} \right)^{-1} \boldsymbol{\Omega}'\boldsymbol{\mathcal{Y}} = \boldsymbol{S}\boldsymbol{\mathcal{Y}}, \quad \boldsymbol{\Omega} = (\boldsymbol{X}; \boldsymbol{Z}),$$

where $\mathbf{D} = \text{diag}\left\{ \mathbf{0}_{(p+1)\times(p+1)}, \boldsymbol{I}_{K\times K} \right\}$. The vector of fitted values (i.e., the EBLUP of $\boldsymbol{\mu}$) is then just

$$\widehat{\boldsymbol{\mu}} = \boldsymbol{\Omega} \left(\boldsymbol{\Omega}'\boldsymbol{\Omega} + \frac{\widehat{\sigma}^2_\epsilon}{\widehat{\sigma}^2_\alpha} \boldsymbol{D} \right)^{-1} \boldsymbol{\Omega}'\boldsymbol{\mathcal{Y}},$$

where $\widehat{\sigma}^2_\epsilon$ and $\widehat{\sigma}^2_\alpha$ are the REML estimates of σ^2_ϵ and σ^2_α, respectively.

As described in Robinson (1991), the LMM has a simple Bayesian analog. For the mixed model representation, Equation (4.3), if $\boldsymbol{\beta}$, σ^2_ϵ, and σ^2_α are all given improper uniform priors, then the posterior of $\boldsymbol{\mu}$ is $\mathcal{N}\left(\widehat{\boldsymbol{\mu}}, \sigma^2_\epsilon \boldsymbol{S}\right)$. Note the use of the variance-covariance matrix $\sigma^2_\epsilon \boldsymbol{S}$ rather than the variance-covariance matrix $\sigma^2_\epsilon \boldsymbol{S}\boldsymbol{S}'$ from Equation (2.16). It is easy to see that $[\boldsymbol{S}]_{ij} \geq [\boldsymbol{S}\boldsymbol{S}']_{ij}$; hence, confidence and prediction intervals based on the Bayesian variance-covariance matrix $\sigma^2_\epsilon \boldsymbol{S}$ are wider than those based on $\sigma^2_\epsilon \boldsymbol{S}\boldsymbol{S}'$. As pointed out by Hastie and Tibshirani (1990), the "extra wideness" is due to the fact that $\boldsymbol{S}$ accounts for squared bias whereas $\boldsymbol{S}\boldsymbol{S}'$ does not.

4.2 Bias-adjusted calibration intervals

As mentioned in Section 2.4.1, the estimated mean response $\widehat{\mu}(x)$ is biased. Inference for P-splines based on the LMM representation, Equation (4.3), differs depending on whether we take the randomness of $\boldsymbol{\alpha}$ into account. Ignoring the

randomness in $\boldsymbol{\alpha}$ leads to the same intervals given in Section 2.4.1. We can account for bias in the confidence and prediction intervals, however, by conditioning on the random effects $\boldsymbol{\alpha}$. To see this, note that the bias vector of $\widetilde{\boldsymbol{\mu}}$, conditional on $\boldsymbol{\alpha}$, is

$$\mathrm{E}\left\{\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}|\boldsymbol{\alpha}\right\} = \boldsymbol{X}\left[\mathrm{E}\left\{\widetilde{\boldsymbol{\beta}}|\boldsymbol{\alpha}\right\} - \boldsymbol{\beta}\right] + \boldsymbol{Z}\left[\mathrm{E}\left\{\widetilde{\boldsymbol{\alpha}}|\boldsymbol{\alpha}\right\} - \boldsymbol{\alpha}\right]. \tag{4.4}$$

From the properties of conditional expectation (Casella and Berger, 2002, pg. 164), we have that $\mathrm{E}\left\{\mathrm{E}\left(\widetilde{\boldsymbol{\beta}}|\boldsymbol{\alpha}\right)\right\} = \mathrm{E}\left\{\widetilde{\boldsymbol{\beta}}\right\}$ and $\mathrm{E}\left\{\mathrm{E}\left(\widetilde{\boldsymbol{\alpha}}|\boldsymbol{\alpha}\right)\right\} = \mathrm{E}\left\{\widetilde{\boldsymbol{\alpha}}\right\}$. But since $\mathrm{E}\left\{\widetilde{\boldsymbol{\beta}}\right\} = \boldsymbol{\beta}$ and $\mathrm{E}\left\{\widetilde{\boldsymbol{\alpha}}\right\} = \mathrm{E}\left\{\boldsymbol{\alpha}\right\} = \mathbf{0}$, the unconditional bias is just

$$\mathrm{E}\left\{\mathrm{E}\left(\widetilde{\boldsymbol{\mu}} - \boldsymbol{\mu}|\boldsymbol{\alpha}\right)\right\} = \boldsymbol{X}\left(\boldsymbol{\beta} - \boldsymbol{\beta}\right) + \boldsymbol{Z}\left(\mathbf{0} - \mathbf{0}\right) = \mathbf{0}.$$

Thus, $\widetilde{\boldsymbol{\mu}}$ is unbiased for $\boldsymbol{\mu}$ when averaged over the distribution of $\boldsymbol{\alpha}$. This is equivalent to the Bayesian approach where (pointwise) confidence and prediction bands use diagonal entries of $\boldsymbol{S}$ in place of the diagonal entries of $\boldsymbol{SS}'$ described in Section 2.4.1.

Let $\boldsymbol{\Omega}_0 = (1, x_0, \ldots, x_0^p, (x_0 - \xi_1)_+^p, \ldots, (x_0 - \xi_K)_+^p)'$ where x_0 is an arbitrary value of the explanatory variable x. A $100(1-\alpha)\%$ (bias-adjusted) confidence interval for $\mu(x_0)$ is given by

$$\widehat{\mu}(x_0) \pm t_{1-\alpha/2,df} \times \widehat{\sigma}_\epsilon \sqrt{\boldsymbol{\Omega}_0 \left(\boldsymbol{\Omega}\boldsymbol{\Omega}' + \frac{\widehat{\sigma}_\epsilon^2}{\widehat{\sigma}_\alpha^2}\boldsymbol{D}\right)^{-1} \boldsymbol{\Omega}_0'},$$

where $df = n - 2\,\mathrm{tr}\left(\boldsymbol{S}\right) + \mathrm{tr}\left(\boldsymbol{SS}'\right)$. In a similar manner, a $100(1-\alpha)\%$ (bias-adjusted) prediction interval for a new observation, $\mathcal{Y} = \mu(x_0) + \epsilon_0$ (independent of current ones), is given as

$$\widehat{\mu}(x_0) \pm t_{1-\alpha/2,df} \times \widehat{\sigma}_\epsilon \sqrt{1 + \boldsymbol{\Omega}_0 \left(\boldsymbol{\Omega}\boldsymbol{\Omega}' + \frac{\widehat{\sigma}_\epsilon^2}{\widehat{\sigma}_\alpha^2}\boldsymbol{D}\right)^{-1} \boldsymbol{\Omega}_0'}. \tag{4.5}$$

Recall the intuition behind the inversion interval. Let $I_{\mathrm{pred}}(x)$ be a $100(1-\alpha)\%$ prediction interval for the future observation $\mathcal{Y}_0$, such that $\Pr\left\{\mathcal{Y}_0 \in I_{\mathrm{pred}}(x) | x_0 = x\right\} = 1 - \alpha$. Then, a confidence interval for x_0 corresponding to an observed $\mathcal{Y}_0$ is just the set $\mathcal{J}_{\mathrm{cal}}(x) = \{x : \mathcal{Y}_0 \in I_{\mathrm{pred}}(x)\}$ since, by construction,

$$\Pr\left\{x_0 \in \mathcal{J}_{\mathrm{cal}}(\mathcal{Y}_0)\right\} = \Pr\left\{\mathcal{Y}_0 \in I_{\mathrm{pred}}(x) | x_0 = x\right\} = 1 - \alpha,$$

(Clark, 1980). In other words, a $100(1-\alpha)\%$ confidence interval for x_0 can be obtained by inverting a corresponding prediction interval for $\mathcal{Y}_0$. However, we need to invert an appropriate prediction interval (i.e., one that adjusts for bias); otherwise, the resulting confidence interval for x_0 will likely not be reliable. To that end, we propose the following bias-adjusted $100(1-\alpha)\%$ calibration interval for the unknown x_0:

$$\widehat{\mathcal{J}}^{\text{ba}}_{\text{cal}}(x_0) = \left\{ x_0 : t_{\alpha/2,df} < \frac{y_0 - \widehat{\mu}(x_0)}{\sqrt{\widehat{\sigma}^2_\epsilon \left[1 + \boldsymbol{\Omega}_0 \left(\boldsymbol{\Omega}'\boldsymbol{\Omega} + \frac{\widehat{\sigma}^2_\epsilon}{\widehat{\sigma}^2_\alpha} \boldsymbol{D} \right)^{-1} \boldsymbol{\Omega}_0' \right]}} < t_{1-\alpha/2,df} \right\}, \tag{4.6}$$

As before, great care should be taken to ensure that $\widehat{\mathcal{J}}^{\text{ba}}_{\text{cal}}(x_0)$ is an interval. This may not be the case, for example, when $\mu(x)$ in not monotonic over the range of interest. Also, as with nonlinear calibration, no closed-form expression exists for $\widehat{\mathcal{J}}^{\text{ba}}_{\text{cal}}(x)$ and iterative techniques are required. In a similar fashion, a $100(1-\alpha)\%$ bias-adjusted regulation interval for x_0 is obtained by inverting a corresponding bias-adjusted confidence interval for the mean response:

$$\widehat{\mathcal{J}}^{\text{ba}}_{\text{reg}}(x_0) = \left\{ x_0 : t_{\alpha/2,df} < \frac{y_0 - \widehat{\mu}(x_0)}{\sqrt{\widehat{\sigma}^2_\epsilon \boldsymbol{\Omega}_0 \left(\boldsymbol{\Omega}'\boldsymbol{\Omega} + \frac{\widehat{\sigma}^2_\epsilon}{\widehat{\sigma}^2_\alpha} \boldsymbol{D} \right)^{-1} \boldsymbol{\Omega}_0'}} < t_{1-\alpha/2,df} \right\}.$$

We designed a small Monte Carlo study to investigate the coverage probability of this technique. The results of this study are presented in Tables 4.1 and 4.2. The true calibration function is the sine wave $\mu(x) = \sin(\pi x - \pi/2)/2 + 1/2$ plotted in Figure 4.1. For data following such a pattern, it would be tempting to fit a nonlinear model, perhaps modeling $\mu(x)$ as a *logistic* function. Inference based on the resulting fit would then be inaccurate since, for example, the true calibration curve does not have any asymptotes. For situations like this, it would be useful to have a semiparametric alternative for which to compare inference.

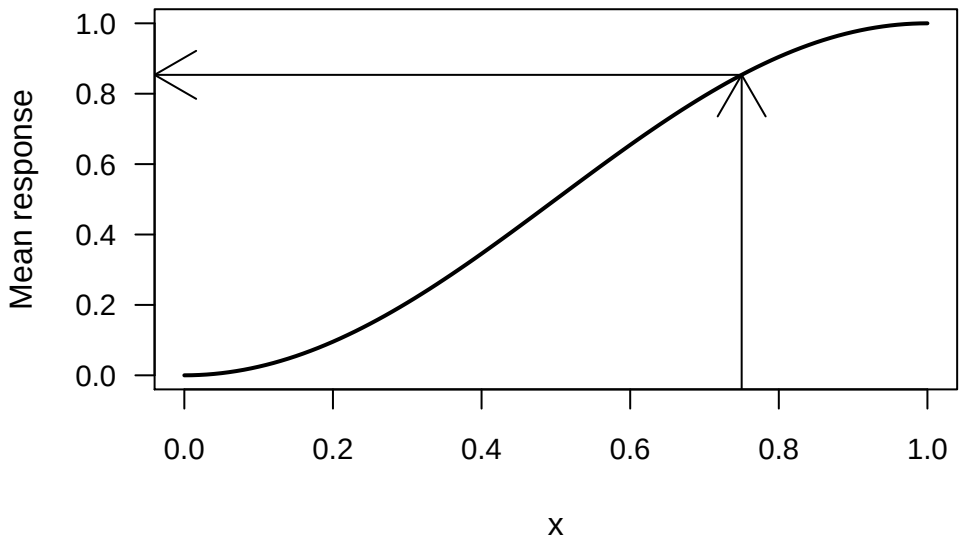


Figure 4.1: Scaled sine wave function.

Notice that the sine wave plotted in Figure 4.1 is scaled to lie in $[0,1]\times[0,1]$ with a period of 2. For each of 1,000 simulations, we used 10, 30, and 50 uniformly spaced designed points on the domain $[0,1]$ with 1, 2, and 3 independent replicates of the response at each design point. The errors were generated as i.i.d. $\mathcal{N}(0, 0.05^2)$ random variates and the true unknown was chosen to be $x_0 = 0.75$ (hence, $\mu_0 \approx 0.8536$).

Tables 4.1 and 4.2 display the estimated coverage probability for this technique applied to the sine wave experiment for both calibration (Table 4.1) and regulation (Table 4.2). For comparison, we also transformed the data to an equivalent linear model and applied Fieller's method (which yields an exact 95% confidence interval for x_0). Clearly, our bias-adjusted intervals performed well for quadratic (degree = 2) and cubic (degree = 3) P-splines with increasing sample size without sacrificing interval length. Notice, however, that the calibration results are somewhat conservative (i.e., the coverage probability is slightly larger than 0.95) while the regulation intervals tend to hit the target coverage $1-\alpha = 0.95$. The reason for this is that the bias in predicting a future observation is the same as that for estimating the mean response,

but the former has larger variance. Therefore, the bias accounts for a smaller portion of the mean-squared error in prediction (i.e., the bias gets "washed out" by a larger variance). Thus, for larger samples (perhaps $n \geq 20$ with $m \geq 1$ replicates at each design point), calibration intervals could be computed with or without adjusting for bias. Regulation intervals, on the other hand, should always be adjusted for bias.

Table 4.1: Coverage and length of 95% calibration intervals for data from the sine wave experiment with $\sigma_\epsilon = 0.05$.

		P-spline ($p = 1$)		P-spline ($p = 2$)		P-spline ($p = 3$)		Fieller	
n	m	CP	Length	CP	Length	CP	Length	CP	Length
10	1	0.93	0.25	0.93	0.21	0.90	0.19	0.95	0.25
10	2	0.96	0.21	0.97	0.20	0.96	0.19	0.96	0.22
10	3	0.97	0.21	0.96	0.20	0.97	0.19	0.96	0.21
30	1	0.95	0.21	0.95	0.20	0.97	0.18	0.96	0.21
30	2	0.96	0.20	0.96	0.19	0.97	0.18	0.95	0.20
30	3	0.97	0.20	0.96	0.19	0.97	0.18	0.96	0.20
50	1	0.97	0.20	0.96	0.19	0.96	0.18	0.96	0.20
50	2	0.97	0.20	0.97	0.19	0.96	0.18	0.95	0.20
50	3	0.98	0.20	0.97	0.19	0.97	0.18	0.96	0.20

Table 4.2: Coverage and length of 95% regulation intervals for data from the sine wave experiment with $\sigma_\epsilon = 0.05$.

		Degree = 1		Degree = 2		Degree = 3		Fieller	
n	m	CP	Length	CP	Length	CP	Length	CP	Length
10	1	0.74	0.13	0.92	0.12	0.92	0.11	0.95	0.10
10	2	0.89	0.09	0.94	0.09	0.94	0.08	0.95	0.06
10	3	0.91	0.07	0.95	0.07	0.94	0.06	0.95	0.05
30	1	0.96	0.07	0.95	0.07	0.94	0.06	0.95	0.05
30	2	0.97	0.06	0.95	0.05	0.94	0.04	0.95	0.04
30	3	0.97	0.05	0.95	0.04	0.95	0.03	0.95	0.03
50	1	0.96	0.06	0.95	0.05	0.95	0.05	0.95	0.04
50	2	0.97	0.04	0.96	0.04	0.95	0.03	0.95	0.03
50	3	0.97	0.04	0.96	0.03	0.95	0.03	0.95	0.02

4.2.1 Whiskey age example.

The point of this example is to demonstrate how our bias-adjusted semiparametric approach to calibration compares against a simpler parametric model (when one is available). To this end, we analyze the whiskey data from Schoeneman et al. (1971) presented in Figure 4.2. The data give the proof (measured as twice the percentage of alcohol by volume, denoted 2ABV) of whiskey stored in a charred oak barrel against time in years; clearly, some curvature is present. Suppose a new sample is obtained from the same barrel and that the proof is observed to be $y_0 = 108$ 2ABV. It is desired to estimate how long this batch has been aged.

```
## Loading required package:  rootSolve
```

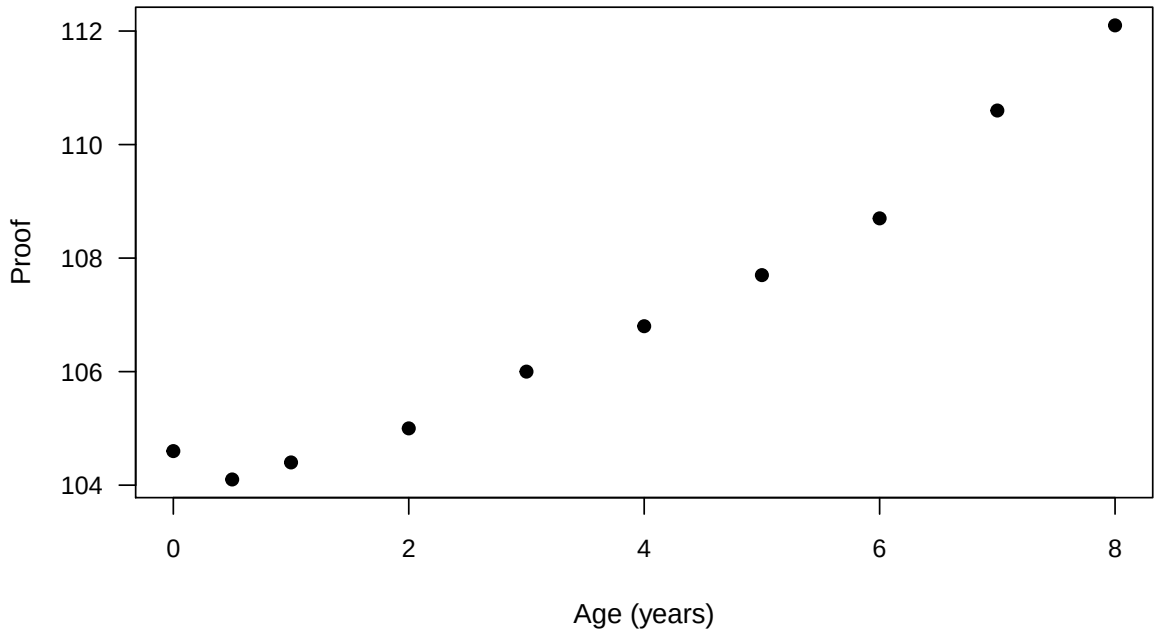


Figure 4.2: Scatterplot of the whiskey data.

We begin by fitting a simple quadratic P-spline with five knots:

$$\mu(\texttt{age}_i) = \beta_0 + \beta_1 \texttt{age}_i + \beta_2 \texttt{proof}_i^2 + \sum_{k=1}^{5} \alpha_k \left(\texttt{proof}_i - \xi_k\right)_+^2 .$$

The number of knots and their placement were determined automatically using the methods described in Section 2.4. The fitted model is depicted in the right-hand side of Figure 4.3. It appears from the scatterplot that a simple quadratic may be sufficient (i.e., $\alpha_i = 0$, $i = 1, \ldots, 5$); this fit is depicted in the left side of Figure 4.3. Both models are different, but the fits are indistinguishable to the human eye. In fact, both models produce identical calibration intervals for x_0 (when rounded to four decimal places). For the linear model, we have $\widehat{x}_0 = 5.2329$ years with a 95% confidence interval for x_0 of $(4.6776, 5.7352)$. Similarly, for the P-spline model, we have $\widehat{x}_0 = 5.2329$ years with a 95% (bias-adjusted) confidence interval for x_0 of $(4.6776, 5.7352)$. The reason we get the same answer from both models is that the polynomial basis part of the penalized spline model is sufficient for modeling the curvature of $\mu(\texttt{age})$, therefore,

the spline basis terms are effectively "zeroed out" (see Figure 4.4). Although our P-spline approach to obtaining calibration intervals requires large samples, in this small sample example, they happen to coincide with the classical inversion limits from the quadratic linear model. If we did not correct for bias in the P-spline model, however, the confidence interval obtained for x_0 would be $(4.7223, 5.7016)$, which is too narrow.

