Some New Methods for the Comparison of Two Linear Regression Models

Wei Liu, Mortaza Jamshidian, Ying Zhang, Frank Bretz, Xiaoliang Han

Abstract

The frequently used approach to the comparison of two linear regression models is to use the partial F test. It is pointed out in this paper that the partial F test has in fact a naturally associated two-sided simultaneous confidence band, which is much more informative than the test itself. But this confidence band is over the entire range of all the covariates. As regression models are true or of interest often only over a restricted region of the covariates, the part of this confidence band outside this region is therefore useless and to ensure $1 - \alpha$ simultaneous coverage probability is therefore wasteful of resources. It is proposed that a narrower and hence more efficient confidence band over a restricted region of the covariates should be used. The critical constant required in the construction of this confidence band can be calculated by Monte Carlo simulation. While this two-sided confidence band is suitable for two-sided comparisons of two linear regression models, a more efficient one-sided confidence band can be constructed in a similar way if one is only interested in assessing whether the mean response of one regression model is higher (or lower) than that of the other in the region. The methodologies are illustrated with two examples.

Southampton Statistical Sciences Research Institute
Methodology Working Paper M05/16

University
of Southampton

SOME NEW METHODS FOR THE COMPARISON OF TWO LINEAR REGRESSION MODELS

Wei Liu

Statistical Sciences Research Institute and School of Maths
The University, Southampton, SO17 1BJ, U.K.

Mortaza Jamshidian[1]
Dept of Maths, California State Univ., Fullerton, CA 92834, U.S.A.

Ying Zhang
Dept of Biostatistics, University of Iowa, Iowa City, IA 52242-1009, U.S.A.

Frank Bretz
Research Unit Bioinfomatics, Univ. of Hannover, D-30419 Hannover, Germany

Xiaoliang Han
Dept of Statistics, Shanghai Univ. of Finance & Economics, P.R.China

September 27, 2005

Abstract

The frequently used approach to the comparison of two linear regression models is to use the partial F test. It is pointed out in this paper that the partial F test has in fact a naturally associated two-sided simultaneous confidence band, which is much more informative than the test itself. But this confidence band is over the entire range of all the covariates. As regression models are true or of interest often only over a restricted region of the covariates, the part of this confidence band outside this region is therefore useless and to ensure $1 - \alpha$ simultaneous coverage probability is therefore wasteful of resources. It is proposed that a narrower and hence more efficient confidence band over a restricted region of the covariates should be used. The critical constant required in the construction of this confidence band can be calculated by Monte Carlo simulation. While this two-sided confidence band is suitable for two-sided comparisons of two linear regression models, a more efficient one-sided confidence band can be constructed in a similar way if one is only interested in assessing whether the mean response of one regression model is higher (or lower) than that of the other in the region. The methodologies are illustrated with two examples.

Keywords: Confidence bands; Linear regression; Multiple comparisons; Simultaneous inference; Statistical simulation.

1   Introduction

Suppose two linear regression models are given by

$$\mathbf{Y}_i = X_i \mathbf{b}_i + \mathbf{e}_i, \quad i = 1, 2$$

where $\mathbf{Y}_i = (y_{i,1}, \cdots, y_{i,n_i})^T$ is a vector of observations, X_i is a $n_i \times (p+1)$ full column-rank design matrix with the first column given by $(1, \cdots, 1)^T$ and the lth $(2 \leq l \leq p+1)$ column given by $(x_{i,1,l-1}, \cdots, x_{i,n_i,l-1})^T$, $\mathbf{b}_i = (b_{i,0}, \cdots, b_{i,p})^T$ is a vector of unknown coefficients, and $\mathbf{e}_i = (e_{i,1}, \cdots, e_{i,n_i})^T$ is a vector of random errors with all the $\{e_{i,j} : j = 1, \cdots, n_i, i = 1, 2\}$ being iid $N(0, \sigma^2)$ random variables, where σ^2 is an unknown parameter. Here $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are two groups of observations that depend on the same p covariates $x_1, \cdots, x_p$ via the classical linear regression model. The two groups may be two gender groups or two treatments etc.

The problem of comparing two linear regression models can be met in many research fields. As an example, Kleinbaum et al. (1998) consider how systolic blood pressure (Y) changes with the covariate age (x_1). From the data they have collected (Kleinbaum et al., 1998, pages 49 and 192) it is clear that the relationship between Y and x_1 can be reasonably described by a linear regression model of Y on x_1 for both the Female and Male gender groups. What they are interested in is whether the two linear regression models for Female and Male are the same, and whether the females tend to have a lower blood pressure than males of the same age.

The frequently used approach to this problem of comparing two linear regression models is to use the partial F test to test the hypotheses

$$H_0 : \ \mathbf{b}_1 = \mathbf{b}_2 \ \ \text{against} \ \ H_a : \ \mathbf{b}_1 \neq \mathbf{b}_2 \tag{1.1}$$

The inferences that can be drawn from this test are that if H_0 is rejected then the two regression models are deemed to be different, and that if H_0 is not rejected then there is not sufficient statistical evidence to conclude that the two regression models are different. But no tangible information on the magnitude of the difference between the two models is provided by this approach of hypotheses testing, whether H_0 is rejected or not. For instance, the partial F test for the blood pressure example results in a P-value smaller than 0.00001 (Kleinbaum et al., 1998, pages 327-330), and so there is a strong statistical indication that the two regression models are different. But no measurement is provided on how different these two regression models are.

