Working Paper 99-72
Statistics and Econometrics Series 25
October 1999

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax 34 - 91- 624.9849

VARIABLE DELETION, CONFIDENCE REGIONS AND BOOTSTRAPPING IN LINEAR REGRESSION

Santiago Velilla*

Abstract

A resampling method is introduced to approximate, when some of the predictors are deleted, the quantiles of the distribution of the usual least squares pivots in linear regression. The approximation is used to construct confidence regions for the parameters of interest of the model.

Key Words

Least squeres estimation; mallows distance; model selection; resampling.

*Department of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain, e-mail: savece@est-econ.uc3m.es; Research partially supported by DGICYT Grant PB96-0111 (Spain).

Variable Deletion, Confidence Regions and Bootstrapping in Linear Regression

Santiago Velilla*

Abstract

A resampling method is introduced to approximate, when some of the predictors are deleted, the quantiles of the distribution of the usual least squares pivots in linear regression. The approximation is used to construct confidence regions for the parameters of interest of the model.

Keywords and phrases: Least squares estimation; Mallows distance; Model selection; Resampling.

1. INTRODUCTION

Consider a linear model of the form

$$\mathbf{Y} = \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \ , \tag{1.1}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of responses, $\mathbf{R}$ is an $n \times M$ full rank matrix of known constants, $\boldsymbol{\gamma}$ is an $M \times 1$ vector of unknown parameters, and $\boldsymbol{\varepsilon} = (\varepsilon_i)$ is an $n \times 1$ vector of i.i.d. errors with zero mean and variance σ^2. If model (1.1) has an intercept, the first column of $\mathbf{R}$ is the vector of ones $\mathbf{1}_n = (1, \overset{(n)}{\ldots}, 1)'$ and the remaining columns correspond to the values of $P = M - 1$ explanatory variables. When P is large, a

*Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903 - Getafe, Madrid, Spain. Research partially supported by $DGICYT$ Grant $PB96-0111$ (Spain).

customary statistical practice is to eliminate some of the carriers in order to gain simplicity and to reduce, at the same time, the cost of handling a, perhaps, large amount of nonessential information. Using a sample-based criterion of variable selection as, for instance, looking for the "best" subset of regressors, applying a stepwise method, ... etc. (see Draper and Smith 1998 chap. 15, or Rawlings, Pantula and Dickey 1998 chap. 7 for excellent updated reviews on variable selection techniques), the design matrix of model (1.1) is partitioned in the form $\mathbf{R} = (\mathbf{X}|\, \mathbf{C})$, where $\mathbf{X}$ is an $n \times m$ matrix that contains the vector of ones and the columns of $\mathbf{R}$ corresponding to p regressors x_1, x_2, ..., x_p considered important, and $\mathbf{C}$ is an $n \times q$ matrix formed with the columns of $\mathbf{R}$ corresponding, in turn, to a group of q nonimportant regressors c_1, c_2, ..., c_q. The orders of the matrices $\mathbf{X}$ and $\mathbf{C}$ are such that $m = 1 + p$ and $P = p + q$. After partitioning $\mathbf{R}$ in the form above, the next step is to write $\boldsymbol{\gamma} = (\boldsymbol{\beta}'\, |\boldsymbol{\delta}'\,)'$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_p)'$ is an $m \times 1$ vector that contains an intercept β_0 and the slope parameters in $\boldsymbol{\gamma}$ corresponding to the variables included in $\mathbf{X}$, and $\boldsymbol{\delta} = (\delta_1, ..., \delta_q)'$ is of $q \times 1$ and contains the remaining slope parameters in $\boldsymbol{\gamma}$. Model (1.1) can be thus reexpressed in the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\delta} + \boldsymbol{\varepsilon}\ , \tag{1.2}$$

where, since the columns of $\mathbf{C}$ are judged of minor importance, all $\delta_j \cong 0$. As a final natural step, and taking into account that the goal is to simplify the model, $\boldsymbol{\delta}$ is set to zero or, equivalently, the variables in $\mathbf{C}$ are deleted. In other words, a reduced or subset model of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}\ + \mathbf{u}\ , \tag{1.3}$$

is considered, where $\mathbf{u}$ is an error term, and the estimator obtained in the least squares fit of (1.3)

$$\widehat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\ , \tag{1.4}$$

is used either for inference purposes on $\boldsymbol{\beta}$ or for predicting the values of future responses.

