Noninformative Prior Bayesian Analysis for Statistical Calibration Problems

Daniel R. Eno

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Keying Ye, Chair
Jesse C. Arnold
Eric P. Smith
George R. Terrell
Robert L. Wheeler

April 9, 1999
Blacksburg, Virginia

Keywords: Reference prior, Probability matching prior, Frequentist coverage, Polynomial regression, Multivariate regression, Heteroscedasticity

Noninformative Prior Bayesian Analysis for Statistical Calibration Problems

Daniel R. Eno

(ABSTRACT)

In simple linear regression, it is assumed that two variables are linearly related, with unknown intercept and slope parameters. In particular, a regressor variable is assumed to be precisely measurable, and a response is assumed to be a random variable whose mean depends on the regressor via a linear function. For the simple linear regression problem, interest typically centers on estimation of the unknown model parameters, and perhaps application of the resulting estimated linear relationship to make predictions about future response values corresponding to given regressor values. The linear statistical calibration problem (or, more precisely, the *absolute* linear calibration problem), bears a resemblance to simple linear regression. It is still assumed that the two variables are linearly related, with unknown intercept and slope parameters. However, in calibration, interest centers on estimating an unknown value of the regressor, corresponding to an observed value of the response variable.

We consider Bayesian methods of analysis for the linear statistical calibration problem, based on noninformative priors. Posterior analyses are assessed and compared with classical inference procedures. It is shown that noninformative prior Bayesian analysis is a strong competitor, yielding posterior inferences that can, in many cases, be correctly interpreted in a frequentist context.

We also consider extensions of the linear statistical calibration problem to polynomial models and multivariate regression models. For these models, noninformative priors are developed, and posterior inferences are derived. The results are illustrated with analyses of published data sets. In addition, a certain type of heteroscedasticity is considered, which relaxes the traditional assumptions made in the analysis of a statistical calibration problem. It is shown that the resulting analysis can yield more reliable results than an analysis of the homoscedastic model.

For Julie, Saffron, and Conifer

Acknowledgements

Twenty-nine years of my life have led to the completion of a Ph.D. degree. The people who have influenced me over that time are too numerous to mention. I extend special thanks to the following people, whose influence has most directly contributed to this dissertation.

- My advisor, Keying Ye, for inspiring me to study Bayesian statistics, and for all of his guidance on my research.
- My committee members, Jesse Arnold, Eric Smith, George Terrell, and Robert Wheeler.
- The Virginia Tech Statistics Department and Statistical Consulting Center, for financial support, excellent instruction, and invaluable experience. I leave Virginia Tech feeling qualified to tackle any applied statistics problem I may encounter, but also with knowledge of the limitations of my trade.
- My parents, Phillip and Elaine Eno, for instilling in me a desire to learn, and my brother, David Eno, for many interesting conversations about mathematics, computer science, and survival in the working world.
- The many students that I have known as colleagues, with whom I have shared insights and ideas.
- Finally, I wish to thank my wife, Julie Mihalcik-Eno, for her love and support, and for bringing our dogs Saffron and Conifer into my life.

Contents

List of Tables **viii**

List of Figures **ix**

1 Background on Calibration Problems **1**

1.1 Introduction 1

1.2 Frequentist Approaches to Calibration Problems 3

1.2.1 The "Classical" Approach 3

1.2.2 The "Inverse" Approach 4

1.3 Bayesian Approaches to Calibration Problems 6

1.3.1 Hoadley's Approach 6

1.3.2 The Approach of Hunter & Lamboy 7

1.4 Multivariate and Polynomial Calibration 8

1.4.1 Multivariate Linear Calibration 8

1.4.2 Polynomial Calibration 8

2 Background on Noninformative Priors **9**

2.1 Introduction 9

2.2 Jeffreys' Prior 10

2.3 The Reference Prior Method 11

2.4 Probability Matching Priors 13

2.5 Noninformative Priors for Calibration Problems 14

3 Reference Prior Analysis of the Univariate Linear Calibration Problem – Constant Variance Case 16

3.1 Statement of the Problem 16

3.2 The Fisher Information Matrix and the Jeffreys' Prior 17

3.3 The Reference Priors 18

3.4 Reference Posterior Analyses 18

3.4.1 Marginal Reference Posteriors for x_0 18

3.4.2 Existence of Moments of $\pi_k^R(x_0|\mathbf{y})$ 20

3.4.3 Gibbs Sampling of the Reference Posteriors 20

3.5 Gibbs Sampling of the Jeffreys' Posterior 22

3.6 An Example of Univariate Linear Calibration 23

3.6.1 Marginal Inferences About x_0 24

3.6.2 Reference Posterior Inferences About All Model Parameters 27

4 Probability Matching Priors for the Univariate Linear Calibration Problem with Constant Variance 31

4.1 Theoretical Development of Probability Matching Priors 31

4.2 Finite Sample Frequentist Coverage Properties of Reference Posterior Inferences Concerning x_0 34

4.2.1 Dependence of the Frequentist Coverage on Model Parameters 35

4.2.2 Simulation Results for the Reference Priors 37

4.3 Simulation Results for Hoadley's Prior 42

5 Noninformative Prior Analysis of the Univariate Linear Calibration Problem – Two Variance Case 46

5.1 The Need for an Extended Calibration Model 46

5.2 Extension of the Univariate Linear Calibration Model 47
5.3 The Fisher Information Matrix and the Jeffreys' Prior 47
5.4 Development of Probability Matching Priors for x_0 48
5.5 Posterior Analysis for a Probability Matching Prior for x_0 50
5.5.1 Integrability of the Posterior 50
5.5.2 Gibbs Sampling of the Posterior 51
5.6 Development of Probability Matching Priors for κ 53
5.7 An Empirical Assessment of the Extended Model Analysis 54
5.8 An Illustrative Example 63

6 Reference Prior Analysis of The Multivariate Linear Calibration Problem 66

6.1 Formulation of the Multivariate Calibration Problem 66
6.2 The Reference Prior for the Multivariate Calibration Problem 67
6.3 Reference Posterior Analysis 68
6.3.1 The Marginal Reference Posterior of $\mathbf{x}_0$ 68
6.3.2 Gibbs Sampling of the Reference Posterior 70
6.4 An Example of Multivariate Linear Calibration 73
6.4.1 The Marginal Posterior for x_0 73
6.4.2 Posterior Inferences via the Gibbs Sampler 78

7 Reference Prior Analysis of the Univariate Polynomial Calibration Problem 79

7.1 Formulation of the Univariate Polynomial Calibration Problem 79
7.2 A Reference Prior for the Univariate Polynomial Calibration Problem 80
7.3 The Marginal Reference Posterior for x_0 81
7.3.1 Derivation of the Marginal Posterior for x_0 81
7.3.2 Integrability of the Reference Posterior 82

7.4 An Illustrative Example of Polynomial Calibration 84

7.4.1 Description of the Experimental Data 84

7.4.2 Posterior Analyses for the Clearance Circle Assay Example 88

Appendix **96**

A.1 The Reference Prior Algorithm 96

A.2 Development of the Reference Priors for the Univariate Linear Calibration Problem with a Single Standard Deviation Parameter 98

A.2.1 The Reference Prior for the Parameter Grouping $(x_0, \alpha, \beta, \sigma)$ 99

A.2.2 The Reference Prior for the Parameter Grouping $(x_0, (\alpha, \beta, \sigma))$ 102

A.3 Integrability of the Posterior for the Extended Univariate Linear Calibration Problem 104

A.4 Development of the Reference Prior for the Multivariate Linear Calibration Problem 109

A.4.1 Evaluation of the Information Matrix 109

A.4.2 Definition of the Compact Sets $\boldsymbol{\Theta}^l$ 112

A.4.3 Derivation of the Reference Prior 112

A.5 Development of the Reference Prior for the Univariate Polynomial Calibration Problem 116

A.5.1 The Likelihood Function and the Fisher Information Matrix 116

A.5.2 Derivation of the Reference Prior for the Grouping $(\mathbf{x}_0, (\alpha, \boldsymbol{\beta}), \sigma)$ . . . 117

Bibliography **121**

Vita **124**

List of Tables

3.1 Data for the plasma example from Aitchison and Dunsmore (1975). 24

3.2 Estimates and 90% intervals for x_0 in the plasma example. 26

6.1 Data for the wheat example from Brown (1982). 74

6.2 Interval estimates for the individual components of $\mathbf{x}_0$, for observation number 18 in table 6.1. 78

7.1 Data for the clearance circle assay from Aitchison and Dunsmore (1975). . . 85

7.2 Summary of the inferences for the clearance circle assay example. 89

List of Figures

3.1 Scatterplot of the data from table 3.1. 25

3.2 Marginal posterior distributions for x_0 corresponding to the reference priors and Hoadley's prior. The response from the second stage of experimentation is assumed to be $y_0 = 3.7$. 26

3.3 Marginal posterior distributions for x_0 corresponding to the reference priors and Hoadley's prior. The response from the second stage of experimentation is assumed to be $y_0 = 5.2$. 27

3.4 Trace of the output of the Gibbs sampler for x_0, corresponding to $y_0 = 3.7$. . 28

3.5 Marginal kernel density estimates for the model parameters α, β, σ, and x_0, corresponding to $y_0 = 3.7$. 29

3.6 Marginal kernel density estimates for the model parameters α, β, σ, and x_0, corresponding to $y_0 = 5.2$. 30

4.1 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors ($x_0 = 0$). 39

4.2 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors ($x_0 = 0.5$). 40

4.3 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors ($x_0 = 1$). 41

4.4 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors ($\beta = 1$). 42

4.5 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distribution of x_0, corresponding to Hoadley's prior ($x_0 = 0$). 43

4.6 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distribution of x_0, corresponding to Hoadley's prior ($x_0 = 0.5$). 44

4.7 Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distribution of x_0, corresponding to Hoadley's prior ($x_0 = 1$). 45

5.1 Estimated frequentist coverage probabilities for the marginal distributions of x_0 ($x_0 = 0$; $\beta = 5$; $\kappa = 0.5, 1$). 57

5.2 Estimated frequentist coverage probabilities for the marginal distributions of x_0 ($x_0 = 0$; $\beta = 5$; $\kappa = 1.5, 2$). 58

5.3 Estimated frequentist coverage probabilities for the marginal distributions of x_0 ($x_0 = 0$; $\beta = 2$; $\kappa = 0.5, 1$). 59

5.4 Estimated frequentist coverage probabilities for the marginal distributions of x_0 ($x_0 = 0$; $\beta = 2$; $\kappa = 1.5, 2$). 60

5.5 Estimated frequentist coverage probabilities for the marginal distributions of x_0 ($x_0 = 1$; $\beta = 5$; $\kappa = 0.5, 1$). 61

5.6 Estimated frequentist coverage probabilities for the marginal distributions of x_0 ($x_0 = 1$; $\beta = 5$; $\kappa = 1.5, 2$). 62

5.7 Scatterplot of the data from Afifi and Azen (1972). 63

5.8 Marginal kernel density estimates for x_0, for the example from Afifi and Azen (1972). 64

5.9 Marginal kernel density estimate for κ, from the extended model analysis of the example from Afifi and Azen (1972). 65

6.1 Scatterplots relating the two regressors and four responses from the data in table 6.1. 75

6.2 Plot of the unnormalized marginal posterior distribution of $\mathbf{x}_0$, for observation number 18 in table 6.1. 76

6.3 Contour plot for the marginal posterior distribution of $\mathbf{x}_0$, for observation number 18 in table 6.1. 77

6.4 Gibbs sampling results for $\mathbf{x}_0$, for observation number 18 in table 6.1. 78

7.1 Scatterplot of the control data displayed in table 7.1. 85

7.2 Scatterplot of the transformed control data from table 7.1. The response is transformed via the square root transformation, and the regressor is transformed via the log transformation. 86

7.3 Scatterplot of the transformed control data from table 7.1. The response is transformed via the square root transformation, and the regressor is transformed via the reciprocal cube root transformation. 87

7.4 Reference prior for the quadratic model, for the clearance circle assay example from Afifi and Azen (1972). 89

7.5 Marginal reference posterior for $\log x_0$, from the quadratic polynomial model. This density corresponds to test preparation #1. 90

7.6 Marginal reference posterior for $\log x_0$, from the quadratic polynomial model. This density corresponds to test preparation #2. 91

7.7 Marginal reference posterior for x_0, from the quadratic polynomial model. This density corresponds to test preparation #1. 92

7.8 Marginal reference posterior for x_0, from the quadratic polynomial model. This density corresponds to test preparation #2. 93

7.9 Marginal reference posterior for x_0, from the linear regression model. This density corresponds to test preparation #1. 94

7.10 Marginal reference posterior for x_0, from the linear regression model. This density corresponds to test preparation #2. 95

Chapter 1

Background on Calibration Problems

1.1 Introduction

The problem of *statistical calibration*, sometimes referred to as *inverse regression*, bears a resemblance to common regression problems. In the univariate case, it is assumed that two variables x and y are related via a mathematical function of a specified functional form, possibly containing one or more unknown parameters. The relationship between the variables x and y is not assumed to be deterministic; rather, the assumption is that, conditional on a fixed value of x, the expected value of y is determined by the regression function:

$$E[y|x] = f(x; \boldsymbol{\beta}),$$

where $\boldsymbol{\beta}$ represents a vector of unknown parameters. Additional assumptions are often incorporated, specifying the form of the deviations in observed y values about their mean values. For example, it is often assumed that the deviations between observed y values and their means are independent and identically distributed normal deviates, with mean zero and common variance σ^2. This assumption introduces an additional model parameter, namely the error standard deviation σ.

Note that in the above description, the variables x and y are treated as fundamentally different types of variables. The variable y, called the *response*, is assumed to be a random variable, while the variable x, called a *regressor* or a *predictor*, is assumed to be non-random. Typically, one assumes that the regressor variable can be controlled by an experimenter.

In regression, the focus of interest is on the estimation of unknown parameters, and possibly on the prediction of future y values corresponding to fixed x values. This type of prediction problem is relatively straightforward, due to the fact that a future y value is a random variable, and probability statements about the value can be made directly.

The problem of statistical calibration differs from that of standard regression, in that a primary point of interest in calibration is the estimation of an *unknown* value of the regressor, say x_0, corresponding to an *observed* value of the response, say y_0 (or perhaps several observed values of the response, with mean y_0). A calibration experiment is carried out in two stages. In the first stage, a sample of pairs $(x_i, y_i), i = 1, 2, \ldots, n$, of data is observed. In the second stage, one or more observations of the response are observed, all corresponding to a single unknown value of the regressor variable. Since the regressor is not assumed to be a random variable, inferences in the calibration problem are inherently different from inferences in a regression (or prediction) problem.

An example of univariate linear calibration is presented by Aitchison and Dunsmore (1975), p. 183. The x variable in this case is the true concentration of a certain enzyme in human blood plasma. This quantity can be measured via a very accurate, but expensive and time consuming, laboratory method. Alternatively, the concentration of the enzyme in a plasma sample can be measured with an autoanalyzer instrument. This autoanalyzer method is less expensive and faster to implement than the laboratory method, but it is also less precise. The determination of the enzyme concentration by the autoanalyzer is denoted as y. In an effort to calibrate the autoanalyzer, the enzyme concentrations of several blood plasma samples are determined by both the laboratory method and the autoanalyzer method. The resulting dataset provides information about the relationship between x and y. At a future date, the autoanalyzer will be used to determine the (unknown) concentration of the enzyme in a blood plasma sample from a patient. Of interest is the true enzyme concentration (x_0) corresponding to the observed response from the autoanalyzer (y_0). This dataset will be considered in more detail in section 3.6.

A univariate calibration problem can be classified as being either a *controlled* calibration problem or a *random* (or *natural*) calibration problem. In a controlled calibration problem, it is assumed that the values of the regressor from the first stage of experimentation are fixed by the researcher. On the other hand, in a random calibration problem, the values $x_1, x_2, \ldots, x_n$ of the regressor variable are randomly generated. In the latter case, although the regressor and response are both random variables, they are still treated asymmetrically. It is assumed that the values of the regressor at the first stage of experimentation are precisely measurable, while the measurements of the response variable are subject to error. This asymmetry is defined by the model and error structure which the relationship between the regressor and the response is assumed to have.

A calibration problem satisfying the above assumptions is more specifically referred to as an *absolute* calibration problem. This can be contrasted with a *comparative* calibration problem, in which measurements of both x and y are subject to error. In a comparative

calibration problem, there is no inherent asymmetry between the x and y variables. In this work, we use the term *calibration* to refer to *absolute calibration*; comparative calibration is not considered.

A review of the literature on statistical calibration is given by Osbourne (1991). Here, some approaches to the solution of the calibration problem are summarized.

1.2 Frequentist Approaches to Calibration Problems

1.2.1 The "Classical" Approach

The paper of Eisenhart (1939) described a solution to the calibration problem that is now called the "classical" solution. Eisenhart assumed that the relationship between x and y was of a simple linear form:

$$E[y|x] = \alpha + \beta x,$$

though the approach can be extended to more complicated models. The defining characteristic of the type of analysis described by Eisenhart is that the line relating y to x is fit by minimizing the residual sum of squares *in the direction of the response* y, using the data collected at the first stage of experimentation. That is, we estimate α and β with the values $\hat{\alpha}$ and $\hat{\beta}$ that minimize

$$SSE_y = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta} x_i)^2.$$

Eisenhart asserts that " ... the fitting should be done in terms of the deviations which actually represent 'error'."

After the slope and intercept parameters are estimated, the unknown regressor value x_0 can be estimated, using the observed response value y_0:

$$\hat{x}_{0,c} = \frac{y_0 - \hat{\alpha}}{\hat{\beta}}.$$

This estimator of x_0 is known as the "classical" estimator. Although this solution does not require a specific distributional assumption about the deviations of the y values from their means, this estimator would be the maximum likelihood estimator of x_0 if independent, identically distributed normal errors were assumed.

Often, a researcher studying a calibration problem will need more information than a simple point estimate of the unknown x_0. An interval estimate may be required, and can be

provided by appealing to methodology closely related to that of Fieller (1954) if the errors are assumed to be distributed as independent identical normal deviates. This method provides an interval estimate in the form of a *fiducial* interval for x_0, and is described in Brown (1993), p. 23–26. The interval may have peculiar properties, depending on the observed value of $\hat{\beta}$. In fact, if $\hat{\beta}$ is not significantly different from zero relative to σ, the region may not even be a connected interval. It is possible that the resulting region will be the union of two disjoint unbounded intervals. Such a result is, of course, of no practical value. In practice, calibration is unlikely to be attempted if β is near zero. In a case where $\hat{\beta}$ is significantly different from zero, an approximate $100 \cdot (1-\gamma)$ percent interval estimate for x_0 (for the case of a single observed response at the second stage of data collection), presented in Brown's book, is

$$\frac{y_0 - \hat{\alpha}}{\hat{\beta}}\left(1 + \frac{\hat{\sigma}^2 t^2}{\hat{\beta}^2 S_{xx}}\right) \pm \frac{\hat{\sigma} t}{\hat{\beta}}\left(1 + \frac{1}{2n} + \frac{(y_0 - \hat{\alpha})^2 + \hat{\sigma}^2 t^2}{2\hat{\beta}^2 S_{xx}}\right), \tag{1.1}$$

where

$$\hat{\sigma} = \sqrt{\frac{SSE_y}{n-2}},$$

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$S_{xx} = \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

and t is the $100 \cdot (1 - \gamma/2)$ percentile of the Student's t distribution with $n - 2$ degrees of freedom.

1.2.2 The "Inverse" Approach

Krutchkoff (1967) suggested that instead of fitting the line relating y to x by minimizing the residual sum of squares in the y direction, it would be preferable to fit the line by minimizing the residual sum of squares in the x direction:

$$SSE_x = \sum_{i=1}^{n} (x_i - \hat{\gamma} - \hat{\delta} y_i)^2.$$

Here, of course, the slope and intercept parameters are different from those discussed in the previous section. At the second stage of experimentation, x_0 can be estimated directly

by substituting y_0 into the fitted equation. We let $\hat{x}_{0,I}$ denote the inverse estimator of x_0. Krutchkoff based his conclusion on the results of a simulation study, in which he found that the mean squared error of estimation for x_0 was uniformly less for this estimator than for the classical estimator.

Following the publication of Krutchkoff's paper, several responses appeared. Some of these responses criticized Krutchkoff's conclusion, while others defended it. Williams (1969) noted that the classical estimator has infinite mean squared error, while the inverse estimator has finite mean squared error if $n \geq 4$. He also demonstrated that any unbiased estimator of x_0 must have infinite mean squared error. Based on these facts, he concluded that mean squared error is not a suitable criterion for the comparison of the estimators.

Berkson (1969) pointed out that the inverse procedure does not yield consistent estimators for α, β, and x_0, while the classical procedure does. He also considered a hypothetical situation in which the sample size at the first stage of experimentation is allowed to tend to infinity. In this case, the estimates of α and β from the classical procedure would be equal to the true parameter values, while the corresponding estimates from the inverse procedure would be biased. For some values of x_0, he continued, the classical procedure would yield smaller mean squared error in estimating x_0 than would the inverse procedure, although the inverse procedure would be superior if x_0 were very near the mean of the x values from the first stage of experimentation. Although it is not possible in practice to allow n to be infinite, Berkson remarks that in practical calibration problems, n is often taken to be very large, so that the asymptotic results are approximately correct. Finally, he demonstrated via simulation that for some parameter values (not considered by Krutchkoff (1967)) the classical procedure can be shown to outperform the inverse procedure. In the same year that Berkson's paper appeared, another paper by Krutchkoff (1969) presented similar simulation results.

One paper that provided some support to the inverse procedure was written by Lwin and Maritz (1982). Their approach was based on a method referred to as *linear compound estimation*. This approach seeks constants k_0 and k_1 so as to minimize the quantity

$$E\left[\sum_{i=1}^{n}(k_0 + k_1 y_i - x_i)^2\right].$$

They point out that "This is an important criterion to be satisfied since in most practical situations one would be concerned with calibration within a fixed range of x values covered by the calibration experiment." Using the linear compound estimation approach, Lwin and Maritz showed that the classical approach is favored if unbiased estimation of x_0 is required, while the inverse approach is favored if unbiasedness is not required.

Other papers followed Krutchkoff (1967); these will not be discussed here. The references can be found in the paper of Osbourne (1991), p. 315–317.

1.3 Bayesian Approaches to Calibration Problems

We begin this section with a discussion of the Bayesian approach to statistical inference. An experiment is to be performed, which will give rise to some data $\mathbf{y} = (y_1, y_2, \ldots, y_n)$. We let $\mathbf{Y}$ denote the space in which the observed data may fall. It is assumed that the probability distribution of the data depends on a set of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)$, which are contained in a *parameter space* $\boldsymbol{\Theta}$. The probability density function of the observed data, conditional on a particular value of $\boldsymbol{\theta}$, can be denoted as $f(\mathbf{y}|\boldsymbol{\theta})$. Often, this is thought of as a function of $\boldsymbol{\theta}$, and is referred to as the *likelihood function*. The goal of statistical inference is to obtain information about $\boldsymbol{\theta}$ from the data $\mathbf{y}$.

A Bayesian analysis combines the information about $\boldsymbol{\theta}$ contained in $\mathbf{y}$ with information that is available about $\boldsymbol{\theta}$ before the experiment is performed. This *prior* information is summarized in a *prior distribution* $\pi(\boldsymbol{\theta})$. The result of a Bayesian analysis is a *posterior distribution* for $\boldsymbol{\theta}$, $\pi(\boldsymbol{\theta}|\mathbf{y})$, which is determined via *Bayes' Rule*:

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta})}{m(\mathbf{y})},$$

where

$$m(\mathbf{y}) = \int_{\boldsymbol{\Theta}} \pi(\boldsymbol{\theta}) f(\mathbf{y}|\boldsymbol{\theta}) \, d\boldsymbol{\theta}$$

is the marginal distribution of $\mathbf{y}$.

1.3.1 Hoadley's Approach

Hoadley (1970) presented a Bayesian solution to the calibration problem. The model proposed by Hoadley was more structured than that discussed above, in that the errors about the regression line were assumed to have a particular parametric form. Specifically, Hoadley assumed the model

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n,$$

and

$$y_i = \alpha + \beta x_0 + \epsilon_i, \qquad i = n+1, n+2, \ldots, n+c, \tag{1.2}$$

where it is assumed that for each $i = 1, 2, \ldots, n + c$, we have

$$\epsilon_i \sim N(0, \sigma^2), \tag{1.3}$$

and that the ϵ_i are mutually independent. The unknown parameters in this parametric model are α, β, σ, and x_0.

In his analysis, Hoadley specified a "noninformative" prior distribution for the parameters (α, β, σ), but left the prior distribution of x_0 unspecified. An assumption in the analysis was that x_0 is *a priori* independent of (α, β, σ), so that the joint prior distribution of $(x_0, \alpha, \beta, \sigma)$ is the product of the specified priors for (α, β, σ) and x_0.

One point of interest from Hoadley's paper was that the "inverse" estimator was shown to be a Bayes estimator (under squared error loss), if a specific prior distribution for x_0 was used. The particular prior distribution for x_0 giving this result is

$$x_0 \sim t_{n-3}\left(\bar{x}, \frac{n+1}{n-3}\frac{S_{xx}}{n}\right),$$

where

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

and

$$S_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Although Hoadley studied this prior distribution, he made no claims about the desirability of resulting posterior inferences, except to note that this might be a reasonable prior distribution to use if there was reason to believe that the unknown x_0 would likely be close to the mean $\bar{x}$ of the regressor values from the first stage of experimentation.

1.3.2 The Approach of Hunter & Lamboy

Hunter and Lamboy (1981) also considered the calibration problem from a Bayesian point of view. A primary difference between their approach and Hoadley's approach was that while Hoadley assumed *a priori* that x_0 and (α, β, σ) were independent, Hunter and Lamboy assumed *a priori* that $E[y_0|x_0]$ and (α, β, σ) were independent. In the discussion following the paper of Hunter and Lamboy, several discussants (Hill (1981), Lawless (1981), Lwin (1981), Orban (1981)) argued that Hoadley's assumption is more likely to be appropriate in practice. Also, Hill (1981) pointed out that the analysis of Hunter and Lamboy could be generated from Hoadley's analysis, by using a particular form for the prior distribution of x_0 (which involved β as well).

1.4 Multivariate and Polynomial Calibration

1.4.1 Multivariate Linear Calibration

Just as simple linear regression can be extended to multiple regression and multivariate regression, the linear calibration problem can be extended to more general settings. The *multivariate linear calibration problem* is similar to the univariate calibration problem, but in the multivariate case it is assumed that q response variables depend linearly on p regressor variables. As before, the data for the calibration experiment are collected in two stages. At the first stage, n observations $(\mathbf{x}_i, \mathbf{y}_i)$ are made. Here, each $\mathbf{x}_i$ is a $p \times 1$ vector, and each $\mathbf{y}_i$ is a $q \times 1$ vector. It is assumed that each element of $\mathbf{x}_i$ is precisely determined, while the elements of $\mathbf{y}_i$ are measurements which are subject to error. At the second stage of data collection, one or more observations of $\mathbf{y}$ are made, and interest centers on the determination of the unknown $\mathbf{x}_0$ vector giving rise to these observations. A new concern comes to light as we extend the univariate calibration problem to the multivariate case, concerning the relative sizes of q and p. Specifically, if the unknown value $\mathbf{x}_0$ is to be uniquely determined, we require that the number of responses q be at least as large as the number of regressors p.

Multivariate calibration was studied extensively by Brown (1982). In his paper, Brown discussed procedures for multivariate calibration that are analogous to the classical and inverse procedures for univariate calibration. In addition, Brown provided an extension of Hoadley's procedure to a Bayesian analysis of the multivariate calibration problem. In this Bayesian solution, Brown used a noninformative prior for all model parameters except the unknown $\mathbf{x}_0$. He studied a certain multivariate Student's t distribution as the prior for $\mathbf{x}_0$. As was the case in Hoadley's earlier work, this prior was studied because the resulting mode of the marginal posterior distribution of $\mathbf{x}_0$ is equal to the inverse estimator of $\mathbf{x}_0$. Brown did not propose any noninformative prior distributions for $\mathbf{x}_0$.

1.4.2 Polynomial Calibration

A further generalization of the multivariate calibration problem can be made, in which the p regressor variables are determined by $\nu < p$ underlying control variables. An example of this would be a model in which ν control variables enter the regression equation in the form of a multivariate polynomial. In this case, we require that q be at least as large as ν, in order for the unknown regressor values at the second stage of calibration to be identifiable. Brown (1982) also considered polynomial calibration models, from frequentist and Bayesian perspectives. In a later chapter, we consider the *univariate* linear calibration problem, in which $q = \nu = 1$.

Chapter 2

Background on Noninformative Priors

2.1 Introduction

As noted in chapter 1, a Bayesian analysis combines prior information about model parameters with information from observed data, thereby generating a posterior distribution. Such an analysis *requires* two inputs, namely the prior distribution and the likelihood function. The prior distribution can be chosen to represent the beliefs of the researcher before observing the results of an experiment; this results in a proper subjective Bayesian analysis. Often, however, it is difficult for a researcher to specify prior beliefs about model parameters, and to cast them into the form of a prior probability distribution.

A noninformative prior is a function which is used in place of a subjective prior distribution when little or no prior information is available. The term "noninformative" is used to connote the lack of subjective beliefs used in formulating such a prior. However, one can think of a noninformative prior as simply being a function that is formally used in place of a subjective prior distribution, for the purpose of accomplishing some goal. For example, the *reference prior method* (Berger and Bernardo (1992)) is based on the notion of maximizing the information provided by the observed data in the resulting analysis. This method attempts to be truly "noninformative" – that is, to generate priors that represent vague prior information about the unknown parameters. *Probability matching priors* (e.g., Welsh and Peers (1963), Stein (1985), Tibshirani (1989)), on the other hand, are selected so that certain quantiles of the resulting posterior distribution match frequentist coverage probabilities (at least asymptotically). Hence, it is perhaps deceptive to call probability matching priors "noninformative". However, this term has historically been used to describe any type of

prior distribution that is not based on a formulation of prior beliefs, so we will use it here.

Kass and Wasserman (1996) provide a review of many methods of generating noninformative priors that have been proposed. Three types of noninformative priors will be considered in this work. These are described below.

2.2 Jeffreys' Prior

One of the earliest methods of defining noninformative priors was based on the *principle of insufficient reason*. This method, sometimes referred to as *Laplace's rule*, prescribes a uniform prior on the parameter space $\boldsymbol{\Theta}$.

If the parameter space is finite, a uniform prior distribution is *proper* (that is, integrable), but if the parameter space is infinite, such a prior may be *improper* (not integrable). The use of an improper prior, although not truly supported by Bayes' Theorem, does not necessarily lead to problems in a Bayesian analysis. If the posterior corresponding to an improper prior is integrable, it can be normalized to be a true density, and interpretation can proceed as if the prior were proper. Hence, if an improper prior is used for obtaining inferences, it is important to verify that the posterior distribution is integrable before making posterior inferences.

Laplace's rule, and the principle of insufficient reason, are intuitively appealing. The reasoning is that if no prior information is available that favors certain parameter values over others, then all parameter values should be considered equally likely. One of the difficulties encountered in interpreting Laplace's rule is that it is not invariant to the choice of parameterization. That is, if $\boldsymbol{\theta}$, which has a uniform distribution, is transformed to a new parameter vector $\boldsymbol{\phi}$ via a one-to-one function, then $\boldsymbol{\phi}$ will not necessarily have a uniform distribution. If there is insufficient reason to favor certain values of $\boldsymbol{\theta}$ over others, then it can be argued that there is also insufficient reason to favor certain values of $\boldsymbol{\phi}$ over others. Hence, the lack of invariance of the uniform prior is considered undesirable. It should be noted, however, that uniform priors have been applied to many problems, and often the results are entirely satisfactory.

Jeffreys (1946) proposed a method of generating noninformative priors which is invariant to transformations of the parameter vector. His method begins by considering the Fisher information matrix,

$$\mathcal{I}(\boldsymbol{\theta}) = -E\left(\frac{\partial}{\partial\boldsymbol{\theta}}\frac{\partial}{\partial\boldsymbol{\theta}'}\log\mathcal{L}\right),$$

where $\mathcal{L}$ denotes the likelihood function and the expectation is over $\mathbf{y} \in \mathbf{Y}$. *Jeffreys' prior* is taken to be

$$\pi^J(\boldsymbol{\theta}) \propto \sqrt{|\mathcal{I}(\boldsymbol{\theta})|}.$$

Jeffreys' prior, like the uniform prior, may be improper. However, Jeffreys' prior is invariant, in the sense that if the Jeffreys' prior in one parameterization is transformed to a different parameterization, then the transformed prior will be the Jeffreys' prior in the new parameterization.

2.3 The Reference Prior Method

The *reference prior method*, introduced by Bernardo (1979) and further described by Berger and Bernardo (1992), is motivated by the notion of maximizing the expected amount of information about $\boldsymbol{\theta}$ provided by the data, $\mathbf{y}$. The amount of information provided by the experiment is quantified by the *Kullback-Liebler divergence*, which is defined by

$$D(g, h) = \int_{\boldsymbol{\Theta}} g(\boldsymbol{\theta}) \log\left(\frac{g(\boldsymbol{\theta})}{h(\boldsymbol{\theta})}\right) d\boldsymbol{\theta},$$

for any two densities g and h.

It would be natural to consider developing a prior which maximizes the expected value (over $\mathbf{y}$) of the information provided by the data:

$$E\bigg(D\big(\pi(\boldsymbol{\theta}|\mathbf{y}), \pi(\boldsymbol{\theta})\big)\bigg). \tag{2.1}$$

However, Berger and Bernardo (1992) point out that the prior distribution that maximizes this expected information is typically a discrete distribution, even when the parameter space is continuous.

The actual reference prior method stems from two modifications of the notion of maximizing the expected information provided by the data. Berger and Bernardo (1992) define $\mathbf{z}_t = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_t)$ to be a vector containing data from t independent replications of the experiment. Then we can replace the notion of maximizing (2.1) with that of maximizing

$$\lim_{t\to\infty} E\bigg(D\big(\pi(\boldsymbol{\theta}|\mathbf{z}_t), \pi(\boldsymbol{\theta})\big)\bigg). \tag{2.2}$$

As t increases, more and more information about $\boldsymbol{\theta}$ is obtained, so that as $t \to \infty$, *all* of the information about $\boldsymbol{\theta}$ that was not included in $\pi(\boldsymbol{\theta})$ would be accounted for in (2.2). However,

as $t \to \infty$, (2.2) typically becomes infinite, so choosing a prior distribution to maximize (2.2) is not possible. The second modification, to actually obtain the reference prior method, is to choose a prior distributions to maximize

$$E\bigg(D\big(\pi(\boldsymbol{\theta}|\mathbf{z}_t), \pi(\boldsymbol{\theta})\big)\bigg),$$

for each t, then define the reference prior to be the limit of these priors, as $t \to \infty$.

The algorithm for generating reference priors, as described by Berger and Bernardo (1992), is included in the appendix (section A.1). In this section, we describe only some of the features of the algorithm, which must be considered before attempting to generate reference priors.

The reference prior method requires the user to divide the individual parameters into ordered groups: $\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \ldots, \boldsymbol{\theta}_{(m)}$, where each group $\boldsymbol{\theta}_{(j)}$ contains one or more of the scalar parameters in $\boldsymbol{\theta}$. The reference prior is developed iteratively, first computing a prior for $\boldsymbol{\theta}_{(m)}$, then for $\boldsymbol{\theta}_{(m-1)}$ given $\boldsymbol{\theta}_{(m)}$, then for $\boldsymbol{\theta}_{(m-2)}$ given $\boldsymbol{\theta}_{(m-1)}$ and $\boldsymbol{\theta}_{(m)}$, etc. Note that in addition to being divided into groups, the parameters in $\boldsymbol{\theta}$ are also ordered. The actual form of a reference prior can depend both on the grouping and the ordering of the parameters. Hence, the reference prior method can generate many different priors for a given problem. Berger and Bernardo (1992) suggest that the ordering of the parameters should be done according to inferential importance, with $\boldsymbol{\theta}_{(1)}$ the most important parameter group and $\boldsymbol{\theta}_{(m)}$ the least important. Also, they suggest that it is usually best to have each of the parameter groups contain only one scalar parameter from $\boldsymbol{\theta}$, so that $m = p$. This recommendation is based on their experience in applying the reference prior method to various applied problems. They state that "we have not yet encountered an example in which the one-at-a-time reference prior is unappealing, and so our pragmatic recommendation is to use this reference prior unless there is a specific reason for using a certain grouping." In particular, in many cases, posterior Bayesian inferences agree with frequentist inferences when this reference prior is used.

Berger and Bernardo (1992) also state, concerning the ordering of the parameters in terms of inferential importance, that " ... beyond putting the 'parameters of interest' first, this is too vague to be of much use." They recommend that, if possible, all reference priors for which the parameters of interest are placed first in the ordering should be computed. This provides a set of prior distributions which can be compared, to assess the sensitivity of the resulting analyses to the choice of prior distribution.

Finally, we note that Jeffreys' prior is a special case of a reference prior. Specifically, Jeffreys' prior corresponds the the reference prior in which all model parameters are treated

in a single group ($m = 1$).

2.4 Probability Matching Priors

An argument that has been used to justify the use of some noninformative priors is that inferences arising from the Bayesian analyses match frequentist inferences. This is not true in general for noninformative prior Bayesian analyses. *Probability matching priors* are designed to have the property that certain posterior quantiles from the resulting Bayesian analyses match frequentist coverage probabilities, at least asymptotically.

The method described in Tibshirani (1989) for generating probability matching priors is based on transforming model parameters, so that the parameter of interest is orthogonal to other model parameters. The term "orthogonal" is interpreted to mean that the Fisher information matrix is block diagonal, with the parameter of interest corresponding to the upper left block (a single element) and the remaining parameters corresponding to the lower right block (see Cox and Reid (1987)). The prior distribution is then taken to be proportional to the square root of the upper left element of the information matrix in the new parameterization. This can be multiplied by any function of the other (orthogonal) parameters. In a situation where $\boldsymbol{\theta}$ is a scalar, the resulting prior is the Jeffreys' prior.

Following Cox and Reid (1987), we denote the model parameters as $(\psi, \phi_1, \phi_2, \ldots, \phi_q)$, where ψ is the parameter of interest. The information matrix in this parameterization is thus

$$\mathcal{I}^* = \begin{bmatrix} i^*_{\psi,\psi} & i^*_{\psi,\phi_1} & \cdots & i^*_{\psi,\phi_q} \\ i^*_{\phi_1,\psi} & i^*_{\phi_1,\phi_1} & \cdots & i^*_{\phi_1,\phi_q} \\ \vdots & \vdots & \ddots & \vdots \\ i^*_{\phi_q,\psi} & i^*_{\phi_q,\phi_1} & \cdots & i^*_{\phi_q,\phi_q} \end{bmatrix}$$

We make a transformation to the new parameters $(\psi, \boldsymbol{\lambda}')$, where $\boldsymbol{\lambda}' = (\lambda_1, \lambda_2, \ldots, \lambda_q)$, and let $\phi_t = \phi_t(\psi, \boldsymbol{\lambda})$ for $t = 1, 2, \ldots, q$.

An orthogonal parameterization may be found by solving the system of differential equations given by (see Cox and Reid (1987))

$$\sum_{r=1}^{q} i^*_{\phi_r,\phi_s} \frac{\partial \phi_r}{\partial \phi_s} = -i^*_{\psi,\phi_s}, \qquad \text{for} \qquad s = 1, 2, \ldots, q.$$

For the univariate linear calibration problem and an extension of it, these equations are derived explicitly and presented in later chapters.

The method of Tibshirani (1989) was based on ideas previously presented by Stein (1985). Tibshirani pointed out that the probability matching property is not rigorously proven for priors derived by solving the above differential equations. Datta and Ghosh (1995) provide a different solution to the problem of finding probability matching priors, which they justify rigorously. Their method provides a necessary and sufficient condition that a prior distribution must satisfy in order to have the probability matching property for a real-valued, twice continuously differentiable function of the model parameters, $t(\boldsymbol{\theta})$. Datta and Ghosh point out their method is more general than Tibshirani's, but will yield equivalent results when the function $t(\boldsymbol{\theta})$ of interest is defined to be the first parameter in an orthogonal parameterization.

2.5 Noninformative Priors for Calibration Problems

Hoadley (1970) noted in his paper that if an improper *uniform* prior distribution were prescribed for x_0, then the posterior distribution would be improper as well. He went on to state that "it seems that a proper prior for $[x_0]$ is a prerequisite for the use of the Bayes solution . . . ". However, in a Bayesian analysis it is often difficult for a researcher to prescribe a prior distribution for unknown parameters, due to a lack of prior information. In such cases, *noninformative* prior distributions can prove useful.

The reference prior algorithm for generating noninformative prior distributions was applied to the linear calibration problem by Kubokawa and Robert (1994). Their formulation of the calibration problem was similar to that studied by Hoadley (1970), but they allowed for more than one response variable to be measured at each setting of the control variable x. (This represents a specific type of *multivariate* calibration problem, with multiple y variables and a single x variable.) An assumption made by Kubokawa and Robert was that the errors in measurement of the response variables were independent. Later, du Plessis, van der Merwe, and Groenewald (1995) extended the work of Kubokawa and Robert (1994). Their model formulation was identical to that used by Kubokawa and Robert, but they considered more general error structures. In particular, they allowed for the errors in the observations of the response variables to be correlated for a particular observation of the response vector.

In each of these works presenting reference priors for calibration problems, only one of the possible parameter groupings was considered. Philippe and Robert (1998) considered reference priors corresponding to two different parameter groupings, and studied the frequentist coverage properties of these priors. In particular, they studied the frequentist coverage properties of *highest posterior density* (HPD) regions for x_0. Their work included a simulation study in which the frequentist coverage probability of HPD regions was evaluated for several

combinations of parameter values.

A comprehensive list of reference priors for the univariate linear calibration problem, considering all parameter groupings in which the x_0 is considered the parameter of interest, was presented by Ghosh, Carlin, and Srivastava (1995). Ghosh et al. (1995) also provided necessary and sufficient conditions for a prior distribution to have the probability matching property for the parameter x_0, using the methodology described by Datta and Ghosh (1995). The reference priors were shown to be probability matching priors, as was Jeffreys' prior.

Chapter 3

Reference Prior Analysis of the Univariate Linear Calibration Problem – Constant Variance Case

3.1 Statement of the Problem

For the univariate linear calibration problem, the data from the first stage of experimentation consist of the n pairs $(x_i, y_i), i = 1, 2, \ldots, n$. For each i, x_i is a known fixed value of the independent variable, and y_i is the corresponding observed response. We assume that each pair (x_i, y_i) is linearly related:

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n, \tag{3.1}$$

where ϵ_i is a random deviation around the underlying straight line relationship.

The data from the second stage of experimentation consist of c observations of the dependent variable, $y_{n+1}, y_{n+2}, \ldots, y_{n+c}$, assumed to be related to a single unknown x_0 via the same relationship:

$$y_i = \alpha + \beta x_0 + \epsilon_i, \qquad i = n+1, n+2, \ldots, n+c. \tag{3.2}$$

It is assumed that for each $i = 1, 2, \ldots, n+c$, we have

$$\epsilon_i \sim N(0, \sigma^2), \tag{3.3}$$

and that the ϵ_i are mutually independent.

Our primary interest is in estimating x_0. However, the regression parameters α and β, and the error standard deviation σ, are also unknown, and must be treated as such in the

analysis. In this chapter, reference priors for the linear calibration problem are presented, and posterior analyses are considered. An example of univariate linear calibration is also considered, and the inferences resulting from the classical procedure are compared with inferences based on various Bayesian analyses.