The first purpose of this paper is to provide a two-sided simultaneous confidence band associated naturally with the partial F test of (1.1), so that the partial F test can be interpreted more intuitively via this simultaneous confidence band. The hypotheses (1.1) can in fact be tested by using this confidence band: the acceptance

or rejection of H_0 is according to whether or not the $Y = 0$ line lies completely inside the confidence band. The advantage of this confidence band approach over the partial F test is that it provides information on the magnitude of the difference between the two regression models, whether or not H_0 is rejected. This is discussed in Section 2. This confidence band is over the entire range $(-\infty, \infty)$ of each covariate however. As a linear regression model is true or of interest often only over a restricted region of the covariates, the part of the confidence band outside this restricted region is useless for inference. It is therefore unnecessary to guarantee the $1 - \alpha$ simultaneous coverage probability over the entire range of each covariate. Furthermore, inferences deduced from the confidence band outside the restricted region, such as the rejection of H_0, may not be valid since the assumed model may be wrong outside the restricted region. This calls for the construction of a $1 - \alpha$ simultaneous confidence band only over this restricted region of the covariates. This confidence band is narrower and so allows sharper inferences than the confidence band associated with the partial F test. This is discussed in Section 3 and forms the second purpose of this paper. These confidence bands are two-sided and suitable for two-sided inferences: whether the two regression models are the same or not (either higher or lower). In many applications, the interest is to assess whether the mean response of one regression model is higher than that of the other, e.g. whether males tend to have higher blood pressure than females in the blood pressure example. Section 4 focuses on one-sided confidence bands which is pertinent for one-sided inferences. This is the third purpose of this paper. Some concluding remarks are contained in Section 5. Finally the Appendix provides some mathematical details.

2   Partial F test from a confidence band viewpoint

The partial F test can be implemented by defining a dummy variable

$$z = \begin{cases} 1 & \text{if } y \text{ is from the model 1} \\ 0 & \text{if } y \text{ is from the model 2} \end{cases}$$

and setting up an overall model for the aggregated data from the two individual models as

$$y = \mathbf{x}^T \mathbf{c}_1 + z\mathbf{x}^T \mathbf{c}_2 + e \tag{2.1}$$

where $\mathbf{x}^T = (1, x_1, \cdots, x_p)$, $\mathbf{c}_1 = \mathbf{b}_2$ and $\mathbf{c}_2 = \mathbf{b}_1 - \mathbf{b}_2$. This overall model implies the two individual models:

$$y = \mathbf{x}^T (\mathbf{c}_1 + \mathbf{c}_2) + e = \mathbf{x}^T \mathbf{b}_1 + e$$

for a y from the individual model 1, and

$$y = \mathbf{x}^T \mathbf{c}_1 + e = \mathbf{x}^T \mathbf{b}_2 + e \tag{2.2}$$

2

for a y from the individual model 2. Now testing the hypotheses (1.1) of the coincidence of the individual models 1 and 2 is the same as testing

$$H_0 : \mathbf{c}_2 = \mathbf{b}_1 - \mathbf{b}_2 = \mathbf{0} \quad \text{against} \quad H_a : \mathbf{c}_2 \neq \mathbf{0}$$

for the overall model (2.1). This can be carried out by using the partial F test: H_0 is rejected if and only if the test statistic

$$\frac{\left[\text{Regression SS of model (2.1)} - \text{Regression SS of model (2.2)}\right]/(p+1)}{\text{MS residual of model (2.1)}}$$

is larger than $f_{p+1,\nu}^\alpha$, the upper α point of an F distribution with $p + 1$ and $\nu = n_1 + n_2 - 2(p+1)$ degrees of freedom. This can be found in most text books on multiple linear regression models; see e.g. Kleinbaum et al. (1998).

On the other hand, to compare the two regression models it is certainly useful to construct a simultaneous confidence band for the difference between the two models

$$\mathbf{x}^T \mathbf{b}_1 - \mathbf{x}^T \mathbf{b}_2 = (1, x_1, \cdots, x_p)\mathbf{b}_1 - (1, x_1, \cdots, x_p)\mathbf{b}_2$$

over the entire range $\{x_l \in (-\infty, \infty) \text{ for } l = 1, \cdots, p\}$ of the covariates. Since $X_i^T X_i$ is assumed to be non-singular, the least squares estimator of $\mathbf{b}_i$ is given by $\hat{\mathbf{b}}_i = (X_i^T X_i)^{-1} X_i^T Y_i, i = 1, 2$. Let $\hat{\sigma}^2$ denote the pooled error mean square, which is simply the mean square residual of the model (2.1). Then $\hat{\sigma}^2$ is independent of the $\hat{\mathbf{b}}_i$'s and has the distribution $\sigma^2 \chi_\nu^2 / \nu$ where $\nu = n_1 + n_2 - 2(p+1)$. So the variance of $\mathbf{x}^T \hat{\mathbf{b}}_1 - \mathbf{x}^T \hat{\mathbf{b}}_2$ is given by $\sigma^2 \mathbf{x}^T \Delta \mathbf{x}$, where $\Delta = (X_1^T X_1)^{-1} + (X_2^T X_2)^{-1}$. We therefore construct the following simultaneous confidence band

$$
\begin{aligned}
\mathbf{x}^T \mathbf{b}_1 - \mathbf{x}^T \mathbf{b}_2 \quad &\in \quad \mathbf{x}^T \hat{\mathbf{b}}_1 - \mathbf{x}^T \hat{\mathbf{b}}_2 \pm c\hat{\sigma}\sqrt{\mathbf{x}^T \Delta \mathbf{x}} \ , \\
\forall \ x_l \quad &\in \quad (-\infty, \infty) \ \text{ for } l = 1, \cdots, p
\end{aligned}
\tag{2.3}
$$

where c is the critical constant chosen so that the simultaneous coverage probability of this confidence band is equal to $1 - \alpha$. It is pointed out in the Appendix that $c = \sqrt{(p+1)f_{p+1,\nu}^\alpha}$.

From the confidence band (2.3) it is clear that if H_0 is true, i.e. the two regression models are the same, then with probability $1 - \alpha$ the zero line $\mathbf{x}^T \mathbf{0}$ lies inside the confidence band for the entire range. Hence an exact size α test of H_0 is to

reject H_0 if and only if $\mathbf{x}^T \mathbf{0}$ is outside the band for at least one $\mathbf{x}$. (2.4)

It is shown in the Appendix that this test is in fact the same as the partial F test, i.e. $\mathbf{x}^T \mathbf{0}$ is outside the band for at least one $\mathbf{x}$ if and only if the partial F test

rejects H_0. So the partial F test can be regarded simply as a side product of the confidence band (2.3). But of course the confidence band provides information on the magnitude of the difference between the two regression models at each point $\mathbf{x}$, which is not available from the partial F test.