The reduced model (1.3) is seen to be a reasonable approximation of (1.2), with an additional advantage of simplicity. However, since the importance of the variables in the sample-based criterion might not be same as in the population, some regressors omitted from the model might have associated coefficients δ_j that, although small in magnitude, are not exactly zero. If this is the case, model (1.3) is misspecified and, as a consequence, deletion of variables is well-known to be done at the cost of introducing some bias. To see this, let $\widehat{\sigma}_R^2 = \mathbf{Y}'(\mathbf{I}-\mathbf{H})\mathbf{Y}/(n-m)$ be the least squares estimate of the variance in the fit of the subset model, where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ is the $n \times n$ orthogonal projection matrix onto the column space of $\mathbf{X}$. As summarized in Hocking (1976, sec. 2), if (1.3) is fitted but the full model (1.2) is assumed to be correct, one has

$$E[\widehat{\boldsymbol{\beta}}_R] = \boldsymbol{\beta} + \mathbf{A}\boldsymbol{\delta} \ , \tag{1.5}$$

where $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}$ is the so called alias $m \times q$ matrix, and

$$E[\widehat{\sigma}_R^2] = \sigma^2 + \frac{\|(\mathbf{I}-\mathbf{H})\mathbf{C}\boldsymbol{\delta}\|^2}{n-m} \geq \sigma^2 \ , \tag{1.6}$$

where $\|.\|$ is the euclidean norm. Observe, however, that, regardless of the values of the δ_j, $\widehat{\boldsymbol{\beta}}_R$ is unbiased for $\boldsymbol{\beta}$ whenever the columns of $\mathbf{X}$ are orthogonal to the columns of $\mathbf{C}$. The introduction of bias is counterbalanced, in some circumstances, by an increase in precision in the estimation of $\boldsymbol{\beta}$. Specifically, if for a given partition $\mathbf{R} = (\mathbf{X} | \mathbf{C})$ the least squares estimate of $\boldsymbol{\gamma} = (\boldsymbol{\beta}' | \boldsymbol{\delta}')'$ in (1.2) is written in the form

$$\widehat{\boldsymbol{\gamma}} = \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\delta}} \end{pmatrix} = (\mathbf{R}'\mathbf{R})^{-1}\mathbf{R}'\mathbf{Y} \ , \tag{1.7}$$

where $\widehat{\boldsymbol{\delta}} = [\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{C}]^{-1}\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{Y}$, and if the $q \times q$ matrix

$$\boldsymbol{\Gamma} = Var[\widehat{\boldsymbol{\delta}}] - \boldsymbol{\delta}\boldsymbol{\delta}' = \sigma^2[\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{C}]^{-1} - \boldsymbol{\delta}\boldsymbol{\delta}' \tag{1.8}$$

is positive semi-definite (p.s.d.), the difference $Var[\widehat{\boldsymbol{\beta}}] - MSE[\widehat{\boldsymbol{\beta}}_R]$ is also p.s.d., where $MSE[\widehat{\boldsymbol{\beta}}_R] = E[(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})']$ is the mean squared error matrix of $\widehat{\boldsymbol{\beta}}_R$. Notice that the condition $\boldsymbol{\Gamma} \geq 0$ implies that the true δ_j are, in magnitude, smaller than the standard deviations of their least squares estimates.

One of the aims of this paper is to study, in the same vein as in the paragraph above, the impact of misspecification on the least squares confidence regions for $\boldsymbol{\beta}$ constructed using a subset model as (1.3). Section 2 illustrates that, even for small values of $\boldsymbol{\delta}$, the actual coverage probabilities can be far from their nominal values. If, nevertheless, the reduced model (1.3) is, for inference purposes, preferred to the full model (1.2), section 3 proposes, as a remedial action, a bootstrap type technique for building confidence regions for $\boldsymbol{\beta}$ that is based on the subset estimators $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\sigma}_R^2$. This technique is shown to produce, in section 4, coverage probabilities close to a given nominal value. Section 5 contains some final comments.

