The Statistician (2003)
52, *Part* 1, *pp.* 15–39

Inferences in multivariate–univariate calibration problems

Denise Benton and K. Krishnamoorthy

University of Louisiana at Lafayette, USA

and Thomas Mathew

University of Maryland Baltimore County, Baltimore, USA

[Received October 2001. Revised August 2002]

Summary. A normally distributed vector response variable is considered, related to a scalar explanatory variable through a linear regression model. The calibration data, i.e. data obtained on the response variable corresponding to known values of the explanatory variable, are to be used for making inferences concerning unknown values of the explanatory variable. The purpose of the paper is a detailed investigation of interval estimation and hypothesis testing problems that arise in this context. Such problems are addressed in the scenario of both single use and multiple use of the calibration data. In the single-use situation, the calibration data are used to make inferences concerning a single unknown value of the explanatory variable. In the multiple-use scenario, the calibration data are used to make inferences concerning a sequence of unknown values of the explanatory variable. A one-sided hypothesis testing problem is addressed and test procedures are developed in the context of both single use and multiple use of the calibration data. Since the test statistic has a distribution that depends on some unknown parameters, a parametric bootstrap procedure is advocated to carry out the test. The parametric bootstrap procedure can be adopted for interval estimation as well. The performance of the parametric bootstrap procedure is numerically investigated and is found to be quite satisfactory. Some examples are used to motivate the problems and to illustrate the applicability of our results. The overall conclusion is that the parametric bootstrap is a simple and satisfactory approach for making inferences in the calibration problem.

Keywords: Bootstrap; Calibration data; Multiple-use confidence region; Multiple-use hypothesis testing; Parametric bootstrap; Percentile bootstrap; Single-use confidence region; Single-use hypothesis testing

1. Introduction

1.1. Statement of the problem

The multivariate–univariate calibration problem involves a $p \times 1$ response variable $\mathbf{y}$ related to a scalar explanatory variable x, where $\mathbf{y}$ is assumed to be random. This paper considers the scenario where x is a non-random quantity and the relationship between $\mathbf{y}$ and x is a normal linear regression model. Let $(\mathbf{y}_i, x_i)$, $i = 1, 2, \ldots, n$, denote n independent observations $\mathbf{y}_i$ on the response variable, corresponding to the value x_i of the explanatory variable. Our assumed model is

$$\mathbf{y}_i \sim N_p(\boldsymbol{\alpha} + \boldsymbol{\beta} x_i, \Sigma), \tag{1.1}$$

Address for correspondence: Thomas Mathew, Department of Mathematics and Statistics, University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250, USA.
E-mail: mathew@math.umbc.edu

 0039–0526/03/52015

$i = 1, 2, \ldots, n$, where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are unknown $p \times 1$ parameter vectors and Σ is an unknown $p \times p$ positive definite matrix. Now let $\mathbf{y}_0$ be a value of the response variable corresponding to an unknown value θ of the explanatory variable. Similarly to model (1.1), we assume that

$$\mathbf{y}_0 \sim N_p(\boldsymbol{\alpha} + \boldsymbol{\beta}\theta, \Sigma), \tag{1.2}$$

where $\mathbf{y}_0$ is assumed to be independent of the $\mathbf{y}_i$s in model (1.1). The problem of calibration or inverse regression consists of statistical inference concerning θ. The $(\mathbf{y}_i, x_i)$s in model (1.1) are referred to as the calibration data. In the context of models (1.1) and (1.2), the problem is known as the multivariate–univariate calibration problem since the response variable is a vector and the explanatory variable (and hence the parameter of interest) is a scalar. The problems addressed in this paper deal with the construction of confidence regions for θ and hypothesis tests concerning θ. Given below are several examples where such multivariate–univariate calibration problems arise in practice and where models (1.1) and (1.2) may be used.

1.2. Some examples

1.2.1. *Example 1: measuring fat content of meat using near infra-red absorbance measurements*

Samples of meat can be analysed for their fat content by using near infra-red absorbance measurements. The data that we shall consider consist of 214 samples of meat and measurements for each sample consist of a 100-channel spectrum of absorbances, and the contents of moisture, fat and protein, as determined by analytical chemistry. (Details of the data are available at `http://lib.stat.cmu.edu/datasets/tecator`; data were obtained by using a Tecator Infratec Food and Feed Analyzer.) The providers of the data suggest using the first 172 samples to fit a model and using the first 13 principal components, scaled to unit variance, to predict the fat content. In fitting a linear model, only three principal components were significant. Suppose that $\mathbf{y} = (y_1, y_2, y_3)'$ represents the vector of these three principal components, and let $\mathbf{y}_i$ denote the value of $\mathbf{y}$ for the ith sample, with a known value x_i for the fat content (as a percentage). We assume model (1.1), where $p = 3$. Furthermore, if $\mathbf{y}_0$ denotes the value of $\mathbf{y}$ for a meat sample with unknown fat content θ, we have model (1.2).

As an example of hypothesis testing, consider

$$H_0 : \theta \leqslant 10\% \textit{ versus } H_1 : \theta > 10\%, \tag{1.3}$$

which is of interest in determining whether a sample of ground meat contains less than 10% fat. Current methods of estimating the fat content of a batch of ground meat involve taking a sample of meat and then frying it in a pan. The amount of fat that drips off is then compared with the weight of the patty to estimate the percentage of fat. The process obviously becomes considerably simpler if the fat content (i.e. θ in model (1.2)) can be estimated by using the $\mathbf{y}_i$s in model (1.1) and $\mathbf{y}_0$ in model (1.2). Confidence intervals and tests (e.g. of the type (1.3)) are certainly of interest in this context.

The problem of estimating fat content in ground meat is also discussed in Martens and Naes (1989), pages 108–112.

1.2.2. *Example 2: the paint finish data*

The paint finish data first appeared in Brown (1982) and were later analysed by Brown and Sundberg (1987), Oman (1988) and Mathew and Zha (1996). In this example, x is a scalar representing the viscosity of the paint samples and $\mathbf{y}$ is a bivariate observation vector consisting of two measurements on certain optical properties of the samples. The viscosity of each paint sample was one of three different values, coded as -1, 0 and 1 in Brown and Sundberg (1987). 27 observations were available for calibration and, if $\mathbf{y}_i$ denotes the observation corresponding to a known viscosity x_i, the model used by Brown and Sundberg (1987) is model (1.1). The problem is to make inferences on the unknown viscosity, say θ, corresponding to a measurement $\mathbf{y}_0$, where model (1.2) applies.

1.2.3. Example 3: the wheat data

The wheat data set also appeared in Brown (1982) and was later used by Fox (1989) to exhibit conditional multivariate calibration. The data set consists of measurements on 21 samples of hard wheat. Accurate percentages of water and protein were obtained along with four derived infra-red reflectance measurements. Our interest is in making inference on the percentage of water in a wheat sample by using the reflectance measurements. It turns out that only the third and fourth measurements are useful in predicting the percentage of water. Thus, if $\mathbf{y}$ is a bivariate vector denoting the third and fourth reflectance measurements, and if x denotes the percentage of water, we have models similar to models (1.1) and (1.2). The problem will be to make inferences concerning θ, the unknown percentage of water.

1.2.4. Example 4: chlorophenol example

Chlorophenols are chemicals that are often used in herbicides and wood preservatives, but they have a high toxicity and are considered to be dangerous to the environment. Methods of detection are expensive and time consuming because they are based on separation analytical techniques. da Silva and Laquipai (1998) discussed using ultraviolet spectrophotometric measurements to predict the concentration of a particular chlorophenol in a solution, resulting in a quick method to screen for chlorophenols. In practice, if the concentration is above a certain level, action must be taken. Thus we have a multivariate–univariate calibration model, where we are interested in testing hypotheses concerning θ, an unknown level of chlorophenol, by using an observed vector of spectrophotometric measurements.

1.2.5. Example 5: olefin example

Near infra-red spectroscopy and multivariate calibration methods are used to estimate the concentrations of olefins in the processing of polyethylene. As a part of the on-line process control, it is of interest to determine whether the olefin concentration is above or below a specification limit. This allows the plant to make a consistent product, minimizing the amount of offgrade produced. For details, see Seasholtz (1999).

1.3. Literature review and summary of the results

It should be clear from the above examples that the multivariate–univariate calibration problem arises in many applications and the problems of point and interval estimation as well as hypothesis testing are relevant in these applications. This paper deals with the problems of interval estimation and hypothesis testing. In the context of models (1.1) and (1.2), the following hypothesis testing problems are considered:

$$H_0 : \theta = c \ \ versus \ \ H_1 : \theta \neq c, \tag{1.4a}$$

$$H_0 : \theta \leqslant c \ \ versus \ \ H_1 : \theta > c, \tag{1.4b}$$

$$H_0 : \theta \geqslant c \ \ versus \ \ H_1 : \theta < c, \tag{1.4c}$$

where c is known (for example, $c = 10\%$ in problem (1.3)).

The literature on calibration distinguishes between two scenarios regarding the use of the calibration data (i.e. the $(\mathbf{y}_i, x_i)$s in model (1.1)), for the purpose of statistical inference concerning θ in model (1.2). The first is the single use of the calibration data, in which the set of $(\mathbf{y}_i, x_i)$s in model (1.1) is used for obtaining confidence regions and hypothesis tests for a single parameter θ corresponding to a single observation $\mathbf{y}_0$ in model (1.2). This results in single-use confidence regions and single-use hypothesis tests. We also have multiple-use confidence regions and multiple-use hypothesis tests, where the calibration data are used repeatedly to construct confidence regions and hypothesis tests for a sequence of parameter values, similar to θ in model (1.2), corresponding to a sequence of observations, similar to $\mathbf{y}_0$ in model (1.2). The distinction between single use and multiple use of the calibration data is clarified and explained in Mee and Eberhardt (1996) for the univariate case. They have also given a clear discussion of the criteria to be used for the construction of single-use and multiple-use confidence regions. Single-use and multiple-use hypothesis tests are discussed in Krishnamoorthy *et al.* (2001), once again for the univariate case. This paper considers both single-use and multiple-use procedures. The criteria to be used for these purposes are given below.

A natural procedure for obtaining a test statistic for testing the hypotheses in expressions (1.4) is first to assume that $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and Σ are known in model (1.2). When this is the case, the classical estimator of θ can be obtained on the basis of model (1.2) and a test statistic can be easily obtained for testing the hypotheses in expressions (1.4). The unknown quantities $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and Σ in the test statistic can now be replaced by estimates obtained by using the calibration data, i.e. by using the $(\mathbf{y}_i, x_i)$s in model (1.1). The same can be done for obtaining a pivot statistic for constructing a confidence region for θ. Though this approach is intuitively appealing, the difficulty that we face is that the resulting test statistic (or pivot statistic) has a distribution that depends on nuisance parameters. In the context of confidence regions, this difficulty is overcome by constructing conservative confidence regions; see Mathew and Zha (1996, 1997, 1998) and Mathew *et al.* (1998). It should be pointed out, however, that in the univariate case (i.e. $p = 1$) there are test statistics and pivot statistics which have distributions free of any unknown parameters, and Fieller's theorem provides an exact single-use confidence region; see Krishnamoorthy *et al.* (2001) where multiple-use hypothesis testing is addressed in the univariate case.

Confidence regions can clearly be used for testing hypothesis (1.4a). The exact confidence region derived in Mathew and Kasala (1994) and the conservative confidence region derived in Mathew and Zha (1996) can both be used for deriving single-use two-sided hypothesis tests. Our numerical study indicates that the test based on the exact confidence region derived in Mathew and Kasala (1994) is biased and should be used only in some special situations (see Section 3.3). A pivot statistic presented in the present paper is similar to that used by Mathew and Zha (1996) to construct confidence regions. Specifically, the square of our pivot is equivalent to that of Mathew and Zha (1996). Thus our pivot statistic can be used for one-sided hypothesis testing. Furthermore, we have used the parametric bootstrap (PB) to take care of the dependence of the distribution of the pivot statistic on the unknown vector of slope parameters.