3.2 The Fisher Information Matrix and the Jeffreys' Prior

We begin by defining some notation. For any indexed quantities v and w, let

$$\mathfrak{s}_v^* = \sum_{i=1}^{n+c} v_i, \qquad \mathfrak{s}_{vv}^* = \sum_{i=1}^{n+c} v_i^2, \qquad \mathfrak{s}_{vw}^* = \sum_{i=1}^{n+c} v_i w_i, \qquad \bar{v}^* = \frac{\mathfrak{s}_v^*}{n+c},$$

$$\mathfrak{s}_v = \sum_{i=1}^{n} v_i, \qquad \mathfrak{s}_{vv} = \sum_{i=1}^{n} v_i^2, \qquad \mathfrak{s}_{vw} = \sum_{i=1}^{n} v_i w_i, \qquad \text{and} \qquad \bar{v} = \frac{\mathfrak{s}_v}{n}.$$

Note that the above sums of squares and cross products are not centered. Later, we will write the analogous centered sums of squares and cross products as

$$S_{vv}^* = \sum_{i=1}^{n+c} (v_i - \bar{v}^*)^2, \qquad S_{vv} = \sum_{i=1}^{n} (v_i - \bar{v})^2 \qquad \text{and} \qquad S_{vw} = \sum_{i=1}^{n} (v_i - \bar{v})(w_i - \bar{w}).$$

The Fisher information matrix for the calibration model with parameters $(x_0, \alpha, \beta, \sigma)$ is

$$\mathcal{I} = \frac{1}{\sigma^2} \begin{bmatrix} c\beta^2 & c\beta & c\beta x_0 & 0 \\ c\beta & n+c & \mathfrak{s}_x^* & 0 \\ c\beta x_0 & \mathfrak{s}_x^* & \mathfrak{s}_{xx}^* & 0 \\ 0 & 0 & 0 & 2(n+c) \end{bmatrix}.$$

In accordance with the above notational convention,

$$\mathfrak{s}_x^* = \sum_{i=1}^{n+c} x_i \quad \text{and} \quad \mathfrak{s}_{xx}^* = \sum_{i=1}^{n+c} x_i^2.$$

Here, we have written x_i in place of x_0, for $i = n+1, n+2, \ldots, n+c$.

From the information matrix, the Jeffreys' prior, as defined in equation (2.2), is easily shown to be

$$\pi^J(x_0, \alpha, \beta, \sigma^2) \propto |\beta| \sigma^{-4}.$$

3.3 The Reference Priors

We present a near exhaustive set of reference priors for the linear calibration problem. In each case, we treat x_0 as the parameter of greatest inferential importance. Given this restriction, however, all possible permutations and groupings of the parameters α, β, and σ are considered. The only aspect that causes differences among the reference priors is the number of parameters in the group involving σ. The results are summarized in the following theorem, whose proof is deferred to the appendix (section A.2).

THEOREM 1 *For the univariate linear calibration problem (as formulated above) the reference priors are*

$$\pi_k^R(x_0, \alpha, \beta, \sigma) \propto \frac{1}{\sigma^k \sqrt{S_{xx} + \frac{cn}{n+c}(x_0 - \bar{x})^2}},$$

where $k \in \{1, 2, 3\}$ is the number of parameters in the group involving σ.

3.4 Reference Posterior Analyses

3.4.1 Marginal Reference Posteriors for x_0

As noted previously, the primary goal of the calibration problem is the estimation of x_0, although inferences about the other model parameters might also be of interest. In this section, we derive the marginal posterior distributions for x_0, corresponding to the three reference priors. Following that, we present conditional distributions from which inferences about all model parameters may be obtained, via the Gibbs sampler.

The reference prior will be of the form

$$\begin{aligned}\pi_k^R(x_0, \alpha, \beta, \sigma) &\propto \sigma^{-k}\left(S_{xx} + \frac{cn}{n+c}(x_0 - \bar{x})^2\right)^{-\frac{1}{2}} \\ &= \frac{1}{\sigma^k \sqrt{S_{xx}^*}}\end{aligned}$$

as presented in Theorem 1. The likelihood function for the $n + c$ observations is

$$\mathcal{L} \propto \frac{1}{\sigma^{n+c}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c}(y_i - \alpha - \beta x_i)^2\right\}.$$

The joint posterior distribution for the parameters x_0, α, β, and σ is thus

$$\pi_k^R(x_0, \alpha, \beta, \sigma|\mathbf{y}) \propto \frac{1}{\sigma^{n+c+k}\sqrt{S_{xx}^*}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c}(y_i - \alpha - \beta x_i)^2\right\}.$$

We first integrate out α. By applying the technique of completing the square to the argument of the exponential function in the posterior, we see that

$$\alpha|\mathbf{y}, x_0, \beta, \sigma \sim N\left(\frac{\mathfrak{S}_{(y-\beta x)}^*}{n+c}, \frac{\sigma^2}{n+c}\right).$$

Hence,

$$\pi_k^R(x_0, \beta, \sigma|\mathbf{y}) \propto \frac{1}{\sigma^{n+c+k-1}\sqrt{S_{xx}^*}} \exp\left\{-\frac{1}{2\sigma^2}\left(\mathfrak{S}_{(y-\beta x)^2}^* - \frac{1}{n+c}\mathfrak{S}_{(y-\beta x)}^{*2}\right)\right\}.$$

Next, consider β. Once again, the technique of completing the square can be applied to the argument of the exponential function, to give

$$\beta|\mathbf{y}, x_0, \sigma \sim N\left(\frac{S_{xy}^*}{S_{xx}^*}, \frac{\sigma^2}{S_{xx}^*}\right).$$

Integrating out β thus yields

$$\pi_k^R(x_0, \sigma|\mathbf{y}) \propto \frac{1}{\sigma^{n+c+k-2}(\sqrt{S_{xx}^*})^2} \exp\left\{-\frac{1}{2\sigma^2}\left(S_{yy}^* - \frac{S_{xy}^{*2}}{S_{xx}}\right)\right\}$$

Finally, we consider σ. First, we make the change of variables $\sigma \to \sigma^2$, to give

$$\pi_k^R(x_0, \sigma^2|\mathbf{y}) \propto \frac{\exp\left\{-\frac{1}{2\sigma^2}\left(S_{yy}^* - \frac{S_{xy}^{*2}}{S_{xx}^*}\right)\right\}}{(\sigma^2)^{\frac{n+k+c-3}{2}+1} S_{xx}^*}.$$

We see now that, conditional on x_0, σ^2 has an Inverse Gamma distribution:

$$\sigma^2|\mathbf{y}, x_0 \sim \mathcal{I}.\mathcal{G}.\left(\frac{n+k+c-3}{2}, 2\left(S_{yy}^* - \frac{S_{xy}^{*2}}{S_{xx}^*}\right)^{-1}\right).$$

Hence, the marginal posterior density of x_0 is

$$\pi_k^R(x_0|\mathbf{y}) \propto \left(S_{yy}^* - \frac{S_{xy}^{*2}}{S_{xx}^*}\right)^{-\frac{n+k+c-3}{2}} \frac{1}{S_{xx}^*} \tag{3.4}$$

$$= \frac{(S_{xx}^*)^{\frac{n+c+k-5}{2}}}{(S_{xx}^* S_{yy}^* - S_{xy}^{*2})^{\frac{n+c+k-3}{2}}}. \tag{3.5}$$

3.4.2 Existence of Moments of $\pi_k^R(x_0|\mathbf{y})$

We now consider the moments of the marginal posterior distribution of x_0, corresponding to the reference priors. Note that S_{xx}^* may be written in the form

$$S_{xx}^* = d_1 + d_2(x_0 - \bar{x})^2,$$

where d_1 and d_2 are quantities which do not depend on x_0. Similarly, it is easily shown that

$$S_{xy}^* = d_3 + d_4(x_0 - \bar{x}),$$

where d_3 and d_4 are quantities which do not depend on x_0. By inspection of equation (3.5), we see that in the numerator of the marginal posterior of x_0, the term $(x_0 - \bar{x})$ has order $n + c + k - 5$, while in the denominator this term has order $n + c + k - 3$. This posterior density is therefore proper, but does not have finite mean or higher order moments.

It should be noted that Hoadley (1970) stated that a proper prior for x_0 is necessary in order for the marginal posterior for x_0 to have a finite mean. However, such a proper prior distribution is not necessary for the posterior to be integrable. The reference priors are not proper, yet the reference posteriors are proper.

3.4.3 Gibbs Sampling of the Reference Posteriors

The marginal posterior distribution for x_0 was derived, up to a normalization constant, in the previous section. However, as one might imagine, inferences about the other model parameters might also be desired. We can study the posterior distributions of the parameters x_0, α, β, and σ by using the Gibbs sampler. The Gibbs sampler is a special case of a Markov Chain Monte Carlo (MCMC) method for generating simulated observations from a multidimensional probability distribution. The first use of the Gibbs sampler for studying the posterior distribution from a Bayesian analysis was by Gelfand and Smith (1990). Gilks, Richardson, and Spiegelhalter (1996) present a general discussion of many aspects of MCMC methodology. Here, we only point out that our purpose in using the Gibbs sampler is to study the posterior distribution for the model parameters $(x_0, \alpha, \beta, \sigma)$. Since it is difficult to study this four dimensional distribution analytically, we instead use the Gibbs sampler to generate a sample from this distribution. The properties of the distribution (e.g., marginal medians, marginal probability density functions, posterior correlations among parameters, etc.) can be evaluated indirectly by studying the simulated data set.

In order to implement the Gibbs Sampler, we need the conditional distributions for the parameters. These are straightforward to calculate. It is convenient to transform the parameter σ to σ^2. (Note that these conditional distributions were presented, in a slightly different

parameterization of the calibration problem, in Kubokawa and Robert (1994).) The full conditional distributions for α, β, and σ^2 are

$$\alpha|\mathbf{y}, x_0, \beta, \sigma \sim N\left(\frac{\mathfrak{s}^*_{(y-\beta x)}}{n+c}, \frac{\sigma^2}{n+c}\right);$$

$$\beta|\mathbf{y}, x_0, \alpha, \sigma \sim N\left(\frac{\mathfrak{s}^*_{x(y-\alpha)}}{\mathfrak{s}^*_{xx}}, \frac{\sigma^2}{\mathfrak{s}^*_{xx}}\right);$$

$$\sigma^2|\mathbf{y}, x_0, \alpha, \beta \sim \mathcal{I}.\mathcal{G}.\left(\frac{n+k+c-1}{2}, \frac{2}{\mathfrak{s}^*_{(y-\alpha-\beta x)^2}}\right).$$

Note that the third distribution may be replaced with the following:

$$\frac{\mathfrak{s}^*_{(y-\alpha-\beta x)^2}}{\sigma^2} \sim \chi^2_{(n+k+c-1)},$$

where $\chi^2_{(n+k+c-1)}$ represents the chi-squared distribution with $n+k+c-1$ degrees of freedom. This distribution is more intuitively appealing than the inverse gamma distribution, and is also easier to simulate.

The conditional density of x_0 is

$$\pi^R_k(x_0|\mathbf{y}, \alpha, \beta, \sigma) \propto \frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=n+1}^{n+c}(y_i - \alpha - \beta x_0)^2\right\}}{\sqrt{S_{xx} + \frac{cn}{n+c}(x_0 - \bar{x})^2}}.$$

Note that the term multiplying the exponential can be bounded above:

$$\frac{1}{\sqrt{S_{xx} + \frac{cn}{n+c}(x_0 - \bar{x})^2}} \leq \frac{1}{\sqrt{S_{xx}}}.$$

Hence, we can use the normal distribution

$$N\left(\frac{1}{\beta}(\bar{y}_0 - \alpha), \frac{\sigma^2}{c\beta^2}\right)$$

as a proposal distribution and simulate x_0 using the rejection method (see, for example, Rubinstein (1981), p. 45–50). Here $\bar{y}_0$ is the mean of the responses corresponding to the unknown x_0.

With the above conditional distributions, the Gibbs sampler can be used to derive posterior inferences for all of the model parameters.

3.5 Gibbs Sampling of the Jeffreys' Posterior

In this section, the conditional posterior distributions of x_0, α, β, and σ, corresponding to the Jeffreys' prior, are derived. These may be used to implement the Gibbs sampler for the study of the Jeffreys' distribution.

The conditional posterior distributions of α and σ corresponding to the Jeffreys' prior are similar to those presented in section 3.4.3 (for the reference priors):

$$\alpha|\mathbf{y}, x_0, \beta, \sigma \sim N\left(\frac{\mathfrak{S}^*_{(y-\beta x)}}{n+c}, \frac{\sigma^2}{n+c}\right);$$

$$\sigma^2|\mathbf{y}, x_0, \alpha, \beta \sim \mathcal{I}.\mathcal{G}.\left(\frac{n+c+3}{2}, \frac{2}{\mathfrak{S}^*_{(y-\alpha-\beta x)^2}}\right).$$

Again, it is easier to implement simulation algorithms if the third distribution may be replaced with the following:

$$\frac{\mathfrak{S}^*_{(y-\alpha-\beta x)^2}}{\sigma^2} \sim \chi^2_{(n+c+3)},$$

where $\chi^2_{(n+c+3)}$ represents the chi-squared distribution with $n+c+3$ degrees of freedom. The conditional posterior distribution of x_0 is

$$x_0|\mathbf{y}, \alpha, \beta, \sigma \sim N\left(\frac{1}{\beta}(\bar{y}_0 - \alpha), \frac{\sigma^2}{c\beta^2}\right),$$

where $\bar{y}_0$ is the mean of the responses observed at the second stage of experimentation.

The conditional posterior distribution of β corresponding to the Jeffreys' prior is not of a common form. Specifically,

$$\pi(\beta|\mathbf{y}, x_0, \alpha, \sigma) \propto |\beta| \exp\left\{-\frac{1}{2}\sum_{i=1}^{n+c}(y_i - \alpha - \beta x_i)^2\right\}. \tag{3.6}$$

If we complete the square in β, let

$$a = \frac{\sigma^2}{\mathfrak{S}^*_{xx}} \quad \text{and} \quad b = \frac{\mathfrak{S}^*_{x(y-\alpha)}}{\mathfrak{S}^*_{xx}},$$

and make the change of variables

$$v = \beta - b,$$

then the unnormalized density function of v is

$$\pi(v|\mathbf{y}, x_0, \alpha, \sigma) = |v + b| \exp\left\{-\frac{1}{2a}v^2\right\}.$$

By the triangle inequality, then,

$$\pi(v|\mathbf{y}, x_0, \alpha, \sigma) \leq |v| \exp\left\{-\frac{1}{2a}v^2\right\} + |b| \exp\left\{-\frac{1}{2a}v^2\right\}. \tag{3.7}$$

The first term in (3.7) has the form of a symmetric Weibull density with parameters 2 and $\sqrt{2a}$, apart from the normalizing constant. This normalizing constant is $1/(2a)$. The second term in (3.7) has the form of a normal density with parameters 0 and a, apart from the normalizing constant. In this case, the normalizing constant is $1/(|b|\sqrt{2\pi a})$. If we set $C = 2a + |b|\sqrt{2\pi a}$, then

$$\begin{aligned}
\frac{1}{C}\pi(v|\mathbf{y}, x_0, \alpha, \sigma) &= \frac{1}{2a + |b|\sqrt{2\pi a}}|v + b| \exp\left\{-\frac{1}{2a}v^2\right\} \\
&\leq \frac{1}{2a + |b|\sqrt{2\pi a}}\left(|v| \exp\left\{-\frac{1}{2a}v^2\right\} + |b| \exp\left\{-\frac{1}{2a}v^2\right\}\right) \\
&= w_1 f_1(v) + w_2 f_2(v),
\end{aligned}$$

where $f_1(v)$ is the proper symmetric Weibull density with parameters 2 and $\sqrt{2a}$, $f_2(v)$ is the proper normal density with parameters 0 and a, and w_1 and w_2 are weights which sum to 1:

$$w_1 = \frac{2a}{C} \quad \text{and} \quad w_2 = \frac{|b|\sqrt{2\pi a}}{C}.$$

Hence,

$$\pi(v|\mathbf{y}, x_0, \alpha, \sigma) \leq C(w_1 f_1(v) + w_2 f_2(v)). \tag{3.8}$$

The right hand side of (3.8), apart from the constant multiple C, is a proper density, comprised of a mixture of the two densities $f_1(v)$ and $f_2(v)$. Furthermore, these two densities are of common forms. It is therefore possible to indirectly simulate β from (3.6), by generating random deviates of v via the rejection method. The proposal distribution for the rejection method is the mixture of the symmetric Weibull and normal distributions with densities $f_1(v)$ and $f_2(v)$.

3.6 An Example of Univariate Linear Calibration

An example of univariate linear calibration is presented by Aitchison and Dunsmore (1975), p. 183. The regressor variable of interest is the concentration of a certain enzyme in blood

plasma samples (measured in meq/l). The enzyme concentration can be measured precisely by a laboratory method that is time consuming and expensive. An alternative method of determining the concentration is via an autoanalyzer – an instrument that is quick and inexpensive to operate, but is not as precise as the laboratory method. Of interest is the calibration of this instrument.

A first stage of experimentation was carried out in which blood plasma samples with enzyme concentrations at nine levels were studied. These levels were selected by researchers performing the experiment. At each level, three blood plasma samples were assessed with the autoanalyzer. The resulting data are shown in table 3.1 (from Aitchison and Dunsmore (1975), p. 184).

Table 3.1: *Data for the plasma example from Aitchison and Dunsmore (1975). The x variable is the laboratory determination of the enzyme concentration in the blood plasma samples. The y variable is the assessment of the enzyme concentration as measured by the autoanalyzer. Note that for this dataset $n = 27$.*

Laboratory determination (x)	Autoanalyzer determinations (y)
3.0	2.3, 2.4, 2.5
3.4	2.6, 2.8, 2.8
3.8	3.0, 3.0, 3.1
4.2	3.2, 3.3, 3.4
4.6	3.7, 3.7, 3.7
5.0	3.9, 4.0, 4.1
5.4	4.2, 4.2, 4.3
5.8	4.6, 4.7, 4.8
6.2	4.9, 5.0, 5.2

The data in table 3.1 are displayed in the scatterplot in figure 3.1. It should be noted that some data points are obscured in this plot, due to replication. From the scatterplot, it appears that a simple linear regression model will be appropriate for relating x and y.

At a second stage of experimentation, the autoanalyzer instrument would be used to assess the (unknown) concentration of the enzyme in a blood plasma sample from a patient. Of interest is a determination of the true concentration of the enzyme in the sample.

3.6.1 Marginal Inferences About x_0

In this section, we will only consider inferences about the parameter of greatest inferential importance, namely x_0. We will perform four types of analysis for this problem. First, inferences about the unknown enzyme concentration will be made via the classical approach (see section 1.2.1). Second, Hoadley's prior will be used in a Bayesian analysis of the problem.

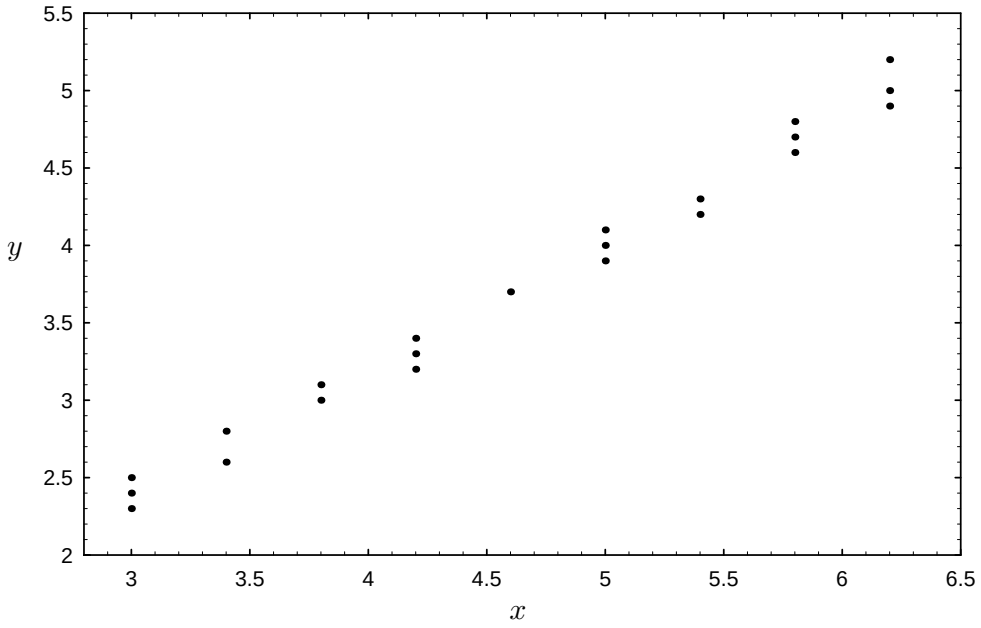

Figure 3.1: *Scatterplot of the data from table 3.1. Note that some data points are obscured due to replication. A simple linear regression model appears to be appropriate for relating x and y.*

The Bayesian analysis will then be repeated with two types of noninformative priors, namely the reference priors and the Jeffreys' prior. For each of these approaches, we will assume that the second stage observation from the autoanalyzer is $y_0 = 3.7$, then we will repeat the analyses assuming that $y_0 = 5.2$. Note that the former value is centrally located in the range of the responses observed at the first stage of experimentation, while the latter is at the upper extreme of the responses observed at the first stage of experimentation.

Some summary statistics for this dataset are provided below. The notations are from section 1.2.1.

$$\hat{\alpha} = -0.0624, \quad \hat{\beta} = 0.8139, \quad \hat{\sigma} = 0.0945, \quad S_{xx} = 28.8.$$

In table 3.2, 90% credible sets for x_0, based on Hoadley's prior, the reference priors, and the Jeffreys' prior, are presented. For the Bayesian analyses based on Hoadley's prior and the reference priors, the estimates are posterior modes, and the intervals are based on the 0.05 and 0.95 posterior quantiles, computed analytically. For the Bayesian analysis based on Jeffreys' prior, the estimate and intervals are based on the 0.50, 0.05, and 0.95 quantiles of samples consisting of ten thousand iterations of the Gibbs sampler. Also included in the table is the classical estimator for x_0, and the 90% fiducial interval, which was computed as described in equation (1.1). The results are included for observed $y_0 = 3.7$ and $y_0 = 5.2$.

Table 3.2: *Estimates and 90% intervals for x_0 in the plasma example. The observed values measured by the autoanalyzer are assumed to be $y_0 = 3.7$ and $y_0 = 5.2$.*

	Results for $y_0 = 3.7$			Results for $y_0 = 5.2$		
	Lower	Estimate	Upper	Lower	Estimate	Upper
Hoadley's prior	4.4217	4.6225	4.8233	6.2322	6.4442	6.6562
Reference prior ($k = 1$)	4.4207	4.6227	4.8249	6.2547	6.4640	6.6819
Reference prior ($k = 2$)	4.4249	4.6227	4.8210	6.2591	6.4641	6.6773
Reference prior ($k = 3$)	4.4288	4.6227	4.8167	6.2632	6.4642	6.6562
Jeffreys' prior	4.430	4.621	4.822	6.246	6.464	6.685
Classical Approach	4.3534	4.6227	4.9047	6.1286	6.4657	6.8206

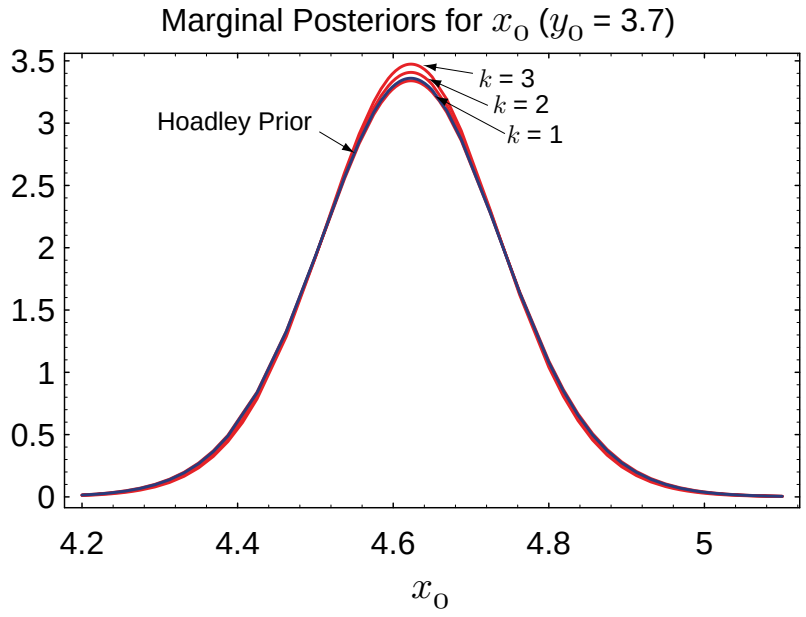

Figure 3.2: *Marginal posterior distributions for x_0 corresponding to the reference priors and Hoadley's prior. The response from the second stage of experimentation is assumed to be $y_0 = 3.7$. Note that this is close to the mean of the responses from the first stage of experimentation. In this case, the result from Hoadley's prior is nearly identical to that from the reference prior with $k = 1$. In the plot, these two resulting posterior densities are nearly indistinguishable.*

The marginal posterior distributions for x_0 on which the results in table 3.2 are based are displayed in figures 3.2 and 3.3. When $y_0 = 3.7$, Hoadley's prior and the reference priors lead to similar inferences about x_0. However, when $y_0 = 5.2$, the marginal posterior distribution for x_0 corresponding to Hoadley's prior is centered closer to the mean of the x values from the first stage of experimentation than are those corresponding to the reference priors. From table 3.2, it can be seen that the results of the classical procedure are most similar to the results for the reference priors, but the intervals resulting from the classical procedure are slightly wider than those based on the reference posteriors.

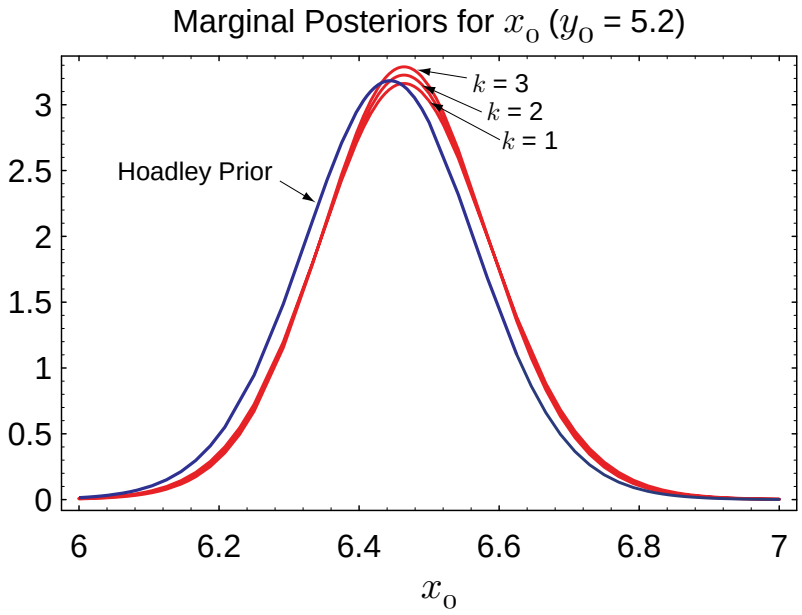

Figure 3.3: *Marginal posterior distributions for x_0 corresponding to the reference priors and Hoadley's prior. The response from the second stage of experimentation is assumed to be $y_0 = 5.2$. Note that this is at the upper extreme of the responses from the first stage of experimentation.*

In chapter 4, Hoadley's prior and the reference priors will be evaluated and compared on the basis of the frequentist coverage properties of posterior inferences about x_0.

3.6.2 Reference Posterior Inferences About All Model Parameters

The results displayed in figures 3.2 and 3.3 were computed analytically. The marginal posterior distribution for x_0 corresponding to Hoadley's prior has the form of a Student's t distribution, as described by Hoadley (1970). The marginal posterior distributions for x_0 corresponding to the reference priors were computed from normalized versions of equation (3.5). (The normalization was done via numerical integration.) It may be of interest for a researcher to develop inferences about the other model parameters as well. For the reference prior with $k = 1$, this was done by implementing the Gibbs sampler, as described in section 3.4.3.

After an initial burn-in period, five thousand iterations of the Gibbs sampler were used to estimate the marginal posterior distributions for α, β, σ, and x_0. This was done for the values $y_0 = 3.7$ and $y_0 = 5.2$. In order to improve the mixing of the Gibbs sampler, the regressor values from the first stage of experimentation were centered about their mean of

$\bar{x} = 4.6$. (This caused the parameters α and β to be uncorrelated.) Figure 3.4 displays the trace plot for x_0 corresponding to $y_0 = 3.7$, after the burn-in period. Similar plots for the other model parameters were studied; they looked similar to that for x_0, and are not presented here.

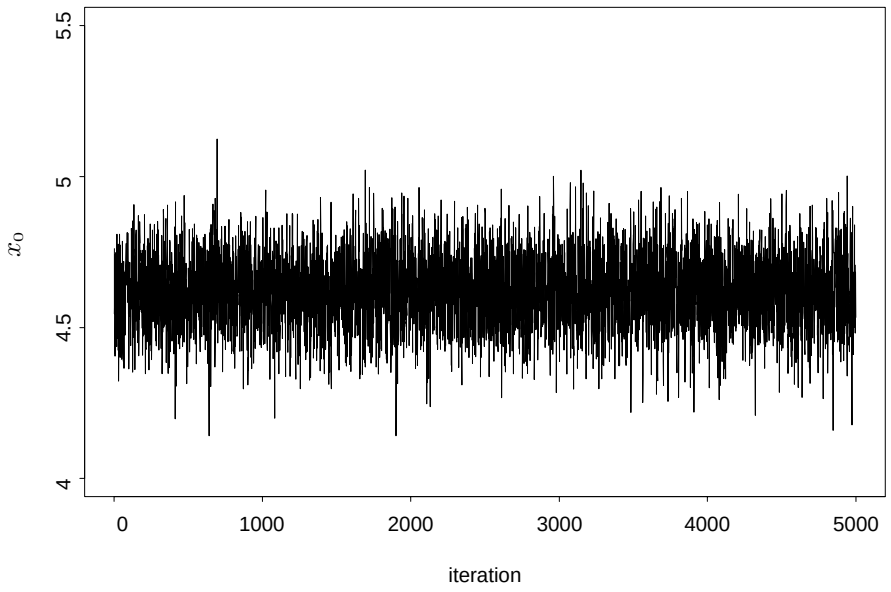

Figure 3.4: *Trace of the output of the Gibbs sampler for x_0, corresponding to $y_0 = 3.7$. On the vertical axis, the simulated values from the marginal posterior distribution of x_0 are displayed; the horizontal axis displays the iteration number in the chain. It appears from this plot that the chain has converged to its stationary distribution, and is mixing well. Similar results were found for other model parameters.*

In figures 3.5 and 3.6, the estimated posterior densities for all model parameters are displayed. Figure 3.5 corresponds to the case $y_0 = 3.7$, and figure 3.6 corresponds to the case $y_0 = 5.2$. The densities were estimated with a kernel density estimator, implemented by the CODA software package [1].

[1]CODA is a freely distributed software package, which performs output analysis and convergence diagnostics for Markov Chain Monte Carlo simulations. We used version 0.3 of the software, written by N. Best (MRC Biostatistics Unit, Instituite of Public Health, Tobinson Way, Cambridge CB2 2SR, UK), M. K. Cowles (University of Nebraska Medical Center), and K. Vines (also of MRC Biostatistics Unit). As of the date of this work, the software and documentation can be obtained from the World Wide Web, at http://www.mrc-bsu.cam.ac.uk/bugs/software/software.html.

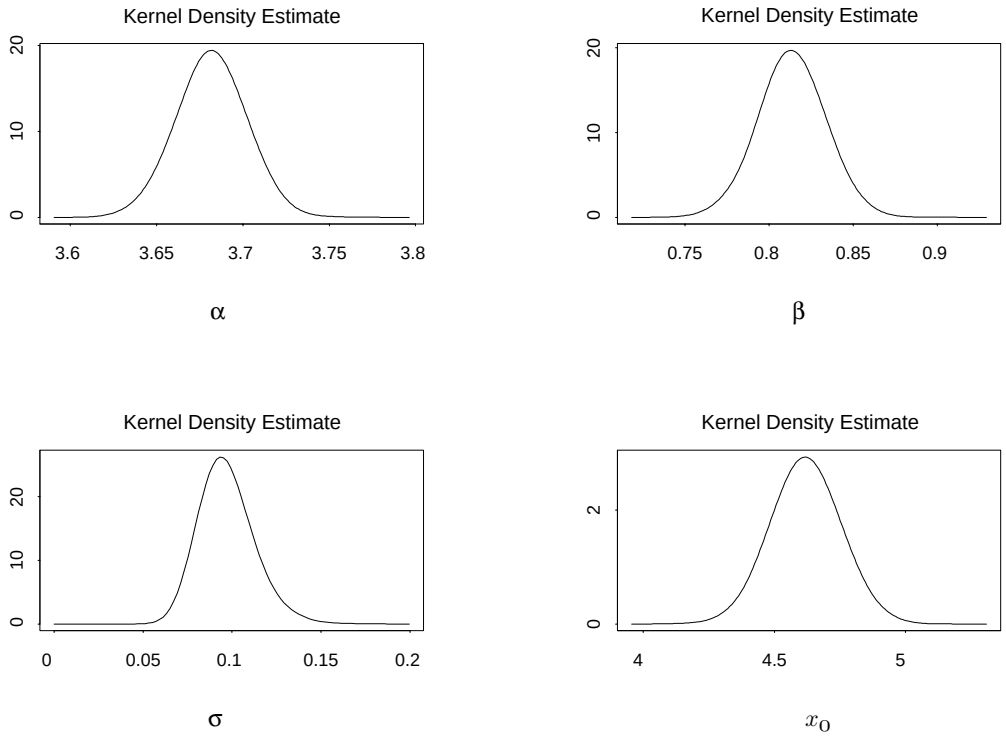

Figure 3.5: *Marginal kernel density estimates for α, β, σ, and x_0, corresponding to $y_0 = 3.7$. Note that the parameter α is the intercept for the centered model (regressors from the first stage of experimentation centered about their mean of $\bar{x} = 4.6$).*

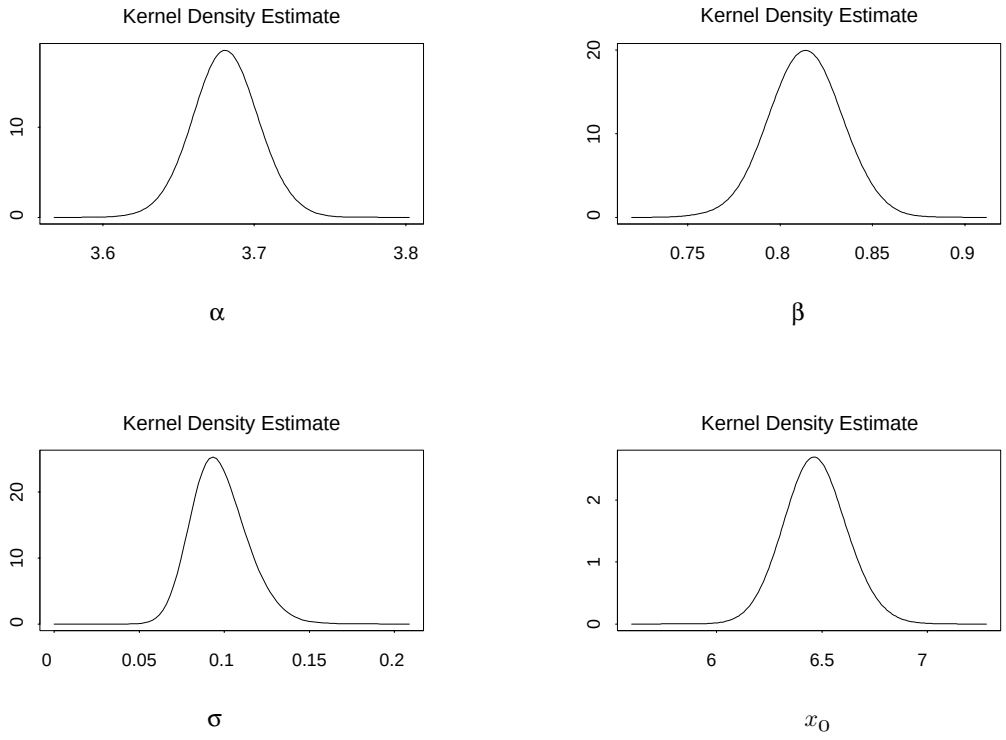

Figure 3.6: *Marginal kernel density estimates for α, β, σ, and x_0, corresponding to $y_0 = 5.2$. Note that the parameter α is the intercept for the centered model (regressors from the first stage of experimentation centered about their mean of $\bar{x} = 4.6$).*

Chapter 4

Probability Matching Priors for the Univariate Linear Calibration Problem with Constant Variance

4.1 Theoretical Development of Probability Matching Priors

In this section, we apply the method described by Tibshirani (1989) to the univariate linear calibration problem. Recall that this method relies on the use of an orthogonal parameterization of the model (the term "orthogonal" is interpreted as described in Cox and Reid (1987)). In particular, we consider x_0 to be the parameter of interest, so we desire a prior which will generate posterior quantiles for this parameter with the correct coverage probability, asymptotically. We write the information matrix as

$$\mathcal{I} = \frac{1}{\sigma^2}\begin{bmatrix} \mathcal{I}_1 & \mathbf{0} \\ \mathbf{0} & 2(n+c) \end{bmatrix},$$

where

$$\mathcal{I}_1 = \begin{bmatrix} c\beta^2 & c\beta & c\beta x_0 \\ c\beta & n+c & \mathfrak{s}_x^* \\ c\beta x_0 & \mathfrak{s}_x^* & \mathfrak{s}_{xx}^* \end{bmatrix}.$$

Note that the parameter σ is already orthogonal x_0. Hence, we need only reparameterize α and β. The prior distribution will then be taken to be proportional to square root of the diagonal element of the information matrix corresponding to x_0, which may be multiplied by an arbitrary function of the other (orthogonal) parameters.

As noted previously, an orthogonal parameterization can be found by solving a system of

partial differential equations. Here, we explicitly derive these equations for the univariate linear calibration problem. The following lemma, whose proof can be found in the text by Bernardo and Smith (1994), will be used.

LEMMA 1 *Consider a model with likelihood $\mathcal{L}$, expressed in terms of a parameter vector $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_l)'$; let $\mathcal{I}_{\boldsymbol{\gamma}}$ be the Fisher information matrix for this model. If the parameters are transformed via a nonsingular transformation to $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_l)'$, then the Fisher information in the new parameterization is*

$$\mathcal{I}_{\boldsymbol{\eta}} = J'\mathcal{I}_{\boldsymbol{\gamma}}J,$$

where

$$J = \left(\frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\eta}}\right)'.$$

For the current problem, we seek a nonsingular transformation of the parameter vector (x_0, α, β) to a new parameter vector $(x_0, \lambda_1, \lambda_2)$ satisfying

$$\mathcal{I}_1^* = J'\mathcal{I}_1 J = \begin{bmatrix} \mathcal{I}_{1,1} & \mathbf{0}' \\ \mathbf{0} & \mathcal{I}_{2,2} \end{bmatrix},$$

where $\mathcal{I}_1^*$ is the information matrix in the parameterization $(x_0, \lambda_1, \lambda_2)$, and J is as in the lemma. Now

$$J = \begin{bmatrix} 1 & 0 & 0 \\ \frac{\partial \alpha}{\partial x_0} & \frac{\partial \alpha}{\partial \lambda_1} & \frac{\partial \alpha}{\partial \lambda_2} \\ \frac{\partial \beta}{\partial x_0} & \frac{\partial \beta}{\partial \lambda_1} & \frac{\partial \beta}{\partial \lambda_2} \end{bmatrix},$$

so the orthogonality condition holds if

$$(J'\mathcal{I}_1 J)_{(1,2)} = 0 \qquad \text{and} \qquad (J'\mathcal{I}_1 J)_{(1,3)} = 0.$$

That is, we require

$$\frac{\partial \alpha}{\partial \lambda_1}\left(c\beta + (n+c)\frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_x^*\frac{\partial \beta}{\partial x_0}\right) + \frac{\partial \beta}{\partial \lambda_1}\left(c\beta x_0 + \mathfrak{s}_x^*\frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_{xx}^*\frac{\partial \beta}{\partial x_0}\right) = 0$$

and

$$\frac{\partial \alpha}{\partial \lambda_2}\left(c\beta + (n+c)\frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_x^*\frac{\partial \beta}{\partial x_0}\right) + \frac{\partial \beta}{\partial \lambda_2}\left(c\beta x_0 + \mathfrak{s}_x^*\frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_{xx}^*\frac{\partial \beta}{\partial x_0}\right) = 0.$$

For the transformation from (x_0, α, β) to $(x_0, \lambda_1, \lambda_2)$ to be nonsingular, the lower right two by two block of J must be nonsingular. Hence, the following system of equations must be satisfied:

$$(n+c)\frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_x^*\frac{\partial \beta}{\partial x_0} = -c\beta \tag{4.1}$$

$$\mathfrak{s}_x^* \frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_{xx}^* \frac{\partial \beta}{\partial x_0} = -c\beta x_0. \tag{4.2}$$

Note that this is the system prescribed by Cox and Reid (1987). It can be verified directly that the following are solutions to these equations:

$$\beta = \lambda_1 g_1(x_0), \qquad \text{and} \qquad \alpha = \lambda_1 g_2(x_0) + \lambda_2,$$

where

$$g_1(x_0) = \exp\left\{ \int c \frac{\mathfrak{s}_x^* - (n+c)x_0}{(n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}} \, dx_0 \right\},$$

and

$$g_2(x_0) = \int c g_1(x_0) \frac{\mathfrak{s}_{xx}^* - x_0 \mathfrak{s}_x^*}{\mathfrak{s}_x^{*2} - (n+c)\mathfrak{s}_{xx}^*} \, dx_0.$$

We now need the upper left entry of $\mathcal{I}_1^*$. This is

$$\begin{aligned}(J'\mathcal{I}_1 J)_{(1,1)} = c\left(\beta^2 + \beta \frac{\partial \alpha}{\partial x_0} + \beta x_0 \frac{\partial \beta}{\partial x_0}\right) + \frac{\partial \alpha}{\partial x_0}\left(c\beta + (n+c)\frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_x^* \frac{\partial \beta}{\partial x_0}\right) \\ + \frac{\partial \beta}{\partial x_0}\left(c\beta x_0 + \mathfrak{s}_x^* \frac{\partial \alpha}{\partial x_0} + \mathfrak{s}_{xx}^* \frac{\partial \beta}{\partial x_0}\right).\end{aligned}$$

Upon substituting in the solutions for α and β given above, and simplifying, this becomes

$$\lambda_1^2 [g_1(x_0)]^2 \frac{n\mathfrak{s}_{xx} - \mathfrak{s}_x^2}{(n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}}.$$

Note that the numerator in the fraction in the last expression does not depend on any parameters. Any prior distribution which has the form of a function $h(\lambda_1, \lambda_2)$, multiplied by the square root of that expression, will have the desired probability matching property. Now, we transform this prior back to the original parameterization (x_0, α, β). The Jacobian of this transformation is $g_1(x_0)$. Hence, in the original parameterization, our probability matching prior will be of the form

$$\begin{aligned}\pi^P(x_0, \alpha, \beta) &\propto h(\lambda_1, \lambda_2) \frac{|\lambda_1|}{\sqrt{(n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}}} \\ &= h(\lambda_1, \lambda_2) \frac{|\beta|}{g_1(x_0)\sqrt{(n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}}}.\end{aligned}$$

Note that

$$\begin{aligned}\int c \frac{\mathfrak{s}_x^* - (n+c)x_0}{(n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}} \, dx_0 &= \int -\frac{1}{2} \frac{\frac{\partial}{\partial x_0}\big((n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}\big)}{(n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}} dx_0 \\ &= -\frac{1}{2} \log\big((n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}\big),\end{aligned}$$

so

$$g_1(x_0) = \left((n+c)\mathfrak{s}^*_{xx} - \mathfrak{s}^{*2}_x\right)^{-\frac{1}{2}}.$$

If we take $h(\lambda_1, \lambda_2) = 1$, then, a prior depending on α, β, and x_0 in the following way will have the probability matching property:

$$\pi^P(x_0, \alpha, \beta) \propto |\beta|. \tag{4.3}$$

If, on the other hand, we take $h(\lambda_1, \lambda_2) = |\lambda_1|^{-1}$, then a prior of the following form will have the probability matching property:

$$\pi^P(x_0, \alpha, \beta) \propto \frac{1}{\sqrt{(n+c)\mathfrak{s}^*_{xx} - \mathfrak{s}^{*2}_x}} \propto \frac{1}{\sqrt{S^*_{xx}}}. \tag{4.4}$$

Of course, any function of the parameter σ can be multiplied by either (4.3) or (4.4) without invalidating the asymptotic probability matching property, since σ is orthogonal to x_0. Thus, the Jeffreys' prior can be seen to have the probability matching property, by multiplying (4.3) by σ^{-4}. Now

$$\begin{aligned} S^*_{xx} &= \mathfrak{s}^*_{xx} - \frac{1}{n+c}\mathfrak{s}^{*2}_x \\ &= \mathfrak{s}_{xx} + cx_o^2 + \frac{1}{n+c}(\mathfrak{s}_x + cx_o^2) \\ &= \mathfrak{s}_{xx} + \frac{nc}{n+c}(x_0^2 - \bar{x}^2) + (n+c)\bar{x}^2. \end{aligned}$$

After completing the square in x_0 and collecting terms, this becomes

$$S^*_{xx} = S_{xx} + \frac{nc}{n+c}(x_0 - \bar{x})^2.$$

Hence, the reference priors developed previously also correspond to probability matching priors. They are obtained from (4.4) by multiplying by σ^{-k}, for $k = 1, 2$ and 3.

