Journal of Statistical Computation and Simulation

ISSN: 0094-9655 (Print) 1563-5163 (Online) Journal homepage: https://www.tandfonline.com/loi/gscs20

Simultaneous bootstrap confidence bands in regression

Peter Hall & Y.E. Pittelkow

To cite this article: Peter Hall & Y.E. Pittelkow (1990) Simultaneous bootstrap confidence bands in regression, Journal of Statistical Computation and Simulation, 37:1-2, 99-113, DOI: 10.1080/00949659008811297

To link to this article: https://doi.org/10.1080/00949659008811297

Published online: 20 Mar 2007.

J. Statist. Comput. Simul., Vol. 37, pp. 99–113
Reprints available directly from the publisher
Photocopying permitted by license only

SIMULTANEOUS BOOTSTRAP CONFIDENCE BANDS IN REGRESSION

PETER HALL and Y. E. PITTELKOW

Australian National University

(*Received 30 November 1989*)

We suggest pivotal methods for constructing simultaneous bootstrap confidence bands in regression. Most attention is given to the problem of simple linear regression, but our techniques admit trivial extension to other cases, including polynomial regression. The advantages of our bootstrap approach are twofold. Firstly, the bootstrap allows a very general distribution for the errors, and secondly, it admits a wide variety of shapes for the confidence band. In our technique the shape of each envelope of the band is determined by a general template, chosen by the experimenter, and bootstrap methods are used to select the scale of the template.

KEY WORDS: Edgeworth expansion, pivot, scale, skewness, template.

1. INTRODUCTION

The problem of constructing simultaneous confidence bands for a regression line dates back to Working and Hotelling (1929). Over the last sixty years there has developed an extensive theory, based on the assumption of normal errors, encompassing constant-width bands (Gafarian, 1964; Bowden and Graybill, 1966), polygonal bands (Graybill and Bowden, 1967), Working–Hotelling bands over general regions (Wynn and Bloomfield, 1971; Uusipaikka, 1983); and one-sided bands (Bohrer and Francis, 1972). In this paper we show how to use the bootstrap to develop simple, practicable generalizations of all these methods. The advantages of the bootstrap approach are twofold. Firstly, it enables the assumption of normal errors to be eliminated; we require only mild regularity conditions on the errors, such as that they have a nonsingular distribution with sufficiently many finite moments. Secondly, it allows the shape of the confidence band to be determined virtually arbitrarily, without concern for the mathematical complexities of theory for a general shape configuration.

In our approach to constructing confidence bands, the shape of the bands is determined by two arbitrary "templates". The templates are general functions which are permitted to depend on the design variables. One template determines the shape of the upper envelope of the band, and the other gives the lower envelope. For the sake of explicitness we devote special attention to two particular templates, one a fixed template giving a straight-line envelope parallel to the estimated regression line, and the other a variable curve giving the shape associated with Working–Hoteling bands. However, there are many other possibilities, including polygonal templates, and our general description of the bootstrap approach allows a great many choices. There is no reason why the template for the upper envelope has to be the same as that for the lower envelope. Indeed, one-sided confidence bands are constructed with one of the templates set equal to infinity.

Once the templates have been selected, the bootstrap is used to choose the scale of each template. The classical normal-based theory imposes an identical scale on either template, reflecting the symmetry of the error distribution. However, in the case of asymmetric errors there can be sound reasons for employing different scales on the templates. We suggest that the scales be chosen to minimize the width of the band; this is quite sensible if the bootstrap is used to determine scale. An alternative approach is to choose scale according to an "equal-tailed" algorithm, in which case the true regression line has equal probability of protruding from either of the two envelopes.

We discuss our techniques in the context of constructing confidence bands over general intervals, finite or infinite, in simple linear regression. The techniques are readily, in fact trivially, extended to other cases, for example to polynomial regression and to non-interval ranges. In the case of classical normal-based methods, these generalizations have been discussed by Wynn and Bloomfield (1971), Gafarian (1978), Uusipaikka (1983) and Wynn (1984). Section 2 presents our methodology in the case of simple linear regression, and Section 3 describes generalizations. Several examples are analyzed in Section 4, involving a variety of different templates and error distributions. Applications to real data and to simulated data are presented, and a Monte Carlo study of coverage accuracy is summarized. Finally, Section 5 describes theoretical properties of the bands. We show that the envelopes of our bootstrap bands are second-order accurate, meaning that in the case of non-normal data they correctly capture second-order terms representing departure from normality. They err only in their capacity for capturing third-order effects.

Seber (1977, p. 183ff) and Kendall and Taylor (1979, p. 388ff) give very accessible accounts of the classical theory of confidence bands for regression lines. In addition we should mention the work of Naiman (1984) and Piegorsch (1984, 1985) on optimality properties of confidence bands.

2. METHODOLOGY FOR SIMPLE LINEAR REGRESSION

We take the model for the observed data $\mathscr{X}=\{(x_i, Y_i),\ 1\leqq i\leqq n\}$ to be

$$Y_i=a+bx_i+e_i, \qquad 1\leqq i\leqq n,$$

where the e_i's are independent and identically distributed with zero mean and variance σ^2. Least-squares estimates of b and a are $\hat{b}=n^{-1}\sigma_x^{-2}\Sigma(x_i-\bar{x})(Y_i-\bar{Y})$ and $\hat{a}=\bar{Y}-\hat{b}\bar{x}$, where $\bar{x}=n^{-1}\Sigma x_i$, $\bar{Y}=n^{-1}\Sigma Y_i$, $\sigma_x^2=n^{-1}\Sigma(x_i-\bar{x})^2$. Our estimate of the regression line $y=ax+b$ is $y=\hat{a}+\hat{b}x$, and our estimate of σ^2 is

$$\hat{\sigma}^2=n^{-1}\sum_{i=1}^{n}(Y_i-\hat{a}-\hat{b}x_i)^2.$$

(Dividing by n or $n-2$ in this formula has no effect on our confidence bands.) The standardized difference between estimated and true regression lines is

$$\Delta(x) = (\hat{a} + \hat{b}x - a - bx)/\hat{\sigma}. \tag{2.1}$$

This "studentizing", or pivoting, of the difference between regression lines is a key feature of our method.

