

**ON THE BOOTSTRAP  
IN MISSPECIFIED  
REGRESSION MODELS**

Santiago Velilla

96-19



WORKING PAPERS

Working Paper 96-19  
Statistics and Econometrics Series 06  
March 1996

Departamento de Estadística y Econometría  
Universidad Carlos III de Madrid  
Calle Madrid, 126  
28903 Getafe (Spain)  
Fax (341) 624-9849

## ON THE BOOTSTRAP IN MISSPECIFIED REGRESSION MODELS

Santiago Velilla\*

### Abstract

---

A resampling method is introduced to approximate the asymptotic distribution of the least squares estimate when the fitted regression model is misspecified. Both theoretical and applied aspects are studied.

---

### Key Words

Asymptotic inference; Confidence ellipsoids; Least squares estimation; Mallows distance; Resampling.

\* Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903 - Getafe, Madrid (Spain). Research partially supported by DGICYT grant PB93-0232 (Spain).

Working Paper 96 - 19

Statistics and Econometric Series (06)

March 1996

## ON THE BOOTSTRAP IN MISSPECIFIED REGRESSION MODELS

Santiago Velilla \*

---

*Abstract:* A resampling method is introduced to approximate the asymptotic distribution of the least squares estimate when the fitted regression model is misspecified. Both theoretical and applied aspects are studied.

---

*Keywords:* Asymptotic inference; Confidence ellipsoids; Least squares estimation; Mallows distance; Resampling.

\* Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903 - Getafe, Madrid (Spain). Research partially supported by DGICYT grant PB93 - 0232 (Spain).

## 1. INTRODUCTION

Consider the linear model

$$Y = X\beta + \varepsilon, \quad (1.1)$$

where  $Y$  is an  $n \times 1$  vector of responses,  $X$  an  $n \times p$  known full rank matrix of constants and  $\varepsilon$  an  $n \times 1$  vector of errors. Suppose that while (1.1) is fitted, the true model instead is

$$Y = X\beta + C\delta_n + \varepsilon, \quad (1.2)$$

where  $C$  is an  $n \times q$  constant matrix whose columns are linearly independent of the columns of  $X$ ,  $\delta_n = \theta n^{-1/2}$ ,  $\theta$  is of  $q \times 1$ , and the errors in  $\varepsilon$  are independent and identically distributed with zero mean and constant variance  $\sigma^2$ . The sequence of models (1.2) is introduced in McKean, Sheather and Hettmansperger (1993). This paper describes a resampling method for studying the effect on the least squares estimator  $\hat{\beta}_n$  of misspecifications of the form (1.2). Background and motivation are given in section 2, while main theoretical results are exposed in section 3. Simulations and applications are presented in sections 4 and 5. Section 6 is devoted to some final comments. For related work of bootstrap in regression, see Efron (1979), Freedman (1981), Wu (1986), Hall (1992), Léger, Politis and Romano (1992) and Efron and Tibshirani (1993, chap. 9).

## 2. RESAMPLING SCHEMES

If model (1.1) is fitted but model (1.2) is the truth, the least squares estimator of  $\beta$  is  $\hat{\beta}_n = (X'X)^{-1}X'Y = \beta + (X'X)^{-1}X'C\delta_n + (X'X)^{-1}X'\varepsilon$  and, therefore,

$$\begin{aligned} n^{1/2}(\hat{\beta}_n - \beta) &= (X'X)^{-1}X'C\theta + n^{1/2}(X'X)^{-1}X'\varepsilon \\ &= \alpha_n + \xi_n. \end{aligned} \quad (2.1)$$

If the  $n \times (p+q)$  matrix  $(X|C)$  is such that  $n^{-1}(X|C)'(X|C)$  converges to a positive definite  $(p+q) \times (p+q)$  matrix, standard asymptotic results in linear regression (see e.g. Freedman (1981)) imply that the limit distribution of (2.1) is

$$N_p[V^{-1}Z\theta; \sigma^2V^{-1}], \quad (2.2)$$

where  $V$  and  $Z$  are, respectively, the limits of  $X'X/n$  and  $X'C/n$ . Two resampling schemes are considered to approximate the asymptotic distribution (2.2).

**2.1 Resampling scheme when both  $C$  and  $\delta_n$  are known.** Consider the residuals of the fit of model (1.1), namely

$$e = (e_i) = Y - X\hat{\beta}_n = (I_n - H)Y = (I_n - H)(C\delta_n + \varepsilon), \quad (2.3)$$

where  $H = X(X'X)^{-1}X'$  is the "hat" matrix, and let  $G_n$  be the empirical distribution function of the  $(e_i)$  centered at its average  $\bar{e} = \sum_{i=1}^n e_i/n$ . As

described in Freedman (1981), when (1.1) is the correct model the standard bootstrap is constructed resampling on a random vector  $\varepsilon^* = (\varepsilon_1^*, \dots, \varepsilon_n^*)'$  of conditionally independent random variables with common distribution  $G_n$ . However, if the true model is (1.2), the bias in the mean function must be incorporated into the resampling scheme in order to approximate the asymptotic distribution (2.2). Specifically, define the starred data

$$Y^* = X\hat{\beta}_n + T^*, \quad (2.4)$$

where, conditionally on  $Y$ ,  $T^*$  is distributed as  $\varepsilon^* + C\delta_n$ . If  $\hat{\beta}_n^* = (X'X)^{-1}X'Y^*$ , the bootstrap principle holds if the conditional law of  $n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n)$  is close, when  $n$  is large, to the law of  $n^{1/2}(\hat{\beta}_n - \beta)$ . Section 3 below shows that this is the case here. Observe that

$$\begin{aligned} n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n) &= (X'X)^{-1}X'C\theta + n^{1/2}(X'X)^{-1}X'\varepsilon^* \\ &= \alpha_n + \xi_n^*, \end{aligned} \quad (2.5)$$

an structure that resembles (2.1). As for the question of variance

estimation, use the starred residuals

$$\mathbf{e}^* = \mathbf{Y}^* - \mathbf{X}\hat{\beta}_n^* = (\mathbf{I}_n - \mathbf{H})\mathbf{Y}^* = (\mathbf{I}_n - \mathbf{H})(\mathbf{C}\delta_n^* + \boldsymbol{\varepsilon}^*), \quad (2.6)$$

and consider the estimate

$$s^{*2} = \mathbf{e}^{*'}\mathbf{e}^*/(n-p), \quad (2.7)$$

the resampling analogue of the mean squared error  $s^2 = \mathbf{e}'\mathbf{e}/(n-p)$  in (1.1).

The conditional mean of  $s^{*2}$  is  $n^{-1} \sum_{i=1}^n (e_i - \bar{e})^2 + [\delta_n' \mathbf{C}'(\mathbf{I}_n - \mathbf{H})\mathbf{C}\delta_n]/(n-p)$ , so

$s^{*2}$  contains a bias component that is negligible for large  $n$ .