4.2 Finite Sample Frequentist Coverage Properties of Reference Posterior Inferences Concerning x_0

In section 4.1, it was shown that the reference priors all correspond to probability matching priors for x_0. That is, each of the three priors will generate posterior credible sets with accurate frequentist coverage properties, *asymptotically*. Of the three reference priors, the prior with $k = 1$ has the appealing feature of also being a probability matching prior for the

parameter σ (this is easily seen by inspection of the information matrix), though estimation of σ may not be of primary importance to a researcher. A remaining issue is how the three reference priors perform for small sample sizes. To assess the small sample performance of the reference priors, a simulation study was undertaken. For relatively small sample sizes, the frequentist coverage probabilities were estimated from simulated datasets. The results of this study are reported in this section.

Before proceeding to the results of the simulation study, a useful result is presented and proven.

4.2.1 Dependence of the Frequentist Coverage on Model Parameters

It will be shown, via the following two lemmas, that for a given experimental design (the set of prescribed values $x_1, x_2, \ldots, x_n$, and the number of observations c at the unknown x_0), the frequentist coverage of posterior quantiles corresponding to any of the reference priors depends only on x_0 and $|\beta|/\sigma$.

LEMMA 2 *Let p_γ denote the γ^{th} posterior quantile of the marginal posterior distribution for x_0, corresponding to any of the reference priors. Then p_γ is a homogeneous function $g(y_1 - \bar{y}^*, y_2 - \bar{y}^*, \ldots, y_{n+c} - \bar{y}^*)$, where $\bar{y}^*$ is the mean over all $n + c$ observations. By homogeneity, we mean that g satisfies $g(kv_1, kv_2, \ldots, kv_{n+c}) = g(v_1, v_2, \ldots, v_{n+c})$ for any constant k.*

PROOF OF LEMMA 2. From equation (3.5), it can be seen that the marginal posterior distribution of x_0 depends on the observed data only through the terms S^*_{xy} and S^*_{yy}. The homogeneity can be seen from the fact that replacing each $y_i - \bar{y}^*$ with $k(y_i - \bar{y}^*)$ in these sums of squares and cross products alters the marginal posterior only by the multiplicative constant $k^{-(n+c+k-3)}$. This term would also appear in the normalizing constant for the density, and would therefore be canceled.

LEMMA 3 *Suppose that g is a homogeneous function. Let*

$$P(g) = P\{x_0 < g(y_1 - \bar{y}^*, y_2 - \bar{y}^*, \ldots, y_{n+c} - \bar{y}^*)\},$$

for fixed values of the parameters x_0, α, β, and σ, and a fixed experimental design. Then $P(g)$ depends only on x_0 and $|\beta|/\sigma$.

PROOF OF LEMMA 3. The quantity of interest, $P(g)$, is defined by

$$P(g) = \int \cdots \int \frac{1}{(2\pi)^{\frac{n+c}{2}}\sigma^{n+c}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c}(y_i - \alpha - \beta x_i)^2\right\} dy_1 dy_2 \cdots dy_{n+c}, \tag{4.5}$$

where the integral is over all $(y_1, y_2, \ldots, y_{n+c})$ such that $x_0 < g(y_1, y_2, \ldots, y_{n+c})$. The integrand of equation (4.5) can be re-expressed as

$$\frac{1}{(2\pi)^{\frac{n+c}{2}}\sigma^{n+c}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n+c}\left(\frac{y_i - \bar{y}^*}{\sigma} - \frac{\beta}{\sigma}(x_i - \bar{x}^*)\right)^2 - \frac{1}{2}\sum_{i=1}^{n+c}\left(\frac{\bar{y}^*}{\sigma} - \frac{\alpha}{\sigma} - \frac{\beta}{\sigma}\bar{x}^*\right)^2\right\},$$

where $\bar{x}^*$ is the mean of the regressor values from both stages of experimentation. We make the following transformation:

$$u_1 = \frac{\bar{y}^*}{\sigma}, \qquad u_i = \operatorname{sgn}(\beta)\frac{y_i - \bar{y}^*}{\sigma} \quad \text{for} \quad i = 2, 3, \ldots, n+c.$$

The Jacobian of this transformation is $J = K_1\sigma^{-(n+c)}$, where K_1 is a constant. Note that

$$\begin{aligned}\frac{y_1 - \bar{y}^*}{\sigma} &= -\frac{1}{\sigma}\sum_{i=2}^{n+c}(y_i - \bar{y}^*) \\ &= -\operatorname{sgn}(\beta)\sum_{i=2}^{n+c} u_i.\end{aligned}$$

The integrand in equation (4.5) thus becomes

$$\begin{aligned}\frac{1}{(2\pi)^{\frac{n+c}{2}}K_1} &\exp\left\{-\frac{1}{2}\left(-\operatorname{sgn}(\beta)\sum_{i=2}^{n+c} u_i - \frac{\beta}{\sigma}(x_1 - \bar{x}^*)\right)^2\right\} \\ &\cdot \exp\left\{-\frac{1}{2}\sum_{i=2}^{n+c}\left(\operatorname{sgn}(\beta)u_i - \frac{\beta}{\sigma}(x_i - \bar{x}^*)\right)^2\right\} \\ &\cdot \exp\left\{-\frac{1}{2}\left(u_1 - \frac{\alpha}{\sigma} - \frac{\beta}{\sigma}\bar{x}^*\right)^2\right\}.\end{aligned}$$

Integrating out u_1 now yields

$$\begin{aligned}&K_2 \exp\left\{-\frac{1}{2}\left(-\operatorname{sgn}(\beta)\sum_{i=2}^{n+c} u_i - \frac{\beta}{\sigma}(x_1 - \bar{x}^*)\right)^2\right\} \exp\left\{-\frac{1}{2}\sum_{i=2}^{n+c}\left(\operatorname{sgn}(\beta)u_i - \frac{\beta}{\sigma}(x_i - \bar{x}^*)\right)^2\right\} \\ &= K_2 \exp\left\{-\frac{1}{2}\left(\sum_{i=2}^{n+c} u_i + \frac{|\beta|}{\sigma}(x_1 - \bar{x}^*)\right)^2\right\} \exp\left\{-\frac{1}{2}\sum_{i=2}^{n+c}\left(u_i - \frac{|\beta|}{\sigma}(x_i - \bar{x}^*)\right)^2\right\},\end{aligned}$$

as the remaining integrand, where K_2 is a constant. Finally, note that

$$\begin{aligned} & g(y_1 - \bar{y}^*, y_2 - \bar{y}^*, \ldots, y_{n+c} - \bar{y}^*) \\ & = g\left(\frac{\text{sgn}(\beta)}{\sigma}(y_1 - \bar{y}^*), \frac{\text{sgn}(\beta)}{\sigma}(y_2 - \bar{y}^*), \ldots, \frac{\text{sgn}(\beta)}{\sigma}(y_{n+c} - \bar{y}^*)\right) \\ & = g\left(-\sum_{i=2}^{n+c} u_i, u_2, \ldots, u_{n+c}\right), \end{aligned}$$

by homogeneity. Hence, for a given design, $P(g)$ depends only on x_0 and $|\beta|/\sigma$.

4.2.2 Simulation Results for the Reference Priors

To assess the small sample frequentist coverage properties of the reference priors, a simulation study was undertaken. Datasets were randomly generated according the model displayed in equations (3.1) and (3.2). Initially, sample sizes of $n = 3$, $n = 9$, and $n = 15$ were used, in each case with $c = 1$. The x values at the first stage of simulated experimentation were uniformly spaced, covering the interval $[-1, 1]$. Four values of the slope parameter were used, namely $\beta = 0.1$, $\beta = 1$, $\beta = 5$, and $\beta = 10$, and in all cases σ was held fixed at $\sigma = 1$, and the intercept was held fixed at $\alpha = 0$. For the second stage of simulated experimentation, a new response y_0 was generated, corresponding to one of the three regressor values $x_0 = 0$, $x_0 = 0.5$, or $x_0 = 1$.

After a data set was simulated (corresponding to a single set of specified values of α, β, σ, and x_0), the marginal posterior distribution for x_0 was approximated, using the Gibbs sampler. The number of random draws of x_0 from the Gibbs sampler ranged from 10,000 to 35,000, after an initial burn-in period of 50 samples (which were discarded). These numbers were chosen by appealing to the methodology of Raftery and Lewis (1992), which provides guidelines on the necessary number of iterations of the Gibbs sampler for estimation of a specified posterior quantile to a desired degree of accuracy. Based on the resulting sample from the Gibbs sampler, the 0.05 and 0.95 posterior quantiles for x_0 were estimated (to within 0.005, with 95% confidence, as described by Raftery and Lewis (1992)), and it was determined whether or not the specified value of x_0 was less than or equal to the estimated quantile.

For each set of parameter values considered, the above procedure was repeated 8,000 times. From these repetitions, the frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles could be estimated. (Note that if the posterior quantile is denoted as q, then the frequentist coverage probability is $P(x_0 \leq q)$, where x_0 is the true value of the regressor variable from the second stage of experimentation, and the probability is taken

over the class of possible datasets that could be generated from the calibration model. This probability can be estimated with the observed proportion of simulated datasets in which x_0 is less than or equal to the observed posterior quantile. The standard error of estimation of the quantile based on the 8,000 simulated datasets can be estimated by $\sqrt{\hat{p}(1-\hat{p})/8000}$, where $\hat{p}$ is the estimated frequentist coverage probability of the posterior quantile of interest. In all cases encountered in the simulation study, this standard error was less than 0.005. For values of $\hat{p}$ near 0.05 or 0.95, this standard error was approximately 0.0025.) The results are displayed in figures 4.1, 4.2, and 4.3.

In each of the figures 4.1, 4.2, and 4.3, eight plots are displayed. The horizontal axis on each plot corresponds to the sample sizes used at the first stage of simulated experimentation, and the vertical axis corresponds to the estimated frequentist coverage probability. The left-hand column of plots displays the frequentist coverage probabilities of 0.05 posterior quantiles, and the right-hand column displays the frequentist coverage probabilities of 0.95 posterior quantiles for x_0. The four rows of plots in each figure correspond to the values of β (0.1, 1, 5, and 10, respectively). The figures 4.1, 4.2, and 4.3 correspond to the regressor values of $x_0 = 0, x_0 = 0.5$, and $x_0 = 1$, respectively. On each plot, three lines are drawn. These display the frequentist coverage probabilities for the three reference priors ($k = 1, 2$, and 3).

Several comments can be made regarding figures 4.1, 4.2, and 4.3. First, the three lines on each plot, corresponding to $k = 1, 2$ and 3, can be distinguished by the fact that in all cases, the case $k = 1$ results in the 0.05 posterior quantile with the smallest frequentist coverage probability and the 0.95 posterior quantile with the largest frequentist coverage probability. The opposite is true of the case $k = 3$. If one is interested in using posterior quantiles to build confidence intervals for x_0, then, the case $k = 1$ would yield the widest (or most conservative) intervals. For the values $\beta = 5$ and $\beta = 10$ of the slope parameter, a 90% confidence interval based on the 0.05 and 0.95 posterior quantiles corresponding to the case $k = 1$ would have accurate frequentist coverage, even when the sample size is as small as $n = 3$. As the sample size increases, all of the reference priors approach the goal of providing accurate frequentist confidence intervals.

The values $\beta = 5$ and $\beta = 10$ of the slope parameter are arguably much more realistic than the values $\beta = 0.1$ and $\beta = 1$. In the latter cases the error standard is equal to or greater than the slope parameter, a situation that is unlikely to occur in the calibration of a real instrument. However, from a theoretical point of view, it is still interesting to consider the performance of the reference priors in the small slope cases. At first glance, the coverage results for these cases seem quite peculiar. As the sample size is increasing, the frequentist coverage probabilities of the posterior quantiles seems to be getting *further* from the nominal

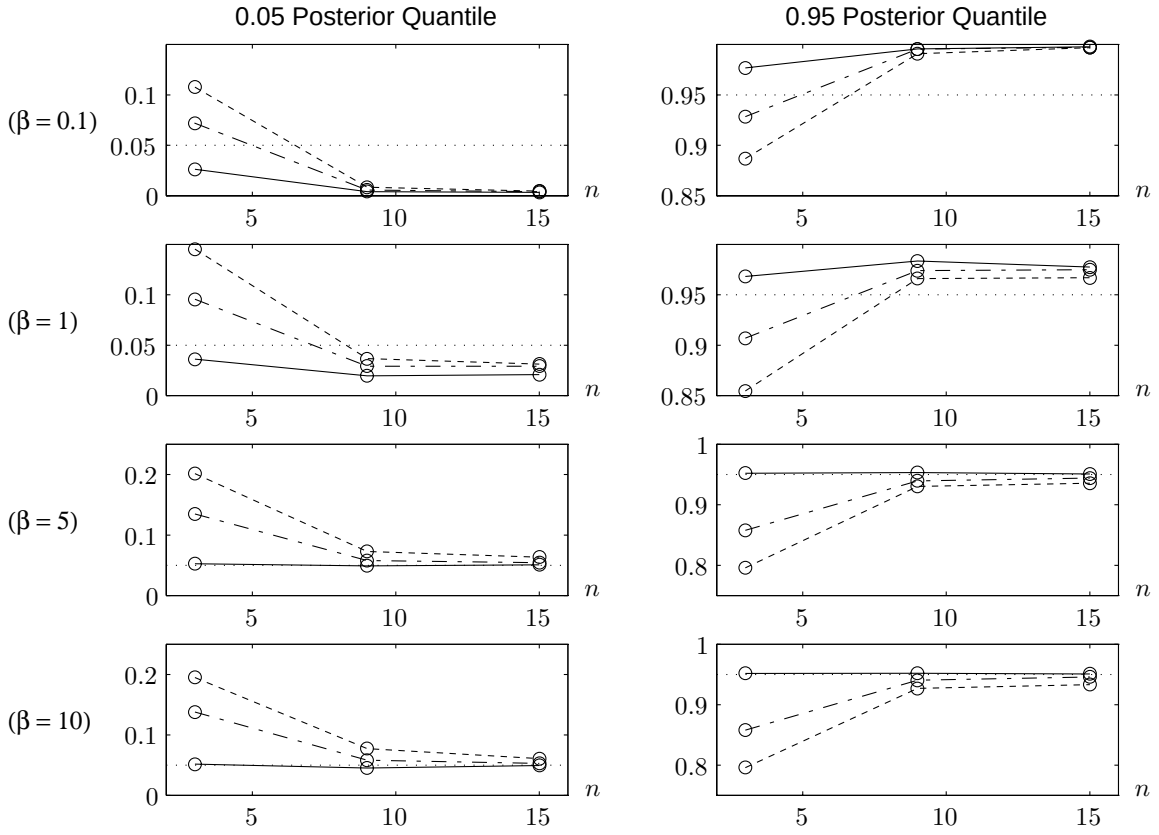

Figure 4.1: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors. The true value of x_0 is 0, and values of β considered are 0.1, 1, 5, and 10. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

levels. Of course, as the sample size increases, the opposite should be true.

To further elucidate the properties of the reference priors, the simulation study described above was extended to include sample sizes of $n = 30$ and $n = 60$ for the value $\beta = 1$ of the slope parameter. This extension was not made for other values of the slope parameter, since to do so would have required an excessive amount of computer time. The results of the full set of simulations for $\beta = 1$, including the new (larger) sample sizes, are displayed in figure 4.4. The three rows of plots in this figure correspond to the values $x_0 = 0$, $x_0 = 0.5$, and $x_0 = 1$, respectively. From these plots, it can be seen that when the sample sizes increase to 30 and 60, the frequentist coverage probabilities do, in fact, approach the nominal levels. Although the larger sample sizes were not studied for the case $\beta = 0.1$, the results shown in figure 4.4 indicate that it is likely that similar results would be found in that case, but even larger sample sizes may be necessary. Since this would require a great deal of computer time

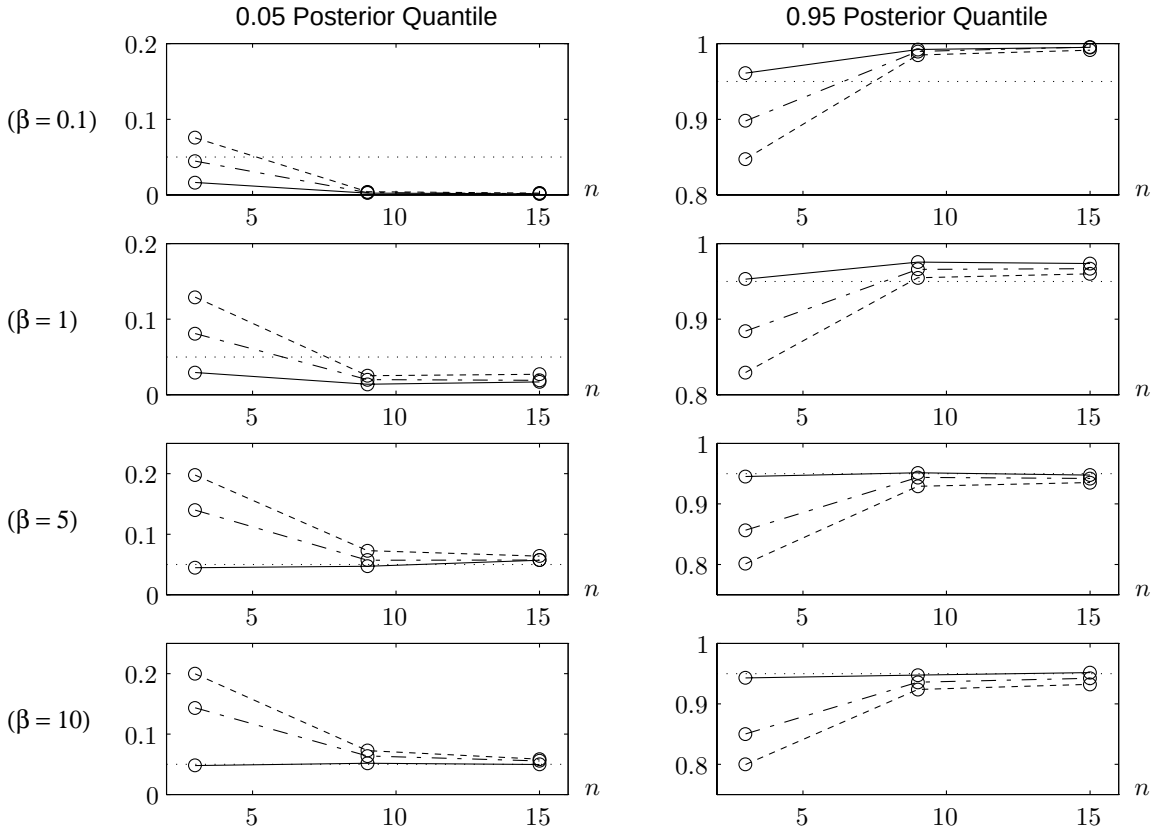

Figure 4.2: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors. The true value of x_0 is 0.5, and values of β considered are 0.1, 1, 5, and 10. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

to study, and since such a case is of little practical interest, no further results were simulated for the case $\beta = 0.1$.

A plausible explanation for the results displayed in figure 4.4 is as follows. For large sample sizes, the asymptotic probability matching properties of the reference priors become approximately correct. For moderate sample sizes (e.g., $n = 9$ and $n = 15$), the fact that the slope of the true regression line is so small relative to the error standard deviation leads the posterior distributions of x_0 to be very diffuse. Hence, an interval based on posterior quantiles would tend to be very wide, and therefore very likely to capture the true x_0 value of 0, 0.5, or 1. For the small sample size of $n = 3$, one can imagine that with non-negligible probability the three responses from the first stage of simulated experimentation would be nearly collinear. In such a case, it would appear to an analyst that the linear model fits the data very well, with a small standard deviation. The posterior inferences would thus be

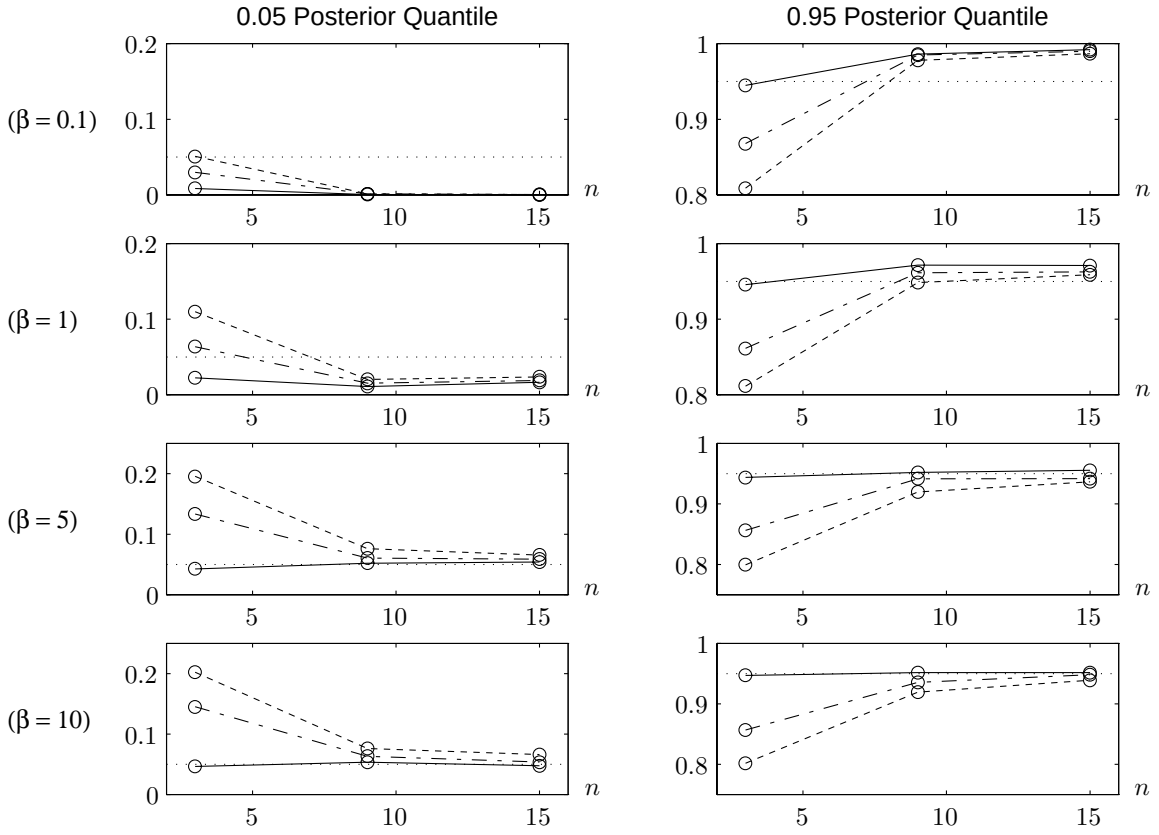

Figure 4.3: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors. The true value of x_0 is 1, and values of β considered are 0.1, 1, 5, and 10. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

unrealistically precise, and would reflect the coincidental structure in the data rather than the true underlying model. This may explain the coverage probabilities for the sample size $n = 3$, shown in figure 4.4. By way of contrast, the sample sizes $n = 9$ and $n = 15$ are large enough so that it is very unlikely that the responses from the first stage of simulated experimentation would be nearly collinear, and therefore unlikely that the analysis would cause a severe underestimation of the error standard deviation (resulting in an unrealistically narrow marginal posterior distribution for x_0).

One final point can be made about the results of the simulation study described above. The coverage properties do not seem to depend on the true value of x_0. The reference posterior distributions for x_0 would generate accurate frequentist confidence intervals for true regressor values in the middle or at the edge of the range used during the first stage of experimentation (for cases in which the magnitude of β is large compared to that of σ).

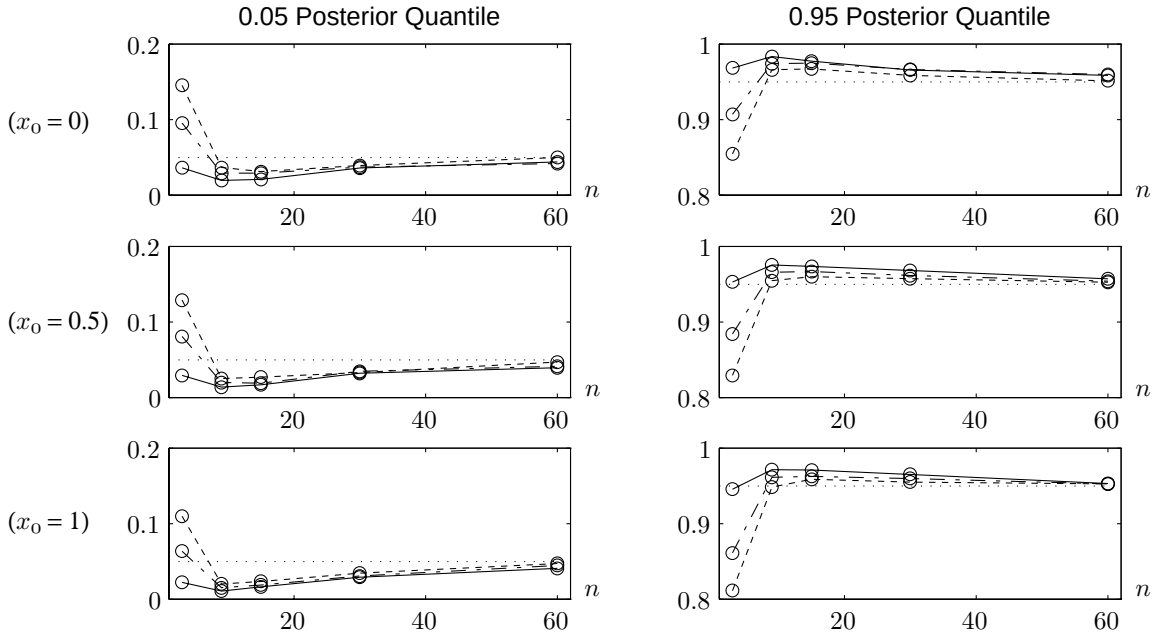

Figure 4.4: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distributions of x_0, corresponding to the reference priors. The value of β is* 1*, and the true values of x_0 are* $0, 0.5$*, and* 1*. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

4.3 Simulation Results for Hoadley's Prior

To compare the performance of the reference priors with that of Hoadley's prior, the simulation study described above was repeated, using Hoadley's prior for x_0. This case was, in fact, easier to handle, because the marginal posterior distribution for x_0 is of a common form, namely that of a Student's t distribution (Hoadley (1970)). Posterior quantiles could therefore be computed directly (without using the Gibbs sampler). The results corresponding to figures 4.1, 4.2, and 4.3, for Hoadley's prior, are shown in figures 4.5, 4.6, and 4.7.

From figures 4.5, 4.6, and 4.7, it can be seen that for the case $x_0 = 0$, Hoadley's prior leads to posterior quantiles with frequentist coverage properties comparable to those of the reference prior with $k = 1$. However, at $x_0 = 1$, the posterior quantiles are far from the nominal frequentist coverage probabilities. In particular, the 0.05 posterior quantile underestimates the 0.05 frequentist coverage point, and the 0.95 posterior quantile underestimates the 0.95 frequentist coverage point. If a 90% interval were made using the 0.05 and 0.95 posterior quantiles, it would be weighted toward the mean of the first stage regressor values (which was equal to zero in the simulation study). This fact is not surprising – Hoadley's prior puts most of the mass, *a priori*, on values of x_0 that are close to the mean of the regressor

values from the first stage of experimentation. Finally, it can be noted that as the sample size increases, the coverage properties of Hoadley's prior do not seem to improve.

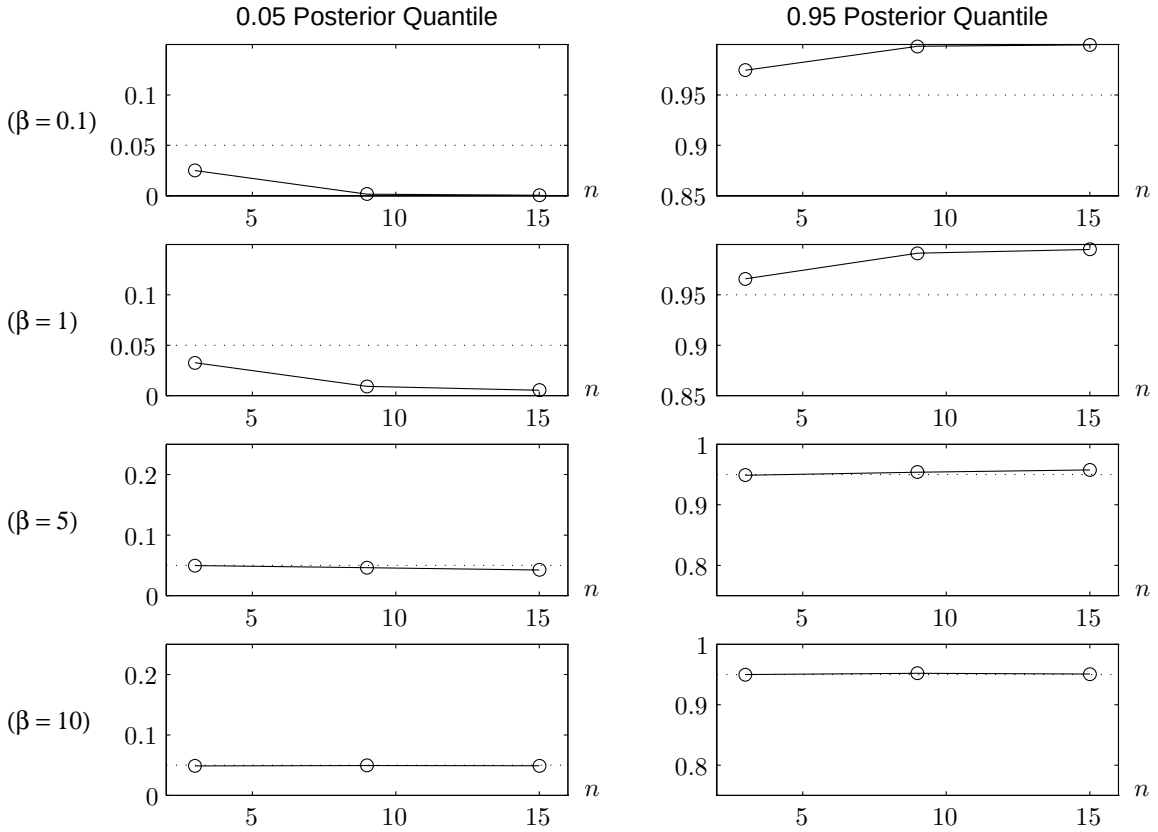

Figure 4.5: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distribution of x_0, corresponding to Hoadley's prior. The true value of x_0 is 0, and values of β considered are 0.1, 1, 5, and 10. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

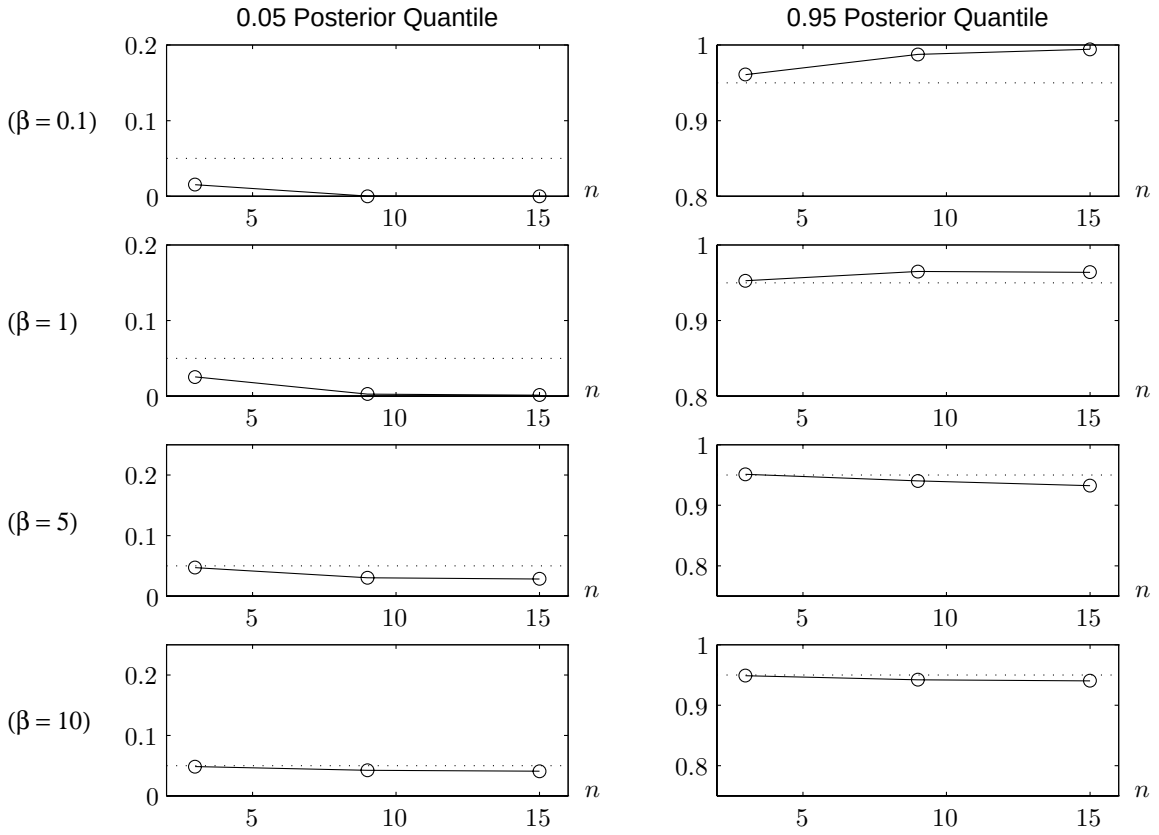

Figure 4.6: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distribution of x_0, corresponding to Hoadley's prior. The true value of x_0 is 0.5, and values of β considered are 0.1, 1, 5, and 10. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

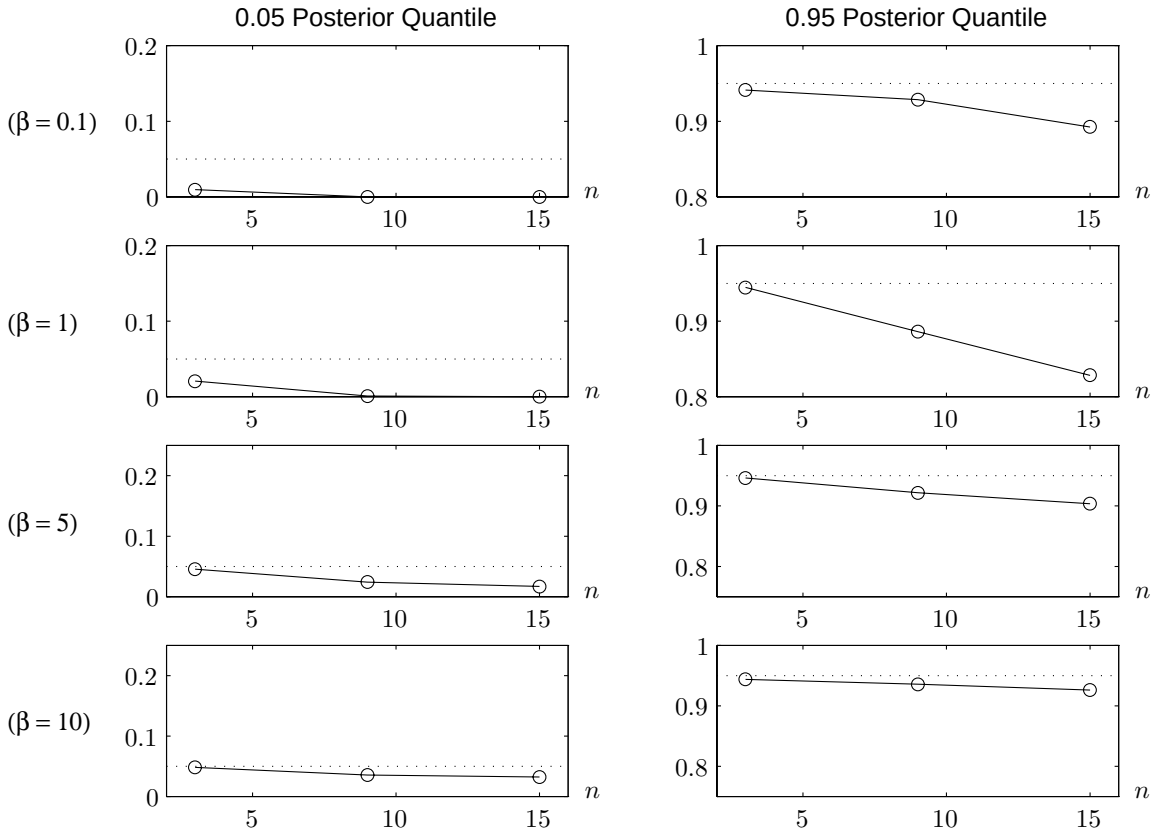

Figure 4.7: *Estimated frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles of the marginal distribution of x_0, corresponding to Hoadley's prior. The true value of x_0 is 1, and values of β considered are 0.1, 1, 5, and 10. The solid, broken, and dashed lines correspond the the values $k = 1$, $k = 2$, and $k = 3$, respectively.*

Chapter 5

Noninformative Prior Analysis of the Univariate Linear Calibration Problem – Two Variance Case

5.1 The Need for an Extended Calibration Model

Berkson (1969) expressed some concerns about the applicability of the linear calibration model, as described in section 3.1, to practical situations. In particular, he argued that the first stage of data collection is likely to take place in a carefully controlled environment, under stable experimental conditions. On the other hand, the second stage of data collection is likely to take place "in the field", under much less strictly controlled conditions.

As an illustration of Berkson's concern, we consider the experiment introduced in section 3.6. For this experiment, the first stage data consisted of twenty-seven observations, for which the regressor values (the true enzyme concentrations in the blood plasma samples) were controlled by researchers. We can presume that an attempt was made to make the samples used during the first stage of data collection as homogeneous as possible, except for the differences in enzyme concentration. In practice, the second stage of data collection is likely to occur outside of a laboratory, in a hospital. Many sources of variability can arise because of this, which were not present during the first stage of data collection. The environment in which the instrument is operated is likely to have changed (e.g., differences in temperature, humidity, barometric pressure), and the operator of the autoanalyzer is likely to have less experience than the researchers who calibrated the instrument. In addition, a blood sample collected from a patient in a hospital is likely to have differences from those used in the laboratory, apart from the difference in the true enzyme concentration. These factors

call into question the applicability of the univariate linear calibration model as previously stated.

5.2 Extension of the Univariate Linear Calibration Model

The data from the first stage of experimentation consist of the n pairs $(x_i, y_i), i = 1, 2, \ldots, n$. For each i, x_i is a known fixed value of the independent variable, and y_i is the corresponding observed response. The data from the second stage of experimentation consist of c observations of the response variable, $y_{n+1}, y_{n+2}, \ldots, y_{n+c}$, assumed to be related to a single unknown regressor value x_0 via the same relationship. Again, we assume a linear relationship between the regressor and the response (for convenience of notation, we set $x_i = x_0$ for $i \geq n + 1$):

$$y_i = \alpha + \beta x_i + \epsilon_i, \qquad i = 1, 2, \ldots, n + c, \tag{5.1}$$

where ϵ_i is a random deviation around the underlying straight line relationship. It is assumed that for $i = 1, 2, \ldots, n$, we have

$$\epsilon_i \sim N(0, \sigma^2), \tag{5.2}$$

for $i = n + 1, n + 2, \ldots, n + c$, we have

$$\epsilon_i \sim N(0, \kappa^2\sigma^2), \tag{5.3}$$

and that the ϵ_i are mutually independent. Based on the arguments from the previous section, we would expect that $\kappa > 1$, though this is not necessarily the case. Note that this extended formulation of the univariate linear calibration model has five unknown parameters: $x_0, \alpha, \beta, \kappa$, and σ. Our primary interest is in estimating x_0. However, the regression parameters α and β, and the dispersion parameters σ and κ, are also unknown, and must be treated as such in the analysis.

5.3 The Fisher Information Matrix and the Jeffreys' Prior

For the extended univariate linear calibration model, the likelihood function is

$$\mathcal{L} \propto \frac{1}{\sigma^{n+c}\kappa^c} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2 - \frac{1}{2\kappa^2\sigma^2} \sum_{i=n+1}^{n+c} (y_i - \alpha - \beta x_0)^2 \right\}.$$

The Fisher information matrix for the parameters $x_0, \alpha, \beta, \kappa$, and σ is thus

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_2 \end{bmatrix},$$

where

$$\mathcal{I}_1 = \frac{1}{\sigma^2}\begin{bmatrix} \frac{c\beta^2}{\kappa^2} & \frac{c\beta}{\kappa^2} & \frac{c\beta x_0}{\kappa^2} \\ \frac{c\beta}{\kappa^2} & n + \frac{c}{\kappa^2} & \mathfrak{s}_x + \frac{cx_0}{\kappa^2} \\ \frac{c\beta x_0}{\kappa^2} & \mathfrak{s}_x + \frac{cx_0}{\kappa^2} & \mathfrak{s}_{xx} + \frac{cx_0^2}{\kappa^2} \end{bmatrix},$$

and

$$\mathcal{I}_2 = \begin{bmatrix} \frac{2c}{\kappa^2} & \frac{2c}{\kappa\sigma} \\ \frac{2c}{\kappa\sigma} & \frac{2(n+c)}{\sigma^2} \end{bmatrix}.$$

In the above expressions, the definitions of $\mathfrak{s}_x$ and $\mathfrak{s}_{xx}$ are identical to the definitions used in chapter 3. The Jeffreys' prior for the extended univariate linear calibration problem is thus

$$\pi^J(x_0, \alpha, \beta, \sigma, \kappa) \propto |\beta|\sigma^{-3}\kappa^{-2}.$$

5.4 Development of Probability Matching Priors for x_0

As in section 4.1, we apply the method described by Tibshirani (1989) to generate priors with the probability matching property. We continue to assume that the parameter of interest is the unknown regressor value x_0. Note that the standard deviation parameters σ and κ are orthogonal x_0, so we need only reparameterize α and β to obtain our orthogonal parameterization.