We base inference on the random process Δ. Assume initially that the distribution of Δ is known; shortly we shall show how to estimate the distribution by means of the bootstrap. To construct a confidence band for the line $y = a + bx$ over an interval I we need two "templates", to describe the shapes of the upper and lower envelopes of the band respectively, and two factors for scaling the templates. The upper and lower templates are represented by two known, nonnegative functions, f_+ and f_- respectively. The templates may depend on the design variables x_i. Let u_+ and u_- be the corresponding scale factors. Then the upper and lower envelopes of the band are the curves $\mathscr{C}_+$ and $\mathscr{C}_-$, defined by

$$\mathscr{C}_\pm = \{(x, y) : x \in I,\ y = \hat{a} + \hat{b}x \pm \hat{\sigma} u_\pm f_\pm(x)\}. \tag{2.2}$$

The confidence band is

$$\{(x, y) : x \in I, \hat{a} + \hat{b}x - \hat{\sigma} u_- f_-(x) \leqq y \leqq \hat{a} + \hat{b}x + \hat{\sigma} u_+ f_+(x)\}.$$

The shape of the band is determined by the templates, and the coverage and "skewness" (relative to the line estimate $y = \hat{a} + \hat{b}x$) are governed by the scale factors. For a given pair of templates, and a given coverage level $1-\alpha$, we must select u_+, u_- such that

$$P\{\hat{a} + \hat{b}x - \hat{\sigma} u_- f_-(x) \leqq a + bx \leqq \hat{a} + \hat{b}x + \hat{\sigma} u_+ f_+(x), \quad \text{all } x \in I\} = 1 - \alpha,$$

i.e. such that

$$P\{-u_+ f_+(x) \leqq \Delta(x) \leqq u_- f_-(x), \quad \text{all } x \in I\} = 1 - \alpha. \tag{2.3}$$

There are two obvious candidates for f_+ and f_-: the fixed template $f_F(x) \equiv 1$, and the root parabolic template

$$f_P(x) = [\operatorname{var}\{n^{1/2}(\hat{a} + \hat{b}x)/\sigma\}]^{1/2} = \{1 + \sigma_x^{-2}(x - \bar{x})^2\}^{1/2}.$$

A third option, suggested by work of Graybill and Bowden (1967), is the V-shaped template

$$f_V(x) = 1 + \sigma_x^{-1}|x - \bar{x}|.$$

One-sided bands would be constructed with either f_+ or f_- set at $+\infty$. In the case of normal errors, Gafarian (1964) treated the case where both f_+ and f_- equal f_F. There the confidence band is determined by two straight lines on either side of and parallel to $y=a+bx$. This is a constant-width confidence band. Working and Hotelling (1929) took both f_+ and f_- to equal f_P, in which case the band has the familiar "double hyperbola" shape. This is a "constant-probability" band, since for large n the pointwise coverage probability

$$P\{\hat{a}+\hat{b}x-\hat{\sigma}u_-f_-(x)\leqq a+bx\leqq \hat{a}+\hat{b}x+\hat{\sigma}u_+f_+(x)\}$$

is virtually independent of x. (The probability is approximately equal to $P\{-n^{1/2}u_+<N(0,1)<n^{1/2}u_-\}$, which does not depend on x.) We should note that Working and Hotelling allowed only the case $I=(-\infty,\infty)$. See Halpern, Rastogi, Ho and Yang (1967), Halperin and Gurian (1971) and Wynn and Bloomfield (1971) for a treatment of $f_\pm=f_P$ for a finite interval I, with normal errors. Of course, the constant width case, $f_\pm=f_F$, is only possible for a finite I.

There is no reason why f_+ and f_- should be the same function. For example, we might take $f_+=f_P$ and $f_-=f_F$; see Section 4 for an example. An advantage of the bootstrap approach to confidence bands is that it readily accommodates markedly asymmetric confidence bands.

Having settled on choices of f_+ and f_- we must select scale factors u_+ and u_- such that relation (2.3) holds. Of course, there is an infinite variety of choices of (u_+,u_-), although once one element of the pair is selected the other is determined by (2.3). We suggest two particular choices of (u_+,u_-): the symmetric choice, where $u_+=u_-$, and the narrowest-width choice, where u_+ and u_- are selected to minimize u_++u_- subject to (2.3). In the case of approximately symmetric errors and identical templates the symmetric choice is attractive. However, in other circumstances the narrowest-width choice has the advantage of producing a narrower band with the same level of coverage.

The equal-tailed method is also an option for defining u_+ and u_-. To implement it, let $u_+(\alpha)$ and $u_-(\alpha)$ be such that

$$P\{-u_+(\alpha)f_+(x)\leqq\Delta(x),\text{ all }x\in I\}=P\{\Delta(x)\leqq u_-(\alpha)f_-(x),\text{ all }x\in I\}=1-(\alpha/2),$$

and define α' to be that value of α such that

$$P\{-u_+(\alpha')f_+(x)\leqq\Delta(x)\leqq u_-(\alpha')f_-(x),\text{ all }x\in I\}=1-\alpha.$$

(Note that $\alpha'>\alpha$.) Take $u_+=u_+(\alpha')$, $u_-=u_-(\alpha')$. For this choice of the scale factors, the confidence band whose boundary curves are defined by (2.2) has the property that the unknown regression line has an equal probability of intersecting either boundary over the interval I.