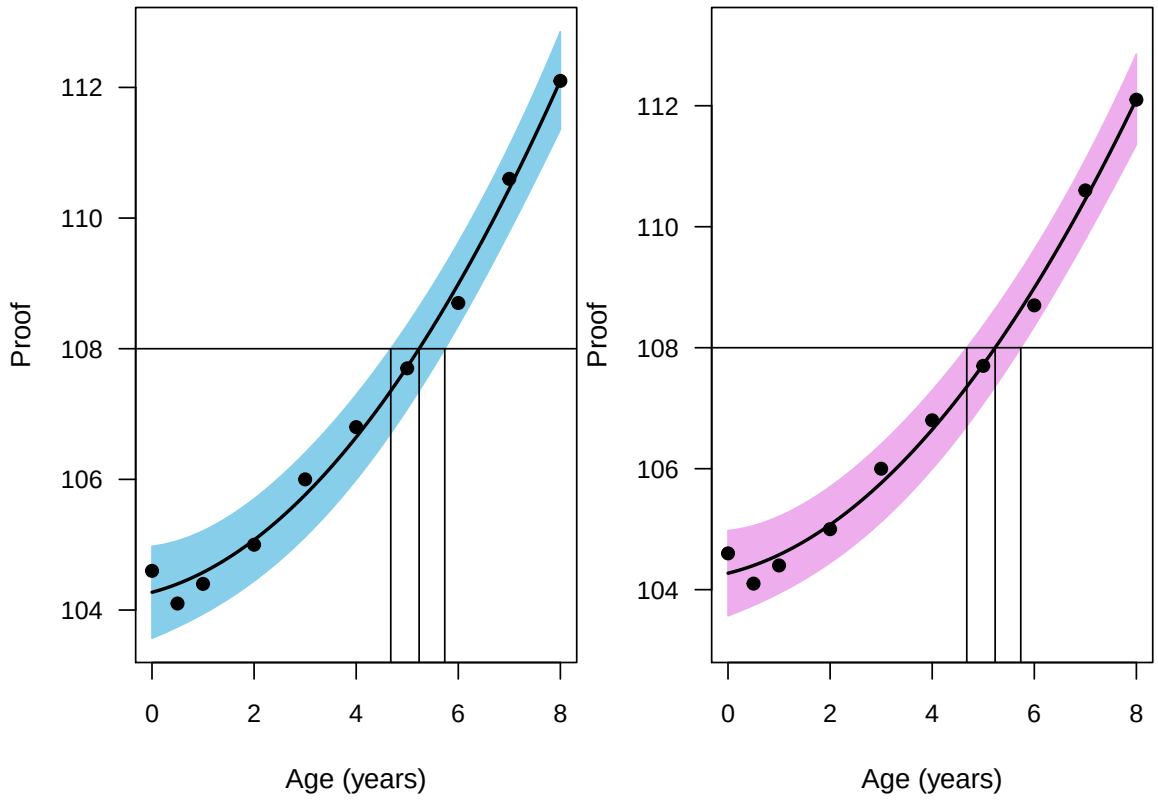


Figure 4.3: Fitted models for the whiskey data. *Left*: Quadratic linear model with 95% (pointwise) prediction band. *Right*: Quadratic P-spline with five knots and 95% (pointwise) bias-adjusted prediction band.

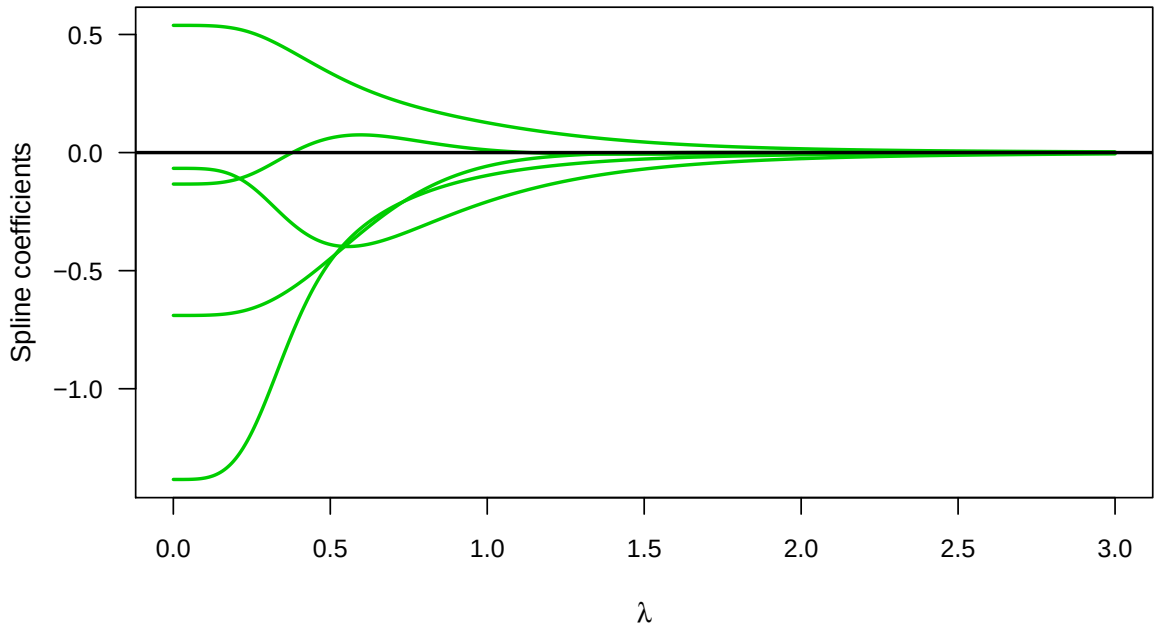


Figure 4.4: Profiles of spline coefficients for the P-spline in Figure 4.3, as the smoothing parameter λ is varied from zero to three. Coefficients are plotted versus the smoothing parameter λ; the REML estimate is $\widehat{\lambda} = 385.2228$.

4.3 Bayesian semiparametric calibration

The mixed model P-spline, Equation (4.3), provides a simple method for estimating the smoothing parameter and accounting for bias in calibration intervals. These intervals, however, do not account for the variability of the estimated smoothing parameter $\widehat{\lambda} = \widehat{\sigma}^2_\epsilon / \widehat{\sigma}^2_\alpha$ (similar to inference in LMMs which typically ignores the variability of $\widehat{\boldsymbol{V}}$). This problem can be remedied by adopting a fully Bayesian approach which we now discuss.

Let $\pi(\cdot)$ denote a probability density function. Following Hoadley (1970), we assume that the calibration experiment contains no information about x_0 and that the priors for x_0 and the calibration experiment are independent; thus,

$$\pi(x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2_\epsilon, \sigma^2_\alpha) = \pi(x_0)\pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma^2_\epsilon, \sigma^2_\alpha).$$

The mixed model representation, Equation (4.3), has $\boldsymbol{\alpha} \sim \mathcal{N}(0, \sigma_\alpha^2 \boldsymbol{I})$. A fully Bayesian approach, however, requires a prior distribution on the parameters $(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\alpha^2, x_0)$. Following standard convention, we assume, a priori, that the fixed-effects are independent and assign vague, independent priors to $(\boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\alpha^2)$. We used the vague (but proper) priors

$$\begin{aligned}
\beta_j &\sim \mathcal{N}\left(0, \sigma_\beta^2\right), \quad j = 1, \ldots, p+1, \\
\sigma_\epsilon^2 &\sim \mathcal{IG}\left(a, b\right), \\
\sigma_\alpha^2 &\sim \mathcal{IG}\left(c, d\right),
\end{aligned}$$

where $\mathcal{IG}$ stands for the inverse gamma distribution. The variance σ_β^2 should be chosen large enough (say 10^6) so that (for all intents and purposes) the β_j's are uniform. Similarly, the parameters a, b, c, and d should be small (say 10^{-6}). These distributions can be considered as an approximate representation of vagueness in the absence of good prior information. In addition, we assume that the prior for x_0, the predictor value of interest, is uniform over the experimental range: $x_0 \sim U[a, b]$.

The (unnormalized) posterior density of $(x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2)$ is given by

$$\begin{aligned}
\pi(x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2 | \mathbf{data}) &= \pi(x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2 | \boldsymbol{y}, \boldsymbol{y}_0) \\
&\propto \pi(\boldsymbol{y}, \boldsymbol{y}_0 | x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2) \pi(x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2) \\
&\propto \pi(\boldsymbol{y}_0 | x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2) \pi(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2) \pi(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2) \pi(x_0) \\
&\propto \pi(\boldsymbol{y}_0 | x_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2) \pi(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2) \pi(\boldsymbol{\beta}) \pi(\boldsymbol{\alpha} | \sigma_\alpha^2) \pi(\sigma_\epsilon^2) \pi(\sigma_\alpha^2) \pi(x_0),
\end{aligned}$$

where $\boldsymbol{y}$ and $\boldsymbol{y}_0$ represent the observed data from the first and second stages of the calibration experiment, respectively. It is relatively straightforward to show that (see Section A.1 in the appendix) the conditional posterior of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is proportional to

$$\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\left\|\boldsymbol{y}_0 - \boldsymbol{X}_0\boldsymbol{\beta} - \boldsymbol{Z}_0\boldsymbol{\alpha}\right\|^2 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}\left\|\boldsymbol{\alpha}\right\|^2\right)\right\},$$

where

$$\boldsymbol{y}_0 = \begin{bmatrix} \boldsymbol{y} \\ y_0 \end{bmatrix}, \quad \boldsymbol{X}_0 = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{x}_0' \end{bmatrix}, \quad \boldsymbol{Z}_0 = \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{z}_0' \end{bmatrix}$$

are augmented data vectors and matrices. Upon completing the square we have that

$$\boldsymbol{\beta}, \boldsymbol{\alpha} | \boldsymbol{y}_0, \boldsymbol{y}, x_0, \sigma_\epsilon^2, \sigma_\alpha^2 \sim \mathcal{N}\left\{ \left(\boldsymbol{\Omega}_0' \boldsymbol{\Omega}_0 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2} \boldsymbol{D} \right)^{-1} \boldsymbol{\Omega}_0' \boldsymbol{y}_0, \sigma_\epsilon^2 \left(\boldsymbol{\Omega}_0' \boldsymbol{\Omega}_0 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2} \boldsymbol{D} \right)^{-1} \right\},$$

where $\boldsymbol{\Omega}_0 = (\boldsymbol{X}_0, \boldsymbol{Z}_0)$. In a similar fashion, the conditional posteriors of σ_ϵ^2 and σ_α^2 are the inverse gammas:

$$\sigma_\epsilon^2 | \boldsymbol{y}_0, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\alpha^2, x_0 \sim \mathcal{IG}\left(a + \frac{n+1}{2}, b + \frac{1}{2} \left\| \boldsymbol{y}_0 - \boldsymbol{X}_0 \boldsymbol{\beta} - \boldsymbol{Z}_0 \boldsymbol{\alpha} \right\|^2 \right),$$
$$\sigma_\alpha^2 | \boldsymbol{y}_0, \boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, x_0 \sim \mathcal{IG}\left(c + \frac{K}{2}, d + \frac{1}{2} \left\| \boldsymbol{\alpha} \right\|^2 \right),$$

where K is the number of knots or random effects (see Section A.2 in the appendix). The conditional posterior of x_0 is more difficult to obtain analytically, however, regardless of the prior $\pi(x_0)$. This makes it difficult (or impossible) to obtain a full Gibbs sampler here, nonetheless, we can sample from the posterior of x_0 using more specialized Markov Chain Monte Carlo methods such as the *Metropolis-Hastings* algorithm (see, for example, Robert and Casella (2004, chap. 7)). Thus, as discussed in (Gelman et al., 2003, pg. 292), we could update the parameters one at a time using Gibbs sampling for $(\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2)$ and a metropolis update for x_0. We illustrate this approach with the following example involving radioimmunoassays. The data were analyzed using the `JAGS` software within `R` via the `rjags` package.

4.3.1 Enzyme-linked immunosorbent assay (ELISA) example.

Rosen and Cohen (1995a) derived a 95% calibration interval for the unknown concentration in a radioimmunoassay problem using the nonparametric bootstrap. We demonstrate our method on the same dataset and compare the results. The data, plotted in Figure 4.5, consists of 23 distinct concentrations with four (independent)

replicates of the response at each concentration. The results from analyzing these data are summarized in Table 4.3 at the end of this section.

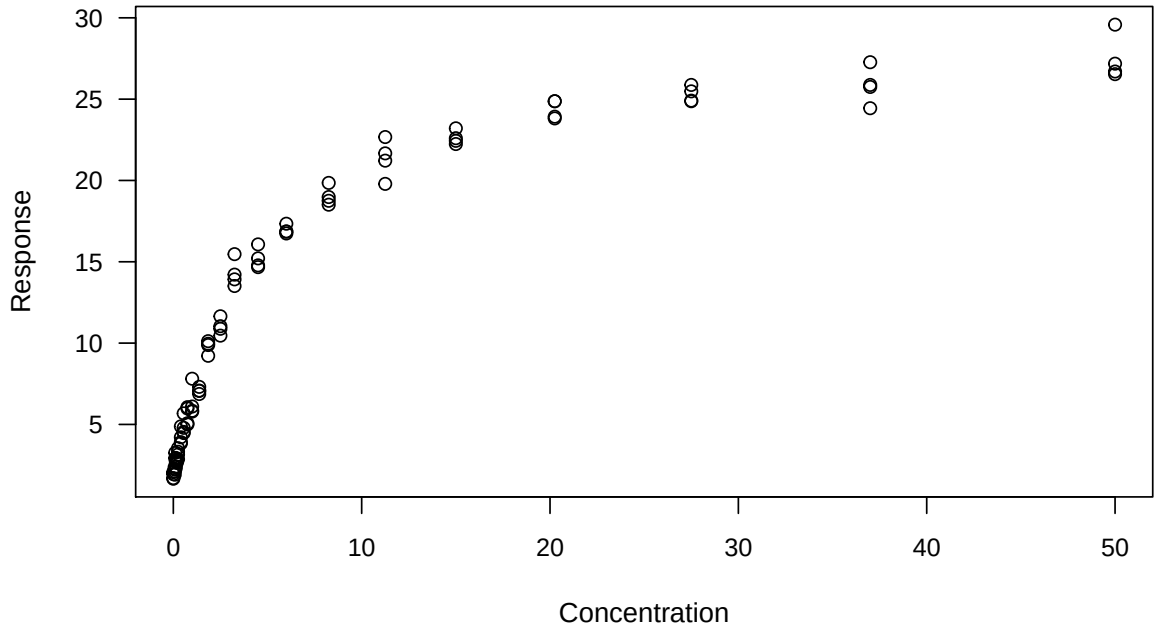


Figure 4.5: Scatterplot of the ELISA data.

The four parameter logistic model,

$$\mathcal{Y}_i = \beta_1 + \frac{\beta_2 - \beta_1}{1 + \exp\{\beta_4(\log x_i - \beta_3)\}} + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right), \quad i = 1, \ldots, n,$$

provides a reasonable parametric fit model to the ELISA data. Although we assume the errors are normal with constant variance, Rosen and Cohen (1995a) more generally assumed $\epsilon_i \overset{iid}{\sim} (0, \sigma_i^2)$, where $\sigma_i = \sigma\left(\mu(x_i, \boldsymbol{\beta})\right)^\theta$. This adds an unnecessary complication to the calibration model and so we simply assume constant variance (standard residual plots do not reveal any serious indication of heteroscedastic errors). Following Rosen and Cohen (1995a), we assume we have a new observation $y_0 = 20$ with unknown concentration x_0. The frequentist (i.e., classical) estimate of x_0 is

rather straightforward to derive

$$\widehat{x}_0 = \exp\left\{\frac{1}{\widehat{\beta}_4}\log\left(\frac{\widehat{\beta}_2 - y_0}{y_0 - \widehat{\beta}_1}\right) + \widehat{\beta}_3\right\} = 9.1837.$$

The 95% inversion limits for x_0 based on Equation (3.15) are $(7.356, 11.657)$—see Figure 4.6.

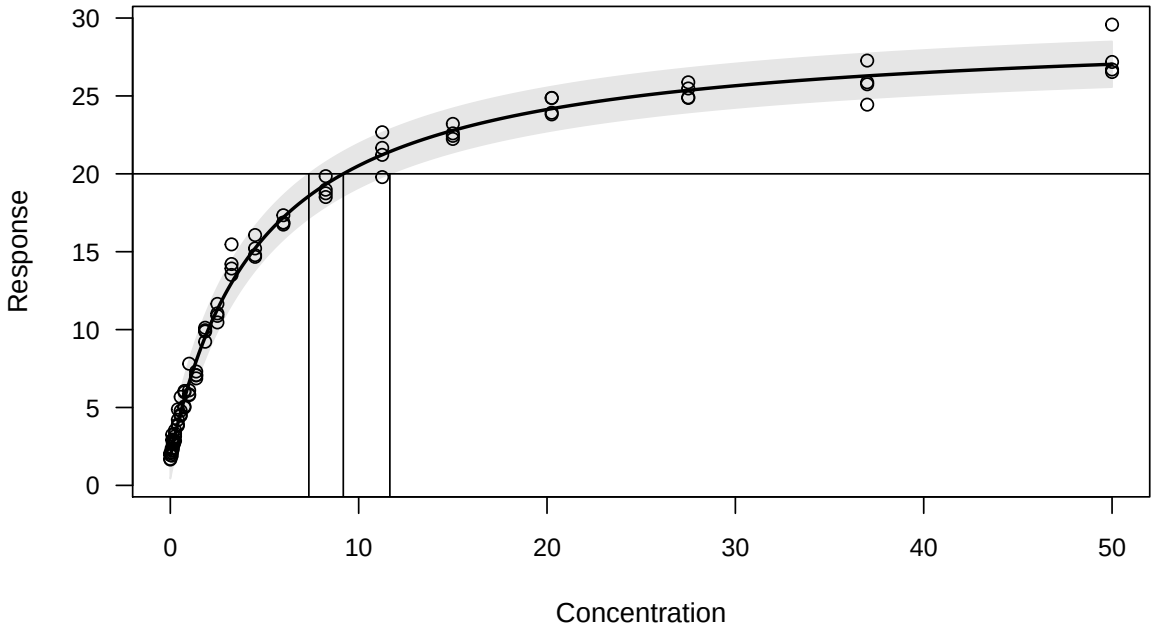


Figure 4.6: Nonlinear least squares fit of the ELISA data to the four parameter logistic model with (pointwise) 95% prediction band.

Figure 4.7 shows a fitted semiparametric calibration curve, a quadratic P-spline with five interior knots. Also shown is the posterior mean based on a fully Bayesian P-spline model. Seeing as how the concentration should not be negative, we used a $\mathcal{U}[0, 50]$ prior for x_0, though, lognormal or gamma priors may also be reasonable. The frequentist estimate of x_0 is obtained by (numerically) inverting the fitted P-spline: $\widehat{x}_0 = 9.345$. A (bias-adjusted) inversion interval for x_0 based on Equation (4.6) is $(7.462, 11.711)$. As previously mentioned, this interval does not account for

the variability of the estimated smoothing parameter $\widehat{\sigma}^2_\epsilon/\widehat{\sigma}^2_\alpha$, hence, we expect the Bayesian credible interval to be slightly wider. The estimator we use for the Bayesian model is the posterior mode of x_0 which is equal to 9.285. There are a number of ways to compute a 95% credible interval from a given posterior; for example, *highest posterior density* (HPD) intervals. Here we simply report the 0.025 and 0.975 quantiles of the posterior which are 7.250 and 11.776, respectively. As noted earlier, these limits are slightly wider than the corresponding frequentist limits, but this extra wideness is likely due to the added uncertainty of the estimated smoothing parameter. Rosen and Cohen (1995a) also obtained calibration intervals for these data. Their approach involved bootstrapping a cross-validated cubic smoothing spline, but did not account for bias in $\widehat{\mu}(x)$. For a brief discussion on bias correction in nonparametric regression using the bootstrap, see Davison and Hinkley (1997, pp. 362-366).

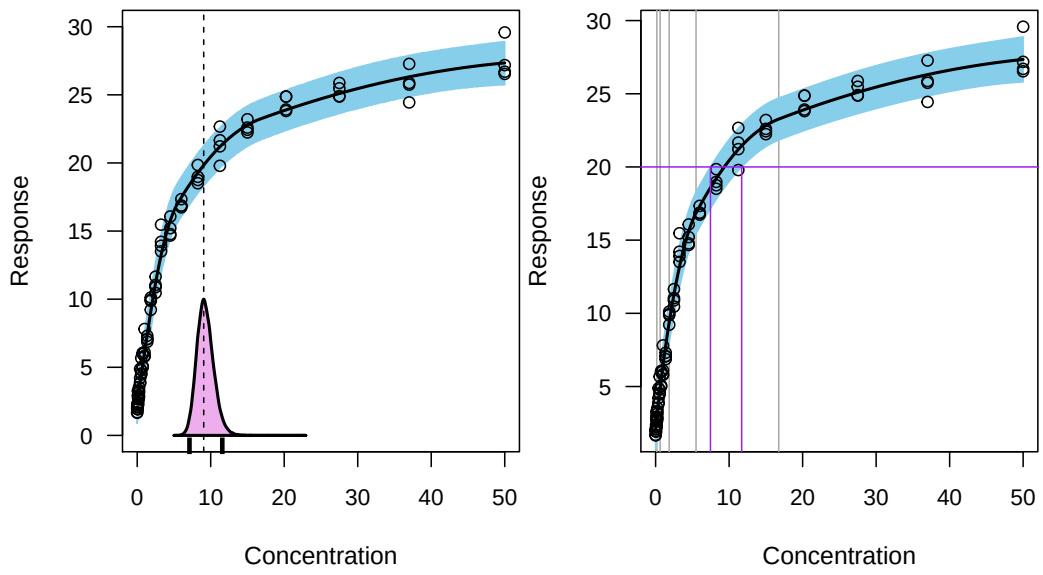


Figure 4.7: Nonparametric calibration for ELISA data. *Left*: Bayesian P-spline with posterior and 95% credible interval for x_0. The shaded blue region represents the (pointwise) prediction band and the density curve represents the posterior of x_0. *Right*: Mixed-effects model P-spline with bias-adjusted 95% calibration interval for x_0. The shaded blue region represents the (pointwise) bias-adjusted predction band.

Table 4.3: Point and interval estimates for x_0 for the ELISA data with $y_0 = 20$.

