TOLERANCE REGIONS AND MULTIPLE-USE CONFIDENCE REGIONS IN MULTIVARIATE CALIBRATION

By Thomas Mathew[1], Manoj Kumar Sharma
and Kenneth Nordström[2]

*University of Maryland Baltimore County, LifeScan, Inc.
and University of Helsinki*

Let $\mathbf{y}_i \sim N(B\mathbf{x}_i, \Sigma)$, $i = 1, 2, \ldots, N$, and $\mathbf{y} \sim N(B\boldsymbol{\theta}, \Sigma)$ be independent multivariate observations, where the $\mathbf{x}_i$'s are known vectors, B, $\boldsymbol{\theta}$ and Σ are unknown, Σ being a positive definite matrix. The calibration problem deals with statistical inference concerning $\boldsymbol{\theta}$ and the problem that we have addressed is the construction of confidence regions. In this article, we have constructed a region for $\boldsymbol{\theta}$ based on a criterion similar to that satisfied by a tolerance region. The situation where $\boldsymbol{\theta}$ is possibly a nonlinear function, say $\mathbf{h}(\boldsymbol{\xi})$, of fewer unknown parameters denoted by the vector $\boldsymbol{\xi}$, is also considered. The problem addressed in this context is the construction of a region for $\boldsymbol{\xi}$. The numerical computations required for the practical implementation of our region are explained in detail and illustrated using an example. Limited numerical results indicate that our regions satisfy the coverage probability requirements of multiple-use confidence regions.

1. Introduction. In this article, we address the problem of constructing confidence regions in the multivariate calibration problem, when the observations follow a general multivariate linear model and are normally distributed. Thus, let $\mathbf{y}_i$ $(i = 1, 2, \ldots, N)$ be independent $p \times 1$ random vectors with

$$(1.1) \qquad \mathbf{y}_i \sim N(\mathbf{a} + B\mathbf{x}_i, \Sigma),$$

where $\mathbf{x}_i$'s are known $m \times 1$ vectors, $\mathbf{a}$ is an unknown $p \times 1$ intercept vector, B is an unknown $p \times m$ matrix, and Σ is an unknown $p \times p$ positive definite matrix. The $\mathbf{x}_i$'s are known values of a controlled variable. The calibration problem deals with statistical inference concerning an unknown value, say $\boldsymbol{\theta}$, of the controlled variable corresponding to an observation vector $\mathbf{y}$, where $\mathbf{y}$ is assumed to be independent of the $\mathbf{y}_i$'s in (1.1) and has a normal distribution similar to (1.1), that is,

$$(1.2) \qquad \mathbf{y} \sim N(\mathbf{a} + B\boldsymbol{\theta}, \Sigma).$$

The problem that we shall address is the construction of a confidence region for the $m \times 1$ unknown vector $\boldsymbol{\theta}$, using the $\mathbf{y}_i$'s in (1.1) and $\mathbf{y}$ in (1.2). We shall

assume that $\text{rank}(\mathbf{1}_N, X) = m + 1$, where $\mathbf{1}_N$ denotes the $N \times 1$ vector of ones and X is the $m \times N$ matrix whose columns are the $\mathbf{x}_i$'s in (1.1). We also assume that $p \geq m$, so that $\boldsymbol{\theta}$ is identifiable in the model (1.2).

The function $\mathbf{a} + B\mathbf{x}$, where $\mathbf{x}$ denotes the explanatory variable, is usually referred to as the *calibration curve*. The observations $\mathbf{y}_i$ ($i = 1, 2, \ldots, N$) will be referred to as the *calibration data*. Depending on the practical application, the confidence region required for $\boldsymbol{\theta}$ is one of two types: single-use confidence region or multiple-use confidence region. If the $\mathbf{y}_i$'s in (1.1) are used to construct a confidence region for a single vector $\boldsymbol{\theta}$, corresponding to a single observation $\mathbf{y}$ in (1.2), we then have a single-use confidence region. If the same set of $\mathbf{y}_i$'s is used to construct confidence regions for a sequence of $\boldsymbol{\theta}$ values, corresponding to a sequence of observation vectors $\mathbf{y}$, we then have multiple-use confidence regions. Multiple-use confidence regions meet the following requirement, expressed in terms of two coverage probabilities, say $1 - \alpha$ and $1 - \beta$. The sequence of confidence regions is such that conditionally given the $\mathbf{y}_i$'s in (1.1), the proportion of confidence regions that will include the corresponding true $\boldsymbol{\theta}$ values is at least $1 - \beta$. The probability that $\mathbf{y}_i$'s following (1.1) will provide such a $100(1 - \beta)\%$ coverage is at least $1 - \alpha$.

In practical applications, it is conceivable that the calibration data, that is, the $\mathbf{y}_i$'s in (1.1), are collected and set aside and will be used repeatedly to construct confidence regions for $\boldsymbol{\theta}$ in (1.2) as the corresponding observation vector $\mathbf{y}$ becomes available. In many situations, it may be difficult, expensive or time-consuming to measure the $\mathbf{x}_i$'s, whereas the $\mathbf{y}_i$'s are more easily obtained. Typically, this is the scenario that gives rise to the calibration problem. Hence, when a few $\mathbf{x}_i$'s are accurately measured, one would like to use the corresponding $\mathbf{y}_i$'s repeatedly. This is the justification for constructing multiple-use confidence regions. We shall shortly describe a specific example that is analyzed in this article.

A model that is more general than (1.2) is where $\boldsymbol{\theta}$ in (1.2) is a nonlinear function of fewer unknown parameters, denoted by an $r \times 1$ vector $\boldsymbol{\xi}$ ($r \leq m$). Let $\mathbf{h}(\boldsymbol{\xi})$ denote this $m \times 1$ vector-valued function. Here the functional form of $\mathbf{h}(\boldsymbol{\xi})$ is assumed to be known. We then have the model

$$(1.3) \qquad\qquad \mathbf{y} \sim N(\mathbf{a} + B\mathbf{h}(\boldsymbol{\xi}), \Sigma).$$

The $\mathbf{x}_i$'s in (1.1) are now the values of $\mathbf{h}(\cdot)$ evaluated at known design points. The problem now is the construction of a confidence region for $\boldsymbol{\xi}$ based on (1.1) and (1.3). Note that the model (1.3) will arise when we have polynomial regression. We would like to point out that (1.2) and (1.3) are used as a general notation. In the context of multiple-use confidence regions, it is understood that there will be a sequence of independent $\mathbf{y}$ values corresponding to a sequence of possibly different $\boldsymbol{\theta}$ values (or $\boldsymbol{\xi}$ values) having the distribution specified in (1.2) [respectively, (1.3)].

Regarding the construction of single-use confidence regions, several solutions are available; see Brown (1982), Fujikoshi and Nishii (1984), Brown and Sundberg (1987), Davis and Hayakawa (1987) and Mathew and Kasala (1994), where the models (1.1) and (1.2) are considered, and Oman (1988) and

Mathew and Zha (1996), where the models (1.1) and (1.3) are considered. On the problem of constructing multiple-use confidence regions in the context of the models (1.1) and (1.2), or the models (1.1) and (1.3), solutions are currently unavailable. If we assume that $\Sigma = \sigma^2 I_p$ (where $\sigma^2 \geq 0$ is unknown), some solutions have been obtained recently by Mathew and Zha (1997). Apart from this article, all the available literature on multiple-use confidence regions deals with the strictly univariate case, that is, $p = m = 1$ in the models (1.1) and (1.2) or $p = r = 1$ in the models (1.1) and (1.3). Some of the major references are Scheffé (1973), Carroll, Spiegelman and Sacks (1988), Mee, Eberhardt and Reeve (1991) and Eberhardt and Mee (1994). For a review of the results on single-use and multiple-use confidence regions in some of the articles cited above, see the papers by Osborne (1991) and Sundberg (1994) and the book by Brown (1993).

Derivation of a multiple-use confidence region, satisfying the coverage probability requirements mentioned earlier, presents some theoretical difficulties even in the univariate case. An approach that has been successfully employed in the univariate case is to first derive conservative simultaneous tolerance intervals and then to invert them to obtain conservative multiple-use confidence regions. For details, see Mee, Eberhardt and Reeve (1991) and Mee and Eberhardt (1996). Recently, Lee and Mathew (1998) have looked at regions obtained by inverting tolerance intervals rather than simultaneous tolerance intervals in the univariate case. The numerical results reported in their article show that the regions so obtained satisfy the requirements of multiple-use confidence regions quite well. That this may be the case is also conjectured in Mee and Eberhardt (1996), Section 5. Our approach in this article is motivated by this. That is, we have derived a region for θ in (1.2), and ξ in (1.3), using a tolerance region condition, rather than the condition that is required of simultaneous tolerance regions. Later in the article, we have investigated numerically whether the region so obtained satisfies the coverage probability requirements of multiple-use confidence regions. This numerical study is carried out in the context of an example. The regions derived in Mathew and Zha (1997) are also based on a tolerance region condition.

The paper is organized as follows. In Section 2, we first give the tolerance region condition that is used in our derivation, along with the actual condition that should be satisfied by a multiple-use confidence region. For the models (1.1) and (1.2), we then prove the existence of a region for θ that satisfies the tolerance region condition. The pivot statistic that we have used in our construction is a natural choice; see (2.4) in the next section. Our procedure is conservative, except in the univariate case, that is, $p = m = 1$ in (1.1) and (1.2). In Section 3, we give a generalization of our procedure applicable to the models (1.1) and (1.3). The computations required for the practical implementation of our procedure are described in Section 4. It turns out that our procedure is computationally involved and closed form solutions are not available. Section 5 deals with an example taken from Oman and Wax (1984), which deals with the estimation of gestational age (i.e., week of pregnancy) based on two fetal bone lengths. The model relating the bone lengths to the

gestational age is given in Oman and Wax (1984); the gestational age enters the model nonlinearly so that the model (1.3) is applicable. The data that are available for this example consist of fetal bone length measurements for several women whose gestational ages are precisely known. This data can be used repeatedly in order to construct confidence regions for the unknown gestational age of women after observing the corresponding fetal bone lengths. In other words, it is required to construct multiple-use confidence regions. It should be mentioned that for this problem, single-use confidence regions have been constructed by Oman (1988) and Mathew and Zha (1996). In Section 5, for the same problem, we have implemented numerically the procedure in this article for the construction of a region that satisfies the tolerance region condition. We have also investigated numerically whether the region derived using the tolerance region condition does satisfy the coverage probability requirements of multiple-use confidence regions. The numerical results indicate that this is the case. Some concluding remarks appear in Section 6.

