Biometrical Journal **46** (2004) 6, 720–730 DOI: 10.1002/bimj.200410072

Exact Confidence Bounds for Comparing Two Regression Lines with a Control Regression Line on a Fixed Interval

Parul Bhargava* and **John D. Spurrier**

Department of Statistics, University of South Carolina, Columbia, SC 29208, USA

Received 26 January 2004, revised 3 September 2004, accepted 21 September 2004

Summary

The problem of finding exact simultaneous confidence bounds for comparing simple linear regression lines for two treatments with a simple linear regression line for the control over a fixed interval is considered. The assumption that errors are iid normal random is considered. It is assumed that the design matrices for the two treatments are equal and the design matrix for the control has the same number of copies of each distinct row of the design matrix for the treatments. The method is based on a pivotal quantity that can be expressed as a function of four t variables. The probability point depends on the size of an angle associated with the interval. We present probability points for various sample sizes and angles.

Key words: Multiple Comparison; Simultaneous Confidence; Probability Point; Multivariate-*t*.

1 Introduction

Suppose we wish to compare the relationship between the predictor and response for $k \geq 3$ groups. The simple linear regression model for the n_i observations from group i is

$$Y_{ij} = \alpha_i + \beta_i x_{ij} + \varepsilon_{ij}\,, \quad \text{for} \quad i = 1, \ldots, k \quad \text{and} \quad j = 1, \ldots, n_i\,. \tag{1}$$

We assume that all error terms are i.i.d. $N(0, \sigma^2)$. Assume there are at least two distinct x_{ij} values for each group and that the mean predictor value for each group is the same. For mathematical convenience, recenter the data, if necessary, such that the mean predictor value is zero.

There is little literature on the multiple comparison of regression models, and all of it deals with two-sided bounds. Royen (1990) used the example of all pairwise comparisons of linear regression models with an unrestricted predictor variable, under the assumption of equal design matrices for each group, as motivation for studying the multivariate studentized range. Spurrier (1999) found exact simultaneous intervals for all contrasts of simple linear regression lines under the same assumptions. Spurrier (2002) found exact simultaneous bounds for all pairwise comparisons of regression lines and for comparisons with a control under certain design matrices restrictions, but with no predictor variable restriction. These bounds are narrower than those in Spurrier (1999).

Liu, Jamshidian, and Zhang (2004) developed a simulation based method for approximating the probability point for simultaneous comparison of regression models. They were able to restrict the method to a specified subset of contrasts and restrict the predictor to a finite interval, resulting in narrower confidence bounds. They also allow for more than one predictor variable and different design matrices for each group. The disadvantage of the method is that the simulation must be repeated for each application.

In this article, we develop exact simultaneous one-sided and two-sided bounds for comparing the simple linear regression models for two groups to a control regression model for $x \in [a, b]$, where a

* Corresponding author: e-mail: parulbhar@yahoo.com

and b are predetermined constants. Denote the control by group 3. The advantage of our results over those of Liu et al. (2004) is that our results are based on exact methods rather than simulation and allow for either one-sided or two-sided bounds. The Liu et al. results have the advantage of allowing for a wider choice of predictor variable values and more than one predictor.

Define the $n_i \times 2$ design matrix $\mathbf{X}_i$ with first column $(1, \ldots, 1)'$ and second column $(x_{i1}, \ldots, x_{\text{in}_i})'$, $i = 1, 2, 3$. Assume $\boldsymbol{X_1} = \boldsymbol{X_2}$ and that each row of $\boldsymbol{X_1}$ appears the same number of times in $\boldsymbol{X_3}$. Let $m = \dfrac{n_1}{(n_1 + n_3)}$. It follows from the assumptions about the design matrices that

$$\frac{\sum_1^{n_1} x_{1j}^2}{\left(\sum_1^{n_1} x_{1j}^2 + \sum_1^{n_3} x_{3j}^2\right)} = m \,. \tag{2}$$

If $\boldsymbol{X_1} = \boldsymbol{X_2} = \boldsymbol{X_3}$, then $m = 1/2$.

Let $\hat{\alpha}_i$ and $\hat{\beta}_i$ denote the least squares estimators of α_i and β_i, $i = 1, 2, 3$, and let $\hat{\sigma}^2$ denote the error mean square with $\nu = 2n_1 + n_3 - 6$ degrees of freedom. The variable $\dfrac{\nu\hat{\sigma}^2}{\sigma^2}$ has a chi-square distribution with ν degrees of freedom and is independent of the least squares estimators.

We discuss algorithms for computing probability points for one-sided bounds for the differences between the treatment group and control group regression models in Section 2 and for two-sided bounds in Section 3. Example data are analyzed in Section 4. Concluding remarks are given in Section 5. Mathematical details are given in the Appendix.

2 Exact One-sided Confidence Bounds

We will discuss an algorithm to compute the probability point c for one-sided lower $(1 - \gamma)$ 100% simultaneous confidence bounds for comparing the regression model in (1) for groups $i = 1, 2$, to the model for group 3, the control. The comparisons of interest are:

$$(\alpha_1 + \beta_1 x) - (\alpha_3 + \beta_3 x) \quad \text{and} \quad (\alpha_2 + \beta_2 x) - (\alpha_3 + \beta_3 x) \quad \text{for all} \quad x \in [a, b] \tag{3}$$

where a and b are predetermined constants.