The differential equations that must be satisfied in the orthogonal parameterization for the extended univariate linear calibration problem, analogous to equations (4.1) and (4.2) for the constant variance case, are as follows:

$$\left(n + \frac{c}{\kappa^2}\right)\frac{\partial\alpha}{\partial x_0} + \left(\mathfrak{s}_x + \frac{cx_0}{\kappa^2}\right)\frac{\partial\beta}{\partial x_0} = -\frac{c\beta}{\kappa^2}$$

$$\left(\mathfrak{s}_x + \frac{cx_0}{\kappa^2}\right)\frac{\partial\alpha}{\partial x_0} + \left(\mathfrak{s}_{xx} + \frac{cx_0^2}{\kappa^2}\right)\frac{\partial\beta}{\partial x_0} = -\frac{c\beta x_0}{\kappa^2}$$

One can verify that the following are solutions to this system of differential equations:

$$\beta = \lambda_1 g_1(x_0, \kappa), \qquad \text{and} \qquad \alpha = \lambda_1 g_2(x_0, \kappa) + \lambda_2;$$

the functions g_1 and g_2 in the above expressions are defined in terms of the following quantities:

$$\mathfrak{a} = \left(\frac{\kappa^2}{c}\mathfrak{s}_{xx} + x_0^2\right), \quad \mathfrak{b} = \left(\frac{n\kappa^2}{c} + 1\right), \quad \mathfrak{c} = \left(\frac{\kappa^2}{c}\mathfrak{s}_x + x_0\right),$$

$$\mathcal{A} = \mathfrak{a} - \frac{\mathfrak{c}^2}{\mathfrak{b}}, \quad \mathcal{B} = \frac{\mathfrak{c}}{\mathfrak{b}} - x_0, \quad \mathcal{C} = \mathfrak{c} - \frac{\mathfrak{a}\mathfrak{b}}{\mathfrak{c}}, \quad \text{and} \quad \mathcal{D} = \frac{\mathfrak{a}}{\mathfrak{c}} - x_0.$$

Specifically,

$$g_1(x_0, \kappa) = \exp\left\{ \int \frac{\mathcal{B}}{\mathcal{A}} \, dx_0 \right\},$$

and

$$g_2(x_0, \kappa) = \int \frac{\mathcal{D}}{\mathcal{C}} g_1(x_0, \kappa) \, dx_0.$$

The upper left element in the transformed Fisher information matrix is

$$\begin{aligned}(J'\mathcal{I}_1 J)_{(1,1)} = \frac{1}{\sigma^2}\Bigg[& \frac{c}{\kappa^2}\left(\beta^2 + \beta\frac{\partial\alpha}{\partial x_0} + \beta x_0 \frac{\partial\beta}{\partial x_0}\right) \\ & + \frac{\partial\alpha}{\partial x_0}\left(\frac{c\beta}{\kappa^2} + \left(n + \frac{c}{\kappa^2}\right)\frac{\partial\alpha}{\partial x_0} + \left(\mathfrak{s}_x + \frac{cx_0}{\kappa^2}\right)\frac{\partial\beta}{\partial x_0}\right) \\ & + \frac{\partial\beta}{\partial x_0}\left(\frac{c\beta x_0}{\kappa^2} + \left(\mathfrak{s}_x + \frac{cx_0}{\kappa^2}\right)\frac{\partial\alpha}{\partial x_0} + \left(\mathfrak{s}_{xx} + \frac{cx_0^2}{\kappa^2}\right)\frac{\partial\beta}{\partial x_0}\right)\Bigg].\end{aligned}$$

If we substitute in the solutions for α and β, given above, we see that this is proportional to

$$\lambda_1^2[g_1(x_0, \kappa)]^2\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right)S_{xx}\right)^{-1},$$

where $S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2$ is the centered sum of squared regressor values from the first stage of data collection, and the proportionality constant does not depend on any of the parameters x_0, λ_1, or λ_2. Probability matching priors, in terms of the original parameters x_0, α, and β, are thus of the form

$$\pi^P(x_0, \alpha, \beta) \propto h(\lambda_1, \lambda_2)\frac{|\beta|}{g_1(x_0, \kappa)}\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}},$$

where $h(\lambda_1, \lambda_2)$ is an arbitrary function of λ_1 and λ_2. This may be multiplied by a function of the other orthogonal parameters (namely κ and σ) as well.

It is straightforward to show that

$$\frac{\partial}{\partial x_0}\mathcal{A} = -2\mathcal{B}.$$

Hence,

$$\begin{aligned} g_1(x_0, \kappa) &= \exp\left\{ -\frac{1}{2}\int \frac{\frac{d}{dx_0}\mathcal{A}}{\mathcal{A}} \, dx_0 \right\} \\ &= \mathcal{A}^{-\frac{1}{2}}. \end{aligned}$$

This, in turn, can be shown to be proportional to

$$\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right) S_{xx}\right)^{-\frac{1}{2}},$$

where the constant of proportionality depends on no parameters except κ and σ. If we take $h(\lambda_1, \lambda_2) = 1$, then, the probability matching prior is

$$\pi(x_0, \alpha, \beta) \propto f(\sigma, \kappa)|\beta|,$$

where $f(\sigma, \kappa)$ is an arbitrary function of the two parameters κ and σ. The Jeffreys' prior is thus seen to be a probability matching prior for x_0. If we instead take $h(\lambda_1, \lambda_2) = |\lambda_1|^{-1}$, then the following are seen to be probability matching priors:

$$\pi(x_0, \alpha, \beta, \sigma, \kappa) \propto f(\sigma, \kappa)\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right) S_{xx}\right)^{-\frac{1}{2}}. \tag{5.4}$$

5.5 Posterior Analysis for a Probability Matching Prior for x_0

In this section, we study the properties of the posterior distribution for the parameters $x_0, \alpha, \beta, \kappa$, and σ, corresponding to certain priors of the form displayed in (5.4). In particular, we take the function $f(\sigma, \kappa)$ to be proportional to $\sigma^{-k}\kappa^{-l}$, so that the prior is

$$\pi(x_0, \alpha, \beta, \sigma, \kappa) \propto \sigma^{-k}\kappa^{-l}\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right) S_{xx}\right)^{-\frac{1}{2}}. \tag{5.5}$$

Note that the likelihood function for the $n+c$ observations, with the error structure described by equations (5.2) and (5.3), may be written

$$\mathcal{L} \propto \sigma^{-(n+c)}\left(\prod_{i=1}^{n+c} w_i^{\frac{1}{2}}\right) \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c} w_i(y_i - \alpha - \beta x_i)^2\right\},$$

where

$$w_i = \begin{cases} 1 & \text{for } i = 1, 2, \ldots, n \\ \kappa^{-2} & \text{for } i = n+1, n+2, \ldots, n+c \end{cases} \tag{5.6}$$

As before, we have denoted x_0 as x_i, for $i = n+1, n+2, \ldots, n+c$.

5.5.1 Integrability of the Posterior

The proof of the following theorem is rather tedious, and is left to the appendix (section A.3).

THEOREM 2 *For the extended univariate linear calibration problem, as described above, the posterior corresponding to a prior of the form*

$$\pi(x_0, \alpha, \beta, \sigma, \kappa) \propto \sigma^{-k}\kappa^{-l}\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}}$$

is proper, providing that $l \geq 1$, $n + k \geq 2l + 1$, *and* $c \geq 2$.

Note that in order for the posterior to be integrable, one of the requirements is that we have at least two observations of the response at the second stage of data collection. Intuitively, this can be attributed to the fact that some information about the value of κ must be provided by the data.

5.5.2 Gibbs Sampling of the Posterior

In the previous section and the appendix , the posterior distributions corresponding to a class of probability matching priors were shown to be proper. However, no marginal posterior distributions were explicitly derived. The joint and marginal posteriors for the parameters $x_0, \alpha, \beta, \sigma$, and κ can be studied via the Gibbs sampler. The development of the relevant conditional distributions will be presented in this section.

As shown in the appendix (section A.3), the conditional posterior for α is

$$\alpha|x_0, \beta, \sigma, \kappa, \mathbf{y} \sim N\left((\bar{y}_w - \beta\bar{x}_w), \frac{\sigma^2}{S_w}\right).$$

where

$$S_w = \sum_{i=1}^{n+c} w_i, \quad \bar{y}_w = \frac{\sum_{i=1}^{n+c} w_i y_i}{S_w}, \quad \bar{x}_w = \frac{\sum_{i=1}^{n+c} w_i x_i}{S_w},$$

and

$$S_{(y-\beta x)w(y-\beta x)} = \sum_{i=1}^{n+c} w_i\big(y_i - \beta x_i - (\bar{y}_w - \beta\bar{x}_w)\big)^2.$$

It can be shown similarly that the conditional posterior for β is

$$\beta|x_0, \alpha, \sigma, \kappa, \mathbf{y} \sim N\left(\frac{\mathfrak{s}_{xw(y-\alpha)}}{\mathfrak{s}_{xwx}}, \frac{\sigma^2}{\mathfrak{s}_{xwx}}\right).$$

In this expression, the quantities labeled $\mathfrak{s}_{(\cdot)}$ are uncentered, weighted sums of squares and products, indexed from 1 to $n + c$.

In implementing the Gibbs sampler, we simulate values of σ indirectly, by first simulating values from the distribution of σ^2. The conditional distribution of σ^2 is

$$\sigma^2|x_0,\alpha,\beta,\kappa,\mathbf{y} \sim \mathcal{IG}.\left(\frac{n+c+k-1}{2}, \frac{2}{\mathfrak{s}_{(y-\alpha-\beta x)w(y-\alpha-\beta x)}}\right),$$

where, again, $\mathfrak{s}_{(\cdot)}$ indicates an uncentered weighted sum indexed from 1 to $n+c$. For the actual implementation of the Gibbs sampler on a computer, it is easier to instead simulate

$$\frac{\mathfrak{s}_{(y-\alpha-\beta x)w(y-\alpha-\beta x)}}{\sigma^2} \sim \chi^2_{n+c+k-1},$$

where $\chi^2_{n+c+k-1}$ denotes the chi-squared distribution with $n+c+k-1$ degrees of freedom.

As with the parameter σ, we simulate values of κ indirectly, by first simulating values of κ^2. The parameter κ enters the posterior through both the prior and the likelihood. If we consider the term κ^{-l} from the prior, combined with the likelihood function, the involvement of κ^2 is in the form of an inverse gamma distribution, with parameters

$$\frac{l+c-1}{2} \quad \text{and} \quad \frac{2\sigma^2}{\mathfrak{s}^-_{(y-\alpha-\beta x)^2}},$$

where $\mathfrak{s}^-_{(y-\alpha-\beta x)^2}$ denotes the uncentered unweighted sum indexed from $n+1$ to $n+c$. This inverse gamma density for κ^2 is multiplied by the other term from the prior, which also involves κ^2. Note that this term is bounded above:

$$\left((x_0-\bar{x})^2+\left(\frac{\kappa^2}{c}+\frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}} \leq \left((x_0-\bar{x})^2+\frac{1}{n}S_{xx}\right)^{-\frac{1}{2}}.$$

Hence, we can use the rejection method to simulate values of κ^2, using the inverse gamma distribution as our proposal distribution. Simulation from this proposal distribution can be simplified by instead simulating

$$\frac{\mathfrak{s}^-_{(y-\alpha-\beta x)^2}}{\sigma^2\kappa^2} \sim \chi^2_{l+c-1},$$

where χ^2_{l+c-1} denotes the chi-squared distribution with $l+c-1$ degrees of freedom.

Simulation from the conditional distribution for x_0 can also be accomplished via the rejection method. The conditional density for x_0 has the form of a normal density:

$$N\left(\frac{1}{\beta}(\bar{y}_0-\alpha), \frac{\sigma^2\kappa^2}{c\beta^2}\right),$$

multiplied by a term from the prior, which is bounded above:

$$\left((x_0-\bar{x})^2+\left(\frac{\kappa^2}{c}+\frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}} \leq \left(\left(\frac{\kappa^2}{c}+\frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}}.$$

The normal distribution can therefore be used as our proposal distribution when invoking the rejection method to simulate values of x_0.

Through the simulation study described in section 5.7, it was found that occasionally a numerical overflow occurred when either σ or κ was allowed to become too small. Hence, in implementing the Gibbs sampler, the simulated values of these parameters were constrained to be greater than 10^{-7}. This constraint required enforcement very rarely, so is unlikely to drastically effected posterior inferences.

5.6 Development of Probability Matching Priors for κ

Although the parameter of greatest interest in a calibration problem is typically the unknown regressor value x_0, it is conceivable that, in the context of the extended univariate linear calibration problem, the parameter κ is of interest. In this section, probability matching priors for this parameter are derived.

We begin by noting that the three parameters x_0, α, and β are already orthogonal to κ. Hence, we need only reparameterize σ to obtain a probability matching prior for κ. The block of the information matrix corresponding to the parameter vector (κ, σ) is

$$\mathcal{I}_2 = \begin{bmatrix} \frac{2c}{\kappa^2} & \frac{2c}{\kappa\sigma} \\ \frac{2c}{\kappa\sigma} & \frac{2(n+c)}{\sigma^2} \end{bmatrix}.$$

Only one differential equation must be solved in order to orthogonalize this:

$$\frac{2(n+c)}{\sigma^2}\frac{\partial\sigma}{\partial\kappa} = -\frac{2c}{\kappa\sigma}.$$

The following is a general solution to this differential equation:

$$\sigma = \lambda\kappa^r,$$

where $r = \frac{c}{n+c}$. Now the matrix

$$J = \begin{bmatrix} 1 & 0 \\ \frac{\partial\sigma}{\partial\kappa} & \frac{\partial\sigma}{\partial\lambda} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ -r\lambda\kappa^{-r-1} & \kappa^{-r} \end{bmatrix}.$$

The first element of $J'I_2J$ is thus

$$(J'\mathcal{I}_2J)_{(1,1)} = 2c\kappa^{-2}(1-r).$$

Probability matching priors for κ will therefore be of the form

$$\pi^P(x_0, \alpha, \beta, \lambda, \kappa) \propto f(x_0, \alpha, \beta, \lambda)\kappa^{-1},$$

where $f(x_0, \alpha, \beta, \lambda)$ is an arbitrary function of the orthogonal parameters. If we transform back to the original parameterization, the Jacobian is proportional to k^{-r}. Hence, in the original parameterization, the probability matching priors will be of the form

$$\pi^P(x_0, \alpha, \beta, \sigma, \kappa) \propto f(x_0, \alpha, \beta, \lambda)\kappa^{-\frac{n}{n+c}}.$$

We note that the probability matching priors for the parameter κ do not agree with those for x_0.

5.7 An Empirical Assessment of the Extended Model Analysis

In this section, we describe a simulation study which evaluates frequentist coverage properties for priors of the form displayed in (5.5), for moderate sample sizes. In light of the results of the simulation study described by Eno and Ye (1998), we choose to study the prior with $k = l = 1$. For comparison, we also study the frequentist coverage properties of the reference prior analysis based on the simpler calibration model (the reference prior is also a probability matching prior for that model). Specifically, the reference prior we choose for the simpler analysis is

$$\pi(x_0, \alpha, \beta, \sigma) \propto \sigma^{-1}\left((x_0 - \bar{x})^2 + \left(\frac{1}{c} + \frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}}.$$

This prior was chosen for comparison because it was shown by Eno and Ye (1998) to attain the probability matching property more quickly than other reference priors.

In the simulation study, two different values of β were considered, namely $\beta = 5$ and $\beta = 2$. In all cases, the intercept α was taken to be zero and the standard deviation σ was taken to be one. Four values of κ were considered. The value $\kappa = 1$ was considered; this represents a situation in which the errors at the two stages of data collection have identical standard deviation. Two values were considered ($\kappa = 1.5$ and $\kappa = 2$) which correspond to situations in which the second stage observations are more dispersed about the linear model than the first stage observations. Finally, one value considered ($\kappa = 0.5$) corresponds to a situation in which the second stage data are less dispersed about the linear model than the first stage data. We note that this last situation seems unlikely to arise in practice. The first stage regressor values were uniformly spaced in the interval $[-1, 1]$, and the second stage regressor value x_0 was initially taken to be zero (the center of the first stage data range). The simulations corresponding to $\beta = 5$ were repeated with $x_0 = 1$ (the edge of the first stage data range).

Several combinations of sample sizes were considered. Specifically, the first stage sample size was taken to be $n = 9, n = 15, n = 21$ and $n = 27$, and the second stage sample size was

taken to be $c = 2, c = 5$, and $c = 8$. In practice, the likely situations to arise would seem to be cases in which n is greater than c, probably much greater.

In evaluating the frequentist coverage properties for the various sets of parameter values, the Gibbs sampler was used to generate posterior inferences. For each set of parameter values considered, a data set was randomly generated according to the extended linear calibration model ((5.1), (5.2), (5.3)). The Gibbs sampler was then used to approximate the marginal posterior distribution for x_0. The number of random draws of x_0 taken from the Gibbs sampler was 5000, after an initial burn-in period of 10 samples (which were discarded). The methodology of Raftery and Lewis (1992), which provides guidelines on the necessary number of iterations of the Gibbs sampler for estimation of a specified posterior quantile to a desired degree of accuracy, was used to guide these decisions. Based on the resulting sample from the Gibbs sampler, the 0.05 and 0.95 posterior quantiles for x_0 were estimated (to within 0.01, with 95% confidence, as described by Raftery and Lewis (1992)), and it was determined whether or not the true specified value of x_0 was less than or equal to the estimated quantile. For each set of parameter values considered, the above procedure was repeated 8,000 times. From these repetitions, the frequentist coverage probabilities of the 0.05 and 0.95 posterior quantiles were estimated.

The above procedure was done for the both the extended linear calibration model and the simpler model (ignoring the heterogeneity of variance), for each simulated data set. The results of the simulation study for $\beta = 5$, $x_0 = 0$ are displayed in figures 5.1 and 5.2, which we discuss below. The corresponding results for $\beta = 2$, $x_0 = 0$ and $\beta = 5$, $x_0 = 1$ are displayed in figures 5.3 through 5.6, at the end of this chapter. Since these results are similar, we do not discuss them at length.

In figure 5.1, (d) through (f) correspond to the case $\kappa = 1$. In this case, we would expect the simpler model to perform better than the extended model, since both models would accurately describe the way in which the data were generated, but the simpler model would essentially fix the value of κ at 1 (correctly, in this case). We see that in this case the 0.05 and 0.95 posterior quantiles for the simpler analysis have approximately the correct coverage probabilities, for all sample sizes. The extended model requires larger values of n and c to approach the nominal coverage probabilities for these quantiles. In particular, $c = 2$ measurements at the second stage of data collection do not seem to provide sufficient information about κ for the extended model to attain the probability matching property. For all values of c, interval estimates of x_0 based on the region between the 0.05 and 0.95 posterior quantiles tend to be wider for the extended model analysis than for the simpler model analysis, though this trait diminishes as sample sizes increase.

Parts (a) through (c) of figure 5.1 correspond to the case $\kappa = 0.5$, in which the second stage

data have less variability than the first stage data. In this case, the simpler model analysis *incorrectly* fixes κ at one, while the extended model analysis accounts for the heterogeneity of variance. The extended model analysis seems to do a fairly good job of attaining the probability matching property, for $c = 5$ and $c = 8$. Again, $c = 2$ second stage observations do not seem to be provide sufficient information about the value of κ in order to attain the probability matching property. In the case of $\kappa = 0.5$, an interval estimate based on the posterior quantiles from the simpler model analysis would be wider than necessary to achieve the nominal frequentist coverage probability.

Figure 5.2 corresponds to cases in which κ is larger than one. Specifically, (a) through (c) correspond to $\kappa = 1.5$, while (d) through (f) correspond to $\kappa = 2$. We can see that the posterior quantiles for x_0 from the simpler analysis do not approach the nominal frequentist coverage probabilities, even as the sample sizes increase. In some cases, the difference is quite severe. On the other hand, the extended model analysis does a much better job of approximating the nominal frequentist coverage probabilities, when $c = 5$ or $c = 8$. (Once again, $c = 2$ second stage observations seem to provide too little information about the value of κ to be reliable. We note, however, that the extended analysis comes closer to correctly approximating to the nominal frequentist coverage probabilities than the simpler analysis, and does so in a way that tends to be more conservative.) Interval estimates based on the simpler model analysis would be narrower than those based on the extended model analysis, and would have lower than the nominal 90% frequentist coverage probability.

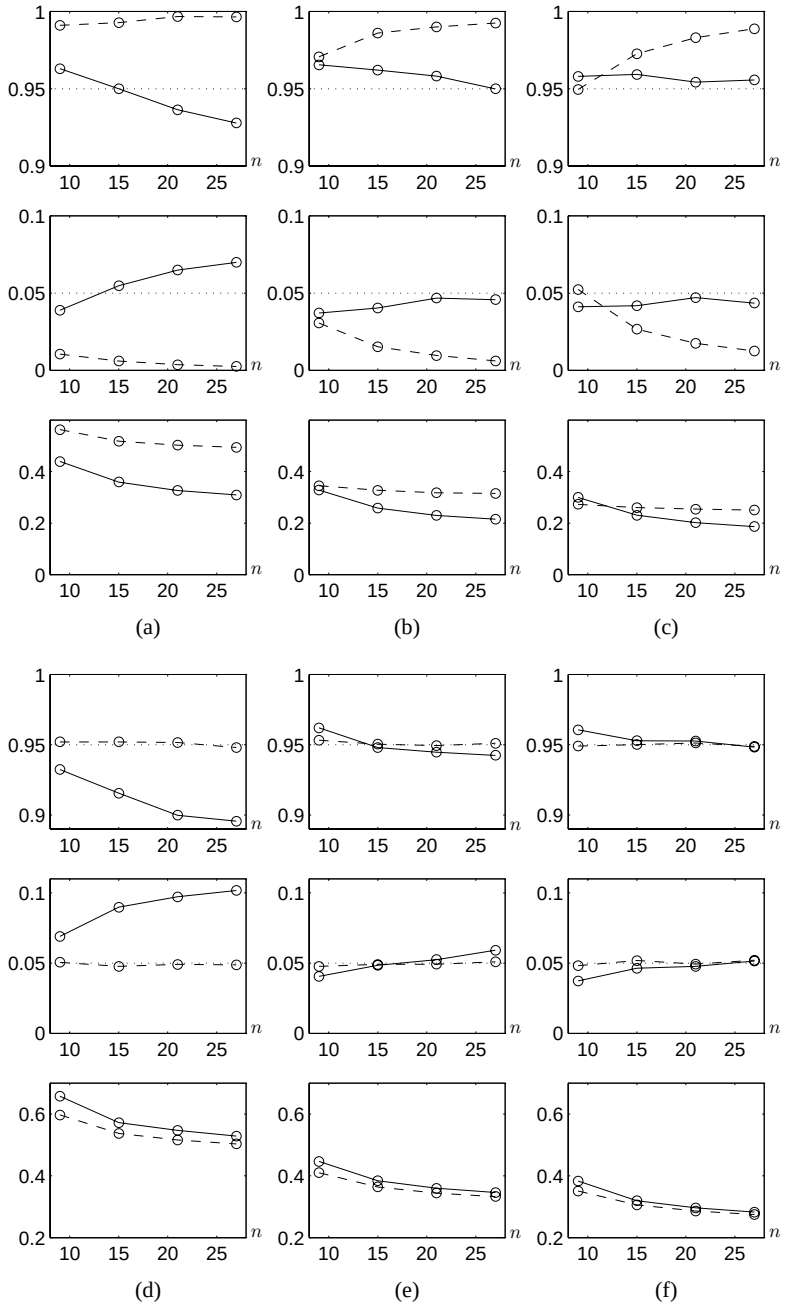

Figure 5.1: *Estimated frequentist coverage probabilities for the marginal distributions of* x_0*. On each plot, the solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler model analysis (ignoring heterogeneity of variance). In all cases, the true value of* x_0 *is 0, and the true value of* β *is 5.* (a), (b), *and* (c) *correspond to* $\kappa = 0.5$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively;* (d), (e), *and* (f) *correspond to* $\kappa = 1.0$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively. In each case, the uppermost two plots display the estimated frequentist coverage probabilities of the 0.95 and 0.05 posterior quantiles, respectively, and the lowermost plot displays the average observed distance between these quantiles.*

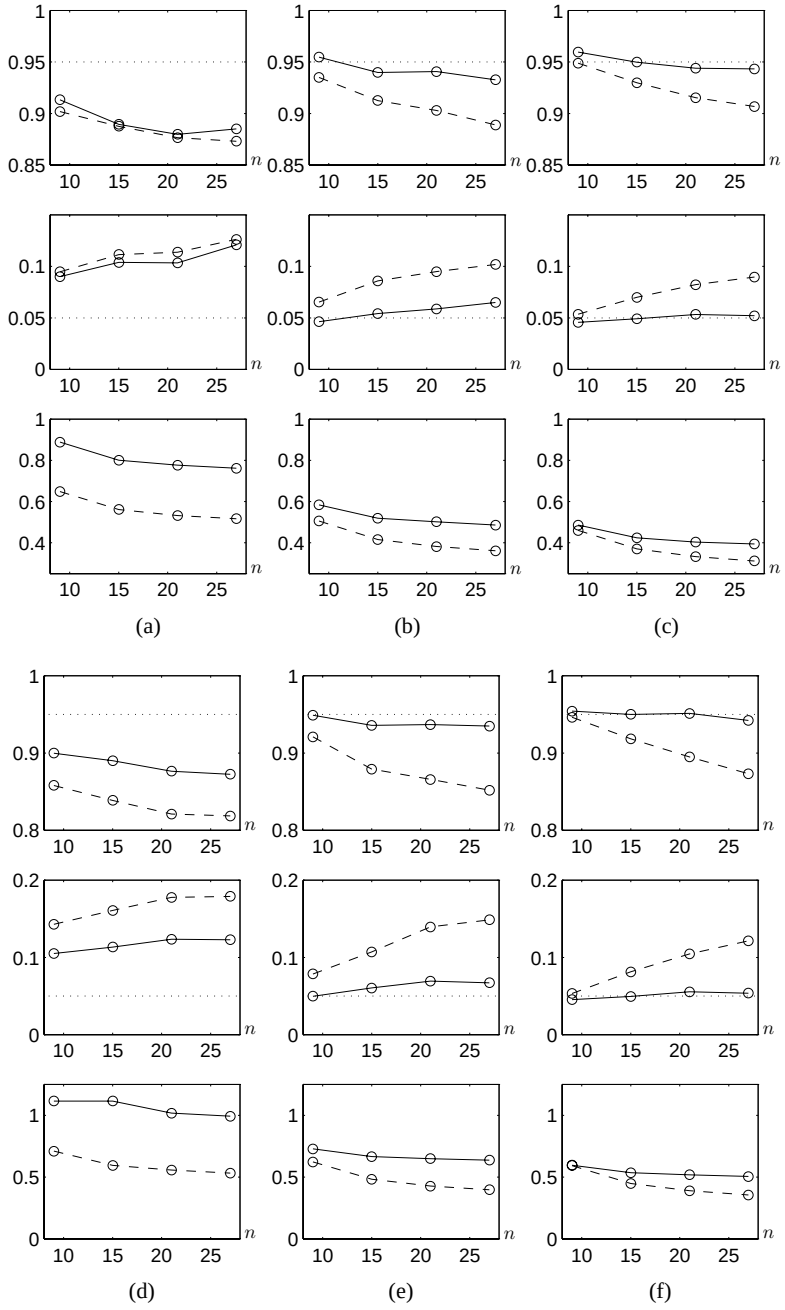

Figure 5.2: *Estimated frequentist coverage probabilities for the marginal distributions of* x_0*. On each plot, the solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler model analysis (ignoring heterogeneity of variance). In all cases, the true value of* x_0 *is 0, and the true value of* β *is 5.* (a), (b), *and* (c) *correspond to* $\kappa = 1.5$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively;* (d), (e), *and* (f) *correspond to* $\kappa = 2.0$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively. In each case, the uppermost two plots display the estimated frequentist coverage probabilities of the 0.95 and 0.05 posterior quantiles, respectively, and the lowermost plot displays the average observed distance between these quantiles.*

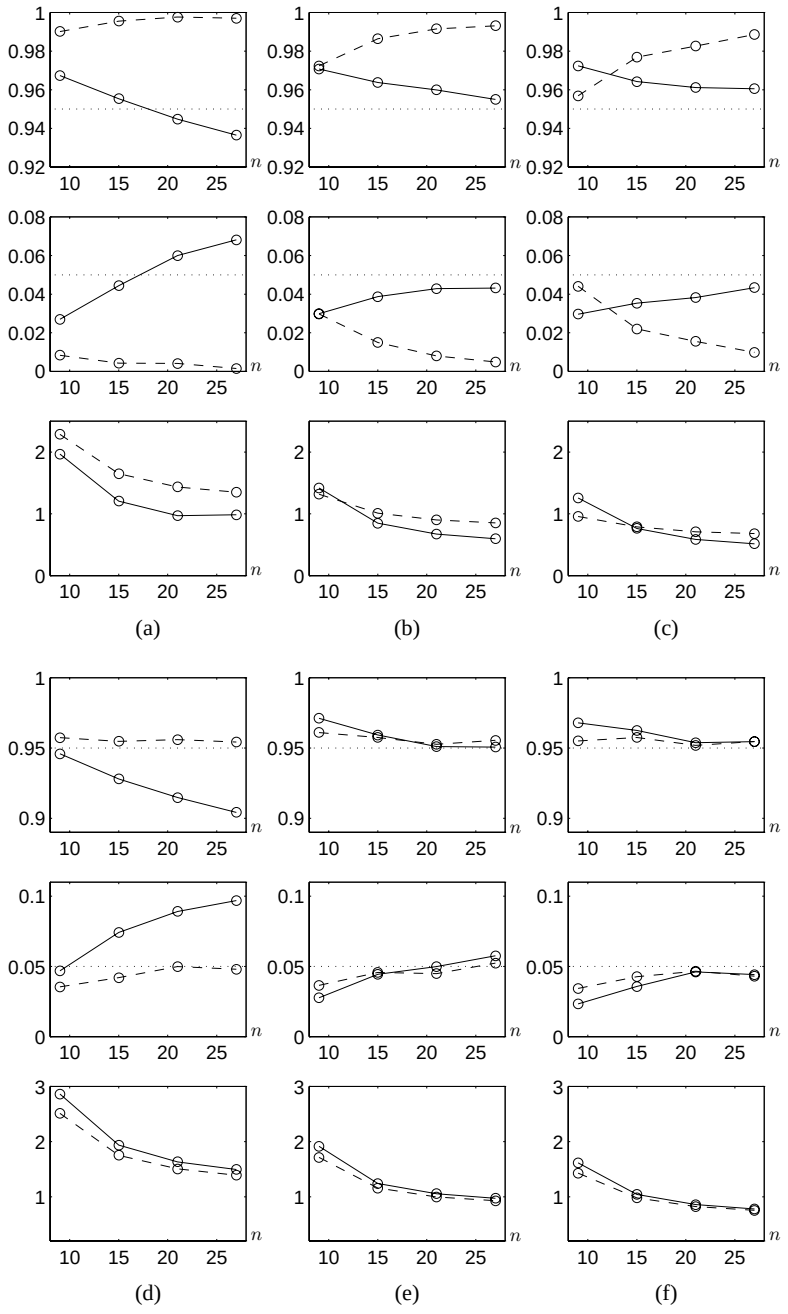

Figure 5.3: *Estimated frequentist coverage probabilities for the marginal distributions of* x_0*. On each plot, the solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler model analysis (ignoring heterogeneity of variance). In all cases, the true value of* x_0 *is 0, and the true value of* β *is 2.* (a), (b), *and* (c) *correspond to* $\kappa = 0.5$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively;* (d), (e), *and* (f) *correspond to* $\kappa = 1.0$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively. In each case, the uppermost two plots display the estimated frequentist coverage probabilities of the 0.95 and 0.05 posterior quantiles, respectively, and the lowermost plot displays the average observed distance between these quantiles.*

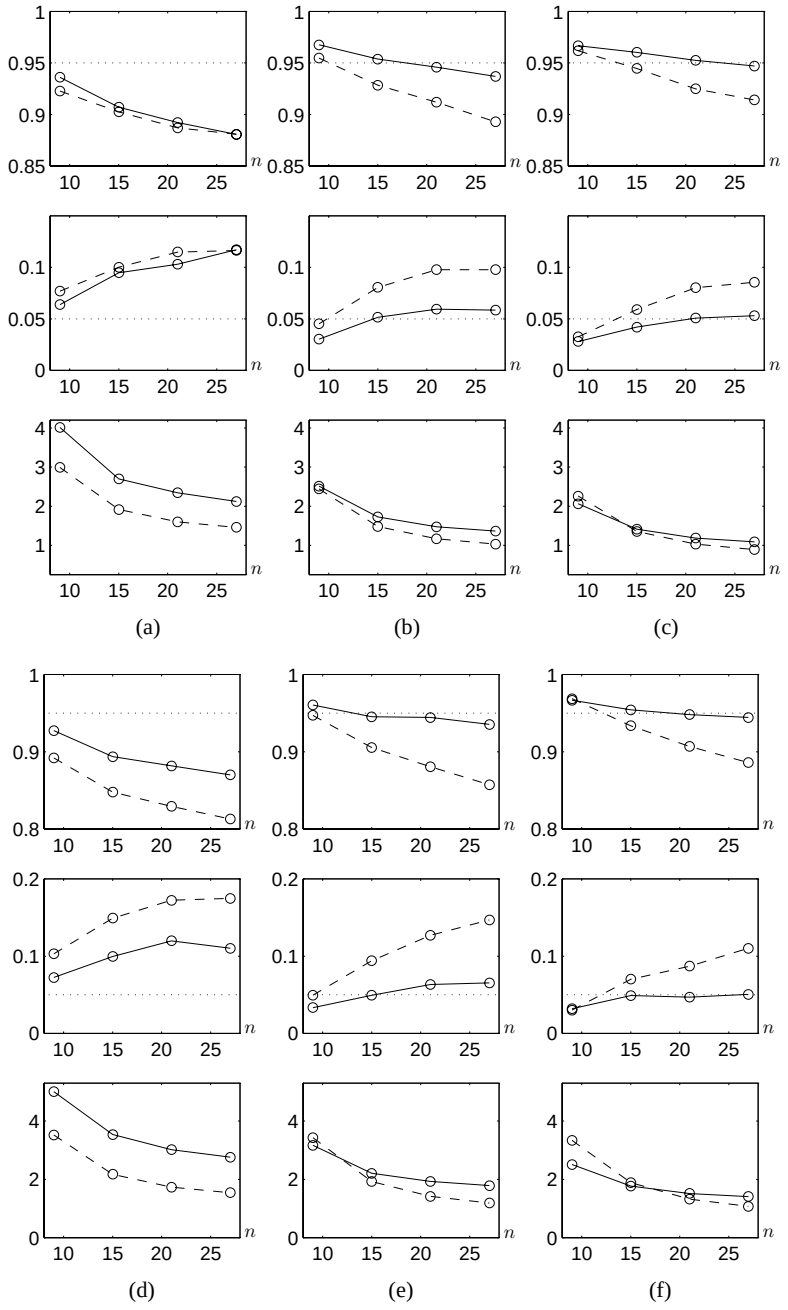

Figure 5.4: *Estimated frequentist coverage probabilities for the marginal distributions of* x_0*. On each plot, the solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler model analysis (ignoring heterogeneity of variance). In all cases, the true value of* x_0 *is 0, and the true value of* β *is 2.* (a), (b), *and* (c) *correspond to* $\kappa = 1.5$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively;* (d), (e), *and* (f) *correspond to* $\kappa = 2.0$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively. In each case, the uppermost two plots display the estimated frequentist coverage probabilities of the 0.95 and 0.05 posterior quantiles, respectively, and the lowermost plot displays the average observed distance between these quantiles.*

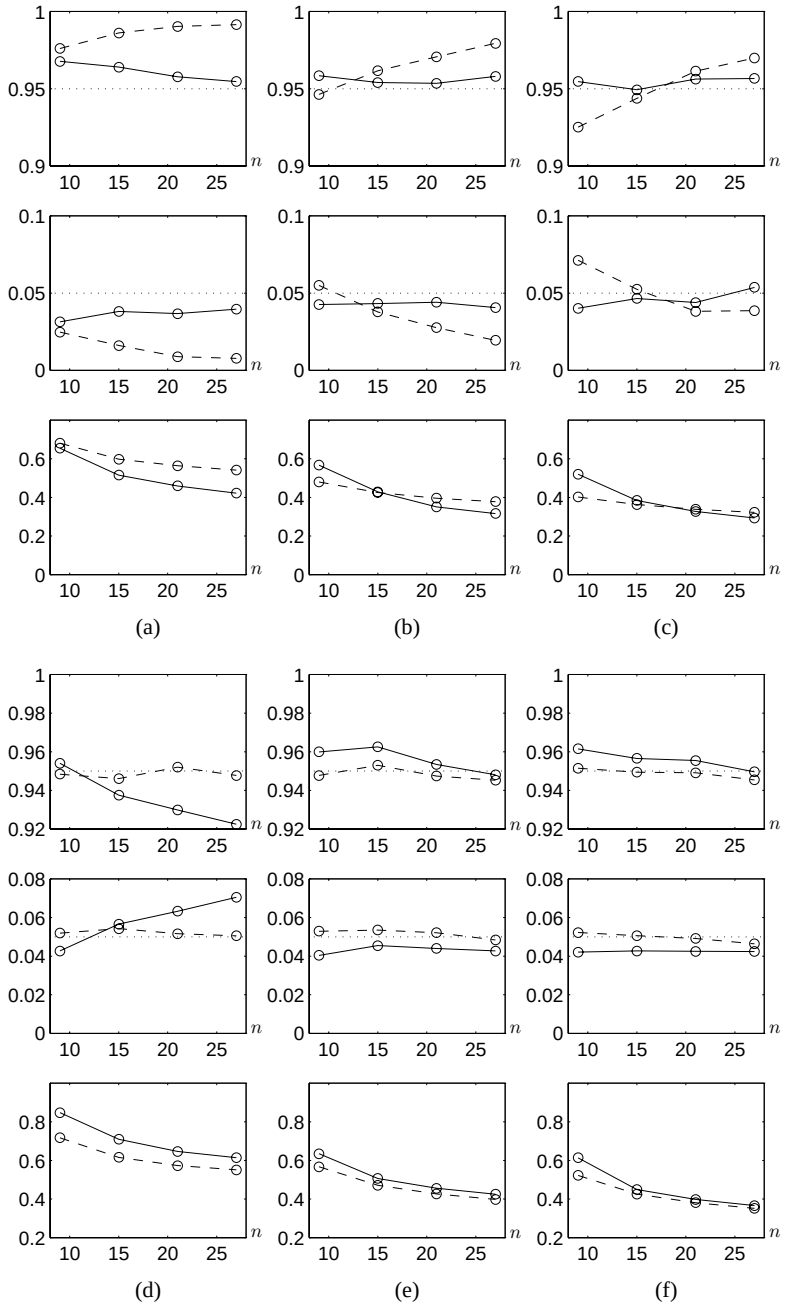

Figure 5.5: *Estimated frequentist coverage probabilities for the marginal distributions of* x_0. *On each plot, the solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler model analysis (ignoring heterogeneity of variance). In all cases, the true value of* x_0 *is 1, and the true value of* β *is 5.* (a), (b), *and* (c) *correspond to* $\kappa = 0.5$, *with* $c = 2$, $c = 5$, *and* $c = 8$, *respectively;* (d), (e), *and* (f) *correspond to* $\kappa = 1.0$, *with* $c = 2$, $c = 5$, *and* $c = 8$, *respectively. In each case, the uppermost two plots display the estimated frequentist coverage probabilities of the 0.95 and 0.05 posterior quantiles, respectively, and the lowermost plot displays the average observed distance between these quantiles.*

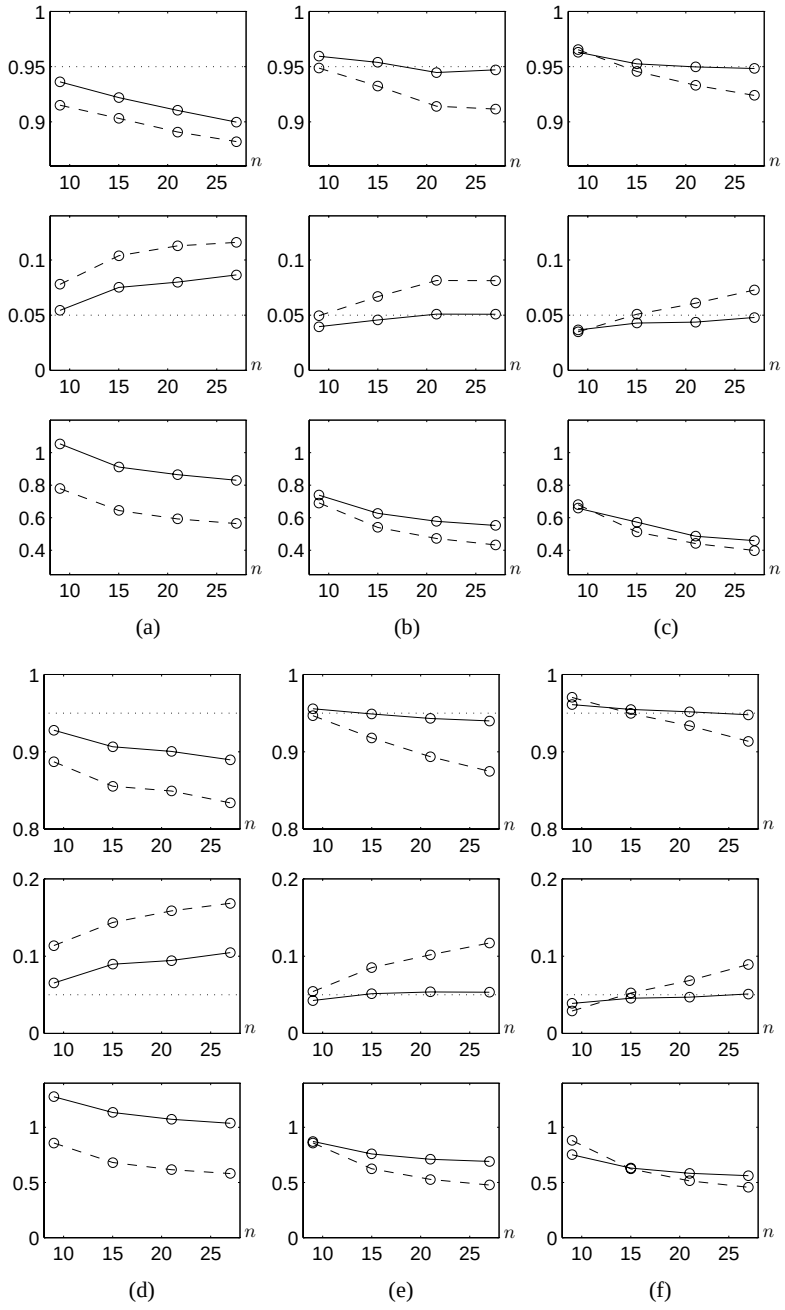

Figure 5.6: *Estimated frequentist coverage probabilities for the marginal distributions of* x_0*. On each plot, the solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler model analysis (ignoring heterogeneity of variance). In all cases, the true value of* x_0 *is 1, and the true value of* β *is 5.* (a), (b), *and* (c) *correspond to* $\kappa = 1.5$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively;* (d), (e), *and* (f) *correspond to* $\kappa = 2.0$*, with* $c = 2$*,* $c = 5$*, and* $c = 8$*, respectively. In each case, the uppermost two plots display the estimated frequentist coverage probabilities of the 0.95 and 0.05 posterior quantiles, respectively, and the lowermost plot displays the average observed distance between these quantiles.*

5.8 An Illustrative Example

As an example of linear calibration, we consider the data presented by Afifi and Azen (1972), page 88. The data consist of twenty values, measured with an instrument, of the concentration of lactic acid in blood samples. These measurements (the responses y_i) correspond to blood samples in which the true lactic acid concentration was fixed at five values (the regressor values x_i). A scatterplot of this data is shown in figure 5.7.

