running title: Multivariate Confidence Tubes

Simultaneous Confidence Tubes in Multivariate Linear Regression

W. Liu[a], Y. Han[b], F. Wan[c], F. Bretz[d,e], A. J. Hayter[f]

[a]S3RI and School of Mathematics

University of Southampton, UK

[b]University of Exeter Medical School, Exeter, UK

[c]Department of Mathematics and Statistics, Lancaster University, UK

[d]Novartis Pharma AG, Basel, Switzerland

[e]Shanghai University of Finance and Economics, Shanghai, China

[f]Dept of Statistics and Operations Technology

University of Denver, USA

Abstract

Simultaneous confidence bands have been shown in the statistical literature as powerful inferential tools in univariate linear regression. While the methodology of simultaneous confidence bands for univariate linear regression has been extensively researched and well developed, no published work seems available for multivariate linear regression. This paper fills this gap by studying one particular simultaneous confidence band for multivariate linear regression. Due to the shape of the band, the word 'tube' is more pertinent and so will be used to replace the word 'band'. It is shown that the construction of the tube is related to the distribution of the largest eigenvalue. A simulation-based method is proposed to compute the $1 - \alpha$ quantile of this eigenvalue. With the computation power of modern computers the simultaneous confidence tube can be computed fast and accurately. A real data example is used to illustrate the method and many potential research problems have been pointed out.

Keywords: multivariate linear regression; multivariate normal distribution; simultaneous confidence band; simultaneous confidence tube; statistical inference; statistical simulation; Wishart distribution.

1   Introduction

Consider the multivariate linear regression model in which $\mathbf{x}_1, \cdots, \mathbf{x}_N$ are a set of N independent p-dimensional observations, with $\mathbf{x}_i$ having the normal distribution $\boldsymbol{N}(\boldsymbol{B}\mathbf{z}_i, \Sigma)$. Here the m-dimensional vectors $\mathbf{z}_i = (1, z_{1i}, \cdots, z_{(m-1)i})'$ are known covariate values, the

unknown $p \times p$ matrix Σ is the covariance matrix of $\mathbf{x}_i$, and the unknown $p \times m$ matrix $\boldsymbol{B} = (\boldsymbol{\beta}_1, \cdots, \boldsymbol{\beta}_p)'$ specifies that the response vector $\mathbf{x}_i$ depends on the covariate vector $\mathbf{z}_i$ via $\mathrm{E}(\mathbf{x}_i) = \boldsymbol{B}\mathbf{z}_i$. This multivariate linear regression model can also be represented as

$$\mathbf{x} = \boldsymbol{B}\mathbf{z} + \boldsymbol{\epsilon} \tag{1}$$

where $\mathbf{x} = (x_1, \cdots, x_p)'$, $\mathbf{z} = (1, z_1, \cdots, z_{(m-1)})'$ and $\boldsymbol{\epsilon}$ are independent $\boldsymbol{N}(\mathbf{0}, \Sigma)$ errors.

Without loss of generality, assume that $\Sigma = (\sigma_{ij})$ is non-singular and $\boldsymbol{Z} = (\mathbf{z}_1, \cdots, \mathbf{z}_N)$ is of full row-rank. Based on the N observations $(\mathbf{x}_1, \mathbf{z}_1), \cdots, (\mathbf{x}_N, \mathbf{z}_N)$ the maximum likelihood estimators of $\boldsymbol{B}$ and Σ are given by $\hat{\boldsymbol{B}} = (\hat{\boldsymbol{\beta}}_1, \cdots, \hat{\boldsymbol{\beta}}_p)' = \boldsymbol{C}\boldsymbol{A}^{-1}$ and $\hat{\Sigma} = \sum_{i=1}^{N}(\mathbf{x}_i - \hat{\boldsymbol{B}}\mathbf{z}_i)(\mathbf{x}_i - \hat{\boldsymbol{B}}\mathbf{z}_i)'/N$, where $\boldsymbol{A} = \sum_{i=1}^{N} \mathbf{z}_i \mathbf{z}_i'$ and $\boldsymbol{C} = \sum_{i=1}^{N} \mathbf{x}_i \mathbf{z}_i'$. Furthermore, assuming $N \geq p + m$ throughout this paper, then we have the following distributional results:

$$
\begin{aligned}
(\hat{\boldsymbol{\beta}}_1', \cdots, \hat{\boldsymbol{\beta}}_p')' &\sim \boldsymbol{N}((\boldsymbol{\beta}_1', \cdots, \boldsymbol{\beta}_p')', \Sigma \otimes \boldsymbol{A}^{-1}) \\
N\hat{\Sigma} &\sim \boldsymbol{W}(\Sigma, n) \quad \text{with} \quad n = N - m, \\
(\hat{\boldsymbol{\beta}}_1', \cdots, \hat{\boldsymbol{\beta}}_p')' &\text{ and } N\hat{\Sigma} \text{ are independent}
\end{aligned}
\tag{2}
$$

where $\boldsymbol{W}(\Sigma, r)$ denotes the Wishart distribution with parameters Σ and r. All these results can be found in the excellent book by Anderson (2003, Section 8.2).

It is clear that the systematic component $\boldsymbol{B}\mathbf{z}$ of model (1) is of interest and can be estimated by $\hat{\boldsymbol{B}}\mathbf{z}$. One can further provide the following exact $1 - \alpha$ confidence set for $\boldsymbol{B}\mathbf{z}$ for a given $\mathbf{z}$

$$\left\{ \frac{(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})'(N\hat{\Sigma})^{-1}(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})}{\mathbf{z}'A^{-1}\mathbf{z}} \leq \frac{p}{n - p + 1} f_{p,n-p+1}^{\alpha} \right\} \tag{3}$$

by noting that $(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})/\sqrt{\mathbf{z}'A^{-1}\mathbf{z}} \sim \boldsymbol{N}(\mathbf{0}, \Sigma)$ and so $(n - p + 1)(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})'(N\hat{\Sigma})^{-1}(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})/ (p\mathbf{z}'A^{-1}\mathbf{z}) \sim F_{p,n-p+1}$ (see e.g. Anderson, 2003, Theorem 5.2.2), where $f_{p,n-p+1}^{\alpha}$ and $F_{p,n-p+1}$

4

denote respectively the upper α-point of, and a random variable having, an F distribution with p and $n - p + 1$ degrees of freedom.

This paper constructs an exact $1 - \alpha$ simultaneous confidence band (SCB) for $\boldsymbol{B}\mathbf{z}$ for all $\tilde{\mathbf{z}} = (z_1, \cdots, z_{m-1})' \in R^{m-1}$. For the special case of simple linear regression (i.e. $p = 1$ and $m = 2$) a solution is given by Working and Hotelling (1929). This result is generalized to multiple linear regression (i.e. $p = 1$ and $m \geq 2$) by Scheffé (1953). The topic of SCBs for $\boldsymbol{B}\mathbf{z}$ in univariate linear regression (i.e. $p = 1$) has generated great interests over the last sixty years since the pioneering work of Working and Hotelling (1929) and Scheffé (1953). Contributions to this topic have been made by numerous authors; see Liu (2010) for a review and the references therein.

To the best of our knowledge, all the published work is confined to univariate regression however. This paper fills this gap by constructing a SCB for $\boldsymbol{B}\mathbf{z}$ for all $\tilde{\mathbf{z}} \in R^{m-1}$ for a general $p \geq 1$. From the confidence set for $\boldsymbol{B}\mathbf{z}$ for a given $\tilde{\mathbf{z}} \in R^{m-1}$ in (3), a natural SCB has the form

$$\left\{ \frac{(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})'(N\hat{\Sigma})^{-1}(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})}{\mathbf{z}'A^{-1}\mathbf{z}} \leq c \ \ \forall \tilde{\mathbf{z}} \in R^{m-1} \right\} \tag{4}$$

where c is a critical constant suitably chosen so that the confidence level is exactly $1 - \alpha$. This SCB is the focus of this paper.