The model in matrix notation can be written as $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$, where,

$$\boldsymbol{Y} = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, Y_{31}, \ldots, Y_{3n_3})' , \qquad \boldsymbol{X} = \begin{pmatrix} X_1 & 0 & 0 \\ 0 & X_2 & 0 \\ 0 & 0 & X_3 \end{pmatrix} , \tag{4}$$

$$\boldsymbol{\beta} = (\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3)' , \qquad \boldsymbol{\varepsilon} = (\varepsilon_{11}, \ldots, \varepsilon_{1n_1}, \varepsilon_{21}, \ldots, \varepsilon_{2n_2}, \varepsilon_{31}, \ldots, \varepsilon_{3n_3})' .$$

We have that $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \text{Cov}\,(\hat{\boldsymbol{\beta}}))$, where

$$\text{Cov}\,(\hat{\boldsymbol{\beta}}) = \sigma^2 \boldsymbol{V}^{-1} , \quad \text{where} \quad \boldsymbol{V} = \text{diag}\left\{ n_1, \sum_{j=1}^{n_1} x_{1j}^2,\ n_2, \sum_{j=1}^{n_2} x_{2j}^2,\ n_3, \sum_{j=1}^{n_3} x_{3j}^2 \right\} . \tag{5}$$

The mean, variance and covariance for the estimated difference between the regression lines for a treatment and a control are

$$E[(\hat{\alpha}_i - \hat{\alpha}_3) + x(\hat{\beta}_i - \hat{\beta}_3)] = (\alpha_i - \alpha_3) + x(\beta_i - \beta_3) , \quad \text{for} \quad i = 1, 2 \tag{6}$$

$$\text{Var}\,[(\hat{\alpha}_i - \hat{\alpha}_3) + x(\hat{\beta}_i - \hat{\beta}_3)] = \sigma^2 (A_{1x} + A_{3x}) , \quad \text{for} \quad i = 1, 2 \tag{7}$$

$$\text{Cov}\,\{[(\hat{\alpha}_1 - \hat{\alpha}_3) + x(\hat{\beta}_1 - \hat{\beta}_3)], [(\hat{\alpha}_2 - \hat{\alpha}_3) + x(\hat{\beta}_2 - \hat{\beta}_3)]\} = \sigma^2 A_{3x} , \tag{8}$$

where

$$A_{1x} = \frac{1}{n_1} + \frac{x^2}{\sum_{j=1}^{n_1} x_{1j}^2} \quad \text{and} \quad A_{3x} = \frac{1}{n_3} + \frac{x^2}{\sum_{j=1}^{n_3} x_{3j}^2} .$$

As the errors are normally distributed, we get

$$\begin{bmatrix} (\hat\alpha_1-\hat\alpha_3)+x(\hat\beta_1-\hat\beta_3) \\ (\hat\alpha_2-\hat\alpha_3)+x(\hat\beta_2-\hat\beta_3) \end{bmatrix} \sim N\left[\begin{pmatrix} (\alpha_1-\alpha_3)+x(\beta_1-\beta_3) \\ (\alpha_2-\alpha_3)+x(\beta_2-\beta_3) \end{pmatrix}, \sigma^2\begin{pmatrix} A_{1x}+A_{3x} & A_{3x} \\ A_{3x} & A_{1x}+A_{3x} \end{pmatrix}\right]. \tag{9}$$

For fixed i and x, we define for $i = 1, 2$

$$T_{ix} = \frac{[(\hat\alpha_i-\hat\alpha_3)+x(\hat\beta_i-\hat\beta_3)]-[(\alpha_i-\alpha_3)+x(\beta_i-\beta_3)]}{\hat\sigma\sqrt{A_{1x}+A_{3x}}}\,, \tag{10}$$

which have marginal t-distributions with $\nu = 2n_1 + n_3 - 6$ degrees of freedom and the numerators of T_{1x} and T_{2x} have a correlation of $\dfrac{A_{3x}}{(A_{1x}+A_{3x})}$.

For the one-sided lower $(1-\gamma)$ 100% simultaneous confidence bounds, we need to find the probability point c such that

$$P\{T_{ix} \le c,\ \text{for all}\ x \in [a,b],\ i = 1,2\} = P\{\sup\,(T_{ix}) \le c,\ \text{for all}\ x \in [a,b],\ i = 1,2\} = 1-\gamma\,. \tag{11}$$

The resulting lower simultaneous confidence bounds for $(\alpha_i - \alpha_3) + x(\beta_i - \beta_3)$ are

$$(\hat\alpha_i-\hat\alpha_3)+x(\hat\beta_i-\hat\beta_3)-c\hat\sigma\sqrt{A_{1x}+A_{3x}} \quad \text{for} \quad i = 1,2 \quad \text{and all} \quad x \in [a,b]\,. \tag{12}$$