Precise computation of u_+ and u_- requires knowledge of the distribution of the random process Δ, and such information is usually only available in the case of normal errors. When the error distribution is unknown, bootstrap methods may be used to estimate u_+ and u_-. To implement the bootstrap, first define the residuals

$$\hat{e}_i = Y_i - (\hat{a} + \hat{b}x_i), \qquad 1 \leq i \leq n,$$

and let $e_1^*, \ldots, e_n^*$ denote a resample drawn at random, with replacement, from $\{\hat{e}_1, \ldots, \hat{e}_n\}$. Put $Y_i^* = \hat{a} + \hat{b}x_i + e_i^*$, $1 \leq i \leq n$. Let $\hat{a}^*$, $\hat{b}^*$, $\hat{\sigma}^{*2}$ denote versions of $\hat{a}$, $\hat{b}$, $\hat{\sigma}^2$ computed for the case where the data set $\mathscr{X}$ is replaced by $\mathscr{X}^* = \{(x_i, Y_i^*), 1 \leq i \leq n\}$. Put

$$\Delta^*(x) = (\hat{a}^* + \hat{b}^* x - \hat{a} - \hat{b}x)/\hat{\sigma}^*. \tag{2.4}$$

We estimate solutions u_+, u_- of (2.3) as solutions $\hat{u}_+, \hat{u}_-$ of

$$P\{-\hat{u}_+ f_+(x) \leq \Delta^*(x) \leq \hat{u}_- f_-(x), \text{ all } x \in I \,|\, \mathscr{X}\} = 1 - \alpha. \tag{2.5}$$

The symmetric choice has $\hat{u}_- = \hat{u}_+$; the narrowest-width choice has $\hat{u}_+$ and $\hat{u}_-$ taken to minimize $\hat{u}_+ + \hat{u}_-$ subject to (2.5); and there is an obvious analogue of the equal-tailed choice. For any of these rules the envelopes of a nominal $(1-\alpha)$-level bootstrap confidence band for $y = a + bx$ over the interval I are given by the curves $\hat{\mathscr{C}}_+$ and $\hat{\mathscr{C}}_-$, defined by

$$\hat{\mathscr{C}}_\pm = \{(x, y) : x \in I, y = \hat{a} + \hat{b}x \pm \hat{\sigma}\hat{u}_\pm f_\pm(x)\}; \tag{2.6}$$

compare (2.2).

In Section 5 we shall give a theoretical account of the accuracy of this bootstrap approximation to the "ideal" confidence band described earlier. We shall show that the band between the two curves $\mathscr{C}_+$ and $\mathscr{C}_-$ is only $O_p(n^{-1/2})$ wide, and that the curves $\hat{\mathscr{C}}_+$ and $\mathscr{C}_+$ are $O_p(n^{-3/2})$ apart, uniformly over any finite range. (The same applies to the curves $\hat{\mathscr{C}}_-$ and $\mathscr{C}_-$.) Thus, the bootstrap confidence bands are second-order accurate, in that they correctly capture second-order features, of size $O_p(n^{-1})$, of the confidence bands. They correct for the major departure from normality of the error distribution, due to skewness.

3. GENERALIZATIONS

The case of polynomial regression may be treated as a special case of univariate multiparameter regression, and so we shall study the latter. We could also examine multivariate, multiparameter regression, but there the confidence region would be particularly awkward to represent.

Take the model for the observed data $\mathscr{X} = \{(\mathbf{x}_i, Y_i), 1 \leq i \leq n\}$ to be

$$Y_i = a + \mathbf{b}^T\mathbf{x}_i + e_i, \qquad 1 \leqq i \leqq n,$$

where $\mathbf{x}_i, \mathbf{b}$ are column vectors of length p, and the e_i's are independent and identically distributed with zero mean and variance σ^2. (To facilitate comparison with simple linear regression we separate out an intercept term. The case where the intercept is constrained to be zero, or any other constant, may be treated similarly.) Least-squares estimates of $\mathbf{b}$ and a are $\hat{\mathbf{b}} = n^{-1}\sigma_x^{-2}\Sigma(\mathbf{x}_i - \bar{\mathbf{x}})(Y_i - \bar{Y})$ and $\hat{a} = \bar{Y} - \hat{\mathbf{b}}^T\bar{\mathbf{x}}$, where $\bar{\mathbf{x}} = n^{-1}\Sigma\mathbf{x}_i$, $\bar{\mathbf{Y}} = n^{-1}\Sigma\mathbf{Y}_i$, $\sigma_x^2 = n^{-1}\Sigma(\mathbf{x}_i - \bar{\mathbf{x}})^T(\mathbf{x}_i - \bar{\mathbf{x}})$. Our estimate of σ^2 is $\hat{\sigma}^2 = n^{-1}\Sigma(Y_i - \hat{a} - \hat{\mathbf{b}}^T\mathbf{x}_i)^2$.

We wish to construct a confidence region for $a + \mathbf{b}^T\mathbf{x}$, all $\mathbf{x} \in R$, where R is a given subset of p-dimensional Euclidean space. The techniques discussed in Section 2 may be applied as before, with only the following minor changes: $a + \mathbf{b}^T\bar{\mathbf{x}}$, $\hat{a} + \hat{\mathbf{b}}^T\bar{\mathbf{x}}$, R replace $a + b\bar{x}$, $\hat{a} + \hat{b}\bar{x}$, I, respectively, at all appearances of the latter; and the templates f_P and f_V are altered to

$$f_P(\mathbf{x}) = \{1 + \sigma_x^{-2}(\mathbf{x} - \bar{\mathbf{x}})^T(\mathbf{x} - \bar{\mathbf{x}})\}^{1/2}, \quad f_V(\mathbf{x}) = 1 + \sigma_x^{-1}\sum_{j=1}^{p} |x^{(j)} - \bar{x}^{(j)}|.$$

In practice it will sometimes be the case that $\mathbf{x}$ is a function of univariate quantity, for example x_1 (the first component of $\mathbf{x}$). If we were representing polynomial regression of order p by multiple regression then we would have $\mathbf{x} = (x_1, \ldots, x_1^p)^T$. Here our interest would generally be in constructing a confidence region for $a + \mathbf{b}^T\mathbf{x}$, $x_1 \in I$, where I is a given interval. We would take $R = \{(x_1, \ldots, x_1^p)^T : x_1 \in I\}$; and recognize that the confidence region over R, say $\{(\mathbf{x}, y) : \mathbf{x} \in R,\ h_1(\mathbf{x}) \leqq y \leqq h_2(\mathbf{x})\}$, served only as a notational surrogate for the band over I, $\{(x_1, y) : x_1 \in I,\ h_1(\mathbf{x}) \leqq y \leqq h_2(\mathbf{x})$ where $\mathbf{x} = (x_1, \ldots, x_1^p)^T\}$.