2. A region for $\boldsymbol{\theta}$ in the model (1.2).   This section deals with the construction of a region for $\boldsymbol{\theta}$ in the models (1.1) and (1.2) using a tolerance region condition. 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 Y and X are, respectively, $p \times N$ and $m \times N$ matrices and we assume $\text{rank}(\mathbf{1}_N, X) = m + 1$. If $\text{vec}(Y)$ denotes the $Np \times 1$ column vector obtained by writing the columns of Y one below the other, (1.1) implies $\text{cov}(\text{vec}(Y)) = I_N \otimes \Sigma$. The model (1.1) can thus be written as

$$(2.1) \qquad\qquad Y \sim N(\mathbf{a}\mathbf{1}'_N + BX, I_N \otimes \Sigma).$$

The least-squares estimators of B and $\mathbf{a}$ are given by

$$(2.2) \quad \hat{B} = Y\left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}'_N\right)X'\left[X\left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}'_N\right)X'\right]^{-1}, \qquad \hat{\mathbf{a}} = \bar{\mathbf{y}} - \hat{B}\bar{\mathbf{x}},$$

where $\bar{\mathbf{y}}$ and $\bar{\mathbf{x}}$ are, respectively, the averages of the column vectors of Y and X. Also, let

$$(2.3) \qquad\qquad S = (Y - \hat{\mathbf{a}}\mathbf{1}'_N - \hat{B}X)(Y - \hat{\mathbf{a}}\mathbf{1}'_N - \hat{B}X)'.$$

Then $S \sim W_p(N - m - 1, \Sigma)$, where $W_p(f, \Sigma)$ denotes the p-dimensional Wishart distribution with $df = f$ and scale matrix Σ. Furthermore, S is distributed independently of $(\hat{\mathbf{a}}, \hat{B})$. It is assumed that $N - m - 1 \geq p$ and $p \geq m$. These assumptions are required for the positive definiteness (with probability one) of S and the identifiability of $\boldsymbol{\theta}$, respectively.

2.1. *The pivot statistic.*   The pivot statistic that we shall use for constructing our region for $\boldsymbol{\theta}$ is motivated by the following observation. From (1.2), it follows that if $\mathbf{a}$, B and Σ are known, the weighted least-squares estimator of $\boldsymbol{\theta}$ is $\tilde{\boldsymbol{\theta}} = (B'\Sigma^{-1}B)^{-1}B'\Sigma^{-1}(\mathbf{y} - \mathbf{a})$ with covariance matrix $(B'\Sigma^{-1}B)^{-1}$. Hence, a confidence region for $\boldsymbol{\theta}$ should be based on the pivot $(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})'B'\Sigma^{-1}B(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta})$. When $\mathbf{a}$, B and Σ are unknown, we shall replace them by the corresponding estima-

tors, namely $\hat{\mathbf{a}}$, $\hat{B}$ and $(1/N - m - 1)S$. Thus, let $\hat{\boldsymbol{\theta}} = (\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1}(\mathbf{y} - \hat{\mathbf{a}})$, and a natural pivot for constructing a confidence region for $\boldsymbol{\theta}$ is

$$
(2.4) \quad T(\boldsymbol{\theta}) = \frac{N - m - p}{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\hat{B}'S^{-1}\hat{B}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{N - m - p}{m}
$$
$$
\times \left[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\right]'S^{-1}\hat{B}(\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1}\left[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\right].
$$

Later in our derivation, it will be necessary to consider the conditional distribution of $T(\boldsymbol{\theta})$, conditionally given the calibration data Y. Note that this is the same as the conditional distribution of $T(\boldsymbol{\theta})$, conditionally given $\bar{\mathbf{y}}$, $\hat{B}$ and S. Also note that the three quantities, $\bar{\mathbf{y}}$, $\hat{B}$ and S, are independently distributed.

2.2. *Coverage probability requirements.* The region that we shall construct, say $\mathscr{C}(\boldsymbol{\theta})$, is given by

$$
(2.5) \quad \mathscr{C}(\boldsymbol{\theta}) = \left\{\boldsymbol{\theta}: T(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta})\right\},
$$

where $k(\boldsymbol{\theta})$ is to be determined subject to appropriate coverage probability requirements. Obviously $\mathscr{C}(\boldsymbol{\theta})$ depends on the data and is a function of only the data; however, for notational convenience, we have suppressed this fact. Note that $k(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}$. (It turns out that a function of $\boldsymbol{\theta}$ is required in order to satisfy our coverage probability condition). We shall compute $k(\boldsymbol{\theta})$ satisfying a condition similar to that required of a tolerance region. The condition is given by

$$
(2.6) \quad P_{\bar{\mathbf{y}}, \hat{B}, S}\left[P_{\mathbf{y}}\{T(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta})|\bar{\mathbf{y}}, \hat{B}, S\} \geq 1 - \beta\right] \geq 1 - \alpha,
$$

for all $\mathbf{a}$, B, Σ and $\boldsymbol{\theta} \in \Theta$, where Θ is the parameter space for $\boldsymbol{\theta}$.

When the calibration curve is used repeatedly for constructing a sequence of confidence regions, the coverage probability requirement that we want is different from that in (2.6). We shall now explain this. Let

$$
(2.7) \quad C(\boldsymbol{\theta}; \bar{\mathbf{y}}, \hat{B}, S) = P_{\mathbf{y}}\{T(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta})|\bar{\mathbf{y}}, \hat{B}, S\}.
$$

Let $\{\boldsymbol{\theta}_i\}$, $i = 1, 2, \ldots, n$, denote a sequence of n values of $\boldsymbol{\theta}$, corresponding to a sequence of n independent $\mathbf{y}$ values. As soon as we observe a $\mathbf{y}$ value, a confidence region for the corresponding $\boldsymbol{\theta}$ can be constructed as in (2.5) using the observed $\mathbf{y}$ value and the calibration data. In this case, $k(\boldsymbol{\theta})$ is to be determined subject to the following condition:

$$
(2.8) \quad P_{\bar{\mathbf{y}}, \hat{B}, S}\left[\frac{1}{n}\sum_{i=1}^{n} C(\boldsymbol{\theta}_i; \bar{\mathbf{y}}, \hat{B}, S) \geq 1 - \beta\right] \geq 1 - \alpha,
$$

for every sequence $\{\boldsymbol{\theta}_i\}$, $i = 1, 2, \ldots, n$. The interpretation of this condition is as follows. If the calibration data is used to construct a sequence of confidence regions for a sequence of possibly different $\boldsymbol{\theta}$ values corresponding to a sequence of independent $\mathbf{y}$ values, then at least a proportion $1 - \beta$ of such confidence regions will include the corresponding true $\boldsymbol{\theta}$ values with a confidence of at least $1 - \alpha$. Thus, when $k(\boldsymbol{\theta})$ is determined subject to (2.8), we shall call the confidence region (2.5) a multiple-use confidence region. The condition (2.8)

is a multivariate analogue of the condition (5) in Mee and Eberhardt (1996). We note that, typically, multiple-use confidence regions are derived under the much stronger requirement

$$(2.9) \qquad P_{\bar{\mathbf{y}}, \hat{B}, S}\Big[\inf_{\boldsymbol{\theta} \in \Theta} C(\boldsymbol{\theta}; \bar{\mathbf{y}}, \hat{B}, S) \geq 1 - \beta\Big] \geq 1 - \alpha,$$

where Θ is the parameter space for θ. In fact, most of the available results in the univariate case deal with the computation of $k(\boldsymbol{\theta})$ satisfying (2.9); see Mee and Eberhardt (1996) for a review. In the multivariate set up, we have not investigated the problem based on (2.8) or (2.9). Instead, our attempt in this article is to derive $k(\boldsymbol{\theta})$ satisfying (2.6). We would like to point out that, as such, the condition (2.6) does not appear to have any practically relevant interpretation for the calibration problem. Nevertheless, (2.6) is analytically more tractable compared to (2.8) or (2.9). Furthermore, the numerical results in the univariate case indicate that (2.6) implies (2.8); see Lee and Mathew (1998). Later, for an example, we have investigated whether $k(\boldsymbol{\theta})$ satisfying (2.6) also satisfies (2.8).

2.3. *A region that satisfies* (2.6). We now turn our attention to (2.6). Theorem 2.1 in this section gives our main result that shows the existence of $k(\boldsymbol{\theta})$ satisfying (2.6). As will be seen later, Theorem 2.1 also facilitates the computation of $k(\boldsymbol{\theta})$. We shall first obtain a lower bound for $P[T(\boldsymbol{\theta}) \leq k|\bar{\mathbf{y}}, \hat{B}, S]$, for any $k > 0$. Note that in order to study (2.6), we can assume $\Sigma = I_p$. This is because we can replace $\mathbf{y}, \bar{\mathbf{y}}, \hat{B}$ and S by $\Sigma^{-1/2}\mathbf{y}, \Sigma^{-1/2}\bar{\mathbf{y}}, \Sigma^{-1/2}\hat{B}$ and $\Sigma^{-1/2}S\Sigma^{-1/2}$, respectively, and this does not change $T(\boldsymbol{\theta})$. The following notation is used in the sequel. We shall use $\chi^2(s; \eta)$ to denote a noncentral chi-square random variable with s degrees of freedom and noncentrality parameter η. A central chi-square random variable with s degrees of freedom will be denoted by $\chi^2(s)$. When there are several independent chi-square random variables involved, we shall use the notation $\chi_i^2(s_i; \eta_i)$ in the noncentral case, and $\chi_i^2(s_i)$ in the central case ($i = 1, 2, \ldots$).

Lemma 2.1 below gives a lower bound for $P[T(\boldsymbol{\theta}) \leq k|\bar{\mathbf{y}}, \hat{B}, S]$. The proof of the lemma is given in the Appendix.

LEMMA 2.1. *Consider the models* (1.1) *and* (1.2) *and assume* $\Sigma = I_p$. *Let* $T(\boldsymbol{\theta})$ *be as defined in* (2.4).

(i) *Suppose* $p > m$. *Let* $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ *denote the ordered nonzero eigenvalues of the matrix* $S^{-1}\hat{B}(\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1}$. *Write*

$$\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)',$$

$$\boldsymbol{\theta}_1 = \left\{X\Big(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N'\Big)X'\right\}^{-1/2}(\boldsymbol{\theta} - \bar{\mathbf{x}}),$$

$$(2.10) \qquad v = \Big(\frac{1}{N} + \boldsymbol{\theta}_1'\boldsymbol{\theta}_1\Big)^{-1}\big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]'$$

$$\times \big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big].$$

Then, for $k > 0$,

$$P\big[T(\boldsymbol{\theta}) \le k | \bar{\mathbf{y}}, \hat{B}, S\big]$$

(2.11)

$$\ge P\bigg[\frac{N-m-p}{m}\bigg\{\lambda_1\chi_1^2\bigg(1; \bigg(\frac{1}{N} + \boldsymbol{\theta}_1'\boldsymbol{\theta}_1\bigg)v\bigg) + \sum_{i=2}^{m}\lambda_i\chi_i^2(1)\bigg\} \le k | \boldsymbol{\lambda}, v\bigg].$$

(ii) *Suppose* $p = m$. *Then*

$$T(\boldsymbol{\theta}) = \frac{N-2m}{m}\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]' S^{-1}\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\big],$$

and (2.11) *holds with* $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ *being the eigenvalues of* S^{-1}.

Note that the definition of $\boldsymbol{\theta}_1$ in (2.10) requires the positive definiteness of the matrix $X(I_N - (1/N)\mathbf{1}_N\mathbf{1}_N')X'$. This positive definiteness follows in view of our assumption $\text{rank}(\mathbf{1}_N, X) = m + 1$.