SCB (4) can be plotted in a three dimensional space in the following way for the special case of $p = 2$ (i.e. the response $\mathbf{x}$ has two components $\mathbf{x} = (x_1, x_2)'$) and $m = 2$ (i.e. there is only one covariate $\tilde{\mathbf{z}} = z_1$) and so each x_i depends on the covariate z_1 via a simple linear regression model. SCB (4) consists of one ellipsoidal disc for $\boldsymbol{B}\mathbf{z}$ in the (x_1, x_2)-plane at each $z_1 \in R^1$; see Figure 1 in Section 3. The centres of all the discs form the straightline $((\hat{\boldsymbol{B}}\mathbf{z})', z_1)$ in

the (x_1, x_2, z_1)-space. SCB (4) stipulates, with confidence level $1 - \alpha$, that $((\boldsymbol{B}\mathbf{z})', z_1)$ for all $z_1 \in R^1$, which form a straightline, is contained completely inside all the discs. SCB (4) for general $p \geq 2$ and $m \geq 2$ is a generalization of Figure 1 which one can only imagine in a four or higher dimensional space as with many other multivariate statistical techniques. From Figure 1 of Section 3, simultaneous confidence tube (SCT) seems more pertinent than SCB for multivariate regression and will be used in the rest of the paper.

This paper is organized as follows. Section 2 considers the determination of the critical constant c in (4). Section 3 provides an illustrative example. Finally Section 4 contains some concluding remarks.

2 Determination of the critical constant c

The key in the construction of SCT (4) is the determination of the critical constant c, which satisfies

$$P\left\{ \max_{\tilde{\mathbf{z}} \in R^{m-1}} g(\mathbf{z}) \leq c \right\} = 1 - \alpha \tag{5}$$

where

$$
\begin{aligned}
g(\mathbf{z}) &= \frac{\left((\hat{\boldsymbol{B}} - \boldsymbol{B})\mathbf{z} \right)' (N\hat{\Sigma})^{-1} \left((\hat{\boldsymbol{B}} - \boldsymbol{B})\mathbf{z} \right)}{\mathbf{z}' A^{-1} \mathbf{z}} \\
&= \frac{\left(\boldsymbol{A}^{-1/2}\mathbf{z} \right)' \left(\Sigma^{-1/2}(\hat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{A}^{1/2} \right)' \left(\Sigma^{-1/2} N\hat{\Sigma}\Sigma^{-1/2} \right)^{-1} \left(\Sigma^{-1/2}(\hat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{A}^{1/2} \right) \left(\boldsymbol{A}^{-1/2}\mathbf{z} \right)}{\left(\boldsymbol{A}^{-1/2}\mathbf{z} \right)' \left(\boldsymbol{A}^{-1/2}\mathbf{z} \right)},
\end{aligned}
$$

where $\boldsymbol{M}^{1/2}$ denotes the square-root matrix of a positive-definite matrix $\boldsymbol{M}$, and $\boldsymbol{M}^{-1/2}$ denotes the inverse matrix of $\boldsymbol{M}^{1/2}$.

First we determine the distribution of the $m \times p$ random matrix $\boldsymbol{U} := \left(\Sigma^{-1/2}(\hat{\boldsymbol{B}} - \boldsymbol{B})\boldsymbol{A}^{1/2} \right)' =$

$\boldsymbol{A}^{1/2} \left(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1, \cdots, \hat{\boldsymbol{\beta}}_p - \boldsymbol{\beta}_p \right) \Sigma^{-1/2}$. It is clear that $E(\boldsymbol{U}) = \boldsymbol{0}$ and its $m \times p$ random elements are jointly normally distributed since they are all linear combinations of the jointly normally distributed random elements of $\hat{\boldsymbol{B}} - \boldsymbol{B}$. To find the covariance matrix of the $m \times p$ random elements of $\boldsymbol{U}$, denote $\Sigma^{-1/2} = (\eta_{ij})_{p \times p}$ and $\hat{\boldsymbol{\gamma}}_i := \boldsymbol{A}^{1/2}(\hat{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)$, $i = 1, \cdots, p$. Hence $\boldsymbol{U} = (\hat{\boldsymbol{\gamma}}_1, \cdots, \hat{\boldsymbol{\gamma}}_p)(\eta_{ij}) = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p)$ where $\boldsymbol{u}_j = \sum_{i=1}^p \eta_{ij}\hat{\boldsymbol{\gamma}}_i$, $j = 1, \cdots, p$. Now the distributions of the $\hat{\boldsymbol{\beta}}_i$'s in (2) give directly $\text{Cov}(\hat{\boldsymbol{\gamma}}_i, \hat{\boldsymbol{\gamma}}_j) = \sigma_{ij}\mathbf{I}_m$ for $1 \le i, j \le p$. This and a few lines of simple manipulation show that $\text{Cov}(\boldsymbol{u}_l, \boldsymbol{u}_k) = \delta_{lk}\mathbf{I}_m$ for $1 \le l, k \le p$, where δ_{lk} is equal to one if $l = k$ and zero otherwise. We have therefore shown that $\boldsymbol{u}_1 \cdots, \boldsymbol{u}_p \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\boldsymbol{0}, \mathbf{I}_m)$, i.e. all the $m \times p$ elements of $\boldsymbol{U}$ are independent $N(0, 1)$ random variables.