Method	**Estimate**	**95% interval**
Parametric (homoscedastic errors)	9.184	$(7.356, 11.657)$
P-spline (bias-adjusted)	9.345	$(7.462, 11.711)$
P-spline (Bayesian)	9.285	$(7.250, 11.776)$

4.4 Discussion

We have proposed a new method for calibration in a semiparametric setting based on the mixed model approach to smoothing that corrects for bias in the smoothed calibration curve. While the idea of using LMMs for penalized smoothing is not new (see, for example, Demidenko (2013, pp. 13 - 17)), this approach has never been adapted for the calibration problem as we have proposed here. By using a simple Monte Carlo experiment, we have demonstrated that bias-adjusted prediction intervals can be used to (numerically) obtain calibration intervals for the unknown x_0 that have good coverage probabilities. We have discussed a situation where this bias-correction is more serious (i.e., regulation, or rather, calibration where the unknown x_0 corresponds to some specified mean response μ_0). Finally, we developed a Bayesian framework for semiparametric calibration by extending the approach originally put forth by Hoadley (1970) for linear calibration (see Section 3.4). The frequentist framework, while fast and simple, has the same disadvantage inherent in LMM inference; that is, the variance of the estimated smoothing parameter (which is a function of the estimated variance components) is ignored. The Bayesian approach handles this by incorporating prior information for all unknown parameters in the model, leading to slightly wider calibration intervals. The Bayesian approach we presented, however, is slower and more difficult to implement in practice.

4.4.1 Priors.

Although we used a uniform prior for x_0 in our examples, such a prior for x is unlikely to be useful in general and we recommend a more careful elicitation of prior information. A sensitivity analysis should also be carried out to see if the results are sensitive to the choice of prior for x_0. Although inverse gamma priors are commonly used in practice as noninformative priors for scale parameters in hierarchical models (e.g., the variance components in a LMM), Gelman (2006) argued against their use. Instead, Gelman adovcated the use of conditionally conjugate priors from a new folded-noncentral-t family.

4.4.2 Future work.

The methods proposed in this chapter open the door to a number of future research opportunities. Perhaps, the most interesting (and logical) next step would be the inclusion of constraints. For example, we have assumed that the calibration curve is monotonic over the range of interest. Fortunately, this is not usually a concern when the data are collected from a carefully designed experiment. The semiparametric fit, however, is not necessarily monotonic for every value of the smoothing parameter which may cause problems when, say, obtaining a bias-adjusted calibration interval. It is possible, however, to incorporate constraints, such as monotonicity, using the general projection method described in Mammen et al. (2001). Until such a constraint can be smoothly incorporated (no pun intended) into our semiparametric approach to calibration, we can instead rely on inference regarding the first derivative of f as described in Ruppert and Wand (2003, pp. 151-156). For instance, if we assume that f is monotonically increasing (decreasing) over the interval $[a, b]$, then a plot of the estimated first derivative of the regression function should lie completely above (below) the x-axis. This derivative function can be estimated in exactly the same way as the regression function itself, that is, using P-splines. Therefore, a thorough

calibration analysis might include a plot of the data with fitted mean response, supplemented by a plot of the estimated first derivative function (possibly with a confidence band). For instance, we might supplement our analysis of the ELISA data with the following graphic:

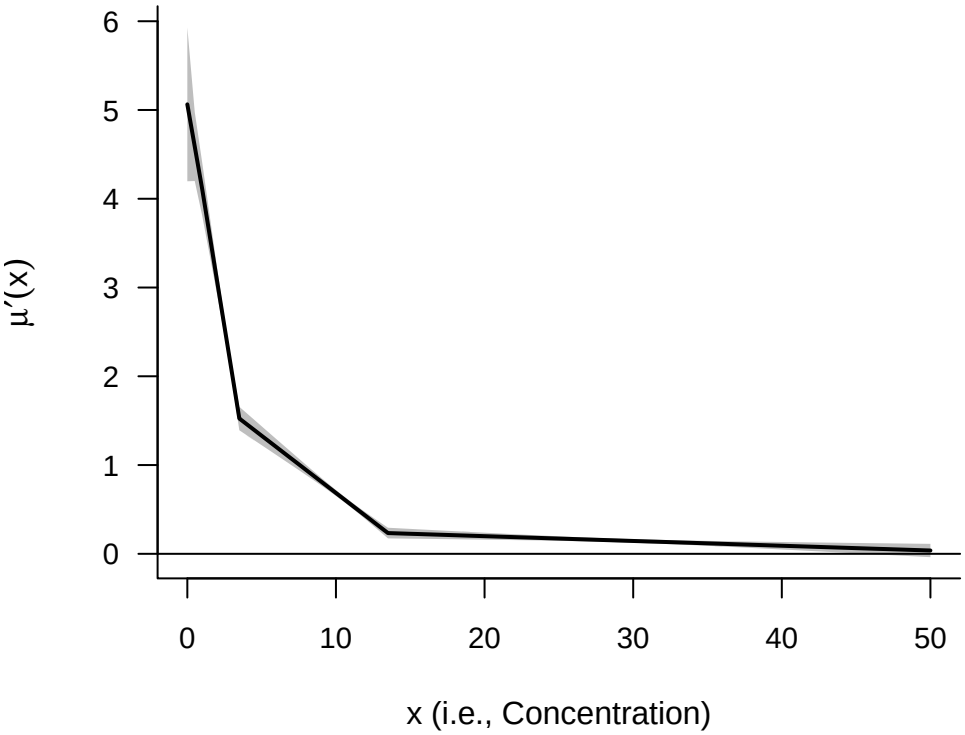


Figure 4.8: Estimate of the first derivative of the regression function for the ELISA example with a 95% global confidence band. A horizontal reference line is displayed at zero on the y-axis. This plot was produced using the `R` package SemiPar.

V. Calibration with Grouped Data

In this chapter, we extend the application of calibration to *grouped data*; that is, data in which the observations are grouped into disjoint classes called clusters or groups. Common examples of grouped data include *repeated measures data* and *longitudinal data*. Groups tend to be homogeneous, therefore, observations belonging to the same group cannot be considered independent. (Although, observations between clusters usually are.) Thus, we need to account for within cluster dependence when modeling this type of data. To our knowledge, other than Oman (1998), very little has been done for calibration with grouped data. Oman considered a simpler model that only allowed for the intercept and slope to vary between groups, whereas we take a more general (and practical) approach that allows for an arbitrary random effects structure. Furthermore, while Oman considers only one type of calibration interval, we discuss four different calibration intervals that can be computed for grouped data, along with some adjustments to improve their accuracy. Moreover, the calibration interval considered by Oman was based on an approximate parametric bootstrap that did not account for the variance attributed by the random variable $\mathcal{Y}_0$.

The LMM was introduced in Section 2.5. In Section 5.1, we discuss a particular useful LMM: the *random coefficient model*. In Section 5.2, we propose a simple method for estimating the unknown x_0 in a mixed model setting. We argue the utility of this approach by showing that, in a particular case, it coincides with the ML solution. In Sections 5.3 and 5.4, we discuss construction of Wald and asymptotic inversion intervals, respectively. We then propose a fully parametric bootstrap approach for controlled calibration in Section 5.5. Unlike Oman (1998), our parametric bootstrap algorithm does take into account the variability attributed

by $\mathcal{Y}_0$. Distribution-free calibration intervals are (briefly) considered in Section 5.6. Finally, Section 5.8 applies the aforementioned techniques to a real dataset taken from Brown (1993).

5.1 LMMs for repeated measures data

As discussed in Section 2.5, the LMM extends the basic LM (2.2) to

$$\boldsymbol{\mathcal{Y}} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{Z}\boldsymbol{\alpha} + \boldsymbol{\epsilon}, \tag{5.1}$$

where $\boldsymbol{X}$ and $\boldsymbol{Z}$ are known design matrices, $\boldsymbol{\beta}$ is a vector of fixed effects, $\boldsymbol{\alpha}$ is a vector of random effects, and $\boldsymbol{\epsilon}$ is a vector of random errors. Since we are using mixed models to analyze grouped data, it would behoove us to decompress model (5.1) into the form introduced by Laird and Ware (1982),

$$\boldsymbol{\mathcal{Y}}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\alpha}_i + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, m, \tag{5.2}$$

where, for the i-th group:

- $\boldsymbol{\mathcal{Y}}_i$ is an $n_i \times 1$ vector of response variables;
- $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known design matrices of dimensions $n_i \times p$ and $n_i \times q$, respectively;
- $\boldsymbol{\beta}$ is a $p \times 1$ vector of fixed effects;
- $\boldsymbol{\alpha}_i$ is a $q \times 1$ vector of random effects;
- $\boldsymbol{\epsilon}_i$ is an $n_i \times 1$ vector of random errors.

Furthermore, it is assumed that the random variables $\{\boldsymbol{\alpha}_i\}_{i=1}^m$ and $\{\boldsymbol{\epsilon}_i\}_{i=1}^m$ are mutually independent and distributed according to

$$\boldsymbol{\alpha}_i \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{G}), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{R}_i\right).$$

For our purposes, we shall assume that $\boldsymbol{R}_i = \boldsymbol{I}$ (an $N \times N$ identity matrix); that is, we are assuming constant variance within groups. Also, for computational purposes (e.g., maximizing the likelihood), it is convenient to reparameterize $\boldsymbol{G}$ as $\sigma_\epsilon^2 \boldsymbol{G}^\dagger$, where $\boldsymbol{G}^\dagger$ is the "scaled" variance-covariance matrix for the random effects. In other words, we have that

$$\boldsymbol{\alpha}_i \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{G}^\dagger\right), \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_\epsilon^2 \boldsymbol{I}\right).$$

Model (5.1) can also be written in marginal form as

$$\boldsymbol{\mathcal{Y}}_i \sim \mathcal{N}\left(\boldsymbol{X}_i \boldsymbol{\beta}, \boldsymbol{V}_i\right),$$

where

$$\boldsymbol{V}_i = \boldsymbol{Z}_i \boldsymbol{G} \boldsymbol{Z}_i' + \sigma_\epsilon^2 \boldsymbol{I}.$$

In long notation (Equation (5.1)), we have

$$\boldsymbol{\mathcal{Y}} = \begin{bmatrix} \boldsymbol{\mathcal{Y}}_1 \\ \vdots \\ \boldsymbol{\mathcal{Y}}_m \end{bmatrix}_{N \times 1}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1 \\ \vdots \\ \boldsymbol{X}_m \end{bmatrix}_{N \times p}, \quad \boldsymbol{Z} = \begin{bmatrix} \boldsymbol{Z}_1 & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{Z}_m \end{bmatrix}_{N \times mq},$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \vdots \\ \boldsymbol{\alpha}_m \end{bmatrix}_{mq \times 1}, \quad \boldsymbol{\epsilon} = \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \vdots \\ \boldsymbol{\epsilon}_m \end{bmatrix}_{N \times 1},$$

where $N = \sum_{i=1}^m n_i$ is the total number of observations. Similarly, this can be summarized in marginal form as $\boldsymbol{\mathcal{Y}} \sim \mathcal{N}\left(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V}\right)$ where

$$\boldsymbol{V} = \sigma_\epsilon^2 \left\{ \underset{\text{diag}}{} \boldsymbol{I}_{n_i} + \boldsymbol{Z}_i \boldsymbol{G}^\dagger \boldsymbol{Z}_i' \right\}_{i=1}^m.$$

Notice, in this model, how the fixed effects are used to model the mean of $\boldsymbol{\mathcal{Y}}$ while the random effects govern the variance-covariance structure of $\boldsymbol{\mathcal{Y}}$. In fact, as pointed out by McCulloch et al. (2008), a key reason for including random effects in a model

is to simplify the otherwise difficult task of dealing with $N(N+1)/2$ unique elements of $\boldsymbol{V}$.

Ignoring constants, the log-likelihood for the data can be written as

$$\mathscr{L}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \boldsymbol{\theta}\right) = -\frac{N}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\sum_{i=1}^{m}\log\left|\boldsymbol{I} + \boldsymbol{Z}_i\boldsymbol{G}^\dagger\boldsymbol{Z}_i'\right| - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)'\left(\boldsymbol{I} + \boldsymbol{Z}_i\boldsymbol{G}^\dagger\boldsymbol{Z}_i'\right)^{-1}\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right), \tag{5.3}$$

where $\boldsymbol{\theta}$ is a vector containing the unique elements of $\boldsymbol{G}^\dagger$. Since $\boldsymbol{G}^\dagger$ is symmetric, it has at most $q(q+1)/2$ unique elements; hence, $\boldsymbol{\theta}$ has a maximum dimension of $q(q+1)/2$. In many practical applications, however, we can restrict $\boldsymbol{G}^\dagger$ (or equivalently $\boldsymbol{G}$) to simpler forms involving only a few parameters. For example, $\boldsymbol{G}^\dagger$ may be a constant multiple of the identity matrix, $\tau\boldsymbol{I}$, which corresponds to uncorrelated random effects with constant variance $\sigma_\epsilon^2\tau$; in this case, $\boldsymbol{\theta} = \tau$ has dimension one. This is the variance-covariance structure we used for the random coefficients in the mixed model approach to P-splines in the previous chapter.

One of the most useful LMMs for repeated measures data is the so-called *random linear trend model*:

$$\mathcal{Y}_{ij} = (\beta_0 + \alpha_{0i}) + (\beta_1 + \alpha_{1i})\,x_{ij} + \epsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n_i, \tag{5.4}$$

where β_0 and β_1 are fixed effects, $\left\{\alpha_{0i}\right\}$ are random intercepts distributed as $\mathcal{N}(0, \sigma_0^2)$, $\left\{\alpha_{1i}\right\}$ are random slopes distributed as $\mathcal{N}(0, \sigma_1^2)$, and $\left\{\epsilon_{ij}\right\}$ are i.i.d. random errors distributed as $\mathcal{N}(0, \sigma_\epsilon^2)$. Also, if we let $\mathrm{Cov}\left\{\alpha_{0i}, \alpha_{1i}\right\} = \sigma_{01}$, then

$$\boldsymbol{G} = \begin{bmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{bmatrix}.$$

We often assume the random intercepts and slopes are independent, that is, $\sigma_1^2 = 0$; this can be formally tested using a *likelihood ratio test*. Also, setting $\alpha_{01} = \cdots = \alpha_{0m} = 0$ or $\alpha_{11} = \cdots = \alpha_{1m} = 0$ yields the *random intercept* and *random slope* models,

respectively (see Figure 5.1). Considerable simplifications arise for the balanced cases (i.e., when $n_i = n$ for all i). For example, for a balanced random intercept model, the matrix $\boldsymbol{Z}$ of Equation (5.1) is just $\boldsymbol{Z} = \boldsymbol{I}_m \otimes \mathbf{1}_n$, where the symbol $\otimes$ denotes the Kronecker product. Estimating the fixed effects and variance components for model (5.4) is also much simpler in the balanced case.

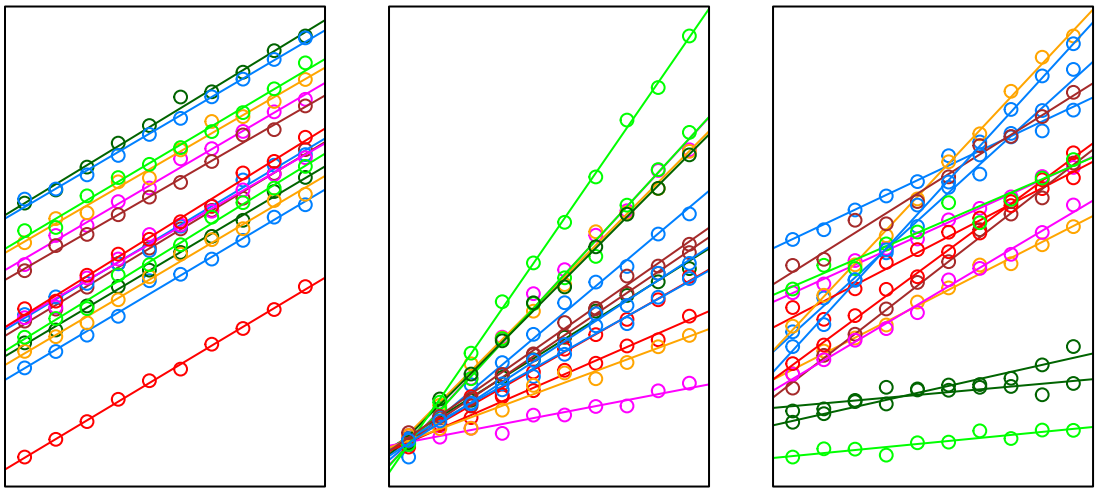


Figure 5.1: Common random coefficient models for grouped data. Each plot consists of ten measurements on 15 subjects—each of which has a positive, linear trend. *Left*: Random intercepts. *Middle*: Random slopes. *Right*: Random intercepts and slopes.

5.1.1 Prediction of future observations.

In the literature for mixed models, prediction of future observations is often overshadowed by the prediction of random effects. Nonetheless, prediction of future observations in mixed models is an important topic—see Jiang (2007) for some motivating examples. For LMMs, there are two kinds of predictions we can make regarding a future observations: (1) predicting a *new* observation *within* an existing group, and (2) predicting a *new* observation in a *new* group. Since longitudinal

studies often aim to make inference for the whole population under study—and not just the groups sampled—we will restrict our attention to case (2). In particular, for calibration, we will assume that a (single) new observation, denoted $\mathcal{Y}_0$, is independent of the current observations and does not belong to an existing group.

5.2 Point estimation

In this section, we discuss point estimation of x_0 in a mixed model setting. Generally, it is difficult to compute the ML estimate of x_0 due to the complex nature of mixed model likelihoods; however, as shown below, some cases yield relatively simple results.

Consider the linear random trend model (Equation (5.4)). For a particular x, we have

$$\mathcal{Y} = \underbrace{\beta_0 + \beta_1 x}_{\text{FIXED}} + \overbrace{\alpha_0 + \alpha_1 x + \epsilon}^{\text{RANDOM}} = \underbrace{\mu\left(x; \boldsymbol{\beta}\right)}_{\text{FIXED}} + \overbrace{R\left(x; \boldsymbol{\alpha}\right)}^{\text{RANDOM}},$$

where $\mu\left(x; \boldsymbol{\beta}\right) = \mathrm{E}\left\{\mathcal{Y}|x\right\}$ is the (population) mean response. For simplicity of notation, let $\mu\left(x\right) = \mu\left(x; \boldsymbol{\beta}\right)$. Solving the equation $\mu(x) = \beta_0 + \beta_1 x$ for x, we get (assuming $\beta_1 \neq 0$)

$$x = \frac{\mu(x) - \beta_0}{\beta_1}.$$

If $\mathcal{Y}_0$ denotes a random observation from a normal distribution with mean $\mu(x_0)$, then $\mathrm{E}\left\{\mathcal{Y}_0\right\} = \mu(x_0) = \beta_0 + \beta_1 x_0$, therefore, a natural estimator of x_0 is

$$\widehat{x}_0 = \frac{\mathcal{Y}_0 - \widehat{\beta}_0}{\widehat{\beta}_1}, \tag{5.5}$$

where $\widehat{\boldsymbol{\beta}} = \left(\widehat{\beta}_0, \widehat{\beta}_1\right)'$ is the EBLUE of $\boldsymbol{\beta} = (\beta_0, \beta_1)'$. Notice how this has the same form as the classical estimator (Equation (3.3)) discussed in Section 3.2. In fact, for the balanced random intercept model, the classical estimator is still the ML estimator

of x_0; in other words, we can compute the ML estimator of x_0 using ordinary least squares with i.i.d. normal errors! To see this, note that for the general case $\boldsymbol{G}^\dagger = \tau \boldsymbol{I}$ and $\boldsymbol{Z}_i = \mathbf{1}_i$ (a column vector of all ones), hence, $\boldsymbol{V}_i = \sigma_\epsilon^2 \left(\boldsymbol{I}_i + \tau \mathbf{1}_i \mathbf{1}_i'\right)$. Therefore, we can write

$$\boldsymbol{\mathcal{Y}}_i \sim \mathcal{N}\left\{\boldsymbol{X}_i \boldsymbol{\beta}, \sigma_\epsilon^2 \left(\boldsymbol{I}_i + \tau \mathbf{1}_i \mathbf{1}_i'\right)\right\}, \quad i = 1, \ldots, m,$$

where $\boldsymbol{X}_i$ is an $n_i \times 2$ design matrix with j-th row equal to $\boldsymbol{X}_{ij}' = (1, x_{ij})$, $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ is a vector of fixed effects, σ_ϵ^2 is the within-subject variance, and $\sigma_\epsilon^2 \tau$ is the variance of the random intercepts. From Equation (5.3), the log-likelihood for the data (ignoring constants) is

$$\begin{aligned}\mathscr{L}_{\mathrm{I}}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau\right) = -\frac{N}{2} \log\left(\sigma_\epsilon^2\right) - \frac{1}{2} \sum_{i=1}^{m} \log\left|\boldsymbol{I} + \tau \mathbf{1}_i \mathbf{1}_i'\right| \\ - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{m} \left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \boldsymbol{\beta}\right)' \left(\boldsymbol{I} + \tau \mathbf{1}_i \mathbf{1}_i'\right)^{-1} \left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \boldsymbol{\beta}\right).\end{aligned}$$

The subscript "I" is there to remind us that this is the likelihood for the standards, the data from the first stage of the calibration experiment (see Section 1). Using the following formulas (Demidenko, 2013, pg. 49),