**2.2 Resampling scheme when  $\mathbf{C}$  is known and  $\delta_n$  unknown.** In general, if  $\mathbf{C}$  is

given, the resampling scheme (2.4) depends on the unknown parameter  $\delta_n$  that

in applications must be replaced by a suitable estimator. The obvious choice

is the least squares estimate of  $\delta_n$

$$\hat{\delta}_n = [\mathbf{C}'(\mathbf{I}_n - \mathbf{H})\mathbf{C}]^{-1}\mathbf{C}'(\mathbf{I}_n - \mathbf{H})\mathbf{Y}, \quad (2.8)$$

where (2.8) follows from the two-step least squares fitting method. (2.4)

has to be modified so that, conditionally on  $\mathbf{Y}$ ,

$$\hat{\mathbf{Y}}^* = \mathbf{X}\hat{\beta}_n^* + \hat{\boldsymbol{\Gamma}}^*, \quad (2.9)$$

where, conditionally on  $\mathbf{Y}$ ,  $\hat{\boldsymbol{\Gamma}}^*$  is distributed as  $\boldsymbol{\varepsilon}^* + \mathbf{C}\hat{\delta}_n$ . Accordingly, (2.6)

and (2.7) must be modified to, respectively,

$$\hat{\mathbf{e}}^* = \hat{\mathbf{Y}}^* - \mathbf{X}\hat{\beta}_n^* = (\mathbf{I}_n - \mathbf{H})\hat{\mathbf{Y}}^* = (\mathbf{I}_n - \mathbf{H})(\mathbf{C}\hat{\delta}_n + \boldsymbol{\varepsilon}^*), \quad (2.10)$$

and

$$\hat{s}^{*2} = \hat{\mathbf{e}}^{*'}\hat{\mathbf{e}}^*/(n-p). \quad (2.11)$$

The resampling estimate is now  $\hat{\beta}_n^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{Y}}^* = \hat{\beta}_n + n^{-1/2}\hat{\alpha}_n + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}^*$ , where  $\hat{\alpha}_n = n^{1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}\hat{\delta}_n$ .

**2.3. Criterion for the size of  $\delta_n$ .** The effectiveness of the resampling

schemes (2.4) and (2.9) depends critically on the size of  $\delta_n$ . It is

desirable a criterion for deciding when  $\delta_n$  is "small" or, in other words,

when is appropriate to model the misspecification component as in (1.2).

Borrowing ideas from standard theory of hypothesis testing in linear models, a natural criterion is

$$U_n = \hat{\delta}_n' [C'(I_n - H)C] \hat{\delta}_n / \hat{\sigma}^2, \quad (2.12)$$

where  $\hat{\sigma}^2 = \text{RSS}/[n - (p+q)]$ , being RSS the residual sum of squares in model (1.2). Small values of  $U_n$  suggest the appropriateness of the family (1.2). If  $\delta_n = 0$  for all  $n$ ,  $U_n$  is asymptotically  $\chi_q^2$ . Observe however that, properly speaking, under the sequence of models (1.2),  $U_n$  converges to a noncentral  $\chi_q^2$  random variable with noncentrality parameter depending on  $\theta$  and  $\sigma$  (see section 3) and, as a result,  $U_n$  is difficult to calibrate.  $U_n$  is a monotonically increasing function of the ratio  $R_n = s^2/\hat{\sigma}^2$  of mean squared errors in models (1.1) and (1.2) respectively. In applications,  $R_n$  is preferred to  $U_n$  because  $E[s^2]/E[\hat{\sigma}^2] = 1 + [\delta_n' C'(I_n - H)C \delta_n]/[\sigma^2(n-p)]$  and, thus, values of  $R_n$  not too far from one support a model of the form (1.2).

### 3. THEORETICAL RESULTS

The distance between two probability distributions  $\mu$  and  $m$  in  $\mathbb{R}^p$  with finite  $i$ th moments ( $1 \leq i < \infty$ ) will be measured with the Mallows distance

$$d_i(\mu, m) = \inf E^{1/i}[\|U - V\|^i],$$

where  $\|\cdot\|$  is the euclidean norm in  $\mathbb{R}^p$  and the infimum is taken over all the pairs of jointly distributed  $p$  dimensional random vectors  $(U, V)$  with marginals  $\mu$  and  $m$ , respectively. For the purposes of this paper, only  $i = 1, 2$  are of interest. Whenever needed, the notation  $d_1(\phi, \psi)$  will be used to denote the distance between the laws of two  $p$  dimensional random vectors  $\phi$  and  $\psi$ . Bickel and Freedman (1981, p. 1281 and ff.) study the properties of the metric  $d_1$ . For example, a necessary and sufficient condition for a sequence of random vectors  $\{\phi_n\}$  to satisfy  $d_1(\phi_n, \phi) \rightarrow 0$  is that both  $\phi_n \xrightarrow{D} \phi$  and  $E[\|\phi_n\|^i] \rightarrow E[\|\phi\|^i]$ . Also, for the case  $i = 2$ , the key

representation below holds

$$d_2^2(\phi, \psi) = d_2^2[\phi - E(\phi), \psi - E(\psi)] + \|E(\phi) - E(\psi)\|^2. \quad (3.1)$$

**3.1 Resampling scheme (2.4).** Comparing (2.1) and (2.5), using identity (3.1), and observing that the conditional mean of  $\varepsilon^*$  is zero,

$$d_2^2[n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n), n^{1/2}(\hat{\beta}_n - \beta)] = d_2^2(\xi_n^*, \xi_n). \quad (3.2)$$

The first argument in the right hand side of (3.2) is the conditional law of  $n^{1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon^*$  given  $\mathbf{Y}$  and the second argument is the unconditional law of  $n^{1/2}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$ . To proof the asymptotic validity of the bootstrap approximation is then just a matter of checking that  $d_2^2(\xi_n^*, \xi_n)$  converges to zero almost surely when  $n$  goes to infinity. This can be done by suitably extending the proof of Freedman (1981) for the least squares estimator assuming that (1.1) is the correct model.

**Theorem 3.1** Under the sequence of models (1.2) and assuming that  $n^{-1}(\mathbf{X}|\mathbf{C})'(\mathbf{X}|\mathbf{C})$  converges to a positive definite  $(p+q) \times (p+q)$  matrix, for almost every sample data  $\mathbf{Y}$ :

- a) The conditional distribution of  $n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n)$  converges weakly to  $N_p[\mathbf{V}^{-1}\mathbf{Z}\theta; \sigma^2\mathbf{V}^{-1}]$ ; and
- b) The conditional distribution of  $s^*$  converges to  $\sigma$ .