Next it is straightforward to check that $\boldsymbol{D} := \Sigma^{-1/2}N\hat{\Sigma}\Sigma^{-1/2} = \sum_{i=1}^n \mathbf{v}_i\mathbf{v}_i'$ where $\mathbf{v}_1, \cdots, \mathbf{v}_n \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\boldsymbol{0}, \mathbf{I}_p)$ since $N\hat{\Sigma} \sim \boldsymbol{W}(\Sigma, n)$ from (2). Hence we have shown that

$$g(\mathbf{z}) = \frac{\left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)' \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}' \left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)}{\left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)' \left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)} \tag{6}$$

where $\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\boldsymbol{0}, \mathbf{I}_m)$, $\mathbf{v}_1, \cdots, \mathbf{v}_n \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\boldsymbol{0}, \mathbf{I}_p)$, and $(\boldsymbol{u}_1, \cdots, \boldsymbol{u}_p)$ and $(\mathbf{v}_1, \cdots, \mathbf{v}_n)$ are independent since $\hat{\boldsymbol{B}}$ and $N\hat{\Sigma}$ are independent from (2). It is clear from (6) that the distribution of $g(\mathbf{z})$ does not depend on the unknown parameters $\boldsymbol{B}$ and Σ of model (1).

Now it follows directly from Anderson (2003, Theorem A.2.4) that

$$\max_{\tilde{\mathbf{z}} \in R^{m-1}} g(\mathbf{z}) = \max_{\tilde{\mathbf{z}} \in R^{m-1}} \frac{\left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)' \boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}' \left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)}{\left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)' \left(\boldsymbol{A}^{-1/2}\mathbf{z}\right)} = l_1,$$

the largest eigenvalue of $\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}'$. Note that a non-zero eigenvalue l of $\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}'$ satisfies $|\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}' - l\mathbf{I}_m| = 0$, which is the same as $0 = |\boldsymbol{D}||\boldsymbol{U}\boldsymbol{D}^{-1}\boldsymbol{U}' - l\mathbf{I}_m| = l^{m-p}|l\boldsymbol{D} - \boldsymbol{U}'\boldsymbol{U}|$ since $|\boldsymbol{D}| \ne 0$ with probability one, where the second equality follows directly from Anderson (2003, Theorem A.3.2). Hence l_1 is also the largest solution l of $|\boldsymbol{Q} - l\boldsymbol{D}| = 0$ where $\boldsymbol{Q} = \boldsymbol{U}'\boldsymbol{U} \sim \boldsymbol{W}(\mathbf{I}_p, m)$, $\boldsymbol{D} \sim \boldsymbol{W}(\mathbf{I}_p, n)$, and $\boldsymbol{Q}$ and $\boldsymbol{D}$ are independent.