4. EXAMPLES

In this section we report an application of different templates and scale factors to some real data, to data simulated from a model with an asymmetric error distribution and to a two-sample regression problem. We also report a Monte Carlo simulation of coverage accuracy. All examples use a balanced resampling algorithm adapted from Gleason (1988).

Example 4.1. Comparison of Different Templates and Scale Factors for Some Blood Pressure Data

The blood pressure data used here are described by Cox and Snell (1981, p. 70) and consist of measurements on 15 patients from a "before" and "after" study following administration of 25 mg of the drug Captopril. The dependent variable is the difference in diastolic blood pressure. The graphs in Figure 1 show the envelopes of simultaneous confidence bands, conditional on the observed values of diastolic blood pressure difference, for various combinations of templates. The

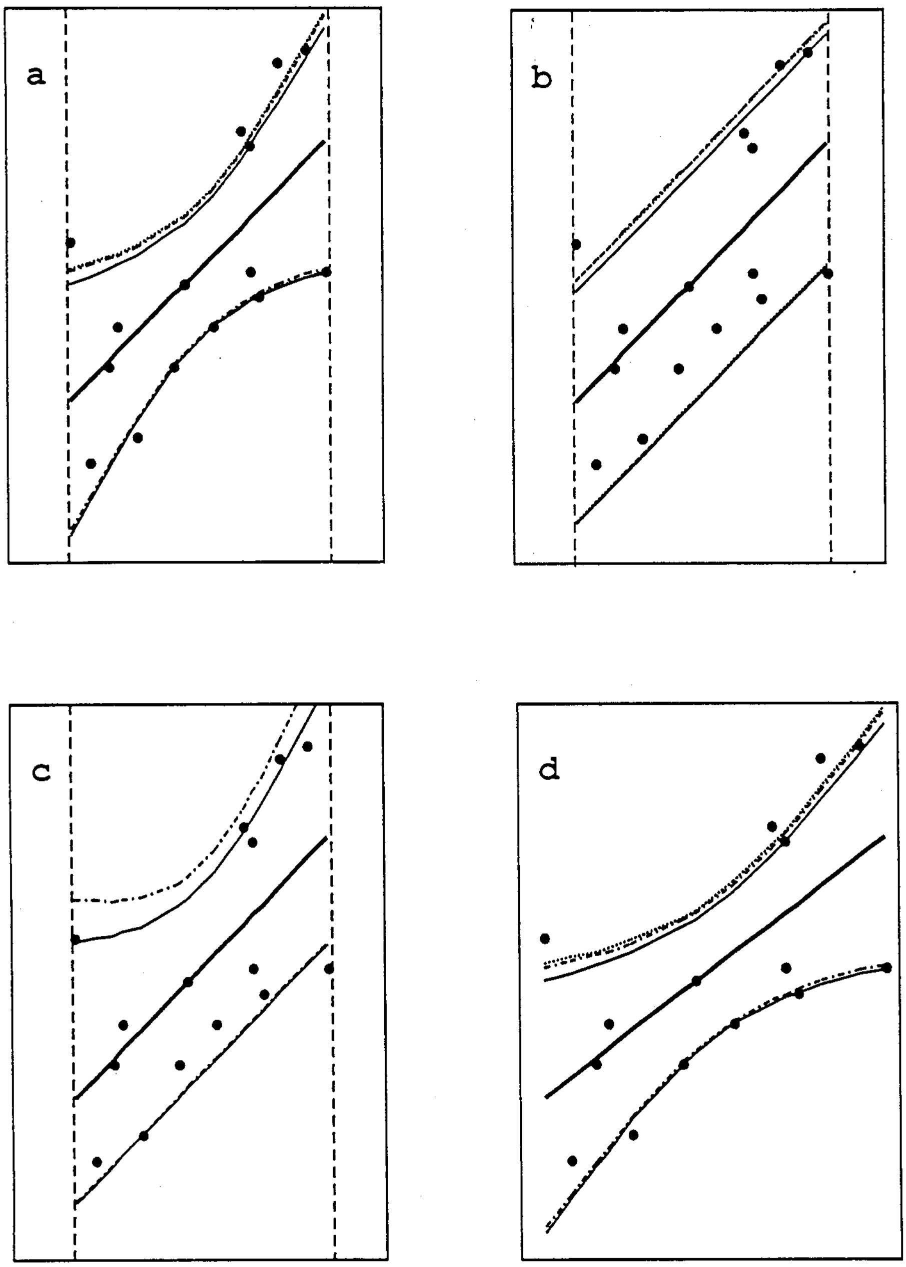

Figure 1 Envelopes for confidence bands for data of Example 4.1, in cases (a) constant-probability bands (i.e. $f_{\pm}=f_P$), $I=[-23,4]$; (b) constant-width bands (i.e. $f_{\pm}=f_F$), $I=[-23,4]$; (c) $f_+=f_P$, $f_-=f_F, I=[-23,4]$; (d) constant-probability bands (i.e. $f_{\pm}=f_P$), $I=(-\infty,\infty)$. In each panel, the line $y=\hat{a}+\hat{b}x$ is bold, ——— = minimum width bootstrap band, $\cdot-\cdot-\cdot$ = "symmetric" bootstrap band (i.e. $u_+=u_-$), $\cdots\cdots$ = normal theory band. Normal theory bands are not readily available in case (c).