**Proof.** Let  $F_n$  be the empirical distribution function of the errors  $\varepsilon_1, \dots, \varepsilon_n$ , and let  $F$  be the distribution of  $\varepsilon_1$ . From lemma 8.9 in Bickel and Freedman (1981),

$$d_2^2(\xi_n^*, \xi_n) \leq n \operatorname{tr}[(\mathbf{X}'\mathbf{X})^{-1}] d_2^2(G_n, F), \quad (3.3)$$

and, since  $d_2$  is a metric,

$$d_2^2(G_n, F) \leq 2[d_2^2(G_n, F_n) + d_2^2(F_n, F)]. \quad (3.4)$$

Each of the two terms in brackets in (3.4) converge to zero almost surely.



The second, by lemma 8.4 of Bickel and Freedman (1981). See the appendix for treatment of the first term as well as for a proof of part b). ■

**3.2 Resampling scheme (2.9).** The asymptotic behaviour of the estimator  $\hat{\delta}_n$  of (2.8) under the sequence of models (1.2) is obtained from the representation  $\hat{\delta}_n = \delta_n + [C'(I_n - H)C]^{-1}C'(I_n - H)\varepsilon$ , lemma 2.3 in Freedman (1981) and the Lindeberg - Feller theorem.

**Proposition 3.2** Under the same hypothesis for the matrix  $(X|C)$  as in theorem 3.1, and assuming that (1.2) is the sequence of true models, the statements below hold as  $n$  goes to infinity:

- a)  $\hat{\delta}_n - \delta_n$  converges to zero almost surely; and
- b)  $n^{1/2}(\hat{\delta}_n - \delta_n)$  converges in distribution to  $N_q[0, \sigma^2(W - Z'V^{-1}Z)^{-1}]$ ,

where  $W$  is the positive definite limit of  $C'C/n$ .

From part b) of proposition 3.2 and (3.1),

$$\begin{aligned} d_2^2[n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n), n^{1/2}(\hat{\beta}_n - \beta)] &= \|\hat{\alpha}_n - \alpha_n\|^2 + d_2^2(\xi_n^*, \xi_n) \\ &= O_p(1) + d_2^2(\xi_n^*, \xi_n), \end{aligned} \quad (3.5)$$

so by (3.3) and (3.4), the bootstrap approximation to the asymptotic distribution (2.2) is correct in this case up to an  $O_p(1)$  term. However, if for a given  $n$  interest lies in approximating the unconditional law of  $\hat{\beta}_n - \beta$ , by (3.5) and the scaling properties of the metric  $d_2$ ,

$$\begin{aligned} d_2^2[\hat{\beta}_n^* - \hat{\beta}_n, \hat{\beta}_n - \beta] &= n^{-1}d_2^2[n^{1/2}(\hat{\beta}_n^* - \hat{\beta}_n), n^{1/2}(\hat{\beta}_n - \beta)] \\ &= n^{-1}\|\hat{\alpha}_n - \alpha_n\|^2 + n^{-1}d_2^2(\xi_n^*, \xi_n) \longrightarrow 0, \end{aligned} \quad (3.6)$$

almost surely. (3.6) follows from part a) of proposition 3.2, the regularity conditions on the matrix  $(X|C)$  and the fact that  $n^{-1}\|\hat{\alpha}_n - \alpha_n\|^2$  is a quadratic form in the coordinates of the random vector  $\hat{\delta}_n - \delta_n$ . As for the variance estimate  $\hat{S}^{*2}$ , the proof of part b) of theorem 3.1 can be suitably modified to obtain  $\hat{S}^* \longrightarrow \sigma$  for almost all samples  $Y$ . Therefore, the

conditional law of  $(\hat{\beta}_n^* - \hat{\beta}_n)/\hat{s}^*$  should approximate the law of  $(\hat{\beta}_n - \beta)/s$ , where, as introduced in section 2 above,  $s^2 = \mathbf{e}'\mathbf{e}/(n-p)$ .

**3.3 Criterion  $U_n$ .** From part b) of proposition 3.2,  $n^{1/2}\hat{\delta}_n^{\Delta}$  converges in distribution to  $\mathbf{Z} \sim N_q[\theta, \sigma^2(\mathbf{W} - \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})^{-1}]$ . Therefore,  $U_n$  converges in distribution to a noncentral  $\chi_q^2(\lambda)$  random variable with non centrality parameter  $\lambda = \theta'(\mathbf{W} - \mathbf{Z}'\mathbf{V}^{-1}\mathbf{Z})\theta/\sigma^2$ .

## 4. SMALL SAMPLE PERFORMANCE

This section studies the performance of the resampling methods of section 2 in two simulated data situations.

**4.1. Linear regression with a quadratic bias term.** Consider fitting a simple linear regression model

$$E[y] = \beta_0 + \beta_1 x, \quad (4.1)$$

to  $n$  data points  $(y_i, x_i)'$  generated independently from the quadratic relation

$$y = 1 + x + \delta x^2 + \varepsilon, \quad (4.2)$$

where  $\mathbf{x} \sim U(-1,3)$  independent of  $\varepsilon \sim N(0,1)$ . By the law of the large numbers,

$$n^{-1}(\mathbf{X}|\mathbf{C})'(\mathbf{X}|\mathbf{C}) = n^{-1} \begin{pmatrix} n & \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 \\ \sum_{i=1}^n x_i & \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 \\ \sum_{i=1}^n x_i^2 & \sum_{i=1}^n x_i^3 & \sum_{i=1}^n x_i^4 \end{pmatrix} \longrightarrow \begin{pmatrix} 1. & 1. & 7./3. \\ 1. & 7./3. & 5.. \\ 7./3. & 5. & 61./5. \end{pmatrix},$$

so, from (2.2), the asymptotic distribution of  $(\hat{\beta}_1 - 1)/s$ , where  $\hat{\beta}_1$  is the

least squares estimate of  $\beta_1$  in (4.1), is

$$N(2\delta, .75/n). \quad (4.3)$$

For the case  $\delta = .1$  and  $n = 30$ , figures 1.a) and 1.b) are the histograms of  $B = 200$  replications from the bootstrap distributions of  $(\hat{\beta}_1^* - \hat{\beta}_1)/s^*$  (scheme (2.4)), and  $(\hat{\beta}_1^* - \hat{\beta}_1)/\hat{s}^*$  (scheme (2.9) with  $\delta$  estimated by  $\hat{\delta} = .1553$ ). The value of criterion  $R_{30}$  is .9955. Superimposed in both figures are the graphs of the normal densities fitted to the histograms. For the case of figure 1.a) the fitted normal curve has mean .2419 and variance .0353, while for figure 1.b) the mean is .3738 and the variance .0369. Despite the small sample size, notice the closeness of these two distributions to the density  $N(.2, .025)$  obtained from (4.3) replacing  $\delta$  and  $n$  by their true values.

Figure 1.a)

Figure 1.b)

As a complement to figures 1.a) and 1.b), figures 2.a) and 2.b) are the histograms of the replicated values of  $s^*$  and  $\hat{s}^*$  that are centered around one.