The next lemma is on the distributions of $\boldsymbol{\lambda}$ and v given in (2.10). It turns out that the distributions of $\boldsymbol{\lambda}$ and v are free of any unknown parameters. When $p = m$, it is obvious that the distribution of $\boldsymbol{\lambda}$ is free of any unknown parameters, since $S \sim W_p(N - m - 1, I_p)$. In the statement of Lemma 2.2, $B(q_1, q_2, I_p)$ denotes the matrix variate beta distribution with parameters q_1, q_2; see (A.8) in the Appendix for the definition of the matrix variate beta distribution.

LEMMA 2.2. *Suppose* $\boldsymbol{\lambda}$ *and* v *are as defined in Lemma* 2.1. *For* $p > m$, *let* $G_1 \sim W_m(N - p - 1, I_m)$, $G_2 \sim W_{p-m}(N - m - 1, I_{p-m})$ *and* $U \sim N(0, I_m \otimes I_{p-m})$, *where* G_1, G_2 *and* U *are independently distributed. Then,* (i) *for* $p > m$, *the distribution of* $\boldsymbol{\lambda}$ *is the same as the distribution of the ordered eigenvalues of* $(I + UG_2^{-1}U')G_1^{-1}$, (ii) *if* $m \le p/2$, *the distribution of* $\boldsymbol{\lambda}$ *is the same as the distribution of the ordered eigenvalues of* $H^{-1}G_1^{-1}$, *where* G_1 *is as defined above and* $H \sim B(N - p + m - 1, p - m, I_m)$, (iii) $v \sim \chi^2(p)$ *and* (iv) $\boldsymbol{\lambda}$ *and* v *are independently distributed.*

The proof of Lemma 2.2 is also given in the Appendix.

We shall now state our main result regarding the existence of $k(\boldsymbol{\theta})$ satisfying (2.6). It turns out that such a $k(\boldsymbol{\theta})$ is a function of $\boldsymbol{\theta}_1'\boldsymbol{\theta}_1$, where $\boldsymbol{\theta}_1$ is given in (2.10). Hence, from now on, we shall use the notation $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$ instead of $k(\boldsymbol{\theta})$.

THEOREM 2.1. *Suppose* $\boldsymbol{\lambda}$ *and* v *are as defined in Lemma* 2.1. *Let* $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$ *satisfy*

(2.12)

$$P_{\boldsymbol{\lambda}, v}\bigg[P\bigg(\frac{N-m-p}{m}\bigg\{\lambda_1\chi_1^2\bigg(1; \bigg(\frac{1}{N} + \boldsymbol{\theta}_1'\boldsymbol{\theta}_1\bigg)v\bigg) + \sum_{i=2}^{m}\lambda_i\chi_i^2(1)\bigg\}$$

$$\le k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)\big|\boldsymbol{\lambda}, v\bigg) \ge 1 - \beta\bigg] = 1 - \alpha.$$

Then

(2.13) $P_{\bar{\mathbf{y}}, \hat{B}, S}\big[P_{\mathbf{y}}\{T(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)|\bar{\mathbf{y}}, \hat{B}, S\} \geq 1 - \beta\big] \geq 1 - \alpha.$

PROOF.   Noting that $\boldsymbol{\lambda}$ and v depend only on $\bar{\mathbf{y}}$, $\hat{B}$ and S, the theorem follows from the inequality (2.11).   □

From (2.13) we conclude that

(2.14) $\mathscr{C}(\boldsymbol{\theta}) = \big\{\boldsymbol{\theta}\colon T(\boldsymbol{\theta}) \leq k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)\big\}$

is a region that satisfies (2.6). Then $\mathscr{C}(\boldsymbol{\theta})$ in (2.14) is conservative. That is, the left-hand side (lhs) of (2.6) is not exactly equal to $1 - \alpha$; it may be more than $1 - \alpha$.

2.4. *Some remarks on Lemma* 2.1 *and Theorem* 2.1.   We shall now make several observations regarding Lemma 2.1 and Theorem 2.1. The main feature to note is that the right-hand side of (2.11) is a random variable that depends on $\boldsymbol{\lambda}$ and v and the distributions of $\boldsymbol{\lambda}$ and v are free of the nuisance parameters $\mathbf{a}$, B and Σ; in fact, the distributions of $\boldsymbol{\lambda}$ and v are free of any unknown parameters (see Lemma 2.2). The conservatism of the region (2.14), already noted above, results from the inequality in (2.11). A careful examination of the proof of Lemma 2.1 (given in the Appendix), shows that inequality occurs in (2.11) for two reasons: (1) inequality occurs in (A.6), and (2) inequality occurs in (A.7), where (A.6) and (A.7) are given in the Appendix. If any one of these is an equality, the region (2.14) will be less conservative. If both (A.6) and (A.7) are equalities, the region (2.14) will be exact, that is, the lhs of (2.13) will be equal to $1 - \alpha$. Let us examine when equality will hold in (A.6) and (A.7). For equality to hold in (A.6), it is necessary to have $m = 1$. For equality to hold in (A.7), we need the condition $\sum_{i=1}^{m} \mathbf{r}_i\mathbf{r}_i' = I_p$, that is, $m = p$, where the $\mathbf{r}_i$'s are given in (A.3). In other words, equality can hold in both (A.6) and (A.7) only in the strictly univariate case, that is, $p = m = 1$. If at least one of the quantities p and m is more than 1, we cannot get an exact region satisfying (2.6), based on $T(\boldsymbol{\theta})$; it will only be conservative. In other words, the lhs of (2.13) will not be exactly equal to $1 - \alpha$ when p and/or m is greater than 1.

Another important aspect to note is that the quantity $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$, to be used in the construction of the region (2.14), depends on $\boldsymbol{\theta}_1$, the parameter of interest. The same is true of the confidence regions derived in Mathew and Zha (1996, 1997). Clearly, in order to implement (2.14) in practice, one needs to know the functional form of $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$. The numerical computation of the functional form of $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$ will be discussed in Section 4. As we shall see, it will be necessary to generate the values of the random variables $\lambda_1, \lambda_2, \ldots, \lambda_m$ and v in order to accomplish the above computation. From the proof of Lemma A.2 given in the Appendix, it follows that as far as the distribution of $\boldsymbol{\lambda}$ is concerned, we can assume $\hat{B} = (I_m, 0)'$. Hence, one realization of $\boldsymbol{\lambda}$ can be obtained by generating a value of the Wishart matrix S and computing the nonzero eigenvalues of $S^{-1}\hat{B}(\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1}$, with $\hat{B} = (I_m, 0)'$ [see (A.2) and (A.3)

in the Appendix]. However, representing the λ_i's as the eigenvalues of the matrix $(I + UG_2^{-1}U')G_1^{-1}$, as given in Lemma 2.2, can be advantageous in the numerical computations. If $m \le p/2$, it follows from Lemma 2.2(ii) that the λ_i's are the eigenvalues of the matrix $H^{-1}G_1^{-1}$, where $H = (I + UG_2^{-1}U')^{-1} \sim B(N - p + m - 1, p - m, I_m)$. In particular, if $m = 1$ and $p > 1$, then $m \le p/2$ and $(I + UG_2^{-1}U')^{-1}$ is a scalar having the univariate beta distribution. Also, when $m = 1$, G_1 is a scalar random variable having a chi-square distribution. Thus, in the special case $m = 1$ and $p > 1$, λ_1 coincides with the scalar $(I + UG_2^{-1}U')G_1^{-1}$, which is the reciprocal of the product of a beta random variable and a chi-square random variable, both independent. In other words, when $m = 1$, generating values of λ_1 is an easy task even if p is large, resulting in a Wishart matrix S having a large dimension. We have indeed used this observation in our example in Section 5.

3. A region for $\boldsymbol{\xi}$ in the model (1.3). This section describes a generalization of the results in Section 2 when the models (1.1) and (1.3) are applicable. That is, we have the models

$$(3.1) \qquad Y \sim N(\mathbf{a}\mathbf{1}_N' + BX, I_N \otimes \Sigma), \qquad \mathbf{y} \sim N(\mathbf{a} + B\mathbf{h}(\boldsymbol{\xi}), \Sigma),$$

where $\boldsymbol{\xi}$ is an $r \times 1$ vector and $\mathbf{h}(\boldsymbol{\xi})$ is an $m \times 1$ vector- valued function of $\boldsymbol{\xi}$. The problem is to derive a region for $\boldsymbol{\xi}$ based on a criterion similar to (2.6). Let $\bar{\mathbf{y}}, \hat{B}$ and S be as in the previous section. Following Oman (1988) and Mathew and Zha (1996, 1997), we assume that the components of $\mathbf{h}(\boldsymbol{\xi})$ are differentiable functions of $\boldsymbol{\xi}$. Under this assumption, let $H(\boldsymbol{\xi})$ be the $p \times r$ matrix defined as

$$(3.2) \qquad H(\boldsymbol{\xi}) = \hat{B}\frac{\partial \mathbf{h}_1(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}.$$

As in Oman [(1988), page 179], we also assume that $H(\boldsymbol{\xi})$ has rank r (with probability one) for all $\boldsymbol{\xi}$ belonging to the appropriate parameter space. Now define

$$(3.3) \qquad \begin{aligned} T(\boldsymbol{\xi}) = \frac{N - m - p}{r}&\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\mathbf{h}(\boldsymbol{\xi}) - \bar{\mathbf{x}})\big]' S^{-1} H(\boldsymbol{\xi}) \\ &\times \big[H(\boldsymbol{\xi})'S^{-1}H(\boldsymbol{\xi})\big]^{-1} H(\boldsymbol{\xi})'S^{-1}\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\mathbf{h}(\boldsymbol{\xi}) - \bar{\mathbf{x}})\big]. \end{aligned}$$

Lemma 2.1, Lemma 2.2 and Theorem 2.1 can now be easily generalized and we get the following inequality, similar to (2.11):

$$(3.4) \qquad \begin{aligned} &P\big[T(\boldsymbol{\xi}) \le k\big|\bar{\mathbf{y}}, \hat{B}, S\big] \\ &\quad\ge P\bigg[\frac{N - m - p}{r}\bigg\{\lambda_1\chi_1^2\bigg(1; \bigg(\frac{1}{N} + \mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi})\bigg)v\bigg) \\ &\qquad\qquad\qquad\qquad + \sum_{i=2}^{r}\lambda_i\chi_i^2(1)\bigg\} \le k\big|\boldsymbol{\lambda}, v\bigg]. \end{aligned}$$

In (3.4),

$$\mathbf{h}_1(\boldsymbol{\xi}) = \left\{ X\left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}_N' \right) X' \right\}^{-1/2} (\mathbf{h}(\boldsymbol{\xi}) - \bar{\mathbf{x}}),$$

(3.5)
$$v = \left(\frac{1}{N} + \mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi}) \right)^{-1} \left[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\mathbf{h}(\boldsymbol{\xi}) - \bar{\mathbf{x}}) \right]'$$

$$\times \left[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\mathbf{h}(\boldsymbol{\xi}) - \bar{\mathbf{x}}) \right]$$

and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)'$, $\lambda_1 > \lambda_2 > \cdots > \lambda_r$, are the nonzero eigenvalues of $S^{-1}H(\boldsymbol{\xi})[H(\boldsymbol{\xi})'S^{-1}H(\boldsymbol{\xi})]^{-1}H(\boldsymbol{\xi})'S^{-1}$. Furthermore, $v \sim \chi^2(p)$ and is distributed independently of $\boldsymbol{\lambda}$, and the distribution of $\boldsymbol{\lambda}$ is the same as the distribution of the ordered eigenvalues of the matrix $(I + UG_2^{-1}U')G_1^{-1}$, with $G_1 \sim W_r(N-m-p+r-1, I_r)$, $G_2 \sim W_{p-r}(N-m, I_{p-r})$, $U \sim N(0, I_r \otimes I_{p-r})$, where G_1, G_2 and U are independently distributed. If $r \le p/2$, the λ_i's are the eigenvalues of $H^{-1}G_1^{-1}$, where $H \sim B(N-m-p+2r-1, p-r, I_r)$ independently of G_1. If we further have $r = 1$ and $p > 1$, then λ_1 is the reciprocal of the product of a univariate beta random variable $B(N-m-p+1, p-1)$, and a central chi-square random variable with $N-m-p$ degrees of freedom. These conclusions follow from the discussion toward the end of Section 2.