2. VARIABLE DELETION AND LEAST SQUARES CONFIDENCE REGIONS BASED ON $\widehat{\boldsymbol{\beta}}_R$

Suppose that interest lies in making inference on a set of linear combinations $\boldsymbol{\Psi} = \mathbf{L}\boldsymbol{\beta}$ defined by a given $s \times m$ matrix $\mathbf{L}$ of rank $r(\mathbf{L}) = s \leq m$. Put $\widehat{\boldsymbol{\Psi}}_R = \mathbf{L}\widehat{\boldsymbol{\beta}}_R$. The effects of underfitting on the confidence regions for $\boldsymbol{\Psi}$ based on $\widehat{\boldsymbol{\beta}}_R$ are studied considering separately the cases of normal and nonnormal errors.

2.1 Normal errors

Put

$$F_{\mathbf{L}} = \frac{Q_{\mathbf{L}}}{s\widehat{\sigma}_R^2} , \tag{2.1}$$

where $Q_{\mathbf{L}} = (\widehat{\boldsymbol{\Psi}}_R - \boldsymbol{\Psi})'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\widehat{\boldsymbol{\Psi}}_R - \boldsymbol{\Psi})$, for the usual least squares pivot for making inferences on $\boldsymbol{\Psi}$ using (1.3). By standard distribution theory in linear regression (see e.g. Seber 1977 sec. 3.4), if the magnitude of $\boldsymbol{\delta}$ is "small", that is if the reduced model is close to the truth, and if the errors are normal, the distribution of $F_{\mathbf{L}}$ is thought to be close to an F distribution with s and $n-m$ degrees of freedom. Therefore, if a $(1-\alpha)\times 100\%$ confidence region is desired for $\boldsymbol{\Psi}$, the deletion approach leads to the elliptical region

$$\{\boldsymbol{\Psi} : F_{\mathbf{L}} \leq F_{s,n-m,\alpha}\}$$

$$= \{\Psi : (\Psi - \widehat{\Psi}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\Psi - \widehat{\Psi}_R) \le s\widehat{\sigma}^2_R F_{s,n-m,\alpha}\} \ , \tag{2.2}$$

where $F_{s,n-m,\alpha}$ is the $(1-\alpha) \times 100\%$ quantile of an F distribution with s and $n-m$ degrees of freedom.

However, as seen in appendix A.1, by application of well-known results relative to the distribution of linear and quadratic forms in random vectors with an spherical normal distribution (see e.g. Arnold 1981, sec. 3.5), if (1.2) is the correct model and $\boldsymbol{\varepsilon} \sim N_n(0, \sigma^2\mathbf{I})$, it can be seen that $Q_{\mathbf{L}}/\sigma^2 \sim \chi^2_s(\boldsymbol{\nu}'\boldsymbol{\nu})$, where $\boldsymbol{\nu} = [\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1/2}\mathbf{L}\mathbf{A}(\boldsymbol{\delta}/\sigma)$, and $(n-m)(\widehat{\sigma}^2_R/\sigma^2) \sim \chi^2_{n-m}(\lambda)$, where $\lambda = [\boldsymbol{\delta}'\mathbf{C}'(\mathbf{I} - \mathbf{H})\mathbf{C}\boldsymbol{\delta}]/\sigma^2$. Moreover, $Q_{\mathbf{L}}$ and $\widehat{\sigma}^2_R$ are independent. In other words, the exact distribution of $F_{\mathbf{L}}$ is given by the ratio

$$F_{\mathbf{L}} \stackrel{D}{=} \frac{\chi^2_s(\boldsymbol{\nu}'\boldsymbol{\nu})/s}{\chi^2_{n-m}(\lambda)/(n-m)} \ , \tag{2.3}$$

of two independent noncentral chi squared distributions scaled by their corresponding degrees of freedom. The right hand side of (2.3) defines a nonpivotal distribution that depends on unknown parameters and that clearly differs from a $F_{s,n-m}$. In particular, the coverage probability of region (2.2) is a function of the ratio $\boldsymbol{\delta}/\sigma$ as in the following simple example.