Several researchers have already applied nonparametric bootstrap techniques to the calibration problem; see Jones and Rocke (1999) for details and further references. Although most of the references in this context deal with the univariate case, Jones and Rocke (1999) also considered a multivariate non-linear calibration problem. In general, the bootstrap method can be applied in both parametric and nonparametric modes. Efron and Tibshirani (1993) devoted a few sections on parametric bootstrapping and illustrated the PB method to compute the standard error of the sample correlation coefficient in the bivariate normal distribution. Lee (1994) concluded that a PB result may be more accurate than its nonparametric version provided that the model assumption is at least approximately correct. In our set-up, the bootstrapping is done parametrically, i.e. under models (1.1) and (1.2). Section 2 provides results along this direction for obtaining confidence regions. Section 3 deals with the PB procedure for obtaining single-use hypothesis tests. In Section 4, we illustrate our results and, in particular, the computational procedures, by using some examples. For this, we have used data sets based on the first three examples in Section 1.2. Section 5 deals with multiple-use hypothesis testing. Some concluding remarks appear in Section 6.

The major thrust of the present paper is to explore the role of the PB in constructing confidence regions and tests in a multivariate–univariate calibration problem. We describe the PB and also numerically study its performance. The numerical results show that the PB procedure is indeed satisfactory. Furthermore, the procedure is applicable for one-sided hypothesis testing as well. As expected, some computational effort is required to compute the PB confidence intervals and hypothesis tests. However, the required computational effort is much less compared with what is required to compute the conservative confidence regions in Mathew and Zha (1996, 1997, 1998) and Mathew *et al.* (1998). Our final conclusions are that the PB method is simple to use and works quite well in the calibration problem.

2. Single-use confidence regions

In this section, we shall construct a single-use confidence region for θ in model (1.2), based on the $(\mathbf{y}_i, x_i)$s in model (1.1) and $\mathbf{y}_0$ in model (1.2). Define

$$\left.\begin{aligned} \bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i, \\ \bar{\mathbf{y}} &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i, \\ s_{xx} &= \sum_{i=1}^{n} (x_i - \bar{x})^2, \\ S &= \sum_{i=1}^{n} (\mathbf{y}_i - \hat{\alpha} - \hat{\beta} x_i)(\mathbf{y}_i - \hat{\alpha} - \hat{\beta} x_i)', \end{aligned}\right\} \tag{2.1}$$

where $\hat{\alpha}$ and $\hat{\beta}$ denote the least squares estimators of α and β based on model (1.1). Clearly, $\mathbf{y}_0, \bar{\mathbf{y}}, \hat{\beta}$ and S are independently distributed with

$$\left.\begin{aligned} \bar{\mathbf{y}} &\sim N_p\left(\alpha + \beta\bar{x}, \frac{1}{n}\Sigma\right), \\ \hat{\beta} &\sim N_p\left(\beta, \frac{1}{s_{xx}}\Sigma\right), \\ S &\sim W_p(n-2, \Sigma), \end{aligned}\right\} \tag{2.2}$$

where $W_p(n-2, \Sigma)$ denotes the p-dimensional Wishart distribution with $n-2$ degrees of freedom and scale matrix Σ.

2.1. A pivot statistic

If β and Σ are known, then the least squares estimator of θ, say $\tilde{\theta}$, is given by

$$\tilde{\theta} = \bar{x} + \frac{\beta'\Sigma^{-1}(\mathbf{y}_0 - \bar{\mathbf{y}})}{\beta'\Sigma^{-1}\beta}.$$

Consequently,

$$Z_0 = \left(\frac{\beta'\Sigma^{-1}\beta}{1+1/n}\right)^{1/2}(\tilde{\theta}-\theta)$$
$$= \frac{\beta'\Sigma^{-1}\{\mathbf{y}_0 - \bar{\mathbf{y}} - \beta(\theta - \bar{x})\}}{\{(1+1/n)\beta'\Sigma^{-1}\beta\}^{1/2}} \sim N(0,1). \quad (2.3)$$

The quantity Z_0 in equation (2.3) can be used as a pivot for obtaining a confidence region for θ when β and Σ are known. Since β and Σ are actually unknown, we replace them by the estimates $\hat{\beta}$ and

$$\hat{\Sigma} = \frac{1}{n-2}S.$$

The resulting pivot statistic, denoted by Q, is

$$Q = \frac{\hat{\beta}'S^{-1}\{\mathbf{y}_0 - \bar{\mathbf{y}} - \hat{\beta}(\theta - \bar{x})\}}{\{(1+1/n)\hat{\beta}'S^{-1}\hat{\beta}\}^{1/2}}, \quad (2.4)$$

where we have ignored the constant $1/\surd(n-2)$ in Q. Let

$$\left.\begin{aligned} \mathbf{u} &= \hat{\beta}\surd s_{xx}, \\ \beta_1 &= \beta\surd s_{xx}, \\ \mathbf{v}_0 &= \frac{\mathbf{y}_0 - \bar{\mathbf{y}}}{\surd(1+1/n)}, \\ \theta_1 &= \frac{\theta - \bar{x}}{\surd\{s_{xx}(1+1/n)\}}. \end{aligned}\right\} \quad (2.5)$$

Then

$$\begin{aligned} \mathbf{u} &\sim N_p(\beta_1, \Sigma), \\ \mathbf{v}_0 &\sim N_p(\beta_1\theta_1, \Sigma), \end{aligned} \quad (2.6)$$

and Q in equation (2.4) can be expressed as

$$Q = \frac{\mathbf{u}'S^{-1}(\mathbf{v}_0 - \mathbf{u}\theta_1)}{(\mathbf{u}'S^{-1}\mathbf{u})^{1/2}}. \quad (2.7)$$

A slight modification of Q is necessary (see Appendix A) to study its distribution. The modified pivot statistic, say T_0, is given by

$$\begin{aligned} T_0 &= \frac{(n-p-1)^{1/2}\mathbf{u}'S^{-1}(\mathbf{v}_0 - \mathbf{u}\theta_1)}{\{(1+r)\mathbf{u}'S^{-1}\mathbf{u}\}^{1/2}}, \\ r &= \mathbf{v}_0'S^{-1}\mathbf{v}_0 - \frac{(\mathbf{v}_0'S^{-1}\mathbf{u})^2}{\mathbf{u}'S^{-1}\mathbf{u}}. \end{aligned} \quad (2.8)$$

Note that T_0 can be easily expressed in terms of the original variables in models (2.1) and (2.2). Note also that it is enough to construct a confidence region for θ_1, since a confidence region for θ can then be obtained by using the relationship between θ_1 and θ, given in equations (2.5). The pivots Q^2 and T_0^2 (or other equivalent forms) have been considered by Williams (1959), Wood (1982), Fujikoshi and Nishii (1984), Davis and Hayakawa (1987) and Mathew and Zha (1996).

The asymptotic results in Fujikoshi and Nishii (1984) and Davis and Hayakawa (1987) and the conservative confidence regions derived by Mathew and Zha (1996) are all based on the pivot statistic T_0^2.

2.2. *The distribution of* T_0

From the distributional results in expressions (2.2) and (2.6), and from the expression for T_0 in equation (2.8), it follows that the distribution of T_0 is invariant under non-singular transformations, i.e. it is invariant under the transformation $\mathbf{u} \rightarrow P\mathbf{u}$, $\mathbf{v}_0 \rightarrow P\mathbf{v}_0$ and $S \rightarrow PSP'$, where P is any non-singular matrix. Hence, without loss of generality, we shall assume throughout this paper that $\Sigma = I_p$. The proof of the following theorem is given in Appendix A.

Theorem 1. Let T_0 and r be as defined in equation (2.8) and assume that $\Sigma = I_p$. Then, conditionally given $\mathbf{u}$ and r, T_0 is distributed as $t_{n-p-1}(\delta_0)$, the non-central t-distribution with $n-p-1$ degrees of freedom and non-centrality parameter

$$\delta_0 = \frac{\mathbf{u}'(\boldsymbol{\beta}_1 - \mathbf{u})\theta_1}{\{\mathbf{u}'\mathbf{u}(1+r)\}^{1/2}}. \tag{2.9}$$

The distribution given in theorem 1 is not directly useful for obtaining confidence regions owing to the dependence on the unknown slope parameter $\boldsymbol{\beta}_1$. Following the arguments of Mathew and Zha (1996), a conservative confidence region can be obtained by using a larger non-centrality parameter than δ_0 in equation (2.9). In particular, using the inequality $\mathbf{u}'(\boldsymbol{\beta}_1 - \mathbf{u})\theta_1/\surd(\mathbf{u}'\mathbf{u}) \leqslant \{\theta_1^2(\boldsymbol{\beta}_1 - \mathbf{u})'(\boldsymbol{\beta}_1 - \mathbf{u})\}^{1/2} = \surd(\theta_1^2 V)$, we have $\delta_0 \leqslant \surd(\theta_1^2 V)$. Noticing that V follows a central χ^2-distribution with p degrees of freedom (under our assumption $\Sigma = I_p$), and replacing δ_0 by $\surd(\theta_1^2 V)$, a conservative confidence region can be obtained, as in Mathew and Zha (1996).

Since developing exact inferential procedures based on T_0 is not possible, and simple approximate inferential procedures based on T_0 seem to be difficult, we resort to the PB method.

2.3. *The parametric bootstrap method for confidence regions*

The PB method can be briefly explained as follows. Let $X_1, \ldots, X_n$ be a sample of independent observations from a population with distribution function $F(x; \eta)$, where η is an unknown parameter. Suppose that we want to estimate a real-valued function $g(\eta)$. Let $\hat{\eta}$ be an estimator of η based on the sample. Then, $F(x; \hat{\eta})$ is called the PB estimate of the population distribution function $F(x; \eta)$. The samples $X_{i1}, \ldots, X_{in}$, $i = 1, \ldots, m$, generated from $F(x; \hat{\eta})$ are called PB samples. The sampling distribution of $g(\hat{\eta})$ can be approximated by the empirical distribution of $g(\hat{\eta}_i)$s, where $\hat{\eta}_i$ is the estimate of η based on the ith PB sample from $F(x; \hat{\eta})$. In many situations, $\hat{\eta}_i$s can be generated directly from the sampling distribution of $\hat{\eta}$. The sampling distribution may be used to make inferences about $g(\eta)$. For example, when m is sufficiently large, $(g(\hat{\eta})_{(\alpha m/2)}, g(\hat{\eta})_{((1-\alpha/2)m)})$, where $g(\hat{\eta})_{(k)}$ denotes the kth smallest of the $g(\hat{\eta}_i)$s, serves as an approximate $100(1-\alpha)\%$ confidence interval for $g(\eta)$. This is referred to as the percentile bootstrap.

We can also check the performance of a PB method for a specific model. For example, let η_0 be a value chosen arbitrarily from the parameter space of η. Let $\hat{\eta}_1, \ldots, \hat{\eta}_m$ be estimates based on m samples generated from $F(x; \eta_0)$. Let $\hat{\eta}_{1i}, \ldots, \hat{\eta}_{ki}$ be estimates generated from $F(x; \hat{\eta}_i)$, and let I_i denote the PB confidence interval for $g(\eta)$, $i = 1, \ldots, m$. If the proportion of I_is containing $g(\eta_0)$ is approximately equal to the confidence level specified, then we conclude that the PB method will yield satisfactory results for the model.

The PB confidence interval based on the pivot statistic T_0 can be obtained as follows.