Figure 1. Several two-sided confidence bands for $\mathbf{x}^T(\mathbf{b}_F - \mathbf{b}_M)$
for the blood pressure example
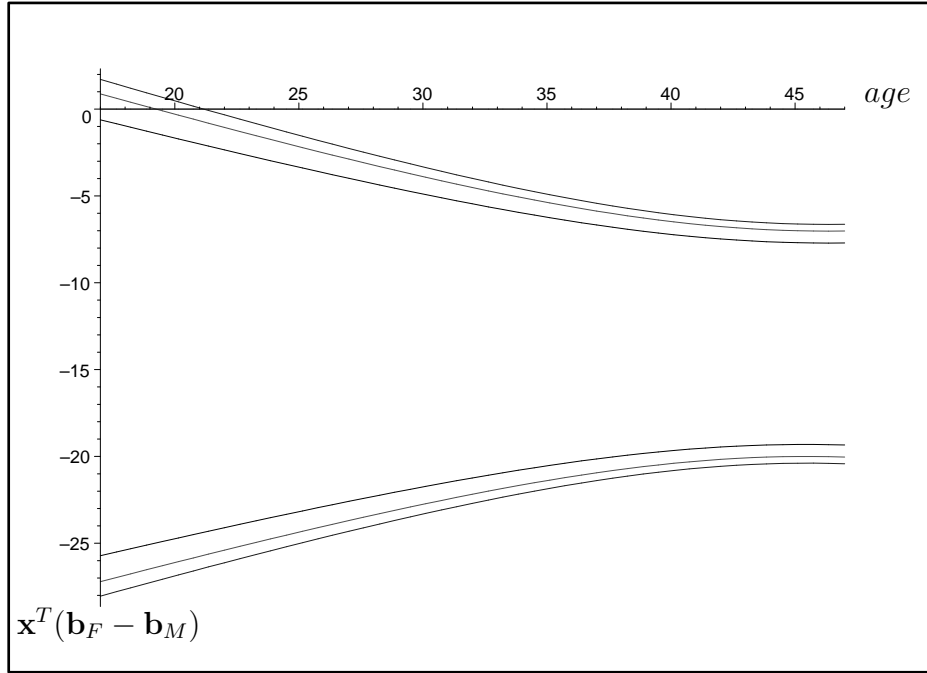
For the blood pressure example, several 99% confidence bands for $\mathbf{x}^T(\mathbf{b}_F - \mathbf{b}_M)$ are depicted in Figure 1. In this case, the critical constant of the band (2.3) is given by $c = \sqrt{2 f_{2,65}^{0.01}} = 3.146$, and the section of the band over the observed range $17 \leq age \leq 47$ is given by the highest and lowest curves in Figure 1. It is clear from this band that H_0 is rejected (at $\alpha = 0.01$) since the zero line is not included in the band for any $21.3 \leq x_1 = age \leq 47$. Furthermore one can infer from the band that females tend to have significantly lower blood pressure than males at all the ages between 21.3 and 47, since the upper curve of the band lies below the zero line for $21.3 \leq x_1 \leq 47$. This inference cannot be made from the partial F test. This illustrates the advantage of using the confidence band (2.3) instead of the partial F test.

Note however the section of the confidence band (2.3) over $-\infty < x_1 \leq 0$ is useless as the covariate, age, cannot take a negative value and hence the linear regression model of Y on x_1 cannot be true over $-\infty < x_1 \leq 0$. More seriously, H_0

4

should not be rejected simply because the zero line is outside this confidence band for an $x_1 \in (-\infty, 0]$. But the partial F test cannot tell whether this may indeed be the reason for the rejection of H_0. This highlights an inherent drawback of the partial F test and therefore the need to construct confidence bands over a restricted region of the covariates, which is the focus of the next section.

3   Confidence bands over a restricted region

Comparison of two regression models over the entire range of the $\mathbf{x}$ is seldom of interest. One reason is that a covariate may only take values in a restricted range. Even if a covariate may take values in the entire range $(-\infty, \infty)$, a linear regression model is seldom suitable for the entire range of the covariate. For example, the covariate, age, in the blood pressure example cannot take a negative value. Furthermore, it is inconceivable that a straight line regression model of blood pressure on age holds for all ages. So a simultaneous confidence band over a suitable restricted region of the covariates, e.g. the observed region of the covariates, is more sensible. The added incentive is that a confidence band over a restricted region is narrower and hence allows sharper inferences than the confidence band over the entire range. In this section we consider the construction of simultaneous confidence bands over a given rectangular region $x_l \in [a_l, b_l]$, $-\infty < a_l \le b_l < \infty$, $l = 1, \cdots, p$.

We construct the following simultaneous confidence band

$$
\begin{aligned}
\mathbf{x}^T\mathbf{b}_1 - \mathbf{x}^T\mathbf{b}_2 &\in \mathbf{x}^T\hat{\mathbf{b}}_1 - \mathbf{x}^T\hat{\mathbf{b}}_2 \pm c\hat{\sigma}\sqrt{\mathbf{x}^T\Delta\mathbf{x}} , \\
\forall\ x_l &\in [a_l, b_l]\ \text{ for } l = 1, \cdots, p.
\end{aligned}
\tag{3.1}
$$

The central question is how the critical constant c can be determined so that the simultaneous confidence level of this band is equal to $1 - \alpha$. The confidence level of the band (3.1) is given by $P\{T < c\}$ where

$$
T = \sup_{x_l \in [a_l, b_l], l = 1, \cdots, p} \frac{|\mathbf{x}^T[(\hat{\mathbf{b}}_1 - \mathbf{b}_1) - (\hat{\mathbf{b}}_2 - \mathbf{b}_2)]|}{\hat{\sigma}\sqrt{\mathbf{x}^T\Delta\mathbf{x}}}.
\tag{3.2}
$$

Furthermore, T is a pivotal random quantity whose distribution does not depend on the unknown parameters $\mathbf{b}_1, \mathbf{b}_2$ and σ. So the critical value c can be determined if the distribution of T can be found. While it is clearly difficult to derive a closed form formula for the distribution of T in this general setting, the method of Liu, Jamshidian and Zhang (2004) can be used to efficiently simulate a T. So one can simulate a sufficiently large number, R say, replicates of the random variable T, and set the $[(1 - \alpha)R]$th largest simulated value as the critical constant c. A geometrical representation of T, which is central to the simulation of T and also relevant to establishing some assertions of Section 2, is given in the Appendix.