Now suppose $k(\mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi}))$ is such that

(3.6)
$$P_{\boldsymbol{\lambda}, v}\left[P\left(\frac{N-m-p}{r}\left\{ \lambda_1\chi_1^2\left(1; \left(\frac{1}{N} + \mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi}) \right)v \right) + \sum_{i=2}^{r}\lambda_i\chi_i^2(1) \right\} \right.\right.$$

$$\left.\left. \le k(\mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi})) \Big| \boldsymbol{\lambda}, v \right) \ge 1 - \beta \right] = 1 - \alpha.$$

Then the region

(3.7)
$$\mathcal{C}(\boldsymbol{\xi}) = \left\{ \boldsymbol{\xi}: T(\boldsymbol{\xi}) \le k(\mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi})) \right\}$$

is the required region for $\boldsymbol{\xi}$, satisfying the condition

$$P_{\bar{\mathbf{y}}, \hat{B}, S}\left[P_{\mathbf{y}}\left\{ T(\boldsymbol{\xi}) \le k(\mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi})) \big| \bar{\mathbf{y}}, \hat{B}, S \right\} \ge 1 - \beta \right] \ge 1 - \alpha.$$

4. Computation of the regions (2.14) and (3.7). In order to implement the region (2.14) for $\boldsymbol{\theta}$, or (3.7) for $\boldsymbol{\xi}$, the major practical problem is the computation of $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$ satisfying (2.12) or $k(\mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi}))$ satisfying (3.6), where $\boldsymbol{\theta}_1$ is given in (2.10) and $\mathbf{h}_1(\boldsymbol{\xi})$ is given in (3.5). Note that the functional forms of these quantities are required. We shall now explain the numerical computation of $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$ satisfying (2.12). The computation of $k(\mathbf{h}_1(\boldsymbol{\xi})'\mathbf{h}_1(\boldsymbol{\xi}))$ satisfying (3.6) will be similar. Let

(4.1)
$$d = \boldsymbol{\theta}_1'\boldsymbol{\theta}_1.$$

In practical applications of calibration where the models (1.1) and (1.2) are applicable, the parameter of interest, namely $\boldsymbol{\theta}$, represents a physical quantity

and a parameter space for $\boldsymbol{\theta}$, and hence for $\boldsymbol{\theta}_1$, will be known. In particular, an upper bound, say δ, will be available on $d = \boldsymbol{\theta}_1'\boldsymbol{\theta}_1$. Thus

$$(4.2) \qquad\qquad 0 \leq d \leq \delta.$$

Our numerical computation of $k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1) = k(d)$ will be for d satisfying (4.2). Note from (2.12) that $k(d)$ is an increasing function of d. Our approach for numerically obtaining the functional form of $k(d)$ is in the same spirit as in Mathew and Zha (1996, 1997), namely, numerically evaluate the value of $k(d)$ satisfying (2.12) for a few specified values of d satisfying (4.2), plot these values of $k(d)$ against d and try to fit a suitable function to the plot. The fitted function will give approximately the functional form of $k(d) = k(\boldsymbol{\theta}_1'\boldsymbol{\theta}_1)$, which can be used to compute the region (2.14).

Thus, our problem is the numerical evaluation of $k(d)$ satisfying (2.12) for various values of $d = \boldsymbol{\theta}_1'\boldsymbol{\theta}_1$, subject to (4.2) for specified values of α and β. For any given value of d, say d_1, we will need a starting value of $k(d_1)$, say $k_*(d_1)$, for evaluating the actual value of $k(d_1)$. The starting value $k_*(d_1)$ that we shall use will be such that $k_*(d_1) < k(d_1)$ and its choice for any $d_1 \geq 0$ will be described shortly. Once $k_*(d_1)$ is chosen for a specified value of d_1, such that $k_*(d_1) < k(d_1)$, we shall increase the value of $k_*(d_1)$ in steps, each time numerically evaluating the lhs of (2.12) until it is equal to $1 - \alpha$ up to a desired level of accuracy. The lhs of (2.12) can be evaluated for a given value of d_1 and $k_*(d_1)$ in the following manner. Generate one set of values of the Wishart matrices G_1 and G_2, the normal matrix U and the chi-square random variable v specified in Lemma 2.2 and compute the ordered eigenvalues $\lambda_1 > \lambda_2 > \cdots > \lambda_m$ of the matrix $(I_m + UG_2^{-1}U')G_1^{-1}$. (If $p = m$, we need to generate only the Wishart matrix S and the λ_i's are the ordered eigenvalues of S^{-1}). For the value of $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and v so obtained, let $i(\boldsymbol{\lambda}, v)$ be an indicator function that takes the value one if

$$(4.3) \quad P\left[\frac{N-m-p}{m}\left\{ \lambda_1 \chi_1^2\left(1; \left(\frac{1}{N} + d_1 \right) v \right) + \sum_{i=2}^{m} \lambda_i \chi_i^2(1) \right\} \leq k_*(d_1) \right] \geq 1 - \beta.$$

If (4.3) does not hold, assign the value zero to $i(\boldsymbol{\lambda}, v)$. Note that the computation of $i(\boldsymbol{\lambda}, v)$ requires the computation of the lhs of (4.3), which is a probability involving linear combinations of independent chi-square random variables. For any specified value of $\boldsymbol{\lambda}$ and v, the lhs of (4.3) can be evaluated using simulation. The values of $\boldsymbol{\lambda}$ and v can be generated a large number of times, and the value of $i(\boldsymbol{\lambda}, v)$ can be evaluated each time. The lhs of (2.12) is the proportion of times $i(\boldsymbol{\lambda}, v)$ takes the value one. If this proportion is less than $1 - \alpha$, then $k_*(d_1) < k(d_1)$ and the value of $k_*(d_1)$ has to be increased in order to get a better approximation of $k(d_1)$. As already pointed out, we increase the value of $k_*(d_1)$ in steps, each time computing the proportion of times $i(\boldsymbol{\lambda}, v)$ takes the value one until this proportion is approximately equal to $1 - \alpha$. Once $k(d)$ is thus computed for a few values of d, they can be plotted in order to obtain the approximate functional form of $k(d)$.

We shall now discuss the choice of the starting value $k_*(d_1)$ satisfying $k_*(d_1) < k(d_1)$. In order to accomplish this, we shall first derive some lower

bounds for $k(0)$. The result is stated in the following lemma; its proof is given in Sharma (1996).

LEMMA 4.1. *Let* $k(d)$ *satisfy* (2.12) *for* $d = \boldsymbol{\theta}_1'\boldsymbol{\theta}_1$ *and let* $\chi_\varepsilon^2(f)$ *and* $F_\varepsilon(f_1, f_2)$, *respectively, denote the* 100ε*th percentile of the central chi-square distribution with* f *degrees of freedom and the* 100ε*th percentile of the central* F *distribution with* (f_1, f_2) *degrees of freedom. Furthermore, let* $\lambda_{\max; \varepsilon}$ *denote the* 100ε*th percentile of the distribution of the maximum eigenvalue of the Wishart matrix* G_1 *specified in Lemma 2.2. Then, for* $d = 0$:

(i) $\quad k(0) \geq \dfrac{N - m - p}{m} \dfrac{\chi_{1-\beta}^2(m)}{\lambda_{\max; \alpha}} \geq \dfrac{N - m - p}{m} \dfrac{\chi_{1-\beta}^2(m)}{\chi_\alpha^2(m(N - p - 1))};$

(ii) $\quad k(0) \geq F_{(1-\beta)(1-\alpha)}(m, N - m - p).$

From Lemma 4.1, it follows that

(4.4)
$$k_*(0) = \max\left[\dfrac{N - m - p}{m} \dfrac{\chi_{1-\beta}^2(m)}{\chi_\alpha^2(m(N - p - 1))}, \right.$$
$$\left. F_{(1-\beta)(1-\alpha)}(m, N - m - p) \right]$$

is a lower bound for $k(0)$ and can be taken as a starting value for computing $k(0)$. The value of $k_*(0)$ can be increased in steps in order to arrive at $k(0)$, as already explained. Now consider a finite sequence of values

(4.5) $$0 < d_1 < d_2 < \cdots < d_s = \delta.$$

Since $k(d)$ is an increasing function of d, we have $k(0) < k(d_1)$. In other words, once $k(0)$ is numerically obtained, it can serve as a starting value for the evaluation of $k(d_1)$. In general, $k(d_i)$ is a starting value for the evaluation of $k(d_{i+1})$ when $d_i < d_{i+1}$.