To find the $\sup\limits_{i,x} T_{ix}$, for $i = 1, 2$ and $x \in [a, b]$ define

$$\begin{aligned} T_1 &= \frac{[(\hat\alpha_1-\alpha_1)-(\hat\alpha_3-\alpha_3)]}{\hat\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_3}}} & T_2 &= \frac{[(\hat\alpha_2-\alpha_2)-(\hat\alpha_3-\alpha_3)]}{\hat\sigma\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_3}}} \\ T_3 &= \frac{[(\hat\beta_1-\beta_1)-(\hat\beta_3-\beta_3)]}{\hat\sigma\sqrt{\dfrac{1}{\sum_1^{n_1} x_{1j}^2}+\dfrac{1}{\sum_1^{n_3} x_{3j}^2}}} & T_4 &= \frac{[(\hat\beta_2-\beta_2)-(\hat\beta_3-\beta_3)]}{\hat\sigma\sqrt{\dfrac{1}{\sum_1^{n_1} x_{1j}^2}+\dfrac{1}{\sum_1^{n_3} x_{3j}^2}}} \end{aligned} \tag{13}$$

where T_1, T_2, T_3 and T_4 have a multivariate-t distribution (Graybill (1976), p. 201) with $\nu = 2n_1 + n_3 - 6$ degrees of freedom and dependence matrix

$$R = \begin{pmatrix} 1 & m & 0 & 0 \\ m & 1 & 0 & 0 \\ 0 & 0 & 1 & m \\ 0 & 0 & m & 1 \end{pmatrix}. \tag{14}$$

The joint density of T_1, T_2, T_3 and T_4 is

$$f(t_1,t_2,t_3,t_4) = \frac{\Gamma\left\{\dfrac{(4+\upsilon)}{2}\right\}|R|^{-1/2}}{\upsilon^2\pi^2\Gamma\left(\dfrac{\upsilon}{2}\right)}\left\{1+\left(\frac{1}{\upsilon}\right)t'R^{-1}t\right\}^{-\frac{(4+\upsilon)}{2}}, \tag{15}$$

where $t' = (t_1, t_2, t_3, t_4)$.

Now, we can represent T_{1x} and T_{2x} as functions of T_1, T_2, T_3 and T_4.

$$T_{1x} = \frac{T_1 + T_3 w_x}{(1+w_x^2)^{1/2}} \quad \text{and} \quad T_{2x} = \frac{T_2 + T_4 w_x}{(1+w_x^2)^{1/2}}$$

where,

$$w_x = x\sqrt{h}, \quad \text{with} \quad h = \frac{\dfrac{1}{\sum_1^{n_1} x_{1j}^2} + \dfrac{1}{\sum_1^{n_3} x_{3j}^2}}{\dfrac{1}{n_1} + \dfrac{1}{n_3}} = \frac{n_3}{\sum_1^{n_3} x_{3j}^2}\,. \tag{16}$$

Differentiating T_{1x} with respect to w_x, we get

$$\frac{\mathrm{d}T_{1x}}{\mathrm{d}w_x} = \frac{T_3 - T_1 w_x}{(1+w_x^2)^{3/2}}, \quad \text{and} \quad \frac{d^2 T_{1x}}{\mathrm{d}w_x^2} = \frac{-T_1}{\left\{1 + \left(\dfrac{T_3^2}{T_1^2}\right)\right\}^{3/2}} \quad \text{at} \quad w_x = \frac{T_3}{T_1}\,. \tag{17}$$

If $T_1 > 0, T_{1x}$ attains the maximum $\sqrt{T_1^2 + T_3^2}$ at $w_x = \dfrac{T_3}{T_1}$ provided that $\dfrac{T_3}{T_1} \in [a\sqrt{h}, b\sqrt{h}]$ and otherwise the maximum occurs at either $w_x = b\sqrt{h}$ or $a\sqrt{h}$. If $T_1 < 0$, we see from (17), that T_{1x} attains the minimum at $w_x = \dfrac{T_3}{T_1}$. Thus, the maximum of T_{1x} occurs at $w_x = b\sqrt{h}$ or $a\sqrt{h}$. Hence in the (T_1, T_3) plane, the region described in equation (11) is

$$\begin{aligned} &\sqrt{T_1^2 + T_3^2} \le c\,, \quad \text{for} \quad T_1 > 0 \quad \text{and} \quad \frac{T_3}{T_1} \in [a\sqrt{h},\ b\sqrt{h}]\,; \quad \text{and} \\ &T_1 + T_3 a\sqrt{h} \le c\sqrt{1 + a^2 h} \quad \text{and} \quad T_1 + T_3 b\sqrt{h} \le c\sqrt{1 + b^2 h}\,, \quad \text{otherwise}\,. \end{aligned} \tag{18}$$

The region of interest in the (T_2, T_4) plane is analogous.

Figure 1 shows the region of interest in the (T_1, T_3) plane for the case of $-\infty < a < 0 < b < \infty$. The region for T_2 and T_4 is identical to that for T_1 and T_3.

Define θ to be the angle from the line $T_3 = a\sqrt{h}\,T_1$ to the line $T_3 = b\sqrt{h}\,T_1$. The angle

$$\theta = \cos^{-1}\left(\frac{1 + abh}{\sqrt{1 + a^2 h}\,\sqrt{1 + b^2 h}}\right). \tag{19}$$

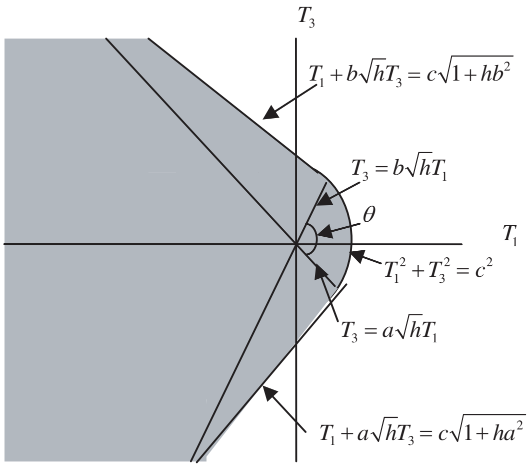

Figure 1 The region of interest in the (T_1, T_3) plane, $-\infty < a < 0 < b < \infty$.