As pointed out in Section 2, for the extreme case of $a_l = -\infty$ and $b_l = \infty$ for $l = 1, \cdots, p$, the critical constant c is given by $\sqrt{(p+1)f^\alpha_{p+1,\nu}}$. On the other hand if $a_l = b_l$ for $l = 1, \cdots, p$ then the critical constant c is given by $t_\nu^{1-\alpha/2}$, the $1 - \alpha/2$ percentile of a t distribution with ν degrees of freedom.
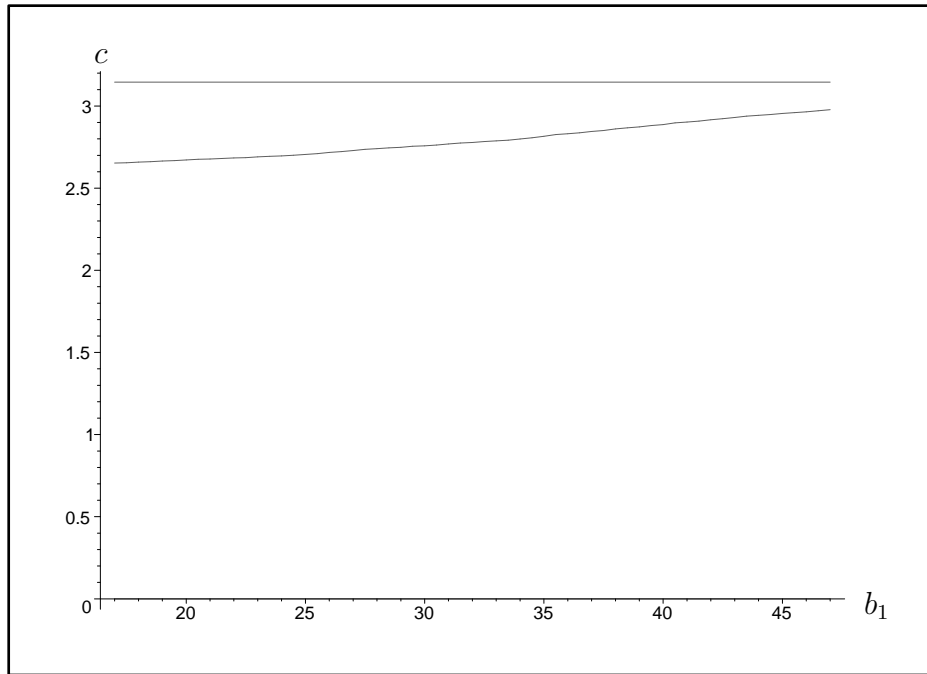
As in Section 2, it is clear that the confidence band (3.1) can be used to test the hypotheses (1.1) in the following way: reject H_0 if and only if

$$\sup_{x_l \in [a_l, b_l], l=1,\cdots,p} \frac{|\mathbf{x}^T(\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2)|}{\hat{\sigma}\sqrt{\mathbf{x}^T \Delta \mathbf{x}}} > c$$

or equivalently

$\quad \mathbf{x}^T \mathbf{0}$ is outside the band (3.1) for at least one $\mathbf{x}$ in the restricted region.   (3.3)

Figure 2. Critical value c of the band (3.1) as a function of b_1
with $a_1 = 17$ for the blood pressure example

For the blood pressure example, if the 99% confidence band over the observed range of 17 $\leq$age$\leq$ 47 is required, then $c = 2.969$ with standard error 0.003 based on $R = 1,000,000$ simulations. This confidence band is given in Figure 1 by the two curves just inside the two curves of the band (2.3). This band is about 6% narrower than the confidence band (2.3) over the range $17 \leq x_1 \leq 47$ and so more

informative. The band (3.1) can be substantially narrower than the band (2.3) if the range $a_1 \leq x_1 \leq b_1$ is narrow enough. In Figure 2, the curve shows how the 99% critical value c of the band (3.1) changes with $17 \leq b_1 \leq 47$ while fixing $a_1 = 17$. The level line gives the 99% critical value $c = 3.146$ for the band (2.3). The lowest point of the curve gives the critical constant $c = t_{64}^{0.995} = 2.654$ of the band (3.1) for the extreme case of $b_1 = a_1$, which is in fact the critical constant of the 99% pointwise t confidence interval. This pointwise t confidence band is given by the two middle curves in Figure 1. Note, however, the pointwise confidence band can be used to make inferences about the regression models at only one x_1 which should be selected before observing the data. On the other hand, the simultaneous confidence band (3.1) can be used to make inferences about the regression models at any point in the restricted region, and the probability that all the inferences are correct simultaneously is at least $1 - \alpha$.

To emphasize that the methodologies of this paper are applicable to any number of covariates, we give another example with $p = 2$ covariates. In the aerobic fitness data from the SAS/STAT User's Guide (1990, p. 1443), the oxygen consumption (OC) in ml per kg body weight per minute was measured for 31 individuals. The covariates of interest are *age* and *run pulse* which is the heart rate while running (at the time the oxygen consumption rate was measured).

It is well known that lack of fitness increases the risk of heart illness. Gardner et al. (1996) studied the risk prediction from the time to run 1.5 miles in a study of 391 cases of heart illness in marine recruits. A recruit unable to run 1.5 miles in 12 minutes had eight times the risk of heart illness in basic training as did one with a faster run time. So we use the threshold value 12 minutes on the time to run 1.5 miles also provided in the aerobic fitness data to divide the 31 individuals into two groups: the first with less than 12 minutes to run 1.5 miles, and the second with more than 12 minutes to run 1.5 miles. We are interested whether the two groups have different OC rate over some range of the covariates *age* and *run pulse*.

Figure 3 is approximately here

The 99% critical constant of the confidence band (2.3) is given by $c = \sqrt{3f_{3,25}^{0.01}} = 3.745$; the confidence band is depicted by the two outside curviplanes in Figure 3. For the comparison of the two groups over the observed range $38 \leq age \leq 57$ and $146 \leq run\ pulse \leq 186$, the 99% critical constant of the confidence band (3.1) is given by $c = 3.689$, which is only marginally smaller than the $c = 3.745$. However, if one is interested in the comparison of the two groups over the range $38 \leq age \leq 45$ and $160 \leq run\ pulse \leq 186$, where 45 and 160 are the commonly known midage and intermediate pulse rate respectively, then the 99% critical constant of the confidence band (3.1) is given by $c = 3.401$. This confidence band is depicted by the

7

two inner curviplanes in Figure 3. Even though both the band (3.1) and the F band (2.3) are not significant from Figure 3 , the band (3.1) is about 9% narrower and so allows sharper inferences than the band (2.3) over the range $38 \leq age \leq 45$ and $160 \leq run\ pulse \leq 186$.