The value of s and the choice of the d_i's in the interval $[0, \delta]$ satisfying (4.5) are clearly subjective. If δ is small, the interval $[0, \delta]$ will be narrow and perhaps the numerical evaluation of $k(d_i)$ for a small number of d_i's may be enough to determine approximately the functional form of $k(d)$. It should be pointed out that δ is expected to be small (in fact, less than one) in many applications. This is for the following reason. It is reasonable to assume that the models (1.1) and (1.2) will be used for statistical inference concerning $\boldsymbol{\theta}$ when $\boldsymbol{\theta}$ is "like" the $\mathbf{x}_i$'s in (1.1). Depending on the particular application, this fact can be expressed in different ways. Here we consider two possibilities: (1) $\boldsymbol{\theta}$ is a convex combination of the $\mathbf{x}_i$'s, (2) $\min_{1 \leq i \leq N} x_{ij} \leq \theta_j \leq \max_{1 \leq i \leq N} x_{ij}$, where θ_j and x_{ij} are, respectively, the jth components of the vectors $\boldsymbol{\theta}$ and $\mathbf{x}_i$ $(1 \leq i \leq N)$. When case (1) holds, we can write $\boldsymbol{\theta} = X\boldsymbol{\psi}$, where $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_N)'$ satisfies $\psi_i \geq 0$ and $\sum_{i=1}^N \psi_i = 1$. Hence $d = \boldsymbol{\theta}_1'\boldsymbol{\theta}_1 = \boldsymbol{\psi}'X'(XX')^{-1}X\boldsymbol{\psi} \leq \boldsymbol{\psi}'\boldsymbol{\psi} \leq 1$. In other words, $\delta = 1$ is an upper bound for d, when case (1) is valid. The actual upper bound may in fact be much less than one. This upper bound can

be easily computed once we know the parameter space for $\boldsymbol{\theta}$ for the particular application one is dealing with. When case (2) holds, an upper bound for $d = \boldsymbol{\theta}'_1 \boldsymbol{\theta}_1$ is the maximum of $\boldsymbol{\theta}'(XX')^{-1}\boldsymbol{\theta}$, where the maximization is done subject to the condition $\min_{1 \leq i \leq N} x_{ij} \leq \theta_j \leq \max_{1 \leq i \leq N} x_{ij}$. This maximization will have to be done numerically. The fact that $\boldsymbol{\theta}'_1 \boldsymbol{\theta}_1 \leq 1$ when $\boldsymbol{\theta}$ is a convex combination of the $\mathbf{x}_i$'s is also pointed out in Mathew and Zha (1997). Note that the computational procedure described here uses only the fact that d is bounded; it does not depend on the assumptions in case (1) and case (2) mentioned above. Needless to say, the experimenter has to determine the upper bound δ on d, based on the physical constraints on $\boldsymbol{\theta}$, dictated by the problem at hand.

One can drastically reduce the numerical computations by using $k(\delta)$ instead of the function $k(d)$. In other words, numerically evaluate the value of $k(d)$ only at $d = \delta$ and, instead of (2.14), compute the region $\{\theta: T(\boldsymbol{\theta}) \leq k(\delta)\}$. Such a region will have a larger volume compared to (2.14) since $k(d) \leq k(\delta)$. However, if $k(\delta)$ is not too large compared to $k(d)$, one can compute and use only $k(\delta)$ to obtain the region. Since $k_*(0) \leq k(d) \leq k(\delta)$ for all $d \in [0, \delta]$, if $k(\delta)$ is not too large compared to $k_*(0)$, one can avoid computing $k(d)$ for other values of d and use $k(\delta)$ to get the region.

Here is a summary of the numerical procedure for the evaluation of $k(d)$.

1. Start with $k_*(0)$ in (4.4) for the evaluation of $k(0)$. Since $k_*(0) \leq k(0)$, the value of $k_*(0)$ can be increased in steps, each time evaluating the lhs of (2.12), until its value is approximately equal to $1 - \alpha$. The numerical evaluation of the lhs of (2.12) is already explained in this section.
2. Fix s values of d satisfying (4.5). For $d_{i+1} > d_i$, $k(d_i)$ can be taken as a starting value for the computation of $k(d_{i+1})$. The value of $k(d_i)$ can be increased in steps, as mentioned before, in order to arrive at $k(d_{i+1})$, $i = 0, 1, 2, \ldots, s-1$.
3. The pairs $(d_i, k(d_i))$, $i = 0, 1, 2, \ldots, s$ (where $d_0 = 0$) can be plotted and a suitable function can be fitted. Since we have a finite interval $[0, \delta]$, a polynomial of appropriate degree should provide a good fit.

So far, we have explained the computation of $k(d)$ satisfying (2.12) for $d = \boldsymbol{\theta}'_1 \boldsymbol{\theta}_1$. For $d = \mathbf{h}_1(\boldsymbol{\xi})' \mathbf{h}_1(\boldsymbol{\xi})$, the computation of $k(d)$ satisfying (3.6) is obviously similar with obvious modifications in the constants. For example, (4.4) should be replaced by

$$
\text{(4.6)} \quad k_*(0) = \max \left[\frac{N - m - p}{r} \frac{\chi^2_{1-\beta}(r)}{\chi^2_\alpha(r(N - m - p + r - 1))}, \right.
$$
$$
\left. F_{(1-\beta)(1-\alpha)}(r, N - m - p) \right]
$$

and we have $k(0) \geq k_*(0)$.

REMARK 4.1. The lower bound $((N - m - p)/m)(\chi^2_{1-\beta}(m)/\lambda_{\max; \alpha})$ for $k(0)$ may turn out to be larger than $k_*(0)$ in (4.4), at least in some cases. It

is certainly preferable to start with a larger lower bound for the numerical evaluation of $k(0)$. However, the practical difficulty in using $((N - m - p)/m)(\chi^2_{1-\beta}(m)/\lambda_{\max;\alpha})$, in case it is larger than $k_*(0)$, is that $\lambda_{\max;\alpha}$ is not readily available and may have to be numerically evaluated. On the other hand, $k_*(0)$ in (4.4) [or in (4.6)] is readily computed.

5. An example and some numerical results. We shall carry out the computational procedure described in the previous section for an example where the models (1.1) and (1.3) are applicable with $r = 1$, that is, the parameter $\boldsymbol{\xi}$ in (1.3) is a scalar, to be denoted by ξ. The data for the example is the gestational age data analyzed in Oman and Wax (1984), Oman (1988) and Mathew and Zha (1996). Here ξ is a scalar representing gestational age (i.e., week of pregnancy). The observations are bivariate, ultrasound measurements on two fetal bone lengths: the femur length (F) and the biparietal diameter (BPD). The model that relates the gestational age ξ to the observation vector $\mathbf{y} = (F, BPD)'$ is

$$(5.1) \qquad \mathbf{y} \sim N(\mathbf{a} + B\mathbf{h}(\xi), \Sigma) \quad \text{where } \mathbf{h}(\xi) = \begin{pmatrix} \xi \\ \xi^2 \end{pmatrix}$$

[see Oman and Wax (1984)]. In (5.1), $\mathbf{a}$ is a 2×1 unknown parameter vector, B is a 2×2 unknown parameter matrix and Σ is an unknown 2×2 positive definite matrix. Thus, we have the model (1.3) with $r = 1$, $m = 2$ and $p = 2$. The parameter space for ξ (in weeks) is the interval $[14, 41]$, and the data analyzed in Oman and Wax (1984) consists of (F, BPD) measurements for 1114 women for whom the value of ξ was precisely known. Let Y be the 2×1114 matrix whose columns are the $(F, BPD)'$ measurements for these women. As in (1.3), we shall use $\mathbf{y}$ to denote the $(F, BPD)'$ measurement for a woman whose gestational age ξ is unknown. Then Y and $\mathbf{y}$ are independent following the models

$$(5.2) \qquad Y \sim N(\mathbf{a}\mathbf{1}'_{1114} + BX, I_{1114} \otimes \Sigma), \qquad \mathbf{y} \sim N(\mathbf{a} + B\mathbf{h}(\xi), \Sigma).$$

The ith column of X in (5.2) is $\mathbf{h}(\xi_i)$, ξ_i being the known gestational age for the ith woman ($i = 1, 2, \ldots, 1114$).

5.1. *Computation of $k(d)$ and the region* (3.7). For the data in Oman and Wax (1984), we have

$$(5.3) \qquad X\left(I_N - \frac{1}{N}\mathbf{1}_N\mathbf{1}'_N\right)X' = \begin{pmatrix} 52877.52 & 2878329 \\ 2878329 & 159145978 \end{pmatrix} = M \text{ (say)} \quad \text{and}$$

$$\bar{\mathbf{x}} = \begin{pmatrix} 28.410233 \\ 854.607720 \end{pmatrix}.$$

Furthermore, from (3.5),

$$(5.4) \qquad \mathbf{h}_1(\xi) = M^{-1/2}(\mathbf{h}(\xi) - \bar{\mathbf{x}}).$$

In our computations, we have chosen $\alpha = 0.05$ and $\beta = 0.05$. $T(\xi)$ is given by (3.3) with $r = 1$, $N - m - p = 1110$ and

$$H(\xi) = \hat{B}\begin{pmatrix} 1 \\ 2\xi \end{pmatrix}.$$

Since $H(\xi)$ is a column vector, $T(\xi)$ becomes

$$(5.5) \qquad T(\xi) = 1110 \times \frac{[(\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\mathbf{h}(\xi) - \bar{\mathbf{x}}))' S^{-1} H(\xi)]^2}{[H(\xi)' S^{-1} H(\xi)]}.$$

For the gestational age data, the matrices $(\hat{\mathbf{a}}, \hat{B})$ and S are given in Oman (1988), page 182. For $d = \mathbf{h}_1(\xi)' \mathbf{h}_1(\xi)$, and $\alpha = \beta = 0.05$, we shall evaluate $k(d)$ satisfying (see (3.6))

$$(5.6) \quad P_{\lambda_1, v}\left[P\left\{ 1110 \times \lambda_1 \chi_1^2\left(1; \left(\frac{1}{1114} + d\right)v\right) \leq k(d) | \lambda_1, v \right\} \geq 0.95 \right] = 0.95,$$

where $v \sim \chi^2(2)$ and $\lambda_1 = (hg_1)^{-1}$ with $h \sim \text{Beta}(1111, 1)$, $g_1 \sim \chi^2(1110)$ and v, h and g_1 are independently distributed. For several values of $\xi \in [14, 41]$, $d = \mathbf{h}_1(\xi)' \mathbf{h}_1(\xi)$ was calculated and $k(d)$ was numerically obtained, as explained in Section 4. The starting value for the computation of $k(0)$ is $k_*(0)$, given in (4.6). Since

$$\frac{N - m - p}{r} \frac{\chi_{1-\beta}^2(r)}{\chi_\alpha^2(N - m - p + r - 1)} = 1110 \times \frac{\chi_{0.95}^2(1)}{\chi_{0.05}^2(1110)}$$

$$= 1110 \times \frac{3.8415}{1032.493} = 4.1299$$

and $F_{(1-\beta)(1-\alpha)}(r, N - m - p) = F_{0.9025}(1, 1110) = 2.7506$, $k_*(0) = 4.1299$. We used this as a starting value for the computation of $k(d)$ for $d = 0.00059$, the smallest value of d that we considered (see Table 1). The computation was carried out as follows. We generated 100,000 pairs of values of (λ_1, v). For a given value of d and $k(d)$ and for a given pair of values of (λ_1, v), we evaluated $P\{1110 \times \lambda_1 \chi_1^2(1; (1/1114 + d)v) \leq k(d) | \lambda_1, v\}$ based on 100,000 simulations. If this probability is at least 0.95, define $i(\lambda_1, v) = 1$, following the notation in Section 4. Otherwise $i(\lambda_1, v) = 0$. The value of $i(\lambda_1, v)$ is computed for each of the 100,000 pairs of values of (λ_1, v). The lhs of (5.6) is the proportion of times $i(\lambda_1, v)$ takes the value one. Here $k(d)$ is determined so as to make this proportion equal to 0.95, approximately. In other words, for computing $k(0.00059)$, we computed the lhs of (5.6) starting with $k_*(0) = 4.1299$. The value of $k_*(0)$ was adjusted suitably and the lhs of (5.6) was evaluated repeatedly, until we arrive at a value of $k(d)$ for which the lhs of (5.6) is 0.95, approximately. The value of $k(d)$ that we finally accepted is such that the lhs of (5.6) is between 0.9490 and 0.9510. The values so obtained are given in Table 1 [the quantity $k_0(d)$ given in Table 1 is explained later]. The value $k_*(0) = 4.1299$ turned out to be a very satisfactory starting value for the

TABLE 1
Values of ξ, d, k(d) satisfying (5.6) and k₀(d) satisfying (5.7)

ξ	d	k(d)	k₀(d)
14	0.01033	4.2572	4.2576
15	0.00747	4.2175	4.2180
16	0.00530	4.1899	4.1911
18	0.00257	4.1610	4.1611
22	0.00096	4.1458	4.1456
26	0.00102	4.1461	4.1461
30	0.00073	4.1435	4.1435
34	0.00059	4.1418	4.1422
38	0.00357	4.1719	4.1716
39	0.00539	4.1939	4.1921
40	0.00784	4.2225	4.2229
41	0.01105	4.2690	4.2684

computation of $k(0.00059)$, since, from Table 1, $k(0.00059) = 4.1418$, which is very close to $k_*(0)$.