- $\left|\boldsymbol{I} + \tau \mathbf{1}_i \mathbf{1}_i'\right| = 1 + n_i \tau$;
- $\left(\boldsymbol{I} + \tau \mathbf{1}_i \mathbf{1}_i'\right)^{-1} = \boldsymbol{I} - \frac{\tau}{1 + n_i \tau} \mathbf{1}_i \mathbf{1}_i'$;

the log-likelihood simplifies to

$$\begin{aligned}\mathscr{L}_{\mathrm{I}}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau\right) = -\frac{N}{2} \log\left(\sigma_\epsilon^2\right) - \frac{1}{2} \sum_{i=1}^{m} \log\left(1 + n_i \tau\right) \\ - \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{m} \left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \boldsymbol{\beta}\right)' \left(\boldsymbol{I} - \frac{\tau}{1 + n_i \tau} \mathbf{1}_i \mathbf{1}_i'\right) \left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i \boldsymbol{\beta}\right).\end{aligned}$$

Similarly, the log-likelihood for the (single) unknown (i.e., the log-likelihood for the data from the second stage of the calibration experiment) is

$$\mathscr{L}_{\mathrm{II}}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau, x_0\right) = -\frac{1}{2} \log\left(\sigma_\epsilon^2\right) - \frac{1}{2} \log\left(1 + \tau\right) - \frac{1}{2\sigma_\epsilon^2 \left(1 + \tau\right)} \left(\mathcal{Y}_0 - \beta_0 - \beta_1 x_0\right)^2.$$

From the independence of $\boldsymbol{\mathcal{Y}}$ and $\mathcal{Y}_0$, the log-likelihood for the pooled data, denoted $\mathscr{L}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau, x_0\right)$, is given by

$$\begin{aligned}\mathscr{L}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau, x_0\right) &= \mathscr{L}_{\mathrm{I}}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau\right) + \mathscr{L}_{\mathrm{II}}\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau, x_0\right) \\ &= -\frac{N+1}{2}\log\left(\sigma_\epsilon^2\right) - \frac{1}{2}\sum_{i=1}^{m}\log\left(1+n_i\tau\right) - \frac{1}{2}\log\left(1+\tau\right) \\ &\quad - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)'\left(\boldsymbol{I} - \frac{\tau}{1+n_i\tau}\mathbf{1}_i\mathbf{1}_i'\right)\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right) \\ &\quad - \frac{1}{2\sigma_\epsilon^2\left(1+\tau\right)}\left(\mathcal{Y}_0 - \beta_0 - \beta_1 x_0\right)^2.\end{aligned}$$

Thus, the full log-likelihood is the sum of two parts: the log-likelihood for the standards, and the log-likelihood for the unknown. Equating to zero the partial derivative of the full log-likelihood with respect to the parameter x_0 results in $\widetilde{x}_0\left(\boldsymbol{\beta}\right) = \left(\mathcal{Y}_0 - \beta_0\right)/\beta_1$. In other words, for any value of $\boldsymbol{\beta}$, $\widetilde{x}_0\left(\boldsymbol{\beta}\right)$ maximizes the likelihood with respect to x_0. Plugging this back into the log-likelihood yields the *profiled log-likelihood*

$$\begin{aligned}\mathscr{L}_p\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau\right) &= -\frac{N+1}{2}\log\left(\sigma_\epsilon^2\right) - \frac{1}{2}\sum_{i=1}^{m}\log\left(1+n_i\tau\right) - \frac{1}{2}\log\left(1+\tau\right) \\ &\quad - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)'\left(\boldsymbol{I} - \frac{\tau}{1+n_i\tau}\mathbf{1}_i\mathbf{1}_i'\right)\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right),\end{aligned}$$

Similarly, equating the partial derivative of $\mathscr{L}_p\left(\boldsymbol{\beta}, \sigma_\epsilon^2, \tau\right)$, with respect to the parameter σ_ϵ^2, to zero yields

$$\widetilde{\sigma}_\epsilon^2\left(\boldsymbol{\beta}, \tau\right) = \frac{1}{N+1}\sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\left(\boldsymbol{I} - \frac{\tau}{1+n_i\tau}\mathbf{1}_i\mathbf{1}_i'\right)\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)'.$$

Substituting this back into the profiled log-likelihood and simplifying (i.e., cancelling common factors and ignoring constants) we get

$$\begin{aligned}\mathscr{L}_p\left(\boldsymbol{\beta}, \tau\right) = -\frac{N+1}{2}\log\left\{\sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)'\left(\boldsymbol{I} - \frac{\tau}{1+n_i\tau}\mathbf{1}_i\mathbf{1}_i'\right)\left(\boldsymbol{\mathcal{Y}}_i - \boldsymbol{X}_i\boldsymbol{\beta}\right)\right\} \\ - \frac{1}{2}\sum_{i=1}^{m}\log\left(1+n_i\tau\right) - \frac{1}{2}\log\left(1+\tau\right),\end{aligned}$$

The parameters x_0 and σ_ϵ^2 have been "profiled out", resulting in a simpler log-likelihood in only $p+1$ parameters. We could continue in this fashion with the parameter $\boldsymbol{\beta}$ as well, although it is quite easy to see that the value of $\boldsymbol{\beta}$ that maximizes $\mathscr{L}_p(\boldsymbol{\beta}, \tau)$ is just the usual GLS estimator, $\widetilde{\boldsymbol{\beta}}$, given by Equation (2.19). Furthermore, it can be shown (Demidenko, 2013) that, for the balanced case, $\widetilde{\boldsymbol{\beta}}$ does not depend on τ and in fact reduces to the ordinary LS estimator $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\mathcal{Y}}$. Thus, the ML estimator of x_0 for the balanced random intercept model is simply

$$\widehat{x}_0 = \widetilde{x}_0\left(\widehat{\boldsymbol{\beta}}\right) = \frac{\mathcal{Y}_0 - \widehat{\beta}_0}{\widehat{\beta}_1},$$

where

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sum_{i=1}^n \sum_{j=1}^m (x_{ij} - \bar{x})^2}, \quad \widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \bar{x}.$$

In general, the ML estimator $\widehat{x}_0$ of x_0 is difficult to obtain analytically. For example, in a random slope model, $\mathcal{Y}_0 \sim \mathcal{N}(\beta_0 + \beta_1 x_0, x_0^2\sigma_\alpha^2 + \sigma_\epsilon^2)$, thus, the unknown x_0 affects both the mean and variance of $\mathcal{Y}_0$. A reasonable, and more practical approach, is to proceed as before, that is, by solving the equation $y_0 = \mu(x_0)$ for the unknown x_0. As shown above, for the balanced random intercept model, this leads to the ML estimate.

We should point out that, in general, the EBLUE of $\boldsymbol{\beta}$ (Equation (2.20)) depends on the estimated variance components through $\widehat{\boldsymbol{V}}$. Furthermore, recall that for the linear calibration problem, the bias-adjusted ML estimator of σ_ϵ^2 (the only variance component) is

$$\widehat{\sigma}_\epsilon^2 = \frac{1}{n-2}\sum_{i=1}^n \left(\mathcal{Y}_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2 + \frac{1}{m-1}\sum_{i=n+1}^{n+m}\left(\mathcal{Y}_i - \overline{\mathcal{Y}}_0\right)^2. \tag{5.6}$$

The first term in Equation (5.6) is the usual unbiased estimator of σ_ϵ^2, and the second term is just the sample variance of the m unknowns $\mathcal{Y}_{01}, \ldots, \mathcal{Y}_{0m}$. If there is only one unknown, then the second term in Equation (5.6) is zero! Hence, for a single unknown, the ML estimator of the variance σ_ϵ^2 (adjusted for bias) is unaffected. We

could extrapolate by making a similar argument for calibration in mixed models. That is, if only a single observation, $\mathcal{Y}_0$, is available from the second stage of the calibration experiment, then the ML estimates of the variance components (adjusted for bias) should be unaffected—this would not be the case for replicate unknowns sharing a single unknown x_0. In this chapter, we only concern ourselves with the case of a single unknown; thus, we make the argument that the usual estimates of the variance components are valid for calibration inference.

For illustration, we generated $n = 30$ observations for each of $m = 15$ groups from a random intercept model with fixed effects $\boldsymbol{\beta} = (0, 1)'$ and variance components $\sigma^2_\alpha = 0.01$ and $\sigma^2_\epsilon = 0.001$. A spaghetti plot of the data is shown in Figure 5.2; notice the apparent linear trend with varying intercepts. For a new observation, $y_0 = 0.75$, the estimate of the corresponding unknown x_0 is $\widehat{x}_0 = \left(0.75 - \widehat{\beta}_0\right)/\widehat{\beta}_1 = 0.7819$. Since these data are balanced, $\widehat{x}_0$ is also the ML estimate of x_0.

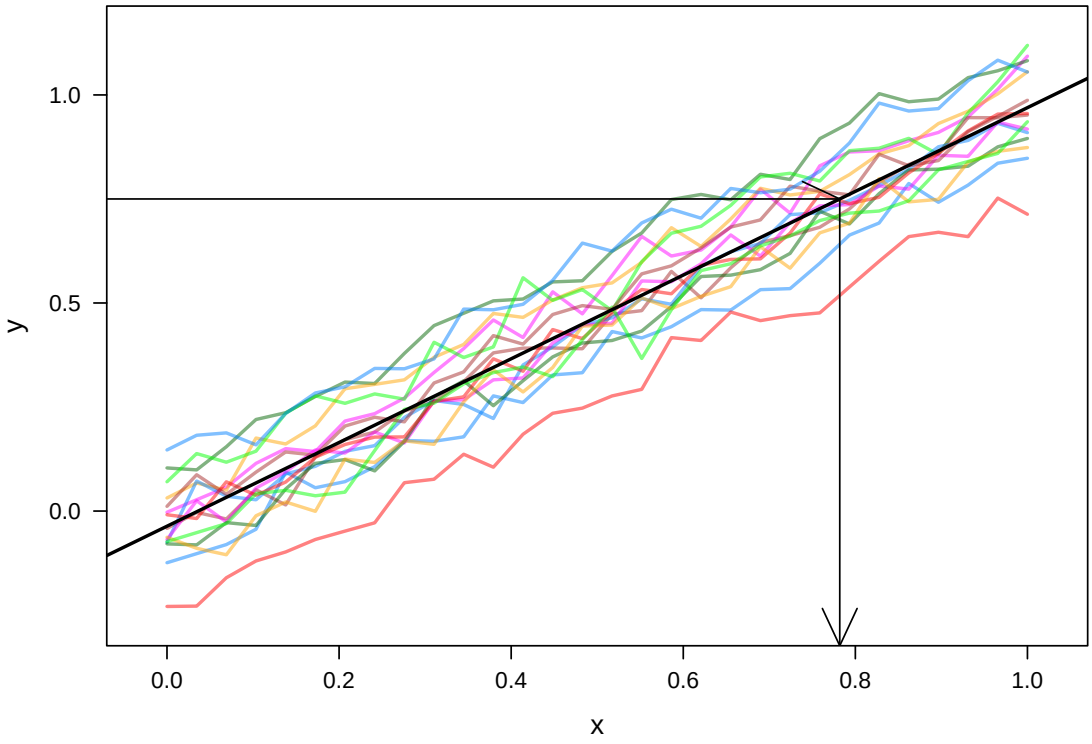


Figure 5.2: Simulated random intercept data.

5.3 Wald interval

The delta method provides the simplest approach to computing confidence intervals, as long as the quantity of interest is a function of random variables that are at least asymptotically normal. Under mild regularity conditions often satisfied in practice, the ML estimator $\widehat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is consistent and asymptotically normal with mean vector $\boldsymbol{\beta}$ and asymptotic variance-covariance matrix $(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$ (Pinheiro, 1994). Furthermore, by assumption, $\mathcal{Y}_0$ is distributed as $\mathcal{N}(\beta_0 + \beta_1 x_0, \sigma_0^2)$, where σ_0^2 denotes the variance of $\mathcal{Y}_0$ which, depending on the random effects structure of the model, may involve the unknown x_0. For instance, if $R(x; \boldsymbol{\alpha}) = \alpha + \epsilon$ (i.e., a random intercept model), then $\sigma_0^2 = \sigma_\alpha^2 + \sigma_\epsilon^2$, whereas if $R(x; \boldsymbol{\alpha}) = \alpha x + \epsilon$ (i.e., a random

slope model), then $\sigma_0^2 = x_0^2\sigma_\alpha^2 + \sigma_\epsilon^2$ which depends on x_0. Although our attention is restricted to LMMs in this dissertation, the simplicity of the Wald-based approach extends to nonlinear mixed-effects models (NLMMs) as well. For an introduction to NLMMs, see Pinheiro (1994) and Pinheiro and Bates (2009).

The Wald statistic, denoted $\mathcal{W}$, is essentially a generalization of the usual $\mathcal{Z}$-statistic. Here it is the point estimator, normalized by an estimate of its standard error which we obtain using Taylor's theorem. By assumption, for large enough N, $\mathcal{W} = \widehat{x}_0^2 / \widehat{\text{Var}}\{\widehat{x}_0\}$ has a $\chi^2(1)$ distribution. As pointed out by Harrell (2001, p. 184), most statistical packages treat $\sqrt{\mathcal{W}}$ as having a t-distribution with degrees of freedom df (rather than a standard normal distribution), however, there is usually no basis for this outside of the ordinary linear model (Gould, 1993).

We discussed the delta method for calibration in Section 3.3.2, where we used a first-order Taylor series expansion to approximate the variance of $\widehat{x}_0$. The same basic procedure holds here for LMMs except now $\text{Var}\{\mathcal{Y}_0\}$ contains additional variability attributed by the random effects. As a simple example, we consider the balanced random intercept model. For this case, we have shown that the ML estimator of x_0 is given by $\widehat{x}_0 = \left(\mathcal{Y}_0 - \widehat{\beta}_0\right)/\widehat{\beta}_1$. We should make clear, however, that even though $\widehat{\boldsymbol{\beta}}$ does not depend on the variance components, its variance-covariance matrix does! Therefore, we still need to compute the variance-covariance matrix of $\widehat{\boldsymbol{\beta}}$ from a mixed model; for example, the variance-covariance matrix obtained from the `SAS/STAT` (SAS Institute Inc., 2011) software's `PROC MIXED` procedure or the `lme` function from the `R` package `nlme` (Pinheiro et al., 2013). The variance-covariance matrix of $\left(\mathcal{Y}_0, \widehat{\boldsymbol{\beta}}\right)$ is

$$\Sigma = \begin{bmatrix} \text{Var}\{\mathcal{Y}_0\} & \mathbf{0} \\ \mathbf{0} & \text{Var}\left\{\widehat{\boldsymbol{\beta}}\right\} \end{bmatrix} = \begin{bmatrix} \sigma_0^2 & \mathbf{0} \\ \mathbf{0} & (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1} \end{bmatrix}, \tag{5.7}$$

Since $\mathcal{Y}_0$ is independent of $\boldsymbol{\mathcal{Y}}$, it is also independent of $\widehat{\boldsymbol{\beta}}$. Recall that our point estimate has the form $x = \mu^{-1}(y; \boldsymbol{\beta})$. Let $\mu_1^{-1}(y; \boldsymbol{\beta})$ and the vector-valued function

$\mu_2^{-1}(y;\boldsymbol{\beta})$ be the partial derivatives of μ^{-1} with respect to the parameters y and $\boldsymbol{\beta}$, respectively. Our point estimator is given by $\mu^{-1}\left(\mathcal{Y}_0;\widehat{\boldsymbol{\beta}}\right)$, where $\mathcal{Y}_0$ is a new observation and $\widehat{\boldsymbol{\beta}}$ is the EBLUE of $\boldsymbol{\beta}$. Using a first-order Taylor-series expansion, an approximate variance for $\widehat{x}_0$ is given by

$$\text{(5.8)}\quad \text{Var}\left\{\widehat{x}_0\right\} = \left[\mu_1^{-1}\left(\mathcal{Y}_0;\widehat{\boldsymbol{\beta}}\right)\right]^2\sigma_0^2 + \left[\mu_2^{-1}\left(\mathcal{Y}_0;\widehat{\boldsymbol{\beta}}\right)\right]'\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left[\mu_2^{-1}\left(\mathcal{Y}_0;\widehat{\boldsymbol{\beta}}\right)\right].$$

Of course, to obtain $\widehat{\text{Var}}\left\{\widehat{x}_0\right\}$, we need to replace σ_0^2 and $\boldsymbol{V}$ in Equation (5.8) with their corresponding estimates $\widehat{\sigma}_0^2$ and $\widehat{\boldsymbol{V}}$, respectively. Once we have $\widehat{\text{Var}}\left\{\widehat{x}_0\right\}$, an approximate $100(1-\alpha)\%$ confidence interval for $\widehat{x}_0$ is given by $\widehat{x}_0 \pm z_{1-\alpha/2}\sqrt{\widehat{\text{Var}}\left\{\widehat{x}_0\right\}}$.

As pointed out by Davidian and Giltinan (1995, p. 283), the first term in Equation (5.8) usually (but not necessarily) dominates the second term, leaving the cruder approximation $\text{Var}\left\{\widehat{x}_0\right\} = \left[\mu_1^{-1}\left(\mathcal{Y}_0;\widehat{\boldsymbol{\beta}}\right)\right]^2\sigma_0^2$. For example, consider the random intercept model discussed earlier. For this model, we have

$$\begin{aligned}\mu^{-1}(y;\boldsymbol{\beta}) &= (y-\beta_0)/\beta_1,\\ \frac{\partial}{\partial y}\mu^{-1}(y;\boldsymbol{\beta}) &= 1/\beta_1,\end{aligned}$$

thus, for the balanced case, the crude approximation turns out to be $\text{Var}_{\text{crude}}\left\{\widehat{x}_0\right\} = \sigma_\epsilon^2(1+\tau)/\widehat{\beta}_1^2$. In general, the approximate variance (Equation (5.8)) can be difficult to compute by hand leading to widespread use of the cruder formula. However, given the ease with which complex computations can be carried out numerically (and with great accuracy), it is good practice to compute and use both terms. In Section 5.8, we illustrate with a real data example how easily this can be done using the software `R` (in particular, see Example 5.8.1).

For the simulated data in Figure 5.2, we get an approximate standard error (i.e., using Taylor's theorem) of 0.1 which produces a Wald-based calibration interval for x_0 of $(0.5859, 0.9778)$. Compare this to the cruder interval $(0.5915, 0.9722)$ obtained by using $\text{Var}_{\text{crude}}\left\{\widehat{x}_0\right\}$ in place of the full Taylor-series estimate (5.8).

Unfortunately, one of the difficulties with inference in LMMs is that

$$\text{Var}\left\{\widehat{\boldsymbol{\beta}}\right\} = \left(\boldsymbol{X}'\widehat{\boldsymbol{V}}^{-1}\boldsymbol{X}\right)$$

involves $\widehat{\boldsymbol{V}}$, but does not take into account its variability. In other words, $\text{Var}\left\{\widehat{\boldsymbol{\beta}}\right\}$ underestimates the true variance of $\widehat{\boldsymbol{\beta}}$ (see, for example, McCulloch et al. (2008, pp. 165-167)). Thus, any inference relying on $\text{Var}\left\{\widehat{\boldsymbol{\beta}}\right\}$, including the Wald interval for calibration just discussed, may be misleading. An alternative is to use the so-called *parametric bootstrap* to compute a bootstrap estimate of the variance-covariance matrix $\widehat{\boldsymbol{\beta}}$ to use in place of $\text{Var}\left\{\widehat{\boldsymbol{\beta}}\right\}$ in Equation (5.7) (see steps (1)-(4) of Algorithm 2 on page 85).

5.4 Inversion interval

We can also obtain a confidence interval for the unknown x_0 by inverting an asymptotic prediction interval for $\mathcal{Y}_0$. Let $\boldsymbol{X}_0$ have the same form as the i-th row of $\boldsymbol{X}$, but with x_{ij} replaced with x_0. For example, if $\mu(x_{ij};\boldsymbol{\beta}) = \beta_0 + \beta_1 x_{ij} + \beta_2 x_{ij}^2$, then $\boldsymbol{X}_0 = (1, x_0, x_0^2)'$.

For brevity, let $\mu_0 = \mu(x_0;\boldsymbol{\beta})$, $\widetilde{\mu}_0 = \mu\left(x_0;\widetilde{\boldsymbol{\beta}}\right)$, and $\widehat{\mu}_0 = \mu\left(x_0;\widehat{\boldsymbol{\beta}}\right)$ where, as before, $\widetilde{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\beta}}$ denote the BLUE and EBLUE of $\boldsymbol{\beta}$, respectively. A new observation, $\mathcal{Y}_0$ say, with unknown x_0, is distributed as $\mathcal{N}(\mu_0, \sigma_0^2)$. Clearly, $\mathcal{Y}_0 - \widetilde{\mu}_0$ is a normally distributed random variable with expectation zero (since $\text{E}\{\mathcal{Y}_0\} = \text{E}\{\widetilde{\mu}_0\} = \mu_0$). Also, note that $\mathcal{Y}_0$ and $\widetilde{\mu}_0$ are independent, hence, $\text{Cov}\{\mathcal{Y}_0, \widetilde{\mu}_0\} = 0$. It therefore follows that the statistic

$$\mathcal{Z} = \frac{\mathcal{Y}_0 - \widetilde{\mu}_0}{\sqrt{\text{Var}\{\mathcal{Y}_0\} + \text{Var}\{\widetilde{\mu}_0\}}} = \frac{\mathcal{Y}_0 - \widetilde{\mu}_0}{\sqrt{\sigma_0^2 + \boldsymbol{X}_0'\left(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X}'\right)^{-1}\boldsymbol{X}_0}} \tag{5.9}$$

is a *pivotal quantity* that has an asymptotic standard normal distribution or, equivalently, $\mathcal{Z}^2 \dot{\sim} \chi_1^2$ (a Chi-squared distribution with a single degree of freedom). The key here is that $\widetilde{\mu}_0$ is a linear function of $\boldsymbol{\mathcal{Y}}$. For example, for a balanced random

intercept model, Equation (5.9) reduces to

$$\mathcal{Z} = \frac{\mathcal{Y}_0 - \widetilde{\beta}_0 - \widetilde{\beta}_1 x_0}{\sqrt{\sigma_\epsilon^2 + \sigma_\alpha^2 + \boldsymbol{X}_0' \left[\boldsymbol{X}' \left(\sigma_\alpha^2 \boldsymbol{I}_m \otimes \boldsymbol{J}_n + \sigma_\epsilon^2 \boldsymbol{I}_N\right)^{-1} \boldsymbol{X}\right]^{-1} \boldsymbol{X}_0}},$$

where $\boldsymbol{J}_n = \mathbf{1}_n \mathbf{1}_n'$ is an $n \times n$ vector of all ones.