4   One-sided Confidence bands

The confidence bands (2.3) and (3.1) are suitable for two-sided inferences. If one is only interested in assessing whether one linear regression model has a lower mean response than another linear regression model, such as whether the females tend to have lower blood pressure than the same-age males over certain age range, one-sided confidence bands are more pertinent. We consider an upper confidence band of the form

$$
\begin{aligned}
\mathbf{x}^T\mathbf{b}_1 - \mathbf{x}^T\mathbf{b}_2 &\leq \mathbf{x}^T\hat{\mathbf{b}}_1 - \mathbf{x}^T\hat{\mathbf{b}}_2 + c\hat{\sigma}\sqrt{\mathbf{x}^T\Delta\mathbf{x}} \ , \\
\forall\ x_l &\in\ [a_l, b_l]\ \ \text{for}\ l = 1, \cdots, p.
\end{aligned}
\tag{4.1}
$$

The critical constant c can again be determined by simulation similar to the two-sided case so that the simultaneous confidence level of this band is equal to $1-\alpha$; the Appendix provides more details on this. From the confidence band (4.1), one can infer whether the mean response of model 1, $\mathbf{x}^T\mathbf{b}_1$, is lower than the mean response of model 2, $\mathbf{x}^T\mathbf{b}_2$, in only some parts or over the whole region of $x_l \in [a_l, b_l]$, $l = 1, \cdots, p$. Note that a lower confidence band can be considered in a similar way and uses the same critical constant c as the upper confidence band (4.1).

For the special case of $a_l = -\infty$ and $b_l = \infty$ for $l = 1, \cdots, p$, it can be shown by using a result of Hochberg and Quade (1975, Lemma) that the critical constant c is the solution to the equation

$$
\frac{1}{2}F(c^2/(p+1); p+1, \nu) + \frac{1}{2}F(c^2/p; p, \nu) = 1 - \alpha
$$

where $F(\cdot; m, \nu)$ is the cumulative distribution function of a central F-variate with m and ν degrees of freedom. For the case of $a_l = b_l$ for $l = 1, \cdots, p$, the critical constant c is given by $t_\nu^{1-\alpha}$, the $1 - \alpha$ percentile of a t distribution with ν degrees of freedom. This corresponds to the pointwise one-sided t confidence band.

For the blood pressure example over the observed region of $x_1 \in [17, 47]$, the 99% level critical constant is calculated to be $c = 2.711$ based on 1,000,000 replications. The critical constant of Hochberg and Quade's (1995) 99% upper confidence band over the entire range is calculated to be 2.973. For the pointwise upper band, the critical constant is given by $t_{65}^{0.99} = 2.385$. These three upper bands are depicted

in Figure 4. Hochberg and Quade's band is given by the highest curve, the band (4.1) is given by the middle curve, and the pointwise band is given by the lowest curve. Again, the pointwise band can be used to make inferences only at one pre-specified point x_1 and so is not very useful. The band (4.1) is always below and so more informative than Hochberg and Quade's band in the restricted region. The band (4.1) is also always below the upper curve of the two-sided band (3.1) and so more informative than the band (3.1) for one-sided inferences. For example, the upper band allows one to infer that the female blood pressure tends to be lower than the male blood pressure for any $17 \leq age \leq 47$, while the two-sided band allows one to infer that the female blood pressure tends to be lower than the male blood pressure for $19.1 \leq age \leq 47$. The curve in Figure 5 shows how the 99% critical constant c of band (4.1) changes with $17 \leq b_1 \leq 47$ while fixing $a_1 = 17$. The level line gives the critical constant $c = 2.973$ of the Hochberg and Quade's band.