A plot of the $(d, k(d))$ values in Table 1 is given in Figure 1. The following function gave a good fit to the plot:

$$(5.7) \qquad k_0(d) = 4.136977 + 8.62437d + 295.4134d^2,$$

where $k_0(d)$ denotes the values on the curve given in (5.7). Figure 1 also gives a plot of $k_0(d)$. The plot in Figure 1 and the values of $k_0(d)$ given in Table 1 show that $k_0(d)$ is a very good approximation to $k(d)$. Thus the region (3.7) is given by

$$(5.8) \qquad \left\{ \xi \colon T(\xi) \leq k_0(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi)) \right\},$$

where $T(\xi)$ is given by (5.5), $\mathbf{h}_1(\xi)$ is given by (5.4) and $k_0(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi))$ is obtained from (5.7).

For a few values of $\mathbf{y} = (F, BPD)'$, the region (5.8) is given in Table 2. These $\mathbf{y}$ values were also considered by Mathew and Zha (1996) for the construction of single-use confidence regions.

We conclude this subsection with the following observations. First, an upper bound for d is $\delta = 0.01105$, which corresponds to $\xi = 41$ (see Table 1). From Table 1, we also have $k(\delta) = 4.2690$. One can certainly use $k(\delta)$, instead of $k_0(d)$, to construct the region (5.8). Such a region is given by $\{ \xi \colon T(\xi) \leq 4.2690 \}$. For the $\mathbf{y}$ values in Table 2, the regions $\{ \xi \colon T(\xi) \leq 4.2690 \}$ are, respectively, (12.978, 16.051), (18.440, 21.982), (23.024, 27.139), (27.530, 32.441) and (31.748, 37.805). As expected, these regions are slightly wider compared to the regions in Table 2 since most of the $k(d)$ and $k_0(d)$ values are somewhat smaller than $k(\delta) = 4.2690$. However, the difference between the above intervals based on $k(\delta)$ and those in Table 2 based on $k_0(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi))$ is rather negligible. The main advantage of using $k(\delta)$ is that one can avoid numerical computation of the other $k(d)$ values. For this example, the use of $k(\delta)$ appears to be quite satisfactory.
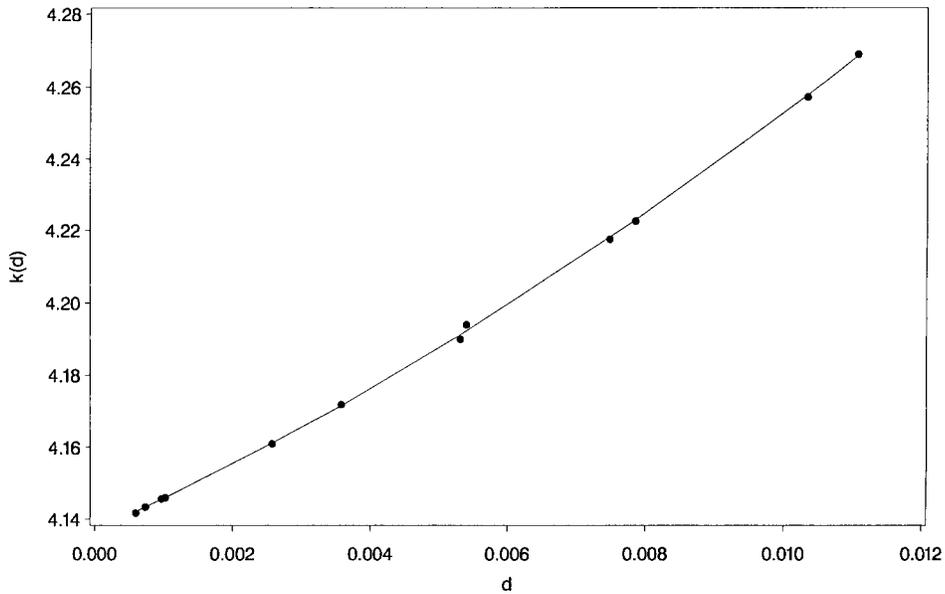
FIG. 1.   *A plot of the $k(d)$ (denoted by the dots) and $k_0(d)$ (the curve) satisfying* (5.6) *and* (5.7), *respectively.*

REMARK 5.1.   As already pointed out, the $k(d)$ values in Table 1 were obtained based on 100,000 simulations of the lhs of (5.6). This is extremely computationally intensive. One such simulation took about 13 hours on a Sun Ultra 1 with 128MB of RAM and 200 MHz UltraSPARC CPU. We used a program written in C for this computation. We also did the same computation using 10,000 simulations instead of 100,000. Each simulation now took only about 10 minutes and the $k(d)$ values agreed with those in Table 1 up to the second decimal place. This appears quite satisfactory to us, for practical purposes, even though we used 100,000 simulations to get a more accurate result.

REMARK 5.2.   Our first interval in Table 2 extends beyond the interval $[14, 41]$, which is the parameter space for ξ. Strictly speaking the region

TABLE 2
The region (5.8) *for ξ using the Oman and Wax* (1984) *data for a few values of* **y**

$\mathbf{y}' = (\boldsymbol{F}, \boldsymbol{BPD})$	Region (5.8)
(14,27)	(12.968, 16.008)
(32,47)	(18.462, 21.955)
(45,62)	(23.052, 27.107)
(56,75)	(27.563, 32.400)
(65,85)	(31.788, 37.763)

should be (5.8) intersected with the parameter space, that is, $\{\xi\colon T(\xi) \leq k_0(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi))\} \cap [14, 41]$. This will obviously bring all the intervals within the parameter space. We note that intersecting with the parameter space will not affect the coverage probability requirements.

5.2. *Conservatism of the region* (5.8).   It should be noted that since $p > 1$ in our example, the region (5.8) will only be conservative. That is, the quantity

$$(5.9) \qquad P_{\bar{\mathbf{y}}, \hat{B}, S}\big[P_{\mathbf{y}}\{T(\xi) \leq k(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi))|\bar{\mathbf{y}}, \hat{B}, S\} \geq 0.95\big]$$

will in general be more than 0.95, even though the lhs of (5.6) is equal to 0.95. In order to study the amount of conservatism of the region (5.8), we need to simulate (5.9) for various parameter values, that is, for various values of B and ξ. Note that we can choose $\Sigma = I_2$ and $\mathbf{a} = 0$, without loss of generality. In this article, we have not reported the simulated values of (5.9); see Sharma (1996) for such numerical results. Our main concern here is to investigate numerically whether the region (5.8), derived using the tolerance region condition (5.6), satisfies the requirements of a multiple-use confidence region, that is, whether it satisfies a condition similar to (2.8).

5.3. *The criterion* (2.8).   For the numerical investigation of (2.8), note that we can choose $\Sigma = I_2$ and $\mathbf{a} = 0$, without loss of generality. In other words, we consider the model $\mathbf{y} \sim N(B\mathbf{h}(\xi), I_2)$. For M given in (5.3), write

$$(5.10) \qquad B_1 = BM^{1/2} \quad \text{and} \quad \hat{B}_1 = \hat{B}M^{1/2}.$$

Thus, under the assumptions $\Sigma = I_2$ and $\mathbf{a} = 0$, we have the models

$$(5.11) \qquad \begin{aligned} \mathbf{y} &\sim N(B_1 M^{-1/2}\mathbf{h}(\xi), I_2), \qquad \bar{\mathbf{y}} \sim N(B_1 M^{-1/2}\bar{\mathbf{x}}, I_2), \\ \hat{B}_1 &\sim N(B_1, I_2 \otimes I_2), \qquad S \sim W(1111, I_2), \end{aligned}$$

where the quantities in (5.11) are also independently distributed. Also, $T(\xi)$ in (5.5) can be written as

$$(5.12) \qquad T(\xi) = 1110 \times \frac{[(\mathbf{y} - \bar{\mathbf{y}} - \hat{B}_1 \mathbf{h}_1(\xi))'S^{-1}H(\xi)]^2}{[H(\xi)'S^{-1}H(\xi)]}.$$

In the simulations that follow, we have used the models in (5.11) with $B_1 = \eta I_2$ for various values of η and ξ.

Following (2.7), let

$$(5.13) \qquad C(\xi; \bar{\mathbf{y}}, \hat{B}_1, S) = P_{\mathbf{y}}\big[T(\xi) \leq k_0(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi))|\bar{\mathbf{y}}, \hat{B}_1, S\big],$$

where $k_0(\mathbf{h}_1(\xi)'\mathbf{h}_1(\xi))$ is the function given in (5.7). Similar to (2.8), the condition to be satisfied by multiple-use confidence regions is

$$(5.14) \qquad P_{\bar{\mathbf{y}}, \hat{B}_1, S}\bigg[\frac{1}{n}\sum_{i=1}^{n} C(\xi_i; \bar{\mathbf{y}}, \hat{B}_1, S) \geq 1 - \beta\bigg] \geq 1 - \alpha,$$

for every sequence $\{\xi_i\}$, $i = 1, 2, \ldots, n$. We shall now investigate numerically whether the confidence region (5.8), derived subject to (5.6), satisfies (5.14)

for $1 - \alpha = 1 - \beta = 0.95$. It is clearly not possible to do this for all possible sequences $\{\xi_i\}$. We have considered only two choices. In view of (3.4) and (3.6), we note that (5.14) will hold if the ξ_i's are all equal in the sequence $\{\xi_i\}$. Consequently, (5.14) is expected to hold when the ξ_i's are nearly equal. Thus it is natural to investigate whether (5.14) will hold when the ξ_i's are as unequal as possible. In any finite sequence, the variance among the ξ_i's is maximum, that is, the ξ_i's are as unequal as possible, when half of the ξ_i's take the minimum value and the remaining half take the maximum value, which are 14 weeks, and 41 weeks, respectively. Thus, consider a sequence where half of the ξ_i's are equal to 14 and the remaining half are equal to 41. In this case, the lhs of (5.14) becomes

$$(5.15) \qquad P_{\bar{\mathbf{y}}, \hat{B}_1, S}\left[\tfrac{1}{2}\left\{C(14; \bar{\mathbf{y}}, \hat{B}_1, S) + C(41; \bar{\mathbf{y}}, \hat{B}_1, S)\right\} \geq 0.95\right].$$

Table 3 gives the simulated values of (5.15) for the models in (5.11) with $B_1 = \eta I_2$ for various values of η. The results are based on 10,000 simulations, carried out as follows. For a given value of η, we generated the values of $\bar{\mathbf{y}}, \hat{B}_1$ and S, based on the model (5.11), and simulated $C(14; \bar{\mathbf{y}}, \hat{B}_1, S)$ and $C(41; \bar{\mathbf{y}}, \hat{B}_1, S)$ based on 10,000 simulations. For this simulation, we used the definition of $C(\xi; \bar{\mathbf{y}}, \hat{B}_1, S)$ given in (5.13). If $\tfrac{1}{2}\{C(14; \bar{\mathbf{y}}, \hat{B}_1, S) + C(41; \bar{\mathbf{y}}, \hat{B}_1, S)\}$ is more than or equal to 0.95, define $i(\bar{\mathbf{y}}, \hat{B}_1, S)$ to be one, and zero otherwise. We then generated 10,000 sets of values of $(\bar{\mathbf{y}}, \hat{B}_1, S)$ and computed the proportion of times $i(\bar{\mathbf{y}}, \hat{B}_1, S)$ assumed the value one. This proportion is the simulated value of (5.15).