Figure 2.a)

Figure 2.b)

**4.2 Polynomial regression.**  $n$  data points  $(y_i, x_i)'$  are generated using the polynomial model

$$y = 1 + x + x^2 + \delta x^3 + \epsilon, \quad (4.4)$$

where  $x \sim U(-1,2)$  and  $\epsilon \sim N(0,1)$ .  $x$  and  $\epsilon$  are independent. The model fitted is

$$E[y] = \beta_0 + \beta_1 x + \beta_2 x^2, \quad (4.5)$$

so the role of  $C$  is taken by the variable  $x^3$ . Analyzing again the structure

of the limit of  $n^{-1}(X|C)'(X|C)$ , is easy to see that, asymptotically, the two centered and scaled slope estimators  $[(\hat{\beta}_1 - 1)/s, (\hat{\beta}_2 - 1)/s]'$  are jointly

$$N_2 I \begin{pmatrix} .6 \\ .15 \end{pmatrix} \delta ; \begin{pmatrix} 3.555 & -2.22 \\ - & 2.22 \end{pmatrix} / n. \quad (4.6)$$

In this example,  $\delta = .2$  and  $n = 50$ .

The figures 3.a) and 3.b) below are the scatter plots of  $B = 300$  bootstrap replications from the bivariate bootstrap distributions of  $[(\hat{\beta}_1^* - \hat{\beta}_1)/s^*, (\hat{\beta}_2^* - \hat{\beta}_2)/s^*]'$  (scheme (2.4)) and  $[(\hat{\beta}_1^* - \hat{\beta}_1)/\hat{s}^*, (\hat{\beta}_2^* - \hat{\beta}_2)/\hat{s}^*]'$  (scheme (2.9) with  $\delta$  estimated by  $\hat{\delta} = .2155$ ). The mean vector and sample covariance matrix for the first cloud are, respectively,  $(.0646, .3070)'$  and  $\begin{pmatrix} .1162 & -.0777 \\ - & .0703 \end{pmatrix}$ , while for the second are  $(.0701, .3305)'$  and  $\begin{pmatrix} .1160 & - \\ -.0776 & .0704 \end{pmatrix}$ . Notice the similarity with the nominal theoretical values  $(.1200, .3000)'$  and  $\begin{pmatrix} .0711 & -.0444 \\ - & .0444 \end{pmatrix}$ , obtained from (4.6).

Figure 3.a)

Figure 3.b)

Figures 4.a) and 4.b) are the histograms for the variance estimates  $s^*$  and  $\hat{s}^*$  that again are centered around one.

Figure 4.a)

Figure 4.b)

## 5. AN APPLICATION IN ASYMPTOTIC INFERENCE

Suppose that a linear regression model of the form

$$Y = X\beta + C\delta + \epsilon, \quad (5.1)$$

is fitted to a set of data, where  $X = (\mathbf{x}_1 | \mathbf{x}_2 | \dots | \mathbf{x}_p)$  and  $C = (\mathbf{c}_1 | \mathbf{c}_2 | \dots | \mathbf{c}_q)$ . Interest lies on making inference on the set of  $\beta$  parameters. Assume that the number  $n$  of data points is large enough, that the distribution of

the errors  $\varepsilon$  is not normal and that there is some indication that the parameters  $\delta$  are "small" according to the values of  $U_n$  of (2.12) so that the assumption  $\delta = \delta_n = \theta n^{-1/2}$  is reasonable. If  $q$  is large relative to  $p$ , a possible idea would be trying to make asymptotic inference on  $\beta$  using information from the fit of the simpler and lower dimensional model

$$Y = X\beta + \varepsilon. \quad (5.2)$$

However, by (2.2), the asymptotic distribution of  $n^{1/2}(\hat{\beta}_n - \beta)$  has a non zero mean, a fact that complicates, for example, the construction of symmetric confidence regions for  $\beta$  centered at the least squares estimate  $\hat{\beta}_n$  in (5.2).

In this section, two possible methods for building an approximate joint confidence region for  $\beta$  are compared: *i*) The standard approach based on the least squares estimate of  $\beta$  in the full model (4.7), namely,  $\tilde{\beta}_n = [X'(I_n - Q)X]^{-1}X'(I_n - Q)Y$ , where  $Q = C(C'C)^{-1}C'$  is the projection matrix onto the orthogonal complement of the linear manifold spanned by the columns of  $C$ , and *ii*) A bootstrap method based on  $\hat{\beta}_n$ .

Method *i*) gives the approximate  $(1-\alpha) \times 100\%$  confidence region formed by all the  $\beta$  points inside the ellipsoid

$$(\beta - \tilde{\beta}_n)'E(\beta - \tilde{\beta}_n) \leq 1, \quad (5.3)$$

where  $E = X'(I_n - Q)X/k$ ,  $k = \hat{\sigma}^2 \chi_{p,\alpha}^2$ ,  $\hat{\sigma}^2$  is the least squares mean squared error in (5.1), and  $\chi_{p,\alpha}^2$  is the  $(1-\alpha) \times 100$  percentile of a  $\chi_p^2$  distribution. Ellipsoid (5.3) is thus centered at  $\tilde{\beta}_n$  and has orientation and shape determined by, respectively, the eigenvectors and eigenvalues of the  $p \times p$  matrix  $X'(I_n - Q)X/k$ .

Method *ii*) starts observing that since  $V[\hat{\beta}_n] = \sigma^2(X'X)^{-1}$ , a natural approximate  $(1-\alpha) \times 100\%$  confidence region alternative to (5.3) is of the form

$$(\beta - \hat{\beta}_n)'F(\beta - \hat{\beta}_n) \leq 1, \quad (5.4)$$

where  $F = X'X/m$ ,  $m = c_{(1-\alpha)} s^2$ ,  $s^2$  is the mean squared error in (5.2), and

$c_{(1-\alpha)}$  is an appropriate boundary constant. Ellipsoid (5.4) is centered at  $\hat{\beta}_n$  and its orientation and shape are given by the eigenvectors and eigenvalues of  $X'X/m$ . By (2.2),  $n^{1/2}(\hat{\beta}_n - \beta)/s$  is approximately  $N_p[(X'X)^{-1}X'C\theta/\sigma; (X'X/n)^{-1}]$  and as a consequence, since the parameters  $\theta$  and  $\sigma$  are unknown, the constant  $c_{(1-\alpha)}$  depends on the quantiles of a nonstandard distribution. A suitable value for  $c_{(1-\alpha)}$  can be found using the results of section 3. Denote by  $\mathbf{P}$  the distribution of  $(\hat{\beta}_n - \beta)/s$  and let  $\mathbf{P}^*$  be the bootstrap distribution of  $(\hat{\beta}_n^* - \hat{\beta}_n)/\hat{s}^*$ , where  $\hat{\beta}_n^*$  and  $\hat{s}^*$  are obtained from the scheme (2.9). Simulate a "large" number  $\mathbf{B}$  of bootstrap samples from  $\mathbf{P}^*$  and obtain a constant  $c_{(1-\alpha)}$  such that  $(1-\alpha)\times 100\%$  of the simulated data satisfy the inequality  $(\hat{\beta}_n^* - \hat{\beta}_n)'(X'X/\hat{s}^{*2})(\hat{\beta}_n^* - \hat{\beta}_n) \leq c_{(1-\alpha)}$ . By the results in section 3 and the law of the large numbers,

$$\begin{aligned} (1-\alpha) &\cong \mathbf{P}^* [(\hat{\beta}_n^* - \hat{\beta}_n)'(X'X/\hat{s}^{*2})(\hat{\beta}_n^* - \hat{\beta}_n) \leq c_{(1-\alpha)}] \\ &\cong \mathbf{P}[(\hat{\beta}_n - \beta)'(X'X/s^2)(\hat{\beta}_n - \beta) \leq c_{(1-\alpha)}], \end{aligned}$$

so  $c_{(1-\alpha)}$  seems a reasonable cut off point to plug into (4.10).