Figure 4. Several 99% upper bands on $\mathbf{x}^T(\mathbf{b}_F - \mathbf{b}_M)$ for the blood pressure example
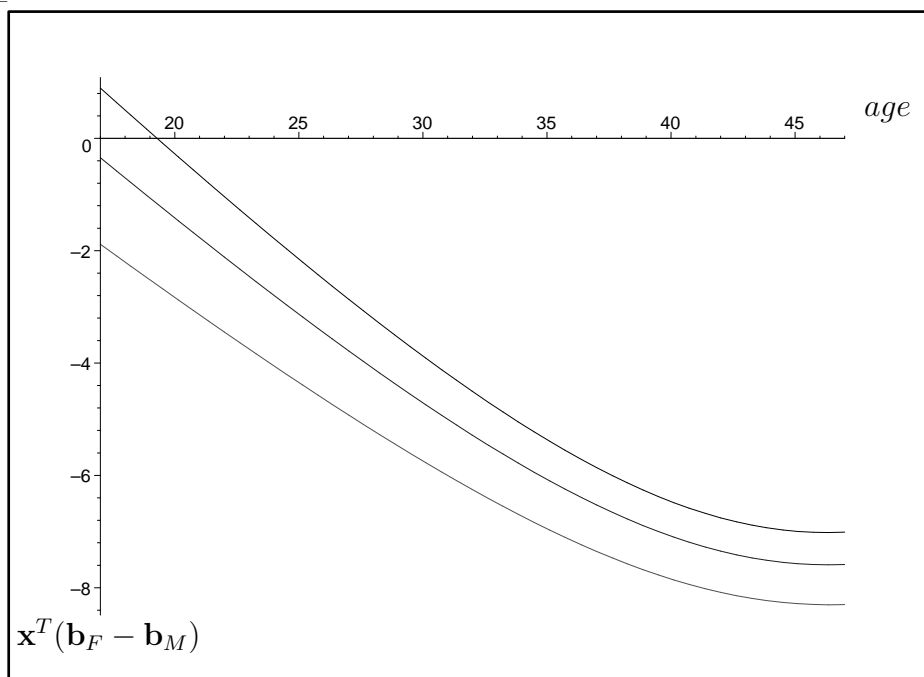
Figure 5. Critical value c of the band (4.1) as a function of b_1
with $a_1 = 17$ for the blood pressure example
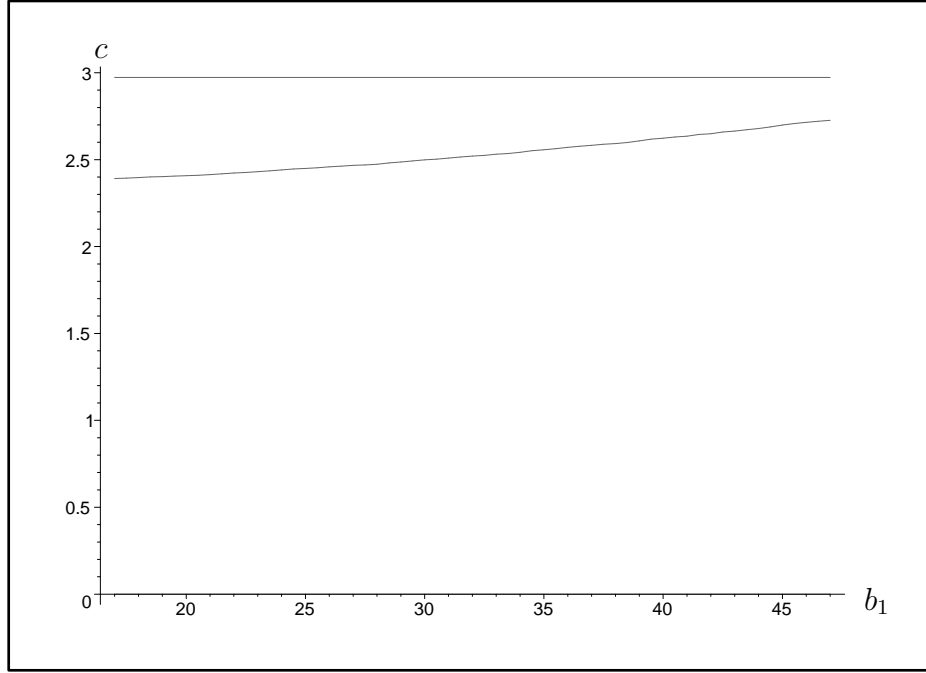
For the aerobic fitness data, it is a common belief that fitter individuals tend to have a larger OC rate than less fit individuals. So one may be interested to known to what extent the data supports this common belief. One-sided inference is appropriate in this case. The 99% critical constant of the lower confidence band over the whole range $-\infty < age < \infty$ and $-\infty < run\ pulse < \infty$ is given by $c = 3.583$, and the lower confidence band is depicted by the lower curviplane in Figure 6. As this plane lies completely below the zero plane from Figure 6, there is no statistical evidence that the fitter group tend to have a larger OC rate than the less fit group for any age and $run\ pulse$. However if one is only interested in the one-sided comparison of the two groups over the range $38 \leq age \leq 45$ and $160 \leq run\ pulse \leq 186$, then the 99% critical constant of the lower confidence band is given by $c = 3.125$. This lower confidence band is depicted by the upper curviplane in Figure 6. Since this plane passes through the zero plane in a small area about $age = 45$ and $run\ pulse = 175$ (inside the range $38 \leq age \leq 45$ and $160 \leq run\ pulse \leq 186$), one can infer that the fitter group do tend to have a larger OC rate than the less fit group in this small area of age and $run\ pulse$, which is a statistically significant result.

Figure 6 is approximately here

10

5   Conclusions

It is proposed in this paper to use confidence bands for the comparisons of two linear regression models. It is pointed out that the usual partial F test has in fact a naturally associated confidence band, which is much more informative than the test itself. But this confidence band is over the entire range of all the covariates. As regression models are true or of interest often only over a restricted region of the covariates, the part of this confidence band outside this region is useless and to guarantee an overall $1 - \alpha$ confidence level is wasteful of resources. A narrower and hence more efficient confidence band is constructed over a restricted region of the covariates. While these confidence bands are suitable for two-sided comparisons of two linear regression models, more efficient one-sided confidence bands can be constructed in a similar way if one is only interested in assessing whether the mean response of one model is higher than that of the other.

It is worth emphasizing that the limits a_l and b_l on the covariates may be chosen in various ways. In standard clinical studies the values of the important covariates (such as age, blood pressure, etc.) are pre-defined by a set of inclusion/exclusion criteria; patients are randomized into the study only if they satisfy the inclusion criteria. Similarly, in survey studies, the range of the important covariates is fixed in advance as well, e.g., when considering risk factors for children the covariate *age* is usually constrained to lie within two thresholds. In these cases it may be natural to choose the pre-defined thresholds as a_l and b_l. On the other hand, sometimes one may be interested in the comparison of two regression models over a region that is only part of region specified by the pre-specified thresholds. In this case the region specified by a_l and b_l should be smaller than the experimental region, which is determined by the pre-defined thresholds, to reflect the interest of statistical inference. Even if pre-specified thresholds or clear-cut region of interest for statistical inference are not available, natural boundaries may still be available and should be used for sharper inferences: e.g., it is reasonable to assume $0 < age < 100$, $50 < rest\ pulse < 100$, and so on. So long as the choices of a_l and b_l are not dependent on the random observations $\mathbf{Y}$, the confidence level $1 - \alpha$ is valid.

Finally, for the comparisons of more than two linear regression models, Spurrier (1999, 2002), Liu, et al. (2004), and Bhargava and Spurrier (2004) proposed various confidence bands. Note however the partial F test for this case does not have a naturally associated set of confidence bands, and so the comparison of more than two linear regression models is characteristically different from the comparison of two linear regression models.

Appendix

In this appendix, we first provide a geometric representation of T in (3.2) which is central to the method of simulating T in Liu et al. (2004). It is then used to show that the partial F test can actually be deduced from the confidence band (2.3).

Note that the partial F test can be expressed as

$$\text{reject } H_0 \text{ if and only if } \frac{(\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2)^T \Delta^{-1} (\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2)}{(p+1)\hat{\sigma}^2} > f_{p+1,\nu}^{\alpha}; \qquad (A1)$$

see Scheffé (1959) for details.