Yet another sequence $\{\xi_i\}$ that we have considered is the following. There are 28 integer values of ξ in the interval $[14, 41]$. Let these be

$$(5.16) \qquad \xi_1 = 14, \qquad \xi_2 = 15, \qquad \xi_3 = 16, \ldots, \xi_{28} = 41.$$

Consider a sequence of ξ values where the above 28 values appear equally often. When this is the case, the lhs of (5.14) becomes

$$(5.17) \qquad P_{\bar{\mathbf{y}}, \hat{B}_1, S}\left[\frac{1}{28}\sum_{i=1}^{28} C(\xi_i; \bar{\mathbf{y}}, \hat{B}_1, S) \geq 0.95\right],$$

where the ξ_i's are as given in (5.16), for $i = 1, 2, \ldots, 28$. Table 4 gives the simulated values of (5.17) for the models in (5.11) with $B_1 = \eta I_2$ for various values of η. The simulation of (5.17) was carried out similar to that of (5.15). However, the results are based on only 2500 simulations. We used fewer simulations since, in order to compute (5.17), we need to simulate the 28 quantities

TABLE 3
Simulated values of (5.15) *for the model* (5.11) *based on* 10,000 *simulations with* $B_1 = \eta I_2$ *for various values of* η

η	0.001	0.01	1	10	100
Simulated value of (5.15)	0.9664	0.9664	0.9696	0.9777	0.9774

TABLE 4

Simulated values of (5.17) *for the model* (5.11) *based on* 2500 *simulations with* $B_1 = \eta I_2$ *for various values of* η

η	0.001	0.01	1	10	100
Simulated value of (5.17)	0.9736	0.9736	0.9740	0.9888	0.9912

$C(\xi_i; \bar{\mathbf{y}}, \hat{B}_1, S)$ $(i = 1, 2, \ldots, 28)$, as opposed to only two such terms in (5.15). That is, the computation of (5.17) requires considerably more computational time than the computation of (5.15). Consequently, we decided to compute (5.17) using only 2500 simulations.

The simulation results in Table 3 and Table 4 show that (5.14) holds at least for the parameter values considered for simulation. In other words, the indication is that confidence regions derived using the tolerance region condition (2.6) do satisfy the requirements of multiple-use confidence regions. Clearly, more extensive simulation is necessary before we can draw a firm conclusion regarding this.

REMARK 5.3. The computations for the above example were carried out as follows. The values of $k(d)$ in Table 1 was computed using a program written in C, as pointed out in Remark 5.1. Figure 1 was obtained using S-Plus. The regions in Table 2 were obtained using MATLAB and the numerical results reported in Tables 3 and 4 were obtained using SAS-IML.

6. Concluding remarks. In this article, we have derived a region for the parameter $\boldsymbol{\theta}$ in the model (1.2) or the parameter $\boldsymbol{\xi}$ in the model (1.3), using a tolerance region condition. The results are applicable to finite samples. For some selected parameter values, we have also investigated numerically whether our region satisfies the coverage probability requirements of a multiple-use confidence region. The numerical results indicate that this is the case. Establishing this fact theoretically appears quite difficult. Hence, extensive numerical results are necessary before one can draw the above conclusion. We believe that this article provides the necessary framework and theoretical tools for the further investigation of this problem.

The implementation of our procedure is computationally involved, since the functional form of $k(d)$ satisfying (2.12), or (3.6), has to be numerically evaluated. One alternative to reduce the computational burden is to replace d by an upper bound, say δ, and compute the region using $k(\delta)$ instead of the function $k(d)$. This will of course result in a region having a larger volume, since $k(d) \leq k(\delta)$. However, if $k(\delta)$ is not too large compared to $k_*(0)$ in (4.4) [or (4.6)], the increase in volume of the region based on $k(\delta)$ will not be that significant. This is indeed the case for the example in Section 5. Another alternative is to explore the possibility of approximating $k(d)$ by a suitable function, say a linear function, and then to determine numerically the coefficients in the function so that the lhs of (2.12) [or (3.6)] is at least $1 - \alpha$. Such

an approximation may also result in a region having a larger volume since the actual function $k(d)$ satisfies (2.12) [or (3.6)] with equality, that is, the lhs of (2.12) and (3.6) are equal to $1 - \alpha$. It should be remembered that in situations that call for the use of multiple-use confidence regions, the same $k(d)$ will be used repeatedly in order to construct a sequence of confidence regions. Consequently, if our regions are to be used as multiple-use confidence regions, it may be worthwhile to do the numerical work and compute the functional form of $k(d)$, following the procedure described in Section 4 of this article. In other words, in a given set-up of the calibration problem, the computation of $k(d)$ has to be done only once. It can then be used repeatedly for the construction of multiple-use confidence regions. Nevertheless, in order to save on computations, it is highly desirable to have good approximations for $k(d)$. We hope that this article will stimulate research in this direction.

APPENDIX

In this Appendix, we shall give the proofs of Lemma 2.1 and Lemma 2.2. In order to prove Lemma 2.1, we shall use the following result.

LEMMA A.1.   *Let $\chi_i^2(s_i; \eta_i)$ denote independent noncentral chi-square random variables with degrees of freedom s_i and noncentrality parameters η_i and let $\chi_i^2(s_i)$ denote independent central chi-square random variables with degrees of freedom s_i $(i = 1, 2, \ldots, m)$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$ be nonnegative scalars. Then $\sum_{i=1}^{m} \lambda_i \chi_i^2(s_i; \eta_i)$ is stochastically smaller than $\lambda_1 \chi_1^2(s_1; \sum_{i=1}^{m} \eta_i) + \sum_{i=2}^{m} \lambda_i \chi_i^2(s_i)$.*

The proof of Lemma A.1 is omitted since it is a special case of more general results in Mathew and Nordström (1997).

PROOF OF LEMMA 2.1.   Recall that we are assuming $\Sigma = I_p$. Note that given $\bar{\mathbf{y}}$, $\hat{B}$ and S,

(A.1) $\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}}) \sim N\big[(B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}}) - (\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}), I_p\big].$

Let R be a $p \times p$ orthogonal matrix and Λ be a $p \times p$ diagonal matrix such that

(A.2) $S^{-1}\hat{B}(\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1} = R\Lambda R'.$

Write

(A.3)
$$R = (\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_p) \quad \text{and}$$
$$\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m, 0, \ldots, 0), \qquad \lambda_1 > \lambda_2 > \cdots > \lambda_m.$$

The $\mathbf{r}_i$'s $(i = 1, 2, \ldots, p)$ are the eigenvectors and the λ_i's $(i = 1, 2, \ldots, m)$ are the nonzero eigenvalues of the matrix on the lhs of (A.2). Note from (A.2) that R depends only on $\hat{B}$ and S and hence is distributed independently of $\mathbf{y}$.

Consequently, given $\bar{\mathbf{y}}$, $\hat{B}$ and S, it follows from (A.1) and the orthogonality of R that

(A.4) $R'\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\big] \sim N\big[R'(B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}}) - R'(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}), I_p\big].$

From (A.2), (A.3) and (A.4), we get

$$T(\boldsymbol{\theta}) = \frac{N - m - p}{m} \sum_{i=1}^{m} \lambda_i \big\{\mathbf{r}_i'\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]\big\}^2.$$

Furthermore, given $\bar{\mathbf{y}}$, $\hat{B}$ and S,

$$\big\{\mathbf{r}_i'\big[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]\big\}^2 \sim \chi_i^2(1; \eta_i),$$

where, for $i = 1, 2, \ldots, m$, $\chi_i^2(1; \eta_i)$ denote independent noncentral chi-square random variables with one d.f. each, noncentrality parameters η_i and

$$\eta_i = \big\{\mathbf{r}_i'\big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]\big\}^2.$$

Thus, conditionally given $\bar{\mathbf{y}}$, $\hat{B}$ and S, we have the representation

$$T(\boldsymbol{\theta}) = \frac{N - m - p}{m} \sum_{i=1}^{m} \lambda_i \chi_i^2(1; \eta_i).$$

Hence, for any $k > 0$,

(A.5)
$$P\big[T(\boldsymbol{\theta}) \le k \big| \bar{\mathbf{y}}, \hat{B}, S\big]$$
$$= P\bigg[\frac{N - m - p}{m} \sum_{i=1}^{m} \lambda_i \chi_i^2(1; \eta_i) \le k \big| \boldsymbol{\lambda}, \eta_1, \ldots, \eta_m\bigg],$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_m)'$. Since $\lambda_1 > \lambda_2 > \cdots > \lambda_m$, we use Lemma A.1 to conclude that

(A.6)
$$P\bigg[\frac{N - m - p}{m} \sum_{i=1}^{m} \lambda_i \chi_i^2(1; \eta_i) \le k \big| \boldsymbol{\lambda}, \eta_1, \ldots, \eta_m\bigg]$$
$$\ge P\bigg[\frac{N - m - p}{m} \bigg\{\lambda_1 \chi_1^2\bigg(1; \sum_{i=1}^{m} \eta_i\bigg) + \sum_{i=2}^{m} \lambda_i \chi_i^2(1)\bigg\} \le k \big| \boldsymbol{\lambda}, \eta_1, \ldots, \eta_m\bigg].$$

Note that

(A.7)
$$\sum_{i=1}^{m} \eta_i = \big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]'$$
$$\times \sum_{i=1}^{m} \mathbf{r}_i \mathbf{r}_i' \big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]$$
$$\le \big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big]'\big[(\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}})\big].$$
$$= \bigg(\frac{1}{N} + \boldsymbol{\theta}_1'\boldsymbol{\theta}_1\bigg)^{-1} v,$$

where $\boldsymbol{\theta}_1$ and v are given in (2.10). The inequality (2.11) now follows, in view of (A.5), (A.6) and (A.7). Finally, when $p = m$, $\hat{B}$ is an $m \times m$ nonsingular matrix (with probability one) and hence we have the simplification $T(\boldsymbol{\theta}_1) = ((N - 2m)/m)[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})]'S^{-1}[\mathbf{y} - \bar{\mathbf{y}} - \hat{B}(\boldsymbol{\theta} - \bar{\mathbf{x}})]$. Furthermore, in the special case when $p = m$, the lhs of (A.2) is S^{-1} and the λ_i's are the eigenvalues of S^{-1}. The proof of Lemma 2.1 is thus complete. $\square$

We shall now prove Lemma 2.2. For this, we need the definitions of the matrix variate beta and F distributions and some related basic results. These definitions and the results given in (A.8) are taken from Eaton [(1983), pages 331 and 332].