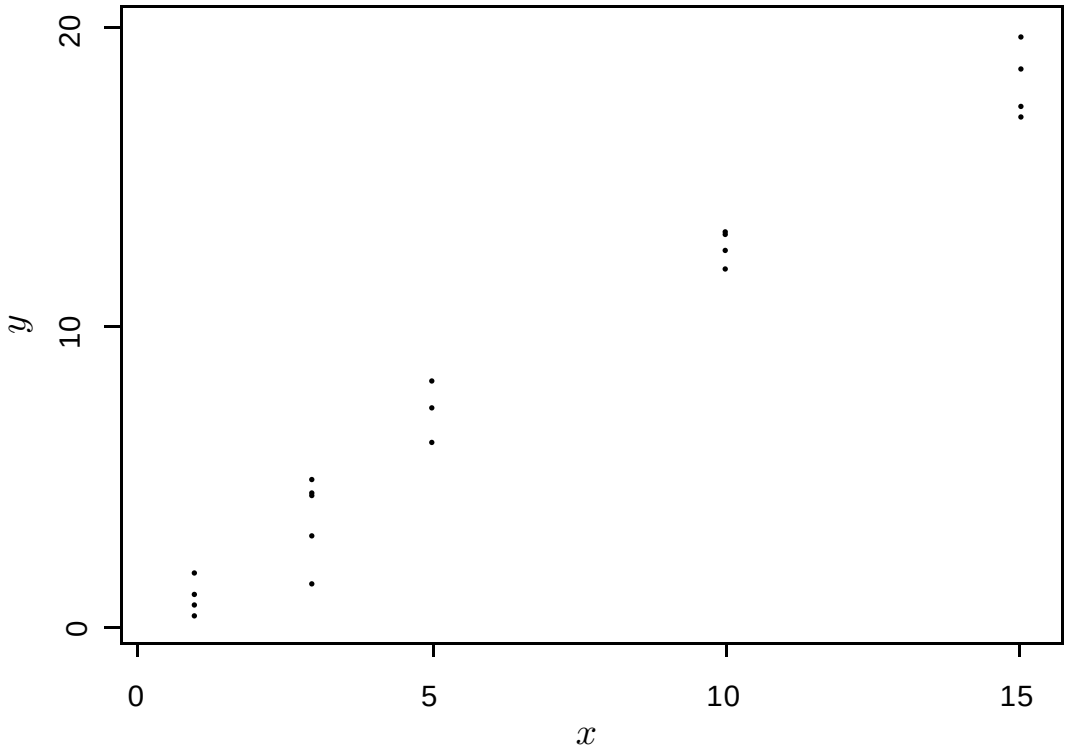

Figure 5.7: *Scatterplot of the data from Afifi and Azen (1972).*

From the scatterplot, it appears that the responses corresponding to $x = 3$ are more dispersed than the observations at other regressor values. Although these observations were measured during the first stage of data collection, we will assume, for the purpose of illustration, that these were second stage data. The remaining observations will be considered first stage data. Hence, for this data set, we have $n = 15$ first stage observations, and $c = 5$ second stage observations. Although these sample sizes are not extremely large, the simulation study described above suggests that the extended model analysis should provide more realistic interval estimates for the "unknown" regressor value.

This data set was analyzed using both the extended model analysis, and the simpler

model analysis. The estimated marginal posterior densities for x_0 (kernel density estimates from 20,000 values from the Gibbs samplers) are displayed in figure 5.8. From these density estimates, it is clear that the extended model analysis results in a more diffuse posterior for x_0. Ninety percent credibility intervals for x_0 for the two analyses are as follows: (1.787, 3.614) for the simple model analysis; (1.548, 3.832) for the extended model analysis. For this set of data, the second stage regressor value was known to be $x_0 = 3$, which is contained in both of the credibility intervals. However, based on the results of the simulation study described in the previous section, it would appear that under repeated sampling, the intervals stemming from the extended model analysis would come much closer to attaining the nominal frequentist confidence level of 90%.

Figure 5.9 displays the estimated marginal posterior density of κ, from the extended model analysis. Again, this is a kernel density estimate, from 20,000 values from the Gibbs sampler. Approximately 80% of the sampled values were greater than 1.0. This provides further evidence of the appropriateness of the extended model analysis.

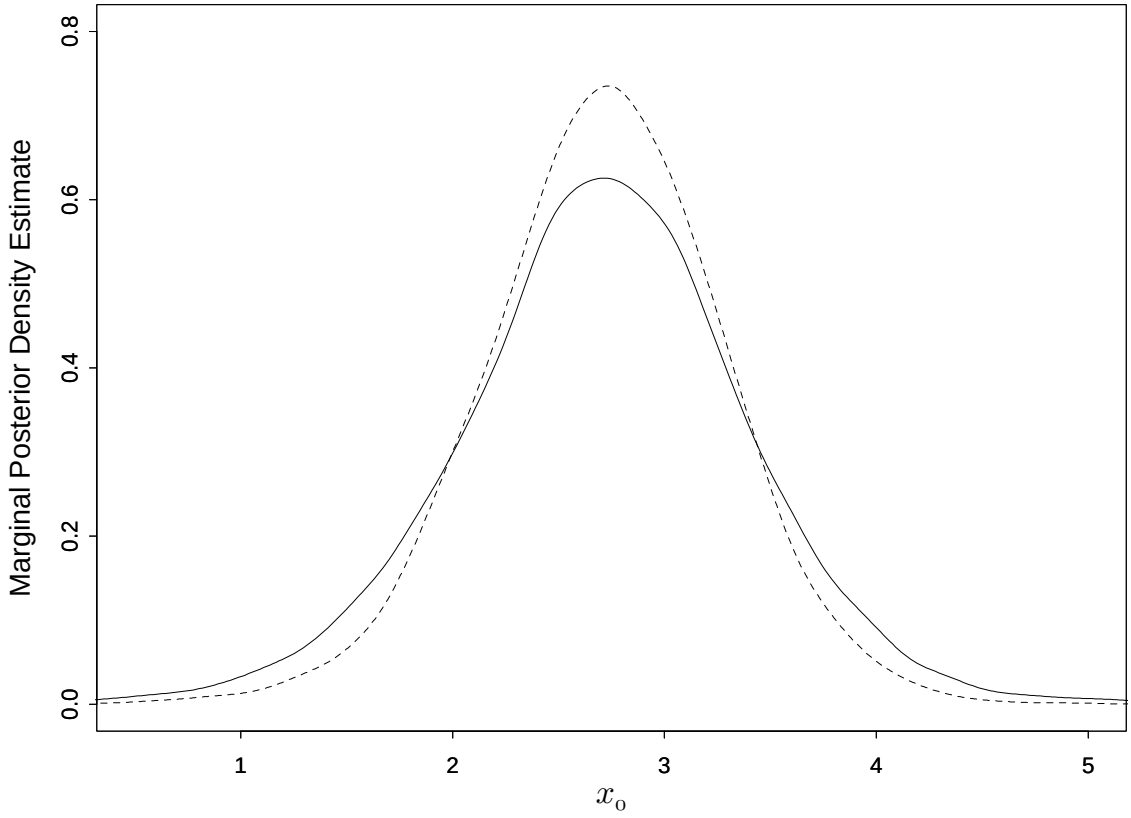

Figure 5.8: *Marginal kernel density estimates for x_0, for the example from Afifi and Azen (1972). The solid line corresponds to the extended model analysis, and the dashed line corresponds to the simpler analysis (ignoring heterogeneity of variance).*

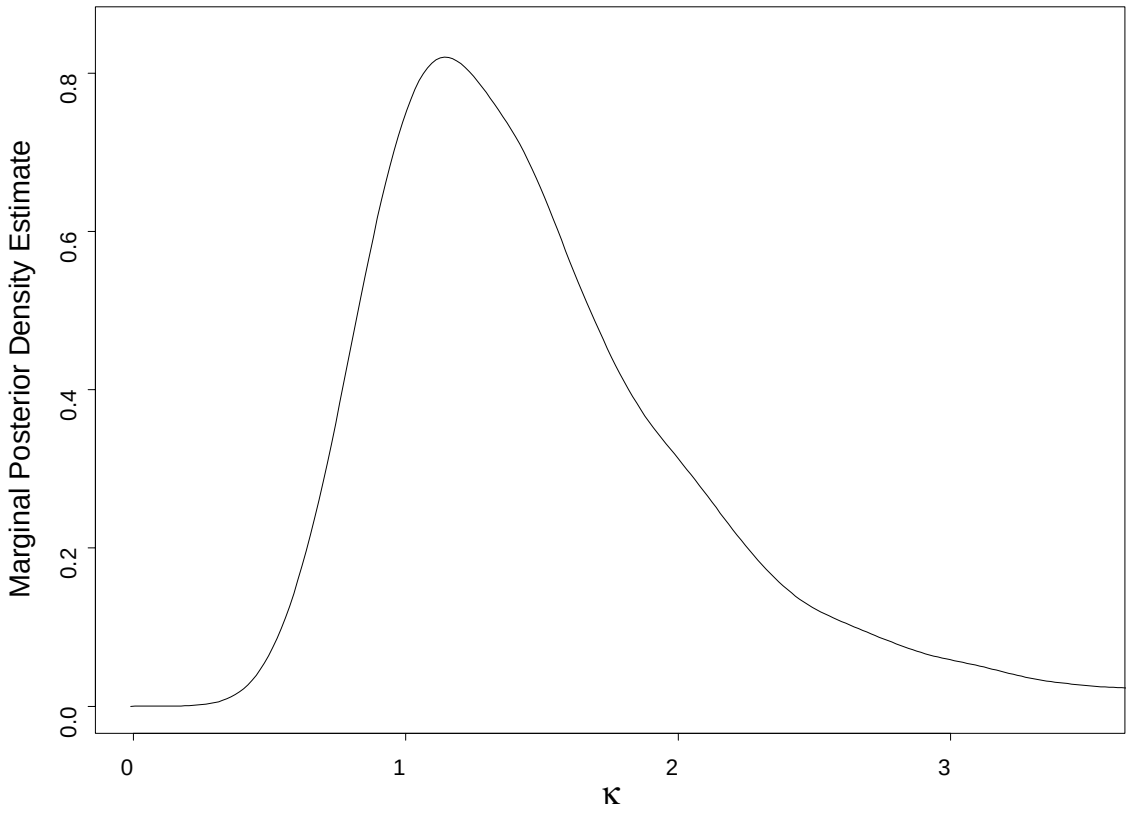

Figure 5.9: *Marginal kernel density estimate for κ, from the extended model analysis of the example from Afifi and Azen (1972). The posterior probability that $\kappa > 1$ is approximately 0.80.*

Chapter 6

Reference Prior Analysis of The Multivariate Linear Calibration Problem

6.1 Formulation of the Multivariate Calibration Problem

In this chapter, we apply the reference prior method to the multivariate multiple calibration problem. As before, the data are collected in two stages. In the first stage of experimentation (the calibration stage), n runs are performed in which a set of control variable values are fixed at known values, and the resulting values of several response variables are observed. In the second stage (the prediction stage), c runs are performed in which the response variables are observed, all corresponding to a single unknown setting of the control variables. It is desired to make inferences about the unknown control variables which gave rise to these c sets of responses. A more specific formulation is described below.

We assume that there are q response variables $y_1, y_2, \ldots, y_q$. These responses are assumed to depend linearly on p regressor variables $x_1, x_2, \ldots, x_p$. A multivariate linear model is assumed:

$$\mathbf{Y} = \mathbf{1}\boldsymbol{\alpha}' + \mathbf{XB} + \mathbf{E}. \tag{6.1}$$

In this model, $\mathbf{Y}$ is an $(n+c) \times q$ matrix containing the observed responses from both stages of experimentation. The j^{th} column of $\mathbf{Y}$ contains the observed values of the j^{th} response variable from the $n+c$ experimental runs. The j^{th} column of $\mathbf{Y}$ will be denoted as $\mathbf{y}_j$.

The vector $\mathbf{1}$ is an $(n+c) \times 1$ column vector of unity elements, and the vector $\boldsymbol{\alpha}' = [\alpha_1, \alpha_2, \ldots, \alpha_q]$ is an unknown $1 \times q$ vector of intercept terms. $\mathbf{B}$ is an unknown $p \times q$

matrix of regression coefficients. We will write $\mathbf{B}$ as $[\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_q]$, so that $\boldsymbol{\beta}_j$ contains the regression coefficients for the j^{th} response variable.

The matrix $\mathbf{X}$ is of dimension $(n+c) \times p$. This is the matrix containing the settings of the regressors from the $n+c$ experimental runs (the i^{th} experimental run in the i^{th} row). Note that the first n rows of $\mathbf{X}$ are known, while the last c rows of $\mathbf{X}$ are unknown. Indeed, the focus of interest of the calibration problem is the estimation of the control variable settings in these last c rows of $\mathbf{X}$. Let the i^{th} row of $\mathbf{X}$ be denoted as $\mathbf{x}_i'$. Note that for $i = n+1, n+2, \ldots, n+c$, the vectors $\mathbf{x}_i'$ are all identical. We will refer to this (unknown) vector as $\mathbf{x}_0'$.

Finally, $\mathbf{E}$ is an $(n+c) \times q$ matrix of errors. If we let $\boldsymbol{\epsilon}_i'$ denote the i^{th} row of $\mathbf{E}$, then $\boldsymbol{\epsilon}_i'$ contains the errors in the responses at the i^{th} experimental run. We assume that the $\boldsymbol{\epsilon}_i$ are all independent and identically distributed, as $N_q(\mathbf{0}, \boldsymbol{\Sigma})$. We further assume that $\boldsymbol{\Sigma}$ is the diagonal matrix $\text{DIAG}(\tilde{\sigma}_1^2, \tilde{\sigma}_2^2, \ldots, \tilde{\sigma}_q^2)$ – that is, that the errors in the q responses are independent of one another. The diagonal elements of $\boldsymbol{\Sigma}$ may be all equal or all different from one another. More generally, we assume that $\boldsymbol{\Sigma} = \text{DIAG}(\sigma_1^2, \ldots, \sigma_1^2, \sigma_2^2, \ldots, \sigma_2^2, \ldots, \sigma_r^2, \ldots, \sigma_r^2)$, where $r \leq q$, and $t_1 + t_2 + \ldots + t_r = q$. Here, t_u is the number of response variables with variance σ_u^2, for $u = 1, 2, \ldots, r$. That is, the first t_1 responses all have the same variance, the next t_2 responses all have the same variance (different from that of the first t_1 responses), etc. We assume that each of these r variances is independent of the other variances, and denote the collection of r standard deviations as $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, \ldots, \sigma_r]'$. Note that each $\tilde{\sigma}_j$ is one of $\sigma_1, \sigma_2, \ldots, \sigma_r$.

The above error structure is quite general, perhaps seemingly unnecessarily so. However, situations can be imagined in which this degree of generality is needed. Suppose, for example, that two responses are being measured with each of two machines, for a total of $q = 4$ response variables. One might reasonably believe that the first two responses (from the first machine) are measured with the same variance, and the second two responses (from the second machine) are measured with the same variance. A case such as this would require the degree of generality described above. In any event, the generality adds little difficulty to the problem, and includes the simpler cases mentioned above (all equal variances or all different variances).

6.2 The Reference Prior for the Multivariate Calibration Problem

In the above formulation, then, there are $q + pq + p + r$ unknown parameters, namely the intercepts $\boldsymbol{\alpha}'$, the regression coefficients $\mathbf{B}$, the unknown regressor values $\mathbf{x}_0'$ and the standard

deviations $\boldsymbol{\sigma}$. We desire a reference prior for these parameters, with $\mathbf{x}_0$ being the parameter of greatest importance. The theorem below presents a reference prior for the multivariate calibration problem described above. The proof of the theorem is left to the appendix (section A.4).

THEOREM 3 *In the above formulation, the reference prior for the grouping* $(\mathbf{x}_0, (\boldsymbol{\alpha}, \mathbf{B}), \boldsymbol{\sigma})$ *is*

$$\pi(\mathbf{x}_0, \boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\sigma}) \propto \left(\prod_{u=1}^{r} \sigma_u^{-1}\right)(1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{-\frac{1}{2}}, \tag{6.2}$$

where $\mathbf{X}_{\alpha,1}$ *consists of the first* n *rows of* $\mathbf{X}$*, augmented on the left with a column of unity elements, and* $\boldsymbol{\xi}_0$ *is the vector* $\begin{bmatrix} 1 & \mathbf{x}_0' \end{bmatrix}'$*. Note that the reference prior may be written in the form*

$$\pi(\mathbf{x}_0, \boldsymbol{\alpha}, \mathbf{B}, \boldsymbol{\sigma}) \propto \left(\prod_{u=1}^{r} \sigma_u^{-1}\right)\left(1 + \frac{cn}{n+c}(\mathbf{x}_0 - \bar{\mathbf{x}})'(\mathbf{X}_1'\mathbf{X}_1)^{-1}(\mathbf{x}_0 - \bar{\mathbf{x}})\right)^{-\frac{1}{2}},$$

where $\mathbf{X}_1$ *is the matrix consisting of the first* n *rows of* $\mathbf{X}$*, and* $\bar{\mathbf{x}}$ *is a vector containing the means of the* p *regressor variables from the first stage of data collection.*

6.3 Reference Posterior Analysis

6.3.1 The Marginal Reference Posterior of $\mathbf{x}_0$

In this section, we compute the marginal posterior distribution of $\mathbf{x}_0$ corresponding to the reference priors. We begin by considering the prior, as given by (6.2). Since the reference prior does not depend on $\boldsymbol{\alpha}$ or $\mathbf{B}$, it is convenient to re-write the model in the following form:

$$\mathbf{Y} = \mathbf{X}_\alpha \boldsymbol{\Phi} + \mathbf{E},$$

where $\mathbf{X}_\alpha$ and $\boldsymbol{\Phi}$ are defined as follows:

$$\mathbf{X}_\alpha = \begin{bmatrix} \mathbf{x}_{\alpha,1}' \\ \mathbf{x}_{\alpha,2}' \\ \vdots \\ \mathbf{x}_{\alpha,n+c}' \end{bmatrix} = \begin{bmatrix} \mathbf{1} & \mathbf{X} \end{bmatrix},$$

and

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_1 & \boldsymbol{\phi}_2 & \cdots & \boldsymbol{\phi}_q \end{bmatrix} = \begin{bmatrix} \boldsymbol{\alpha}' \\ \mathbf{B} \end{bmatrix}.$$

The likelihood function for the $n+c$ observations is (see, for example, Press (1972), p. 228)

$$\begin{aligned}\mathcal{L} &\propto \left(\prod_{j=1}^{q}\tilde{\sigma}_j^{-1}\right)^{n+c}\exp\left\{-\frac{1}{2}\mathrm{tr}\big((\mathbf{Y}-\mathbf{X}_\alpha\mathbf{\Phi})\mathbf{\Sigma}^{-1}(\mathbf{Y}-\mathbf{X}_\alpha\mathbf{\Phi})'\big)\right\}\\ &\propto \left(\prod_{j=1}^{q}\tilde{\sigma}_j^{-1}\right)^{n+c}\prod_{j=1}^{q}\exp\left\{-\frac{1}{2\tilde{\sigma}_j^2}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)\right\}.\end{aligned}$$

Here, "tr" denotes the trace. The joint posterior distribution for all unknown parameters is thus

$$\begin{aligned}\pi(\mathbf{x}_0,\mathbf{\Phi},\boldsymbol{\sigma}|\mathbf{Y}) &\propto \frac{\left(\prod_{u=1}^{r}\sigma_u^{-1}\right)\left(\prod_{j=1}^{q}\tilde{\sigma}_j^{-(n+c)}\right)}{\sqrt{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha 1}'\mathbf{X}_{\alpha 1})^{-1}\boldsymbol{\xi}_0)}}\prod_{j=1}^{q}\exp\left\{-\frac{1}{2\tilde{\sigma}_j^2}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)\right\}\\ &= \frac{\prod_{u=1}^{r}\sigma_u^{-t_u(n+c)-1}}{\sqrt{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha 1}'\mathbf{X}_{\alpha 1})^{-1}\boldsymbol{\xi}_0}}\prod_{j=1}^{q}\exp\left\{-\frac{1}{2\tilde{\sigma}_j^2}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)\right\}.\end{aligned}$$

Note that, conditional on $\mathbf{x}_0$, posterior inferences for each $\boldsymbol{\phi}_j$ are independent of $\boldsymbol{\phi}_{j'}$ $(j'\neq j)$. Since the primary goal of the calibration problem is the estimation of $\mathbf{x}_0$, we will integrate out the other parameters to obtain the marginal posterior for $\mathbf{x}_0$.

Now

$$\begin{aligned}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j) &= \mathbf{y}_j'\mathbf{y}_j+\boldsymbol{\phi}_j'\mathbf{X}_\alpha'\mathbf{X}_\alpha\boldsymbol{\phi}_j-2\boldsymbol{\phi}_j'\mathbf{X}_\alpha'\mathbf{y}_j\\ &= \mathbf{y}_j'\mathbf{y}_j+\boldsymbol{\phi}_j'\mathbf{X}_\alpha'\mathbf{X}_\alpha\boldsymbol{\phi}_j-2\boldsymbol{\phi}_j'\mathbf{X}_\alpha'\mathbf{X}_\alpha(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j\\ &= (\boldsymbol{\phi}_j-(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j)'\mathbf{X}_\alpha'\mathbf{X}_\alpha(\boldsymbol{\phi}_j-(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j)\\ &\quad-\mathbf{y}_j'\mathbf{X}_\alpha(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j+\mathbf{y}_j'\mathbf{y}_j\\ &= (\boldsymbol{\phi}_j-(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j)'\mathbf{X}_\alpha'\mathbf{X}_\alpha(\boldsymbol{\phi}_j-(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j)\\ &\quad+\mathbf{y}_j'(I-H)\mathbf{y}_j,\end{aligned}$$

where $H=\mathbf{X}_\alpha(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'$ and I is the $(n+c)\times(n+c)$ identity matrix. After integrating out the parameters in $\mathbf{\Phi}$, then, the posterior becomes

$$\pi(\mathbf{x}_0,\boldsymbol{\sigma}|Y)\propto\frac{\prod_{u=1}^{r}\sigma_u^{-(t_u(n+c-p-1)+1)}}{|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^{\frac{q}{2}}\sqrt{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha 1}'\mathbf{X}_{\alpha 1})^{-1}\boldsymbol{\xi}_0}}\prod_{u=1}^{r}\exp\left\{-\frac{1}{2\sigma_u^2}\sum_{\{j:\tilde{\sigma}_j=\sigma_u\}}\mathbf{y}_j'(I-H)\mathbf{y}_j\right\}.$$

We make the change of variables $\sigma_u\to\sigma_u^2$, and let $\boldsymbol{\sigma}^2=[\sigma_1^2,\sigma_2^2,\ldots,\sigma_r^2]$. We now have

$$\pi(\mathbf{x}_0,\boldsymbol{\sigma}^2|Y)\propto\frac{\prod_{u=1}^{r}\left(\sigma_u^2\right)^{-\left(\frac{t_u(n+c-p-1)}{2}+1\right)}}{|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^{\frac{q}{2}}\sqrt{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha 1}'\mathbf{X}_{\alpha 1})^{-1}\boldsymbol{\xi}_0}}\prod_{u=1}^{r}\exp\left\{-\frac{1}{2\sigma_u^2}\sum_{\{j:\tilde{\sigma}_j=\sigma_u\}}\mathbf{y}_j'(I-H)\mathbf{y}_j\right\}.$$

We see now that, conditional on $\mathbf{x}_0$, each σ_u^2 has an inverse gamma distribution. Specifically,

$$\sigma_u^2|\mathbf{x}_0 \sim \mathcal{IG}\left(\frac{t_u(n+c-p-1)}{2}, \frac{2}{\sum_{j:\tilde{\sigma}_j=\sigma_u} \mathbf{y}_j'(I-H)\mathbf{y}_j}\right).$$

Upon integrating out each σ_u^2, then, we have the marginal posterior distribution of $\mathbf{x}_0$:

$$\pi(\mathbf{x}_0|\mathbf{Y}) \propto \frac{1}{|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^{\frac{q}{2}}\sqrt{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha 1}'\mathbf{X}_{\alpha 1})^{-1}\boldsymbol{\xi}_0}} \prod_{u=1}^{r}\left\{\sum_{\{j:\tilde{\sigma}_j=\sigma_u\}} \mathbf{y}_j'(I-H)\mathbf{y}_j\right\}^{-\frac{t_u(n+c-p-1)}{2}}. \quad (6.3)$$

In this expression, the dependence on $\mathbf{x}_0$ is through $\mathbf{X}_\alpha'\mathbf{X}_\alpha$, $\sqrt{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha 1}'\mathbf{X}_{\alpha 1})^{-1}\boldsymbol{\xi}_0}$ and H. Brown (1982) demonstrated the integrability of the likelihood function for $q > 1$; the reference posterior is therefore integrable in such cases as well. Integrability of the reference posterior for the univariate case was addressed in chapter 3.

6.3.2 Gibbs Sampling of the Reference Posterior

The marginal posterior distribution for $\mathbf{x}_0$ was computed analytically in section 6.3.1, up to a normalization constant. From this unnormalized distribution, it is possible to determine the posterior mode for $\mathbf{x}_0$ for a given dataset. However, it is not possible to derive interval (or multidimensional credible set) estimates for $\mathbf{x}_0$. Also, this marginal posterior does not allow for inferences of the other model parameters, which may be of some interest.

In this section, the conditional distributions of all parameters in the reference posterior for the multivariate linear calibration problem are computed. Using these conditional distributions, the Gibbs sampler can be employed to obtain inferences about all of the model parameters. This includes posterior credible set estimates.

It can be seen from the results of section 6.3.1 that the conditional distribution for the regression parameter matrix $\boldsymbol{\Phi}$ is defined by

$$\boldsymbol{\phi}_j|\mathbf{Y}, \boldsymbol{\sigma}, \mathbf{x}_0 \sim N\left((\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}_j, \tilde{\sigma}_j^2(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\right),$$

where $\boldsymbol{\phi}_j$ denotes the j^{th} column of $\boldsymbol{\Phi}$ ($j = 1, 2, \ldots, q$). When invoking the Gibbs sampler, we can simulate the parameters in $\boldsymbol{\Phi}$ in groups, each group representing a column of $\boldsymbol{\Phi}$.

The conditional distributions for the elements of $\boldsymbol{\sigma}$ are defined in term of the variances σ_u^2, which have inverse gamma distributions:

$$\sigma_u^2|\mathbf{Y}, \boldsymbol{\Phi}, \mathbf{x}_0 \sim \mathcal{IG}\left(\frac{t_u(n+c)}{2}, \frac{2}{\sum_{\{j:\tilde{\sigma}_j=\sigma_u\}}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)}\right).$$

It is perhaps easier to implement the Gibbs sampler in terms of

$$\frac{1}{\sigma_u^2}\sum_{\{j:\tilde{\sigma_j}=\sigma_u\}}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)\sim\chi^2_{t_u(n+c)},$$

where $\chi^2_{t_u(n+c)}$ denotes the chi-squared distribution with $t_u(n+c)$ degrees of freedom. When invoking the Gibbs sampler, we simulate the standard deviations in $\boldsymbol{\sigma}$ sequentially.

The elements of $\mathbf{x}_0$ will also be simulated sequentially. The likelihood function for the multivariate linear calibration problem is

$$\mathcal{L}\propto\left(\prod_{j=1}^{q}\tilde{\sigma}_j^{-1}\right)^{n+c}\exp\left\{\sum_{j=1}^{q}-\frac{1}{2\tilde{\sigma}_j^2}(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)'(\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j)\right\}.$$

Note that

$$\mathbf{y}_j-\mathbf{X}_\alpha\boldsymbol{\phi}_j=\begin{bmatrix}\mathbf{y}_{j,1}-\mathbf{1}_{(n\times1)}\alpha_j-\mathbf{X}_1\beta_j\\ \mathbf{y}_{j,2}-\mathbf{1}_{(c\times1)}\alpha_j-\mathbf{X}_2\beta_j\end{bmatrix}.$$

Here $y_{j,1}$ contains the first n elements of the j^{th} column of $\mathbf{Y}$, and $y_{j,2}$ contains the elements in the remaining c rows. Similarly, $\mathbf{X}_1$ contains the first n rows of $\mathbf{X}$ and $\mathbf{X}_2$ contains the remaining c rows. The j^{th} term in the sum in the exponent thus contains two parts, only one of which depends on $\mathbf{x}_0$:

$$\begin{aligned}
&-\frac{1}{2\tilde{\sigma}_j^2}(\mathbf{y}_{j,2}-\mathbf{1}_{(c\times1)}\alpha_j-\mathbf{X}_2\beta_j)'(\mathbf{y}_{j,2}-\mathbf{1}_{(c\times1)}\alpha_j-\mathbf{X}_2\beta_j)\\
&\quad=-\frac{1}{2\tilde{\sigma}_j^2}\sum_{i=n+1}^{n+c}(y_{i,j}-\alpha_j-\mathbf{x}_0'\boldsymbol{\beta})^2\\
&\quad=-\frac{1}{2\tilde{\sigma}_j^2}\sum_{i=n+1}^{n+c}\left(x_{0,k^*}\beta_{k^*,j}-\left(y_{i,j}-\alpha_j-\sum_{\{k:k\neq k^*\}}x_{0,k}\beta_{k,j}\right)\right)^2\\
&\quad=-\frac{\beta_{k^*,j}^2}{2\tilde{\sigma}_j^2}\sum_{i=n+1}^{n+c}(x_{0,k^*}-\mu_{i,k^*,j})^2,
\end{aligned}$$

where

$$\mu_{i,k^*,j}=\frac{1}{\beta_{k^*,j}}\left(y_{i,j}-\alpha_j-\sum_{\{k:k\neq k^*\}}x_{0,k}\beta_{k,j}\right).$$

Here, $k^*\in\{1,2,\ldots,p\}$ and x_{0,k^*} is the element of $\mathbf{x}_0$ to be updated. If we complete the square in x_{0,k^*}, this becomes

$$-\frac{c\beta_{k^*,j}^2}{2\tilde{\sigma}_j^2}(x_{0,k^*}-\mu_{k^*,j})^2+K_1,$$

where K_1 is a term not involving x_{0,k^*}, and

$$\begin{aligned}\mu_{k^*,j} &= \frac{1}{c}\sum_{i=n+1}^{n+c}\mu_{i,k^*,j} \\ &= \frac{1}{c}\sum_{i=n+1}^{n+c}\frac{1}{\beta_{k^*,j}}\left(y_{i,j}-\alpha_j-\sum_{\{k:k\neq k^*\}}x_{0,k}\beta_{k,j}\right) \\ &= \frac{1}{\beta_{k^*,j}}\left(\bar{y}_{\cdot j}-\alpha_j-\sum_{\{k:k\neq k^*\}}x_{0,k}\beta_{k,j}\right),\end{aligned}$$

Apart from K_1, we re-write this as

$$-\frac{1}{2\tilde{\sigma}^2_{k^*,j}}(x_{0,k^*}-\mu_{k^*,j})^2,$$

where

$$\tilde{\sigma}^2_{k^*,j}=\frac{\tilde{\sigma}^2_j}{c\beta^2_{k^*,j}}.$$

Now the exponent in the likelihood function contains q terms involving x_{0,k^*}:

$$\sum_{j=1}^{q}-\frac{1}{2\tilde{\sigma}^2_{k^*,j}}(x_{0,k^*}-\mu_{k^*,j})^2.$$

Again, we complete the square in x_{0,k^*}, yielding

$$-\frac{a}{2}(x_{0,k^*}-\mu_{k^*})^2+K_2,$$

where

$$a=\sum_{j=1}^{q}\frac{1}{\tilde{\sigma}^2_{k^*,j}},$$

$$\mu_{k^*}=\frac{1}{a}\sum_{j=1}^{q}\frac{\mu_{k^*,j}}{\tilde{\sigma}^2_{k^*,j}},$$

and K_2 is a term not involving x_{0,k^*}. In summary, the likelihood function, considered as a function of x_{0,k^*} only (conditional on fixed values of the other parameters), has the form of a normal distribution with mean μ_{k^*} and variance $1/a$. Note that the mean of this normal distribution is a weighted average of the conditional maximum likelihood estimates of x_{0,k^*} from the q response variables, with the weights corresponding to the inverses of the variances

of these estimates. The marginal posterior distribution of x_{0,k^*} has the form of this normal density, multiplied by the term

$$\frac{1}{\sqrt{1+\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0}}$$

from the reference prior. Since this term is less than or equal to one, we can use the normal distribution as a proposal distribution and simulate from the conditional distributions of x_{0,k^*} via the rejection method.

6.4 An Example of Multivariate Linear Calibration

We consider an example of multivariate calibration that appeared in the paper of Brown (1982) (p. 290 and p. 299). In this example, $p = 2$ and $q = 4$. The two regressor variables are the percentage of water (x_1) and the percentage of protein (x_2) in samples of wheat. Twenty-one samples of wheat were studied during the first stage of experimentation. Although the regressors were not controlled by a researcher, they were accurately determined for these 21 samples. (Hence, this is an example of random calibration.) The four responses (y_1, y_2, y_3, y_4) were derived as differences between infrared reflectance measurements at different wavelengths. At a second stage of experimentation, a researcher would be interested in using these responses to determine the unknown values of water content and protein content in a new sample of wheat. The data collected at the first stage of experimentation are displayed in table 6.1.

For the analysis presented here, observation number 18 will be omitted from the first stage data set. We will assume that the responses y_1, y_2, y_3, and y_4 from observation number 18 correspond to unknown values of x_1 and x_2. The goal of the analysis will be to estimate these unknown regressor values. Also, we will assume that each of the four responses has a different error standard deviation (so $r = 4$ and $t_u = 1$ for $u = 1, 2, 3, 4$).

The data from table 6.1 are displayed in the scatterplots shown in figure 6.1. From these plots, it can be seen that the regressor x_1 primarily effects the responses y_3 and y_4 while the regressor x_2 seems to have an effect on all four responses. In our analysis, we will assume the multivariate linear regression model, as in equation (6.1).

6.4.1 The Marginal Posterior for x_0

The marginal posterior distribution for $\mathbf{x}_0$ is displayed in figure 6.2. Note that this distribution was computed using equation (6.3), and has not been normalized. Figure 6.3 displays the

Table 6.1: *Data for the wheat example from Brown (1982). The x variables are the laboratory determinations of water content and protein content in wheat samples. The y variables are differences in infrared reflectance at different frequencies. Observation number 18 is omitted in the analysis presented here; the responses from this observation will be used as second stage data. The resulting first stage sample size is* $n = 20$.

Sample	% Water (x_1)	% Protein (x_2)	y_1	y_2	y_3	y_4
1	9.00	10.73	361	108	96	243
2	8.94	11.05	361	107	98	245
3	9.12	9.86	362	110	94	241
4	9.06	11.41	362	105	94	246
5	10.02	11.57	362	104	70	221
6	10.06	9.42	367	113	75	221
7	9.52	10.93	366	108	82	233
8	9.32	11.61	360	104	86	236
9	9.56	8.82	362	113	85	229
10	9.10	11.81	360	103	90	242
11	9.14	12.33	351	97	88	238
12	9.70	12.93	353	95	73	227
13	9.60	12.69	352	97	77	228
14	10.62	13.13	355	96	52	206
15	10.04	10.41	357	106	69	216
16	10.00	13.57	351	93	69	222
17	9.46	9.26	363	113	88	231
18	8.86	9.82	363	110	101	248
19	9.78	9.46	366	114	79	224
20	9.34	12.85	350	96	85	235
21	10.12	12.81	355	97	63	216

same distribution, in a contour plot. The posterior mode for $\mathbf{x}_0$ is located at approximately $(x_{0,1}, x_{0,2}) = (8.822, 10.117)$. This is quite close to the true value of $(x_{0,1}, x_{0,2}) = (8.86, 9.82)$.

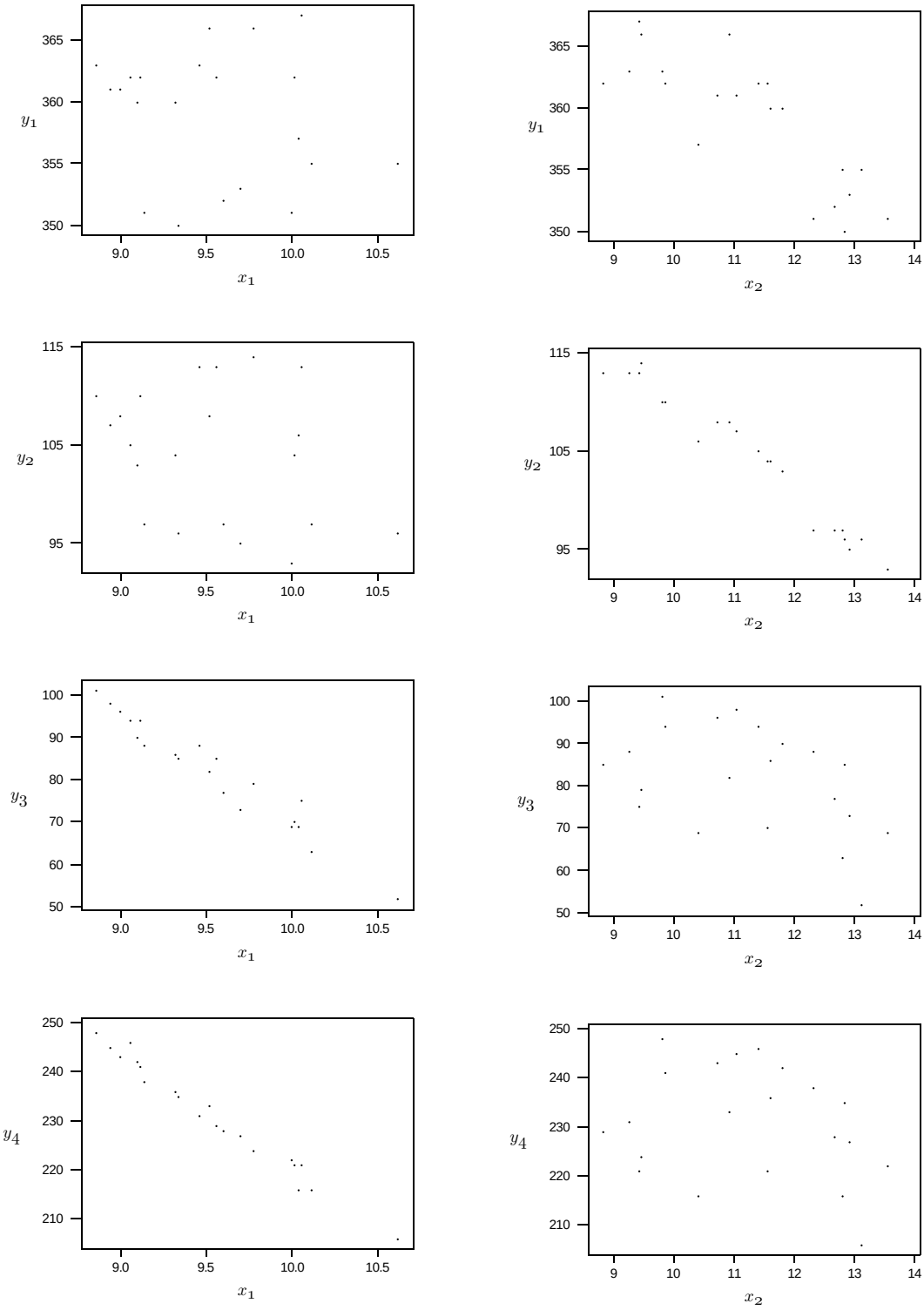

Figure 6.1: *Scatterplots relating the two regressors and four responses from the data in table 6.1.*

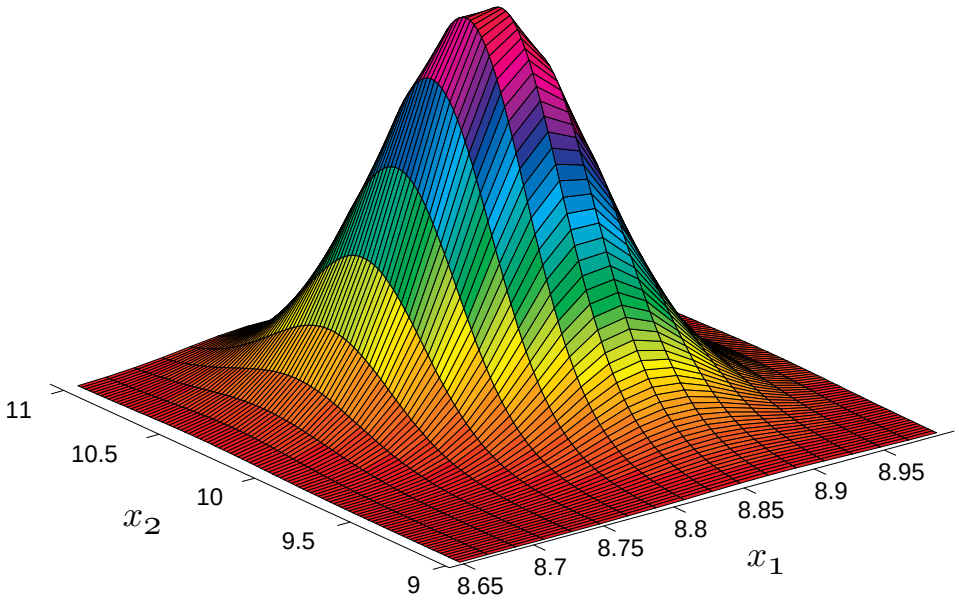

Figure 6.2: *Plot of the unnormalized marginal posterior distribution of* $\mathbf{x}_0$*, for observation number 18 in table 6.1.*

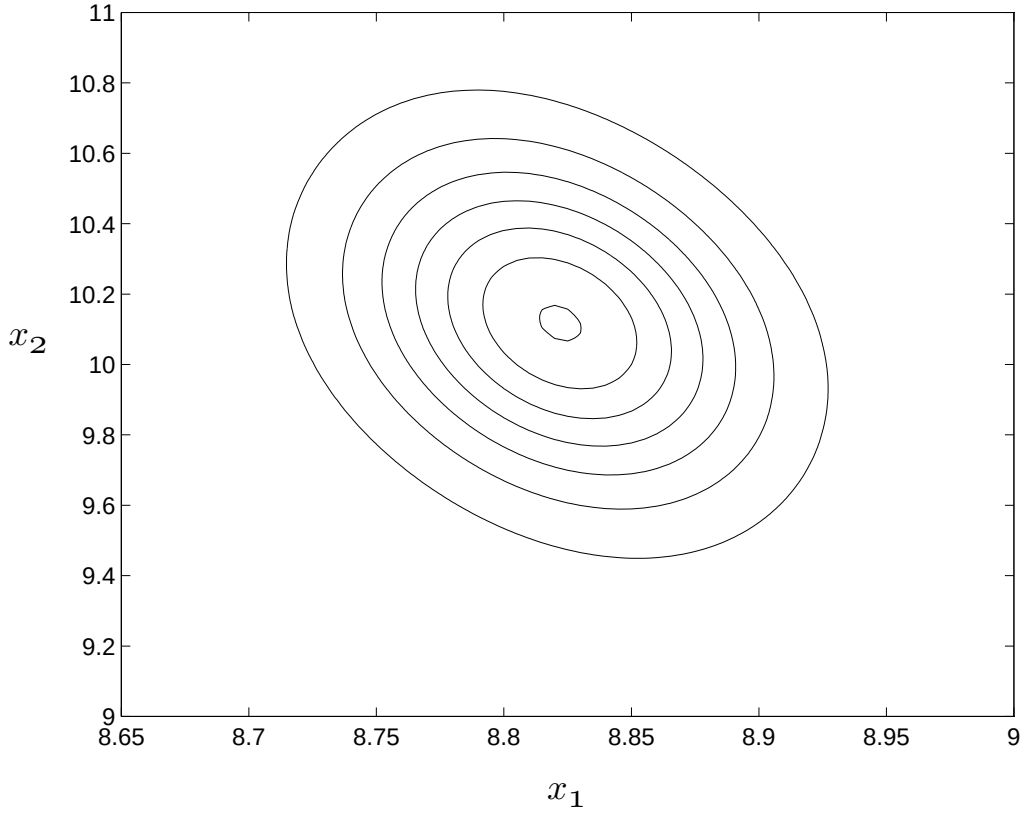

Figure 6.3: *Contour plot for the marginal posterior distribution of* $\mathbf{x}_0$*, for observation number 18 in table 6.1. The mode is located at approximately* $(x_{0,1}, x_{0,2}) = (8.822, 10.117)$.