Example 1 Suppose that, in the notation of (1.2), the matrices $\mathbf{X}$ and $\mathbf{C}$ are

$$\begin{aligned} \mathbf{X}' &= \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -3 & -2 & -1 & 0 & 1 & 2 & 3 \end{pmatrix} \\ \mathbf{C}' &= \begin{pmatrix} -2 & -4 & -2 & 0 & 2 & 4 & 2 \end{pmatrix} , \end{aligned} \tag{2.4}$$

so that $\boldsymbol{\beta} = (\beta_0, \beta_1)'$ and δ is a real number. If a 95% confidence interval based on $\widehat{\boldsymbol{\beta}}_R = (\widehat{\beta}_{0,R}, \widehat{\beta}_{1,R})'$ is desired for β_1, the region (2.2) reduces to the confidence interval

$$\widehat{\beta}_{1,R} \pm \widehat{\sigma}_R \sqrt{x^{11}} t_{5,.025} \ , \tag{2.5}$$

where $\widehat{\sigma}_R$ is the positive squared root of $\widehat{\sigma}^2_R$, x^{11} is the appropriate diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$, and $t_{5,.025}$ is the upper .975 quantile of a central Student's t distribution with five degrees of freedom. To analyze the dependence on δ/σ of the coverage probability of interval (2.5), an experiment is conducted simulating, for every point in a grid of values of δ in the interval $[-.3, .3]$, $N = 2000$ independent replications of an structure of the form

$$\mathbf{Y} = \mathbf{X} \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mathbf{C}\delta + \boldsymbol{\varepsilon} \ , \tag{2.6}$$

where the components ε_i, $i = 1, \ldots, 7$, of $\boldsymbol{\varepsilon}$ are i.i.d. $N(0, 1)$. The variance σ^2 of the errors is taken to be one so that the coverage probability of (2.5) depends solely on δ. Also, the range of values of δ is selected such that the matrix $\boldsymbol{\Gamma}$ of (1.8) is p.s.d. or, equivalently, such that the difference $[\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{C}]^{-1} - \delta^2$ is nonnegative. For the specific values of $\mathbf{X}$ and $\mathbf{C}$ given above in (2.4), this requires $|\delta| \leq 1/\sqrt{\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{C}} = .2958 \cong .3$. Notice that this is precisely the range of values of δ that lead, when the column $\mathbf{C}$ is dropped out from the model, to an increasing precision in the estimation of $\boldsymbol{\beta}$.

For each simulated sample, the interval $\widehat{\beta}_{1,R} \pm \widehat{\sigma}_R\sqrt{x^{11}}t_{5,.025} = \widehat{\beta}_{1,R} \pm .4858\widehat{\sigma}_R$ is computed. For any given value of δ in the grid, the observed proportion of times that the interval (2.5) covers the target value $\beta_1 = 1$ is an estimate of its true coverage probability. As displayed in the symmetrically shaped figure 1, with the exception of the values of δ in a narrow interval around $\delta = 0$, the observed estimates are well below .95. For example, for $\delta = \pm.25$, around one out of five intervals of the form (2.5) do not cover, on average, the target value $\beta_1 = 1$. As a conclusion, even for the range of values of δ that lead to a better estimation of $\boldsymbol{\beta}$, the coverage probabilities of (2.5) can offer a poor approximation of the desired nominal confidence coefficient. ∎

Figure 1

2.2 Nonnormal errors

When the errors are not normal, the asymptotic alternative to (2.2) is the region

$$\{\boldsymbol{\Psi} : Q_{\mathbf{L}}/\widehat{\sigma}_R^2 \leq \chi^2_{s,\alpha}\}$$

$$= \{\Psi : (\Psi - \widehat{\Psi}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\Psi - \widehat{\Psi}_R) \leq \widehat{\sigma}_R^2 \chi^2_{s,\alpha}\} \ , \tag{2.7}$$