Also note that there exists a $(p+1) \times (p+1)$ non-singular matrix P such that

$$(X_1^T X_1)^{-1} + (X_2^T X_2)^{-1} = P^T P.$$

Let $\mathbf{Z} = (P^T)^{-1}(\hat{\mathbf{b}}_1 - \mathbf{b}_1 - \hat{\mathbf{b}}_2 + \mathbf{b}_2)/\sigma$ then $\mathbf{Z}$ is a $p+1$ dimensional standard normal random vector independent of $\hat{\sigma}$. Hence T is given by

$$T = \sup_{x_l \in (a_l, b_l), l=1,\cdots,p} \frac{|(P\mathbf{x})^T \mathbf{Z}|}{(\hat{\sigma}/\sigma)\sqrt{(P\mathbf{x})^T (P\mathbf{x})}} = \frac{\|\mathbf{Z}\|}{(\hat{\sigma}/\sigma)} Q, \qquad (A2)$$

where

$$Q = \sup_{x_l \in (a_l, b_l), l=1,\cdots,p} \frac{|(P\mathbf{x})^T \mathbf{Z}|}{\|P\mathbf{x}\| \|\mathbf{Z}\|} = \sup_{\mathbf{v} \in V} \frac{|\mathbf{v}^T \mathbf{Z}|}{\|\mathbf{v}\| \|\mathbf{Z}\|} \qquad (A3)$$

where

$$\begin{aligned} V &= \{P\mathbf{x} : x_l \in (a_l, b_l), l = 1, \cdots, p\} \\ &= \{\mathbf{p}_0 + x_1 \mathbf{p}_1 + \cdots + x_p \mathbf{p}_p : x_l \in (a_l, b_l), l = 1, \cdots, p\}. \end{aligned}$$

with $P = (\mathbf{p}_0, \mathbf{p}_1, \cdots, \mathbf{p}_p)$.

To simulate a T from (A2), the key is to calculate Q in (A3) which involves the maximization of a p-variate function over a hyper-rectangle region. Note that $\mathbf{v}^T \mathbf{Z}/(\|\mathbf{v}\| \|\mathbf{Z}\|)$ is simply the cosine of the angle between $\mathbf{v}$ and $\mathbf{Z}$. In order to find Q one only needs to find the smallest angle between either $\mathbf{Z}$ and $\mathbf{v}$ or $-\mathbf{Z}$ and $\mathbf{v}$ while $\mathbf{v}$ varies in the cone V. Liu, et al. (2004) gave an efficient method for the computation of Q.

For the band (2.3), the cone V is the p dimensional hyper-plane that passes through $\mathbf{p}_0$. So Q is equal to one for each given $\mathbf{Z}$ and hence $T = \|\mathbf{Z}\|/(\hat{\sigma}/\sigma)$. By noting that

$$T^2/(p+1) = \frac{\|\mathbf{Z}\|^2/(p+1)}{(\hat{\sigma}/\sigma)^2}$$

has an F distribution with degrees of freedom $p + 1$ and ν, the critical constant c for the band (2.3) is clearly given by $\sqrt{(p+1)f^{\alpha}_{p+1,\nu}}$. Furthermore, by noting that $\mathbf{Z} = (P^T)^{-1}(\hat{\mathbf{b}}_1 - \mathbf{b}_1 - \hat{\mathbf{b}}_2 + \mathbf{b}_2)/\sigma$, we have

$$
\begin{aligned}
\frac{T^2}{(p+1)} &= \frac{\| \mathbf{Z} \|^2 / (p+1)}{(\hat{\sigma}/\sigma)^2} \\
&= \frac{(\hat{\mathbf{b}}_1 - \mathbf{b}_1 - \hat{\mathbf{b}}_2 + \mathbf{b}_2)^T \Delta^{-1} (\hat{\mathbf{b}}_1 - \mathbf{b}_1 - \hat{\mathbf{b}}_2 + \mathbf{b}_2)}{(p+1)\hat{\sigma}^2}.
\end{aligned}
\tag{A4}
$$

From the derivation it is clear that the algebraic equation (A4) holds for any $\mathbf{b}_1$ and $\mathbf{b}_2$, and so for $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{0}$ in particular. Now note that the zero line $\mathbf{x}^T\mathbf{0}$ lies inside the band (2.3) if and only if $T^2/(p+1) \leq f^{\alpha}_{p+1,\nu}$ with $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{0}$ in T. This observation and (A4) with $\mathbf{b}_1 = \mathbf{b}_2 = \mathbf{0}$ imply that the zero regression line $\mathbf{x}^T\mathbf{0}$ lies inside the band (2.3) if and only if

$$
\frac{(\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2)^T \Delta^{-1} (\hat{\mathbf{b}}_1 - \hat{\mathbf{b}}_2)}{(p+1)\hat{\sigma}^2} \leq f^{\alpha}_{p+1,\nu}.
$$

Hence the partial F test is the same as the test (2.4).

For the one-sided confidence band (4.1), its simultaneous confidence level is given by $P\{U < c\}$ where

$$
\begin{aligned}
U &= \sup_{x_l \in (a_l, b_l), l=1, \cdots, p} \frac{\mathbf{x}^T[(\hat{\mathbf{b}}_1 - \mathbf{b}_1) - (\hat{\mathbf{b}}_2 - \mathbf{b}_2)]}{\hat{\sigma}\sqrt{\mathbf{x}^T \Delta \mathbf{x}}} \\
&= \sup_{x_l \in (a_l, b_l), l=1, \cdots, p} \frac{(P\mathbf{x})^T \mathbf{Z}}{(\hat{\sigma}/\sigma)\sqrt{(P\mathbf{x})^T(P\mathbf{x})}} \\
&= \frac{\| \mathbf{Z} \|}{(\hat{\sigma}/\sigma)} \sup_{\mathbf{v} \in V} \frac{\mathbf{v}^T \mathbf{Z}}{\| \mathbf{v} \| \| \mathbf{Z} \|}.
\end{aligned}
$$

The last superiem can be found in a similar but simpler way as for Q in (A3), since one only needs to find the smallest angle between $\mathbf{Z}$ and $\mathbf{v}$ while $\mathbf{v}$ varies in the cone V.