In practice, σ_0^2, $\boldsymbol{V}$, and x_0 are usually unknown and need to be estimated from the data. In such cases, an approximate pivot (essentially a Wald statistic), denoted $\mathcal{Q}$, can be obtained by replacing σ_0^2, $\boldsymbol{V}$, and x_0 in the above equation with their respective estimates $\widehat{\sigma}_0^2$, $\widehat{\boldsymbol{V}}$, and $\widehat{x}_0$. This suggests an approximate $100(1-\alpha)\%$ confidence interval for x_0 of

$$\widehat{\mathcal{J}}_{\text{cal}}(x) = \left\{x : z_{\alpha/2} < \mathcal{Q} < z_{1-\alpha/2}\right\}, \tag{5.10}$$

where $z_{\alpha/2} = z_{1-\alpha/2}$ denote the $\alpha/2$ and $1-\alpha/2$ quantiles of a standard normal distribution, respectively. Similar to the approximate predictive pivot (Equation (3.14)), it is unlikely that Equation (5.10) will yield closed-form solutions, thus, the solution must be obtained numerically. For the simulated data example, a 95% inversion interval based on Equation (5.10), corresponding to $y_0 = 0.75$, is given by $(0.5859, 0.9779)$, which is very similar to the Wald-based intervals obtained earlier.

5.5 Parametric bootstrap

In Section 3.3.3, we discussed how to calculate calibration intervals based on the nonparametric bootstrap. A crucial assumption for the ordinary nonparametric bootstrap, however, is that the data are independent; for reasons discussed earlier, this assumption is typically not valid for grouped data. Nonetheless, a different kind of bootstrap, called the *parametric bootstrap*, has shown promise as a serious inferential tool. For examples, see McCulloch et al. (2008, pg. 342) and Efron (2011). When applicable, the parametric bootstrap typically gives answers similar to a Bayesian analysis with uninformative priors, however, it is much faster than

MCMC simulations (bootstrap simulations usually only require a few thousand iterations whereas traditional MCMC simulations may require tens, or even hundreds of thousands, of iterations). The parametric bootstrap essentially entails sampling from the fitted model itself, rather than sampling (with replacement) from the data. For controlled calibration in a mixed model setting, we propose the following algorithm (essentially a parametric version of Algorithm 1 based on the LMM instead of the ordinary LM):

Algorithm 2: Parametric bootstrap for controlled calibration in LMMs.

for $r = 1$ **to** R **do**

(1) generate q new values of the random effects, denoted $\boldsymbol{\alpha}_r^\star$, from a $\mathcal{N}\left(\mathbf{0}, \widehat{\boldsymbol{G}}\right)$ distribution;

(2) generate N new errors, denoted $\boldsymbol{\epsilon}_r^\star$, from a $\mathcal{N}\left(\mathbf{0}, \widehat{\sigma}_\epsilon^2 \boldsymbol{I}\right)$ distribution;

(3) set $\boldsymbol{y}_r^\star = \boldsymbol{X}\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}\boldsymbol{\alpha}_r^\star + \boldsymbol{\epsilon}_r^\star$;

(4) update the original model using $\boldsymbol{y}_r^\star$ as the response vector to obtain $\widehat{\boldsymbol{\beta}}_r^\star$;

(5) generate $y_{0r}^\star$ from a $\mathcal{N}\left(y_0, \widehat{\sigma}_0^2\right)$ distribution;

(6) compute $\widehat{x}_{0r}^\star = \mu^{-1}\left(y_{0r}^\star; \widehat{\boldsymbol{\beta}}_r^\star\right)$.

end

Note that only steps (5) and (6) are specific to calibration. Similar parametric bootstrap schemes have also been proposed for mixed models. For example, we can condition on the current values of the random effects by ignoring step (1) of Algorithm 2 and using the current EBLUP, $\widehat{\boldsymbol{\alpha}}$, in place of $\boldsymbol{\alpha}^\star$. Semiparametric variants of Algorithm 2 that involve sampling directly from the EBLUP and residuals have also been proposed, but Morris (2002) considers this to be bad practice because it consistently underestimates the true variation in the data.

We applied Algorithm 2 to the simulated data from Figure 5.2. A histogram of the $R = 9,999$ bootstrap replicates of $\widehat{x}_0$ is shown in Figure 5.3. Not surprisingly,

the distribution is reasonably symmetric and approximately normal; the normal Q-Q plot also confirms this. These bootstrap replicates were used to produce the last two confidence intervals in Table 5.1. For comparison, we have also included the calibration intervals computed in the previous sections. The results are all very similar and there is little reason here for choosing one interval over another. The Wald-based intervals are symmetric, but, as can be seen from Figure 5.3, symmetry is not unrealistic for this example (this is not the case for the example given in the next section). The inversion interval is not symmetric about $\widehat{x}_0$, as well as the bootstrap intervals, however, the bootstrap approach has the advantage of providing an estimate of the entire sampling distribution of $\widehat{x}_0$. It should be noted, though, that the parametric bootstrap assumes that the model specified for the data is correct! If, however, the data were not normal, then all of these intervals would likely produce misleading results. In the next section, we discuss a potential remedy that can be used for non-Gaussian LMMs, that is, LMMs that do not assume a specific distribution for the random effects or the errors.

Table 5.1: Approximate 95% calibration intervals for the simulated balanced random intercept example. The intervals based on the parametric bootstrap are labeled (PB).

Interval	Estimate	Lower 2.5%	Upper 97.5%	Length	SE
Wald	0.7819	0.5859	0.9778	0.3920	0.0999
Crude interval	0.7819	0.5915	0.9722	0.3807	0.0971
Inversion	0.7819	0.5859	0.9779	0.3920	NA
Normal (PB)	0.7819	0.5873	0.9742	0.3870	0.0987
Percentile (PB)	0.7819	0.5895	0.9789	0.3893	0.0987

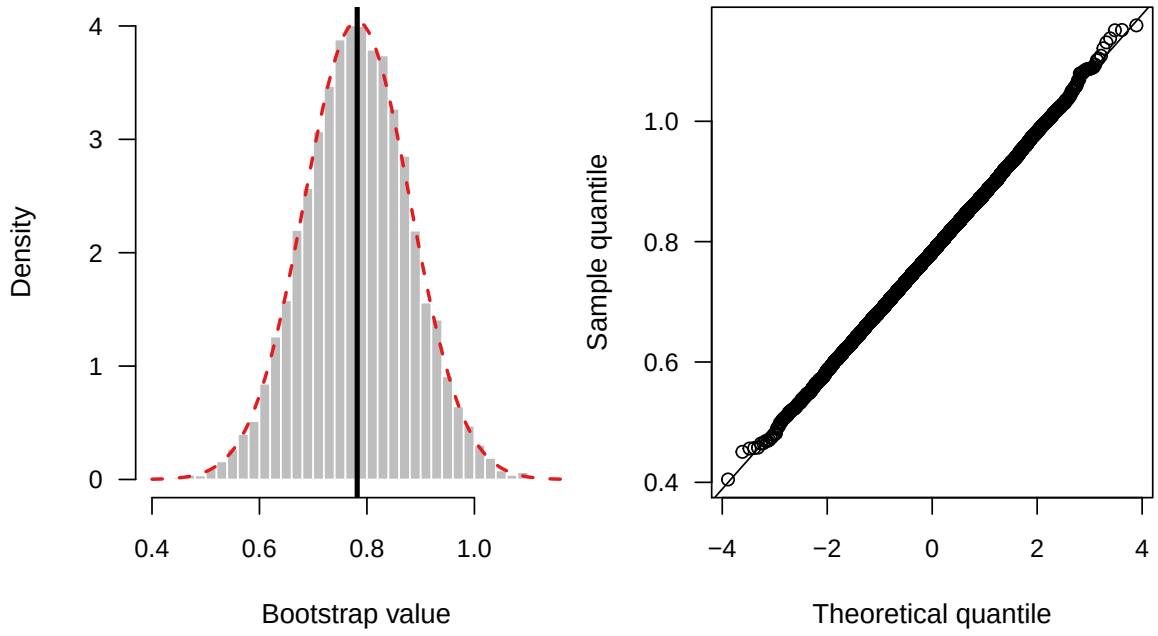


Figure 5.3: Bootstrap distribution of $\widehat{x}_0$ obtained using Algorithm 2. The dotted red curve represents a normal distribution with mean $\widehat{x}_0$ and standard deviation estimated from the bootstrap replicates $\widehat{x}^{\star}_{0r}$. The vertical black line indicates the position of $\widehat{x}_0$.

5.5.1 Parametric bootstrap adjusted inversion interval.

Although we favor the bootstrap confidence intervals obtained directly from the R bootstrap replicates of $\widehat{x}_0$, researchers are likely more familiar with the inversion and Wald-based intervals discussed in the previous two sections. These intervals, however, use the quantiles from a standard normal distribution (i.e., rely on normal approximations). The parametric bootstrap can be used to improve upon these intervals by replacing the standard normal quantiles with more accurate ones. For instance, for the inversion interval, at each run in Algorithm 2, we compute

$$\mathcal{Q}^{\star} = \frac{y_0^{\star} - \mu\left(\widehat{x}_0; \widehat{\boldsymbol{\beta}}^{\star}\right)}{\sqrt{\widehat{\sigma}_0^{2\star} + \boldsymbol{X}_0' \left(\boldsymbol{X}' \widehat{\boldsymbol{V}}^{*-1} \boldsymbol{X}'\right)^{-1} \boldsymbol{X}_0}},$$

where the denominator is evaluated at $x_0 = \widehat{x}_0$. As a result, we obtain the R bootstrap values $\mathcal{Q}_r^\star$. Let $\gamma_{\alpha/2}^\star$ and $\gamma_{1-\alpha/2}^\star$ denote the sample $\alpha/2$ and $1-\alpha/2$ quantiles of $\mathcal{Q}_r^\star$, respectively. A bootstrap adjusted inversion interval for x_0 is then given by

$$\widehat{\mathcal{J}}_{\text{cal}}^\star(x) = \left\{x : \gamma_{\alpha/2}^\star < \mathcal{Q} < \gamma_{1-\alpha/2}^\star\right\}. \tag{5.11}$$

We illustrate this on the bladder volume example in Section 5.8. A similar adjustment can also be made for the Wald-based interval as well; this is very similar to the studentized bootstrap procedure outlined in steps (6)-(7) of Algorithm 1.

5.6 Distribution free calibration interval

For certain cases, we can easily obtain an asymptotic prediction interval for a future observation that does not require normality for the random effects or errors. These intervals are called *distribution-free* prediction intervals; for details, the interested reader is pointed to Jiang and Zhang (2002) or Jiang (2007). This suggests the possibility of a distribution-free calibration interval for x_0 by inverting a corresponding distribution-free prediction interval.

We consider only the case of *standard* LMMs. Following Jiang and Zhang (2002) and Jiang (2007), an LMM is said to be standard if each $\boldsymbol{Z}_i$ of Equation (5.2) consists of only 0's and 1's, such that each row contains exactly one 1 and each column has at least one 1. The random intercept model (balanced or unbalanced case) is standard in this sense; the random slope model, however, is not.

The method for standard LMMs turns out to be quite simple. Compute the ordinary LS estimate $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$, then obtain the residual vector as $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$. Denote the $\alpha/2$ and $1-\alpha/2$ quantiles of the residuals as $e_{\alpha/2}$ and $e_{1-\alpha/2}$, respectively. Then, a distribution-free prediction interval for a new observation, y_0, with asymptotic coverage probability $1-\alpha$ is given by $\left[\widehat{\mu}(x_0) + e_{\alpha/2}, \widehat{\mu}(x_0) + e_{1-\alpha/2}\right]$, where $\widehat{\mu}(x_0)$ is the empirical best predictor of y_0. For example, for a random intercept

model, the interval in question is simply $\left(\widehat{\beta}_0 + \widehat{\beta}_1 x_0 + e_{\alpha/2}, \widehat{\beta}_0 + \widehat{\beta}_1 x_0 + e_{1-\alpha/2}\right)$. If y_0 is observed and x_0 is the unknown, then this formula can be inverted (in closed-form) to produce an asymptotic $100(1-\alpha)\%$ distribution-free confidence interval for x_0 of

$$\left[\frac{(y_0 - e_{1-\alpha/2}) - \widehat{\beta}_0}{\widehat{\beta}_1}, \frac{(y_0 - e_{\alpha/2}) - \widehat{\beta}_0}{\widehat{\beta}_1}\right]. \tag{5.12}$$

Using Equation (5.12), a 95% distribution-free calibration interval for the simulated random intercept example is $(0.6121, 0.9852)$. However, since these data are normal (we know because we generated the data), the intervals given in Table 5.1 are probably more accurate.

5.7 Simulation study

To illustrate the practical performance of the confidence interval procedures discussed in Sections 5.3-5.5, we ran a small simulation study in which $(x_{ij}, \mathcal{Y}_{ij})$, $i = 1, 2, \ldots, 30$, $j = 1, 2, \ldots, 20$, were generated from both

$$\mathcal{Y}_{ij} = (\alpha_{0j} + \beta_0) + (\alpha_{1j} + \beta_1)\, x_{ij} + \epsilon_{ij}$$

and

$$\mathcal{Y}_{ij} = (\alpha_{0j} + \beta_0) + (\alpha_{1j} + \beta_1)\, x_{ij} + \beta_2 x_{ij}^2 + \epsilon_{ij}.$$

In the first model, $(\beta_0, \beta_1) = (0, 2)$, $\alpha_{0j} \overset{iid}{\sim} \mathcal{N}(0, 0.001)$, $\alpha_{1j} \overset{iid}{\sim} \mathcal{N}(0, 0.05)$, and $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, 0.001)$. Similarly, in the second model, $(\beta_0, \beta_1, \beta_2) = (0, 3, -1)$, $\alpha_{0j} \overset{iid}{\sim} \mathcal{N}(0, 0.0001)$, $\alpha_{1j} \overset{iid}{\sim} \mathcal{N}(0, 0.05)$, and $\epsilon_{ij} \overset{iid}{\sim} \mathcal{N}(0, 0.001)$. In both models, the random variables $\{\alpha_{0j}\}$, $\{\alpha_{1j}\}$, and $\{\epsilon_{ij}\}$ are mutually independent. We computed separate 95% calibration intervals using each method corresponding to five different unknowns: $\mathcal{Y}_0 = \{0, 0.5, 1, 1.5, 2\}$. An example data set from each model is displayed in Figure 5.4.

Due to the large computational time involved in bootstrapping and fitting mixed models, the Monte Carlo sample size was limited to $N = 1,000$ and the bootstrap

sample size was set to $R = 999$. Furthermore, because we used a confidence level of $1 - \alpha = 0.95$, the standard deviation of the coverage probability is approximately $\sqrt{0.95\,(1 - 0.95)\,/1000} = 0.007$; thus, we only report two decimal places of precision in Tables 5.2 and 5.3.

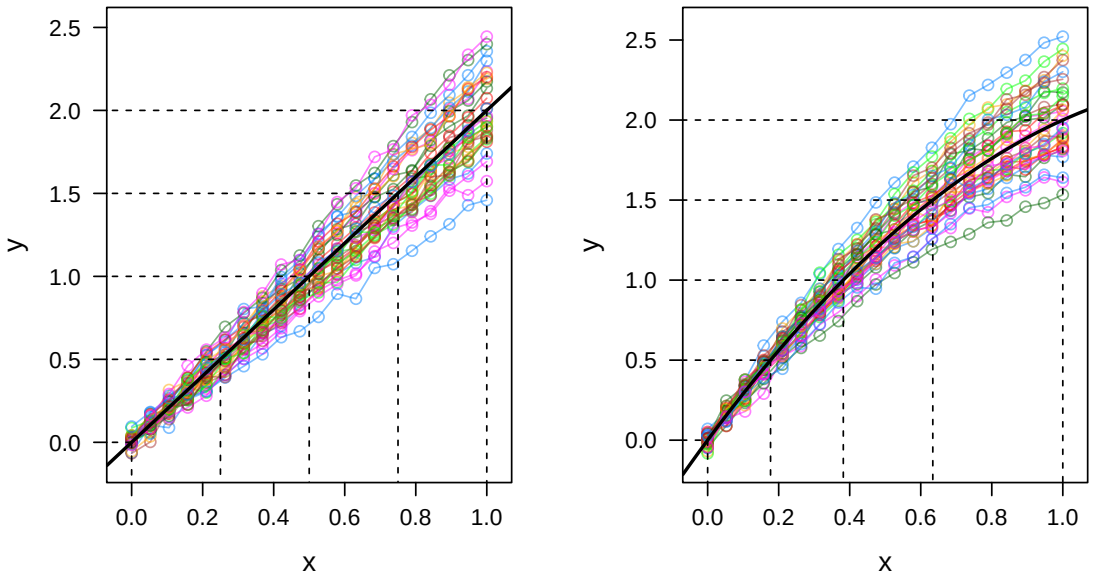


Figure 5.4: Scatterplots of simulated data. *Left*: Data from a balanced random intercept and slope model. *Right*: Data from a balanced random intercept and slope model with a fixed quadratic term. The mean response is shown as a solid black curve and the dashed lines indicate the positions of the true unknowns $(x_0, \mathcal{Y}_0)$.

As expected, all four procedures perform well for the balanced random intercept and slope model, with coverage estimates close to 0.95 (see Table 5.2). For the model with the quadratic term, the methods still performed well, however, the Wald-based interval performed much worse for $\mathcal{Y}_0 = 2$, with coverage probability well below 0.95 (see Table 5.3). A decrease in coverage was anticipated since we do not expect the sampling distribution of the inverse estimator, $\widehat{x}_0$, to be approximately normal (or

Table 5.2: Coverage probability and length estimates of calibration intervals corresponding to various values of $\mathcal{Y}_0$ in a balanced random intercept and slope model. The Monte Carlo sample size was $N = 1,000$ and the bootstrap sample size was $R = 999$.

	$\mathcal{Y}_0 =$	0	0.5	1	1.5	2
Coverage	Wald	0.95	0.94	0.93	0.94	0.93
	Inversion	0.95	0.95	0.94	0.93	0.95
	Percentile bootstrap	0.95	0.94	0.93	0.93	0.94
	Adjusted inversion	0.96	0.94	0.94	0.94	0.96
Length	Wald	0.09	0.14	0.24	0.35	0.45
	Inversion	0.09	0.14	0.24	0.34	0.45
	Percentile bootstrap	0.09	0.14	0.24	0.34	0.45
	Adjusted inversion	0.09	0.14	0.24	0.36	0.47

Table 5.3: Coverage probability and length estimates of 95% calibration intervals corresponding to various values of $\mathcal{Y}_0$ in a balanced random intercept and slope model containing a quadratic term. The Monte Carlo sample size was $N = 1,000$ and the bootstrap sample size was $R = 999$.

	$\mathcal{Y}_0 =$	0	0.5	1	1.5	2
Coverage	Wald	0.95	0.96	0.94	0.94	0.87
	Inversion	0.96	0.95	0.94	0.93	0.93
	Percentile bootstrap	0.94	0.93	0.93	0.93	0.92
	Adjusted inversion	0.94	0.94	0.94	0.93	0.92
Length	Wald	0.04	0.08	0.17	0.35	1.09
	Inversion	0.04	0.08	0.16	0.38	1.36
	Percentile bootstrap	0.04	0.08	0.16	0.37	0.63
	Adjusted inversion	0.04	0.08	0.17	0.37	1.37

even symmetric) in this portion of the curve. Surprisingly, the inversion interval seems to perform as well as the intervals based on the parametric bootstrap with respect to coverage. However, the interval based on the percentile bootstrap has the advantage of providing an estimate of the entire sampling distribution of $\widehat{x}_0$, rather than just a confidene interval, as well as good width. Also, the Wald-based and inversion intervals do not account for the uncertainty in the estimated variance components and rely on large-sample sizes, therefore, these intervals would not be expected to perform as well on smaller data sets. A more extensive simulation study allowing for different sample size configurations would be necessary to further substantiate these claims.

5.8 Bladder volume example

In this section, we discuss an example involving a real dataset taken from Brown (1993) where the author states:

> "A series of 23 women patients attending a urodynamic clinic were recruited for the study. After successful voiding of the bladder, sterile water was introduced in additions of 10, 15, and then 25 ml increments up to a final cumulative total of 175 ml. At each volume a measure of height (`H`) in mm and depth (`D`) in mm of largest ultrasound bladder images were taken. The product `H` $\times$ `D` was taken as a measure of liquid volume..."

The product of `H` and `D` is plotted against true volume in Figure 5.5 for each of the 23 subjects. As can be seen from this plot, each subject has a slightly nonlinear trend and there are some missing observations (i.e., the data are unbalanced).