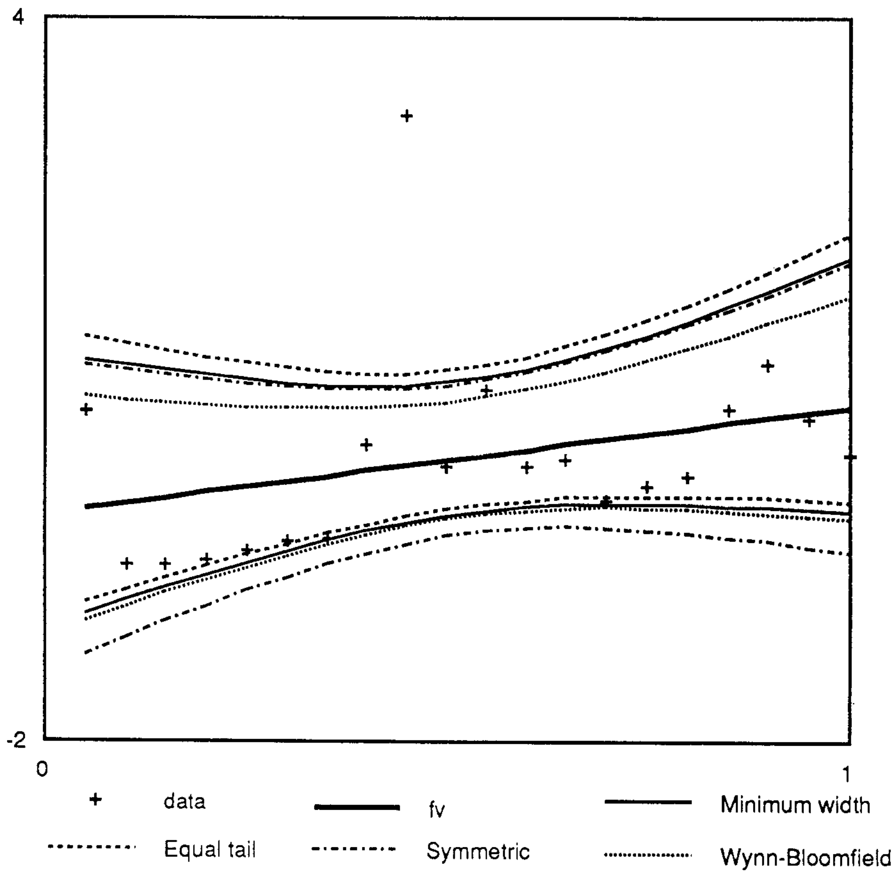

Figure 2 Envelopes of confidence bands for the data of Example 4.2. Error $|N(0,1)|^2-1$; skewness 5.634.

bootstrap bands were computed from 499 resamples, and the scale factor was estimated to provide a coverage level of 95%. The confidence bands based on the assumption of normality were generally very close to the symmetric bootstrap bands for these data and mostly cannot be distinguished from each other in the graphs. The bands based on normality are as follows—the Wynn–Bloomfield band in panel *a*, the Working–Hotelling bands in panel d, and the constant width band described by Bowden and Graybill (1966, p. 182) in panel b. Where the normal theory band is restricted to an interval a dashed vertical line indicates the upper and lower limits of the interval.

Example 4.2. Constant Probability Bands for Simulated Data Set

Figure 2 shows the simultaneous constant probability envelopes for 20 points simulated under a model $y_i=x_i+e_i$ where $e_i \sim |N(0,1)|^2-1$ and $x_i=\frac{1}{20},\frac{2}{20},\ldots,1.$

Table 1 Coverage probabilities for several different confidence intervals, estimated by Monte Carlo methods. See Example 4.3 for further details

$n=10$

Error distribution *Band*	*Constant probability*		*Constant width*		*Wynn–Bloomfield*
	Symmetric	*Min. width*	*Symmetric*	*Min. width*	
$N(0,1)$	0.967	0.955	0.965	0.957	0.971
$\|N(0,1)\|$	0.940	0.935	0.950	0.947	0.934
$\|N(0,1)\|^2$	0.909	0.902	0.918	0.923	0.887
$\|U(0,1)\|^2$	0.953	0.947	0.949	0.945	0.943

$\mathbf{x}=\{\frac{1}{10},\frac{2}{10},\ldots,1\}$

$n=15$

Error distribution *Band*	*Constant probability*		*Constant width*		*Wynn–Bloomfield*
	Symmetric	*Min. width*	*Symmetric*	*Min. width*	
$N(0,1)$	0.958	0.948	0.954	0.951	0.956
$\|N(0,1)\|$	0.973	0.956	0.948	0.940	0.942
$\|N(0,1)\|^2$	0.925	0.924	0.937	0.932	0.905
$\|U(0,1)\|^2$	0.951	0.950	0.949	0.948	0.948

$\mathbf{x}=\{\frac{1}{15},\frac{2}{15},\ldots,1\}$

$n=20$

Error distribution *Band*	*Constant probability*		*Constant width*		*Wynn–Bloomfield*
	Symmetric	*Min. width*	*Symmetric*	*Min. width*	
$N(0,1)$	0.955	0.943	0.943	0.942	0.953
$\|N(0,1)\|$	0.953	0.951	0.953	0.952	0.948
$\|N(0,1)\|^2$	0.917	0.917	0.921	0.923	0.896
$\|U(0,1)\|^2$	0.949	0.947	0.949	0.939	0.944

$\mathbf{x}=\{\frac{1}{20},\frac{2}{20},\ldots,1\}$

This figure shows the envelopes for the minimum width, equal tail, "symmetric" and Wynn–Bloomfield bands for data when there is marked asymmetry in the residuals.

Example 4.3. Comparison of Coverage Errors for Different Templates and Scale Factors

Monte Carlo simulations were used to estimate coverage probabilities for the simultaneous bands. The results are summarized in Table 1. The model $y_i=a+bx_i+e_i$, $1\leq i\leq n$, with $a=0$ and $b=1$ was used with four error distributions, each standardized to have mean zero and unit variance. They were regularly spaced such that for a sample of size n, $x=1/n$, $2/n,\ldots,n/n$. The scale was chosen to provide a (nominal) coverage level of 95%.

The results in Table 1 for sample sizes of 10, 15 and 20 are based on 1000 Monte Carlo simulations and 499 resamples computed using the NAG subroutine library on a PYRAMID 9825.

Example 4.4. Pairs of Regression Lines

A common problem in applied regression is the comparison of two (or more) regression lines. Simultaneous bootstrap confidence bands can be used to examine this problem without making any more assumptions about the residuals than that they are independent and identically distributed.