Observe the simplicity of the resampling approach in comparison to the relative numerical complexity involved in (5.3), particularly in obtaining the matrix  $\mathbf{I}_n - \mathbf{Q}$ . The volumes of regions (5.3) and (5.4) are proportional, respectively, to  $|X'(\mathbf{I}_n - \mathbf{Q})X|^{-1/2}$  and  $|X'X|^{-1/2}$ . By a well known result in matrix theory (Rao (1973, p. 70)),  $|X'(\mathbf{I}_n - \mathbf{Q})X| \leq |X'X|$  so it can be reasonably conjectured that the bootstrap method tends to give "smaller" symmetric confidence regions. The ideas above can be easily modified to carry out the construction of confidence regions for linear combinations of the parameters  $\beta$  and, in particular, for parameter subsets.

**EXAMPLE** (Attitude Survey Data). An illustration for the theory above is given by the Attitude Survey Data where, as described in Chatterjee and Price (1991, p. 69 and ff.), a response  $y$ , an overall rating of a supervisor

given by the employees of a large financial institution, is related to six regressors  $x_1, \dots, x_6$ . The partial t statistics indicate that only the coefficients associated to  $x_1$  and  $x_3$  are statistically significant, so a possible model for this data set could be

$$Y = \mathbf{1}_n \beta_0 + X\beta + C\delta_n + \varepsilon, \quad (5.5)$$

where  $\mathbf{1}_n$  is the  $n \times 1$  vector whose coordinates are all one,  $X = (x_1 | x_3)$ ,  $C = (x_2 | x_4 | x_5 | x_6)$ , and  $\delta_n$  is of  $4 \times 1$ . Criterion  $R_n$  takes the value .9628. The normal probability plot of the studentized residuals shows a slight indication of nonnormality of the distribution of the errors. The size of the data set is  $n = 30$ . Interest lies on building a confidence ellipse for the slope parameters  $\beta = (\beta_1, \beta_3)'$  in (4.11). For this example,  $\tilde{\beta}_n = (.5852, .3000)'$ ,  $\hat{\sigma}^2 = 48.2639$ , and  $\chi_{2,.05}^2 = 5.9915$ , so the inner product matrix of the 95% least squares approximate confidence ellipse of (4.9) is

$$E = \begin{pmatrix} 6.9112 & 1.7990 \\ 1.7990 & 7.0821 \end{pmatrix}. \quad (5.6)$$

The eigenvalues and eigenvectors of  $E$  are  $\lambda_1 = 5.1956$ ,  $\lambda_2 = 8.7977$  and  $\gamma_1 = (.7237, -.6901)'$ ,  $\gamma_2 = (.6901, .7237)'$ , respectively. The area of the region defined by (5.6) is .4647.

The bootstrap confidence region is based on the fit of the reduced model  $E[Y] = \mathbf{1}_n \beta_0 + X\beta$ . For the Attitude Survey Data,  $\hat{\beta}_n = (.6435, .2112)'$ ,  $\hat{\beta}_0 = 9.8709$ ,  $\hat{\delta}_n = (-.0282, .0769, .0415, -.1864)'$  and  $s^2 = 46.4685$ . Using  $B = 1000$  bootstrap replications from the resampling scheme (2.9) and  $\alpha = .05$ , the value of the boundary constant is  $c_{.95} = 7.1801$ , and therefore,

$$F = \begin{pmatrix} 15.4091 & 8.1056 \\ 8.1056 & 11.9736 \end{pmatrix}, \quad (5.7)$$

The eigenvalues and eigenvectors of  $F$  are  $\lambda_1 = 5.4057$ ,  $\lambda_2 = 21.9770$  and  $\gamma_1 = (.6296, -.7770)'$ ,  $\gamma_2 = (.7770, .6296)'$ , respectively. The area of the elliptical region associated to (5.7) is .2882, roughly 40% less than the

region defined by (5.6). Figure 5 shows ellipses (5.6) and (5.7) in the  $\beta = (\beta_1, \beta_3)'$  plane.

Figure 5

## 6. FINAL COMMENTS

This paper describes a bootstrap technique for approximating the distribution of the least squares estimator when the underlying regression model contains an asymptotically negligible misspecification component. A simulation study analyzes the accuracy of the approximation in some special cases. An application with real data illustrates the use of the ideas of section 3 in the construction of approximate confidence regions for the parameters of interest.

The schemes (2.4) and (2.9) depend on the residuals of the fit of a lower dimensional model (1.1). Therefore, the resampling ideas of section 2 are perhaps conceptually simpler than dealing with a model with a higher number of regressors like for example in the case of the Attitude Survey Data. Recall finally that (2.4) and (2.9) can be effective when, in (1.2), either both  $C$  and  $\delta_n$  are known or when  $C$  is given and  $\delta_n$  is estimated. Finding a resampling scheme that does not depend on specific assumptions on the bias part of the model remains as an open problem.

## ACKNOWLEDGEMENTS

The author is grateful to Ismael Sánchez for helpful computing assistance.



APPENDIX

**Complements to the proof of theorem 3.1. a)** Let  $(U,V)$  be jointly distributed with mass  $1/n$  at  $(e_i, \varepsilon_i)$ ,  $i = 1, \dots, n$ . From (3.1),

$$\begin{aligned} d_2^2(G_n, F_n) &= d_2^2(U - E[U], V) = d_2^2(U - E[U], V - E[V]) + \left( \sum_{i=1}^n \varepsilon_i/n \right)^2 \\ &= d_2^2(U, V) - \left[ \sum_{i=1}^n (e_i - \varepsilon_i)/n \right]^2 + \left( \sum_{i=1}^n \varepsilon_i/n \right)^2 \\ &\leq \sum_{i=1}^n (e_i - \varepsilon_i)^2/n - \left[ \sum_{i=1}^n (e_i - \varepsilon_i)/n \right]^2 + \left( \sum_{i=1}^n \varepsilon_i/n \right)^2. \end{aligned} \quad (A.1)$$