6.4.2 Posterior Inferences via the Gibbs Sampler

The Gibbs sampler was applied to the current problem, to allow for more precise inferences about the unknown $\mathbf{x}_0$. Ten thousand iterations of the Gibbs sampler yielded the sample displayed in the scatterplot and histograms in figure 6.4. Individual 95% credibility intervals for $x_{0,1}$ and $x_{0,2}$ are displayed in table 6.2.

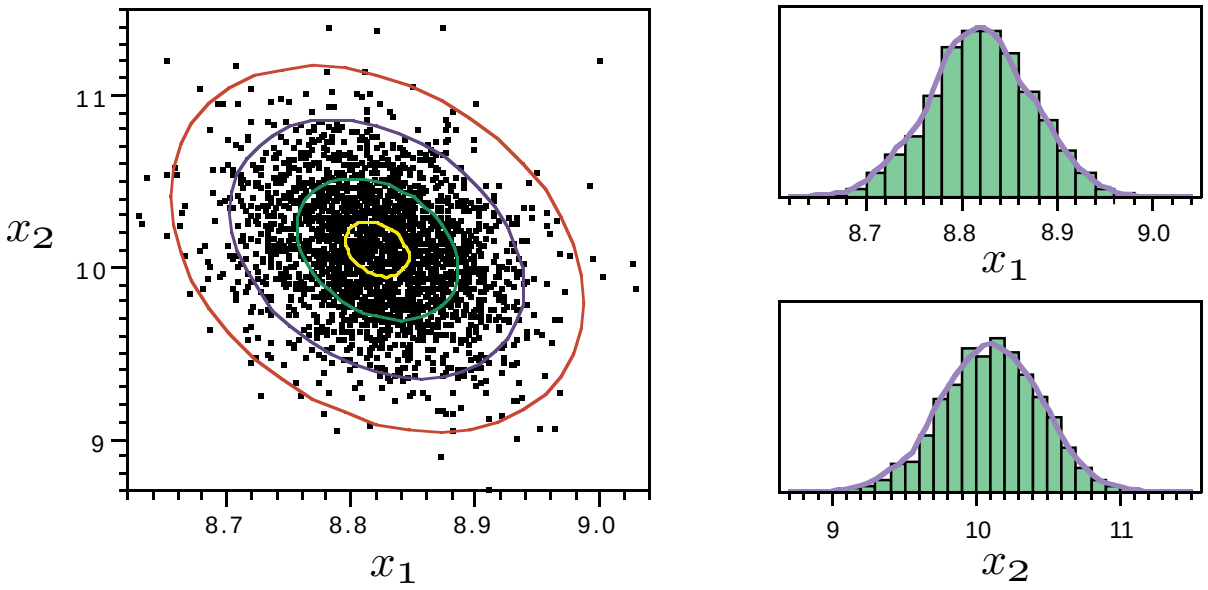

Figure 6.4: *Gibbs sampling results for* $\mathbf{x}_0$*, for observation number 18 in table 6.1. Ten thousand iterations of the Gibbs sampler were used.*

Table 6.2: *Interval estimates for the individual components of* $\mathbf{x}_0$*, for observation number 18 in table 6.1.*

	True	Posterior 0.025	Posterior Median	Posterior 0.975
$x_{0,1}$	8.86	8.71	8.82	8.93
$x_{0,2}$	9.82	9.42	10.11	10.77

Chapter 7

Reference Prior Analysis of the Univariate Polynomial Calibration Problem

7.1 Formulation of the Univariate Polynomial Calibration Problem

In this chapter, another extension of the univariate linear calibration problem is considered. In particular, we develop a reference prior for the *univariate polynomial calibration problem*.

Again, the data are collected in two stages. At the first stage, n pairs (x_i, y_i) are observed. As in the linear case, we assume that the x values are either controlled by the experimenter, or, at least, measured without appreciable error. At the second stage of data collection, c values of the response are observed, all corresponding to a single unknown value of the regressor, x_0. The response variable y is assumed to be related to the regressor x via a polynomial function of order p:

$$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \cdots + \beta_p x_i^p + \epsilon_i, \quad \text{for} \quad i = 1, 2, \ldots, n,$$

and

$$y_i = \alpha + \beta_1 x_0 + \beta_2 x_0^2 + \cdots + \beta_p x_0^p + \epsilon_i, \quad \text{for} \quad i = n+1, n+2, \ldots, n+c.$$

We assume that the errors ϵ_i are independent and identically distributed normal deviates, with mean 0 and standard deviation σ. The model thus contains $p+3$ unknown parameters: the intercept α, the regression coefficients $\beta_1, \beta_2, \ldots, \beta_p$, the error standard deviation σ, and the unknown regressor value x_0. It is convenient to write x_i in place of x_0, for $i =$

$n+1, n+2, \ldots, n+c$, and define

$$\mathbf{x}_i = (x_i, x_i^2, \ldots, x_i^p)' \quad \text{and} \quad \boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$$

so that the model may be re-written as

$$y_i = \alpha + \boldsymbol{\beta}'\mathbf{x}_i + \epsilon_i, \quad \text{for} \quad i = 1, 2, \ldots, n+c.$$

Of primary interest in this problem is the estimation of the unknown regressor value x_0. A feature that distinguishes the polynomial calibration problem from the linear calibration problem is that, since a polynomial function need not be monotonic, more than one value of x_0 may give rise to a particular mean response $\bar{y}_0$.

7.2 A Reference Prior for the Univariate Polynomial Calibration Problem

The univariate polynomial calibration model, in matrix notation, is

$$\mathbf{y} = \mathbf{1}\alpha + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y}$ is a vector containing the $n+c$ responses, $\mathbf{1}$ is an $(n+c)\times 1$ vector of unity elements, $\mathbf{X}$ is the $(n+c)\times p$ matrix whose i^{th} row is $\mathbf{x}_i'$, and $\boldsymbol{\epsilon}$ is the $(n+c)\times 1$ vector of errors. For this model, the likelihood function is

$$\mathcal{L} \propto \frac{1}{\sigma^{n+c}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c}(y_i - \alpha - \boldsymbol{\beta}'\mathbf{x}_i)^2\right\}.$$

We set $\boldsymbol{\zeta}_0 = (1, 2x_0, 3x_0^2, \ldots, px_0^{p-1})'$. The Fisher information matrix for the parameters α, β, σ, and x_0 can be shown to be

$$\mathcal{I} = \frac{1}{\sigma^2}\begin{bmatrix} \mathcal{I}_1 & \mathbf{0} \\ \mathbf{0}' & \mathcal{I}_2 \end{bmatrix},$$

where

$$\mathcal{I}_1 = \begin{bmatrix} c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & c\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\mathbf{x}_0' \\ c\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & n+c & \sum_{i=1}^{n+c}\mathbf{x}_i' \\ c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\mathbf{x}_0 & \sum_{i=1}^{n+c}\mathbf{x}_i & \mathbf{X}'\mathbf{X} \end{bmatrix},$$

and

$$\mathcal{I}_2 = 2(n+c).$$

The sub-matrix $\mathcal{I}_1$ of the information matrix may be re-written as

$$\mathcal{I} = \frac{1}{\sigma^2}\begin{bmatrix} c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\boldsymbol{\xi}_0' \\ c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\boldsymbol{\xi}_0 & \mathbf{X}_\alpha'\mathbf{X}_\alpha \end{bmatrix},$$

where $\boldsymbol{\xi}_0 = (1, \mathbf{x}_0')'$, and $\mathbf{X}_\alpha$ is comprised of the matrix $\mathbf{X}$, augmented on the left with a column of unity elements.

A reference prior for the univariate polynomial calibration problem, as described above, is given in the following theorem. The proof of this theorem is left to the appendix (section A.5).

THEOREM 4 *For the univariate polynomial calibration problem, as described above, the reference prior for the parameter grouping* $(x_0, (\alpha, \boldsymbol{\beta}), \sigma)$ *is*

$$\pi^R(x_0, \alpha, \boldsymbol{\beta}, \sigma) \propto \sigma^{-1}\left(\frac{\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0}{1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}\right)^{\frac{1}{2}},$$

where $\mathbf{X}_{\alpha,1}$ *is the matrix consisting of the first* n *rows of* $\mathbf{X}_\alpha$.

We note the similarity between this reference prior and the result described in theorem 3, for the multivariate linear calibration problem. The distinguishing feature of the reference prior for the polynomial model is the inclusion of the term $\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0$, which arises from the dependence of the elements of $\mathbf{x}_0$ on the individual parameter x_0.

7.3 The Marginal Reference Posterior for x_0

In this section, we derive the marginal posterior distribution for x_0, corresponding to the reference prior described in theorem 4. We also consider the integrability of this posterior.

7.3.1 Derivation of the Marginal Posterior for x_0

The likelihood function for the $n + c$ observations may be written as

$$\mathcal{L} \propto \frac{1}{\sigma^{n+c}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{1}\alpha - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{1}\alpha - \mathbf{X}\boldsymbol{\beta})\right\}.$$

If we let $\boldsymbol{\phi} = (\alpha, \boldsymbol{\beta}')'$, we may re-write this as

$$\mathcal{L} \propto \frac{1}{\sigma^{n+c}}\exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}_\alpha\boldsymbol{\phi})'(\mathbf{y} - \mathbf{X}_\alpha\boldsymbol{\phi})\right\}.$$

Completing the quadratic form in the exponent yields

$$(\mathbf{y}-\mathbf{X}_\alpha\boldsymbol{\phi})'(\mathbf{y}-\mathbf{X}_\alpha\boldsymbol{\phi}) = (\boldsymbol{\phi}-(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y})'\mathbf{X}_\alpha'\mathbf{X}_\alpha(\boldsymbol{\phi}-(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}) + \mathbf{y}'(I-H)\mathbf{y},$$

where $H = \mathbf{X}_\alpha(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'$. Hence

$$\boldsymbol{\phi}|x_0,\sigma,\mathbf{y} \sim N_{p+1}\bigg((\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_\alpha'\mathbf{y}, \sigma^2(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\bigg).$$

After integrating out $\boldsymbol{\phi}$ from the likelihood function, we are left with

$$\frac{1}{\sigma^{n+c-p-1}}|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^{-\frac{1}{2}}\exp\bigg\{-\frac{1}{2\sigma^2}\mathbf{y}'(I-H)\mathbf{y}\bigg\}.$$

We next consider the standard deviation parameter σ, which appears in both the reference prior and the likelihood function. We make the transformation $\sigma \to \sigma^2$. The portion of the posterior involving σ^2 is

$$\big(\sigma^2\big)^{-\frac{n+c-p-1}{2}+1}\exp\bigg\{-\frac{1}{2\sigma^2}\mathbf{y}'(I-H)\mathbf{y}\bigg\},$$

so

$$\sigma^2|x_0,\mathbf{y} \sim \mathcal{I}.\mathcal{G}.\bigg(\frac{n+c-p-1}{2}, \frac{2}{\mathbf{y}'(I-H)\mathbf{y}}\bigg).$$

If we integrate out σ^2, and combine all terms from the reference prior and the likelihood function that involve x_0, we are left with

$$\begin{aligned}\pi(x_0|\mathbf{y}) &\propto \bigg(\frac{\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0}{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}\bigg)^{\frac{1}{2}}|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^{-\frac{1}{2}}\big(\mathbf{y}'(I-H)\mathbf{y}\big)^{-\frac{n+c-p-1}{2}} \\ &= \frac{(\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0)^{\frac{1}{2}}}{|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^{\frac{1}{2}}(1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{\frac{1}{2}}}\big(\mathbf{y}'(I-H)\mathbf{y}\big)^{-\frac{n+c-p-1}{2}}. \qquad (7.1)\end{aligned}$$

We note that the parameter x_0 is involved in this density through $\boldsymbol{\zeta}_0$, $\boldsymbol{\xi}_0$, $\mathbf{X}_\alpha$, and H.

7.3.2 Integrability of the Reference Posterior

We begin by noting that $\mathbf{y}'(I-H)\mathbf{y}$ has the form of an error sum of squares. Specifically, for any particular value of x_0, this is the error sum of squares that would arise from a regression model (fit via least squares) in which that value was the true regressor value corresponding to the second stage responses. If we let $\tilde{y}_i$ denote a predicted response from such a regression

model, and $\hat{y}_i$ denote a predicted response from a similar regression model based on only the first stage data, we have the following inequality:

$$\begin{aligned}\mathbf{y}'(I-H)\mathbf{y} &= \sum_{i=1}^{n+c}(y_i-\tilde{y}_i)^2 \\ &\geq \sum_{i=1}^{n}(y_i-\tilde{y}_i)^2 \\ &\geq \sum_{i=1}^{n}(y_i-\hat{y}_i)^2.\end{aligned}$$

Hence, the term $\mathbf{y}'(I-H)\mathbf{y}$ is bounded below by the error sum of squares for a regression model fit to the first stage data. If we assume that the first stage data do not perfectly fit to the chosen polynomial model (which will be the case, with probability 1, if $n > p+1$), then

$$\left(\mathbf{y}'(I-H)\mathbf{y}\right)^{-\frac{n+c-p-1}{2}}$$

will be bounded above, provided that the exponent is negative. Of course, if $n > p+1$, this provision holds automatically.

We next consider the term $|\mathbf{X}'_\alpha\mathbf{X}_\alpha|$. If we assume that the first stage regressor values constitute a full rank linear model, then $\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1}$ is a (symmetric) positive definite matrix. We let A denote a square root of this matrix. Now

$$\begin{aligned}|\mathbf{X}'_\alpha\mathbf{X}_\alpha| &= |\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1} + c\boldsymbol{\xi}_0\boldsymbol{\xi}'_0| \\ &= |A||(I + cA^{-1}\boldsymbol{\xi}_0\boldsymbol{\xi}'_0A^{-1})||A| \\ &= |A|^2(1 + c\boldsymbol{\xi}'_0(A^{-1})^2\boldsymbol{\xi}_0) \\ &\geq |A|^2.\end{aligned} \tag{7.2}$$

Hence, the term $|\mathbf{X}'_\alpha\mathbf{X}_\alpha|^{\frac{1}{2}}$ in the denominator of (7.1) is bounded away from zero.

The integrability of (7.1) will thus depend on the order of the parameter x_0, as $x_0 \to \pm\infty$. In the term $(\boldsymbol{\zeta}'_0\boldsymbol{\zeta}_0)^{\frac{1}{2}}$ the order of x_0 is $p-1$; in the term $(1 + c\boldsymbol{\xi}'_0(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{\frac{1}{2}}$ the order of x_0 is p; in the term $|\mathbf{X}'_\alpha\mathbf{X}_\alpha|^{\frac{1}{2}}$ the order of x_0 is p (as seen from (7.2)). Hence, the overall order of x_0 in these terms is $-(p+1)$, and the posterior will be integrable for any polynomial model of linear or higher order. Also, the marginal posterior of x_0 will have finite moments up to order $p-1$. We note that for the case $p = 1$, the posterior is integrable but does not have finite mean, as was demonstrated directly for the linear case in chapter 3.

7.4 An Illustrative Example of Polynomial Calibration

7.4.1 Description of the Experimental Data

As an illustration of polynomial calibration, we consider the data set presented by Aitchison and Dunsmore (1975), page 210. These data resulted from an assay of an antibiotic, based on the "clearance circle" technique. The goal of such an experiment is to estimate the concentration of the active constituent in a particular test preparation of the antibiotic.

In a clearance circle assay, the regressor variable x is a precise measure of the concentration of active constituent in a preparation of antibiotic. This concentration is controlled in a laboratory experiment, where it is set at several different values by diluting a known full-strength antibiotic preparation to varying degrees. Each response y_i is obtained by placing a drop of antibiotic solution (of a specified volume) on a petri dish which is uniformly infected with bacteria. The actual response variable y_i is the measured diameter of the circle which has been disinfected by the antibiotic preparation after a specified period of time. It is expected that the diameter of this clearance circle depends on the concentration of the active constituent in an antibiotic solution. Based on the known dilutions of the standard preparation, this regression relationship can be estimated.

At the same time that the clearance circles corresponding to the known antibiotic concentrations are measured, clearance circles are also measured for the test preparation whose unknown antibiotic concentration is of interest.

The data corresponding to the controlled preparations for the present example are displayed in table 7.1. The regressor variable, labeled "dilution", measures the strength of the antibiotic solution. In particular, a solution with a dilution of d is $\frac{1}{d}$ as strong as the full-strength standard preparation. The responses at each dilution correspond to measured clearance circle diameters. A scatterplot of this data set is displayed in figure 7.1. Two different test preparations were also considered; six clearance circles were measured for each of these preparations. The resulting measurements are also displayed in table 7.1.

From figure 7.1, it is evident that the relationship between antibiotic concentration and the diameter of the resulting clearance circles is not linear. In addition, it appears that the variance of the responses is not constant across the different dilutions. To help stabilize the variability, a square root transformation can be applied to the response variable (clearance circle diameter). A transformation of the regressor (antibiotic dilution) can be used to improve the fit of a linear or polynomial regression function.

We consider two transformations of the antibiotic dilution. If the log transformation is

Table 7.1: *Data for the clearance circle assay from Aitchison and Dunsmore (1975). For each known dilution of the antibiotic, four clearance circles were measured. For each of the two test preparations, six clearance circles were measured.*

Dilution	Clearance Circle Diameters (mm)
1	24, 22, 26, 25
2	17, 16, 14, 18
4	10, 14, 12, 14
8	8, 8, 10, 7
16	6, 5, 7, 6
32	4, 4, 5, 5
Test #1	15, 13, 12, 14, 15, 15
Test #2	10, 10, 11, 8, 8, 9

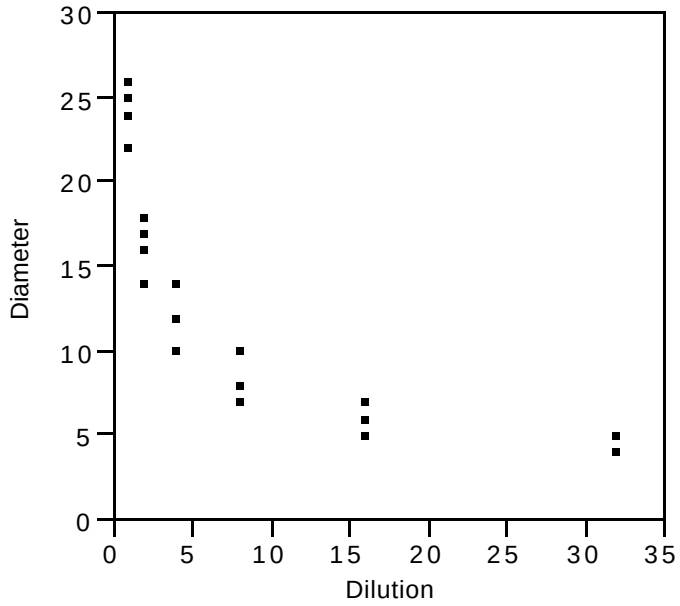

Figure 7.1: *Scatterplot of the control data displayed in table 7.1. From this plot, it is evident that the relationship between dilution and clearance circle diameter is not linear. Also, it appears that the variability in the response variable is dependent on the regressor value.*

applied to this regressor, the resulting relationship between regressor and response is still nonlinear, but a quadratic polynomial seems to fit the data well. A scatterplot of the transformed data is displayed in figure 7.2.

The other transformation of the regressor that we consider is the reciprocal cube root. This is an unusual transformation, but it has the property that a straight line model fits the data well when this transformation is applied. A scatterplot displaying this relationship is shown in figure 7.3.

The data may now be analyzed in two different ways. Of particular interest is the quadratic polynomial model, since that will provide an illustration of the polynomial reference prior analysis. We consider the linear model as well, for comparison.

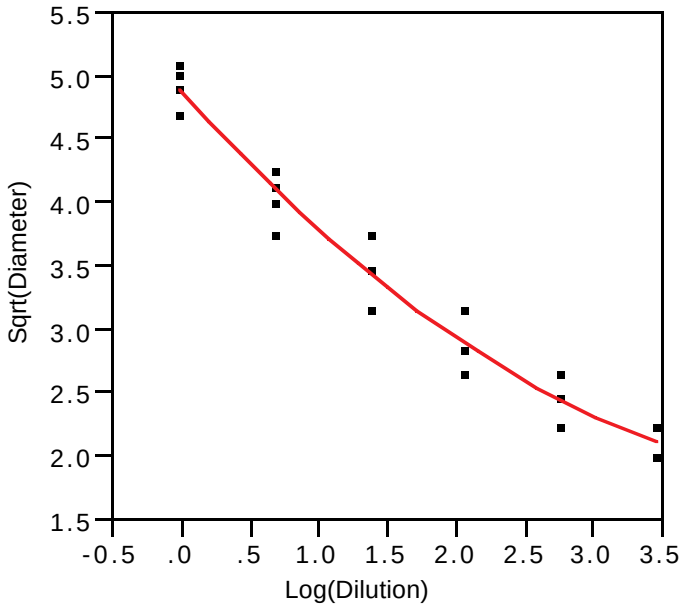

Figure 7.2: *Scatterplot of the transformed control data from table 7.1. The response is transformed via the square root transformation, and the regressor is transformed via the log transformation. The line displayed on the plot is the quadratic polynomial fit via least-squares. This polynomial fits the data quite well.*

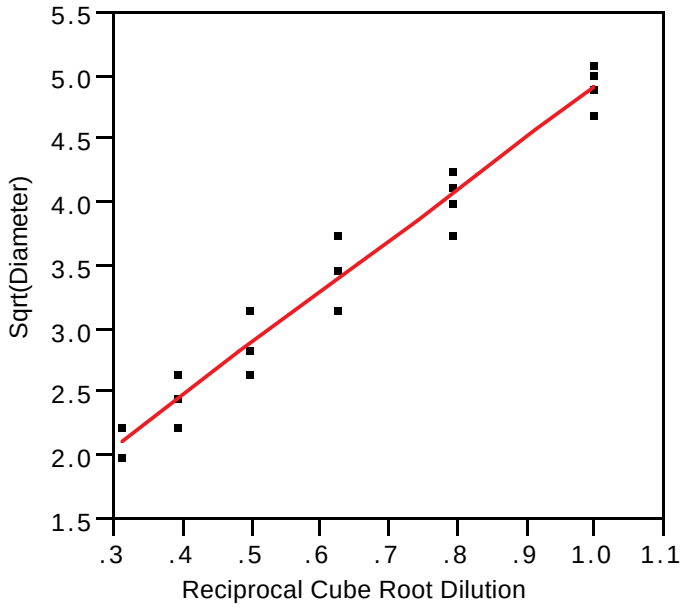

Figure 7.3: *Scatterplot of the transformed control data from table 7.1. The response is transformed via the square root transformation, and the regressor is transformed via the reciprocal cube root transformation. The line displayed on the plot is the straight line fit via least-squares. This line fits the data quite well.*

7.4.2 Posterior Analyses for the Clearance Circle Assay Example

Two posterior analyses of the data displayed in table 7.1 are considered. First, we consider the quadratic polynomial model, relating $\sqrt{y}$ and $\log x$. (In order to eliminate collinearity between the linear and quadratic portions of the regression function, the regressor values were centered about the mean of the transformed control dilutions.) The reference prior for $\log x$, as defined by theorem 4, is displayed in figure 7.4. Of primary importance in the analysis of this data is the estimation of the concentrations of active constituent in the two test preparations. Hence, it will be important to transform the marginal posteriors for $\log x_0$, resulting from the reference prior analyses, back to the original dilution scale (x_0). Before doing this, however, we consider the form of the marginal reference posteriors for $\log x_0$. The densities are displayed in figures 7.5 and 7.6. We note that in both cases, the densities are bimodal – this is due to the fact that for a quadratic polynomial model, we may have two regressor values corresponding to a given mean response.

From figure 7.2, we can see that within the range of standard preparations considered, the relationship between the (transformed) regressor and the response is monotonic. When the marginal posterior densities for $\log x_0$ are transformed to the original dilution scale (x_0), the shorter of the two modes of each posterior is transformed to a location well beyond the range of the controlled regressor values. Since the model is not likely to be reliable outside the range of the controlled regressor values, we truncate the range of the posterior densities. The truncated range is taken to be $(0, 40)$; this extends only slightly beyond the range of the controlled regressor values. The resulting re-normalized marginal posterior densities for x_0, for the two test preparations, are displayed in figures 7.7 and 7.8.

We next consider the straight line model, as suggested by figure 7.3. Again, reference prior analyses are presented, for the two test preparations. The resulting marginal posteriors for x_0 are displayed in figures 7.9 and 7.10. (We note that these posterior densities have been transformed from the densities for $x_0^{-\frac{1}{3}}$, which were the direct result of the reference prior analyses.)

For comparison, the results of all of the above analyses are summarized in table 7.2. From this table, we can see that the two types of analyses yield similar, though not identical, results. Also included in table 7.2 are results obtained from the classical calibration procedure, as described in chapter 1. In particular, (1.1) is easily adapted to the case $c > 1$; this was used to generate interval estimates. We can see from the table that the classical result produces intervals that are significantly wider than those from both the liner and the quadratic Bayesian reference prior analyses.

Table 7.2: *Summary of the inferences for the clearance circle assay example. Posterior modes for the Bayesian reference prior analyses are presented, along with 90% highest posterior density (HPD) interval estimates. Classical results are also presented; after construction, the "fiducial" interval estimates based on the linear and quadratic models were transformed back to the original dilution scale.*

	Results for Test Prep. #1			Results for Test Prep. #2		
	Lower	Est.	Upper	Lower	Est.	Upper
Reference Prior Analysis, Linear Model	2.38	2.78	3.25	5.08	6.22	7.71
Reference Prior Analysis, Quadratic Model	2.41	2.85	3.39	5.10	6.40	8.12
Classical Analysis, Linear Model	2.40	2.81	3.30	5.20	6.34	7.86
Classical Analysis, Quadratic Model	2.45	2.90	3.48	5.19	6.46	8.13

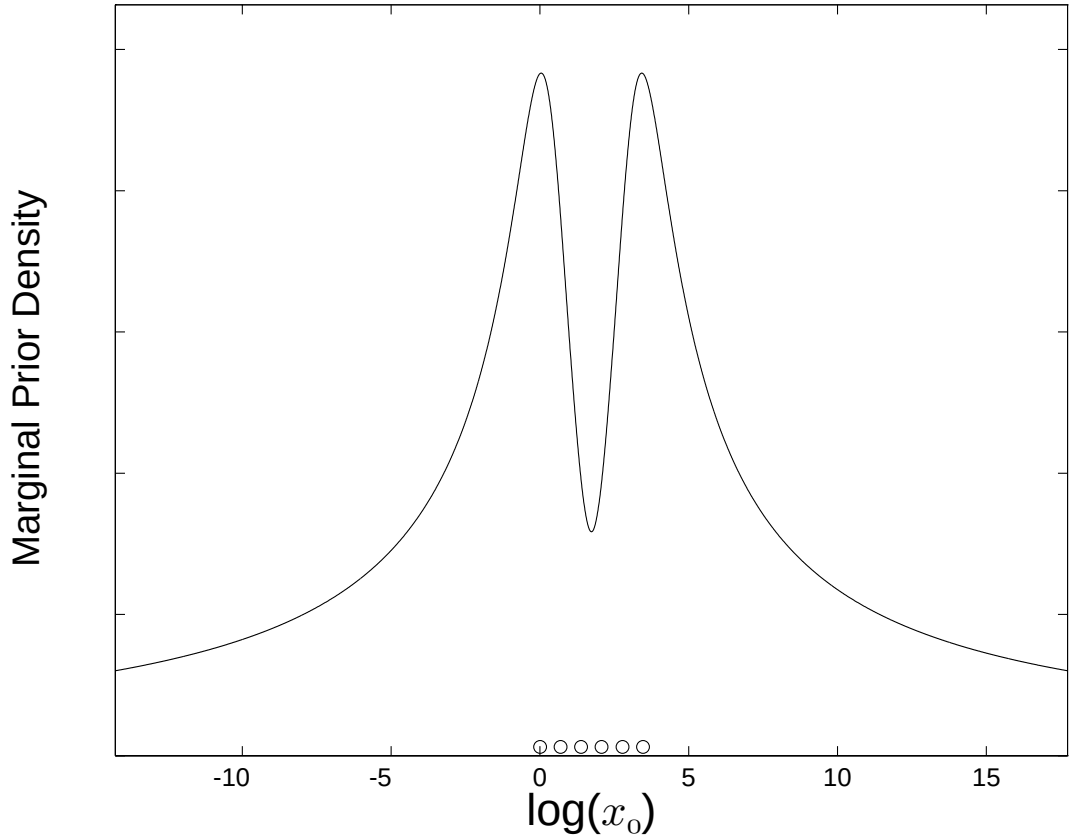

Figure 7.4: *Reference prior for the quadratic model, for the clearance circle assay example. The circles represent the locations of the (transformed) regressor settings; each circle corresponds to four responses.*

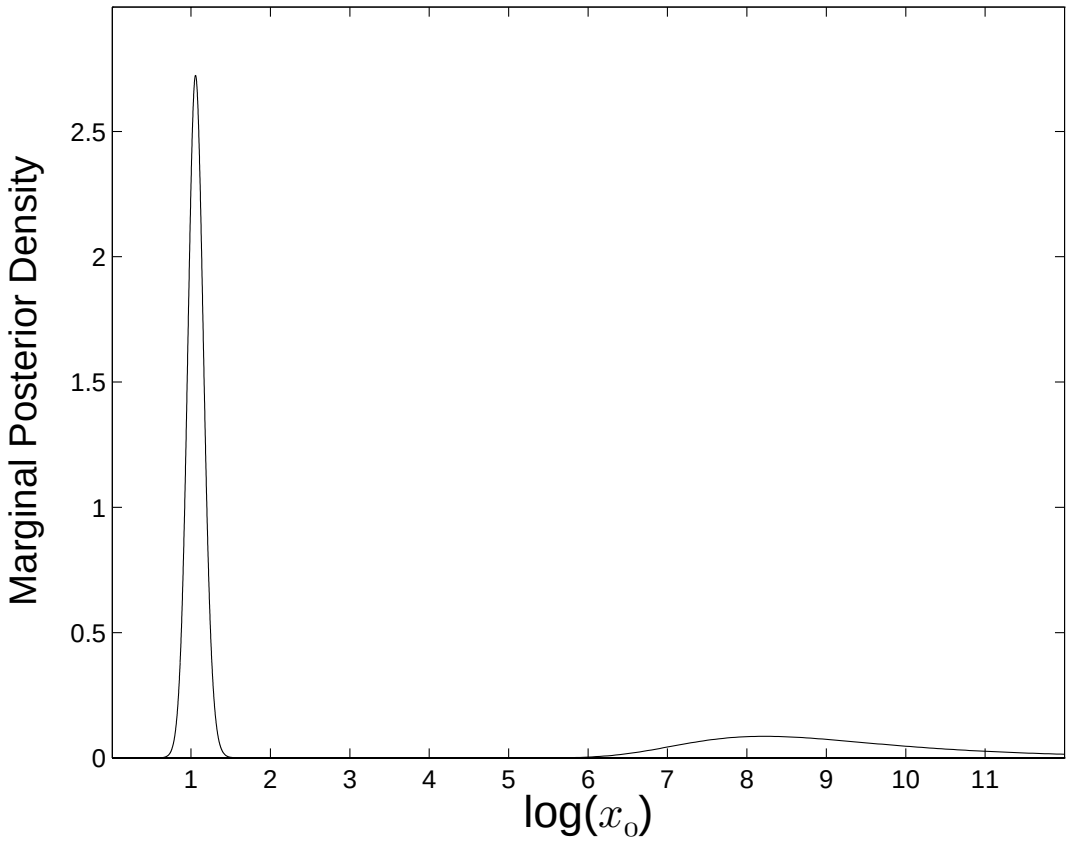

Figure 7.5: *Marginal reference posterior for* $\log x_0$*, from the quadratic polynomial model. This density corresponds to test preparation #1.*

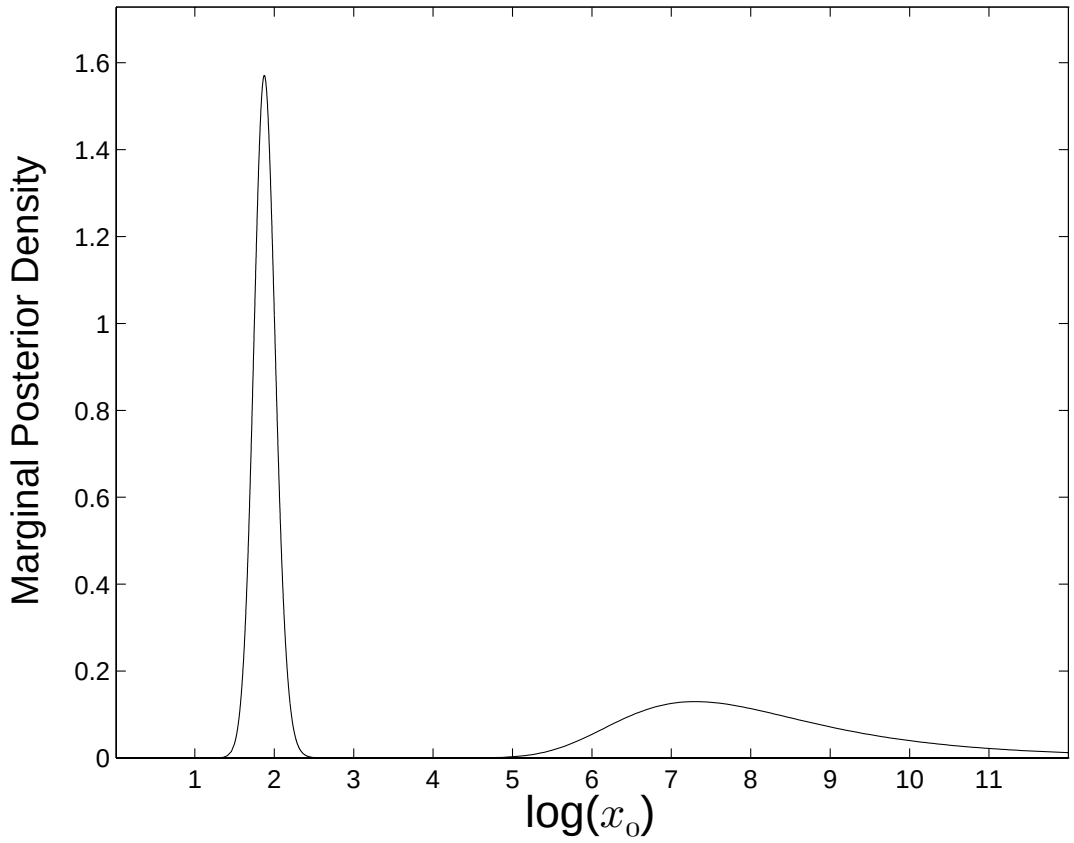

Figure 7.6: *Marginal reference posterior for* $\log x_0$*, from the quadratic polynomial model. This density corresponds to test preparation #2.*

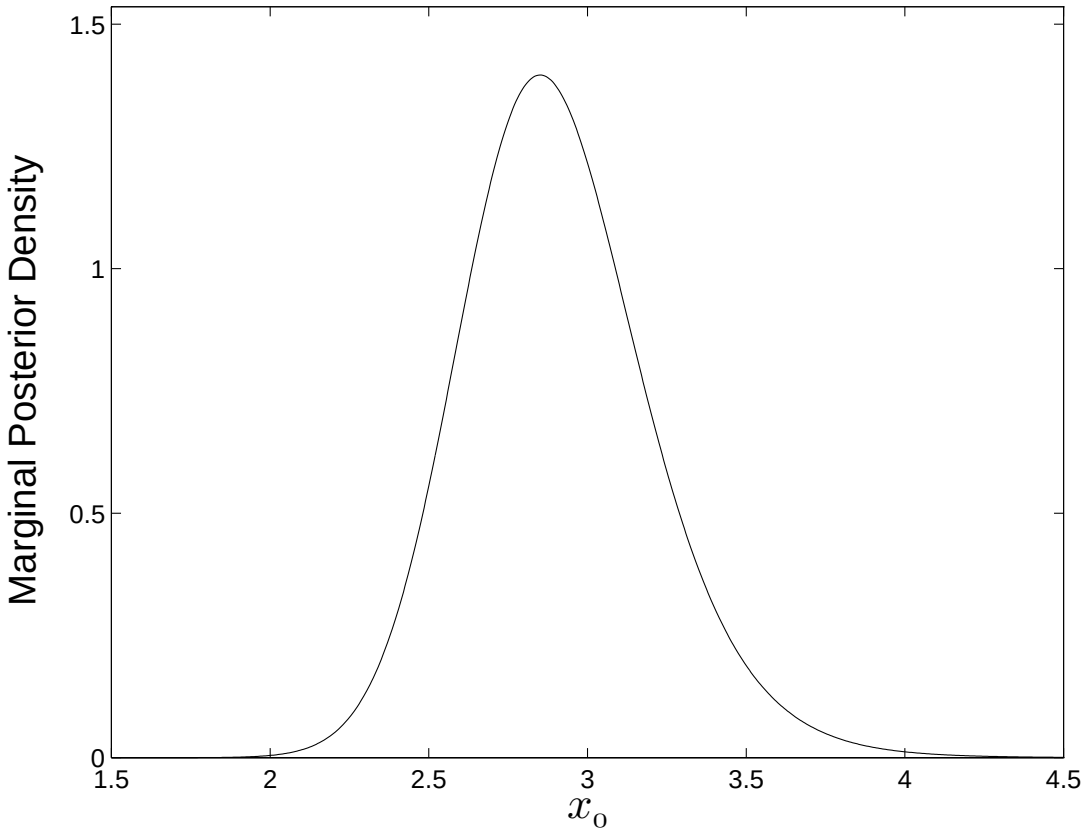

Figure 7.7: *Marginal reference posterior for x_0, from the quadratic polynomial model. This density corresponds to test preparation #1. The range of x_0 has been restricted to include only values near the range of first stage dilutions.*

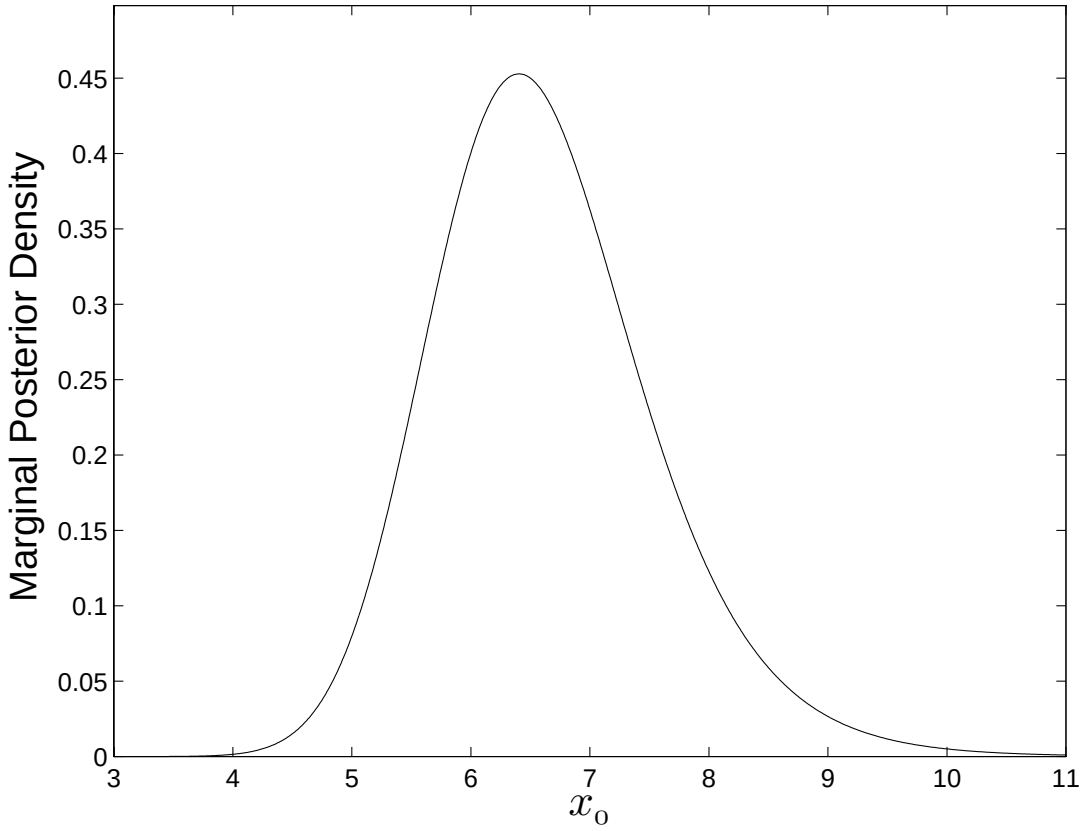

Figure 7.8: *Marginal reference posterior for x_0, from the quadratic polynomial model. This density corresponds to test preparation #2. The range of x_0 has been restricted to include only values near the range of first stage dilutions.*

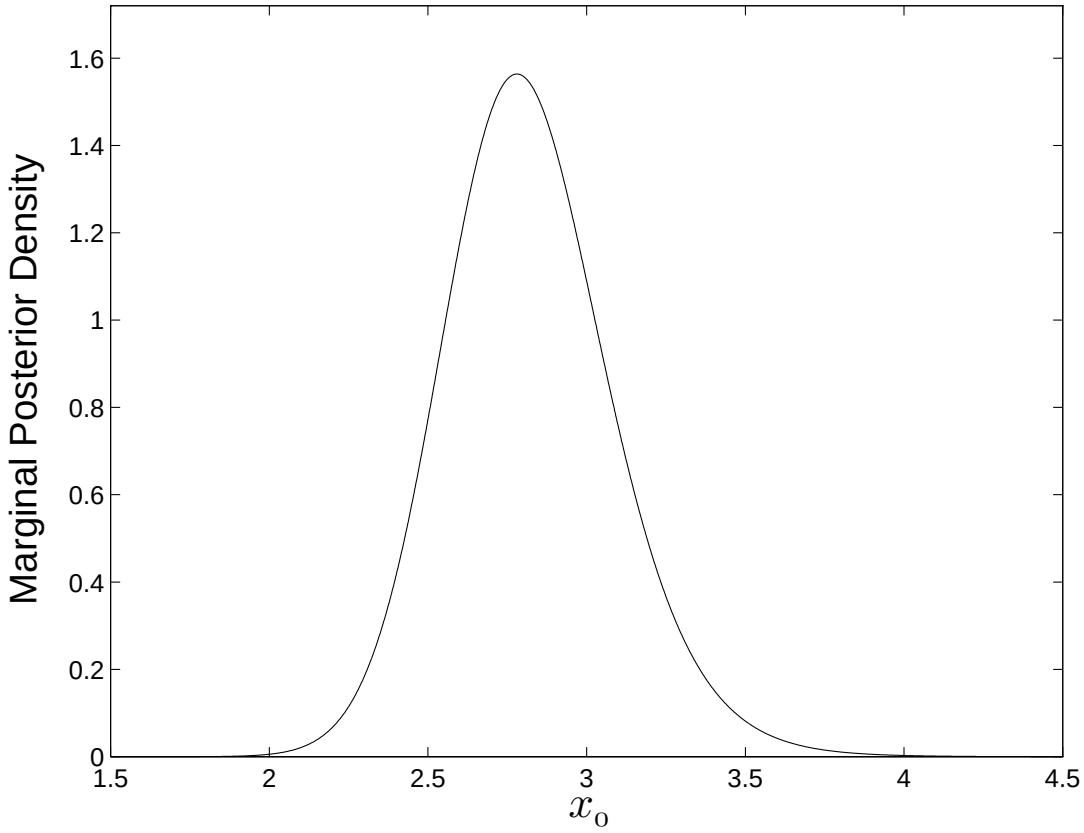

Figure 7.9: *Marginal reference posterior for x_0, from the quadratic polynomial model. This density corresponds to test preparation #1.*

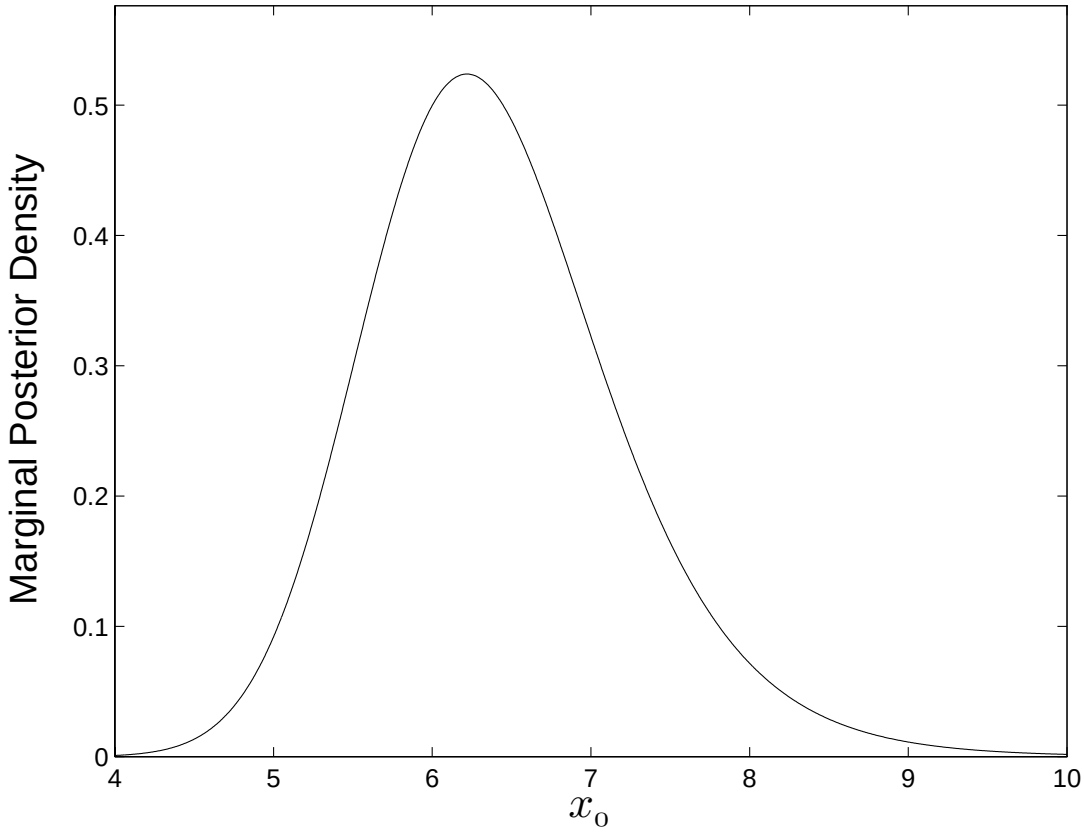

Figure 7.10: *Marginal reference posterior for x_0, from the quadratic polynomial model. This density corresponds to test preparation #2.*

Appendix

A.1 The Reference Prior Algorithm

The algorithm for constructing reference priors is described in detail in Berger and Bernardo (1992). The details of this algorithm are presented in this section.