(a) For a given confidence level $1-\alpha$, let $k_1(\theta_1)$ and $k_2(\theta_1)$ be such that

$$P\{k_1(\theta_1) \leqslant T_0 \leqslant k_2(\theta_1)\} = 1-\alpha.$$

(b) By inverting the inequalities in step (a) for θ_1, we obtain the lower and upper confidence limits, say θ_{1l} and θ_{1u} respectively, as

$$\theta_{1l} = \hat{\theta}_1 + k_1(\theta_1)\left\{\frac{1+r}{(n-p-1)\mathbf{u}'S^{-1}\mathbf{u}}\right\}^{1/2},$$

$$\theta_{1u} = \hat{\theta}_1 + k_2(\theta_1)\left\{\frac{1+r}{(n-p-1)\mathbf{u}'S^{-1}\mathbf{u}}\right\}^{1/2},$$

where

$$\hat{\theta}_1 = \frac{\mathbf{u}'S^{-1}\mathbf{v}_0}{\mathbf{u}'S^{-1}\mathbf{u}}. \tag{2.10}$$

(c) To estimate the values of $k_1(\theta_1)$ and $k_2(\theta_2)$, compute $\mathbf{u}$ and $\hat{\theta}_1$ using calibration data and $\mathbf{y}_0$. For $i=1, 100000$, generate $\mathbf{u}_i \sim N_p(\mathbf{u}, I_p)$, $\mathbf{v}_{0i} \sim N_p(\mathbf{u}\hat{\theta}_1, I_p)$ and $S_i \sim W_p(n-2, I_p)$. Set

$$T_{0i} = \frac{(n-p-1)^{1/2}\mathbf{u}_i'S_i^{-1}(\mathbf{v}_{0i} - \mathbf{u}_i\hat{\theta}_1)}{\{(1+r_i)\mathbf{u}_i'S_i^{-1}\mathbf{u}_i\}^{1/2}},$$

where r_i denotes the value of r in equation (2.8) computed by using $\mathbf{u}_i$, $\mathbf{v}_{0i}$ and S_i respectively in place of $\mathbf{u}$, $\mathbf{v}_0$ and S.

(d) Let $T_0(a)$ denote the 100ath percentile of the T_{0i}s in step (c). Then, $k_1(\theta_1) \approx T_0(\alpha/2)$ and $k_2(\theta_1) \approx T_0(1-\alpha/2)$.

Note that we can also construct a confidence interval for θ_1 by using the percentile bootstrap, i.e. based on the PB sampling distribution of $\hat{\theta}_1$ in equation (2.10). The procedure is as follows.

(a) For $i=1, 100000$, generate $\mathbf{u}_i \sim N_p(\mathbf{u}, I_p)$, $\mathbf{v}_{0i} \sim N_p(\mathbf{u}\hat{\theta}_1, I_p)$ and $S_i \sim W_p(n-2, I_p)$. Let $\hat{\theta}_{1i} = \mathbf{u}_i'S_i^{-1}\mathbf{v}_{0i}/\mathbf{u}_i'S_i^{-1}\mathbf{u}_i$.

(b) Let $\hat{\theta}_1(a)$ be the 100ath percentile of the $\hat{\theta}_{1i}$s generated in step (a). Then, for a given confidence level $1-\alpha$, $(\hat{\theta}_1(\alpha/2), \hat{\theta}_1(1-\alpha_2))$ is an approximate $100(1-\alpha)\%$ confidence interval for θ_1.

2.4. Performance of the parametric bootstrap method

To understand the validity of the PB procedure, and to compare the PB results based on $\hat{\theta}_1$ and the pivot statistic T_0, we estimated the coverage probabilities of the PB confidence intervals by using the evaluation method described in Section 2.3. For example, for a given $\boldsymbol{\beta}_1$ and θ_1, the coverage probability of the confidence interval based on T_0 can be estimated as follows.

2.4.1. *Estimation of the coverage probabilities of the parametric bootstrap confidence intervals based on T_0*

(a) For $i=1, 1000$, generate $\mathbf{u}_i \sim N_p(\boldsymbol{\beta}_1, I_p)$ and $\mathbf{v}_{0i} \sim N_p(\boldsymbol{\beta}_1\theta_1, I_p)$ and $S_i \sim W_p(n-2, I_p)$.

(b) Compute T_{0i} by formula (2.8), using $\mathbf{u}_i$, $\mathbf{v}_{0i}$ and S_i in place of $\mathbf{u}$, $\mathbf{v}_0$ and S respectively. Consider this as an observed value of the test statistic T_0.

(c) For $j = 1, 1000$, generate $\mathbf{u}_{ij} \sim N_p(\mathbf{u}_i, I_p)$, $\mathbf{v}_{0ij} \sim N_p(\mathbf{u}_i\theta_1, I_p)$ and $S_j \sim W_p(n-2, I_p)$.
(d) Compute T_{0ij} by formula (2.8), using $\mathbf{u}_{ij}$, $\mathbf{v}_{0ij}$ and S_j in place of $\mathbf{u}$, $\mathbf{v}_0$ and S respectively. Note that T_{0ij} also depends on i. (End the j-loop.)
(e) For a fixed i ($i = 1, 2, \ldots, 1000$), let I_i denote the PB confidence interval for θ_1, obtained by using the percentiles of the T_{0ij}s, $j = 1, 2, \ldots, 1000$.
(f) The proportion of I_is that contain θ_1 provides an estimate of the coverage probability of the PB confidence interval based on T_0.

The coverage of the confidence interval based on the percentile bootstrap (i.e. based on the bootstrap distribution of $\hat{\theta}_1$) can be similarly estimated. Table 1 gives the estimated coverage probabilities and expected lengths of the PB confidence intervals based on T_0 and $\hat{\theta}_1$, for $\beta_1 = (1, 2, 3)'$, $n = 10, 20, 30$ and various values of θ_1. We observe from Table 1 that the coverage probabilities of the PB confidence intervals based on T_0 and $\hat{\theta}_1$ are almost equal to the specified level when $|\theta_1| \leqslant 1$. In most applications, the values of $|\theta_1|$ should not exceed 1; see remark 1 below. Furthermore, for $|\theta_1| \leqslant 1$, the expected lengths of the PB intervals based on T_0 and $\hat{\theta}_1$ are virtually equal.

Remark 1. It appears reasonable to assume that $|\theta_1| \leqslant 1$ in most practical applications. This can be explained as follows. Note that $v_x^2 = s_{xx}/(n-1)$ is the variance among the x_is, where s_{xx} is defined in expression (2.1). In carefully designed experiments, the range of the x_is will cover the range of practical interest, and θ will be within this range. At the very least, we can assume that $|\theta - \bar{x}| \leqslant 3v_x$. From the expression for θ_1 in equations (2.5), it now follows that, when $|\theta - \bar{x}| \leqslant 3v_x$, we have $|\theta_1| \leqslant 1$ for $n \geqslant 10$. If we assume that $|\theta - \bar{x}| \leqslant 2v_x$, then $|\theta_1| \leqslant 1$ for $n \geqslant 5$.

3. Single-use hypothesis tests

Since the single-use confidence region in the previous section can also provide a test procedure in the case of two-sided alternatives, i.e. hypotheses (1.4a), in this section we shall consider tests

Table 1. Estimated coverage probabilities and expected lengths of PB confidence intervals based on T_0 and $\hat{\theta}_1$ for $p = 3$ and $\beta_1' = (1, 2, 3)$†

θ_1	*Coverage probabilities and lengths for the following values of n:*					
	n = 10		*n = 20*		*n = 30*	
	T_0	$\hat{\theta}_1$	T_0	$\hat{\theta}_1$	T_0	$\hat{\theta}_1$
−2.0	0.93 (2.3)	0.93 (1.9)	0.92 (1.8)	0.92 (1.7)	0.93 (1.7)	0.93 (1.7)
−1.5	0.93 (1.9)	0.94 (1.6)	0.93 (1.5)	0.93 (1.4)	0.93 (1.4)	0.92 (1.3)
−1.0	0.95 (1.4)	0.95 (1.2)	0.94 (1.2)	0.95 (1.1)	0.94 (1.1)	0.95 (1.1)
−0.5	0.94 (1.1)	0.95 (1.0)	0.94 (0.9)	0.96 (0.9)	0.95 (0.9)	0.96 (0.9)
0.0	0.96 (0.9)	0.96 (1.0)	0.96 (0.8)	0.96 (0.8)	0.94 (0.8)	0.96 (0.8)
0.5	0.95 (1.2)	0.96 (1.0)	0.95 (0.9)	0.94 (0.9)	0.93 (0.9)	0.94 (0.9)
1.0	0.94 (1.4)	0.94 (1.2)	0.95 (1.2)	0.94 (1.1)	0.95 (1.1)	0.94 (1.1)
1.5	0.93 (1.9)	0.94 (1.6)	0.92 (1.5)	0.92 (1.4)	0.92 (1.4)	0.92 (1.3)
2.0	0.92 (2.4)	0.93 (1.9)	0.91 (1.9)	0.90 (1.7)	0.90 (1.7)	0.91 (1.7)

†Expected lengths are given in parentheses.

only for one-sided alternatives, i.e. the hypotheses (1.4b). Hypotheses (1.4c) can be handled similarly. Define

$$c_1 = \frac{c - \bar{x}}{\surd\{s_{xx}(1 + 1/n)\}}. \tag{3.1}$$

Hypotheses (1.4b) can now be expressed as

$$H_0 : \theta_1 \leqslant c_1 \text{ versus } H_1 : \theta_1 > c_1. \tag{3.2}$$

3.1. The tests

Let

$$\begin{aligned} Z_1 &= \frac{1}{\surd(1+1/n)} \frac{\boldsymbol{\beta}'\Sigma^{-1}\{\mathbf{y}_0 - \bar{\mathbf{y}} - \boldsymbol{\beta}(\bar{x} - c)\}}{(\boldsymbol{\beta}'\Sigma^{-1}\boldsymbol{\beta})^{1/2}}, \\ T_1 &= \frac{(n-p-1)^{1/2}\mathbf{u}'S^{-1}(\mathbf{v}_0 - \mathbf{u}c_1)}{\{(1+r)\mathbf{u}'S^{-1}\mathbf{u}\}^{1/2}}, \end{aligned} \tag{3.3}$$

where the various quantities in equations (3.3) are as defined in Section 2. We note that Z_1 in equations (3.3) is the quantity Z_0 in equation (2.3) evaluated at $\theta = c$. Similarly, T_1 in equations (3.3) is obtained by replacing θ_1 by c_1 in the expression for T_0 in equation (2.8). The statistic T_1 in equations (3.3) is a possible candidate for testing hypotheses (3.2). The use of T_1 is motivated by the same arguments that justified the use of T_0 in equation (2.8) for constructing confidence intervals. That is, Z_1 in equations (3.3) is the test statistic to be used for testing hypotheses (1.4b), i.e. hypotheses (3.2), when $\boldsymbol{\beta}$ and Σ are known, and T_1 is obtained from Z_1 by replacing $\boldsymbol{\beta}$ and Σ by their estimates.

The use of Z_1 in equations (3.3) is meaningful for testing hypotheses (3.2), since the distribution of Z_1 is stochastically increasing in θ_1, as can be easily verified. However, we cannot claim a similar property for T_1. Hence the performance of the test based on T_1 must be numerically evaluated. From theorem 1, it follows that when $\theta_1 = c_1$, and conditionally given $\mathbf{u}$ and r, T_1 is distributed as $t_{n-p-1}(\delta_1)$, with non-centrality parameter

$$\delta_1 = \frac{\mathbf{u}'(\boldsymbol{\beta}_1 - \mathbf{u})c_1}{\{(1+r)\mathbf{u}'\mathbf{u}\}^{1/2}}.$$

The test based on T_1 rejects H_0 in hypotheses (3.2) for large values of T_1.