Table 1 Values of c for 95% simultaneous confidence bounds for comparing 2 regression lines to a control for different θ, $m = 1/2$.

		One-sided			Two-sided		
		θ			θ		
n_1	n_3	0.1π	0.5π	π	0.1π	0.5π	π
3	3	3.184	3.922	4.582	4.152	4.915	5.136
4	4	2.499	2.960	3.332	3.040	3.509	3.657
5	5	2.323	2.717	3.021	2.765	3.159	3.289
6	6	2.242	2.607	2.882	2.642	3.002	3.123
7	7	2.196	2.544	2.802	2.572	2.913	3.029
8	8	2.167	2.504	2.751	2.528	2.856	2.969
9	9	2.146	2.476	2.716	2.497	2.816	2.927
10	10	2.131	2.455	2.690	2.474	2.787	2.896
20	20	2.073	2.377	2.592	2.388	2.677	2.780
50	50	2.045	2.339	2.545	2.347	2.625	2.724
∞	∞	2.029	2.317	2.517	2.323	2.594	2.691

It is shown in the Appendix that the probability in equation (11) can be written as

$$\int_{-\infty}^{c} \int_{-\infty}^{c} \int_{l_2}^{\infty} \int_{l_1}^{\infty} f(w_1, w_2, w_3, w_4)\, \mathrm{d}w_1\, \mathrm{d}w_2 \mathrm{d}w_3\, \mathrm{d}w_4$$

where,

$$l_1 = \begin{cases} -\sqrt{c^2 - w_3^2} & \text{if} \quad w_3 > c\cos(\theta) \\ \dfrac{\cos(\theta)\, w_3 - c}{\sin(\theta)} & \text{otherwise} \end{cases} \quad \text{and} \quad l_2 = \begin{cases} -\sqrt{c^2 - w_4^2} & \text{if} \quad w_4 > c\cos(\theta) \\ \dfrac{\cos(\theta)\, w_4 - c}{\sin(\theta)} & \text{otherwise} \end{cases}. \tag{20}$$

We can evaluate (20) by using Gaussian quadrature and then using the secant method to find the critical point c. Table 1 gives the values of c for exact simultaneous one-sided confidence bounds for comparing the regression models for 2 treatments to a control regression model; for $n_1 = n_3$, $\theta = 0.1\pi$, 0.5π, and π, $\gamma = 0.05$. The unrestricted interval $(-\infty, \infty)$ gives $\theta = \pi$. The calculations for $n_1 = n_3 = \infty$ use the 4-variate normal density in place of 4-variate t density. As the limits for W_3 and W_4 are $(-\infty, c)$, the limits were broken into two intervals: $(-20, -2)$ and $(-2, c)$. Also the limit for W_1 is (l_1, ∞), the limit for W_1 considered was $(l_1, 20)$. The same was done for W_2. The probability outside the -20 and $+20$ cutoffs is negligible.

3 Exact Two-sided Confidence Bounds

For the model described above we will find the exact two-sided $(1 - \gamma)$ 100% simultaneous confidence bounds for the restricted case $x \in [a, b]$. Recall that Spurrier (2002) presented probability points for the unrestricted case. To find the two-sided confidence bounds we need to find the probability point c such that

$$P\{|T_{ix}| \le c, \text{ for all } x \in [a, b],\ i = 1, 2\} = P\{\sup (T_{ix}^2) \le c^2, \text{ for all } x \in [a, b],\ i = 1, 2\} = 1 - \gamma\,. \tag{21}$$

The resulting two-sided simultaneous confidence bounds are

$$(\hat{\alpha}_i - \hat{\alpha}_3) + x(\hat{\beta}_i - \hat{\beta}_3) \pm c\hat{\sigma}\sqrt{A_{1x} + A_{3x}} \quad \text{for} \quad i = 1, 2 \quad \text{and all} \quad x \in [a, b]\,. \tag{22}$$

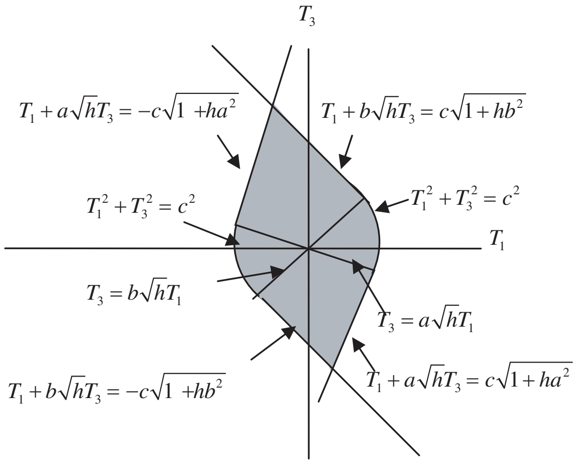

Figure 2 The region of interest for two-sided confidence bounds.