Recalling (2.3), the first summand in (A.1) is  $n^{-1} \|e - \varepsilon\|^2 = n^{-1} \|(\mathbf{I}_n - \mathbf{H})\mathbf{C}\delta_n - \mathbf{H}\varepsilon\|^2 \leq n^{-1} \|\mathbf{C}\delta_n\|^2 + n^{-1} \varepsilon' \mathbf{H} \varepsilon = n^{-1} [\theta' (\mathbf{C}'\mathbf{C}/n) \theta] + n^{-1} \varepsilon' \mathbf{H} \varepsilon$ . The term in brackets is bounded by the assumption on  $\mathbf{C}$  so the first summand converges to zero almost surely. Since, by lemma 2.3 in Freedman (1981)  $n^{-1} \mathbf{X}' \varepsilon \rightarrow 0$  and  $n^{-1} \varepsilon' \mathbf{H} \varepsilon = n^{-1} \varepsilon' \mathbf{X} (\mathbf{X}' \mathbf{X}/n)^{-1} n^{-1} \mathbf{X}' \varepsilon$ , the second goes also to zero. Using Cauchy - Schwartz inequality,

$$\begin{aligned} \left[ \sum_{i=1}^n (e_i - \varepsilon_i)/n \right]^2 &= [(\mathbf{1}_n/n)' (e - \varepsilon)]^2 = [(\mathbf{1}_n/n)' (\mathbf{I}_n - \mathbf{H})\mathbf{C}\delta_n - \mathbf{H}\varepsilon]^2 \\ &\leq n^{-1} \|(\mathbf{I}_n - \mathbf{H})\mathbf{C}\delta_n - \mathbf{H}\varepsilon\|^2 \leq n^{-1} \|\mathbf{C}\delta_n\|^2 + n^{-1} \varepsilon' \mathbf{H} \varepsilon, \end{aligned}$$

so the second summand in (A.1) converges to zero almost surely. The third summand in (A.1) goes to zero by the law of the large numbers, so the proof of a) is finished; b) Let  $q = (n-p)^{-1/2} \|\varepsilon\|$  and  $q^* = (n-p)^{-1/2} \|\varepsilon^*\|$ . By the triangle inequality,

$$d_1(s^*, \sigma) \leq d_1(s^*, q^*) + d_1(q^*, \sigma),$$

so it is enough to show that each of the summands goes to zero as  $n$  goes to infinity. For the first summand,

$$\begin{aligned} d_1^2(s^*, q^*) &\leq (n-p)^{-1} E^2[|\|e^*\| - \|\varepsilon^*\||/Y] \leq (n-p)^{-1} E[\|e^* - \varepsilon^*\|^2/Y] \\ &= [\delta_n' \mathbf{C}' (\mathbf{I}_n - \mathbf{H}) \mathbf{C} \delta_n] / (n-p) + [p/(n-p)] n^{-1} \sum_{i=1}^n (e_i - \bar{e})^2 \\ &= [\delta_n' \mathbf{C}' (\mathbf{I}_n - \mathbf{H}) \mathbf{C} \delta_n] / (n-p) + [p/(n-p)] \hat{\sigma}_n^2, \end{aligned} \quad (A.2)$$

say. The first term above is bounded by  $(n-p)^{-1} \|\mathbf{C}\delta_n\|^2 = (n-p)^{-1} [\theta' (\mathbf{C}'\mathbf{C}/n) \theta]$

$\rightarrow 0$ . If  $\sigma_n^2 = n^{-1} \sum_{i=1}^n (\varepsilon_i - \bar{\varepsilon})^2$ , it can be seen that  $|\hat{\sigma}_n - \sigma_n|^2 \leq n^{-1} \sum_{i=1}^n (e_i - \varepsilon_i)^2 \rightarrow 0$  as in part a) above. Since  $\sigma_n^2 \rightarrow \sigma^2$ ,  $\hat{\sigma}_n^2 \rightarrow \sigma^2$  and, therefore,  $d_1(s, q)$  converges to zero almost surely. From lemma 8.5 of Bickel and Freedman (1981), to get  $d_1(q, \sigma) \rightarrow 0$ , it is enough to prove  $d_1(q^{*2}, \sigma^2) \rightarrow 0$ . But

$$d_1(q^{*2}, \sigma^2) \leq d_1(q^{*2}, q^2) + d_1(q^2, \sigma^2).$$

By the law of the large numbers,  $d_1(q^2, \sigma^2)$  goes to zero. On the other hand, by lemma 8.6 in Bickel and Freedman (1981),  $d_1(q^{*2}, q^2) = [n/(n-p)]d_1(\varepsilon_1^{*2}, \varepsilon_1^2)$ , and this goes to zero as in Freedman (1981, p. 1224). ■

#### REFERENCES

- Bickel, P.J., and Freedman, D.A. (1981), "Some Asymptotic Theory for the Bootstrap", *The Annals of Statistics*, 9, 1196 - 1217.
- Chatterjee, S. and Price, B. (1991), *Regression Analysis By Example*, 2nd Ed., New York: John Wiley.
- Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1 - 26.
- Efron, B., and Tibshirani, R.J. (1993), *An Introduction to the Bootstrap*, New York: Chapman and Hall.
- Freedman, D.A. (1981), "Bootstrapping Regression Models," *The Annals of Statistics*, 9, 1218 - 1228.
- Hall, P. (1992), *The Bootstrap and Edgeworth Expansions*, New York: Springer Verlag.
- Léger, C., Politis, N., and Romano, J.P. (1992), "Bootstrap Technology and Applications," *Technometrics*, 34, 378 - 398.
- McKean, J.W., Sheather, S.J., and Hettmansperger, T.P. (1993), "The Use and Interpretation of Residuals Based on Robust Estimation," *Journal of*

*the American Statistical Association*, **88**, 1254 - 1263.

Rao, C.R.(1973), *Linear Statistical Inference and Its Applications*, 2nd Ed.,  
New York: John Wiley.

Wu, C.F.J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in  
Regression Analysis," *The Annals of Statistics*, **14**, 1261 - 1350.