A different test statistic can be obtained by using the following argument. When $\theta_1 = c_1$, $\boldsymbol{\beta}_1$ can be estimated by using $\mathbf{u}$ and $\mathbf{v}_0$ in equations (2.6). The resulting estimator $\hat{\boldsymbol{\beta}}_{1c}$ is given by

$$\hat{\boldsymbol{\beta}}_{1c} = \frac{\mathbf{u} + \mathbf{v}_0 c_1}{1 + c_1^2}.$$

Replacing $\boldsymbol{\beta}_1$ with $\hat{\boldsymbol{\beta}}_{1c}$ and Σ with $S/(n-2)$ in equation (2.3), and after some simplifications, we obtain a statistic that is proportional to the quantity

$$Q_* = \frac{\mathbf{u}_*' S^{-1}\mathbf{v}_{0*}}{(\mathbf{u}_*' S^{-1}\mathbf{u}_*)^{1/2}},$$

where

$$\mathbf{u}_* = \mathbf{u} + \mathbf{v}_0 c_1,$$
$$\mathbf{v}_{0*} = (\mathbf{v}_0 - \mathbf{u} c_1)/\surd(1 + c_1^2). \tag{3.4}$$

Note that the expression for Q_* is similar to the expression for Q in equation (2.7). To arrive at the distribution of Q_*, we shall consider the following modified statistic T_2 (see Appendix B):

$$T_2 = \frac{(n-p-1)^{1/2}\mathbf{u}_*' S^{-1}\mathbf{u}_{0*}}{\{(1+r_*)\mathbf{u}_*' S^{-1}\mathbf{u}_*\}^{1/2}},$$
$$r_* = \mathbf{v}_{0*}' S^{-1}\mathbf{v}_{0*} - \frac{(\mathbf{v}_{0*}' S^{-1}\mathbf{u}_*)^2}{\mathbf{u}_*' S^{-1}\mathbf{u}_*}. \tag{3.5}$$

Note that T_2 (and also T_1 in equations (3.3)) can be easily expressed in terms of the original variables in expressions (2.1) and (2.2). A pivot statistic equivalent to T_2^2 has been suggested by Mathew and Kasala (1994) for obtaining single-use confidence regions. The distribution of T_2 is given in the following theorem. The proof of theorem 2 is given in Appendix B.

Theorem 2. Let T_2 and r_* be as defined in equations (3.5) and assume that $\Sigma = I_p$. Then, conditionally given $\mathbf{u}_*$ and r_*, T_2 is distributed as $t_{n-p-1}(\delta_2)$, with non-centrality parameter

$$\delta_2 = \frac{\mathbf{u}_*' \boldsymbol{\beta}_1(\theta_1 - c_1)}{\{\mathbf{u}_*'\mathbf{u}_*(1+r_*)(1+c_1^2)\}^{1/2}}. \tag{3.6}$$

Now consider the test that rejects H_0 in hypothesis (3.2) for large values of T_2. Since $\mathbf{u}_* \sim N_p\{(1+\theta_1 c_1)\boldsymbol{\beta}_1, (1+c_1^2)I_p)\}$, $\mathbf{u}_*$ is more likely fall in the direction of $\boldsymbol{\beta}_1$ (provided that $\theta_1 c_1 > -1$), and hence $\mathbf{u}_*'\boldsymbol{\beta}_1$ is expected to be positive. If this happens, then from the expression for the non-centrality parameter in equation (3.6) we can conclude that the test that rejects H_0 in hypothesis (3.2) for large values of T_2 attains its maximum size when $\theta_1 = c_1$. From equation (3.6), it follows that, when $\theta_1 = c_1$, T_2 follows a central t-distribution with $n-p-1$ degrees of freedom. Thus, for a given level α, the test that rejects the null hypothesis (3.2) whenever $T_2 > t_{n-p-1,1-\alpha}$ is expected to have size at most α. Here $t_{n-p-1,1-\alpha}$ is the $100(1-\alpha)$th percentile of a central t-distribution with $n-p-1$ degrees of freedom.

These arguments show that the test based on T_2 can be easily carried out by using an exact central t-distribution. However, this is not so for the test based on T_1, since the distribution of T_1 depends on $\boldsymbol{\beta}_1$. Thus we shall use the PB to carry out the test based on T_1, just as we did for the confidence interval. Note that, if we intend to employ the PB, $\hat{\theta}_1$ in equation (2.10) can also be used as a test statistic, with large values of $\hat{\theta}_1$ providing the rejection region. The PB tests based on T_1 and $\hat{\theta}_1$ can be carried out as follows.

3.1.1. Parametric bootstrap test based on the test statistic T_1

The PB testing procedure can be described as follows.

(a) Using the $\mathbf{y}_i$s and $\mathbf{y}_0$ in models (1.1) and (1.2), compute the test statistic

$$T_1 = \frac{(n-p-1)^{1/2}\mathbf{u}' S^{-1}(\mathbf{v}_0 - \mathbf{u} c_1)}{\{(1+r)\mathbf{u}' S^{-1}\mathbf{u}\}^{1/2}},$$

where the quantities S, $\mathbf{u}$, $\mathbf{v}_0$, r and c_1 are given respectively in equations (2.1), (2.5), (2.8) and (3.1). Denote the computed value by T_{10}.

(b) For $i = 1, 100000$, generate $\mathbf{u}_i \sim N_p(\mathbf{u}, I_p), \mathbf{v}_{0i} \sim N_p(\mathbf{u}\hat{\theta}_1, I_p)$ and $S_i \sim W_p(n-2, I_p)$. Compute the test statistic by using the formula for T_1 in step (a), by replacing S, $\mathbf{u}$ and $\mathbf{v}_0$ by S_i, $\mathbf{u}_i$ and $\mathbf{v}_{0i}$ respectively, and call it T_{1i}.

(c) For a given level α, if the proportion of T_{1i}s that are greater than T_{10} is less than α, reject the null hypothesis (3.2). Here T_{10} is the computed value of T_1 in step (a).

The PB test based on $\hat{\theta}_1$ can be carried out similarly.

3.2. Sizes and powers of the tests based on T_1

For testing hypotheses (3.2), the size and power of the PB test based on T_1 can be evaluated as follows.

(a) For $i = 1, 1000$, generate $\mathbf{u}_i \sim N_p(\boldsymbol{\beta}_1, I_p), \mathbf{v}_{0i} \sim N_p(\boldsymbol{\beta}_1\theta_1, I_p)$ and $S_i \sim W_p(n-2, I_p)$. Compute the value of T_1 by using the formula given above (see step (a) in Section 3.1.1), by replacing S, $\mathbf{u}$ and $\mathbf{v}_0$ by S_i, $\mathbf{u}_i$ and $\mathbf{v}_{0i}$ respectively. Denote the computed value by T_{1i}.

(b) For $j = 1, 1000$, generate $\mathbf{u}_{ij} \sim N_p(\mathbf{u}_i, I_p), \mathbf{v}_{0ij} \sim N_p(\mathbf{u}_i c_1, I_p)$ and $S_j \sim W_p(n-2, I_p)$. Compute the value of T_1 again using $\mathbf{u}_{ij}$, $\mathbf{v}_{0ij}$ and S_j in place of $\mathbf{u}$, $\mathbf{v}_0$ and S respectively, and denote the computed value by T_{1ij}. Note that T_{1ij} also depends on i. (End the j-loop.)

(c) Compute the proportion of T_{1ij}s greater than T_{1i}. If this proportion is less than the specified α, set $P_i = 1$. (End the i-loop.)

(d) If $\theta_1 = c_1$, then $\Sigma_{i=1}^{1000} P_i/1000$ is the estimated size of the PB test based on T_1. If $\theta_1 > c_1$, then $\Sigma_{i=1}^{1000} P_i/1000$ is the estimated power of the PB test based on T_1.

The size and power of the PB test based on $\hat{\theta}_1$ can be estimated similarly. Table 2 gives the estimated sizes and powers of the PB tests based on T_1 and $\hat{\theta}_1$ for testing hypotheses (3.2). Table 3 gives the same when the hypotheses of interest are

$$H_0 : \theta_1 \geqslant c_1 \textit{ versus } H_1 : \theta_1 < c_1. \tag{3.7}$$

From the numerical results in Table 2 and Table 3, we see that the PB test based on T_1 is satisfactory for $|\theta_1| < 1.5$, whereas the PB test based on $\hat{\theta}_1$ is satisfactory only for small values of $|\theta_1|$. In particular, the sizes of the PB test based on $\hat{\theta}_1$ is smaller than the specified level for $\theta_1 > 0.5$, and larger than the specified level for $\theta_1 < -0.5$. Our limited simulations show that $\hat{\theta}_1$ should not be used for testing hypotheses (3.2) and (3.7). This is in agreement with what is known about the percentile bootstrap (i.e. based on the bootstrap distribution of $\hat{\theta}_1$). That is, the use of a pivot statistic should be preferred to the use of $\hat{\theta}_1$ for constructing confidence intervals and tests; see Hall and Wilson (1991) and Hall (1992).

Other numerical studies for some other values of α, $p = 4$ and $\boldsymbol{\beta}_1$ yielded similar conclusions; see Benton (2000) for some additional tables. We note from these numerical studies that the power of the PB tests increases as

$$\|\boldsymbol{\beta}_1\| = \sqrt{\sum_{i=1}^{p} \hat{\beta}_{1i}^2}$$

increases.

3.3. Power comparison of the tests based on T_1 and T_2

Table 4 gives the simulated powers (100000 runs) of the exact test based on T_2 in equation (3.5) and the PB test based on T_1, for testing hypotheses (3.2) and (3.7). The power of the PB

Table 2. Estimated powers of the PB tests based on T_1 and $\hat{\theta}_1$ for testing hypotheses (3.2) for $\alpha = 0.05, p = 3$ and $\beta_1' = (2, 1, 5)$

c_1	θ_1	*Powers for the following values of n:*					
		n = 10		*n = 20*		*n = 30*	
		T_1	$\hat{\theta}_1$	T_1	$\hat{\theta}_1$	T_1	$\hat{\theta}_1$
1.5	1.5	0.08	0.04	0.05	0.02	0.06	0.03
	2.0	0.26	0.16	0.24	0.16	0.27	0.17
	2.5	0.48	0.43	0.52	0.46	0.54	0.45
	3.0	0.72	0.74	0.75	0.75	0.80	0.76
1.0	1.0	0.07	0.03	0.07	0.04	0.06	0.04
	1.5	0.30	0.22	0.31	0.23	0.32	0.24
	2.0	0.66	0.63	0.65	0.61	0.72	0.67
	2.3	0.79	0.81	0.81	0.83	0.86	0.84
0.5	0.5	0.06	0.04	0.06	0.04	0.05	0.04
	1.0	0.40	0.34	0.42	0.32	0.43	0.36
	1.5	0.80	0.83	0.84	0.86	0.88	0.86
	2.0	0.93	0.96	0.96	0.97	0.97	0.98
0.0	0.0	0.05	0.05	0.05	0.05	0.04	0.05
	0.5	0.44	0.48	0.48	0.50	0.52	0.51
	1.0	0.86	0.90	0.90	0.94	0.93	0.95
	2.0	0.97	0.99	0.99	0.99	0.99	0.99
−0.5	−0.5	0.06	0.06	0.06	0.08	0.05	0.07
	0.0	0.35	0.48	0.43	0.54	0.43	0.53
	0.5	0.75	0.88	0.84	0.92	0.88	0.93
	1.0	0.92	0.98	0.98	0.99	0.99	1.00
−1.0	−1.0	0.06	0.10	0.06	0.12	0.05	0.10
	−0.5	0.29	0.41	0.32	0.47	0.35	0.50
	0.0	0.62	0.81	0.75	0.88	0.76	0.87
	0.5	0.87	0.97	0.95	0.99	0.95	0.99
−1.5	−1.5	0.05	0.13	0.06	0.12	0.05	0.10
	−1.0	0.26	0.41	0.28	0.41	0.29	0.45
	−0.5	0.56	0.74	0.60	0.75	0.61	0.80
	0.0	0.83	0.94	0.88	0.95	0.89	0.96

test based on T_1 was evaluated following the procedure outlined in Section 3.2. Regarding the power of the exact test based on T_2 in equation (3.5), note that $\mathbf{u}_*$ and $\mathbf{v}_{0*}$ in equations (3.4) are independent with $\mathbf{u}_* \sim N_p\{(1+\theta_1 c_1)\boldsymbol{\beta}_1, (1+c_1^2)I_p\}$ and $\mathbf{v}_{0*} \sim N_p\{\boldsymbol{\beta}_1(\theta_1 - c_1), I_p\}$. It is easy to see that the expected value of the numerator of T_2 is proportional to

$$(\theta_1 - c_1) + \theta_1 c_1(\theta_1 - c_1) = (\theta_1 - c_1)(1 + \theta_1 c_1).$$

Therefore, if $\theta_1 c_1 < -1$, then the test may not be able to detect the difference even if θ_1 is much larger than c_1. For example, for the hypotheses $H_0 : \theta_1 \leqslant -0.2$ *versus* $H_1 : \theta_1 > -0.2$, the power of the test based on T_2 declines for $\theta_1 > 1.5$ (see Table 4). A similar observation also holds for the hypotheses $H_0 : \theta_1 \geqslant 1$ *versus* $H_1 : \theta_1 < 1$. As a consequence, the test may incorrectly accept the null hypothesis. Further, the confidence interval based on T_2 is likely to contain false values for θ_1, a fact already observed by Mathew and Kasala (1994) in their numerical studies. Nevertheless, it is a powerful test if both θ_1 and c_1 are positive and $H_1 : \theta_1 > c_1$, or if they are both negative and $H_1 : \theta_1 < c_1$. Otherwise, the test based on T_1 should be preferred.