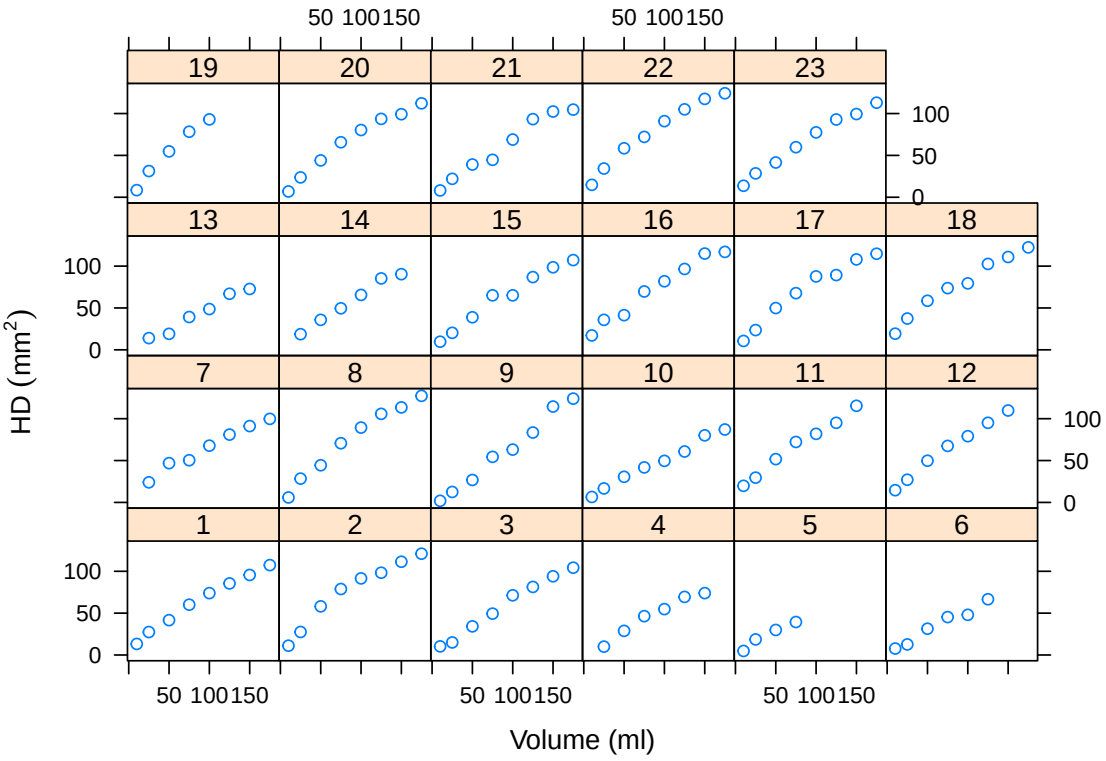


Figure 5.5: Scatterplot of the bladder volume data.

Finding an appropriate random effects structure for the model can be difficult. An informal, but simple, approach is to fit the same model to the data for each subject and compare the estimated coefficients. To this end, we fit a simple quadratic model, $\mu(x) = \beta_0 + \beta_1 x + \beta_2 x^2$, to each of the 23 subjects measurements. Plots of the individual 95% confidence intervals are displayed in Figure 5.6. We should point out that we used orthogonal polynomials to obtain each fit; this was done to remove the correlation between the estimated coefficients. Clearly, the intercept and slope parameters, β_0 and β_1, vary greatly between subjects, while β_2 remains relatively constant throughout. This suggests an LMM with random effects for the intercept

and linear terms. In particular, we used the following random coefficient model:

$$\texttt{HD}_{ij} = \beta_0 + \alpha_{0i} + (\beta_1 + \alpha_{1i})\,\texttt{volume}_{ij} + \beta_3 \texttt{volume}_{ij}^2. \tag{5.13}$$

Notice this is just the linear random trend model (Equation (5.4)) with an additional quadratic term. Figure 5.7 shows the subject-specific fits based on model (5.13). Clearly, the model does a good job describing the data. Figure 5.7 also shows how each subject deviates from the overall fitted mean response $\widehat{\mu}\,(x) = \widehat{\beta}_0 + \widehat{\beta}_1 x + \widehat{\beta}_2 x^2$.

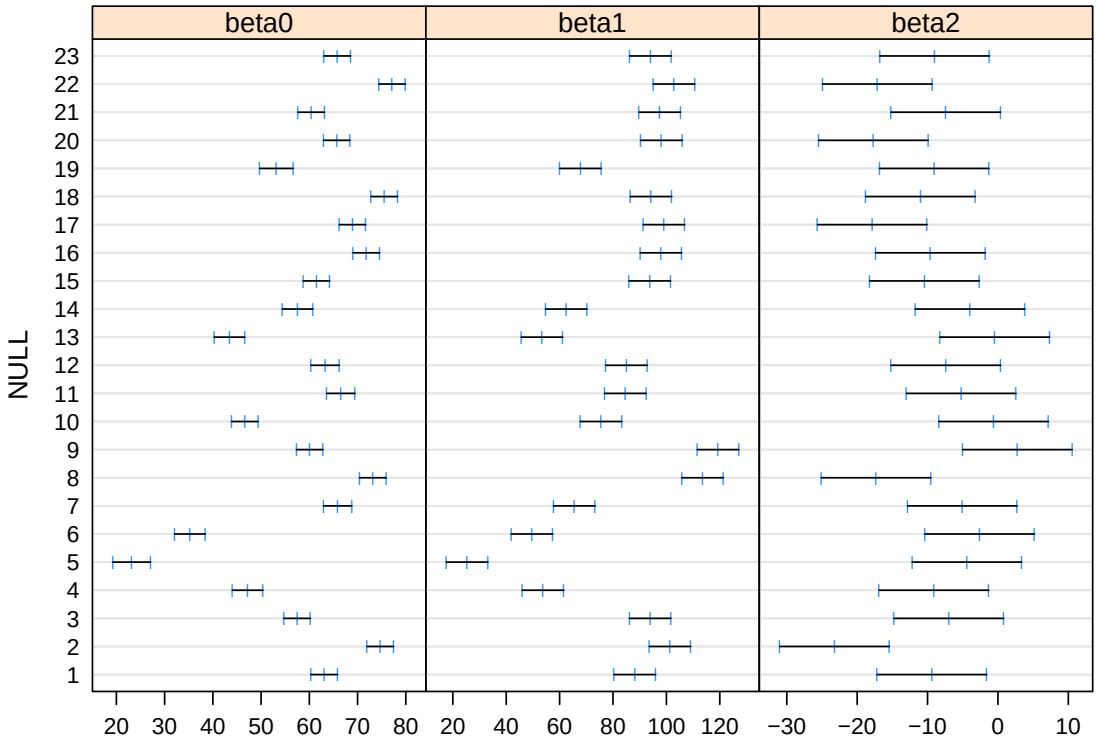


Figure 5.6: 95% confidence intervals for regression coefficients from subject-specific fits to the bladder volume data. Note that we used orthogonal polynomials in order to remove the correlation between the estimated coefficients.

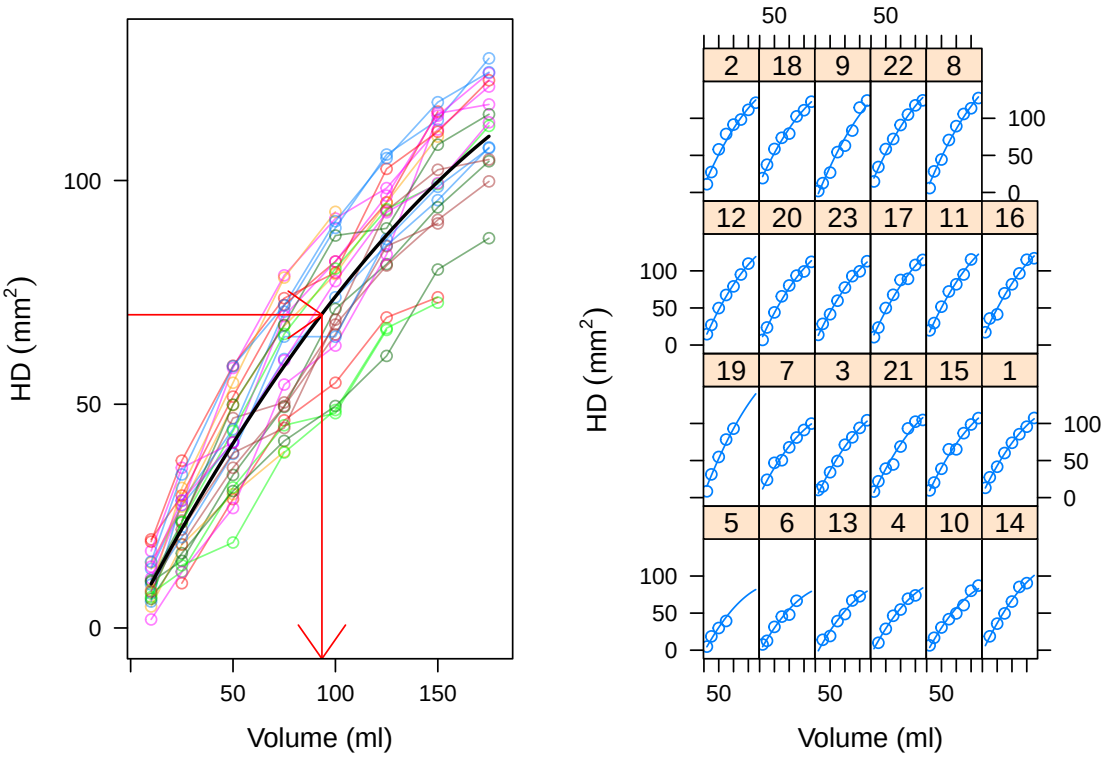


Figure 5.7: Scatterplot of the bladder volume data with fitted mean response. *Left*: Scatterplot of the bladder volume data with fitted mean response (solid black curve). The horizontal red arrows indicate the positions of the unknown $\texttt{HD}_0 = 70$ mm^2 and the point estimate $\widehat{\texttt{volume}}_0$ obtained from inverting the fitted mean response. *Right*: Fitted curves from a quadratic model fit to the bladder volume data. The model includes (uncorrelated) random effects for the intercept and linear term for each subject.

Suppose we obtain a new observation, $\texttt{HD}_0 = 70$ mm^2, for which the true volume is unknown. To estimate the unknown volume, denoted $\texttt{volume}_0$, we proceed as discussed at the end of Section 5.2. In particular, we solve the equation

$$\widehat{\mu}\left(\texttt{volume}_0\right) = \widehat{\beta}_0 + \widehat{\beta}_1 \texttt{volume}_0 + \widehat{\beta}_2 \texttt{volume}_0^2 = 70 \text{ mm}^2$$

for $\texttt{volume}_0$ using the quadratic formula. The point estimate obtained is $\widehat{\texttt{volume}}_0 = 93.4124$ ml (see Figure 5.7). Since $\widehat{\texttt{volume}}_0$ is not the ML estimate, we cannot compute

an approximate ML interval as we could for the balanced random intercept example. However, the following snippet of R code shows how to use the well-known car package (Fox and Weisberg, 2011) to compute the approximate standard error based on the first-order Taylor series estimate given in Equation (5.8). Note that our model has $\text{Var}\{\mathcal{Y}_0\} = \sigma_0^2 + x_0^2\sigma_1^2 + \sigma_\epsilon^2$ where σ_0^2 and σ_1^2 are the variances of the random intercept and linear terms, respectively. If the random effects were correlated, there would be an additional term $x_0\sigma_{01}$ where $\sigma_{01} = \text{Cov}\{\alpha_{0i}, \alpha_{1i}\}$.

R Example 5.8.1.

```
## Obtain fitted model
mod <- lme(HD ~ volume + I(volume^2), random = pdDiag(~volume),
    data = Bladder)
b <- as.numeric(fixef(mod))  # vector of fixed effects
## Set up variance-covariance matrix from Equation (5.7)
var.y0 <- getVarCov(mod)[1, 1] + x0.est^2 * getVarCov(mod)[2, 2] +
    summary(mod)$sigma^2
covmat <- diag(4)
covmat[1:3, 1:3] <- vcov(mod)
covmat[4, 4] <- var.y0
## Call the deltaMethod() function from package car
dm <- car:::deltaMethod(c(b0 = b[1], b1 = b[2], b2 = b[3], y0 = 70),
    g = "(-b1+sqrt(b1^2-4*b2*(b0-y0)))/(2*b2)", vcov. = covmat)
rownames(dm) <- ""
dm

##  Estimate    SE
##     93.41 19.63
```

The resulting standard error is 19.63 which yields an approximate (Wald-based) 95% confidence interval for $\texttt{volume}_0$ of $(54.94, 131.89)$. For comparison, we computed the same interval assuming the data are cross-sectional, that is, by ignoring the grouped structure of the data and using the methods discussed in Section 3.3.2. The resulting interval is $(56.43, 130.77)$ which, as expected, is narrower than the one previously obtained.

Obtaining the inversion interval (Equation (5.10)) is less straightforward; we have to write our own prediction function in `R` that will also return the standard errors of the fitted values. Example 5.8.2 shows the minimal `R` code necessary to obtain $\widehat{\mathcal{J}}_{\text{cal}}(x)$ for the bladder volume example. The first line simply extracts the point estimate, $\widehat{\texttt{volume}}_0$, obtained in Example 5.8.1. The next few lines of code define a new prediction function, `predFun`, that simply calls the built-in prediction function, but additionally returns the standard errors of the fitted values. The last block of code finds the roots to the equation $\mathcal{Q}^2 - z^2_{1-\alpha/2} = 0$ where $\mathcal{Q}$ is the approximate pivot described in Section 5.4. The resulting interval, $(58.24, 137.05)$, is only slightly larger than the interval based on the delta method. Notice it is also not symmetric about the point estimate $\widehat{\texttt{volume}}_0 = 93.41$ ml which, given the nonlinear trend and increasing variation in the data, is more realistic.

R Example 5.8.2.

```
## Extract point estimate from previous example
x0.est <- dm[["Estimate"]]
## Function to compute fitted values and their standard errors
predFun <- function(x) {
    z <- list(volume = x)
    fit <- predict(mod, newdata = z, level = 0)
    se.fit <- sqrt(diag(cbind(1, unlist(z), unlist(z)^2) %*%
        mod$varFix %*% t(cbind(1, unlist(z), unlist(z)^2))))
    list(fit = fit, se.fit = se.fit)
}
## Invert approximate prediction bounds (numerically)
invBounds <- function(x) {
    z <- list(volume = x)
    (70 - predFun(x)$fit)^2/(var.y0 + predFun(x)$se.fit^2) -
        qnorm(0.975)^2
}
c(uniroot(invBounds, interval = c(10, x0.est), tol = 1e-10)$root,
    uniroot(invBounds, interval = c(x0.est, 175), tol = 1e-10)$root)

## [1]  58.24 137.05
```

In addition, we can also obtain the bootstrap adjusted inversion interval (Equation (5.11)) described at the end of Section 5.5. Figure 5.8 shows a histogram of the $R = 9,999$ bootstrap replicates of W. As can be seen, the estimated sampling distribution of W is reasonably normal. The necessary quantiles to compute $\widehat{\mathcal{J}}^{\star}_{\text{cal}}(x)$ are $q^{\star}_{0.025} = -2$ and $q^{\star}_{0.975} = 1.98$ (compare these to the corresponding quantiles from a standard normal distribution; i.e., ± 1.96)). Making the necessary adjustments to the code in Example 5.8.2 yields $\widehat{\mathcal{J}}^{\star}_{\text{cal}}(x) = (57.94, 138.16)$, which is only slightly wider than the unadjusted inversion interval based on the normal approximation. In short, the normal approximation works quite well for these data.

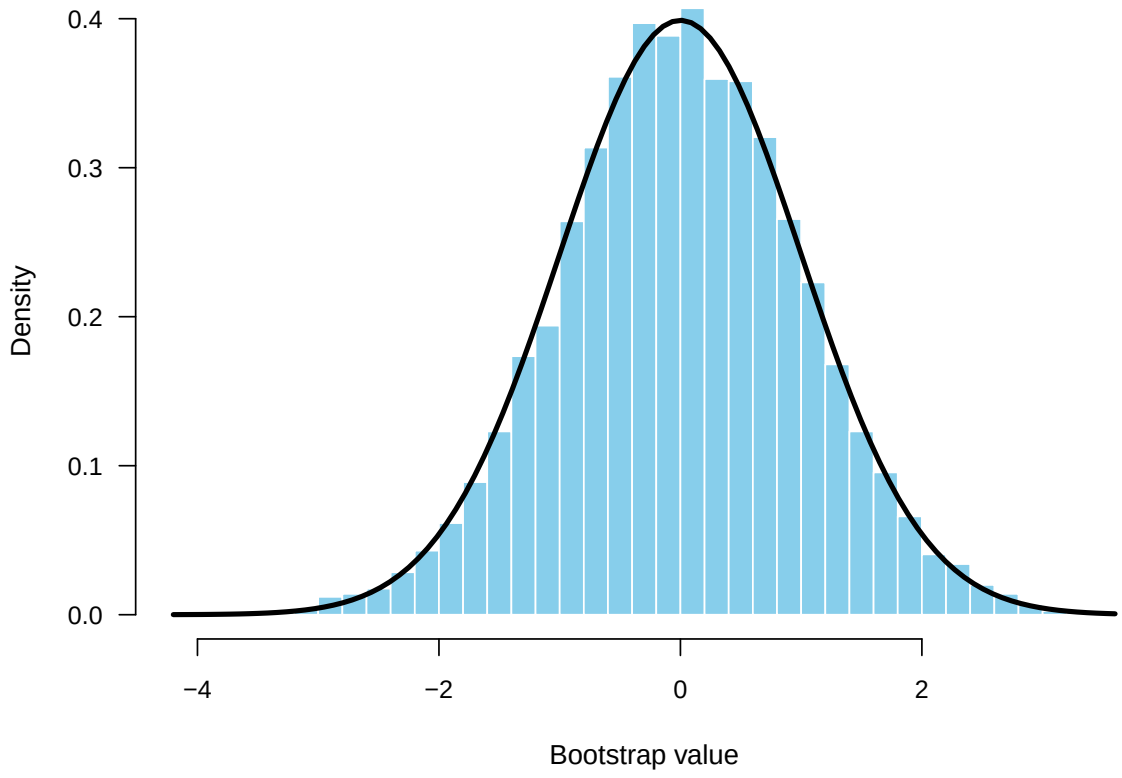


Figure 5.8: Bootstrap distribution of W for the bladder volume example. A standard normal distribution (solid black curve) is shown for comparison.

Finally, we compute the bootstrap distribution of $\widehat{\texttt{volume}}_0$ directly using Algorithm 2. Our results are based on a bootstrap simulation of size $R = 9,999$. Since we do not expect the sampling distribution of $\widehat{\texttt{volume}}_0$ to be symmetric, we provide only a 95% percentile bootstrap interval, that is, the interval obtained by taking the lower 2.5 and upper 97.5 percentiles of the bootstrap distribution. The resulting interval, $(58.37, 136.02)$, is quite close to the inversion interval previously obtained. A histogram of the $9,999$ bootstrap replicates of $\widehat{\texttt{volume}}_0$ is given in Figure 5.9. As indicated by the normal Q-Q plot in Figure 5.9, the bootstrap distribution is positively skewed (the estimated skewness is 0.3011), providing further support for the inversion and percentile intervals over the symmetric Wald-based interval obtained earlier using the delta method.

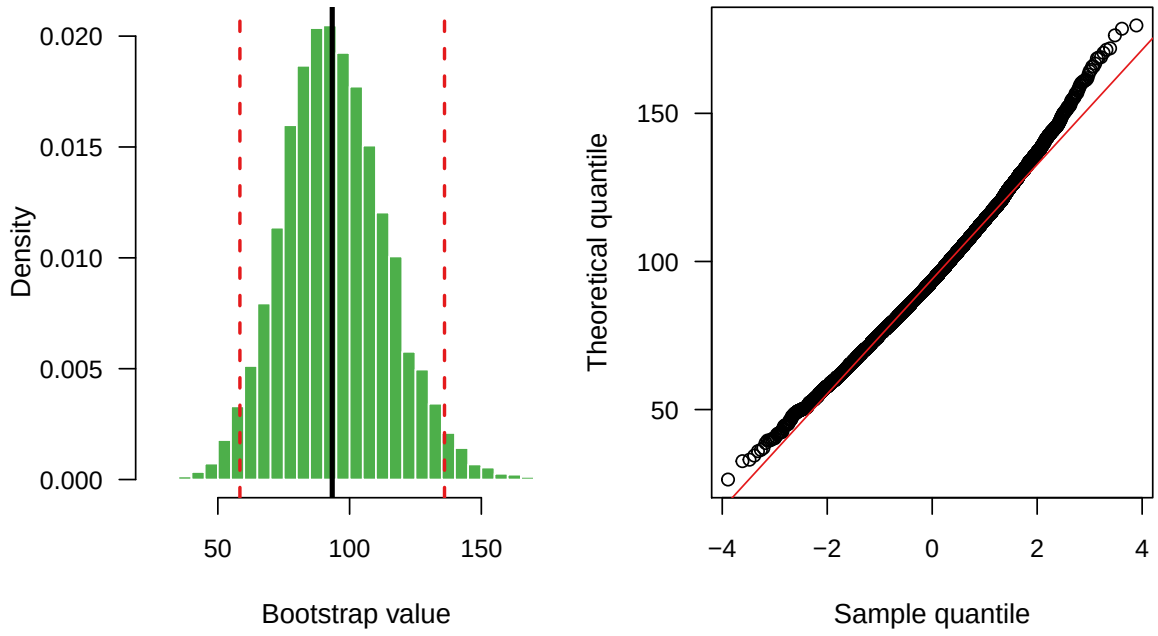


Figure 5.9: Bootstrap distribution of $\widehat{\texttt{volume}}_0$ obtained using Algorithm 2. *Left*: Bootstrap distribution of $\widehat{\texttt{volume}}_0$ obtained using Algorithm 2. The solid black line indicates the original point estimate and the dotted red lines indicate the positions of the sample 2.5 and 97.5 quantiles of the bootstrap distribution. *Right*: Normal Q-Q plot of the bootstrap replicates of $\widehat{\texttt{volume}}_0$.

