Available online at www.sciencedirect.com
ScienceDirect
Journal of Statistical Planning and Inference 137 (2007) 1226–1235
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Single use conservative confidence regions in multivariate controlled calibration

K.K. Jose[a,*], Jikcey Isaac[b]

[a]Department of Statistics, St. Thomas College, Pala, Kerala-686574, India
[b]Department of Statistics, Marian College, Kuttikkanam, Kerala-685531, India

Received 30 August 2005; accepted 14 February 2006
Available online 2 May 2006

Abstract

We consider a multivariate linear model for multivariate controlled calibration, and construct some conservative confidence regions, which are nonempty and invariant under nonsingular transformations. The computation of our confidence region is easier compared to some of the existing procedures. We illustrate the results using two examples. The simulation results show the closeness of the coverage probability of our confidence regions to the assumed confidence level.

MSC: Primary 62F25; secondary 62H99

Keywords: Multivariate calibration; Single use conservative confidence region; Linear models; Coverage probability; Noncentral F distribution

1. Introduction

Let $\mathbf{y}$ be a $p \times 1$ response variable and $\mathbf{x}$ be a $q \times 1$ explanatory variable. Here we assume $p \geqslant q$. The multivariate calibration problem deals with statistical inference concerning an unknown value of $\mathbf{x}$ corresponding to a future value of $\mathbf{y}$ by modelling an appropriate relationship between $\mathbf{y}$ and $\mathbf{x}$, using available data. In this paper we construct a confidence region for the unknown value of $\mathbf{x}$, when the explanatory variable $\mathbf{x}$ is assumed to be fixed (controlled). Here we are considering only a multivariate linear model relationship between $\mathbf{y}$ and $\mathbf{x}$, and assume that $\mathbf{y}$ follows a multivariate normal distribution. Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N$ denote independent realisations of $\mathbf{y}$, corresponding to the values of $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$ of $\mathbf{x}$. We assume $\mathbf{y}_i \sim N(A\mathbf{x}_i, \Sigma)$, where A is an unknown $p \times q$ parameter matrix and Σ is an unknown $p \times p$ positive definite matrix.

Let $Y = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N)$ and $X = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N)$, then

$$E(Y) = AX \quad \text{and} \quad \text{Cov (Vec }(Y)) = I_N \otimes \Sigma. \tag{1}$$

The $q \times N$ matrix X is assumed to be of rank q. Now consider another $p \times 1$ normally distributed random vector $\mathbf{y}$, corresponding to an unknown value θ of $\mathbf{x}$ and independent of Y in (1) and assume the same multivariate linear model

* Corresponding author. Tel.: +91 482 2201288; fax: +91 482 2216313.
E-mail addresses: kkjstc@rediffmail.com, statpala@yahoo.com (K.K. Jose).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2006.02.005

as in (1). Then we get

$$E(\mathbf{y}) = A\theta \quad \text{and} \quad \mathrm{Cov}(\mathbf{y}) = \Sigma. \tag{2}$$

The problem discussed here is the construction of a confidence region for θ. We shall also consider the situation where θ is a non-linear function of fewer unknown parameters, denoted by an $s \times 1$ vector η ($s \leqslant q$). Then we have the model

$$E(\mathbf{y}) = A\mathbf{t}(\eta) \quad \text{and} \quad \mathrm{Cov}(\mathbf{y}) = \Sigma, \tag{3}$$

where $\mathbf{t}(\eta)$ is a $q \times 1$ vector-valued function of η. Now the problem is the construction of a confidence region for η. Polynomial regression is an example of model (3). Oman (1988) considered model (3) and a specific application can be found in Oman and Wax (1984). Osborne (1991) gives an excellent review of statistical calibration.

The models considered here do not have an intercept term. However, models with the intercept term appears in many applications. Such models can be reduced to those without the intercept; see Remark 2.2 in Mathew and Kasala (1994). Several authors have considered the problem of constructing exact or conservative confidence regions for θ in model (2). We refer to Brown (1993), Sundberg (1994) and Mathew and Zha (1996) for a review of the literature.

In this paper we derive some conservative confidence regions modifying the ideas in Mathew and Zha (1996). First, we describe the procedure for models (1) and (2), and then we extend our procedure to model (3). Computational details are discussed in Sections 3 and 4 and two examples are considered in Sections 5 and 6. Both examples are also discussed in Mathew and Sharma (2002a,b) as well as Mathew et al. (1998). The first example given in Section 5, is based on the paint finish data analysed in Brown (1982), Brown and Sundberg (1987) and Mathew and Zha (1996). In this example our procedure gave confidence regions with coverage probability very close to the assumed confidence level.

Our second example given in Section 6, is based on the problem considered in Oman and Wax (1984) and Mathew and Zha (1996) dealing with the estimation of gestational age (i.e., week of pregnancy) using ultrasound measurements of two fetal bone lengths. Oman and Wax (1984) modelled a quadratic relationship between bone lengths and gestational age. The problem is to predict the unknown gestational age corresponding to a measurement of the bone lengths. For this problem, our results are very similar to those in Mathew and Zha (1996) with the additional property that our confidence region is very easy to compute. Some concluding remarks are noted in Section 7.

2. Canonical forms and confidence regions

We shall first discuss models (1) and (2) and use the following canonical forms derived in Mathew and Kasala (1994). Let

$$\begin{aligned} Y_1 &= YX'(XX')^{-1/2}, \quad S = Y(I - X'(XX')^{-1}X)Y', \\ A_1 &= A(XX')^{1/2}, \quad \theta_1 = (XX')^{-1/2}\theta. \end{aligned} \tag{4}$$

Then

$$Y_1 \sim N(A_1, I_q \otimes \Sigma), \quad \mathbf{y} \sim N(A_1\theta_1, \Sigma), \quad S \sim W_p(\Sigma, N - q), \tag{5}$$

where we assume $N - q \geqslant p$. Also, Y_1, $\mathbf{y}$ and S are independently distributed. We work with the canonical form (5). First we obtain a confidence region for θ_1, then we use the transformation in (4) to obtain a confidence region for θ.

Following Mathew and Zha (1996), the confidence region that we shall construct will be based on the statistic

$$(\mathbf{y} - Y_1\theta_1)'S^{-1}Y_1(Y_1'S^{-1}Y_1)^{-1}Y_1'S^{-1}(\mathbf{y} - Y_1\theta_1) = (\hat{\theta}_1 - \theta_1)'Y_1'S^{-1}Y_1(\hat{\theta}_1 - \theta_1), \tag{6}$$

where $\hat{\theta}_1 = (Y_1'S^{-1}Y_1)^{-1}Y_1'S^{-1}\mathbf{y}$.

Clearly it is reasonable to use (6) to obtain a confidence region for θ_1. The possibility of using (6) is mentioned in Williams (1959) and Wood (1982). Fujikoshi and Nishii (1984) and Davis and Hayakawa (1987) discussed asymptotic