The distributions of the solutions l of $|\boldsymbol{Q} - l\boldsymbol{D}| = 0$ have been studied by Fisher (1939), Girshick (1939), Hsu (1939), Roy (1939) and Mood (1951) among others, and relevant results are summarized in Anderson (2003, Section 13.2). In particular, for $m \geq p$ (and $n \geq p$ since we have assumed $N \geq p+m$), the joint probability density function (pdf) of all the solutions $0 \leq l_p \leq \cdots \leq l_1$ is given in Theorem 13.2.2, and for $m < p$, the joint pdf of the nonzero solutions $0 < l_m < \cdots < l_1$ is given by Theorem 13.2.3. In theory, for a given c, one can find $P\{\max_{\tilde{\mathbf{z}} \in R^{m-1}} g(\mathbf{z}) \leq c\} = P\{l_1 \leq c\}$ by integrating the joint pdf of the l_i's over the region $\{l_1 \leq c\}$. A standard numerical searching algorithm, such as the bisection method, will then find the required critical constant c. However, high dimensional numerical integration may not be straightforward.

There is also an extensive literature on the distribution of l_1 (or, equivalently, $f_1 = l_1/(l_1+1)$). For, example, for $m \geq p$, Roy (1945, 1957) provides some expressions for $P\{f_1 \leq f\}$ for $p = 2, 3$ and 4; see Anderson (2003, pp.334) and the references therein. Table B.4 of Anderson (2003) provides the quantiles of f_1 base on the approximation by Pillai (1967).

We recommend a simulation method to compute the critical constant c. Note that the required c is just the $100(1 - \alpha)$th percentile of the random variable l_1, and so can be approximated by the sample percentile using simulation in the following way. We simulate a large number R of independent replicates of $l_1 : l_{11}, \cdots, l_{1R}$, and use the $\langle(1 - \alpha)R\rangle$th largest l_{1i} value as c, where $\langle a \rangle$ denotes the integer part of a. It is well known that this approximation approaches c almost surely as R approaches infinity; some methods of assessing the accuracy of this approximation can be found in Edwards and Berry (1987) and Liu et $al.$ (2005). In each simulation of l_1, we first generate $\mathbf{v}_1, \cdots, \mathbf{v}_{m+n} \overset{\text{i.i.d.}}{\sim} \boldsymbol{N}(\mathbf{0}, \mathbf{I}_p)$, then compute $\boldsymbol{Q} = \sum_{i=1}^{m} \mathbf{v}_i\mathbf{v}_i'$ and $\boldsymbol{D} = \sum_{i=m+1}^{m+n} \mathbf{v}_i\mathbf{v}_i'$, and finally solve l_1 from $|\boldsymbol{Q} - l\boldsymbol{D}| = 0$. From our

experience with various configurations of $p \leq 5$, $m \leq 5$ and $n \leq 200$, the computation of c using $R = 1,000,000$ simulations takes only a few seconds on an ordinary PC and the results agree to at least three decimal points with those computed using Roy's (1945, 1957) expression for $m \geq p = 2$ and one dimensional numerical quadrature. More information on computation time and accuracy is provided in Section 3.

Before finishing this section, we show that the theoretical result derived above implies the results of Working and Hotelling (1929) and Scheffé (1953) for the special case of $p = 1$. For $p = 1$, $\boldsymbol{Q}$ and $\boldsymbol{D}$ are independent chi-square random variables χ^2_m and χ^2_n, respectively. Hence the only solution $l = l_1$ of $|\boldsymbol{Q} - l\boldsymbol{D}| = 0$ is $l_1 = \chi^2_m/\chi^2_n$ and so the critical constant c is given by $c = (m/n)f^\alpha_{m,n}$ as in Scheffé (1953).

3   Example

The Matlab software (version R2012a) includes the dataset Flu as an example for fitting multivariate linear regression model. It is used in this section to illustrate the construction of SCT (4) considered in this paper. The dataset has nine response variables x_i which are predicted regional flu estimates based on Google queries in nine US regions including NE, $MidAtl$ and $ENCentral$ etc, and only one covariate $z_1 = WtdILI$ which is the flu prediction of the National Centres for Disease Control and Prevention. It has $N = 52$ observations on the responses and covariate.