5.9 Discussion

We have described a number of techniques for controlled calibration in a (linear) mixed model setting. The Wald-based interval is the simplest, but relies on the asymptotic normality of $\widehat{x}_0$ along with a Taylor-series approximation of its variance. Perhaps the biggest drawback to using a Wald-based calibration interval is that it is always symmetric about $\widehat{x}_0$. While this is appealing to many researchers, it is not very realistic in standard situations where, say, the data exhibit nonlinear behavior (possibly due to horizontal asymptotes) and nonconstant variance. The asymptotic normality of $\widehat{x}_0$ may also be questioned when $\widehat{x}_0$ is not the ML estimate (as in the bladder volume example). This is akin to using the Wald-based method for nonlinear calibration problems (the software JMP does this). Nonetheless, the estimated sampling distribution of $\widehat{x}_0$ displayed in Figures 5.3 and 5.9 using the parametric bootstrap are both reasonably normal. Thus, the Wald-based approach may still produce reliable inference when the distribution of $\widehat{x}_0$ can be considered symmetric. Although more difficult to obtain, the inversion and parametric bootstrap intervals are presumably superior to the Wald-based interval.

All the methods we have proposed rely on certain assumptions, for example, normality and large sample size. These assumptions may or may not hold in practice. If the random effects and errors are not normally distributed, then it is possible to fit a non-Gaussian mixed model (Jiang, 2007, p. 8). In this situation, we can still calculate a calibration interval by inverting a corresponding distribution-free prediction interval with asymptotic coverage probability $1 - \alpha$. As we have shown in Section 5.6, this is rather simple to do for standard LMMs. If, on the other hand, the data are not normal and the sample size is rather small, then there is not much we can do with the methods discussed in this chapter.

Future work on data like these might combine the nonparametric method of calibration proposed in Chapter 4 with the application to grouped data discussed in this chapter. In particular, we would recognize the grouped structure of the data but allow each specimen to have its own (nonlinear) trend by introducing subject-specific spline terms which may or may not have corresponding random effects. This may seem far-fetched at first, but remember that the specific approach we took for nonparametric calibration is already based on an LMM (4.2). This is an obvious, and likely promising, area of future research, and we discuss it, along with some other ideas, in the conclusion to this dissertation.

VI. Conclusions and Suggestions for Further Research

Statistical calibration is an important application of regression in many areas of science, for example: bioassays, chemometrics, and calibrating laboratory equipment. For many of these applications, the data are inherently nonlinear with no known parametric form, or sometimes the data are collected in such a way that the observations can not be considered as independent. It is necessary, then, to have simple and general methods available for calibration in these situations. This has been the main goal of our research.

6.1 Conclusions

We discussed (controlled) semiparametric calibration in Chapter 4. We provided a frequentist approach to obtaining calibration intervals that involved inverting bias-corrected prediction intervals based on the simple LMM-based smoother described in Ruppert and Wand (2003). The coverage probability and length of these intervals were investigated using a small Monte Carlo experiment. This experiment showed that these intervals do in fact obtain coverage probability close to the nominal $1-\alpha$ level without sacrificing length. The experiment also highlighted that correcting for bias is more serious for calibration with respect to a mean response (i.e., regulation). A simple Bayesian analog was also proposed that has the benefit of providing the entire posterior distribution of x_0. We illustrated these methods using real data analysis examples.

In Chapter 5, we extended the usual methods of (controlled) calibration (i.e., point estimation and obtaining Wald-based/inversion intervals) for grouped data using the LMM. We also proposed a parametric bootstrap algorithm for controlled calibration in the LMM. This algorithm can be used to obtain confidence intervals for the unknown x_0 directly from the estimated sampling distribution of the point

estimator $\widehat{x}_0$, or by improving the inversion interval by removing the normality constraint on the approximate predictive pivot $\mathcal{W}$. These strategies were illustrated using real data analysis examples. We also briefly described how to use a distribution-free prediction interval to obtain a distribution-free inversion interval for the unknown x_0 in closed form for the random intercept model.

Calibration will always remain an important topic in statistics. We list here some possible topics for future research based on extending the ideas in Chapters 4 and 5:

- Semiparametric calibration with constraints;
- Prior selection for x_0 in semiparametric calibration;
- Bootstrap for (controlled) semiparametric calibration;
- Calibration in NLMMs;
- Semiparametric calibration with random coefficients.

For the most part, these topics were discussed in the conclusions to Chapters 4 and 5.

Appendix A: Proofs

In this appendix, we provide the "extra steps" for deriving some of the mathematical results presented in this dissertation.

A.1 Conditional posterior of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$

Note that the kernel of a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$ is

$$K\left(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) \propto \exp\left\{-\frac{1}{2}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)' \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right)\right\}.$$

Furthermore, let $\boldsymbol{x}$ and $\boldsymbol{\theta}$ be $n \times 1$ vectors, and $\boldsymbol{A}$ be an invertible $n \times n$ symmetric matrix. We can complete the square for the quadratic form $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - \boldsymbol{\theta}'\boldsymbol{x}$ by writing

$$\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} - \boldsymbol{\theta}'\boldsymbol{x} = \left(\boldsymbol{x}-\boldsymbol{\mu}\right)' \boldsymbol{A}^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}\right) + C,$$

where

$$\boldsymbol{\mu} = \frac{1}{2}\boldsymbol{A}^{-1}\boldsymbol{\theta} \quad \text{and} \quad C = -\frac{1}{4}\boldsymbol{\theta}'\boldsymbol{A}^{-1}\boldsymbol{\theta}.$$

Let the vectors $\boldsymbol{x}_0$ and $\boldsymbol{z}_0$ have the same form as the i-th rows of $\boldsymbol{X}$ and $\boldsymbol{Z}$ in Equation (4.2), respectively, but with x_i replaced with x_0. Ignoring the constant of proportionality, the conditional posterior of $(\boldsymbol{\beta}, \boldsymbol{\alpha})$ is

$$\begin{aligned}
\pi\left(\boldsymbol{\beta}, \boldsymbol{\alpha} | \boldsymbol{y}, y_0, \sigma_\epsilon^2, \sigma_\alpha^2, x0\right) &\propto \pi\left(\boldsymbol{y} | \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2, x_0\right) \pi\left(y_0 | \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, \sigma_\alpha^2, x_0\right) \pi\left(\boldsymbol{\beta}\right) \pi\left(\boldsymbol{\alpha} | \sigma_\alpha^2\right) \\
&\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}||^2 - \frac{1}{2\sigma_\alpha^2}||\boldsymbol{\alpha}||^2 - \frac{1}{2\sigma_\epsilon^2}\left[y_0 - \mu(x_0)\right]^2\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}||^2 + \left[y_0 - \mu(x_0)\right]^2 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}||\boldsymbol{\alpha}||^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}||^2 + \left(y_0 - \boldsymbol{x}_0'\boldsymbol{\beta} - \boldsymbol{z}_0'\boldsymbol{\alpha}\right)^2 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}||\boldsymbol{\alpha}||^2\right)\right\} \\
&= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(||\boldsymbol{y}_0 - \boldsymbol{X}_0\boldsymbol{\beta} - \boldsymbol{Z}_0\boldsymbol{\alpha}||^2 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}||\boldsymbol{\alpha}||^2\right)\right\},
\end{aligned}$$

where

$$\boldsymbol{y}_0 = \begin{bmatrix} \boldsymbol{y} \\ y_0 \end{bmatrix}, \quad \boldsymbol{X}_0 = \begin{bmatrix} \boldsymbol{X} \\ \boldsymbol{x}_0' \end{bmatrix}, \quad \boldsymbol{Z}_0 = \begin{bmatrix} \boldsymbol{Z} \\ \boldsymbol{z}_0' \end{bmatrix}$$

are augmented data vectors and matrices. To show that the conditional posterior of $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\alpha}')'$ is normal, note that

$$\begin{aligned} &\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(||\boldsymbol{y}_0 - \boldsymbol{X}_0\boldsymbol{\beta} - \boldsymbol{Z}_0\boldsymbol{\alpha}||^2 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}||\boldsymbol{\alpha}||^2\right)\right\} \\ &= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y}_0^* - \boldsymbol{X}_0^*\boldsymbol{\beta} - \boldsymbol{Z}_0^*\boldsymbol{\alpha}||^2\right\} \\ &= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y}_0^* - \boldsymbol{\Omega}_0^*\boldsymbol{\theta}||^2\right\}, \end{aligned}$$

where, similar to before,

$$\boldsymbol{y}_0^* = \begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{X}_0^* = \begin{bmatrix} \boldsymbol{X}_0 \\ \boldsymbol{0} \end{bmatrix}, \quad \boldsymbol{Z}_0^* = \begin{bmatrix} \boldsymbol{Z}_0 \\ (\sigma_\epsilon^2/\sigma_\alpha^2)\,\boldsymbol{I} \end{bmatrix}$$

are augmented data vectors and matrices and $\boldsymbol{\Omega}_0^* = (\boldsymbol{X}_0^*; \boldsymbol{Z}_0^*)$. Now, using basic matrix multiplication,

$$\begin{aligned} \exp\left\{-\frac{1}{2\sigma_\epsilon^2}||\boldsymbol{y}_0^{*\prime} - \boldsymbol{\Omega}_0^*\boldsymbol{\theta}||^2\right\} &= \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{y}_0^{*\prime}\boldsymbol{y}_0^* - 2\boldsymbol{y}_0^{*\prime}\boldsymbol{\Omega}_0^*\boldsymbol{\theta} + \boldsymbol{\theta}'\boldsymbol{\Omega}_0^{*\prime}\boldsymbol{\Omega}_0^*\boldsymbol{\theta}\right)\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{\theta}'\boldsymbol{\Omega}_0^{*\prime}\boldsymbol{\Omega}_0^*\boldsymbol{\theta} - 2\boldsymbol{y}_0^{*\prime}\boldsymbol{\Omega}_0^*\boldsymbol{\theta}\right)\right\}. \end{aligned}$$

Upon completing the square, we get

$$\exp\left\{-\frac{1}{2\sigma_\epsilon^2}\left(\boldsymbol{\theta}'\boldsymbol{\Omega}_0^{*\prime}\boldsymbol{\Omega}_0^*\boldsymbol{\theta} - 2\boldsymbol{y}_0^{*\prime}\boldsymbol{\Omega}_0^*\boldsymbol{\theta}\right)\right\} = \exp\left\{-\frac{1}{2}\left(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}\right)'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\left(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}\right)\right\},$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\theta}} = \left(\boldsymbol{\Omega}_0^{*\prime}\boldsymbol{\Omega}_0^*\right)^{-1}\boldsymbol{\Omega}_0^{*\prime}\boldsymbol{y}_0^* = \left(\boldsymbol{\Omega}_0{}'\boldsymbol{\Omega}_0 + \frac{\sigma_\epsilon^2}{\sigma_\alpha^2}\boldsymbol{D}\right)^{-1}\boldsymbol{\Omega}_0{}'\boldsymbol{y}_0, \quad \boldsymbol{D} = \mathbf{diag}\left\{\boldsymbol{0}_{p\times p}, \boldsymbol{I}_{q\times q}\right\},$$

and

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = \sigma_\epsilon^2\left(\boldsymbol{\Omega}_0^{*\prime}\boldsymbol{\Omega}_0^*\right)^{-1} = \sigma_\epsilon^2\left(\boldsymbol{\Omega}_0{}'\boldsymbol{\Omega}_0\right)^{-1}.$$

Thus, the conditional posterior of the coefficients $\boldsymbol{\theta}$ is multivariate normal with mean vector $\boldsymbol{\mu}_{\boldsymbol{\theta}}$ and variance-covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$.

A.2 Conditional posteriors of σ_ϵ^2 and σ_α^2

Note that since $\sigma_\epsilon^2 \sim \mathcal{IG}(a, b)$, then

$$\pi(\sigma_\epsilon^2) = \frac{b^a}{\Gamma(a)} \left(\sigma_\epsilon^2\right)^{-(a+1)} \exp\left\{-\frac{b}{\sigma_\epsilon^2}\right\}.$$

Now, ignoring the constant of proportionality, the conditional posterior of σ_ϵ^2 is

$$\begin{aligned}
\pi(\sigma_\epsilon^2|\boldsymbol{y}, y_0, \boldsymbol{\beta},\boldsymbol{\alpha}, \sigma_\alpha^2, x_0) &\propto \pi(\boldsymbol{y}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2)\pi(y_0|\boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, x_0)\pi(\sigma_\epsilon^2) \\
&\propto \left(\sigma_\epsilon^2\right)^{-(a+1)} \exp\left\{-\frac{||\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta} - \boldsymbol{Z}\boldsymbol{\alpha}||^2}{2\sigma_\epsilon^2} - \frac{(y_0 - \boldsymbol{x}_0\boldsymbol{\beta} - \boldsymbol{z}_0\boldsymbol{\alpha})^2}{2\sigma_\epsilon^2} - \frac{b}{\sigma_\epsilon^2}\right\} \\
&= \left(\sigma_\epsilon^2\right)^{-(a+1)} \exp\left\{\frac{\left(\frac{1}{2}||\boldsymbol{y}_0 - \boldsymbol{X}_0\boldsymbol{\beta} - \boldsymbol{Z}_0\boldsymbol{\alpha}||^2 + b\right)}{\sigma_\epsilon^2}\right\},
\end{aligned}$$

which is proportional to the density function of a $\mathcal{IG}\left(a, \frac{1}{2}||\boldsymbol{y}_0 - \boldsymbol{X}_0\boldsymbol{\beta} - \boldsymbol{Z}_0\boldsymbol{\alpha}||^2 + b\right)$ distribution. The proof for the conditional posterior $\pi(\sigma_\alpha^2|\boldsymbol{y}, y_0, \boldsymbol{\beta}, \boldsymbol{\alpha}, \sigma_\epsilon^2, x_0)$ follows in an analogous manner.

A.3 LMM log-likelihood

For the LMM (Equation (5.1)), we have that

$$\boldsymbol{\mathcal{Y}} \sim \mathcal{N}(\boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{V}),$$

hence, the density function is

$$f(\boldsymbol{y}) = (2\pi)^{-N/2} |\boldsymbol{V}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{V}^{-1} (\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta})\right\}.$$

Taking the logarithm of $f(\boldsymbol{y})$ gives the log-likelihood

$$\ell = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log(|\boldsymbol{V}|) - \frac{1}{2}(\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta})' \boldsymbol{V}^{-1} (\boldsymbol{\mathcal{Y}} - \boldsymbol{X}\boldsymbol{\beta}).$$

Let $\boldsymbol{V}$ have the block diagonal form $\boldsymbol{V} = \sigma_\epsilon^2 \left\{ \underset{\text{diag}}{\boldsymbol{I}_{n_i}} + \boldsymbol{Z}_i \boldsymbol{G}^* \boldsymbol{Z}_i' \right\}_{i=1}^{m}$. Since the determinant of a block diagonal matrix is just the product of the determinant of

the block diagonals, then

$$\begin{aligned}
\log\left(|\boldsymbol{V}|\right) &= \log\left\{\prod_{i=1}^{m}\left|\sigma_\epsilon^2\left(\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'\right)\right|\right\} \\
&= \log\left\{\prod_{i=1}^{m}\left(\sigma_\epsilon^2\right)^{n_i}\left|\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'\right|\right\} \\
&= \sum_{i=1}^{m} n_i \log\left(\sigma_\epsilon^2\right) + \sum_{i=1}^{m}\log\left(|\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'|\right) \\
&= N\log\left(\sigma_\epsilon^2\right) + \sum_{i=1}^{m}\log\left(|\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'|\right).
\end{aligned}$$

Similarly, since the inverse of a block diagonal matrix is another block diagonal matrix, composed of the inverse of each block, it follows that

$$\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{V}^{-1}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{X}\boldsymbol{\beta}\right) = \sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i-\boldsymbol{X}_i\boldsymbol{\beta}\right)'\boldsymbol{V}_i^{-1}\left(\boldsymbol{\mathcal{Y}}_i-\boldsymbol{X}_i\boldsymbol{\beta}\right),$$

where $\boldsymbol{V}_i = \sigma_\epsilon^2\left(\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'\right)$. Therefore, the log-likelihood becomes

$$\begin{aligned}
\ell &= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left(|\boldsymbol{V}|\right) - \frac{1}{2}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{X}\boldsymbol{\beta}\right)'\boldsymbol{V}^{-1}\left(\boldsymbol{\mathcal{Y}}-\boldsymbol{X}\boldsymbol{\beta}\right) \\
&= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma_\epsilon^2) - \frac{1}{2}\sum_{i=1}^{m}\log\left(|\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'|\right) \\
&\qquad - \frac{1}{2\sigma_\epsilon^2}\sum_{i=1}^{m}\left(\boldsymbol{\mathcal{Y}}_i-\boldsymbol{X}_i\boldsymbol{\beta}\right)'\left(\boldsymbol{I}_{n_i}+\boldsymbol{Z}_i\boldsymbol{G}^*\boldsymbol{Z}_i'\right)^{-1}\left(\boldsymbol{\mathcal{Y}}_i-\boldsymbol{X}_i\boldsymbol{\beta}\right).
\end{aligned}$$

Appendix B: The R Package investr

In this appendix, we describe the investr package for the R statistical software. The name investr stands for inverse estimation in R. It is currently listed on CRAN at http://cran.r-project.org/web/packages/investr/index.html. The source code for additional functions (e.g., `pspline`) can be obtained from the GitHub development site at https://github.com/w108bmg/Research/tree/master/Rcode. To install the package, simply open an R terminal and type:

```
install.packages("investr", dependencies = TRUE)  # install package
library(investr)  # load package
```

Once the package is loaded, we have access to all of its functions, datasets, and examples. The three main functions are described in Table B.

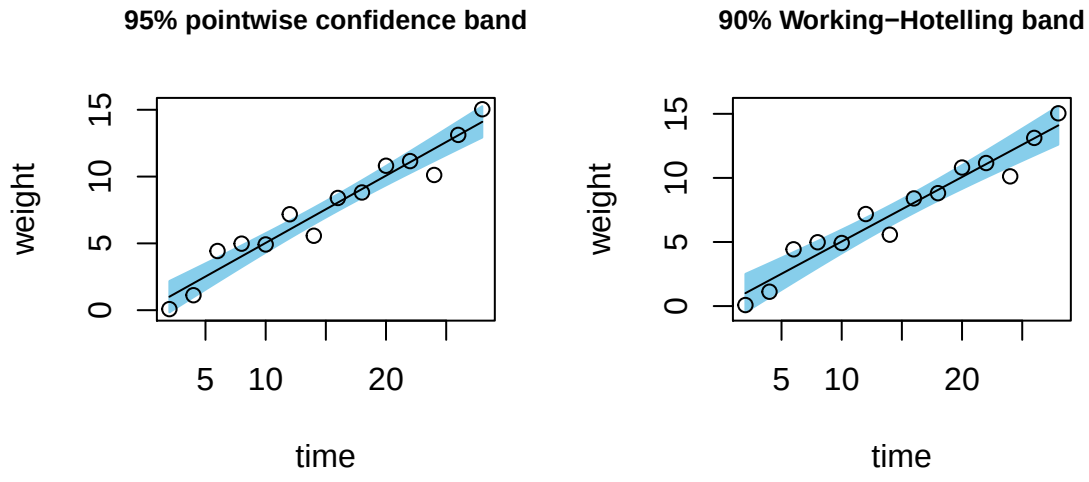
Function	Description
`calibrate`	For a vector of m response values with unknown predictor value x_0, computes the classical estimate (i.e., ML estimate) $\widehat{x}_0$ and a corresponding Wald or inversion interval for the simple linear calibration problem.
`invest`	For a vector of m response values with unknown predictor value x_0, computes the classical estimate $\widehat{x}_0$ and a corresponding Wald or inversion interval for polynomial and nonlinear calibration problems.
`plotFit`	For plotting fitted regression models with or without confidence/prediction bands.

Table B.1: Main functions from the investr package.