Using equation (17), it follows that T_{ix}^2 is maximized at $w_x = \dfrac{T_3}{T_1}$ if $\dfrac{T_3}{T_1} \in [a\sqrt{h}, b\sqrt{h}]$ and when $T_3/T_1 \notin [a\sqrt{h}, b\sqrt{h}]$ then $T_{ix}^2 \le c^2$ iff

$$-c \le \frac{T_1 + T_3 a\sqrt{h}}{\sqrt{1+a^2h}} \le c \quad \text{and} \quad -c \le \frac{T_1 + T_3 b\sqrt{h}}{\sqrt{1+b^2h}} \le c\,. \tag{23}$$

Hence, the region of interest in the (T_1, T_3) plane is

$$\sqrt{T_1^2 + T_3^2} \le c \quad \text{for} \quad \frac{T_3}{T_1} \in [a\sqrt{h}, b\sqrt{h}]\,;$$

and

$$-c \le \frac{T_1 + T_3 a\sqrt{h}}{\sqrt{1+a^2h}} \le c \quad \text{and} \quad -c \le \frac{T_1 + T_3 b\sqrt{h}}{\sqrt{1+b^2h}} \le c\,, \quad \text{otherwise}\,. \tag{24}$$

The region of interest in the (T_2, T_4) plane is analogous.

Figure 2 shows the region of interest in the (T_1, T_3) plane for the case $-\infty < a < 0 < b < \infty$. Comparing Figure 2 to Figure 1 shows how the regions differ for two-sided and one-sided bounds.

It is shown in the Appendix that the probability in equation (24) can be expressed as

$$\int_{-c}^{c} \int_{-c}^{c} \int_{l_2}^{u_2} \int_{l_1}^{u_1} f(w_1, w_2, w_3, w_4)\, \mathrm{d}w_1\, \mathrm{d}w_2\, \mathrm{d}w_3\, \mathrm{d}w_4$$

where,

$$l_i = \begin{cases} -\sqrt{c^2 - w_{i+2}^2} & \text{if} \quad w_{i+2} > c\cos(\theta) \\ \dfrac{\cos(\theta)\, w_{i+2} - c}{\sin(\theta)} & \text{otherwise} \end{cases}$$

and

$$u_i = \begin{cases} \sqrt{c^2 - w_{i+2}^2} & \text{if} \quad w_{i+2} < -c\cos(\theta) \\ \dfrac{\cos(\theta)\, w_{i+2} + c}{\sin(\theta)} & \text{otherwise} \end{cases} \qquad \text{for} \quad i = 1, 2\,. \tag{25}$$

We can evaluate (25) by using Gaussian quadrature and then using the secant method to find the critical point c. Table 1 gives the values of c for exact simultaneous two-sided confidence bounds for comparing the regression models for 2 treatments to a control regression model, for various values of $n_1 = n_3$, $\theta = 0.1\pi$, 0.5π and π, $\gamma = 0.05$.

4 An Example

Masuda, Saito, and Inui (1997) studied the effects of three treatments (control, presence of indomethacin, presence of ketoprofen) on the accumulation of the chemotherapy agent [^{3}H]methotrexate in rat renal tissue as a function of time. The data suggest a linear relationship between accumulation and time for each treatment in the time interval used in the study. Thus, one can compare the three treatments by performing multiple comparisons of the three regression models.

Three observations were collected at three different time points: 15, 30, and 60 minutes for each treatment. We might be interested in the interval [0, 60] minutes. The average time value for each treatment is 35 minutes. Hence the re-centered predictor variable is $x = \text{time} - 35$ minutes. We have $a = 0 - 35 = -35$ and $b = 60 - 35 = 25$. Also, the value of $h = \frac{n_3}{\sum x_{3j}^2} = \frac{9}{3((-20)^2 + (-5)^2 + (25)^2)} = \frac{3}{1050}$ and $\theta = \cos^{-1}\left(\frac{1 + abh}{\sqrt{1 + a^2h}\,\sqrt{1 + b^2h}}\right) = 0.6392\pi$ radians.

Using the four-dimensional integral method in (20), the 95% simultaneous one-sided confidence bounds for this example are calculated to be

$$(\hat{\alpha}_i - \hat{\alpha}_3) + x(\hat{\beta}_i - \hat{\beta}_3) - 2.555\hat{\sigma}\sqrt{2\left(\frac{1}{9} + \frac{x^2}{3150}\right)}\,. \tag{26}$$

This gives us lower confidence bounds that are 5.93% closer to the difference in the sample regression lines than the unrestricted $(a = -\infty, b = \infty)$ bounds, where the critical point is 2.716. From Table 1 we can see the exact critical value 2.555 is between 2.476 and 2.716.

Using the four-dimensional integral in (25) of two-sided confidence bounds, the critical point for 95% simultaneous confidence intervals for this example is 2.879. Hence the one-sided interval gives us a lower confidence allowance that is 11.25% narrower than the two-sided intervals with the restricted predictor variable.

Recall that Spurrier (2002) found the critical point for the 95% two-sided simultaneous confidence intervals for unrestricted predictor variable. The values for critical points in Table 1 for $\theta = \pi$ match the critical values given in his paper. For the Masuda et al. (1997) example, with $a = -\infty, b = \infty$ Spurrier found the critical point to be 2.927. The critical point with $a = -35$, $b = 25$ gives us a 1.64% narrower two-sided bound. We can see from all the tables that the confidence bounds get narrower as the angle θ gets smaller.