The data used for this example are described in Aitkin *et al.* (1988, p. 70). The dependent variable is the time taken by 24 children to construct designs in an intelligence test, and the independent variable is a measure of field dependence. The children were divided into two groups of 12 each and given different instructions before starting the test.

95% simultaneous bootstrap confidence bands for these two groups may be computed using the generalizations in Section 3, or by noting that if inference is conducted conditional on numbers m, n, the sample sizes in each of the two groups, the two different samples may be treated as independent, putting $1-\alpha_1=(0.95)^{n/(m+n)}$ and $1-\alpha_2=(0.95)^{m/(m+n)}$.

Figure 3 shows the 95% symmetric constant probability bands for these data computed for $\alpha_1=\alpha_2=0.025$ using 519 resamples in each group. From the extent of overlap of the confidence bands for the two groups one would conclude, as did Aitkin *et al.*, that the two regression lines are not different over the interval shown.

5. THEORY

For the sake of simplicity we shall give a theoretical account only in the case of simple linear regression. The generalizations described in Section 3 may be treated similarly.

Recall from Section 2 that ideally, the scale factors u_+ and u_- would be determined from the equation $\pi(u_+, u_-)=1-\alpha$, where

$$\pi(u_+, u_-)=P\{-u_+f_+(x)\leqq\Delta(x)\leqq u_-f_-(x), \text{ all } x\in I\}$$

and $\Delta(x)=(\hat{a}+\hat{b}x-a-bx)/\hat{\sigma}$; see (2.1) and (2.3). Usually the function π would be unknown, and we have suggested that it be replaced by its bootstrap estimate

$$\hat{\pi}(u_+, u_-)=P\{-u_+f_+(x)\leqq\Delta^*(x)\leqq u_-f_-(x), \text{ all } x\in I|\mathscr{X}\},$$

where $\Delta^*(x)=(\hat{a}^*+\hat{b}^*x-\hat{a}-\hat{b}x)/\hat{\sigma}^*$; see (2.4) and (2.5). In this section we shall prove that π and $\hat{\pi}$ agree to first and second orders, differing only in third-order terms. That is, $\hat{\pi}-\pi=O_p(n^{-1})$ uniformly in $u_+, u_->0$. If we were to make a normal approximation to the distribution of Δ then our estimate of π would differ from the true value in terms of order $n^{-1/2}$, not n^{-1}. Therefore the bootstrap

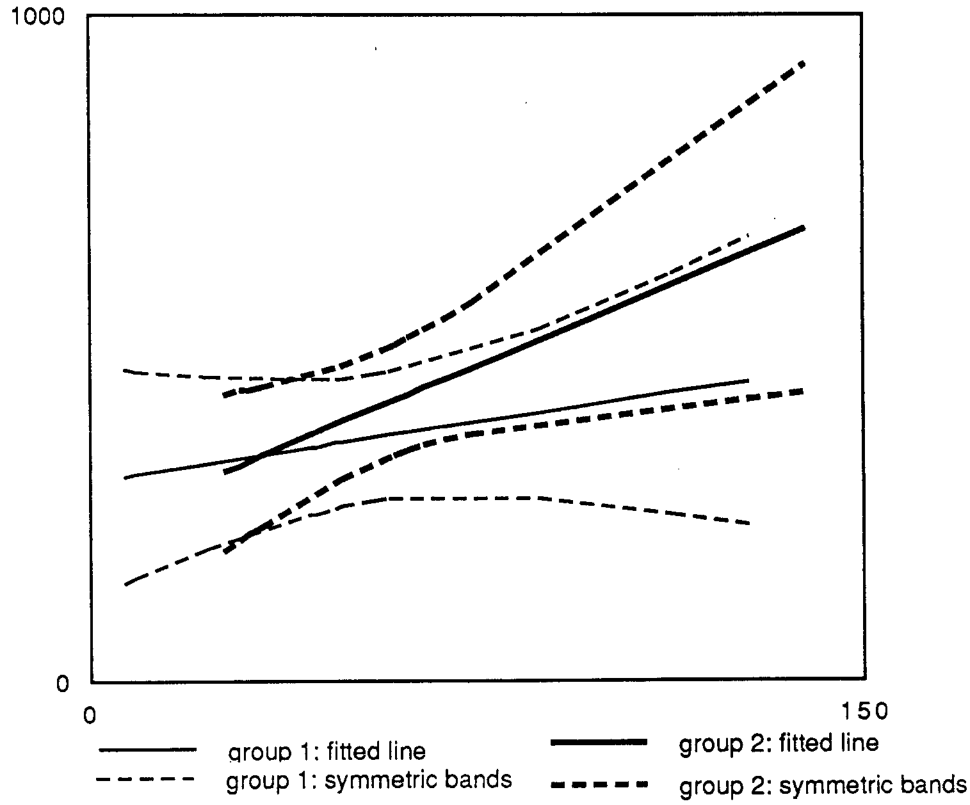

Figure 3 Envelopes of symmetric confidence bands for the data of Example 4.4.

approximation correctly captures the main effects, of size $n^{-1/2}$ and due to skewness, of departures from the assumption of normal errors.