Table 3. Estimated powers of the PB tests based on T_1 and $\hat{\theta}_1$ for testing hypotheses (3.7) for $\alpha = 0.05, p = 3$ and $\beta_1' = (2, 1, 5)$

c_1	θ_1	Powers for the following values of n:					
		n = 10		n = 20		n = 30	
		T_1	$\hat{\theta}_1$	T_1	$\hat{\theta}_1$	T_1	$\hat{\theta}_1$
1.5	1.5	0.06	0.11	0.05	0.11	0.07	0.13
	1.0	0.22	0.34	0.26	0.41	0.27	0.42
	0.5	0.51	0.72	0.61	0.79	0.61	0.79
	0.0	0.79	0.92	0.88	0.96	0.87	0.95
1.0	1.0	0.05	0.09	0.05	0.10	0.06	0.11
	0.5	0.29	0.41	0.32	0.47	0.36	0.53
	0.0	0.67	0.84	0.73	0.87	0.77	0.88
	−0.5	0.87	0.97	0.94	0.99	0.97	0.99
0.5	0.5	0.07	0.09	0.05	0.08	0.06	0.08
	0.0	0.32	0.45	0.44	0.56	0.44	0.55
	−0.5	0.72	0.88	0.86	0.93	0.90	0.95
	−1.0	0.89	0.99	0.98	0.99	0.98	0.99
0.0	0.0	0.04	0.05	0.06	0.06	0.05	0.06
	−0.5	0.37	0.42	0.48	0.50	0.52	0.52
	−1.0	0.78	0.88	0.90	0.93	0.92	0.94
	−1.5	0.90	0.98	0.98	0.99	0.99	0.99
−0.5	−0.5	0.06	0.03	0.06	0.04	0.06	0.03
	−1.0	0.34	0.30	0.38	0.32	0.44	0.37
	−1.5	0.70	0.77	0.81	0.81	0.87	0.85
	−2.0	0.84	0.94	0.95	0.98	0.98	0.99
−1.0	−1.0	0.07	0.05	0.06	0.03	0.07	0.04
	−1.5	0.27	0.22	0.34	0.25	0.35	0.24
	−2.0	0.57	0.56	0.67	0.63	0.73	0.67
	−2.5	0.74	0.83	0.89	0.91	0.93	0.94
−1.5	−1.5	0.06	0.04	0.07	0.04	0.07	0.04
	−1.0	0.21	0.15	0.26	0.17	0.27	0.19
	−0.5	0.43	0.38	0.53	0.46	0.57	0.48
	0.0	0.63	0.68	0.77	0.75	0.79	0.77

Table 4. Estimated powers of the PB tests based on T_1 and the exact test based on T_2 for $n = 15, \alpha = 0.05, p = 3$ and $\beta_1' = (2, 1, 5)$

H_0	H_1	θ_1	T_1	T_2	H_0	H_1	θ_1	T_1	T_2
$\theta_1 \leqslant 1.0$	$\theta_1 > 1.0$	1.0	0.04	0.05	$\theta_1 \geqslant -0.5$	$\theta_1 < -0.5$	−0.5	0.03	0.05
		1.5	0.30	0.30			−1.0	0.34	0.40
		2.0	0.65	0.72			−1.5	0.79	0.86
		2.5	0.86	0.95			−2.0	0.96	0.99
							−2.5	0.99	1.0
$\theta_1 \leqslant -0.2$	$\theta_1 > -0.2$	0.5	0.66	0.62	$\theta_1 \geqslant 1.0$	$\theta_1 < 1.0$	1.0	0.06	0.05
		1.0	0.93	0.87			0.5	0.31	0.28
		1.5	0.97	0.90			0	0.70	0.59
		2.0	0.99	0.88			−0.5	0.93	0.57
		3.0	1.0	0.74			−1.0	0.97	0.25

4. Some examples

We shall now illustrate the computation of the confidence regions and tests, developed in Sections 2 and 3, by using three data sets: the meat data, the paint finish data and the wheat data. For background information on these data sets and the corresponding inference problems, see Section 1.2.

4.1. The meat data

Recall that the hypotheses of interest in the meat data problem are given in expression (1.3), where θ represents the fat content in a sample of meat. Among the data available on 214 meat samples, 172 samples were used to fit model (1.1). This leaves 42 samples which can be used for illustrating the PB procedure for testing hypotheses. Note that for these 42 samples the value of θ is known. We have selected six of the 42 samples for testing purposes; these six samples have some values of fat content below 10 and some above.

On the basis of the 172 samples used to fit the models, the various quantities have values $s_{xx} = 27501.17, \bar{x} = 18.38, \bar{\mathbf{y}}' = (-0.0167, 0.0073, 0.0113), \hat{\boldsymbol{\beta}}' = (0.0379, -0.0479, -0.0407)$ and

$$S = \begin{pmatrix} 133.78134 & 61.48432 & 36.68384 \\ 61.48432 & 122.48032 & -60.26106 \\ 36.68384 & -60.26106 & 127.53256 \end{pmatrix}.$$

Hence $\hat{\boldsymbol{\beta}}_1' = \hat{\boldsymbol{\beta}}' \sqrt{s_{xx}} = (6.285, \ -7.943, \ -6.749)$. Table 5 gives the six values of $\mathbf{y}_0$, the true values of θ and the p-values for testing hypotheses (1.3). These p-values are based on the PB test using T_1 and are computed on the basis of 100000 iterations (see Section 3.1.1).

We note that, among the results reported in Table 5, the wrong decisions are reached in two cases (corresponding to the third and fourth $\mathbf{y}_0$-values), if we use a significance level of 0.05.

4.2. The paint finish data

To demonstrate how the proposed PB methods compare with other methods, we apply them to the paint data described in example 2 of Section 1.2. In this example, x_i represents viscosity coded -1, 0 and 1, and $\mathbf{y}_i$ comprises two measurements of optical properties of the surface covered with the paint, corresponding to viscosity x_i, $i = 1, \ldots, 27$. Among the 27 x_i-values, there were nine values corresponding to each of the coded viscosity values -1, 0 and 1.

The calibration data set consists of the 27 observations used by Brown and Sundberg (1987)

Table 5. Simulated p-values of the test based on T_1 for testing the hypotheses $H_0 : \theta \leqslant 10.0$ *versus* $H_1 : \theta > 10.0$

$\mathbf{y}_0'$	θ	*p-value*
(−1.183, −0.523, 1.042)	4.8	0.74
(−0.078, −0.036, 1.592)	8.6	0.77
(1.254, 0.070, 1.448)	16.0	0.10
(2.180, 2.827, −1.143)	17.0	0.06
(0.207, 0.135, 0.306)	19.4	0.01
(1.155, −1.041, 1.525)	24.8	0.00

and Mathew and Zha (1996). Computations based on the calibration data gave $s_{xx} = 18$, $\bar{x}=0$, $\hat{\alpha} = \bar{\mathbf{y}} = (1.7478, 37.9363)'$, $\hat{\beta}' = (-0.1278, \ -1.6922)$ and

$$S^{-1} = \begin{pmatrix} 6.86285 & 0.03052 \\ 0.03052 & 0.02299 \end{pmatrix}.$$

In Table 6, we provide the PB confidence intervals based on T_0 in equation (2.8) and based on $\hat{\theta}_1$, along with the other intervals given in Mathew and Zha (1996). The PB intervals are calculated by using the method outlined in Section 2. We first computed PB intervals for θ_1, which were then modified to obtain intervals for θ. The PB interval estimates are based on 100000 PB samples and are denoted by PB(T_0) and PB($\hat{\theta}$) in Table 6. Table 6 also gives the exact confidence region from Brown (1982), the conservative confidence region from Mathew and Zha (1996) and the likelihood-based confidence region from Brown and Sundberg (1987), denoted by 'likelihood' in Table 6.

It should be recalled that Mathew and Zha's (1996) conservative confidence interval is also based on the statistic T_0. For this example, their confidence intervals are in good agreement with the PB(T_0) intervals. This conclusion is quite similar to that in Jones and Rocke (1999), namely that, in the univariate case, bootstrapping an appropriate t-statistic produces results that are in agreement with the standard method. An advantage of the PB method is that it is simple to use in comparison with Mathew and Zha's method which is computationally quite involved. We also notice from Table 6 that the PB intervals based on the sampling distribution of $\hat{\theta}$ are close to Brown's intervals. It should also be noted that, even though the likelihood-based confidence region due to Brown and Sundberg (1987) appears to be the narrowest in Table 6, the coverage of this interval can be below 95%; see the discussion in Mathew and Zha (1996), page 719.

4.3. *The wheat data*

For the wheat data, the $\mathbf{y}_i$s are bivariate infra-red reflectance measurements on 21 samples of hard wheat and the x_is represent the percentage of water in the wheat samples. We use the first 15 observations to fit the model. The PB methods are demonstrated on the remaining observations with the exception of observation 20, for which the values of the variables of interest are identical with those of observation 19.

The relevant estimates and other summary statistics, based on the first 15 observations, are $s_{xx} = 3.4013$, $\bar{x} = 9.52$, $\bar{\mathbf{y}}' = (81.933, 231.500)$, $\hat{\beta}' = (-24.918, \ -23.601)$,

$$S = \begin{pmatrix} 120.910 & 22.075 \\ 22.075 & 45.072 \end{pmatrix}$$

Table 6. 95% confidence intervals for θ for the paint finish data

$\mathbf{y}_0$	θ	*Confidence intervals from the following methods:*				
		Brown (1982)	*Likelihood*	*Mathew and Zha (1996)*	*PB(T_0)*	*PB($\hat{\theta}$)*
(1.78, 38.73)	−1	(−1.65, 0.92)	(−1.27, 0.57)	(−1.41, 0.67)	(−1.35, 0.68)	(−1.59, 0.89)
(1.79, 39.83)	0	(−1.93, 0.57)	(−1.60, 0.27)	(−1.76, 0.37)	(−1.64, 0.37)	(−1.90, 0.57)
(1.52, 35.65)	1	(0.39, 3.12)	(0.71, 2.68)	(0.61, 2.97)	(0.60, 2.70)	(0.39, 3.03)
(1.94, 34.09)	Unknown	Empty set	(−1.34, 1.16)	(−1.39, 1.23)	(−1.38, 1.16)	(−1.32, 1.16)

and $\hat{\beta}'_1 = \hat{\beta}'\sqrt{s_{xx}} = (-45.955, -43.526)$. In Table 7, we provide the PB single-use confidence intervals and the p-values for testing hypotheses based on five values of $\mathbf{y}_0$, following model (1.2), and where θ is known. For illustration, a test of $H_0 : \theta \leqslant 9.25$ *versus* $H_1 : \theta > 9.25$ is conducted. The single-use confidence intervals in Table 7 were computed in the same way as for the PB(T_0) intervals in the previous example. The p-values in Table 7 correspond to the PB test based on T_1 and are computed on the basis of 100000 bootstrap samples.