Suppose we are only interested in how the first two responses $x_1 = NE$ and $x_2 = MidAtl$ depend on the covariate z_1 in terms of the multivariate linear model (1). Based on the 52

observations, we have computed the estimates

$$\hat{\boldsymbol{B}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}_1' \\ \hat{\boldsymbol{\beta}}_2' \end{pmatrix} = \begin{pmatrix} 0.186 & 0.662 \\ 0.243 & 0.633 \end{pmatrix}, \quad N\hat{\Sigma} = \begin{pmatrix} 1.675 & 1.182 \\ 1.182 & 1.199 \end{pmatrix}$$

with $p = 2$, $m = 2$, $N = 52$ and $n = N - m = 50$. Note that $\hat{\boldsymbol{\beta}}_i$ is the same as the estimate one would get by fitting an individual linear regression model of x_i on z_1, $i = 1, 2$. The two diagonal elements of $N\hat{\Sigma}$ are the residual sums of squares of the two individual linear regression models, respectively. However, the multivariate linear regression model of $\mathbf{x} = (x_1, x_2)'$ on z_1 takes into consideration the possible correlation between x_1 and x_2 and so is more informative than the two individual linear regression models of x_1 and x_2 on z_1.

With $\alpha = 0.05$ and the given values of p, m and n, our Matlab program computed the critical constant c in (4), giving $c = 0.1899$ with $R = 1,000,000$ simulations. This took twelve seconds on an ordinary PC (Intel(R) Core(TM)i5-2400 CPU@3.10GHz 4.00GB). Using the expression for $P\{f_1 \le f\}$ given in Roy (1945, 1957) for $p = 2$ and numerical quadrature, we computed $P\{l_1 \le 0.1899\} = 0.9502$. Also using this expression and numerical quadrature, we computed $c = 0.1897$. These indicate that the critical constant c computed using simulation is very accurate, more than adequate for most applications.

For each given z_1, all the $\boldsymbol{B}\mathbf{z}$ that satisfy $(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})'(N\hat{\Sigma})^{-1}(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z}) \le c\mathbf{z}'A^{-1}\mathbf{z}$ is given by an ellipsoidal disc in the (x_1, x_2)-plane with $(x_1, x_2)' = \boldsymbol{B}\mathbf{z}$. Its centre is $(x_1, x_2)' = \hat{\boldsymbol{B}}\mathbf{z}$, its shape is determined by $(N\hat{\Sigma})^{-1}$, and its size depends on $c\mathbf{z}'A^{-1}\mathbf{z}$. All the centres for $z_1 \in R^1$ form a straightline in the (x_1, x_2, z_1)-space with $(x_1, x_2)' = \hat{\boldsymbol{B}}\mathbf{z}$. This straightline is our estimate of the unknown multivariate regression line $\{((\boldsymbol{B}\mathbf{z})', z_1) : z_1 \in R^1\}$. SCT (4) is the union of all these discs, one at each $z_1 \in R^1$. A collection of such discs and so SCT (4) are plotted in Figure 1 (in red colour), with the centres of the discs being given by the straightline.

10

Figure 1: *The 95% SCT for $p = 2$ and $m = 2$. The observed data points: stars; the fitted regression model: straightline; the SCT: union of all the discs.*
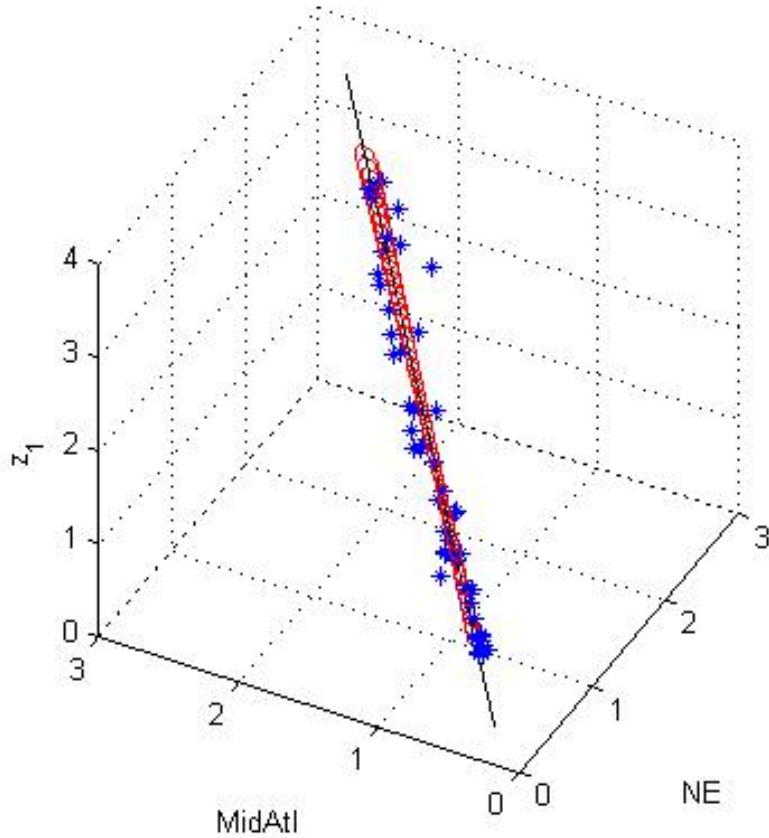
SCT (4) tells us that the unknown multivariate regression line $\{((\boldsymbol{Bz})', z_1) : z_1 \in R^1\}$ lies completely inside the tube, with $1 - \alpha$ confidence. In Matlab, one can view the tube (by using the Figure 1 in Matlab format as the online supplemental document at the journal's website) from different angles to get a better feeling of the tube.