B.1 The `plotFit` function

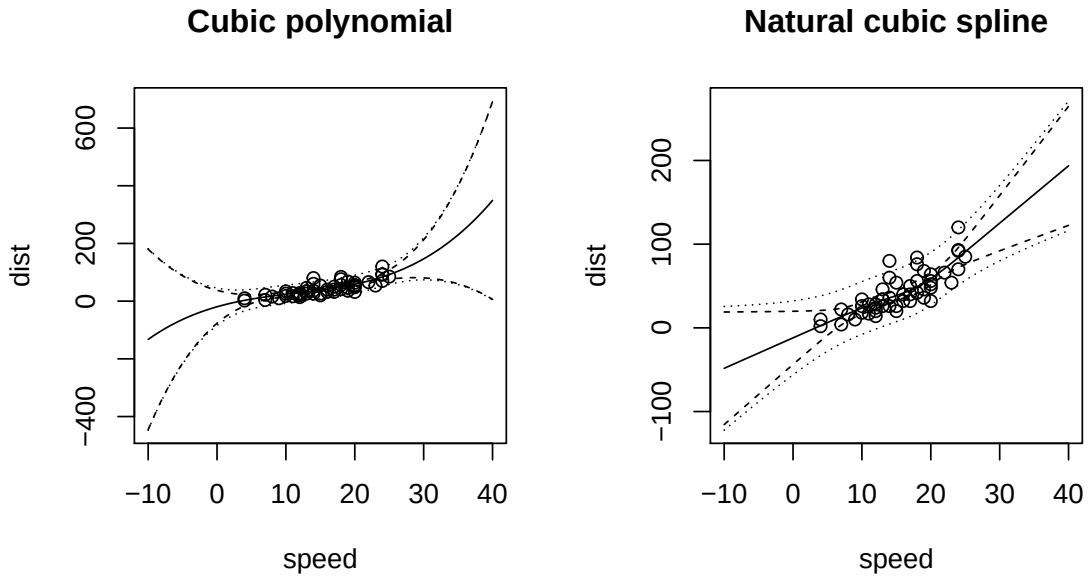
The `plotFit` function is a general purpose function that is also useful outside of statistical calibration problems. Its sole purpose is to plot fitted models for `R` objects of class `lm` or class `nls` with the option of adding a confidence and/or prediction band. For example, the following snippet of `R` code fits a simple linear regression model to the `crystal` data frame and plots the data along with the fitted regression line and (pointwise) 95% confidence band. Of course, we can change the default 95% confidence level by specifying, for example, `level=0.9`. Additionally, we can also specify an adjustment for simultaneous inference such as *Scheffé*, *Bonferroni*, or *Working-Hotelling*. The second call to `plotFit` in the code below illustrates the use of both of these options.

```
par(mfrow = c(1, 2), cex.main = 0.8)  # side-by-side plots
crystal.lm <- lm(weight ~ time, data = crystal)  # fit model
plotFit(crystal.lm, interval = "confidence", shade = T,
    col.conf = "skyblue", main = "95% pointwise confidence band")
plotFit(crystal.lm, interval = "confidence", shade = T,
    col.conf = "skyblue", level = 0.9, adjust = "W-H",
    main = "90% Working-Hotelling band")
```



More elaborate models can also be plotted in the same way. For example, the following snippet of code fits a simple linear, quadratic, cubic, and natural cubic spline model to the well-known `cars` data frame and then plots the corresponding fits with both confidence and prediction bands at the 95% level.

```
data(cars, package = "datasets")  # load cars data frame
library(splines)  # load splines package
## Fit models
cars.lm1 <- lm(dist ~ poly(speed, degree = 3), data = cars)
cars.lm2 <- lm(dist ~ ns(speed, df = 3), data = cars)
## Plot models
par(mfrow = c(1, 2))  # 2-by-2 grid of plots
plotFit(cars.lm1, interval = "both", xlim = c(-10, 40),
    ylim = c(-50, 150), main = "Cubic polynomial")
plotFit(cars.lm2, interval = "both", xlim = c(-10, 40),
    ylim = c(-50, 150), main = "Natural cubic spline")
```



B.2 The `calibrate` function

The most basic calibration problem, the one often encountered in more advanced regression texts, is the simple linear calibration problem for which

$$\mathcal{Y}_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \epsilon_i \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right), \quad i = 1, \ldots, n,$$
$$\mathcal{Y}_{0j} = \beta_0 + \beta_1 x_0 + \epsilon_{0j}, \quad \epsilon_{0j} \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_\epsilon^2\right), \quad j = 1, \ldots, m.$$

For example, consider the arsenic data introduced in Section 3.2.4. The following snippet of code obtains a 95% inversion interval and 95% Wald-based interval for the unknown x_0 corresponding to $y_0 = 3$ based on Equations (3.12) and (3.18), respectively:

```
calibrate(arsenic.lm, y0 = 3, interval = "inversion")

## estimate    lower    upper
##    2.931    2.537    3.325

calibrate(arsenic.lm, y0 = 3, interval = "Wald")

## estimate    lower    upper       se
##   2.9314   2.5374   3.3255   0.1929
```

If instead we were interested in the unknown x_0 corresponding to a fixed mean response of $\mu_0 = 3$ (i.e., a regulation problem) we would instead use

```
calibrate(arsenic.lm, y0 = 3, interval = "inversion",
    mean.response = TRUE)

## estimate    lower    upper
##    2.931    2.860    3.002
```

B.3 The `invest` function

In this section, we describe the more general function, `invest`, which can be used for more complex univariate calibration problems such as polynomial and nonlinear calibration.

For the quadratic linear model in the whiskey age example of Section4.2.1, we used the following code to obtain a 95% inversion interval for the unknown age corresponding to sample with a known proof of 108:

```
whiskey <- data.frame(age = c(0, 0.5, 1, 2, 3, 4, 5,
    6, 7, 8), proof = c(104.6, 104.1, 104.4, 105, 106,
    106.8, 107.7, 108.7, 110.6, 112.1))
whiskey.lm <- lm(proof ~ age + I(age^2), data = whiskey)
invest(whiskey.lm, y0 = 108)

## estimate    lower    upper
##    5.233    4.678    5.735
```

As for a nonlinear regression example, we consider the nasturtium example of Section 3.4.1. The following snippet of code fits the log-logistic regression function

$$\mu(x;\beta_1,\beta_2,\beta_3) = \begin{cases} \beta_1, & x = 0 \\ \beta_1 / \left[1 + \exp\left\{\beta_2 + \beta_3 \ln(x)\right\}\right], & x > 0 \end{cases}.$$

to the data and obtains both a 95% inversion interval and 95% Wald-based interval for the unknown concentration corresponding to the three unknowns 309, 296, and 419:

```
nas.nls <- nls(weight ~ ifelse(conc == 0, theta1, theta1/(1 +
    exp(theta2 + theta3 * log(conc)))), data = nasturtium,
    start = list(theta1 = 1000, theta2 = 0, theta3 = 1))
invest(nas.nls, y0 = c(309, 296, 419), interval = "inversion")

## estimate    lower    upper
##    2.264    1.772    2.969

invest(nas.nls, y0 = c(309, 296, 419), interval = "Wald")

## estimate    lower    upper       se
##   2.2639   1.6889   2.8388   0.2847
```

Bootstrap approaches to obtaining calibration intervals are currently not available in the investr package, however, a future release is likely to contain some

bootstrap functionality. Until such time, the well-known `boot` package can be used to obtain the bootstrap calibration intervals described in Chapter 3. The `bootMer` function from the `R` package `lme4` ($\geq 1.0 - 5$) can be used to obtain the parametric bootstrap calibration intervals discussed in Chapter 5.

Bibliography

J. Aitchison and L. R. Dunsmore. *Statistical Prediction Analysis*. Cambridge University Press, Cambridge England New York, 1980.

M. A. Ali and M. Ashkar. The calibration problem revisited. *Communications in Statistics - Theory and Methods*, 31(10):1733–1741, 2002.

M. A. Ali and N. Singh. An alternative estimator in inverse regression. *J. Statist. Comput. Simul.*, 14:1–15, 1981.

J. Berkson. Estimation of a linear function for a calibration line; consideration of a recent proposal. *Technometrics*, 11(4):649–660, Nov. 1969. ISSN 0040-1706, 1537-2723.

P. Brown. *Measurement, Regression, and Calibration*. Clarendon Press Oxford University Press, Oxford England New York, 1993.

P. J. Brown. Multivariate calibration. *Journal of the Royal Statistical Society - Series B*, 44(3):287–321, 1982.

P. J. Brown and R. Sundberg. Confidence and conflict in multivariate calibration. *Journal of the Royal Statistical Society - Series B*, 49(1):46–57, 1987.

B. A. Brumback, D. Ruppert, and M. P. Wand. Comment on shively, kohn, and wood. *Journal of the American Statistical Association*, 94:794–797, 1999.

J. P. Buonaccorsi. Design considerations for calibration. *Technometrics*, 28(2):149–155, May 1986. ISSN 0040-1706, 1537-2723.

R. J. Carroll and C. H. Spiegelman. The effect of ignoring small measurement errors in precision instrument calibration. *J. Quality. Tech.*, 18(3):170–173, 1986.

G. Casella and R. L. Berger. *Statistical Inference*. Thomson Learning, Australia Pacific Grove, CA, second edition, 2002. ISBN 0534243126.

R. M. Clark. Calibration, cross-validation and carbon-14 i. *Journal of the Royal Statistical Society - Series A*, 142(1):47–62, 1979.

R. M. Clark. Calibration, cross-validation and carbon-14 ii. *Journal of the Royal Statistical Society - Series A*, 143(2):177–194, 1980.

C. M. Crainiceanu, D. Ruppert, and M. P. Wand. Bayesian analysis for penalized spline regression using winbugs. *Journal of Statistical Software*, 14(14):1–24, 2005.

R. C. Dahiya and J. J. McKeon. Modified classical and inverse regression estimators in calibration. *Sankhyā: The Indian Journal of Statistics - Series B*, 53(1):48–55, 1991.

M. Davidian and D. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1995.

A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge University Press, Cambridge New York, NY, USA, 1997.

E. Demidenko. *Mixed Models: Theory and Applications with R*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 2013.

R. Dorfman. A note on the *delta*-method for finding variance formulae. *The Biometric Bulletin*, 1:129–137, 1938.

N. R. Draper and H. Smith. *Applied Regression Analysis*. Wiley, New York, 1998. ISBN 0471170828.

J. L. Du Plessis and A. J. Van Der Merwe. Bayesian calibration in the estimation of the age of rhinoceros. *Ann. Inst. Statist. Math.*, 48(1):17–28, 1996.

L. R. Dunsmore. A bayesian approach to calibration. *Journal of the Royal Statistical Society*, 30(2):396–405, Oct. 1967.

B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7 (1):1–26, 1979.

B. Efron. The bootstrap and markov-chain monte carlo. *Journal of Biopharmaceutical Statistics*, 21(6):1052–1062, 2011.

B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman & Hall, New York, 1994.

C. Eisenhart. The interpretation of certain regression methods and their use in biological and industrial research. *Annals of Mathematical Statistics*, 10:162–186, 1939.

E. C. Fieller. The distribution of the index in a normal bivariate population. *Biometrika*, 24(3):428–440, 1932.

E. C. Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 16:175–185, 1954.

J. Fox and S. Weisberg. *An R Companion to Applied Regression*. Sage, Thousand Oaks CA, second edition, 2011. URL http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534, 2006.

A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis, Second Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2003.

W. Gould. Confidence intervals in logistic and probit models. *Stata Technical Bulletin*, 14:26–28, 1993.

F. A. Graybill. *Theory and Application of the Linear Model*. Duxbury Classic Series. Duxbury, Pacific Grove, California, 1976.

F. A. Graybill and H. K. Iyer. *Regression Analysis : Concepts and Applications*. Duxbury Press, Belmont, Calif, 1994.

M.-A. Gruet and E. Jolivet. Calibration with a nonlinear calibration curve: How to do it? *Computational Statistics*, 9:249–276, 1993.

P. Hall. *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York, 1992. ISBN 0387977201.

M. Halperin. On inverse estimation in linear regression. *Technometrics*, 12(4):727–736, Nov. 1970. ISSN 0040-1706, 1537-2723.

M. Hamada, A. Pohl, C. Spiegelman, and J. Wendelberger. A bayesian approach to calibration intervals and properly calibrated tolerance intervals. *Journal of Quality Technology*, 35(2):194–205, 2003.

F. E. Harrell. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. Graduate Texts in Mathematics. Springer, New York, NY, 2001.

T. Hastie and R. Tibshirani. *Generalized Additive Models*. Chapman & Hall, London, 1990.

C. R. Henderson. Sire evaluation and genetic trends. In *Proceedings of the Animal Breeding and Genetics Symposium in Honor of Dr. Jay L. Lush*, pages 10–41, 1973.

B. M. Hill. Discussion. *Technometrics*, 23(4):335–338, Nov. 1981.

D. V. Hinkley. On the ratio of two correlated normal random variables. *Biometrika*, 56(3):635–639, 1969.

B. Hoadley. A bayesian look at inverse linear regression. *Journal of the American Statistical Association*, 65(329):356–369, 1970.

S. Huet, A. Bouvier, M.-A. Poursat, and E. Jolivet. *Statistical Tools for Nonlinear Regression: A Practical Guide with S-PLUS and R Examples*. Springer, New York, 2004. ISBN 0387400818.

W. G. Hunter and W. Lamboy. Bayesian analysis of ratios. Technical Report 587, University of Wisconsin, Statistics Department, 1979a.

W. G. Hunter and W. Lamboy. Making inferences about ratios: Some examples. Technical Report 588, University of Wisconsin, Statistics Department, 1979b.

W. G. Hunter and W. Lamboy. A bayesian analysis of the linear calibration problem. *Technometrics*, 23(4):323–328, Nov. 1981. ISSN 0040-1706, 1537-2723.

J. Jiang. *Linear and Generalized Linear Mixed Models and Their Applications*. Springer Series in Statistics. Springer, New York, NY, 2007.

J. Jiang and W. Zhang. Distribution-free prediction intervals in mixed linear models. *Statistica Sinica*, 12:537553, 2002.

G. Jones and P. Lyons. Approximate graphical methods for inverse regression. *Journal of Data Science*, 7:61–72, 2009.

G. Jones and D. M. Rocke. Bootstrapping in controlled calibration experiments. *Technometrics*, 41(3):224–233, Aug. 1999.

A. J. Kalotay. Structural solution to the linear calibration problem. *Technometrics*, 13(4):761–769, 1971.

N. Kannan, J. P. Keating, and R. L. Mason. A comparison of classical and inverse estimators in the calibration problem. *Communications in Statistics - Theory and Methods*, 36(1):83–95, 2007.

T. Krivobokova and G. Kauermann. A note on penalized spline smoothing with correlated errors. *Journal of the American Statistical Association*, 102(480):1328–1337, 2007.

R. G. Krutchkoff. Classical and inverse regression methods of calibration. *Technometrics*, 9(3):425–439, Aug. 1967.

R. G. Krutchkoff. Classical and inverse regression methods of calibration in extrapolation. *Technometrics*, 11(3):605–608, Aug. 1969.

R. G. Krutchkoff. The calibration problem and closeness. *Journal of Statistical Computation and Simulation*, 1(1):87–95, Jan. 1972. ISSN 0094-9655, 1563-5163.

N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982.

J. F. Lawless. Discussion. *Technometrics*, 23(4):334–335, 1981.

D. J. Lunn, D. Spiegelhalter, A. Thomas, and N. Best. The bugs project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067, 2009.

J. F. Lwin. Discussion. *Technometrics*, 23(4):339–341, 1981.

T. Lwin and J. S. Maritz. An analysis of the linear-calibration controversy from the perspective of compound estimation. *Technometrics*, 24(3):235–242, Aug. 1982. ISSN 0040-1706, 1537-2723.

E. Mammen, J. S. Marron, B. A. Turlach, and M. P. Wand. A general framework for constrained smoothing. *Statistical Science*, 16(3):232–248, 2001.

C. E. McCulloch, S. R. Searle, and J. M. Neuhaus. *Generalized, Linear, and Mixed Models*. Wiley Series in Probability and Statistics. Wiley, Hoboken, NJ, 2008.

R. G. Miller. Tan unbalanced jackknife. *The Annals of Statistics*, 2(5), 1974.

C. E. Minder and J. B. Whitney. A likelihood analysis of the linear calibration problem. *Technometrics*, 17(4):463–471, Nov. 1975. ISSN 0040-1706, 1537-2723.

A. M. Mood, F. A. Graybill, and D. C. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, Auckland, U.A, 1974. ISBN 0070854653.

J. S. Morris. The blups are not "best" when it comes to bootstrapping. *Statistics & Probability Letters*, 56:425–430, 2002.

L. J. Naszdi. Elimination of the bias in the course of calibration. *Technometrics*, 20 (2):201–205, May 1978. ISSN 0040-1706, 1537-2723.

S. D. Oman. An exact formula for the mean squared error of the inverse estimator in the linear calibration problem. *Journal of Statistical Planning and Inference*, 11 (2):189–196, 1985.

S. D. Oman. Calibration with random slopes. *Biometrika*, 85(2):439–449, 1998.

J. E. Orban. Discussion. *Technometrics*, 23(4):342–343, 1981.

T. Pham-Gia, N. Turkkan, and E. Marchand. Density of the ratio of two normal random variables and applications. *Communications in Statistics - Theory and Methods*, 35:1569–1591, 2006.

J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, and R Core Team. *nlme: Linear and Nonlinear Mixed Effects Models*, 2013. R package version 3.1-110.

J. C. Pinheiro. *Topics in Mixed-Effects Models*. PhD thesis, University of Wisconsin, Madison, WI., 1994.

J. C. Pinheiro and D. Bates. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer, New York, NY, 2009.

E. J. G. Pitman Hobart. The "closest" estimates of statistical parameters. *Communications in Statistics - Theory and Methods*, 20(11):3423–3437, 1937.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL http://www.R-project.org/. ISBN 3-900051-07-0.

A. Racine-Poon. A bayesian approach to nonlinear calibration problems. *Journal of the American Statistical Association*, 83(403):650–656, Sept. 1988.

C. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer, 2004.

G. K. Robinson. That blup is a good thing: The estimation of random effects. *Statistical Science*, 6(1):15–51, 1991.

O. Rosen and A. Cohen. Constructing a bootstrap confidence interval for the unknown concentration in radioimmunoassay. *Statistics in Medicine*, 14(9):935–952, 1995a.

O. Rosen and A. Cohen. Constructing a bootstrap confidence interval for the unknown concentration in radioimmunoassay. *Statistics in Medicine*, 14(9):935–952, 1995b.

D. Ruppert. Selecting the number of knots for penalized splines. *Journal of Computational and Graphical Statistics*, 11(4):735–757, 2002.

D. Ruppert and M. P. Wand. *Semiparametric Regression*. Cambridge University Press, New York, NY, 2003.

SAS Institute Inc. *SAS/STAT Software, Version 9.3*. Cary, NC, 2011. URL http://www.sas.com/.

R. Schoeneman, R. Dyer, and E. Earl. Analytical profile of straight bourbon whiskies. *Journal of the Association of Official Analytical Chemists*, 54:1247–1261, 1971.

J. R. Schwenke and G. A. Millikem. On the calibration problem extended to nonlinear models. *Biometrics*, 47:563–574, 1991.

G. A. F. Seber and C. J. Wilde. *Nonlinear Regression*. John Wiley & Sons, Hoboken, NJ, 2003.

G. K. Shukla. On the problem of calibration. *Technometrics*, 14(3):547–553, Aug. 1972. ISSN 0040-1706, 1537-2723.

V. K. Srivastava and N. Singh. Small-disturbance asymptotic theory for linear calibration estimators. *Technometrics*, 31(3):373–378, 1989.

J. R. Trout and W. H. Swallow. Regular and inverse interval estimation of individual observations using uniform confidence bands. *Technometrics*, 21(4):567–574, Nov. 1979. ISSN 0040-1706, 1537-2723.

J. M. VerHoef. Who invented the delta method? *The American Statistician*, 66(2): 124–127, 2012.

E. J. Williams. A note on regression methods in calibration. *Technometrics*, 11(1): 189–192, Feb. 1969. ISSN 0040-1706, 1537-2723.

Q. Zeng and M. Davidian. Bootstrap-adjusted calibration confidence intervals for immunoassay. *Journal of the American Statistical Association*, 92(437):278–290, Mar. 1997a.

Q. Zeng and M. Davidian. Bootstrap-adjusted calibration confidence intervals for immunoassay. *Journal of the American Statistical Association*, 92(437):278–290, 1997b.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
27–03–2014	Doctoral Dissertation	Oct 2011–Mar 2014

4. TITLE AND SUBTITLE

Topics in Statistical Calibration

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Greenwell, Brandon M., Civilian

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
WPAFB, OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENC-DS-14-M-01

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Office of Scientific Research (AFOSR/RTA)
Dr. David S. Stargel
875 N. Randolph Street, Suite 325, Room 3112
Arlington, VA 22203-1768
david.stargel@afosr.af.mil

10. SPONSOR/MONITOR'S ACRONYM(S)

AFOSR/RTA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

13. SUPPLEMENTARY NOTES

This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

14. ABSTRACT

Calibration, more generally referred to as inverse estimation, is an important and controversial topic in statistics. In this work, both semiparametric calibration and the application of calibration to grouped data is considered, both of which may be addressed through the use of the linear mixed-effects model. A method is proposed for obtaining calibration intervals that has good coverage probability when the calibration curve has been estimated semiparametrically and is biased. The traditional Bayesian approach to calibration is also expanded by allowing for a semiparametric estimate of the calibration curve. The usual methods for linear calibration are then extended to the case of grouped data, that is, where observations can be categorized into a finite set of homogeneous clusters. Observations belonging to the same cluster are often similar and cannot be considered as independent; hence, we must account for within-subject correlation when making inference. Estimation techniques begin by extending the familiar Wald-based and inversion methods using the linear mixed-effects model. Then, a simple parametric bootstrap algorithm is proposed that can be used to either obtain calibration intervals directly, or to improve the inversion interval by relaxing the normality constraint on the approximate predictive pivot. Many of these methods have been incorporated into the R package, `investr`, which has been developed for analyzing calibration data.

15. SUBJECT TERMS

Calibration, Bootstrap, Linear mixed-effects model, Smoothing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Christine M. Schubert Kabban, AFIT/ENC
U	U	U	UU	134	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636 x4549 christine.schubertkabban@afit.edu