Suppose that inferences are desired about a vector of parameters $\boldsymbol{\theta}$, which is partitioned into m groups:

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \ldots, \boldsymbol{\theta}_{(m)}).$$

Each of the m groups contains one or more parameters, and it is assumed that the groups are listed in decreasing order of inferential importance – that is, that $\boldsymbol{\theta}_{(1)}$ contains the parameters of greatest interest, and $\boldsymbol{\theta}(m)$ contains the parameters of least interest. Define $\boldsymbol{\theta}_{[i]} = (\boldsymbol{\theta}_{(1)}, \boldsymbol{\theta}_{(2)}, \ldots, \boldsymbol{\theta}_{(i)})$, and $\boldsymbol{\theta}_{[-i]} = (\boldsymbol{\theta}_{(i+1)}, \boldsymbol{\theta}_{(i+2)}, \ldots, \boldsymbol{\theta}_{(m)})$. Let $\boldsymbol{\Theta}$ denote the domain of $\boldsymbol{\theta}$.

We let $\mathcal{I}$ denote the Fisher information matrix, $S = \mathcal{I}^{-1}$, and write S in partitioned form as

$$S = \begin{bmatrix} S_{1,1} & \cdots & S_{1,m} \\ \vdots & \ddots & \vdots \\ S_{m,1} & \cdots & S_{m,m} \end{bmatrix}.$$

We define, for $i = 1, 2, \ldots, m$, S_i and H_i as follows:

$$S_i = \begin{bmatrix} S_{1,1} & \cdots & S_{1,i} \\ \vdots & \ddots & \vdots \\ S_{i,1} & \cdots & S_{i,i} \end{bmatrix},$$

$$H_i = S_i^{-1} = \begin{bmatrix} H_{1,1} & \cdots & H_{1,i} \\ \vdots & \ddots & \vdots \\ H_{i,1} & \cdots & H_{i,i} \end{bmatrix}.$$

Further, we define $h_i(\boldsymbol{\theta})$ to be the determinant of $H_{i,i}$.

Suppose that $\boldsymbol{\Theta}^1 \subset \boldsymbol{\Theta}^2 \subset \cdots$ is a nested sequence of compact subsets of $\boldsymbol{\Theta}$, with $\bigcup_{l=1}^{\infty} \boldsymbol{\Theta}^l = \boldsymbol{\Theta}$. We define

$$\boldsymbol{\Theta}^l_{\boldsymbol{\theta}_{[i]}} = \{\boldsymbol{\theta}_{(i+1)} : (\boldsymbol{\theta}_{[i]}, \boldsymbol{\theta}_{(i+1)}, \boldsymbol{\theta}^*_{[-(i+1)]}) \in \boldsymbol{\Theta}^l \text{ for some } \boldsymbol{\theta}^*_{[-(i+1)]}\}.$$

We also define $I_\Omega(x) = 1$ if $x \in \Omega$, $I_\Omega(x) = 0$ if $x \notin \Omega$, and $I^l(x\big|\boldsymbol{\theta}_{[i]}) = I_{\boldsymbol{\Theta}^l_{\boldsymbol{\theta}_{[i]}}}(x)$.

To implement the reference prior algorithm, we begin by defining

$$\pi^l_m(\boldsymbol{\theta}_{[-(m-1)]}\big|\boldsymbol{\theta}_{[m-1]}) = \frac{|h_m(\boldsymbol{\theta})|^{\frac{1}{2}} I^l(\boldsymbol{\theta}_{(m)}\big|\boldsymbol{\theta}_{[m-1]})}{\int |h_m(\boldsymbol{\theta})|^{\frac{1}{2}} I^l(\boldsymbol{\theta}_{(m)}\big|\boldsymbol{\theta}_{[m-1]})\ d\boldsymbol{\theta}_{(m)}},$$

where the integral is over the range $\boldsymbol{\Theta}^l_{\boldsymbol{\theta}_{[m-1]}}$. We then proceed iteratively, defining, for $i = m-1, m-2, \ldots, 1$, π^l_i in terms of the previously defined π^l_j $(j = i+1, i+2, \ldots, m)$. The definition is as follows:

$$\pi^l_i(\boldsymbol{\theta}_{[-(i-1)]}\big|\boldsymbol{\theta}_{[i-1]}) = \frac{\pi^l_{i+1}(\boldsymbol{\theta}_{[-i]}\big|\boldsymbol{\theta}_{[i]}) \exp\{\frac{1}{2} E^l_{\boldsymbol{\theta}}[\log|h_i|\big|\boldsymbol{\theta}_{[i]}]\} I^l(\boldsymbol{\theta}_{(i)}\big|\boldsymbol{\theta}_{[i-1]})}{\int \exp\{\frac{1}{2} E^l_{\boldsymbol{\theta}}[\log|h_i|\big|\boldsymbol{\theta}_{[i]}]\} I^l(\boldsymbol{\theta}_{(i)}\big|\boldsymbol{\theta}_{[i-1]})\ d\boldsymbol{\theta}_{(i)}},$$

where the integral is over the range $\boldsymbol{\Theta}^l_{\boldsymbol{\theta}_{[i-1]}}$, and where

$$E^l_{\boldsymbol{\theta}}[g(\boldsymbol{\theta})\big|\boldsymbol{\theta}_{[i]}] = \int_{\{\boldsymbol{\theta}_{[-i]} : (\boldsymbol{\theta}_{[i]}, \boldsymbol{\theta}_{[-i]}) \in \boldsymbol{\Theta}^l\}} g(\boldsymbol{\theta}) \pi^l_{i+1}(\boldsymbol{\theta}_{[-i]}\big|\boldsymbol{\theta}_{[i]})\ d\boldsymbol{\theta}_{[-i]}.$$

In the above, we interpret $\boldsymbol{\theta}_{[0]}$ as vacuous, and write $\pi^l(\boldsymbol{\theta}) = \pi^l_1(\boldsymbol{\theta}_{[-0]}\big|\boldsymbol{\theta}_{[0]})$. Finally, we define the reference prior as

$$\pi^R(\boldsymbol{\theta}) = \lim_{l \to \infty} \frac{\pi^l(\boldsymbol{\theta})}{\pi^l(\boldsymbol{\theta}^*)},$$

provided it exists, where $\boldsymbol{\theta}^*$ is some point in $\boldsymbol{\Theta}^1$.

A.2 Development of the Reference Priors for the Univariate Linear Calibration Problem with a Single Standard Deviation Parameter

As noted in section 3.2, the information matrix for the parameter vector $(x_0, \alpha, \beta, \sigma)$ is

$$\mathcal{I} = \frac{1}{\sigma^2}\begin{bmatrix} c\beta^2 & c\beta & c\beta x_0 & 0 \\ c\beta & n+c & \mathfrak{s}_x^* & 0 \\ c\beta x_0 & \mathfrak{s}_x^* & \mathfrak{s}_{xx}^* & 0 \\ 0 & 0 & 0 & 2(n+c) \end{bmatrix},$$

where

$$\mathfrak{s}_x^* = \sum_{i=1}^{n+c} x_i \quad \text{and} \quad \mathfrak{s}_{xx}^* = \sum_{i=1}^{n+c} x_i^2.$$

As before, we have written x_i in place of x_0, for $i = n+1, n+2, \ldots, n+c$. The inverse of this information matrix is

$$\mathcal{S} = \sigma^2 \begin{bmatrix} \frac{1}{c n \beta^2 A}(cB^2 + (n+c)A) & -\frac{1}{n\beta A}(A + \mathfrak{s}_x B) & \frac{B}{\beta A} & 0 \\ -\frac{1}{n\beta A}(A + \mathfrak{s}_x B) & \frac{1}{nA}(A + \mathfrak{s}_x^2) & -\frac{1}{A}\mathfrak{s}_x & 0 \\ \frac{B}{\beta A} & -\frac{1}{A}\mathfrak{s}_x & \frac{n}{A} & 0 \\ 0 & 0 & 0 & \frac{1}{2(n+c)}, \end{bmatrix}$$

where

$$\mathfrak{s}_x = \sum_{i=1}^{n} x_i, \qquad \mathfrak{s}_{xx} = \sum_{i=1}^{n} x_i^2, \qquad A = n\mathfrak{s}_{xx} - \mathfrak{s}_x^2, \qquad \text{and} \qquad B = \mathfrak{s}_x - n x_0.$$

The derivation of the reference prior depends on the definition of the sets $\boldsymbol{\Theta}^l$, described previously. For the univariate linear calibration problem, we take these sets to be Cartesian products of compact intervals in the appropriate spaces. Specifically, for $\boldsymbol{\Theta}^l$ we take $\alpha \in [\alpha_1, \alpha_2]$, $\beta \in [\beta_1, \beta_2]$, $\sigma \in [\sigma_1, \sigma_2]$, and $x_0 \in [x_{0,1}, x_{0,2}]$, where $\alpha_2, \beta_2, \sigma_1, \sigma_2$, and $x_{0,2}$ are positive real numbers (with $\sigma_1 < 1, \sigma_2 > 1$), and α_1, β_1, and $x_{0,1}$ are negative real numbers. Although this notation does not explicitly express the dependence of the intervals on l, as $l \to \infty$, we let $\alpha_2, \beta_2, \sigma_2$, and $x_{0,2}$ increase to $+\infty$, α_1, β_1 and $x_{0,1}$ decrease to $-\infty$, and σ_1 decrease to 0. All integrals evaluated in the derivation of the reference priors will be over the appropriate compact intervals.

Here, we will show the derivations of the reference priors for the parameter groupings $(x_0, \alpha, \beta, \sigma)$ and $(x_0, (\alpha, \beta, \sigma))$. The derivations of the reference priors for other parameter groupings are similar. In the derivations shown below, all quantities labeled with K's

represent either constants (possibly depending on the ranges of the parameters in $\mathbf{\Theta}^l$), or functions depending only on indicated parameters (e.g., $K_1(x_0)$ is a function depending on the parameter x_0 and constant terms, but not on other parameters).

A.2.1 The Reference Prior for the Parameter Grouping $(x_0, \alpha, \beta, \sigma)$

Note that in this case, $m = 4$. The following quantities can be computed directly from the information matrix (or it's inverse):

$$h_1 = \frac{cn\beta^2 A}{\sigma^2(cB^2 + (n+c)A)} = K_1(x_0)\frac{\beta^2}{\sigma^2};$$

$$h_2 = \frac{nA(cB^2 + (n+c)A)}{\sigma^2[(cB^2 + (n+c)A)(A + \mathfrak{s}_x^2) - c(A + \mathfrak{s}_x B)^2]} = K_2(x_0)\frac{1}{\sigma^2};$$

$$h_3 = \frac{\mathfrak{s}_{xx}^*}{\sigma^2};$$

$$h_4 = \frac{2(n+c)}{\sigma^2}.$$

Step 1. Note that

$$\begin{aligned}\int |h_4|^{\frac{1}{2}} I_{[\sigma_1,\sigma_2]}(\sigma)\, d\sigma &= \int_{\sigma_1}^{\sigma_2} (2(n+c))^{\frac{1}{2}} \sigma^{-1}\, d\sigma \\ &= (2(n+c))^{\frac{1}{2}}(\log\sigma_2 - \log\sigma_1),\end{aligned}$$

so

$$\begin{aligned}\pi_4^l(\sigma|x_0,\alpha,\beta) &= \frac{|h_4|^{\frac{1}{2}} I_{[\sigma_1,\sigma_2]}(\sigma)}{\int |h_4|^{\frac{1}{2}} I_{[\sigma_1,\sigma_2]}(\sigma)\, d\sigma} \\ &= \frac{1}{\sigma(\log\sigma_2 - \log\sigma_1)} I_{[\sigma_1,\sigma_2]}(\sigma) \\ &= K_3\sigma^{-1} I_{[\sigma_1,\sigma_2]}(\sigma).\end{aligned}$$

Step 2. Now

$$\begin{aligned}E_{\boldsymbol{\theta}}^l[\log|h_3|\big|x_0,\alpha,\beta] &= \int \log\left|\frac{\mathfrak{s}_{xx}^*}{\sigma^2}\right| K_3\sigma^{-1} I_{[\sigma_1,\sigma_2]}(\sigma)\, d\sigma \\ &= \log|\mathfrak{s}_{xx}^*| - K_3(\log\sigma_2)^2 + K_3(\log\sigma_1)^2 \\ &= K_4(x_0),\end{aligned}$$

and

$$\int \exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[log|h_3|\big|x_0,\alpha,\beta]\right\}I_{[\beta_1,\beta_2]}(\beta)\ d\beta = \exp\left\{\frac{1}{2}K_4(x_0)\right\}(\beta_2-\beta_1),$$

so

$$\begin{aligned}\pi^l_3(\beta,\sigma|x_0,\alpha) &= \frac{\pi^l_4(\sigma|x_0,\alpha,\beta)\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_3|\big|x_0,\alpha,\beta]\right\}I_{[\beta_1,\beta_2]}(\beta)}{\int \exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_3|\big|x_0,\alpha,\beta]\right\}I_{[\beta_1,\beta_2]}(\beta)\ d\beta}\\ &= \frac{K_3}{\beta_2-\beta_1}\sigma^{-1}I_{[\beta_1,\beta_2]}(\beta)I_{[\sigma_1,\sigma_2]}(\sigma).\end{aligned}$$

Step 3. Next, note that

$$\begin{aligned}E^l_{\boldsymbol{\theta}}[\log|h_2|\big|x_0,\alpha] &= \iint \log\left|\frac{K_2(x_0)}{\sigma^2}\right|\frac{K_3}{\beta_2-\beta_1}\sigma^{-1}I_{[\sigma_1,\sigma_2]}(\sigma)I_{[\beta_1,\beta_2]}(\beta)d\sigma d\beta\\ &= K_3[(\log\sigma_2-\log\sigma_1)\log|K_2(x_0)|] - 2K_3(\log\sigma_2)^2 + 2K_3(\log\sigma_1)^2\\ &= K_5(x_0),\end{aligned}$$

so

$$\int \exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_2|\big|x_0,\alpha]\right\}I_{[\alpha_1,\alpha_2]}(\alpha)\ d\alpha = (\alpha_2-\alpha_1)\exp\left\{\frac{1}{2}K_5(x_0)\right\},$$

and

$$\begin{aligned}\pi^l_2(\alpha,\beta,\sigma|x_0) &= \frac{\pi^l_3(\beta,\sigma|x_0,\alpha)\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_2|\big|x_0,\alpha]\right\}I_{[\alpha_1,\alpha_2]}(\alpha)}{\int \exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_2|\big|x_0,\alpha\right\}I_{[\alpha_1,\alpha_2]}(\alpha)\ d\alpha}\\ &= \frac{K_3}{(\alpha_2-\alpha_1)(\beta_2-\beta_1)}\sigma^{-1}I_{[\alpha_1,\alpha_2]}(\alpha)I_{[\beta_1,\beta_2]}(\beta)I_{[\sigma_1,\sigma_2]}(\sigma)\\ &= K_6\sigma^{-1}I_{[\alpha_1,\alpha_2]}(\alpha)I_{[\beta_1,\beta_2]}(\beta)I_{[\sigma_1,\sigma_2]}(\sigma).\end{aligned}$$

Step 4. In this final step,

$$\begin{aligned}
E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0] &= \iiint \log\left|K_1(x_0)\frac{\beta^2}{\sigma^2}\right| K_6\sigma^{-1}I_{[\sigma_1,\sigma_2]}(\sigma)I_{[\beta_1,\beta_2]}(\beta)I_{[\alpha_1,\alpha_2]}(\alpha)d\sigma d\beta d\alpha \\
&= K_6(\alpha_2-\alpha_1)\Bigg[(\beta_2-\beta_1)(\log\sigma_2-\log\sigma_1)\log K_1(x_0) + \\
&\quad 2(\log\sigma_2-\log\sigma_1)\int_{\beta_1}^{\beta_2}\log|\beta|\ d\beta - 2(\beta_2-\beta_1)\int_{\sigma_1}^{\sigma_2}\sigma^{-1}\log\sigma\ d\sigma\Bigg] \\
&= K_6(\alpha_2-\alpha_1)\Bigg[(\beta_2-\beta_1)(\log\sigma_2-\log\sigma_1)\log\left(\frac{cnA}{cB^2+(n+c)A}\right) + \\
&\quad 2(\log\sigma_2-\log\sigma_1)(\beta_2\log\beta_2-\beta_1\log|\beta_1|-\beta_2-\beta_1) - \\
&\quad (\beta_2-\beta_1)[(\log\sigma_2)^2-(\log\sigma_1)^2]\Bigg] \\
&= K_7\log(cB^2+(n+c)A)+K_8.
\end{aligned}$$

It is easily verified that $K_7=-1$, and K_8 depends only on the ranges of the parameters in $\boldsymbol{\Theta}^l$, not on the parameters themselves. Now

$$\begin{aligned}
\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0]\right\} &= \exp\left\{\frac{1}{2}[(-1)\log(cB^2+(n+c)A)+K_8]\right\} \\
&= \exp\left\{\frac{1}{2}K_8\right\}(cB^2+(n+c)A)^{-\frac{1}{2}}.
\end{aligned}$$

Let

$$\begin{aligned}
K_9 &= \int \exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0]\right\}I_{[x_{0,1},x_{0,2}]}(x_0)\ dx_0 \\
&= \int_{x_{0,1}}^{x_{0,2}} \exp\left\{\frac{1}{2}K_8\right\}(cB^2+(n+c)A)^{-\frac{1}{2}}\ dx_0.
\end{aligned}$$

Note that K_9 is finite as long as A is greater than zero (that is, as long as the first stage of the calibration experiment includes at least two distinct x values). Hence,

$$\begin{aligned}
\pi^l_1(\alpha,\beta,\sigma|x_0) &= \frac{\pi^l_2(\beta,\sigma,\alpha|x_0)\exp\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0]\}I_{[x_{0,1},x_{0,2}]}(x_0)}{\int\exp\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0\}I_{[x_{0,1},x_{0,2}]}(x_0)\ dx_0} \\
&= K_9^{-1}K_6\sigma^{-1}\exp\left\{\frac{1}{2}K_8\right\}(cB^2+(n+c)A)^{-\frac{1}{2}} \\
&\quad \cdot I_{[\alpha_1,\alpha_2]}(\alpha)I_{[\beta_1,\beta_2]}(\beta)I_{[\sigma_1,\sigma_2]}(\sigma)I_{[x_{0,1},x_{0,2}]}(x_0) \\
&= K_{10}(cB^2+(n+c)A)^{-\frac{1}{2}}\sigma^{-1}I_{[\alpha_1,\alpha_2]}(\alpha)I_{[\beta_1,\beta_2]}(\beta)I_{[\sigma_1,\sigma_2]}(\sigma)I_{[x_{0,1},x_{0,2}]}(x_0).
\end{aligned}$$

We now take $\boldsymbol{\theta}^* = (0, 0, 0, 1)$, so the reference prior is

$$\begin{aligned}\pi^R(x_0, \alpha, \beta, \sigma) &\propto \lim_{l\to\infty} \frac{K_{10}(cB^2 + (n+c)A)^{-\frac{1}{2}}\sigma^{-1}}{K_{10}(\mathfrak{s}_{xx}^2 + (n+c)A)^{-\frac{1}{2}}} \\ &\propto ((cB^2 + (n+c)A)^{-\frac{1}{2}}\sigma^{-1} \\ &\propto \frac{1}{\sigma\sqrt{(n+c)(\mathfrak{s}_{xx} - n\bar{x}^2) + cn(x_0 - \bar{x})^2}},\end{aligned}$$

as claimed.

A.2.2 The Reference Prior for the Parameter Grouping $(x_0, (\alpha, \beta, \sigma))$

For the parameter grouping $(x_0, (\alpha, \beta, \sigma))$, we have $m = 2$, and

$$h_1 = \frac{cn\beta^2 A}{\sigma^2(cB^2 + (n+c)A)} = K_1(x_0)\frac{\beta^2}{\sigma^2};$$

$$h_2 = \frac{2(n+c)}{\sigma^6}((n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2}).$$

Step 1. Note that

$$\begin{aligned}&\iiint |h_2|^{\frac{1}{2}} I_{[\sigma_1,\sigma_2]}(\sigma) I_{[\beta_1,\beta_2]}(\beta) I_{[\alpha_1,\alpha_2]}(\alpha) d\sigma d\beta d\alpha \\ &= \int_{\alpha_1}^{\alpha_2}\int_{\beta_1}^{\beta_2}\int_{\sigma_1}^{\sigma_2} (2(n+c))^{\frac{1}{2}}((n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2})^{\frac{1}{2}}\sigma^{-3}\, d\sigma d\beta d\alpha \\ &= -2(2(n+c))^{\frac{1}{2}}(\alpha_2 - \alpha_1)(\beta_2 - \beta_1)(\sigma_2^{-2} - \sigma_1^{-2})((n+c)\mathfrak{s}_{xx}^* - \mathfrak{s}_x^{*2})^{\frac{1}{2}}.\end{aligned}$$

So,

$$\begin{aligned}\pi_2^l((\alpha, \beta, \sigma)|x_0) &= \frac{|h_2|^{\frac{1}{2}} I_{[\sigma_1,\sigma_2]}(\sigma) I_{[\beta_1,\beta_2]}(\beta) I_{[\alpha_1,\alpha_2]}(\alpha)}{\iiint |h_2|^{\frac{1}{2}} I_{[\sigma_1,\sigma_2]}(\sigma) I_{[\beta_1,\beta_2]}(\beta) I_{[\alpha_1,\alpha_2]}(\alpha) d\sigma d\beta d\alpha} \\ &= \frac{-1}{2(\alpha_2 - \alpha_1)(\beta_2 - \beta_1)(\sigma_2^{-2} - \sigma_1^{-2})}\sigma^{-3} I_{[\sigma_1,\sigma_2]}(\sigma) I_{[\beta_1,\beta_2]}(\beta) I_{[\alpha_1,\alpha_2]}(\alpha) \\ &= K_3\sigma^{-3} I_{[\sigma_1,\sigma_2]}(\sigma) I_{[\beta_1,\beta_2]}(\beta) I_{[\alpha_1,\alpha_2]}(\alpha).\end{aligned}$$

Step 2. We have

$$\begin{aligned}
E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0] &= \iiint K_3 \log\left|K_1(x_0)\frac{\beta^2}{\sigma^2}\right|\sigma^{-3}I_{[\sigma_1,\sigma_2]}(\sigma)I_{[\beta_1,\beta_2]}(\beta)I_{[\alpha_1,\alpha_2]}(\alpha)d\sigma d\beta d\sigma \\
&= K_3(\alpha_2-\alpha_1)\Bigg[-2(\beta_2-\beta_1)(\sigma_2^{-2}-\sigma_1^{-2})\log\left(\frac{cnA}{cB^2+(n+c)A}\right) - \\
&\quad 4(\sigma_2^{-2}-\sigma_1^{-2})\int_{\beta_1}^{\beta_2}\log|\beta|d\beta - 2(\beta_2-\beta_1)\int_{\sigma^1}^{\sigma_2}\sigma^{-3}\log\sigma\ d\sigma\Bigg] \\
&= K_3(\alpha_2-\alpha_1)\Bigg[-2(\beta_2-\beta_1)(\sigma_2^{-2}-\sigma_1^{-2})\log\left(\frac{cnA}{cB^2+(n+c)A}\right) - \\
&\quad 4(\sigma_2^{-2}-\sigma_1^{-2})(\beta_2\log\beta_2-\beta_1\log|\beta_1|-\beta_2-\beta_1) + \\
&\quad \frac{1}{2}(\beta_2-\beta_1)(\sigma_2^{-2}+2\sigma_2^{-2}\log\sigma_2-\sigma_1^{-2}-2\sigma_1^{-2}\log\sigma_1)\Bigg] \\
&= K_4\log(cB^2+(n+c)A)+K_5.
\end{aligned}$$

It is easily verified that $K_4=-1$, and K_5 depends only on the ranges of the parameters in $\boldsymbol{\Theta}^l$, not on the parameters themselves. Now

$$\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0]\right\} = \exp\left\{\frac{1}{2}K_5\right\}(cB^2+(n+c)A)^{-\frac{1}{2}}.$$

Let

$$K_6 = \int \exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0]\right\}I_{[x_{0,1},x_{0,2}]}(x_0)\ dx_0,$$

and note that this is finite as long as A is greater than zero. Then

$$\begin{aligned}
\pi_1^l((\alpha,\beta,\sigma)|x_0) &= \frac{\pi_2^l((\beta,\sigma,\alpha)|x_0)\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0]\right\}I_{[x_{0,1},x_{0,2}]}(x_0)}{\int\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}[\log|h_1|\big|x_0\right\}I_{[x_{0,1},x_{0,2}]}(x_0)\ dx_0} \\
&= K_6^{-1}K_3\sigma^{-3}\exp\left\{\frac{1}{2}K_6\right\}(cB^2+(n+c)A)^{-\frac{1}{2}}I_{[x_{0,1},x_{0,2}]}(x_0) \\
&\quad \cdot I_{[\alpha_1,\alpha_2]}(\alpha)I_{[\beta_1,\beta_2]}(\beta)I_{[\sigma_1,\sigma_2]}(\sigma).
\end{aligned}$$

If we take $\boldsymbol{\theta}^* = (0,0,0,1)$, then, we see that the reference prior is

$$\begin{aligned}
\pi^R(x_0,\alpha,\beta,\sigma) &\propto \sigma^{-3}(cB^2+(n+c)A)^{-\frac{1}{2}} \\
&\propto \frac{1}{\sigma^3\sqrt{(n+c)(\mathfrak{s}_{xx}-n\bar{x}^2)+cn(x_0-\bar{x})^2}},
\end{aligned}$$

as claimed.

A.3 Integrability of the Posterior for the Extended Univariate Linear Calibration Problem

In this section, we demonstrate the integrability of the posterior for the extended univariate linear calibration problem, described in chapter 5. The prior is taken to be of a particular form from the class of probability matching priors. In particular, we take

$$\pi(x_0, \alpha, \beta, \sigma, \kappa) \propto \sigma^{-k}\kappa^{-l}\left((x_0 - \bar{x})^2 + \left(\frac{\kappa^2}{c} + \frac{1}{n}\right)S_{xx}\right)^{-\frac{1}{2}}.$$

The likelihood function for the $n + c$ observations, with the error structure described by equations (5.2) and (5.3), is

$$\mathcal{L} \propto \sigma^{-(n+c)}\left(\prod_{i=1}^{n+c} w_i^{\frac{1}{2}}\right)\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c} w_i(y_i - \alpha - \beta x_i)^2\right\},$$

where

$$w_i = \begin{cases} 1 & \text{for } i = 1, 2, \ldots, n \\ \kappa^{-2} & \text{for } i = n+1, n+2, \ldots, n+c \end{cases} \tag{A.1}$$

As before, we have denoted x_0 as x_i, for $i = n+1, n+2, \ldots, n+c$. The sum in the exponent can be re-written, by completing the square in α, as

$$S_w\big(\alpha - (\bar{y}_w - \beta\bar{x}_w)\big)^2 + S_{(y-\beta x)w(y-\beta x)},$$

where

$$S_w = \sum_{i=1}^{n+c} w_i, \quad \bar{y}_w = \frac{\sum_{i=1}^{n+c} w_i y_i}{S_w}, \quad \bar{x}_w = \frac{\sum_{i=1}^{n+c} w_i x_i}{S_w},$$

and

$$S_{(y-\beta x)w(y-\beta x)} = \sum_{i=1}^{n+c} w_i\big(y_i - \beta x_i - (\bar{y}_w - \beta\bar{x}_w)\big)^2.$$

Hence,

$$\alpha|x_0, \sigma, \kappa, \beta, \mathbf{y} \sim N\left((\bar{y}_w - \beta\bar{x}_w), \frac{\sigma^2}{S_w}\right).$$

Integrating α out from the likelihood function thus yields the remaining term

$$\sigma^{-(n+c-1)}\left(\prod_{i=1}^{n+c} w_i^{\frac{1}{2}}\right)S_w^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\sigma^2}S_{(y-\beta x)w(y-\beta x)}\right\}.$$

Considering the term $S_{(y-\beta x)w(y-\beta x)}$, we complete the square in β, yielding

$$S_{(y-\beta x)w(y-\beta x)} = S_{xwx}\left(\beta - \frac{S_{xwy}}{S_{xwx}}\right)^2 + S_{ywy} - \frac{S_{xwy}^2}{S_{xwx}},$$

where

$$S_{ywy} = \sum_{i=1}^{n+c} w_i(y_i - \bar{y}_w)^2, \quad S_{xwx} = \sum_{i=1}^{n+c} w_i(x_i - \bar{x}_w)^2,$$

and

$$S_{xwy} = \sum_{i=1}^{n+c} w_i(x_i - \bar{x}_w)(y_i - \bar{y}_w).$$

Hence,

$$\beta|x_0, \sigma, \kappa, \mathbf{y} \sim N\left(\frac{S_{xwy}}{S_{xwx}}, \frac{\sigma^2}{S_{xwx}}\right).$$

Integrating β out from the likelihood function thus yields the remaining term

$$\sigma^{-(n+c-2)}\left(\prod_{i=1}^{n+c} w_i^{\frac{1}{2}}\right) S_w^{-\frac{1}{2}} S_{xwx}^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(S_{ywy} - \frac{S_{xwy}^2}{S_{xwx}}\right)\right\}.$$

Now the involvement of the parameter σ in the posterior is through both the prior and the likelihood. In particular, the portion of the posterior involving σ is

$$\sigma^{-(n+c+k-2)} \exp\left\{-\frac{1}{2\sigma^2}\left(S_{ywy} - \frac{S_{xwy}^2}{S_{xwx}}\right)\right\}.$$

If we transform the parameter σ to σ^2, the above expression in terms of the transformed parameter is

$$\left(\sigma^2\right)^{-(\frac{n+c+k-3}{2}+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(S_{ywy} - \frac{S_{xwy}^2}{S_{xwx}}\right)\right\},$$

from which it can be seen that

$$\sigma^2|x_0, \kappa, \mathbf{y} \sim \mathcal{I}.\mathcal{G}.\left(\frac{n+c+k-3}{2}, 2\left(S_{ywy} - \frac{S_{xwy}^2}{S_{xwx}}\right)^{-1}\right).$$

The remaining posterior after integrating out σ^2, including all terms from the prior and the likelihood which involve x_0 and κ, is thus

$$\pi(x_0, \kappa|\mathbf{y}) \propto \kappa^{-l}\left((x_0 - \bar{x})^2 + \frac{(n\kappa^2 + c)}{nc} S_{xx}\right)^{-\frac{1}{2}} \left(\prod_{i=1}^{n+c} w_i^{\frac{1}{2}}\right) S_w^{-\frac{1}{2}} S_{xwx}^{-\frac{1}{2}} \left(S_{ywy} - \frac{S_{xwy}^2}{S_{xwx}}\right)^{-\frac{n+c+k-3}{2}}$$

$$= \kappa^{-l}\left((x_0 - \bar{x})^2 + \frac{(n\kappa^2 + c)}{nc} S_{xx}\right)^{-\frac{1}{2}} \left(\prod_{i=1}^{n+c} w_i^{\frac{1}{2}}\right) S_w^{-\frac{1}{2}} S_{xwx}^{\frac{n+c+k-4}{2}} \left(S_{xwx} S_{ywy} - S_{xwy}^2\right)^{-\frac{n+c+k-3}{2}}.$$

Now we replace the weights $w_i, i = 1, 2, \ldots, n + c$ with the specific quantities described by (A.1). The following equalities are straightforward to verify:

$$\prod_{i=1}^{n+c} w_i^{\frac{1}{2}} = \kappa^{-c}; \quad S_w = \frac{n\kappa^2 + c}{\kappa^2};$$

$$\bar{x}_w = \frac{1}{n\kappa^2 + c}\Big(n\kappa^2\bar{x} + cx_0\Big); \quad \bar{y}_w = \frac{1}{n\kappa^2 + c}\Big(n\kappa^2\bar{y} + c\bar{y}_0\Big);$$

where $\bar{x}$ and $\bar{y}$ are the means of the regressor and response values from the first stage of experimentation, respectively, and $\bar{y}_0$ is the mean of the responses from the second stage of experimentation;

$$(x_0 - \bar{x}_w) = \frac{n\kappa^2}{n\kappa^2 + c}(x_0 - \bar{x});$$

$$(x_i - \bar{x}_w) = (x_i - \bar{x}) - \frac{c}{n\kappa^2 + c}(x_0 - \bar{x});$$

$$(\bar{y}_0 - \bar{y}_w) = \frac{n\kappa^2}{n\kappa^2 + c}(\bar{y}_0 - \bar{y});$$

$$(y_i - \bar{y}_w) = (y_i - \bar{y}) - \frac{c}{n\kappa^2 + c}(\bar{y}_0 - \bar{y}).$$

Note also that

$$\begin{aligned}
S_{ywy} &= \sum_{i=1}^{n}(y_i - \bar{y}_w)^2 + \frac{1}{\kappa^2}\sum_{i=n+1}^{n+c}(y_i - \bar{y}_w)^2 \\
&= \sum_{i=1}^{n}\left((y_i - \bar{y}) - \frac{c}{n\kappa^2 + c}(\bar{y}_0 - \bar{y})\right)^2 + \frac{1}{\kappa^2}\sum_{i=n+1}^{n+c}\left((y_i - \bar{y}_0) + (\bar{y}_0 - \bar{y}_w)\right)^2 \\
&= \sum_{i=1}^{n}(y_i - \bar{y})^2 + \frac{nc^2}{(n\kappa^2 + c)^2}(\bar{y}_0 - \bar{y})^2 + \frac{1}{\kappa^2}S_{yy}^{-} + \frac{n^2c\kappa^2}{(n\kappa^2 + c)^2}(\bar{y}_0 - \bar{y})^2 \\
&= S_{yy} + \frac{1}{\kappa^2}S_{yy}^{-} + \frac{nc}{n\kappa^2 + c}(\bar{y}_0 - \bar{y})^2,
\end{aligned}$$

where $\bar{y}_0$ is the mean of the responses from the second stage of experimentation, $S_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$, and $S_{yy}^{-} = \sum_{i=n+1}^{n+c}(y_i - \bar{y}_0)^2$. Similarly,

$$S_{xwx} = S_{xx} + \frac{nc}{n\kappa^2 + c}(x_0 - \bar{x})^2, \quad \text{and} \quad S_{xwy} = S_{xy} + \frac{nc}{n\kappa^2 + c}(x_0 - \bar{x})(\bar{y}_0 - \bar{y}),$$

where S_{xx} and S_{xy} are the centered, unweighted sums of squares and products from the first stage data.

For simplicity we write $v_0 = \bar{y}_0 - \bar{y}$, and make the transformation of x_0 and κ defined by

$$u_0 = x_0 - \bar{x}, \quad \rho = \frac{n\kappa^2 + c}{c}.$$

The Jacobian of this transformation is $J = (\rho - 1)^{-\frac{1}{2}}$. Note that the range for u_0 is $(-\infty, \infty)$, and the range for ρ is $(1, \infty)$. We can re-write the weighted sums of squares and products in terms of u_0, v_0, and ρ, yielding

$$S_{xwx} = S_{xx} + \frac{nu_0^2}{\rho};$$

$$S_{ywy} = S_{yy} + \frac{nv_0^2}{\rho} + \frac{n}{c(\rho - 1)} S_{yy}^{-};$$

$$S_{xwy} = S_{xy} + \frac{nu_0 v_0}{\rho}.$$

Also,

$$S_w = \frac{\rho}{\rho - 1}, \quad \text{and} \quad \prod_{i=1}^{n+c} w_i^{\frac{1}{2}} = (\rho - 1)^{-\frac{c}{2}}.$$

Now the joint posterior for u_0 and ρ is

$$\begin{aligned}
\pi(u_0, \rho | \mathbf{y}) &\propto \rho^{-1} (\rho - 1)^{-\frac{2l+c-1}{2}} \left(S_{xx} + \frac{nu_0^2}{\rho} \right)^{\frac{n+c+k-5}{2}} \\
&\quad \cdot \left(\left(S_{xx} + \frac{nu_0^2}{\rho} \right) \left(S_{yy} + \frac{nv_0^2}{\rho} + \frac{nS_{yy}^{-}}{c(\rho - 1)} \right) - \left(S_{xy} + \frac{nu_0 v_0}{\rho} \right)^2 \right)^{-\frac{n+c+k-3}{2}} \\
&\propto (\rho - 1)^{-\frac{2l+c-1}{2}} \left(\frac{\rho S_{xx}}{n} + u_0^2 \right)^{-1} \\
&\quad \cdot \left(\left(S_{yy} + \frac{nv_0^2}{\rho} + \frac{nS_{yy}^{-}}{c(\rho - 1)} \right) - \frac{(S_{xy} + \frac{nu_0 v_0}{\rho})^2}{S_{xx} + \frac{nu_0^2}{\rho}} \right)^{-\frac{n+c+k-3}{2}}.
\end{aligned}$$

We now consider the last term in the above expression. This may be re-written as

$$\left(\frac{nS_{yy}^{-}}{c(\rho - 1)} + \frac{\rho(S_{xx}S_{yy} - S_{xy}^2) + n(u_0^2 S_{yy} + v_0^2 S_{xx} - 2u_0 v_0 S_{xy})}{\rho S_{xx} + nu_0^2} \right)^{-\frac{n+c+k-3}{2}} \tag{A.2}$$

$$= \left(\frac{nS_{yy}^{-}}{c(\rho - 1)} + \frac{A\rho + B}{C\rho + D} \right)^{-\frac{n+c+k-3}{2}}, \tag{A.3}$$

where

$$A = S_{xx}S_{yy} - S_{xy}^2,$$

$$B = n(u_0^2 S_{yy} + v_0^2 S_{xx} - 2u_0 v_0 S_{xy}),$$

$$C = S_{xx}, \quad \text{and} \quad D = nu_0^2.$$

If we consider the expression

$$\frac{A\rho + B}{C\rho + D} \tag{A.4}$$

as a function of ρ, and examine the first derivative of this function, we can easily see that the function is decreasing in ρ for $\rho > 0$ as long as $AD - BC < 0$. In fact,

$$\begin{aligned} AD - BC &= nu_0^2(S_{xx}S_{yy} - S_{xy}^2) - nS_{xx}(u_0^2 S_{yy} + v_0^2 S_{xx} - 2u_0 v_0 S_{xy}) \\ &= -n(u_0 S_{xy} - v_0 S_{xx})^2 \\ &< 0. \end{aligned}$$

Hence, if we replace (A.4) in (A.3) with

$$\frac{A}{C} = \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}},$$

this increases the value of (A.3). We have, then, that up to a normalizing constant (which will not effect integrability)

$$\begin{aligned} \int_{u_0=-\infty}^{\infty} \pi(u_0, \rho|\mathbf{y})\, du_0 &\leq \int_{u_0=-\infty}^{\infty} (\rho-1)^{-\frac{2l+c-1}{2}} \left(\frac{\rho S_{xx}}{n} + u_0^2\right)^{-1} \\ &\quad \cdot \left(\frac{nS_{yy}^-}{c(\rho-1)} + \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}}\right)^{-\frac{n+c+k-3}{2}} du_0 \\ &= (\rho-1)^{-\frac{2l+c-1}{2}} \left(\frac{nS_{yy}^-}{c(\rho-1)} + \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}}\right)^{-\frac{n+c+k-3}{2}} \\ &\quad \cdot \int_{u_0=-\infty}^{\infty} \frac{1}{\frac{\rho S_{xx}}{n} + u_0^2}\, du_0 \\ &\propto \rho^{-\frac{1}{2}} (\rho-1)^{-\frac{2l+c-1}{2}} \left(\frac{nS_{yy}^-}{c(\rho-1)} + \frac{S_{xx}S_{yy} - S_{xy}^2}{S_{xx}}\right)^{-\frac{n+c+k-3}{2}}. \end{aligned}$$

The integrability of this resulting function of ρ must be studied as $\rho \to \infty$ and as $\rho \to 1^+$. We first consider the behavior of this function as $\rho \to \infty$. Note that as $\rho \to \infty$, ρ has order $-\frac{2l+c}{2}$. Hence, the function will be integrable for any $l \geq 1$ and $c \geq 1$.