We note that the test procedure results in correct decisions in four of the five cases if we use a 5% significance level. At the 10% level, the test results in correct decisions in all cases.

The three examples given above illustrate the applicability and usefulness of the PB procedure for obtaining confidence intervals and tests.

5. Multiple-use one-sided hypothesis testing

Recall that, in the multiple-use case, the calibration data set is used repeatedly to test a sequence of hypotheses, one at a time, corresponding to a sequence of unknown θ-values, after observing the corresponding value of $\mathbf{y}_0$ in model (1.2). Thus, consider a sequence of observations $\{\mathbf{y}_{0j}\}$, $j = 1, 2, 3, \ldots,$ following the model

$$\mathbf{y}_{0j} \sim N_p(\boldsymbol{\alpha} + \boldsymbol{\beta}\theta_j, \Sigma),$$

where the $\mathbf{y}_{0j}$s are also independent. Problems of confidence regions and hypothesis tests are considerably more difficult to solve in the multiple-use case, as opposed to the single-use situation. The difficulty arises from the fact that two types of uncertainty need to be taken into account in the multiple-use case—the uncertainty in the calibration data and the uncertainty in the sequence $\{\mathbf{y}_{0j}\}$. The problem of deriving multiple-use confidence regions has been attempted by several researchers: see Mee *et al.* (1991) and Mee and Eberhardt (1996) for the univariate case, and Mathew and Zha (1997) and Mathew *et al.* (1998) for the multivariate case. The multiple-use confidence regions in Mathew *et al.* (1998) can be used for testing two-sided hypotheses, i.e. hypotheses of the form

$$H_{0j} : \theta_{0j} = c_j \textit{ versus } H_{1j} : \theta_{0j} \neq c_j,$$

where the c_js are known constants. Here we shall only consider one-sided hypotheses, i.e.

$$H_{0j} : \theta_{0j} \leqslant c \textit{ versus } H_{1j} : \theta_{0j} > c, \tag{5.1}$$

where c is known. Even in the context of test (5.1), a more general problem is one where different constants c_j are used instead of a common constant c. Recall that, for the meat example in Section 1.2, it is indeed meaningful and practically useful to test the sequence of hypotheses in

Table 7. 95% confidence intervals for θ, and p-values for testing $H_0 : \theta \leqslant 9.25$ *versus* $H_1 : \theta > 9.25$, for the wheat data

$\mathbf{y}'_0$	θ	$\hat{\theta}$	*Confidence interval*	*p-value*
(101.0, 248.0)	8.86	8.81	(8.61, 9.00)	0.99
(85.0, 235.0)	9.34	9.38	(9.19, 9.56)	0.08
(88.0, 231.0)	9.46	9.48	(9.27, 9.69)	0.01
(69.0, 222.0)	10.00	9.95	(9.76, 10.14)	0.00
(63.0, 216.0)	10.12	10.20	(10.00, 10.40)	0.00

expression (5.1), with $c = 10\%$; see model (1.3). Motivated by the definition of Z_0 in equation (2.3) and T_0 and T_1 in equations (2.8) and (3.3) respectively, let

$$T_{3j} = \frac{(n-2)^{1/2}\hat{\beta}' S^{-1}(\mathbf{y}_{0j} - \hat{\alpha} - \hat{\beta}c)}{(\hat{\beta}' S^{-1}\hat{\beta})^{1/2}}, \tag{5.2}$$

where $\hat{\alpha}$ and $\hat{\beta}$ denote the least squares estimators of α and β based on model (1.1), and S is defined in equations (2.1). Suppose that we reject H_{0j} when

$$T_{3j} > k(c),$$

where $k(c)$ is to be determined.

5.1. *The criterion for multiple-use hypothesis testing*

We shall now describe the condition to be used for the computation of $k(c)$. Let

$$\Delta_j = \begin{cases} 1, & \text{if } T_{3j} > k(c), \\ 0, & \text{otherwise.} \end{cases} \tag{5.3}$$

If a sequence of N tests is carried out for $j = 1, 2, \ldots, N$, then the proportion of H_{0j}s that are rejected is

$$\frac{1}{N}\sum_{j=1}^{N} \Delta_j.$$

The quantity $k(c)$ is to be determined so that, for all values of θ_j under H_{0j} ($j = 1, 2, \ldots, N$), this proportion is small with a high probability. Let

$$p_j = P(T_{3j} > k(c) | \hat{\alpha}, \hat{\beta}, S). \tag{5.4}$$

Note that, conditionally given $\hat{\alpha}$, $\hat{\beta}$ and S, the Δ_js in equation (5.3) are independent Bernoulli random variables with success probability p_j given in equation (5.4). By the strong law of large numbers, we obtain

$$\frac{1}{N}\sum_{j=1}^{N} \Delta_j \to \frac{1}{N}\sum_{j=1}^{N} p_j$$

for sufficiently large N. Therefore, we need to determine $k(c)$ subject to the condition

$$P_{\hat{\alpha},\hat{\beta},S}\left\{ \max_{H_{01},\ldots,H_{0N}} \left(\frac{1}{N}\sum_{j=1}^{N} p_j \right) \leqslant \alpha \right\} = \gamma, \tag{5.5}$$

where α is chosen to be small and γ is chosen to be large (e.g. $\alpha = 0.05$ and $\gamma = 0.95$).

We shall further simplify equation (5.5). For this, note that, conditionally given $\hat{\alpha}$, $\hat{\beta}$ and S,

$$\frac{T_{3j}}{\surd(n-2)} \sim N\left\{ \frac{\hat{\beta}' S^{-1}(\alpha + \beta\theta_j - \hat{\alpha} - \hat{\beta}c)}{(\hat{\beta}' S^{-1}\hat{\beta})^{1/2}}, \frac{\hat{\beta}' S^{-2}\hat{\beta}}{\hat{\beta}' S^{-1}\hat{\beta}} \right\},$$

where we have used the assumption $\Sigma = I_p$. Hence p_j in equation (5.4) can be expressed as

$$p_j = P\left\{ Z > \left(\frac{\hat{\beta}' S^{-1}\hat{\beta}}{\hat{\beta}' S^{-2}\hat{\beta}} \right)^{1/2} \frac{k(c)}{\surd(n-2)} - \frac{\hat{\beta}' S^{-1}(\alpha + \beta\theta_j - \hat{\alpha} - \hat{\beta}c)}{(\hat{\beta}' S^{-2}\hat{\beta})^{1/2}} \middle| \hat{\alpha}, \hat{\beta}, S \right\}, \tag{5.6}$$

where $Z \sim N(0,1)$. From equation (5.6), it follows that the maximum value of

$$\frac{1}{N}\sum_{j=1}^{N} p_j$$

under H_{0j} (i.e. under the condition $\theta_j \leqslant c$, $j = 1, 2, \ldots, N$) is attained at $\theta_j = c$, provided that $\hat{\beta}' S^{-1}\beta > 0$. It can be argued that $\hat{\beta}' S^{-1}\beta$ is very often likely to be positive; see Appendix C. Assuming this to be so, we immediately obtain

$$\max_{H_{01},\ldots,H_{0N}}\left(\frac{1}{N}\sum_{j=1}^{N} p_j\right) = P\left\{Z > \left(\frac{\hat{\beta}' S^{-1}\hat{\beta}}{\hat{\beta}' S^{-2}\hat{\beta}}\right)^{1/2}\frac{k(c)}{\sqrt{(n-2)}} - \frac{\hat{\beta}' S^{-1}(\alpha+\beta c-\hat{\alpha}-\hat{\beta}c)}{(\hat{\beta}' S^{-2}\hat{\beta})^{1/2}}\,\middle|\,\hat{\alpha},\hat{\beta},S\right\}.$$

Thus, assuming that $\hat{\beta}' S^{-1}\beta > 0$, condition (5.5) can be written as

$$P\left\{\left(\frac{\mathbf{u}' S^{-1}\mathbf{u}}{\mathbf{u}' S^{-2}\mathbf{u}}\right)^{1/2}\frac{k_1(c_*)}{\sqrt{(n-2)}} - \frac{\mathbf{u}' S^{-1}(\mathbf{h}_1-\mathbf{h}_2 c_*)}{(\mathbf{u}' S^{-2}\mathbf{u})^{1/2}} \geqslant z_{1-\alpha}\right\} = \gamma, \tag{5.7}$$

where $z_{1-\alpha}$ is the $100(1-\alpha)$th percentile of the standard normal distribution and

$$\beta_1 = \beta\sqrt{s_{xx}},$$
$$\mathbf{u} = \hat{\beta}\sqrt{s_{xx}},$$
$$\mathbf{h}_1 = \alpha + \beta\bar{x} - \bar{\mathbf{y}} \sim N\left(0, \frac{1}{n}I_p\right),$$
$$\mathbf{h}_2 = \mathbf{u} - \beta_1 \sim N(0, I_p),$$
$$c_* = (c-\bar{x})/\sqrt{s_{xx}}.$$

Also, we have used the notation $k_1(c_*)$ instead of $k(c)$. Thus, it follows from equation (5.7) that $k_1(c_*)$ is to be determined such that

$$P\left\{\frac{(n-2)^{1/2}\{z_{1-\alpha}(\mathbf{u}' S^{-2}\mathbf{u})^{1/2} + \mathbf{u}' S^{-1}(\mathbf{h}_1-\mathbf{h}_2 c_*)\}}{(\mathbf{u}' S^{-1}\mathbf{u})^{1/2}} \leqslant k_1(c_*)\right\} = \gamma. \tag{5.8}$$

Since it is difficult to obtain $k_1(c_*)$ analytically, we have used the PB for computing $k_1(c_*)$.

5.2. Parametric bootstrap method for computing $k_1(c_*)$

The computational procedure for the PB is as follows.

(a) Compute $\hat{\beta}$, $\bar{x}$ and s_{xx} on the basis of the calibration data. Set $c_* = (c-\bar{x})/\sqrt{s_{xx}}$ and $\mathbf{u} = \hat{\beta}\sqrt{s_{xx}}$.

(b) For $i = 1, 100000$, generate $\mathbf{u}_i \sim N_p(\mathbf{u}, I_p)$, $\mathbf{h}_{1i} \sim N(0, I_p/n)$ and $S_i \sim W_p(n-2, I_p)$. Set $\mathbf{h}_{2i} = \mathbf{u}_i - \hat{\beta}_1$ and compute

$$X_i = \frac{(n-2)^{1/2}\{z_{1-\alpha}(\mathbf{u}_i' S_i^{-2}\mathbf{u}_i)^{1/2} + \mathbf{u}_i' S_i^{-1}(\mathbf{h}_{1i}-\mathbf{h}_{2i} c_*)\}}{(\mathbf{u}_i' S_i^{-1}\mathbf{u}_i)^{1/2}}. \tag{5.9}$$

Table 8. PB estimates of $k_1(c_*)$ satisfying condition (5.8) for $\alpha = 0.05$, $\gamma = 0.95$, $p = 3$ and $\hat{\beta}_1' = (2, -3, 5)$

c_*	Estimates of $k_1(c_*)$ for the following values of n:			
	$n = 15$	$n = 20$	$n = 25$	$n = 30$
−2.0	7.44	6.79	6.48	6.26
−1.5	6.28	5.65	5.39	5.22
−1.0	5.05	4.55	4.29	4.15
−0.5	3.97	3.50	3.27	3.14
0.0	3.28	2.82	2.58	2.44
0.5	3.49	3.07	2.86	2.76
1.0	4.15	3.70	3.51	3.38
1.5	4.90	4.43	4.20	4.06
2.0	5.70	5.18	4.92	4.76

Table 9. PB estimates of the left-hand side of equation (5.8) for $\alpha = 0.05, \gamma = 0.95, p = 3$ and $\beta_1' = (2, -3, 5)$

c_*	Estimates for the following values of n:			
	$n = 15$	$n = 20$	$n = 25$	$n = 30$
−2.0	0.94	0.95	0.94	0.95
−1.5	0.94	0.95	0.96	0.94
−1.0	0.95	0.96	0.95	0.95
−0.5	0.95	0.95	0.95	0.95
0.0	0.95	0.95	0.94	0.95
0.5	0.94	0.95	0.95	0.94
1.0	0.95	0.94	0.96	0.96
1.5	0.94	0.95	0.95	0.95
2.0	0.95	0.94	0.96	0.95

(c) The 100γth percentile of the X_is in step (b) is a PB estimate of $k_1(c_*)$.