where $Q_{\mathbf{L}}$ is as introduced after (2.1), and $\chi^2_{s,\alpha}$ is the upper $(1-\alpha) \times 100\%$ quantile of a χ^2_s distribution. The rationale behind (2.7) can be sketched briefly as follows. From (1.2) and (1.4), one has $\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\delta} + \boldsymbol{\xi}$, where $\boldsymbol{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$ and $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}$ is as in (1.5). If the sequence of matrices $\mathbf{X}'\mathbf{X}/n$ is assumed to converge to a positive definite (p.d.) matrix $\mathbf{V}$, an standard application of the Lindeberg condition leads, when $n \to \infty$, to $\sqrt{n}\boldsymbol{\xi} = \sqrt{n}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \xrightarrow{D} N_m[0 \ , \ \sigma^2\mathbf{V}^{-1}]$ (see, for example, lemma 3.1 in Miller 1974) and, as corollary, to $\mathbf{X}'\boldsymbol{\varepsilon}/n \xrightarrow{P} 0$. Therefore, if the columns of $\mathbf{C}$ are ignored, that is if the reduced model (1.3) is considered approximately correct, $\widehat{\Psi}_R - \Psi = \mathbf{L}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) \cong \mathbf{L}\boldsymbol{\xi} \overset{D}{\cong} N_s[0 \ , \ \sigma^2\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']$ and $\widehat{\sigma}_R^2 \cong \boldsymbol{\varepsilon}'(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}/(n-m) \xrightarrow{P} \sigma^2$. Hence, it is accepted that, for n large enough, the ratio $Q_{\mathbf{L}}/\widehat{\sigma}_R^2 = (Q_{\mathbf{L}}/\sigma^2)/(\widehat{\sigma}_R^2/\sigma^2) \overset{D}{\cong} \chi^2_s$ and, as a consequence, region (2.7) is expected to have a coverage probability close to $(1-\alpha)$.

An alternative asymptotic analysis, taking into account the variables associated to the matrix $\mathbf{C}$ in (1.2), offers a different message. Specifically, if "smallness" of the parameter $\boldsymbol{\delta}$ is formulated by embedding model (1.2) as the *nth* term in the sequence of contiguous models introduced in McKean, Sheather and Hettmansperger (1993, p. 1256)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\delta}_n + \boldsymbol{\varepsilon} \ , \tag{2.8}$$

where all the elements are as defined previously and $\boldsymbol{\delta}_n = \boldsymbol{\theta}/\sqrt{n}$ for some fixed vector $\boldsymbol{\theta}$ of $q \times 1$, and if the regularity condition below is imposed:

(C) As n goes to infinity,

$$\frac{1}{n}\mathbf{R}'\mathbf{R} = \frac{1}{n}\begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{C} \\ \mathbf{C}'\mathbf{X} & \mathbf{C}'\mathbf{C} \end{pmatrix} \to \boldsymbol{\Pi} = \begin{pmatrix} \mathbf{V} & \mathbf{Z} \\ \mathbf{Z}' & \mathbf{W} \end{pmatrix} , \tag{2.9}$$

where $\boldsymbol{\Pi}$ is a finite p.d. $M \times M$ matrix,

it can be seen that the ratio $Q_{\mathbf{L}}/\widehat{\sigma}_R^2$ converges to a noncentral chi squared distribution with noncentrality parameter depending on the parameter vector $\boldsymbol{\theta}/\sigma$. Details can be seen in appendix A.2. As a conclusion, region (2.7) above is using, similarly as region (2.2), the quantile of an incorrect distribution and should not be expected to have coverage probability close to $(1-\alpha)$. This conjecture is confirmed by the simulation results presented in tables 2 and 3 of section 4.

3. VARIABLE DELETION AND BOOTSTRAP CONFIDENCE REGIONS BASED ON $\widehat{\beta}_R$

Section 2 describes some anomalies of the confidence regions for $\mathbf{\Psi} = \mathbf{L}\boldsymbol{\beta}$ based on the fit of the reduced model (1.3). An obvious remedial action would be to replace, both in (2.2) and (2.7), the incorrect quantiles $F_{s,n-m,\alpha}$ and $\chi^2_{s,\alpha}$ by the correct ones. However, and as seen previously, the distribution of the ratio $Q_{\mathbf{L}}/\widehat{\sigma}^2_R$, either exact or asymptotic, depends, under a model of the form (1.2), on unknown parameters and the same is true for its quantiles. Therefore, this approach is of no practical use. As an alternative, and in order to circumvent the parametric dependence of the distributions involved, a resampling technique can be developed to approximate the quantiles of interest.