From the fact that

$$\hat{\pi} - \pi = O_p(n^{-1}) \tag{5.1}$$

it is readily proved that the bootstrap estimates $\hat{u}_+, \hat{u}_-$ of solutions u_+, u_- to the equation $\pi(u_+, u_-) = 1 - \alpha$, differ from the true u_+ and u_- only in terms of order $n^{-3/2}$. To see how this is done, let us assume for the sake of definiteness that we are constructing bands for the "symmetric" case, where $u_+ = u_-$. It may be deduced from the definition of Δ that the process $n^{1/2}\Delta(x)$ is asymptotically distributed as $Z_1 + \sigma_x^{-1}(x - \bar{x})Z_2$, where Z_1 and Z_2 are independent standard normal random variables. Put

$$G_n(v) = P\{-vf_+(x) \leqq n^{1/2}\Delta(x) \leqq vf_-(x), \text{ all } x \in I\},$$

$$G_{\infty}(v)=P\{-vf_{+}(x)\leqq Z_1+\sigma_x^{-1}(x-\bar{x})Z_2\leqq vf_{-}(x),\text{ all } x\in I\},$$

$$=\lim_{n\to\infty} G_n(v).$$

Then

$$\pi(u-\delta,u-\delta)=G_n\{n^{1/2}(u-\delta)\}=G_n(n^{1/2}u)-(n^{1/2}\delta)G_n'(n^{1/2}u)+O\{(n^{1/2}\delta)^2\}$$

$$=\pi(u,u)-(n^{1/2}\delta)G_{\infty}'(n^{1/2}u)+o(n^{1/2}\delta).$$

(These formulae are readily justified by Edgeworth expansions of the type which we shall give shortly.) Suppose $\hat{u}(\alpha)$ and $u(\alpha)=\hat{u}(\alpha)-\delta$ are chosen such that

$$\hat{\pi}(\hat{u}(\alpha),\hat{u}(\alpha))=\pi(u(\alpha),u(\alpha))=1-\alpha,$$

and let $v_{\infty}(\alpha)$ be the solution of the equation $G_{\infty}(v_{\infty}(\alpha))=1-\alpha$. Then $n^{1/2}\hat{u}(\alpha)$, $n^{1/2}u(\alpha)$ both converge to $v_{\infty}(\alpha)$. Furthermore,

$$O=\hat{\pi}(\hat{u}(\alpha),\hat{u}(\alpha))-\pi(u(\alpha),u(\alpha))$$

$$=\hat{\pi}(\hat{u}(\alpha),\hat{u}(\alpha))-\pi(\hat{u}(\alpha),\hat{u}(\alpha))+n^{1/2}\delta G_{\infty}'(n^{1/2}u(\alpha))+o_p(n^{1/2}\delta),$$

whence

$$\delta=\hat{u}(\alpha)-u(\alpha)\sim n^{-1/2}G_{\infty}'(v_{\infty}(\alpha))^{-1}\{\pi(\hat{u}(\alpha),\hat{u}(\alpha))-\hat{\pi}(\hat{u}(\alpha),\hat{u}(\alpha))\}$$

$$=O_p(n^{-3/2}),$$

as had to be shown.

This analysis also demonstrates that the curves $\hat{\mathscr{C}}_{+},\hat{\mathscr{C}}_{-}$, defining the envelope of the bootstrap confidence band and given by formula (2.6), are order $n^{-1/2}$ apart, since $\hat{u}(\alpha)\sim n^{-1/2}v_{\infty}(\alpha)$. The fact that $\hat{u}(\alpha)-u(\alpha)=O_p(n^{-3/2})$ shows that $\hat{\mathscr{C}}_{+}$ and $\mathscr{C}_{+}$ (and also $\hat{\mathscr{C}}_{-}$ and $\mathscr{C}_{-}$) are $O_p(n^{-3/2})$ apart.

We shall derive (5.1) by first developing an Edgeworth expansion of the probability π, then deriving an analogous formula for $\hat{\pi}$, and finally, subtracting these two expressions. Note that the value of π is invariant under changes of the scale σ; this is crucial to our argument and is a consequence of the fact that we studentized when constructing the statistic Δ, so as to make it pivotal. We may assume without loss of generality that $\sigma^2=1$.

To expand the probability π, observe first that

$$n^{1/2}\Delta(x)=(S_1+yS_3)(1+n^{-1/2}S_2-n^{-1}S_1^2-n^{-1}S_3^2)^{-1/2},$$

where $y=\sigma_x^{-1}(x-\bar{x})$, $S_1=n^{-1/2}\Sigma e_i$, $S_2=n^{-1/2}\Sigma(e_i^2-1)$, $S_3=n^{-1/2}\sigma_x^{-1}\Sigma(x_i-\bar{x})e_i$. That is, $n^{1/2}\Delta(x)=D(S_1,S_2,S_3;y)$, where

$$D(s_1,s_2,s_3;y)=(s_1+ys_3)(1+n^{-1/2}s_2-n^{-1}s_1^2-n^{-1}s_3^2)^{-1/2}. \tag{5.2}$$

Then $\pi(u_+,u_-)=\int_Q d\mu$, where μ is the joint probability measure of (S_1,S_2,S_3) and

$$Q=\{(s_1,s_2,s_3)\in\mathbb{R}^3:-n^{1/2}u_+f_+(x)\leqq D(s_1,s_2,s_3;\sigma_x^{-1}(x-\bar{x}))$$
$$\leqq n^{1/2}u_-f_-(x),\quad \text{all } x\in I\}.$$

For any practical choice of f_+ and f_-, in particular for the versions f_+ and f_- discussed in Section 2, Q may be expressed in terms of a finite number of unions and intersections of convex sets, the number of such operations being bounded uniformly in all choices $n\geqq 2$ and $u_+,u_->0$. Under mild assumptions on the process generating the data (e.g. that the errors have a nonsingular distribution with finite 8th moment, and that the design variables x_i are either regularly spaced in an interval or come from a continuous distribution on an interval), the measure μ admits a short Edgeworth expansion uniformly on convex sets $C\subseteq\mathbb{R}^3$:

$$\mu(C)=\int_C\{1+n^{-1/2}p(\mathbf{s})\}\phi(\mathbf{s})\,d\mathbf{s}+O(n^{-1}),$$

where ϕ is the density of a trivariate normal distribution having the same mean and covariance structure as $\mathbf{S}=(S_1,S_2,S_3)^T$, and p is an odd, third degree polynomial. The coefficients of p depend on moments of the error distribution up to and including the 6th, and on the design variables x_i, but are bounded as $n\to\infty$. See Hall (1989, p. 268) for an account of results of this type. Combining the results of this paragraph we conclude that

$$\pi(u_+,u_-)=\int_Q\{1+n^{-1/2}p(\mathbf{s})\}\phi(\mathbf{s})\,ds+O(n^{-1}) \tag{5.3}$$

uniformly in $u_+,u_->0$.