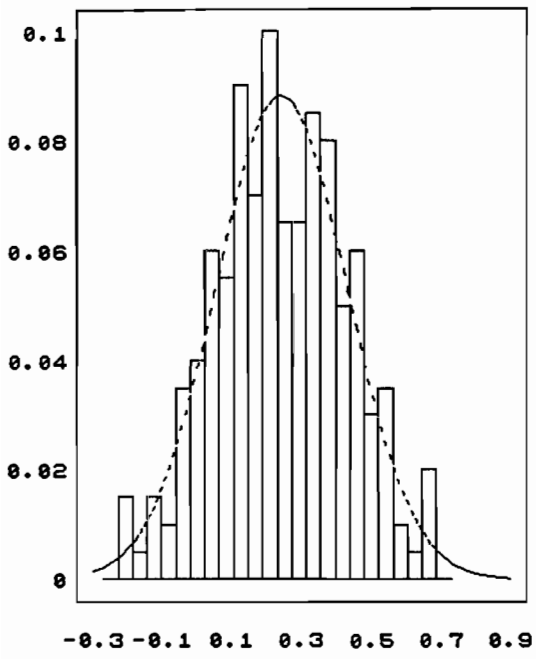


Figure 1.a)

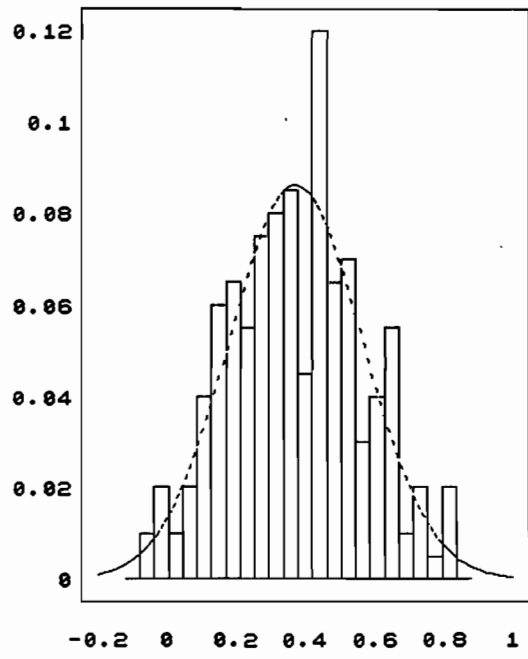


Figure 1.b)

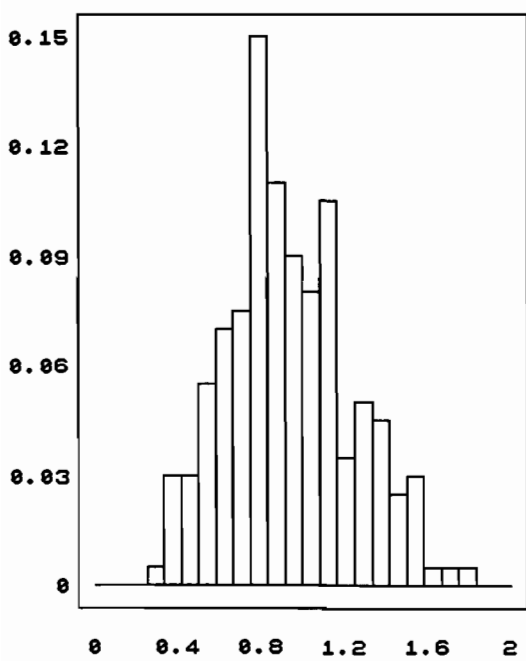


Figure 2.a)

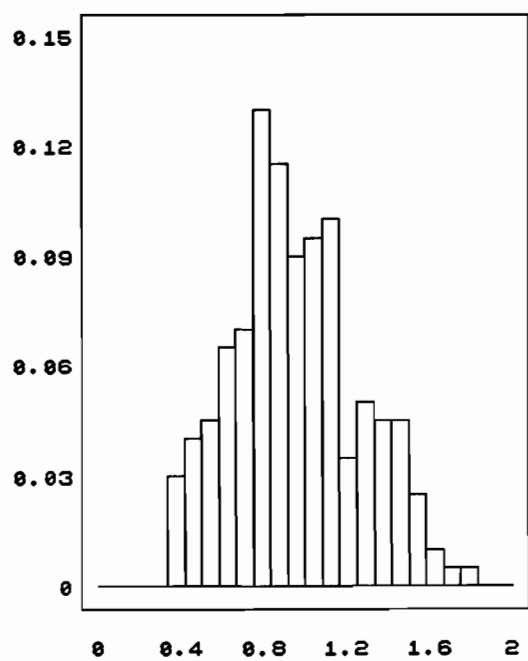


Figure 2.b)

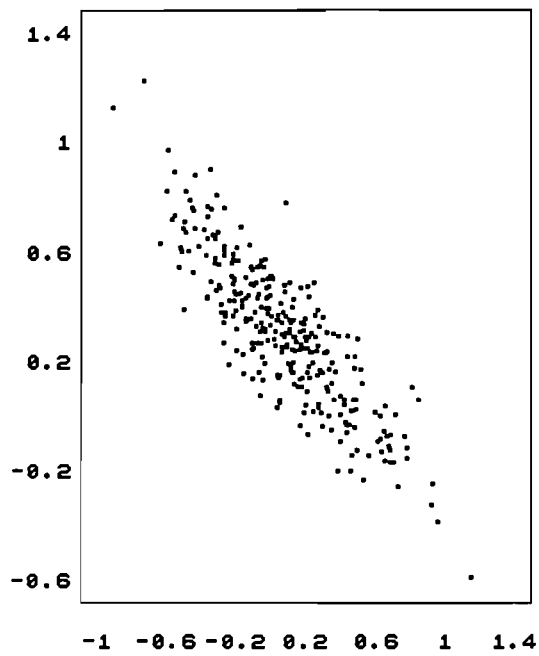


Figure 3.a)

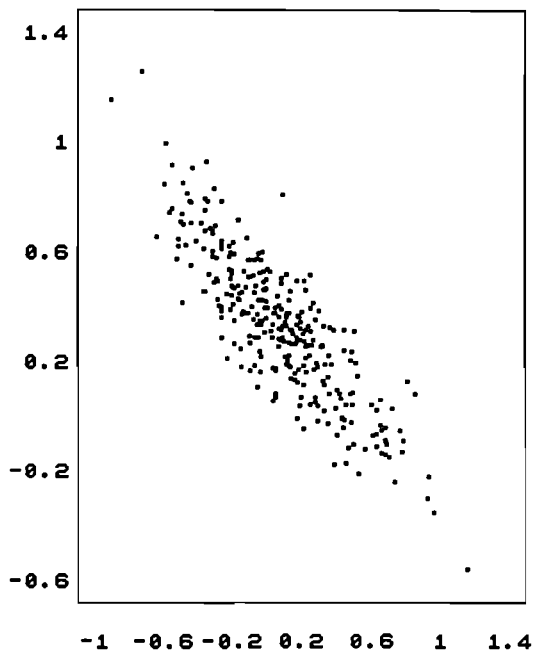


Figure 3.b)

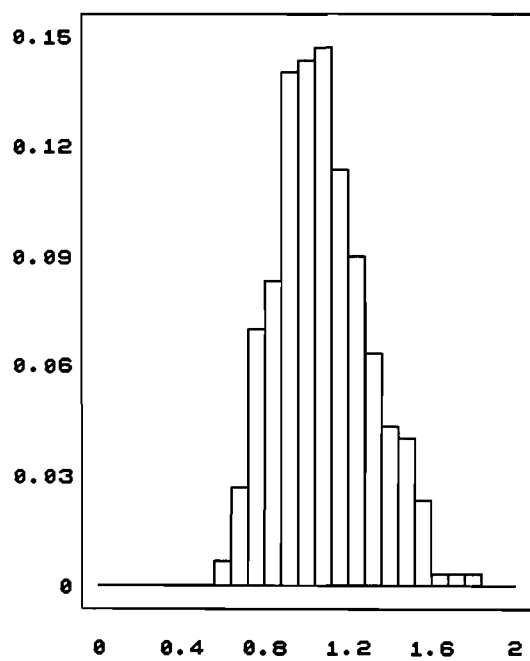


Figure 4.a)