In considering the behavior of the function as $\rho \to 1^+$, we treat two cases separately. First, if $c = 1$, then $S_{yy}^- = 0$, so the order of the term $(\rho - 1)$ is $-2l$. Hence, the function will not be integrable for any $l \geq 1$. If, on the other hand, $c > 1$, then S_{yy}^- will be positive with probability 1, in which case the order of the term $(\rho - 1)$ will be $\frac{n+k-2l-2}{2}$. Hence, if $c > 1$ the function will be integrable if $n + k \geq 2l + 1$.

A.4 Development of the Reference Prior for the Multivariate Linear Calibration Problem

A.4.1 Evaluation of the Information Matrix

For the current problem, we have the parameter vector $\boldsymbol{\theta} = (\mathbf{x}_0, \alpha_1, \boldsymbol{\beta}_1', \alpha_2, \boldsymbol{\beta}_2', \ldots \alpha_q, \boldsymbol{\beta}_q', \boldsymbol{\sigma}')$. There are $p + q + pq + r$ individual parameters in $\boldsymbol{\theta}$. The likelihood function for the $n + c$ observations is

$$\mathcal{L} = \prod_{i=1}^{n+c} \frac{1}{(2\pi)^{q/2}|\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2}(\mathbf{y}_i' - \boldsymbol{\alpha}' - \mathbf{x}_i'\mathbf{B})\boldsymbol{\Sigma}^{-1}(\mathbf{y}_i' - \boldsymbol{\alpha}' - \mathbf{x}_i'\mathbf{B})' \right\}.$$

Here, $\mathbf{y}_i'$ denotes the i^{th} row of the data matrix $\mathbf{Y}$. The log of the likelihood can be written as

$$\log \mathcal{L} = -\frac{(n+c)q}{2}\log(2\pi) - (n+c)\sum_{j=1}^{q}\log(\tilde{\sigma}_j) - \frac{1}{2}\sum_{i=1}^{n+c}\sum_{j=1}^{q}\frac{(y_{ij} - \alpha_j - \mathbf{x}_i'\boldsymbol{\beta}_j)^2}{\tilde{\sigma}_j^2},$$

where $\tilde{\sigma}_j$ is the standard deviation of the j^{th} response variable. Recall that each $\tilde{\sigma}_j$ will be one of $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$.

We now compute the information matrix. For $u = 1, 2, \ldots, r$,

$$\frac{\partial^2}{\partial \sigma_u^2}\log \mathcal{L} = \frac{(n+c)t_u}{\sigma_u^2} - \frac{3}{\sigma_u^4}\sum_{i=1}^{n+c}\sum_{\{j:\tilde{\sigma}_j = \sigma_u\}}(y_{ij} - \alpha_j - \mathbf{x}_i'\boldsymbol{\beta}_j)^2,$$

so

$$-E\left[\frac{\partial}{\partial \sigma_u}\frac{\partial}{\partial \sigma_u}\log \mathcal{L}\right] = \frac{2(n+c)t_u}{\sigma_u^2}.$$

If j $(j = 1, 2, \ldots, q)$ is such that $\tilde{\sigma}_j = \sigma_u$, then

$$\frac{\partial}{\partial \alpha_j}\frac{\partial}{\partial \sigma_u}\log \mathcal{L} = -\frac{2}{\sigma_u^3}\sum_{i=1}^{n+c}(y_{ij} - \alpha_j - \mathbf{x}_i'\boldsymbol{\beta}_j),$$

so

$$-E\left[\frac{\partial}{\partial\alpha_j}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}\right]=0,$$

and

$$\frac{\partial}{\partial\boldsymbol{\beta}_j}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}=-\frac{2}{\sigma_u^3}\sum_{i=1}^{n+c}(y_{ij}-\alpha_j-\mathbf{x}_i'\boldsymbol{\beta}_j)(\mathbf{x}_i),$$

so

$$-E\left[\frac{\partial}{\partial\boldsymbol{\beta}_j}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}\right]=\mathbf{0}_{p\times1}.$$

Note that if j $(j=1,2,\ldots,q)$ is such that $\tilde{\sigma}_j\neq\sigma_u$, then

$$\frac{\partial}{\partial\alpha_j}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}=0,$$

and

$$\frac{\partial}{\partial\boldsymbol{\beta}_j}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}=\mathbf{0}_{p\times1}.$$

Also,

$$\frac{\partial}{\partial\mathbf{x}_0}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}=-\frac{2}{\sigma_u^3}\sum_{i=n+1}^{n+c}\sum_{\{j:\tilde{\sigma}_j=\sigma_u\}}(y_{ij}-\alpha_j-\mathbf{x}_0'\boldsymbol{\beta}_j)\mathbf{x}_0,$$

so

$$-E\left[\frac{\partial}{\partial\mathbf{x}_0}\frac{\partial}{\partial\sigma_u}\log\mathcal{L}\right]=\mathbf{0}_{p\times1}.$$

Next, for $j=1,2,\ldots,q$, we have

$$\frac{\partial^2}{\partial\alpha_j^2}\log\mathcal{L}=-\frac{1}{\tilde{\sigma}_j^2}\sum_{i=1}^{n+c}1,$$

so

$$-E\left[\frac{\partial^2}{\partial\alpha_j^2}\log\mathcal{L}\right]=\frac{n+c}{\tilde{\sigma}_j^2},$$

and

$$\frac{\partial}{\partial\boldsymbol{\beta}_j}\frac{\partial}{\partial\alpha_j}\log\mathcal{L}=-\frac{1}{\tilde{\sigma}_j^2}\sum_{i=1}^{n+c}\mathbf{x}_i,$$

so

$$-E\left[\frac{\partial}{\partial \boldsymbol{\beta}_j}\frac{\partial}{\partial \alpha_j}\log\mathcal{L}\right] = \frac{1}{\tilde{\sigma}_j^2}\sum_{i=1}^{n+c}\mathbf{x}_i.$$

Note that if $j' \neq j$, then

$$\frac{\partial}{\partial \alpha_{j'}}\frac{\partial}{\partial \alpha_j}\log\mathcal{L} = 0,$$

and

$$\frac{\partial}{\partial \boldsymbol{\beta}_{j'}}\frac{\partial}{\partial \alpha_j}\log\mathcal{L} = \mathbf{0}_{p\times 1}.$$

Also,

$$\frac{\partial}{\partial \mathbf{x}_0}\frac{\partial}{\partial \alpha_j}\log\mathcal{L} = -\frac{1}{\tilde{\sigma}_j^2}\sum_{i=n+1}^{n+c}\boldsymbol{\beta}_j,$$

so

$$-E\left[\frac{\partial}{\partial \mathbf{x}_0}\frac{\partial}{\partial \alpha_j}\log\mathcal{L}\right] = \frac{c}{\tilde{\sigma}_j^2}\boldsymbol{\beta}_j.$$

Now for $j = 1, 2, \ldots, q$,

$$\frac{\partial}{\partial \boldsymbol{\beta}_j}\frac{\partial}{\partial \boldsymbol{\beta}_j'}\log\mathcal{L} = -\frac{1}{\tilde{\sigma}_j^2}\sum_{i=n+1}^{n+c}\mathbf{x}_i\mathbf{x}_i',$$

so

$$-E\left[\frac{\partial}{\partial \boldsymbol{\beta}_j}\frac{\partial}{\partial \boldsymbol{\beta}_j'}\log\mathcal{L}\right] = \frac{1}{\tilde{\sigma}_j^2}\mathbf{X}'\mathbf{X}.$$

For $j' \neq j$,

$$\frac{\partial}{\partial \boldsymbol{\beta}_{j'}}\frac{\partial}{\partial \boldsymbol{\beta}_j'}\log\mathcal{L} = \mathbf{0}_{p\times p}.$$

Finally,

$$\frac{\partial}{\partial \mathbf{x}_0}\frac{\partial}{\partial \mathbf{x}_0'}\log\mathcal{L} = \sum_{i=n+1}^{n+c}\sum_{j=1}^{q}\frac{1}{\tilde{\sigma}_j^2}(y_{ij} - \alpha_j - \mathbf{x}_i'\boldsymbol{\beta}_j)\frac{\partial}{\partial \mathbf{x}_0}\boldsymbol{\beta}_j' - \sum_{i=n+1}^{n+c}\sum_{j=1}^{q}\frac{1}{\tilde{\sigma}_j^2}\boldsymbol{\beta}_j\boldsymbol{\beta}_j',$$

so

$$\begin{aligned}-E\left[\frac{\partial}{\partial \mathbf{x}_0}\frac{\partial}{\partial \mathbf{x}_0'}\log\mathcal{L}\right] &= c\sum_{j=1}^{q}\frac{1}{\tilde{\sigma}_j^2}\boldsymbol{\beta}_j\boldsymbol{\beta}_j' \\ &= c\mathbf{B}\Sigma^{-1}\mathbf{B}'.\end{aligned}$$

The information matrix for the parameters $(\mathbf{x}_0, (\alpha_1, \boldsymbol{\beta}_1), \ldots, (\alpha_q, \boldsymbol{\beta}_q), \boldsymbol{\sigma})$ is thus

$$\mathcal{I} = \begin{bmatrix} \mathcal{I}_1 & \mathbf{0} \\ \mathbf{0} & \mathcal{I}_2 \end{bmatrix}.$$

In this expression, $\mathcal{I}_2$ is the diagonal matrix $\textsc{Diag}(\frac{2(n+c)t_1}{\sigma_1^2}, \ldots, \frac{2(n+c)t_r}{\sigma_r^2})$, and

$$\mathcal{I}_1 = \begin{bmatrix} \mathcal{I}_{1,1} & \mathcal{I}_{1,2} \\ \mathcal{I}'_{1,2} & \mathcal{I}_{2,2} \end{bmatrix},$$

where $\mathcal{I}_{1,1} = c\mathbf{B}\boldsymbol{\Sigma}^{-1}\mathbf{B}'$, $\mathcal{I}_{1,2} = [\frac{c}{\sigma_1^2}\boldsymbol{\beta}_1\xi_0', \ldots, \frac{c}{\sigma_q^2}\boldsymbol{\beta}_q\xi_0']$, and $\mathcal{I}_{2,2}$ is the block diagonal matrix $\Sigma^{-1} \otimes \mathbf{X}'_\alpha\mathbf{X}_\alpha$. Here, $\mathbf{X}_\alpha$ is the $\mathbf{X}$ matrix, augmented on the left by a column of unity elements, $\boldsymbol{\xi}_0 = \begin{bmatrix} 1 & \mathbf{x}_0' \end{bmatrix}'$, and $\otimes$ denotes the Kronecker product.

A.4.2 Definition of the Compact Sets $\boldsymbol{\Theta}^l$

We define the sets $\boldsymbol{\Theta}^l$ as Cartesian products of intervals corresponding to $x_{0,1}, \ldots, x_{0,p}$, $\alpha_1, \ldots, \alpha_q$, $\beta_{1,1}, \ldots, \beta_{p,q}$, and $\sigma_1, \ldots, \sigma_q$. In particular, we define

$$\begin{aligned} \boldsymbol{\Theta}^l =& [x_{0,1}^{l,1}, x_{0,1}^{l,2}] \times \cdots \times [x_{0,p}^{l,1}, x_{0,p}^{l,2}] \times [\alpha_1^{l,1}, \alpha_1^{l,2}] \times \cdots \times [\alpha_q^{l,1}, \alpha_q^{l,2}] \\ & \times [\beta_{1,1}^{l,1}, \beta_{1,1}^{l,2}] \times \cdots \times [\beta_{p,q}^{l,1}, \beta_{p,q}^{l,2}] \times [\sigma_1^{l,1}, \sigma_1^{l,2}] \times \cdots \times [\sigma_q^{l,1}, \sigma_q^{l,2}]. \end{aligned}$$

As $l \to \infty$, the upper endpoints $x_{0,1}^{l,2}, \ldots x_{0,p}^{l,2}$, $\alpha_1^{l,2}, \ldots \alpha_q^{l,2}$, $\beta_{1,1}^{l,2}, \ldots \beta_{p,q}^{l,2}$, and $\sigma_1^{l,2}, \ldots \sigma_q^{l,2}$ tend monotonically to $+\infty$, the lower endpoints $x_{0,1}^{l,1}, \ldots x_{0,p}^{l,1}$, $\alpha_1^{l,1}, \ldots \alpha_q^{l,1}$, and $\beta_{p,q}^{l,1}, \ldots \beta_{p,q}^{l,1}$ tend monotonically to $-\infty$, and the lower endpoints $\sigma_1^{l,1}, \ldots, \sigma_q^{l,1}$ tend monotonically to zero.

A.4.3 Derivation of the Reference Prior

To implement the reference prior algorithm for the multivariate linear calibration problem described above, we must first compute the quantities h_i. For the parameter grouping $(\mathbf{x}_0, (\boldsymbol{\alpha}, \mathbf{B}), \boldsymbol{\sigma})$, we have $m = 3$. Note that $S_3 = \mathcal{I}^{-1}$, so $H_3 = \mathcal{I}$. It follows that $H_{3,3} = \mathcal{I}_2$, and

$$h_3 = |\mathcal{I}_2| = (2(n+c))^r \prod_{u=1}^{r} t_u \sigma_u^{-2}.$$

Now $S_2 = \mathcal{I}_1^{-1}$, so $H_2 = \mathcal{I}_1$. It follows that $H_{2,2} = \Sigma^{-1} \otimes \mathbf{X}'_\alpha\mathbf{X}_\alpha$, and

$$h_2 = |\mathbf{X}'_\alpha\mathbf{X}_\alpha|^q \prod_{u=1}^{r} \sigma_u^{-2t_u(p+1)}.$$

To find h_1, we first need to find the upper left block of the matrix $\mathcal{I}^{-1}$ (which is equivalent to the upper left block of $\mathcal{I}_1^{-1}$). We do so using the familiar lemma below (see, for example, Basilevsky (1983), page 179).

LEMMA 4 *If*

$$M = \begin{bmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{bmatrix}$$

is an invertible square matrix, written in partitioned form such that $A_{1,1}$ and $A_{2,2}$ are square matrices, then

$$M^{-1} = \begin{bmatrix} C_{1,1} & C_{1,2} \\ C_{2,1} & C_{2,2} \end{bmatrix},$$

where $C_{1,1} = (A_{1,1} - A_{1,2}A_{2,2}^{-1}A_{2,1})^{-1}$.

In the notation of the lemma, applied to $\mathcal{I}_1$ from the information matrix, we have $A_{1,1} = c\mathbf{B}\Sigma^{-1}\mathbf{B}'$, $A_{1,2} = [\frac{c}{\tilde{\sigma}_1^2}\boldsymbol{\beta}_1\boldsymbol{\xi}_0', \frac{c}{\tilde{\sigma}_2^2}\boldsymbol{\beta}_2\boldsymbol{\xi}_0', \ldots, \frac{c}{\tilde{\sigma}_q^2}\boldsymbol{\beta}_q\boldsymbol{\xi}_0']$, $A_{2,1} = A_{1,2}'$, and $A_{2,2} = \Sigma^{-1} \otimes \mathbf{X}_\alpha'\mathbf{X}_\alpha$. Hence,

$$\begin{aligned} C_{1,1} &= (c\mathbf{B}\Sigma^{-1}\mathbf{B}' - c^2\xi_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0\mathbf{B}\Sigma^{-1}\mathbf{B}')^{-1} \\ &= (c\mathbf{B}\Sigma^{-1}\mathbf{B}')^{-1}(1 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0)^{-1}. \end{aligned}$$

We can re-write this using the following result.

LEMMA 5 *In the context of the current problem, let $\mathbf{X}_\alpha$ be written in partitioned form as*

$$\mathbf{X}_\alpha = \begin{bmatrix} \mathbf{X}_{\alpha,1} \\ \mathbf{X}_{\alpha,2} \end{bmatrix},$$

where $\mathbf{X}_{\alpha,1}$ contains the first n rows of $\mathbf{X}_\alpha$, and $\mathbf{X}_{\alpha,2}$ contains the remaining c rows. Then

$$(1 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0)^{-1} = (1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0).$$

PROOF OF LEMMA 5 We verify the statement by direct multiplication. Note that since each row of the matrix $\mathbf{X}_{\alpha,2}$ is equal to $\boldsymbol{\xi}_0'$, it follows that $\mathbf{X}_{\alpha,2}'\mathbf{X}_{\alpha,2} = c\boldsymbol{\xi}_0\boldsymbol{\xi}_0'$. Now

$$\begin{aligned} &(1 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0)(1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0) \\ &= 1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0 - c^2\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0 \\ &= 1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\mathbf{X}_{\alpha,2}'\mathbf{X}_{\alpha,2}(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0. \end{aligned}$$

The last term of this expression may be re-expressed as

$$\begin{aligned}
&- c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}(\mathbf{X}_\alpha'\mathbf{X}_\alpha - \mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0 \\
&= -c\boldsymbol{\xi}_0'[(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1} - (\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}]\boldsymbol{\xi}_0. \\
&= -c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0 + c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0 \\
&= 0.
\end{aligned}$$

The desired result immediately follows.

Using this result, we can express H_1, which is equal to $C_{1,1}^{-1}$, as

$$H_1 = \frac{c\mathbf{B}\Sigma^{-1}\mathbf{B}'}{1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}.$$

It follows that

$$h_1 = \frac{c|\mathbf{B}\Sigma^{-1}\mathbf{B}'|}{1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}.$$

We are now ready to begin applying the reference prior algorithm, described in a previous section. In the following steps, each K represents a constant (depending possibly on the domains $\boldsymbol{\Theta}^l$, but not on any parameters), and all integrals are over the appropriate compact intervals defined previously.

Step 1. To begin the sequential development of the reference prior, we find

$$\begin{aligned}
\int_{\boldsymbol{\sigma}} |h_3|^{\frac{1}{2}}\, d\boldsymbol{\sigma} &= \int_{\boldsymbol{\sigma}} (2(n+c))^{\frac{q}{2}} \prod_{u=1}^{r} t_u \sigma_u^{-1}\, d\boldsymbol{\sigma} \\
&= (2(n+c))^{\frac{q}{2}} K_1.
\end{aligned}$$

Hence

$$\pi_3^l(\boldsymbol{\sigma}|\mathbf{x}_0, (\boldsymbol{\alpha}, \mathbf{B})) = \frac{1}{K_1} \prod_{u=1}^{r} t_u \sigma_u^{-1},$$

for $\boldsymbol{\sigma} \in [\sigma_1^{l,1}, \sigma_1^{l,2}] \times \ldots \times [\sigma_r^{l,1}, \sigma_r^{l,2}]$.

Step 2. Note that

$$\begin{aligned}
E_{\boldsymbol{\theta}}^l\big[\log|h_2|\big|\mathbf{x}_0, (\boldsymbol{\alpha}, \mathbf{B})\big] &= \int_{\boldsymbol{\sigma}} \frac{1}{K_1} \prod_{u=1}^{r} t_u \sigma_u^{-1} \log\left(\frac{|\mathbf{X}'\mathbf{X}|^q}{\prod_{u=1}^{r} \sigma_u^{2t_u(p+1)}}\right) d\boldsymbol{\sigma} \\
&= \log(|\mathbf{X}_\alpha'\mathbf{X}_\alpha|^q) + K_2.
\end{aligned}$$

So

$$\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}\big[\log|h_2|\big|\mathbf{x}_0,(\boldsymbol{\alpha},\mathbf{B})\big]\right\} = |\mathbf{X}'_\alpha\mathbf{X}_\alpha|^{\frac{q}{2}}\exp(K_2/2),$$

and

$$\int_{\boldsymbol{\alpha},\mathbf{B}}\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}\big[\log|h_2|\big|\mathbf{x}_0,(\boldsymbol{\alpha},\mathbf{B})\big]\right\}\,d\boldsymbol{\alpha}\;d\mathbf{B} = K_3|\mathbf{X}'_\alpha\mathbf{X}_\alpha|^{\frac{q}{2}}\exp(K_2/2).$$

Hence

$$\pi^l_2((\boldsymbol{\alpha},\mathbf{B}),\boldsymbol{\sigma}|\mathbf{x}_0) = \frac{1}{K_1K_3}\prod_{u=1}^r t_u\sigma_u^{-1},$$

for $(\boldsymbol{\alpha},\mathbf{B},\boldsymbol{\sigma})$ with $\boldsymbol{\sigma}\in[\sigma_1^{l,1},\sigma_1^{l,2}]\times\cdots\times[\sigma_r^{l,1},\sigma_r^{l,2}]$, $\boldsymbol{\alpha}\in[\alpha_1^{l,1},\alpha_1^{l,2}]\times\cdots\times[\alpha_q^{l,1},\alpha_q^{l,2}]$, and $\mathbf{B}$ such that $(\beta_{1,1},\ldots,\beta_{p,q})\in[\beta_{1,1}^{l,1},\beta_{1,1}^{l,2}]\times\cdots\times[\beta_{p,q}^{l,1},\beta_{p,q}^{l,2}]$.

Step 3. Note that

$$\begin{aligned}
E^l_{\boldsymbol{\theta}}\big[\log|h_1|\big|\mathbf{x}_0\big] &= \int_{\boldsymbol{\alpha},\mathbf{B},\boldsymbol{\sigma}}\frac{1}{K_1K_3}\left(\prod_{u=1}^r t_u\sigma_u^{-1}\right)\log\left(\frac{c|\mathbf{B}\Sigma^{-1}\mathbf{B}'|}{1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}\right)\,d\boldsymbol{\alpha}\;d\mathbf{B}\;d\boldsymbol{\sigma}\\
&= \int_{\boldsymbol{\alpha},\mathbf{B},\boldsymbol{\sigma}}\frac{1}{K_1K_3}\left(\prod_{u=1}^r t_u\sigma_u^{-1}\right)\log(c|\mathbf{B}\Sigma^{-1}\mathbf{B}'|)\,d\boldsymbol{\alpha}\;d\mathbf{B}\;d\boldsymbol{\sigma}\\
&\quad-\int_{\boldsymbol{\alpha},\mathbf{B},\boldsymbol{\sigma}}\frac{1}{K_1K_3}\left(\prod_{u=1}^r t_u\sigma_u^{-1}\right)\log(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)\,d\boldsymbol{\alpha}\;d\mathbf{B}\;d\boldsymbol{\sigma}\\
&= K_4-\log(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0),
\end{aligned}$$

so

$$\exp\left\{\frac{1}{2}E^l_{\boldsymbol{\theta}}\big[\log|h_1|\big|\mathbf{x}_0\big]\right\} = \exp(K_4)(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{-\frac{1}{2}}.$$

If we let

$$K_5 = \int_{\mathbf{x}_0}(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{-\frac{1}{2}}\,d\mathbf{x}_0,$$

then

$$\pi^l_1(\mathbf{x}_0,(\boldsymbol{\alpha},\mathbf{B}),\boldsymbol{\sigma}) = \frac{1}{K_1K_3K_5}\left(\prod_{u=1}^r t_u\sigma_u^{-1}\right)(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{-\frac{1}{2}}.$$

To define $\boldsymbol{\theta}^*$, we take $\mathbf{x}_0^*=\mathbf{0}$, $\boldsymbol{\alpha}^*=\mathbf{0}$, $\mathbf{B}^*=\mathbf{0}$, and $\boldsymbol{\sigma}^*=\mathbf{1}$, where $\mathbf{1}$ is a vector of unity elements. Hence,

$$\pi(\mathbf{x}_0,(\boldsymbol{\alpha},\mathbf{B}),\boldsymbol{\sigma})\propto\left(\prod_{u=1}^r\sigma_u^{-1}\right)(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)^{-\frac{1}{2}}.$$

A.5 Development of the Reference Prior for the Univariate Polynomial Calibration Problem

In this section, the derivation of the reference prior for the univariate polynomial calibration problem, as described in chapter 7, is shown. In particular, the parameter grouping $(x_0, (\alpha, \boldsymbol{\beta}), \sigma)$ is considered.

A.5.1 The Likelihood Function and the Fisher Information Matrix

The likelihood function for the univariate polynomial calibration problem as described above is proportional to

$$\mathcal{L} \propto \frac{1}{\sigma^{n+c}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n+c}(y_i - \alpha - \boldsymbol{\beta}'\mathbf{x}_i)^2\right\}.$$

We let

$$\boldsymbol{\zeta}_0 = (1, 2x_0, 3x_0^2, \ldots, px_0^{p-1})',$$

and

$$\boldsymbol{\xi}_0 = (1, \mathbf{x}_0')'.$$

The Fisher information matrix for the $n + c$ observations is

$$\mathcal{I} = \frac{1}{\sigma^2}\begin{bmatrix} \mathcal{I}_1 & \mathbf{0}_{(p+2)\times 1} \\ \mathbf{0}_{1\times(p+2)} & 2(n+c) \end{bmatrix},$$

where

$$\begin{aligned}\mathcal{I}_1 &= \begin{bmatrix} c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & c\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\mathbf{x}_0' \\ c\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & n+c & \sum_{i=1}^{n+c}\mathbf{x}_i' \\ c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\mathbf{x}_0 & \sum_{i=1}^{n+c}\mathbf{x}_i & \mathbf{X}'\mathbf{X} \end{bmatrix} \\ &= \begin{bmatrix} c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0 & c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\boldsymbol{\xi}_0' \\ c\boldsymbol{\beta}'\boldsymbol{\zeta}_0\boldsymbol{\xi}_0 & \mathbf{X}_\alpha'\mathbf{X}_\alpha \end{bmatrix},\end{aligned}$$

and $\mathbf{X}$ is the $(n + c) \times p$ matrix whose i^{th} row is $\mathbf{x}_i'$, and $\mathbf{X}_\alpha$ is the $(n + c) \times (p + 1)$ matrix whose i^{th} row is $(1, \mathbf{x}_i')'$.

In the derivation of the reference prior, we will need the upper left entry of $\mathcal{I}^{-1}$. Note that this is also the upper left entry of $\mathcal{I}_1^{-1}$, apart from the factor σ^2. We again invoke lemma 4

to compute this:

$$\begin{aligned}\mathcal{I}_{1,1}^{-1} &= \left(c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0 - c^2\boldsymbol{\beta}'\boldsymbol{\zeta}_0\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0\boldsymbol{\zeta}_0'\boldsymbol{\beta}\right)^{-1} \\ &= \left((c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0)(1 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0)\right)^{-1} \\ &= \frac{(1 - c\boldsymbol{\xi}_0'(\mathbf{X}_\alpha'\mathbf{X}_\alpha)^{-1}\boldsymbol{\xi}_0)^{-1}}{c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0} \\ &= \frac{1 + c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}{c\boldsymbol{\zeta}_0'\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{\zeta}_0},\end{aligned}$$

where $\mathbf{X}_{\alpha,1}$ is the matrix consisting of the first n rows of $\mathbf{X}_\alpha$. Note that the last equality in the above expression follows from lemma 5.

A.5.2 Derivation of the Reference Prior for the Grouping $(\mathbf{x}_0, (\alpha, \boldsymbol{\beta}), \sigma)$

We define the compact sets $\boldsymbol{\Theta}^l$ as Cartesian products of the following sets:

$$x_0 \in [x_{0,1}, x_{0,2}], \quad \alpha \in [\alpha_1, \alpha_2], \quad \sigma \in [\sigma_1, \sigma_2], \quad \text{and} \quad \boldsymbol{\beta} \in \{\boldsymbol{\beta} : \boldsymbol{\beta}'\boldsymbol{\beta} \leq M\}.$$

Although not explicit in the notation, these sets depend on l.

Step 1. Note that

$$\begin{aligned}\int_\sigma |h_3|^{\frac{1}{2}}\, d\sigma &= \int_\sigma \sqrt{2(n+c)}\sigma^{-1}\, d\sigma \\ &= \sqrt{2(n+c)}(\log\sigma_2 - \log\sigma_1),\end{aligned}$$

so

$$\pi_3^l(\sigma|x_0, (\alpha, \boldsymbol{\beta})) = K_1\sigma^{-1},$$

for $\sigma \in [\sigma_1, \sigma_2]$. Here, and in the remainder of this section, quantities labeled K represent constants, depending possibly on the ranges of parameters but not on the parameters themselves. In particular,

$$K_1 = \frac{1}{\log\sigma_2 - \log\sigma_1}.$$

Also, all integrations are over the appropriate compact sets described by $\boldsymbol{\Theta}^l$.

Step 2. Note that

$$\begin{aligned}E_\theta^l[\log|h_2|\big|x_0, (\alpha, \boldsymbol{\beta})] &= \int_\sigma K_1\sigma^{-1}\log\left|\frac{\mathbf{X}_\alpha'\mathbf{X}_\alpha}{\sigma^{2(p+1)}}\right| d\sigma \\ &= \log|\mathbf{X}_\alpha'\mathbf{X}_\alpha| + K_2,\end{aligned}$$

where

$$K_2 = -2(p+1)\int_\sigma \sigma^{-1}\log\sigma \; d\sigma.$$

Hence

$$\exp\left\{\frac{1}{2}E_\theta^l[\log|h_2|\big|x_0,(\alpha,\boldsymbol{\beta})]\right\} = K_3|\mathbf{X}'_\alpha\mathbf{X}_\alpha|,$$

where

$$K_3 = \exp\left\{\frac{1}{2}K_2\right\}.$$

Now

$$\int_\alpha\int_{\boldsymbol{\beta}}\exp\left\{\frac{1}{2}E_\theta^l[\log|h_2|\big|x_0,(\alpha,\boldsymbol{\beta})]\right\}\,d\alpha d\boldsymbol{\beta} = \frac{K_3}{K_4}|\mathbf{X}'_\alpha\mathbf{X}_\alpha|,$$

where

$$K_4 = \frac{1}{(\alpha_2-\alpha_1)\int_{\boldsymbol{\beta}}\; d\boldsymbol{\beta}}.$$

We thus have

$$\pi_2^l((\alpha,\boldsymbol{\beta}),\sigma\big|x_0) = K_1K_4\sigma^{-1},$$

for $((\alpha,\boldsymbol{\beta}),\sigma)$ with $\alpha\in[\alpha_1,\alpha_2]$, $\sigma\in[\sigma_1,\sigma_2]$, and $\boldsymbol{\beta}$ such that $\boldsymbol{\beta}'\boldsymbol{\beta}\le M$.

Step 3. We note that

$$E_\theta^l[\log|h_1|\big|x_0] = \int_\alpha\int_{\boldsymbol{\beta}}\int_\sigma K_1K_4\sigma^{-1}\log\left|\frac{c(\boldsymbol{\zeta}_0'\boldsymbol{\beta})^2}{\sigma^2(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0)}\right| d\alpha d\boldsymbol{\beta} d\sigma$$

$$= 2K_4(\alpha_2-\alpha_1)\int_{\boldsymbol{\beta}}\log|\boldsymbol{\zeta}_0'\boldsymbol{\beta}|\;d\boldsymbol{\beta} - 2K_1\int_\sigma\sigma^{-1}\log\frac{\sigma}{\sqrt{c}}\;d\sigma - \log(1+c\boldsymbol{\xi}_0'(\mathbf{X}'_{\alpha,1}\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0). \tag{A.5}$$

The second integral in the above expression does not depend on x_0, and so will not effect the reference prior. The first integral, however, does depend on x_0, and therefore requires attention.

We begin by making the change of variables defined by

$$u = \boldsymbol{\zeta}_0'\boldsymbol{\beta} = \beta_1 + \boldsymbol{\zeta}'_{0,2}\boldsymbol{\beta}_2$$
$$\mathbf{v} = \boldsymbol{\beta}_2,$$

where $\boldsymbol{\beta}_2$ is the vector containing the last $p-1$ elements of $\boldsymbol{\beta}$, and $\boldsymbol{\zeta}_2$ is the vector containing the last $p-1$ elements of $\boldsymbol{\zeta}$. In matrix notation, this transformation may be written

$$\begin{bmatrix} u \\ \mathbf{v} \end{bmatrix} = \begin{bmatrix} 1 & \boldsymbol{\zeta}_{0,2}' \\ \mathbf{0}_{p\times 1} & I_{p-1} \end{bmatrix} \boldsymbol{\beta},$$

The Jacobian of this transformation is seen to be 1.

The range of integration for $\boldsymbol{\beta}$ includes all values for which $\boldsymbol{\beta}'\boldsymbol{\beta} \leq M$. In terms of the transformed variables, this range includes values of u, $\mathbf{v}$ with

$$\begin{aligned} M &\geq (u - \boldsymbol{\zeta}_{0,2}'\mathbf{v})^2 + \mathbf{v}'\mathbf{v} \\ &= u^2 - 2u\boldsymbol{\zeta}_{0,2}'\mathbf{v} + \mathbf{v}'\boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}'\mathbf{v} + \mathbf{v}'\mathbf{v} \\ &= u^2 + (\mathbf{v} - u(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2})'(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')(\mathbf{v} - u(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2}) \\ &\quad - u^2\boldsymbol{\zeta}_{0,2}'(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2}. \end{aligned} \tag{A.6}$$

We note that

$$\begin{aligned} I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}' &= \begin{vmatrix} I & \boldsymbol{\zeta}_{0,2} \\ \boldsymbol{\zeta}_{0,2}' & 1 \end{vmatrix} \\ &= \begin{vmatrix} I & \boldsymbol{\zeta}_{0,2} \\ \mathbf{0}' & 1 + \boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2} \end{vmatrix} \\ &= 1 + \boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}. \end{aligned}$$

We may therefore replace the quantity $1 - \boldsymbol{\zeta}_{0,2}'(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2}$, which appears as the multiplier of u^2 in (A.6), with

$$1 - \frac{\boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}}{1 + \boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}} = \frac{1}{1 + \boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}}.$$

The range for $(u, \mathbf{v})$ can thus be written as

$$\frac{u^2}{1 + \boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}} + (\mathbf{v} - u(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2})'(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')(\mathbf{v} - u(I + \boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2}) \leq M. \tag{A.7}$$

The first integral in (A.5), in terms of u and $\mathbf{v}$, can now be written

$$\int_{u,\mathbf{v}} \log|u| \; du d\mathbf{v}, \tag{A.8}$$

over the range defined by (A.7). We emphasize that this range depends on x_0, through $\boldsymbol{\zeta}_{0,2}$.

We now make another change of variables, defined by

$$s = \frac{u}{\sqrt{1+\boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}}}$$
$$\mathbf{t} = (I+\boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{\frac{1}{2}}\big(\mathbf{v} - u(I+\boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{-1}\boldsymbol{\zeta}_{0,2}\big),$$

where

$$(I+\boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')^{\frac{1}{2}}$$

is a square root of the matrix $(I+\boldsymbol{\zeta}_{0,2}\boldsymbol{\zeta}_{0,2}')$. Since s depends on u only, while $\mathbf{t}$ depends on both u and $\mathbf{v}$, Jacobian of this transformation is easily shown to be 1. The integral (A.8), in terms of s and $\mathbf{t}$, is over the range of values of s and $\mathbf{t}$ satisfying

$$s^2 + \mathbf{t}'\mathbf{t} \leq M,$$

and the integral is

$$\int_{s,\mathbf{t}} \log\left(s\sqrt{1+\boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}}\right)\, dsd\mathbf{t} = \log\left(\sqrt{1+\boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}}\right)\int_{s,\mathbf{t}}\, dsd\mathbf{t} + \int_{s,\mathbf{t}} \log s\; dsd\mathbf{t}.$$

Since

$$\int_{s,\mathbf{t}}\, dsd\mathbf{t} = \int_{\boldsymbol{\beta}} d\boldsymbol{\beta},$$

the term including the first integral in (A.5) is equal to $\log(1+\boldsymbol{\zeta}_{0,2}'\boldsymbol{\zeta}_{0,2}) = \log(\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0)$. We thus have

$$\exp\left\{\frac{1}{2}E_\theta^l[\log|h_1|\big|x_0]\right\} = K_5\left(\frac{\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0}{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}\right)^{\frac{1}{2}},$$

where K_5 depends only on the first integral in (A.5). Now

$$\pi_\theta^l(x_0,(\alpha,\boldsymbol{\beta}),\sigma) \propto \sigma^{-1}\left(\frac{\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0}{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}\right)^{\frac{1}{2}},$$

for $(x_0,(\alpha,\boldsymbol{\beta}),\sigma) \in \boldsymbol{\Theta}^l$.

We now let $\alpha_2, \sigma_2, x_{0,2}$, and M tend to $+\infty$, α_2 and $x_{0,2}$ tend to $-\infty$, and σ_1 tend to 0. The reference prior is thus

$$\pi^R(x_0,(\alpha,\boldsymbol{\beta}),\sigma) \propto \sigma^{-1}\left(\frac{\boldsymbol{\zeta}_0'\boldsymbol{\zeta}_0}{1+c\boldsymbol{\xi}_0'(\mathbf{X}_{\alpha,1}'\mathbf{X}_{\alpha,1})^{-1}\boldsymbol{\xi}_0}\right)^{\frac{1}{2}}.$$

Bibliography

Afifi, A. A. and Azen, S. P. (1972). *Statistical Analysis: A Computer Oriented Approach*. Academic Press.

Aitchison, J. and Dunsmore, I. (1975). *Statistical Prediction Analysis*. Cambridge University Press.

Basilevsky, A. (1983). *Applied Matrix Algebra in the Statistical Sciences*. North-Holland.

Berger, J. O. and Bernardo, J. M. (1992). On the development of reference priors. In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*. (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith eds.). Clarendon Press: Oxford. 35–60.

Berkson, J. (1969). Estimation of a linear function for a calibration line; consideration of a recent proposal. *Technometrics*. **11**, 649–660.

Bernardo, J. M. (1979). Reference posterior distributions for Bayesian inference. *Journal of the Royal Statistical Society B*. **41**, 113–147.

Bernardo, J. M. and Smith, A. F. M. (1994). *Bayesian Theory*. John Wiley and Sons.

Brown, P. J. (1993). *Measurement, Regression, and Calibration*. Oxford University Press.

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

Cox, D. and Reid, N. (1987). Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society B*. **49**, 1–39.

Datta, G. S. and Ghosh, J. K. (1995). On priors providing frequentist validity for Bayesian inference. *Biometrika*. **82**, 37–45.

du Plessis, J., van der Merwe, A., and Groenewald, P. (1995). Reference priors for the multivariate calibration problem. *South African Statistical Journal*. **29**, 155–168.

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

Eno, D. R. and Ye, K. (1998). Noninformative priors for calibration problems. In *Proceedings of the 1998 Joint Statistical Meetings, Biometrics section*.

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

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*. **85**, 398–409.

Ghosh, M., Carlin, B. P., and Srivastava, M. S. (1995). Probability matching priors for linear calibration. *Test*. **4**, 333–357.

Gilks, W. R., Richardson, D. J., and Spiegelhalter, D. J. (1996). *Markov Chain Monte Carlo In Practice*. Chapman and Hall.

Hill, B. M. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics*. **23**, 335–338.

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

Hunter, W. and Lamboy, W. (1981). A Bayesian analysis of the linear calibration problem. *Technometrics*. **23**, 323–328.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Statistical Society of London, Series A*. **186**, 453–461.

Kass, R. E. and Wasserman, L. (1996). The selection of prior distributions by formal rules. *Journal of the American Statistical Association*. **91**, 1343–1370.

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

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

Kubokawa, T. and Robert, C. P. (1994). New perspectives on linear calibration. *Journal of Multivariate Analysis*. **51**, 178–200.

Lawless, J. F. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics*. **23**, 334–335.

Lwin, T. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics*. **23**, 339–341.

Lwin, T. and Maritz, J. S. (1982). An analysis of the linear calibration controversy from the perspective of compound estimation. *Technometrics*. **24**, 235–242.

Orban, J. E. (1981). Discussion of Hunter and Lamboy's paper. *Technometrics*. **23**, 342–343.

Osbourne, C. (1991). Statistical calibration: A review. *International Statistical Review*. **59**, 309–336.

Philippe, A. and Robert, C. P. (1998). A note on the confidence properties of reference priors for the calibration model. *Test*. **7**, 147–160.

Press, S. J. (1972). *Applied Multivariate Analysis*. Holt, Rinehart, and Winston, Inc.

Raftery, A. E. and Lewis, S. M. (1992). How many iterations of the Gibbs sampler? In *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*. (J.M. Bernardo, J.O. Berger, A.P. Dawid, and A.F.M. Smith eds.). Clarendon Press: Oxford. 763–773.

Rubinstein, R. Y. (1981). *Simulation and the Monte Carlo Method*. John Wiley and Sons.

Stein, C. M. (1985). On the coverage probability of confidence sets based on a prior distribution. In *Sequential Methods in Statistics*. Banach Center Publications. **16**. Warsaw: PWN – Polish Scientific Publishers. 485–514.

Tibshirani, R. (1989). Noninformative priors for one parameter of many. *Biometrika*. **76**, 604–608.

Welsh, B. L. and Peers, H. W. (1963). On formulae for confidence points based on intervals of weighted likelihoods. *Journal of the Royal Statistical Society B*. **25**, 318–329.

Williams, E. J. (1969). A note on regression methods in calibration. *Technometrics*. **11**, 189–192.

Vita

Daniel R. Eno, son of Phillip and Elaine Eno, was born in Oswego, New York on August 26, 1969. He grew up in Martville, New York, with his parents and his brother, David Eno. In 1987, he graduated as salutatorian from Hannibal Central High School.

Between 1987 and 1991, Daniel attended Alfred University, where he received a Bachelor of Arts degree with majors in Physics and Mathematics in 1991. He holds a Master of Science degree in Mathematics from Virginia Tech (1993) and Master of Science degree in Statistics from Virginia Tech (1996). He was recipient of the Boyd Harshbarger Award (1996) and the Klaus Hinkelmann Award (1998) from the Virginia Tech Statistics Department. He will be awarded the degree of Doctor of Philosophy in Statistics from Virginia Tech in May, 1999.