5 Conclusions

We have shown how to construct exact simultaneous one-sided and two-sided confidence bounds for comparing two regression lines to a control regression line over a fixed interval. The two-sided bounds are an improvement over those presented by Spurrier (2002) in that the predictor variable can be restricted to a predetermined interval. For the Masuda et al. (1997) example, by restricting the exposure time to the interval from 0 to 60 minutes, we are able to get 1.64% narrower confidence bounds compared to those of Spurrier (2002). If one restricts the predictor variable to a subset of this interval, the confidence bounds will be even narrower.

The one-sided bounds are particularly useful when one wants to give a lower bound on the amount of improvement that a test treatment offers relative to the control. All previous research in this area has dealt with two-sided bounds. By using one-sided bounds we get much better lower bounds at the expense of having no upper bound.

Appendix

Mathematical details for one-sided bounds

We will consider the region of interest for five cases:

When $-\infty < a < 0 < b$, $-\infty < T_1 < c$ and T_3 is bounded below and above by a line or arc depending on the value of T_1. It follows that the probability in equation (11) equals

$$\int_{-\infty}^{c} \int_{-\infty}^{c} \int_{l_1}^{u_1} \int_{l_2}^{u_2} f(t_1, t_2, t_3, t_4)\,\mathrm{d}t_4\,\mathrm{d}t_3\,\mathrm{d}t_2\,\mathrm{d}t_1$$

where,

$$l_i = \begin{cases} \dfrac{c\sqrt{1+a^2h} - t_i}{a\sqrt{h}} & \text{if} \quad t_i < \dfrac{c}{\sqrt{1+a^2h}} \\ -\sqrt{c^2 - t_i^2} & \text{otherwise} \end{cases}$$

and

$$u_i = \begin{cases} \dfrac{c\sqrt{1+b^2h} - t_i}{b\sqrt{h}} & \text{if} \quad t_i < \dfrac{c}{\sqrt{1+b^2h}} \\ \sqrt{c^2 - t_i^2} & \text{otherwise} \end{cases} \quad \text{for} \quad i = 1, 2\,. \tag{27}$$

Similarly we can write the probability in equation (11) for the other four cases. For $0 < a < b < \infty$, $-\infty < T_1 < \infty$ and T_3 is bounded below by $-\infty$ and above by a line or arc depending on the value of T_1 (see Figure 3a). By symmetry, we can get the probability for $-\infty < a < b < 0$, $-\infty < T_1 < \infty$ and T_3 is bounded above by $+\infty$ and below by a line or arc depending on the value of T_1. For $a = 0$, $b = \infty$, $-\infty < T_1 < c$ and T_3 is bounded below by $-\infty$ and above by a line or arc depending on the value of T_1 (see Figure 3b). By symmetry, we can get the probability for $a = -\infty$, $b = 0$, $-\infty < T_1 < c$ and T_3 is bounded above by $+\infty$ and below by a line or arc depending on the value of T_1.

We can simplify the above cases, by reducing the number of variables from *a, b,* and *h* to the angle θ. Rotate the axes such that the lines $T_3 = T_1 a\sqrt{h}$ in the (T_1, T_3) plane and $T_4 = T_2 a\sqrt{h}$ in the (T_2, T_4) plane become axes. This can be achieved by the transformation

$$W_1 = \frac{T_1 a\sqrt{h} - T_3}{\sqrt{1+a^2h}}\,, \qquad W_2 = \frac{T_2 a\sqrt{h} - T_4}{\sqrt{1+a^2h}}\,, \qquad W_3 = \frac{T_3 a\sqrt{h} + T_1}{\sqrt{1+a^2h}}\,,$$

$$\text{and} \quad W_4 = \frac{T_4 a\sqrt{h} + T_2}{\sqrt{1+a^2h}}\,. \tag{28}$$

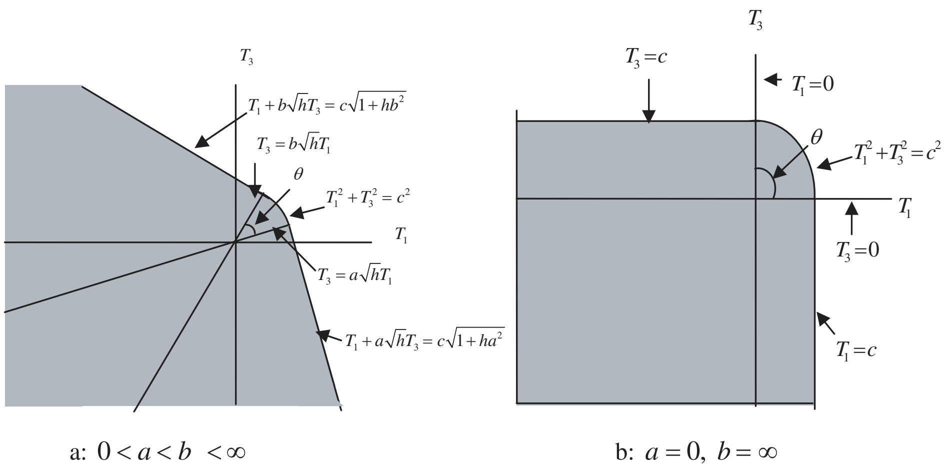

Figure 3 The region of interest for one-sided confidence bounds in the (T_1, T_3) plane.