One can project the three dimensional tube into the (x_1, z_1)-plane and the (x_2, z_1)-plane. The projection in the (x_1, z_1)-plane contains the first individual regression line, and the projection in the (x_2, z_1)-plane contains the second individual regression lines, with a simultaneous

11

confidence level of at least $1 - \alpha$. However, the original tube cannot be recovered from, and so contains more information than, the two projections.

Now suppose that we are interested in the multivariate linear regression model of the first three responses $x_1 = NE$, $x_2 = MidAtl$ and $x_3 = ENCentral$ on the covariate z_1. We can compute the estimates

$$\hat{\boldsymbol{B}} = \begin{pmatrix} \hat{\boldsymbol{\beta}}'_1 \\ \hat{\boldsymbol{\beta}}'_2 \\ \hat{\boldsymbol{\beta}}'_3 \end{pmatrix} = \begin{pmatrix} 0.186 & 0.662 \\ 0.243 & 0.633 \\ 0.245 & 0.657 \end{pmatrix}, \quad N\hat{\Sigma} = \begin{pmatrix} 1.675 & 1.182 & 1.603 \\ 1.182 & 1.199 & 1.451 \\ 1.603 & 1.451 & 2.173 \end{pmatrix}$$

with $p = 3$, $m = 2$, $N = 52$ and $n = N - m = 50$. For $\alpha = 0.05$ and $R = 1,000,000$, the critical constant c in (4) was computed as 0.2453 from our Matlab program, which works for general p, m and n. This took fifteen seconds. For both $p = 2$ and 3, we have tried different random seeds in simulations and the critical constant c changes only in the fourth decimal places.

For each given z_1, all the $\boldsymbol{B}\mathbf{z}$ that satisfy $(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z})'(N\hat{\Sigma})^{-1}(\hat{\boldsymbol{B}}\mathbf{z} - \boldsymbol{B}\mathbf{z}) \leq c\mathbf{z}'A^{-1}\mathbf{z}$ is now given by an ellipsoidal ball in the (x_1, x_2, x_3)-space with $(x_1, x_2, x_3)' = \boldsymbol{B}\mathbf{z}$. Its centre is $(x_1, x_2, x_3)' = \hat{\boldsymbol{B}}\mathbf{z}$, and all the centres for $z_1 \in R^1$ form a straightline in the (x_1, x_2, x_3, z_1)-space. SCT (4) is the union of all these balls, one at each $z_1 \in R^1$. Of course this four dimensional SCT still tells us the whereabout of the unknown multivariate regression line, even though the SCT can only be imagined in one's mind. Again, projections of the four dimensional SCT into three or two dimensional space can help us to view the original SCT from some particular angles; but the projections contain less information than the original SCT.

12

4 Conclusions

SCBs have been shown in the statistical literature to be powerful inferential tools in univariate regression, supplementing the standard approaches of estimation and hypotheses testing of the unknown parameters. While the methodology of SCBs for univariate linear regression has been extensively researched and well developed (cf. Liu, 2010), no published work seems available for multivariate linear regression. This paper is a first effort to fill this gap by studying the SCT in (4).