REFERENCES

Bhargava, P and Spurrier, J.D. (2004). Exact confidence bounds for comparing two regression lines with a control regression line on a fixed interval. *Biometrical Journal*, 46(6): 720-730.

Gardner, J.W., Kark, J.A., Karnei, K., Sanborn, J.S., Gastaldo, E., Burr, P., and Wenger, C.B. (1996). Risk factors predicting exertional heat illness in male Marine Corps recruits. *Medicine and Science in Sports and Exercise*, 28(8), 939-944.

Hochberg, Y. and Quade, D. (1975). One-sided simultaneous confidence bounds on regression surfaces with intercepts. *J. of the American Statistical Association*, 70, 889-891.

Kleinbaum, D.G., Kupper, L.L., Muller, K.E. and Nizam, A. (1998). *Applied Regression Analysis and Other Multivariable Methods, 3rd Edition*. Duxbury Press.

Liu, W., Jamshidian, M. and Zhang, Y. (2004). Multiple comparison of several linear regression models. *J. of the American Statistical Association*, 99, 395-403.

SAS Institute Inc. (1990). *SAS/STAT User's Guide, Version 6, Fourth edition, Volume 2* SAS InstituteInc., Cary, NC.

Scheffé, H. (1959). *The Analysis of Variance*. Wiley.

Spurrier, J.D. (1999). Exact Confidence Bounds for All Contrasts of Three or More Regression Lines. *J. of the American Statistical Association*, 94, 483-88.

Spurrier, J.D. (2002). Exact multiple comparisons of three or more regression lines: Pairwise comparisons and comparisons with a control. *Biometrical Journal*, 44(7), 801-812.

Figure 3: The two-sided restricted and unrestricted 99% confidence bands on $x^T(b_1-b_2)$ for the aerobic data, Age and pulse are restricted in $[38,45] * [160,186]$, $c = 3.7452$, $c(\text{restricted}) = 3.401$. The shaded part in the included zero plane indicates the restricted region.

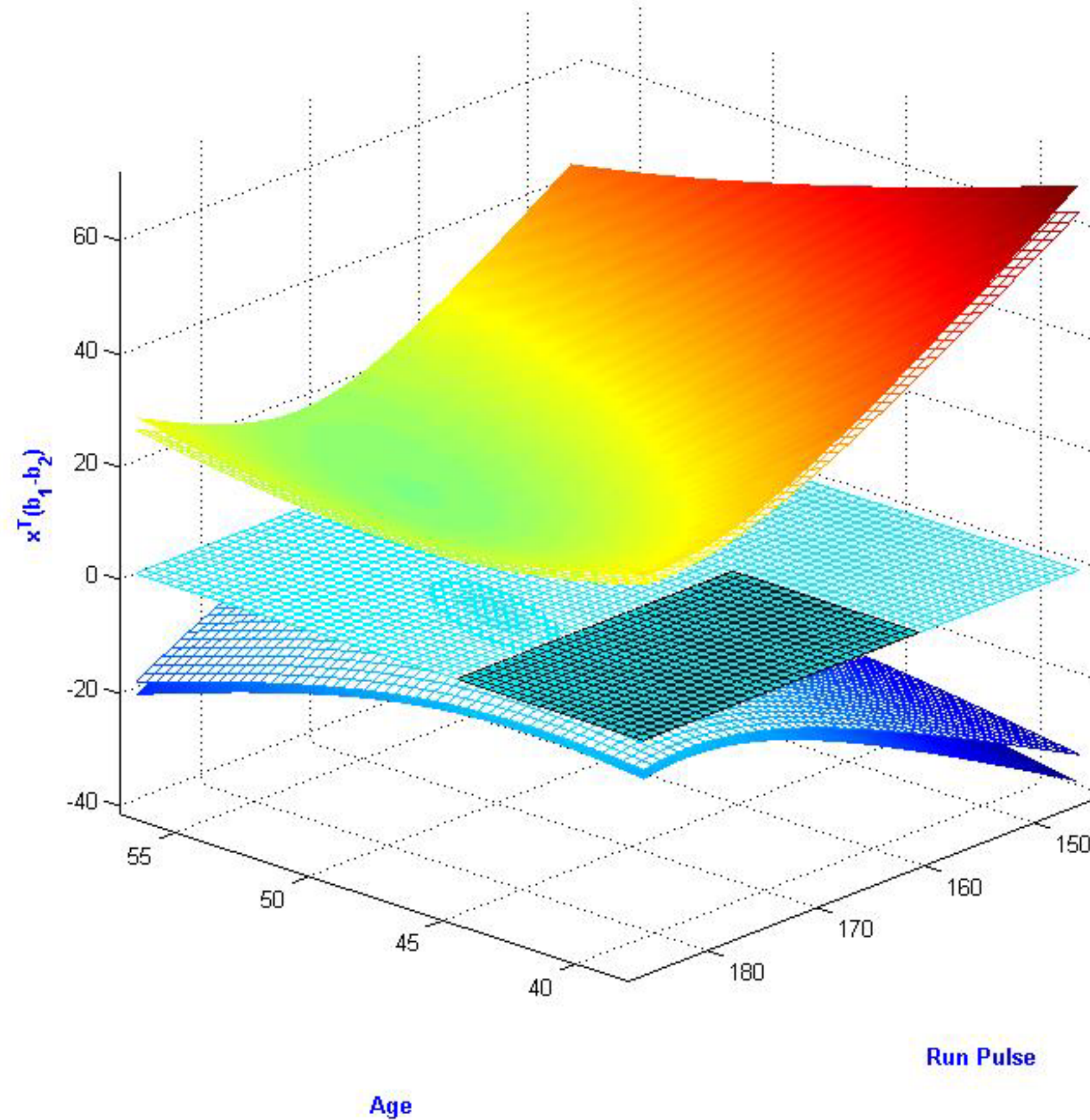Figure 6: The lower restricted and unrestricted 99% confidence bands on $x^T(b_1 - b_2)$ for the aerobic data.

Age and pulse are restricted in $[38,45] * [160,186]$, $c = 3.583$, $c(\text{restricted}) = 3.125$.
The shaded part in the included zero plane indicates the restricted region.
Hidden: The unrestricted band does not pass through the zero plane.