For illustration, we provide PB estimates (based on 100000 runs) of $k_1(c_*)$ when $\hat{\beta}_1' = (2, -3, 5)$ and for various values of c_* and n in Table 8.

To evaluate the performance of the PB estimates of $k_1(c_*)$, we conducted a numerical study to verify the probability requirement in equation (5.8). The numerical study was carried out in the following manner. We assumed that β_1 is known and the calibration data are standardized so that $\bar{x} = 0$ and $\bar{y} = 0$. As a result $\hat{\alpha} = 0$. Furthermore, we set $\beta_1' = (2, -3, 5)$, $\alpha = 0.05$ and $\gamma = 0.95$. We then estimated the left-hand side of equation (5.8) for several values of n and c_*, and the results are presented in Table 9. We observe from Table 9 that the estimates are in good agreement with the specified value of γ, and hence we conclude that the PB estimates of $k_1(c_*)$ are very satisfactory for practical use.

Table 10. Values of the multiple-use test statistic T_{3j}, for testing $H_{0j} : \theta_j \leqslant 9.25$ *versus* $H_{1j} : \theta_j > 9.25$, for the wheat data†

$\mathbf{y}_0$	θ	T_{3j}
(101.0, 248.0)	8.86	−5.98
(85.0, 235.0)	9.34	1.71
(88.0, 231.0)	9.46	3.15
(69.0, 222.0)	10.00	9.41
(63.0, 216.0)	10.12	12.79

†$\alpha = 0.05$ and $\gamma = 0.95$.

5.3. *The wheat data (continued)*

In Section 4.3, we analysed the wheat data and considered single-use hypothesis tests for testing the hypotheses $H_0 : \theta \leqslant 9.25$ *versus* $H_1 : \theta > 9.25$. Now suppose that the calibration data (consisting of 15 samples) will be used to test a sequence of such hypotheses, i.e. we wish to test $H_{0j} : \theta_j \leqslant 9.25$ *versus* $H_{1j} : \theta_j > 9.25$, $j = 1, 2, 3, \ldots$, where the θ_js represent the percentage of water in a sequence of hard wheat samples. For the wheat data, the values of $\bar{x}$, $\bar{\mathbf{y}}$, s_{xx}, $\hat{\boldsymbol{\beta}}$ and S are given in Section 4.3, and $c_* = (c - \bar{x})/\sqrt{s_{xx}} = (9.25 - 9.52)/\sqrt{3.4013} = -0.1464$. Table 10 gives the values of T_{3j} (computed by using equation (5.2)) for the five $\mathbf{y}_0$-values in Table 7.

We computed the value of $k_1(c_*)$ following the PB procedure given in Section 5.2, and the value turned out to be $k_1(c_*) = 2.95$ when $\alpha = 0.05$ and $\gamma = 0.95$. In Table 10, if a value of T_{3j} exceeds $k_1(c_*) = 2.95$, then $H_{0j} : \theta_j \leqslant 9.25$ is rejected for the corresponding sample. Note that the correct decision is reached in all cases except for the second observation in Table 10. If we use $\alpha = 0.10$ and $\gamma = 0.95$, the value of $k_1(c_*)$ is 2.64. Note that at this level an incorrect decision is also made for the second observation. In comparison, the single-use test (see Section 4.3, Table 7) results in correct decisions in all the cases at the 10% level. This is not surprising because a multiple-use hypothesis test has a wider acceptance region than its corresponding single-use test (see the description in Section 1.3).

Remark 2. Mathew *et al.* (1998) have derived a conservative multiple-use confidence region in the multivariate calibration problem, based on a pivot that is equivalent to T_{3j} in equation (5.2). Though not reported here, it is possible to apply the PB procedure and to derive multiple-use confidence regions.

6. Concluding remarks

The work of Brown (1982) is one of the earliest papers dealing with the multivariate calibration problem. It has brought out the significance of multivariate calibration in applied work, and it generated considerable interest among researchers, as evidenced by the subsequent published work on the topic; see Brown (1993) for a review. The present paper deals with the special case of multivariate–univariate calibration problems, and we have investigated the applicability of the PB for deriving confidence regions and hypothesis tests. Numerical results suggest that the PB is an attractive option for making inferences in such calibration problems. We have extensively investigated various aspects of both interval estimation and one-sided hypothesis testing. Even though some procedures are available for constructing confidence regions in the calibration problem, and hence for carrying out two-sided hypothesis tests, it is not clear how to handle the one-sided hypothesis testing problem if we want to avoid using the bootstrap. Yet, the one-sided testing problem is relevant in many practical applications, as should be clear from the several examples in this paper.

For obtaining confidence regions, the PB is easily adapted to the general multivariate calibration problem, where the parameter of interest is a vector. Indeed, Mathew and Zha (1996, 1997, 1998) and Mathew *et al.* (1998) have studied a general multivariate calibration problem in the context of a multivariate linear model. Parametric bootstrapping can be used to obtain single-use and multiple-use confidence regions based on the pivot statistics in these references. The basic ideas and the necessary computations are similar to what we have done in this paper. As noted by Jones and Rocke (1999), the bootstrap procedure is applicable to non-linear calibration problems as well, both univariate and multivariate. Our overall conclusion is that, under the models assumed, parametric bootstrapping is an excellent option for drawing inferences in the calibration problem.

Acknowledgement

The authors are grateful to two reviewers for providing valuable comments and suggestions.

Appendix A: Proof of theorem 1

The proof of theorem 1 is based on technical arguments that are similar to those in Srivastava and Khatri (1979), section 4.4. Similar techniques have also been used in Fujikoshi and Nishii (1984), Davis and Hayakawa (1987), Mathew and Kasala (1994) and Mathew and Zha (1996).

Recall that we are assuming $\Sigma = I_p$, and $\mathbf{u}$ and $\mathbf{v}_0$ have the distributions given in expression (2.6). Let

$$\Gamma = \left(\frac{\mathbf{u}}{\|\mathbf{u}\|}, \Gamma_1\right)$$

be a $p \times p$ orthogonal matrix so that $\mathbf{u}'\Gamma = \|\mathbf{u}\|(1, \mathbf{0})$. Here and elsewhere $\|\cdot\|$ denotes the Euclidean norm. Write

$$\Gamma' S \Gamma = W = \begin{pmatrix} w_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & W_{22} \end{pmatrix} \sim W_p(n-2, I_p),$$

where $w_{11} = \mathbf{u}'S^{-1}\mathbf{u}/\|\mathbf{u}\|^2$ is a scalar. Since $\mathbf{u}$ and $\mathbf{v}_0$ are independent, we see that conditionally given Γ (i.e. conditionally given $\mathbf{u}$)

$$\mathbf{v} = \Gamma'\mathbf{v}_0 = \begin{pmatrix} v_1 \\ \mathbf{v}_2 \end{pmatrix} \sim N_p(\boldsymbol{\psi}, I_p) = N_p\left\{\begin{pmatrix} \psi_1 \\ \boldsymbol{\psi}_2 \end{pmatrix}, I_p\right\}, \tag{A.1}$$

where $v_1 = \mathbf{u}'\mathbf{v}_0/\|\mathbf{u}\|$ and $\psi_1 = \mathbf{u}'\boldsymbol{\beta}_1\theta_1/\|\mathbf{u}\|$. For computing ψ_1, we have used the fact that $E(\mathbf{v}_0) = \boldsymbol{\beta}_1\theta_1$; see expression (2.6). Let Q be as defined in equation (2.7). Using the orthogonal transformation Γ, Q can be written as

$$Q = \frac{\mathbf{u}'\Gamma(\Gamma'S\Gamma)^{-1}(\Gamma'\mathbf{v}_0 - \Gamma'\mathbf{u}\theta_1)}{\{\mathbf{u}'\Gamma(\Gamma'S\Gamma)^{-1}\Gamma'\mathbf{u}\}^{1/2}} = \frac{\|\mathbf{u}\|(1,\mathbf{0})W^{-1}\left\{\mathbf{v} - \|\mathbf{u}\|\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix}\theta_1\right\}}{\left\{\|\mathbf{u}\|^2(1,\mathbf{0})W^{-1}\begin{pmatrix}1\\ \mathbf{0}\end{pmatrix}\right\}^{1/2}}.$$

Using the inverse of a partitioned matrix (see Rao (1973), page 33), and simplifying, we obtain

$$Q = \frac{v_1 - \|\mathbf{u}\|\theta_1 - \mathbf{w}_{12}W_{22}^{-1}\mathbf{v}_2}{\sqrt{w_{11.2}}}, \tag{A.2}$$

where $w_{11.2} = w_{11} - \mathbf{w}_{12}W_{22}^{-1}\mathbf{w}_{21}$. Note that $W_{22}^{-1/2}\mathbf{w}_{21} \sim N_p(\mathbf{0}, I_{p-1})$, independent of W_{22} and $w_{11.2}$. Also, given $\mathbf{u}$,

$$v_1 - \|\mathbf{u}\|\theta_1 \sim N(\psi_1 - \|\mathbf{u}\|\theta_1, 1),$$

where ψ_1 is given above. Writing

$$\mathbf{w}_{12}W_{22}^{-1}\mathbf{v}_2 = \mathbf{w}_{12}W_{22}^{-1/2}W_{22}^{-1/2}\mathbf{v}_2,$$

we see that, conditionally given $W_{22}^{-1/2}\mathbf{v}_2$,

$$\mathbf{w}_{12}W_{22}^{-1}\mathbf{v}_2 \sim N(0, \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2).$$

Hence, conditionally given $\mathbf{u}$ and $\mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2$,

$$v_1 - \|\mathbf{u}\|\theta_1 - \mathbf{w}_{12}W_{22}^{-1}\mathbf{v}_2 \sim N(\psi_1 - \|\mathbf{u}\|\theta_1, 1 + \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2).$$

Consequently, conditionally given $\mathbf{u}$ and $\mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2$,

$$\frac{v_1 - \|\mathbf{u}\|\theta_1 - \mathbf{w}_{12}W_{22}^{-1}\mathbf{v}_2}{(1 + \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2)^{1/2}} \sim N\left\{\frac{\psi_1 - \|\mathbf{u}\|\theta_1}{(1 + \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2)^{1/2}}, 1\right\}. \tag{A.3}$$

Also, $w_{11.2}$ follows a central χ^2-distribution with $n - p - 1$ degrees of freedom. In view of expression (A.3), we modify Q in equation (A.2) and define

$$T_0 = \frac{(n - p - 1)^{1/2}(v_1 - \|\mathbf{u}\|\theta_1 - \mathbf{w}_{12}W_{22}^{-1}\mathbf{v}_2)}{\{w_{11.2}(1 + \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2)\}^{1/2}}. \tag{A.4}$$

It then follows that, conditionally given $\mathbf{u}$ and $\mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2$, T_0 follows a non-central t-distribution with $n-p-1$ degrees of freedom and non-centrality parameter δ_0, given by

$$\delta_0 = \frac{\psi_1 - \|\mathbf{u}\|\theta_1}{(1 + \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2)^{1/2}}. \tag{A.5}$$

The proof is complete once we show that T_0 in equation (A.4) coincides with the expression for T_0 in equations (2.8), and δ_0 in equation (A.5) is the same quantity as in equation (2.9). Towards this, note that

$$\begin{aligned}\psi_1 - \|\mathbf{u}\|\theta_1 &= (\mathbf{u}'\boldsymbol{\beta}_1\theta_1 - \mathbf{u}'\mathbf{u}\theta_1)/\|\mathbf{u}\|,\\ \mathbf{v}_2'W_{22}^{-1}\mathbf{v}_2 &= \mathbf{v}_0'\Gamma_1(\Gamma_1'S\Gamma_1)^{-1}\Gamma_1'\mathbf{v}_0\\ &= \mathbf{v}_0'S^{-1}\mathbf{v}_0 - \frac{(\mathbf{v}_0'S^{-1}\mathbf{u})^2}{\mathbf{u}'S^{-1}\mathbf{u}} = r,\end{aligned}$$

where r is the quantity given in equations (2.8). This completes the proof.