It has been shown that the construction of SCT(4) hinges on the distribution of the largest eigenvalue l_1. While the distribution of l_1 has been studied by many researchers, we have proposed a simulation-based method to compute the $1 - \alpha$ quantile c of l_1. With the computation power of modern computers, this method computes c very fast and accurately. Also, the method works for general values of p, m and n and is easy to understand. A Matlab program is written which allows the SCT to be computed easily.

It is also interesting to observe the following relationship between SCT (4) and Roy's (1953) test when applied for testing $H_0 : \boldsymbol{B} = \boldsymbol{B}^*$, where $\boldsymbol{B}^*$ is given. Roy's test rejects H_0 if and only if $l_1 > c$, which is the same as $\max_{\tilde{\mathbf{z}} \in R^{m-1}} g(\mathbf{z}) > c$, but with $\boldsymbol{B}$ replaced with $\boldsymbol{B}^*$ in $g(\mathbf{z})$, as shown in this paper. The latter means the regression function $\boldsymbol{B}^* \mathbf{z}$ for $\tilde{\mathbf{z}} \in R^{m-1}$ is not contained completely inside SCT (4). Hence Roy's test is just the intuitive test implied by SCT (4): a plausible candidate of the true model $\boldsymbol{B}\mathbf{z}$ for $\tilde{\mathbf{z}} \in R^{m-1}$ is contained completely inside SCT (4) with probability $1 - \alpha$.

This work generalizes the results of Working and Hotelling (1929) and Scheffé (1953). Note,

however, the bulk of the published work on SCBs for univariate linear regression is on SCBs over a restricted covariate region and of various shapes (cf. Liu, 2010). Construction of SCTs for multivariate linear regression in these two directions, in addition to many other problems, warrants further research.

REFERENCES

Anderson, T.W. (2003). *An Introduction to Multivariate Statistical Analysis, 3rd Edition*, Wiley, New York.

Edwards, D. & Berry, J.J. (1987). The efficiency of simulation-based multiple comparisons. *Biometrics*, 43, 913-928.

Fisher, R.A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Annals of Eugenics*, 9, 238-249.

Girshick, M.A. (1939). On the sampling theory of roots of determinantal equations. *Annals of Mathematical Statistics*, 10, 203-224.

Hsu, P.L. (1939). On the distribution of the roots of certain determinantal equations. *Annals of Eugenics*, 9, 250-258.

Liu, W. (2010). *Simultaneous inference in regression*. Chapman and Hall, New York.

Liu, W., Jamshidian, M., Zhang, Y. & J. Donnelly (2005). Simulation-based simultaneous

confidence bands in multiple linear regression with predictor variables constrained in intervals. *Journal of Computational and Graphical Statistics, 14(2), 459-484.*

Mood, A.M. (1951). On the distribution of characteristic roots of normal second-moment matrices. *Annals of Mathematical Statistics*, 22, 266-273.

Pillai, K.C.S. (1967). Upper percentage points of the largest root of a matrix in multivariate analysis. *Biometrika*, 54, 189-194.

Roy, S.N. (1939). p-statistics or some generalizations in analysis of variance approximate to multivariate problems. *Sankhya*, 4, 381-396.

Roy, S.N. (1945). The individual sampling distribution of the maximum, the minimum and any intermediate of the p-statistics on the null-hypothesis. *Sankhya*, 7, 133-158.

Roy, S.N. (1953). On the heuristic method of test construction and its use in multivariate analysis. *Annals of Mathematical Statistics*, 24, 220-238.

Roy, S.N. (1957). *Some Aspects of Multivariate Analysis*, Wiley, New York.

Scheffé, H. (1953). A method for judging all contrasts in analysis of variance. *Biometrika*, 40, 87-104

Working, H. & Hotelling, H. (1929). Applications of the theory of error to the interpretation of trends. *J. Amer. Stat. Assoc.*, 24, 73-85.

Corresponding auther: Wei Liu, School of Maths, University of Southampton, SO17 1BJ, UK, w.liu@maths.soton.ac.uk