The equations of interest in (T_1, T_3) plane get transformed as follows:

$$\begin{array}{ll} T_1^2 + T_3^2 = c^2 & W_1^2 + W_3^2 = c^2 \\ T_3 = T_1 a\sqrt{h} & W_1 = 0 \\ T_1 + T_3 a\sqrt{h} = c\sqrt{1 + a^2 h} & W_3 = c \\ T_3 = T_1 b\sqrt{h} & (a\sqrt{h} - b\sqrt{h})\, W_3 - (1 + abh)\, W_1 = 0 \\ T_1 + T_3 b\sqrt{h} = c\sqrt{1 + b^2 h} & (a\sqrt{h} - b\sqrt{h})\, W_1 + (1 + abh)\, W_3 = c\sqrt{1 + a^2 h}\,\sqrt{1 + b^2 h}\,. \end{array} \tag{29}$$

The region of interest in the (W_1, W_3) plane simplifies to three cases based on the size of θ.

The equations in (29) involving a, b, and h can be written in terms of θ as follows:

$$\begin{array}{ll} (a\sqrt{h} - b\sqrt{h})\, W_3 - (1 + abh)\, W_1 = 0 & \sin(\theta) W_3 + \cos(\theta)\, W_1 = 0 \\ (a\sqrt{h} - b\sqrt{h})\, W_1 + (1 + abh)\, W_3 = c\,\sqrt{1 + a^2 h}\,\sqrt{1 + b^2 h} & -\sin(\theta) W_1 + \cos(\theta)\, W_3 = c\,. \end{array} \tag{30}$$

The different cases for the region of interest can be seen from the Figure 4.

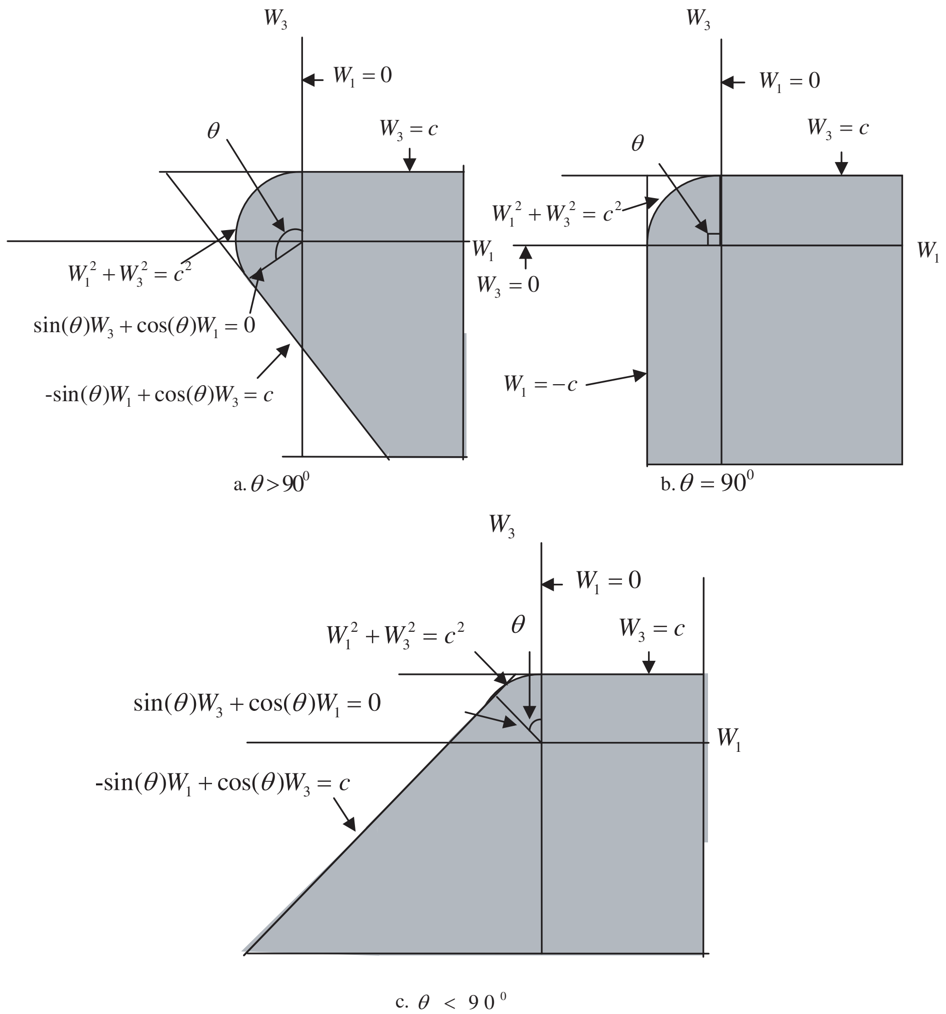

Figure 4 The region of interest in terms of θ for one-sided confidence bounds.

For, the probability in equation (11), the limits in the (W_1, W_3) plane are $-\infty < W_3 < c$ and upper limit for W_1 is $+\infty$ but the lower limit depends on cos (θ). When $W_3 \geq c \cos(\theta)$, then $W_1 \geq -\sqrt{c^2 - W_3^2}$, otherwise $W_1 \geq \dfrac{\cos(\theta) W_3 - c}{\sin(\theta)}$. Similarly we can get limits for (W_2, W_4).

The probability density function for W_1, W_2, W_3 and W_4 is given by $f(w_1, w_2, w_3, w_4)$, where f is defined in equation (15). Thus, the probability in equation (11) as a function of θ equals display (20).