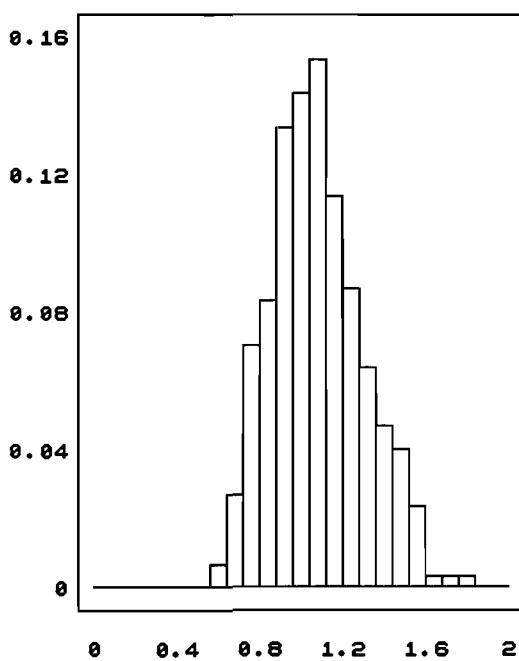
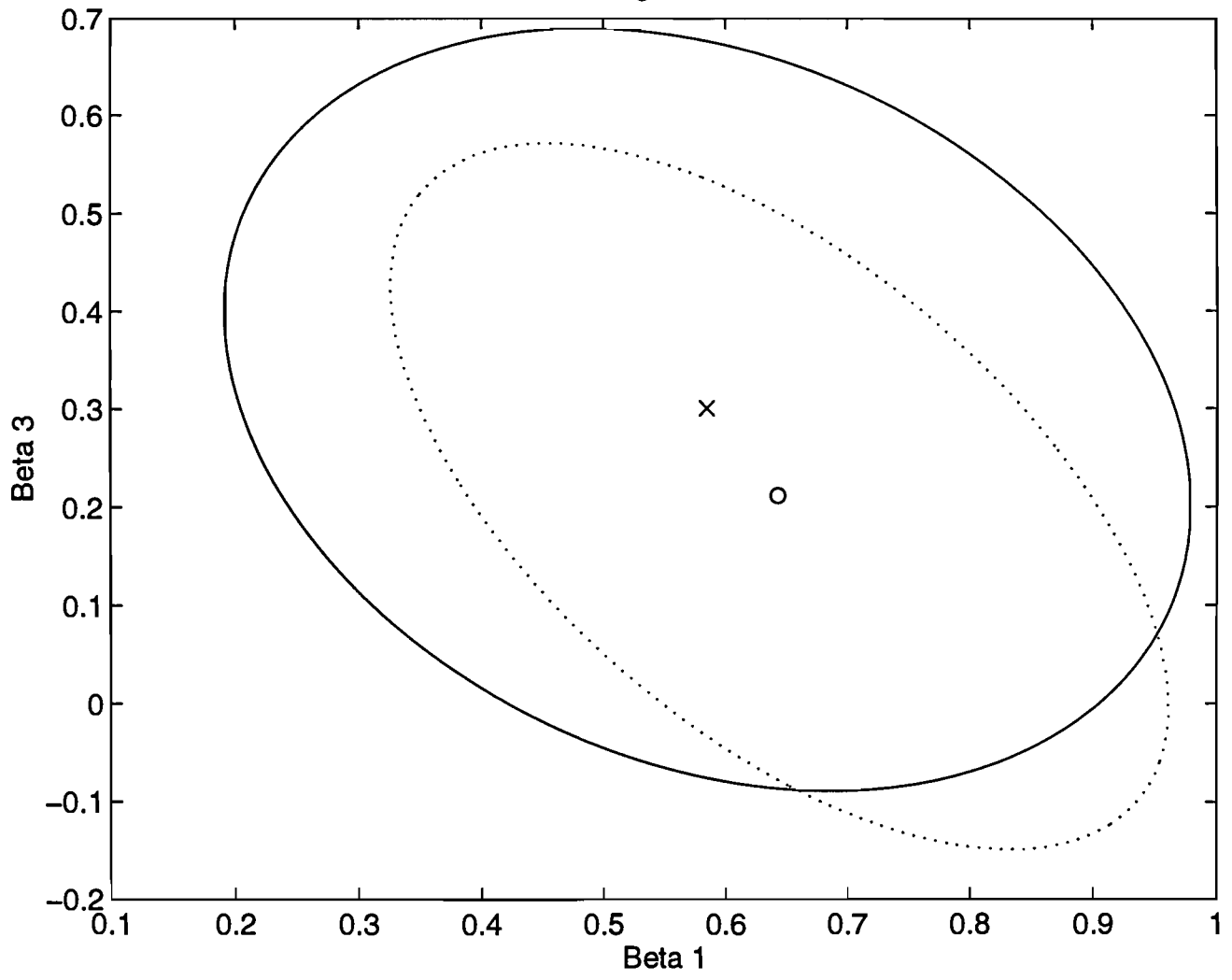


Figure 4.b)

Figure 5



## CAPTIONS FOR FIGURES

**Figure 1.** Simple linear regression: a) Histogram of  $B = 200$  replications from the bootstrap distribution of  $(\hat{\beta}_1^* - \hat{\beta}_1)/s^*$ ; b) Histogram of  $B = 200$  replications from the bootstrap distribution of  $(\hat{\beta}_1^* - \hat{\beta}_1)/\hat{s}^*$ .

**Figure 2.** Simple linear regression: a) Histogram of  $B = 200$  replications from  $s^*$ ; b) Histogram of  $B = 200$  replications from  $\hat{s}^*$ .

**Figure 3.** Polynomial regression: a) Scatter cloud of  $B = 300$  replications from the bootstrap distribution of  $[(\hat{\beta}_1^* - \hat{\beta}_1)/s^*, (\hat{\beta}_2^* - \hat{\beta}_2)/s^*]$ ; b) Scatter cloud of  $B = 300$  replications from the bootstrap distribution of  $[(\hat{\beta}_1^* - \hat{\beta}_1)/\hat{s}^*, (\hat{\beta}_2^* - \hat{\beta}_2)/\hat{s}^*]$ .

**Figure 4.** Polynomial regression: Histogram of  $B = 300$  replications from  $s^*$ ; b) Histogram of  $B = 300$  replications from  $\hat{s}^*$ .

**Figure 5.** 95% joint confidence ellipses for the Attitude Survey Data. Continuous contour: least squares region. Dashed contour: bootstrap region. Center of the least squares region marked with 'x'. Center of the bootstrap region marked with 'o'.