Appendix B: Proof of theorem 2

The proof of theorem 2 is very similar to that of theorem 1, and we shall give a very brief outline. Assume without loss of generality that $\Sigma = I_p$ and let $\mathbf{u}_*$ and $\mathbf{v}_{0*}$ be as defined in expression (3.4). Let

$$\Gamma_* = \left(\frac{\mathbf{u}_*}{\|\mathbf{u}_*\|}, \Gamma_{1*}\right)$$

be a $p \times p$ orthogonal matrix so that $\mathbf{u}_*'\Gamma_* = \|\mathbf{u}_*\|(1, \mathbf{0})$. Write

$$\Gamma_*'S\Gamma_* = W_* = \begin{pmatrix} w_{11*} & \mathbf{w}_{12*} \\ \mathbf{w}_{21*} & W_{22*}\end{pmatrix} \sim W_p(n-2, I_p).$$

Let

$$\mathbf{v}_* = \Gamma_*'\mathbf{v}_{0*} = \begin{pmatrix} v_{1*} \\ \mathbf{v}_{2*}\end{pmatrix},$$

where v_{1*} is a scalar. Following the arguments in Appendix A, it can be shown that T_2 in equations (3.5) can be expressed as

$$T_2 = \frac{(n-p-1)^{1/2}(v_{1*} - \mathbf{w}_{12*}W_{22*}^{-1}\mathbf{v}_{2*})}{\{w_{11.2*}(1+\mathbf{v}_{2*}'W_{22*}^{-1}\mathbf{v}_{2*})\}^{1/2}},$$

where $w_{11.2*} = w_{11*} - \mathbf{w}_{12*}W_{22*}^{-1}\mathbf{w}_{21*}$. As in the proof of theorem 1 in Appendix A, it can now be verified that, given $\mathbf{u}_*$ and $r_* = \mathbf{v}_{2*}'W_{22*}^{-1}\mathbf{v}_{2*}$, T_2 follows a non-central t-distribution with $n-p-1$ degrees of freedom and non-centrality parameter

$$\frac{\mathbf{u}_*'\beta_1(\theta_1 - c_1)}{\|\mathbf{u}_*\|\{(1+c_1^2)(1+r_*)\}^{1/2}}.$$

It can also be verified that r_* coincides with the quantity given in equations (3.5).

Appendix C: An approximate lower bound for $P(\hat{\beta}'S^{-1}\beta > 0)$

Once again, we assume that $\Sigma = I_p$. Note that, conditionally given S, $\hat{\beta}'S^{-1}\beta \sim N(\beta'S^{-1}\beta, \beta'S^{-2}\beta/s_{xx})$. Therefore,

$$\begin{aligned} P(\hat{\beta}'S^{-1}\beta > 0) &= E_S\{P(\hat{\beta}'S^{-1}\beta > 0|S)\} \\ &= E_S\left[P\left\{Z > -\sqrt{s_{xx}}\frac{\beta'S^{-1}\beta}{(\beta'S^{-2}\beta)^{1/2}}\middle|S\right\}\right] \\ &\approx P\left[Z > -\sqrt{s_{xx}}E\left\{\frac{\beta'S^{-1}\beta}{(\beta'S^{-2}\beta)^{1/2}}\right\}\right], \end{aligned} \tag{C.1}$$

where $Z \sim N(0, 1)$. To evaluate the expectation in expression (C.1), let Γ_2 be an orthogonal matrix such that $\Gamma_2'\beta = \|\beta\|(1, 0, \ldots, 0)'$. Then $\Gamma_2'S\Gamma_2 \sim W_p(n-2, I_p)$. Let

$$\Gamma_2'S\Gamma_2 = \begin{pmatrix} s_{11} & \mathbf{s}_{12} \\ \mathbf{s}_{21} & S_{22} \end{pmatrix}, \tag{C.2}$$

where s_{11} is a scalar. Since $\Gamma_2'S\Gamma_2 \sim W_p(n-2, I_p)$, we also have $S_{22} \sim W_{p-1}(n-2, I_{p-1})$. The expectation in expression (C.1) can be written as

$$\begin{aligned} E\left\{\frac{\beta'S^{-1}\beta}{(\beta'S^{-2}\beta)^{1/2}}\right\} &= E\left[\frac{\beta'\Gamma_2(\Gamma_2'S\Gamma_2)^{-1}\Gamma_2'\beta}{\{\beta'\Gamma_2(\Gamma_2'S\Gamma_2)^{-2}\Gamma_2'\beta\}^{1/2}}\right] \\ &= E\left\{\frac{\|\beta\|}{(1+\mathbf{s}_{12}S_{22}^{-2}\mathbf{s}_{21})^{1/2}}\right\} \\ &> \frac{\|\beta\|}{\{1+E(\mathbf{s}_{12}S_{22}^{-2}\mathbf{s}_{21})\}^{1/2}}. \end{aligned} \tag{C.3}$$

To arrive at inequality (C.3), we have used the fact that $\Gamma_2'\beta = \|\beta\|(1, 0, \ldots, 0)'$ and we have also used the expression for the inverse of the partitioned matrix in expression (C.2). Inequality (C.3) follows from the fact that $1/\sqrt{(1+x)}$ is a convex function. Now we use the following properties: $S_{22}^{-1/2}\mathbf{s}_{21} \sim N_{p-1}(0, I_{p-1})$ and is independent of S_{22}, and

$$E(S_{22}^{-1}) = \frac{1}{n-p-2}I_{p-1}$$

(see lemma 7.7.1 in Anderson (1984)). Using these results, we obtain

$$\begin{aligned} E(\mathbf{s}_{12}S_{22}^{-2}\mathbf{s}_{21}) &= \mathrm{tr}\{E(S_{22}^{-1}S_{22}^{-1/2}\mathbf{s}_{21}\mathbf{s}_{12}S_{22}^{-1/2})\} \\ &= \mathrm{tr}\{E(S_{22}^{-1})\,E(S_{22}^{-1/2}\mathbf{s}_{21}\mathbf{s}_{12}S_{22}^{-1/2})\} \\ &= \mathrm{tr}\left(\frac{1}{n-p-2}I_{p-1}\right) = \frac{p-1}{n-p-2}. \end{aligned} \tag{C.4}$$

From expressions (C.1), (C.3) and (C.4), we obtain

$$P(\hat{\beta}' S^{-1}\beta > 0) > P\left[Z > -\sqrt{s_{xx}}\frac{\|\beta\|}{\{1+(p-1)/(n-p-2)\}^{1/2}}\right] \tag{C.5}$$

approximately. For large n, $s_{xx} = \Sigma_{i=1}^{n}(x_i - \bar{x})^2$ tends to be large, and hence the right-hand side of inequality (C.5) will be large.

References

Anderson, T. W. (1984) *An Introduction to Multivariate Analysis*, 2nd edn. New York: Wiley.

Benton, D. (2000) Inferences in multivariate calibration. *PhD Thesis*. University of Louisiana, Lafayette.

Brown, P. J. (1982) Multivariate calibration (with discussion). *J. R. Statist. Soc.* B, **44**, 287–321.

Brown, P. J. (1993) *Measurement, Regression and Calibration*. New York: Oxford University Press.

Brown, P. J. and Sundberg, R. (1987) Confidence and conflict in multivariate calibration. *J. R. Statist. Soc.* B, **49**, 46–57.

Davis, A. and Hayakawa, T. (1987) Some distribution theory relating to confidence regions in multivariate calibration. *Ann. Inst. Statist. Math.*, **39**, 141–152.

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

Fox, D. R. (1989) Conditional multivariate calibration. *Communs Statist. Theory Meth.*, **18**, 2311–2330.

Fujikoshi, Y. and Nishii, R. (1984) On the distribution of a statistic in multivariate inverse regression analysis. *Hirosh. Math. J.*, **14**, 215–225.

Hall, P. (1992) *The Bootstrap and Edgeworth Expansion*. New York: Springer.

Hall, P. and Wilson, S. R. (1991) Two guidelines for bootstrap hypothesis testing. *Biometrics*, **47**, 757–762.

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

Krishnamoorthy, K., Kulkarni, P. M. and Mathew, T. (2001) Multiple use one-sided hypothesis testing in univariate linear calibration. *J. Statist. Planng Inf.*, **93**, 211–223.

Lee, S. M. N. (1994) Optimal choice between parametric and nonparametric bootstrap estimates. *Math. Proc. Camb. Philos. Soc.*, **115**, 335–363.

Martens, H. and Naes, T. (1989) *Multivariate Calibration*. New York: Wiley.

Mathew, T. and Kasala, S. (1994) An exact confidence region in multivariate calibration. *Ann. Statist.*, **22**, 94–105.

Mathew, T., Sharma, M. K. and Nordström, K. (1998) Tolerance regions and multiple use confidence regions in multivariate calibration. *Ann. Statist.*, **26**, 1989–2013.

Mathew, T. and Zha, W. (1996) Conservative confidence regions in multivariate calibration. *Ann. Statist.*, **24**, 707–725.

Mathew, T. and Zha, W. (1997) Multiple use confidence regions in multivariate calibration. *J. Am. Statist. Ass.*, **92**, 1141–1150.

Mathew, T. and Zha, W. (1998) Some single use confidence regions in a multivariate calibration problem. In *Statistical Inference and Related Topics, a Festschrift in Honor of A. K. Md. E. Saleh* (eds S. E. Ahmed, M. Ahsanullah and B. K. Sinha), pp. 351–363. New York: Nova Science.

Mee, R. W. and Eberhardt, K. R. (1996) A comparison of uncertainty criteria for calibration. *Technometrics*, **38**, 221–229.

Mee, R. W., Eberhardt, K. R. and Reeve, C. P (1991) Calibration and simultaneous tolerance intervals for regression. *Technometrics*, **33**, 211–219.

Oman, S. D. (1988) Confidence regions in multivariate calibration. *Ann. Statist.*, **16**, 174–187.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*. New York: Wiley.

Seasholtz, M. B. (1999) Making money with chemometrics. *Chemometr. Intell. Lab. Syst.*, **45**, 55–63.

da Silva, E. and Laquipai, M. (1998) Method of rapid screening of chlorophenols using a reduced calibration set of UV spectra and multivariate calibration techniques. *Analyt. Lett.*, **31**, 2549–2563.

Srivastava, M. S. and Khatri, C. G. (1979) *An Introduction to Multivariate Statistics*. New York: North-Holland.

Williams, E. J. (1959) *Regression Analysis*. New York: Wiley.

Wood, J. T. (1982) Estimating the age of an animal: an application of Multivariate Calibration. In *Proc. 11th Int. Biometric Conf.*