The four-dimensional integral in (20) can also be written as a linear combination of 10 three-dimensional integrals. The derivation is too lengthy to include. Both methods give the same results for c up to the third decimal place.

Mathematical details for two-sided bounds

Using transformation (28) the cases for the region of interest $\theta \leq 90°$ and $> 90°$, can be seen from the Figure 5. Note that $\theta = 180°$ is the unrestricted predictor variable case.

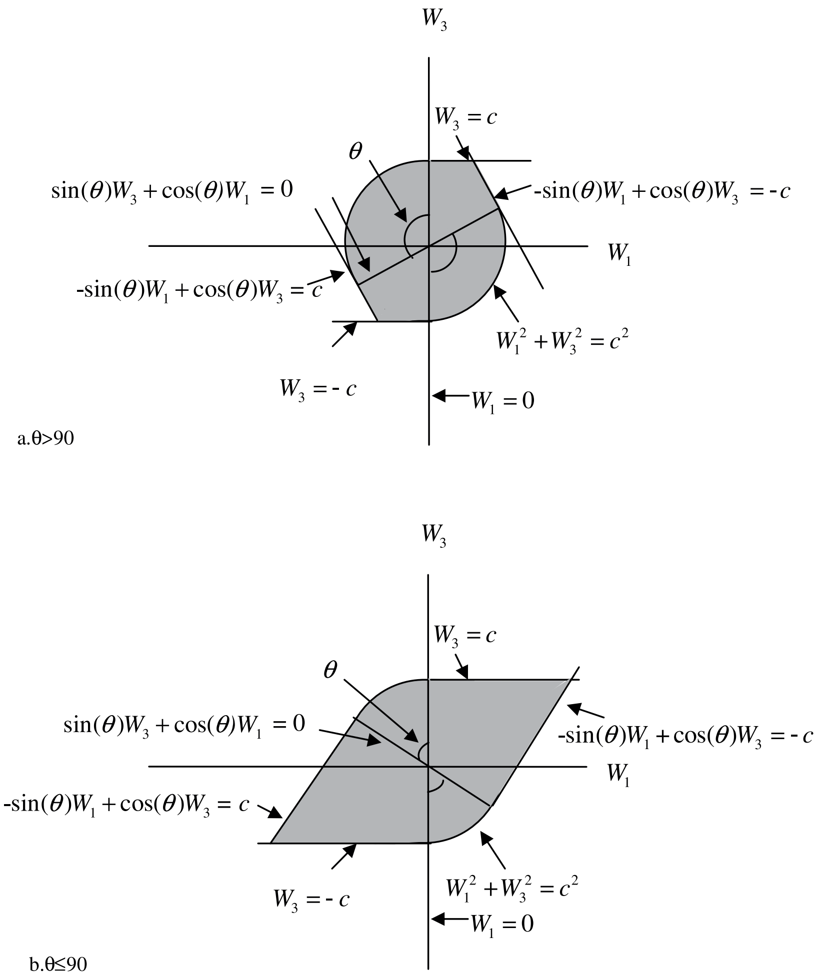

Figure 5 The region of interest in terms of θ for two-sided confidence bounds.

Hence, for two-sided $(1-\gamma)$ 100% simultaneous confidence bounds for comparing the regression models for two treatments to a control regression model, the limits in the (W_1, W_3) are $-c < W_3 < c$ and limits for W_1 depend on cos (θ). When $W_3 > c \cos(\theta)$, then W_1 is bounded below by the arc and the lower limit is given by $-\sqrt{c^2 - W_3^2}$, otherwise it is bounded by a line and the lower limit is given by $\dfrac{\cos(\theta)\, W_3 - c}{\sin(\theta)}$. When $W_3 < -c \cos(\theta)$, then W_1 is bounded above by the arc and the upper limit is given by $\sqrt{c^2 - W_3^2}$, otherwise it is bounded by a line and the upper limit is given by $\dfrac{\cos(\theta) W_3 + c}{\sin(\theta)}$. Similarly we can get limits for (W_2, W_4). Thus, the probability of interest for two-sided bounds can be represented as a function of θ by display (25).

Acknowledgements The authors thank the Editor, Associate Editor, and a referee for helpful comments.

References

Graybill, Franklin A. (1976). *Theory and Application of the Linear Model*. Wadsworth, Pacific Grove, CA, USA.

Liu, W., Jamshidan, M., and Zhang, Y. (2004). Multiple Comparison of Several Linear Regression Models. *Journal of the American Statistical Association*, **99**, 395–403.

Masuda, S., Saito, H., and Inui, K.-I. (1997). Interactions of Nonsteroidal Anti-inflammatory Drugs with Rat Renal Organic Anion Transporter, OAT-K1. *The Journal of Pharmacology and Experimental Therapeutics,* **283**, 1039–1042.

Royen, T. (1990). Tables for Studentized Multivariate Maximum Ranges and their Application by Maximum Range Tests. *Biometrical Journal*, **32**, 643–680.

Spurrier, J. D. (1999). Exact Confidence Bounds for All Contrasts of Three or More Regression Lines. *Journal of the American Statistical Association*, **94**, 483–488.

Spurrier, J. D. (2002). Exact Pairwise Comparisons of Three or More Regression Lines: Pairwise Comparisons and Comparisons with a Control. *Biometrical Journal,* **44**, 801–812.