The function ϕ appearing in (5.4) is the density of the normal distribution with zero mean and variance matrix

$$\Sigma=\begin{bmatrix}1 & \mu_3 & 0\\ \mu_3 & \mu_4-1 & 0\\ 0 & 0 & 1\end{bmatrix}. \tag{5.4}$$

where $\mu_j=E(e_i/\sigma)^j$. Indicate this fact by writing ϕ as $\phi(\cdot,\Sigma)$.

Next we develop a bootstrap analogue of (5.3). Let $\hat{p}$ be obtained from p on replacing each appearance of $E(e_i^j)$ by the corresponding sample moment $n^{-1}\Sigma\hat{e}_i^j$ for $j=1,2,\ldots$. Likewise, define $\hat{\mu}_j=n^{-1}\Sigma_i(\hat{e}_i/\hat{\sigma})^j$ and let $\hat{\Sigma}$ be the 3×3 matrix obtained on replacing (μ_3,μ_4) by $(\hat{\mu}_3,\hat{\mu}_4)$ in (5.4). Then we have instead of (5.3)

$$\hat{\pi}(u_+, u_-) = \int_Q \{1 + n^{-1/2}\hat{p}(\mathbf{s})\}\phi(\mathbf{s}, \hat{\Sigma})\, d\mathbf{s} + O_p(n^{-1}) \tag{5.5}$$

with probability one as $n \to \infty$; see Hall (1989, p. 268). Since the sample moments of the residual sequence are within $O_p(n^{-1/2})$ of the respective population moments of the errors then $\hat{p} - p = O_p(n^{-1/2})$ and $\hat{\Sigma} - \Sigma = O_p(n^{-1/2})$. We may now deduce from (5.5) that

$$\hat{\pi}(u_+, u_-) = \int_Q \phi(\mathbf{s}, \hat{\Sigma})\, d\mathbf{s} + n^{-1/2} \int_Q \phi(\mathbf{s}, \Sigma) p(\mathbf{s})\, d\mathbf{s} + O_p(n^{-1}).$$

Now, $\phi(\mathbf{s}, \hat{\Sigma})$ and $\phi(\mathbf{s}, \Sigma)$ differ only by $O_p(n^{-1/2})$, and then only in terms involving s_2. Since s_2 enters the formula for $D(s_1, s_2, s_3; y)$ only through the term $n^{-1/2}\ s_2$ appearing in (5.2), then

$$\int_Q \{\phi(\mathbf{s}, \hat{\Sigma}) - \phi(\mathbf{s}, \Sigma)\}\, d\mathbf{s} = O_p(n^{-1/2} n^{-1/2}) = O_p(n^{-1}).$$

Therefore

$$\begin{aligned}\hat{\pi}(u_+, u_-) &= \int_Q \phi(\mathbf{s}, \Sigma)\, d\mathbf{s} + n^{-1/2} \int_Q \phi(\mathbf{s}, \Sigma) p(\mathbf{s})\, d\mathbf{s} + O_p(n^{-1}) \\ &= \pi(u_+, u_-) + O_p(n^{-1}),\end{aligned}$$

the last line coming from (5.3). This establishes the desired result (5.1).

References

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989). *Statistical Modelling in* GLIM. Clarendon Press, Oxford.

Bohrer, R. and Francis, G. K. (1972). Sharp one-sided confidence bands for linear regression over intervals. *Biometrika* **59,** 99–107.

Bowden, D. C. and Graybill, F. A. (1966). Confidence bands of uniform and proportional width for linear models. *J. Amer. Statist. Assoc.* **61,** 182–198.

Cox, D. R. and Snell, E. J. (1981). *Applied Statistics: Principles and Examples.* Chapman and Hall, New York.

Gafarian, A. V. (1964). Confidence bands in straight-line regression. *J. Amer. Statist. Assoc.* **59,** 182–213.

Gafarian, A. V. (1978). Confidence bands in multivariate polynomial regression. *Technometrics* **20,** 141–150.

Gleason, J. (1988). Algorithms for balanced bootstrap simulations. *The American Statistician* **42,** 4, 263–266.

Graybill, F. A. and Bowden, D. C. (1967). Linear segment confidence bands for simple linear models. *J. Amer. Statist. Assoc.* **62,** 403–408.

Hall, P. (1989). Unusual properties of bootstrap confidence intervals in regression problems. *Probab. Th. Related Fields* **81,** 247–273.

Halperin, M. and Gurian, J. (1971). A note on estimation in straight-line regression when both variables are subject to error. *J. Amer. Statist. Assoc.* **66,** 587–589.

Halperin, M., Rastogi, S. C., Ho, I. and Yang, Y. Y. (1967). Shorter confidence bands in linear regression. *J. Amer. Statist. Assoc.* **62,** 1050–1067.

Kendall, M. G. and Taylor, A. (1979). *The Advanced Theory of Statistics*, Vol. 2, 4th edn. Griffin, London.

Naimann, D. Q. (1984). Average width optimality of simultaneous confidence bounds. *Ann. Statist.* **12,** 1199–1214.

Piegorsch, W. W. (1984). Admissible and optimal confidence bands in linear regression. Ph.D. Thesis, Cornell University.

Piegorsch, W. W. (1985). Admissible and optimal confidence bands in simple linear regression. *Ann. Statist.* **13,** 801–810.

Seber, G. A. F. (1977). *Linear Regression Analysis.* Wiley, New York.

Uusipaikka, E. (1983). Exact confidence bands for linear regression over intervals. *J. Amer. Statist. Assoc.* **78,** 638–644.

Working, H. and Hotelling, H. (1929). Application of the theory of error to the interpretation of trends. *J. Amer. Statist. Assoc. Suppl.* **24,** 73–85.

Wynn, H. P. (1984). An exact confidence band for one-dimensional polynomial regression. *Biometrika* **71,** 375–380.

Wynn, H. P. and Bloomfield P. (1971). Simultaneous confidence bounds in regression analysis. *J. Roy. Statist. Soc. Ser. B* **33,** 202–217.