Let $S_1 \sim W_p(q_1, I_p)$ and $S_2 \sim W_p(q_2, I_p)$ be independently distributed. Then:

(A.8)
1. If $q_1 + q_2 \geq p$, $(S_1 + S_2)^{-1/2}S_1(S_1 + S_2)^{-1/2} \sim B(q_1, q_2, I_p)$, the matrix variate beta distribution with parameters q_1, q_2.
2. If $q_2 \geq p$, $S_2^{-1/2}S_1 S_2^{-1/2} \sim F(q_1, q_2 - p + 1, I_p)$, the matrix variate F distribution with parameters q_1 and $q_2 - p + 1$.
3. If U is an $r \times p$ matrix ($r \leq p$) distributed independently of S_1 as $U \sim N(0, I_r \otimes I_p)$ and if $q_1 \geq p$, $US_1^{-1}U' \sim F(p, q_1 - p + 1, I_r)$.
4. If $F \sim F(s_1, s_2, I_p)$ then $(I + F)^{-1} \sim B(p + s_2 - 1, s_1, I_p)$.

The following result will also be used for proving Lemma 2.2.

LEMMA A.2.   *Let* $S \sim W_p(q, I_p)$, $q \geq p$. *Let* Z *be a random* $p \times r$ *matrix* ($r < p$) *distributed independently of* S *and satisfying* $Z'Z = I_r$ *(with probability one). Then*:

(i) *The distribution of* $(Z'S^{-2}Z)^{-1}(Z'S^{-1}Z)$ *is the same as the distribution of* $G_1(I + UG_2^{-1}U')^{-1}$, *where* $G_1 \sim W_r(q - p + r, I_r)$, $G_2 \sim W_{p-r}(q, I_{p-r})$, $U \sim N(0, I_r \otimes I_{p-r})$, *and* G_1, G_2, *and* U *are independently distributed*.

(ii) *If* $r \leq p/2$, *the distribution of* $(Z'S^{-2}Z)^{-1}Z'S^{-1}Z$ *is the same as that of* G_1H, *where* G_1 *is as defined above*, $H \sim B(q - p + 2r, p - r, I_r)$ *and* G_1 *and* H *are independently distributed*.

PROOF.   Let Z_1 be a $p \times (p - r)$ matrix such that $Z_0 = (Z : Z_1)$ is a $p \times p$ random orthogonal matrix. Then $Z_0'SZ_0 \sim W_p(q, I_p)$. Using the orthogonality of Z_0, we also get $(Z_0'SZ_0)^{-1} = Z_0'S^{-1}Z_0$ and $(Z_0'SZ_0)^{-2} = Z_0'S^{-2}Z_0$. Thus $Z'S^{-1}Z$, which is the top left-hand corner $r \times r$ submatrix of $Z_0'S^{-1}Z_0$, is also the top left-hand corner $r \times r$ submatrix of the inverse of a $W_p(q, I_p)$ matrix. Similarly, $Z'S^{-2}Z$ is the top left-hand corner $r \times r$ submatrix of the square of the inverse of the same Wishart matrix. In other words, in order to prove Lemma A.2, we can assume $Z = (I_r : 0)'$. Now partition S as $S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$, where S_{11} is $r \times r$, S_{12} is $r \times (p - r)$, S_{22} is $(p - r) \times (p - r)$ and $S_{21} = S_{12}'$. Using the expression for the inverse of a partitioned matrix, we conclude the following: the top left-hand corner $r \times r$ submatrix of S^{-1} is $S_{11.2}^{-1}$, where

$S_{11.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}$. Furthermore, writing $S^{-2} = S^{-1}S^{-1}$ and using the expression for S^{-1} in the partitioned form, the top left-hand corner $r \times r$ submatrix of S^{-2} is seen to be $S_{11.2}^{-2} + S_{11.2}^{-1}S_{12}S_{22}^{-2}S_{21}S_{11.2}^{-1}$. Consequently, with $Z = (I_r : 0)'$, the distribution of $(Z'S^{-2}Z)^{-1}Z'S^{-1}Z$ mentioned in Lemma A.2 is the same as the distribution of the quantity

$$(\text{A.8}) \qquad \left[S_{11.2}^{-2} + S_{11.2}^{-1}S_{12}S_{22}^{-2}S_{21}S_{11.2}^{-1}\right]^{-1}S_{11.2}^{-1} = S_{11.2}\left[I_r + S_{12}S_{22}^{-2}S_{21}\right]^{-1}.$$

We shall now use the following distributional results [see Muirhead (1982), Theorem 3.2.10]: (i) $S_{22} \sim W_{p-r}(q, I_{p-r})$, (ii) $S_{11.2} \sim W_r(q - p + r, I_r)$ and is distributed independently of S_{12} and S_{22} and (iii) $S_{12}S_{22}^{-1/2} \sim N(0, I_r \otimes I_{p-r})$ and is distributed independently of S_{22}. Taking $G_1 = S_{11.2}$, $G_2 = S_{22}$ and $U = S_{12}S_{22}^{-1/2}$, part (i) of Lemma A.2 follows from the representation (A.9). Part (ii) follows by noting that when $r \leq p - r$, that is, when $r \leq p/2$, $UG_2^{-1}U' \sim F(p - r, q - p + r + 1, I_r)$ [using (iii) in (A.8)], and hence $H = (I + UG_2^{-1}U')^{-1} \sim B(q - p + 2r, p - r, I_r)$, using (iv) in (A.8)]. This completes the proof of Lemma A.2. □

PROOF OF LEMMA 2.2. Recall that $\boldsymbol{\lambda}$ is the vector of ordered nonzero eigenvalues of the matrix $S^{-1}\hat{B}(\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1}$. Note that

$$S^{-1}\hat{B}(\hat{B}'S^{-1}\hat{B})^{-1}\hat{B}'S^{-1} = S^{-1}Z(Z'S^{-1}Z)^{-1}Z'S^{-1},$$

where $Z = \hat{B}(\hat{B}'\hat{B})^{-1/2}$. Since $Z'Z = I_m$ and since the nonzero eigenvalues of $S^{-1}Z(Z'S^{-1}Z)^{-1}Z'S^{-1}$ are the same as those of $(Z'S^{-1}Z)^{-1}Z'S^{-2}Z$, parts (i) and (ii) of Lemma 2.2 follows from Lemma A.2. In order to prove (iii), note that since $\bar{\mathbf{y}} \sim N(\mathbf{a} + B\bar{\mathbf{x}}, (1/N)I_p)$ and $\hat{B} \sim N[B, \{X(I_N - (1/N)\mathbf{1}_N\mathbf{1}_N')X'\}^{-1} \otimes I_p]$, we immediately get

$$(\text{A.9}) \qquad (\bar{\mathbf{y}} - \mathbf{a} - B\bar{\mathbf{x}}) - (B - \hat{B})(\boldsymbol{\theta} - \bar{\mathbf{x}}) \sim N\left[0, \left(\frac{1}{N} + \boldsymbol{\theta}_1'\boldsymbol{\theta}_1\right)I_p\right],$$

where $\boldsymbol{\theta}_1$ is given in (2.10). The definition of v in (2.10) along with (A.10) shows that $v \sim \chi^2(p)$. Finally, we need to show that v and $\boldsymbol{\lambda}$ are independently distributed. From the proof of (i) given above, it follows that $\boldsymbol{\lambda}$ is distributed independently of $\hat{B}$. Furthermore, S is distributed independently of $\hat{B}$ and $\bar{\mathbf{y}}$. Using the definition of v, we now conclude that v and $\boldsymbol{\lambda}$ are independently distributed. This concludes the proof of Lemma 2.2. □

REFERENCES

BROWN, P. J. (1982). Multivariate calibration (with discussion). *J. Roy. Statist. Soc. Ser. B* **44** 287–321.

BROWN, P. J. (1993). *Measurement, Regression, and Calibration*. Oxford Univ. Press.

BROWN, P. J. and SUNDBERG, R. (1987). Confidence and conflict in multivariate calibration. *J. Roy. Statist. Soc. Ser. B* **49** 46–57.

CARROLL, R. J., SPIEGELMAN, C. H. and SACKS, J. (1988). A quick and easy multiple-use calibration-curve procedure. *Technometrics* **30** 137–141.

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

EATON, M. L. (1983). *Multivariate Statistics, A Vector Space Approach*. Wiley, New York.

EBERHARDT, K. R. and MEE, R. W. (1994). Constant-width calibration intervals for linear regression. *J. Quality Technology* **26** 21–29.

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

LEE, Y.-T. and MATHEW, T. (1998). Tolerance intervals and calibration in linear regression. Unpublished manuscript.

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

MATHEW, T. and NORDSTRÖM, K. (1997). Inequalities for the probability content of a rotated ellipse and related stochastic domination results. *Ann. Appl. Probab.* **7** 1106–1117.

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. Amer. Statist. Assoc.* **92** 1141–1150.

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.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York.

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

OMAN, S. D. and WAX, Y. (1984). Estimating fetal age by ultrasound measurements: an example of multivariate calibration. *Biometrics* **40** 947–960.

OSBORNE, C. (1991). Statistical calibration: a review. *Internat. Statist. Rev.* **59** 309–336.

SCHEFFÉ, H. (1973). A statistical theory of calibration. *Ann. Statist.* **1** 1–37.

SHARMA, M. K. (1996). Multiple use and simultaneous confidence regions in calibration. Ph.D. dissertation, Univ. Maryland, Baltimore County.

SUNDBERG, R. (1994). Most modern calibration is multivariate. In *Seventeenth International Biometric Conference, Hamilton, Canada* 395–405.

T. MATHEW
DEPARTMENT OF MATHEMATICS
  AND STATISTICS
UNIVERSITY OF MARYLAND
1000 HILLTOP CIRCLE
BALTIMORE, MARYLAND 21250
E-MAIL: mathew@math.umbc.edu

M. K. SHARMA
LIFESCAN, INC.
1000 SOUTH GIBRALTAR AVENUE
MILPITAS, CALIFORNIA 95035
E-MAIL: msharma@lfsus.jnj.com

K. NORDSTRÖM
DEPARTMENT OF STATISTICS
UNIVERSITY OF HELSINKI
P.O BOX 54
FIN–000014 HELSINKI
FINLAND
E-MAIL: nordstro@noppa.helsinki.fi