For reasons of technical convenience, the derivations are presented in terms of the random quantity

$$T_{\mathbf{L}} = \frac{\sqrt{Q_{\mathbf{L}}/n}}{\widehat{\sigma}_R} , \tag{3.1}$$

where $\widehat{\sigma}_R$ is the positive squared root of $\widehat{\sigma}^2_R$. Notice that, if $c_n(\mathbf{L},\alpha)$ is the $(1-\alpha)\times$ 100% quantile of $T_{\mathbf{L}}$, one has, since $Q_{\mathbf{L}}/\widehat{\sigma}^2_R = nT^2_{\mathbf{L}}$,

$$\{\mathbf{\Psi} : T_{\mathbf{L}} \leq c_n(\mathbf{L},\alpha)\} = \{\mathbf{\Psi} : Q_{\mathbf{L}}/\widehat{\sigma}^2_R \leq nc^2_n(\mathbf{L},\alpha)\} , \tag{3.2}$$

so if $c_n(\mathbf{L},\alpha)$ were known, the region on the right-hand side of (3.2) would be a $(1-\alpha)\times 100\%$ confidence region for $\mathbf{\Psi} = \mathbf{L}\boldsymbol{\beta}$ based on statistics computed from the subset model. Unfortunately, $c_n(\mathbf{L},\alpha)$ is difficult to determine. However, as seen next in subsection 3.1, the distribution of $T_{\mathbf{L}}$ admits a tractable bootstrap type approximation that can be used to get an estimate of $c_n(\mathbf{L},\alpha)$. Replacing this estimate for $c_n(\mathbf{L}, \alpha)$ in the right hand side of (3.2), leads to a feasible confidence region for $\boldsymbol{\Psi}$ based on the ratio $Q_{\mathbf{L}}/\widehat{\sigma}_R^2$.

3.1 A bootstrap approximation for the distribution of $T_{\mathbf{L}}$

The first step is to design a data based resampling scheme, based in as much as possible on $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\sigma}_R^2$, that mimics the structure of the misspecified model (1.3) $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where, taking into account (1.2), $\mathbf{u} = \mathbf{C}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$. The natural replacement for $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}}_R$. On the other hand, consider a random vector $\boldsymbol{\varepsilon}^* = (\varepsilon_1^*, ..., \varepsilon_n^*)'$ of conditionally independent components with common distribution G_n, the empirical distribution of the residuals obtained using $\widehat{\boldsymbol{\beta}}_R$

$$\mathbf{e}_R = (e_{i,R}) = \mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_R = (\mathbf{I} - \mathbf{H})\mathbf{Y} , \tag{3.3}$$

where $\mathbf{H}$ is the projection matrix introduced in the second paragraph of section 1. Notice that, if $\mathbf{X}$ has an intercept and the residuals $e_{i,R}$ are rescaled multiplying by the factor $\sqrt{n/(n-m)}$, one has $E^*[\sqrt{n/(n-m)}\,\boldsymbol{\varepsilon}^*] = \mathbf{0}$ and, since $\widehat{\sigma}_R^2 = \|\mathbf{e}_R\|^2/(n-m)$, $Var^*[\sqrt{n/(n-m)}\,\boldsymbol{\varepsilon}^*] = \widehat{\sigma}_R^2 \mathbf{I}_n$, where $E^*[.]$ and $Var^*[.]$ are, respectively, the mean and covariance operators under G_n. Therefore, it seems natural to substitute $\boldsymbol{\varepsilon}$ by $\sqrt{n/(n-m)}\,\boldsymbol{\varepsilon}^*$. Finally, the first summand in $\mathbf{u}$ is replaced by $\mathbf{C}\widehat{\boldsymbol{\delta}}$, where $\widehat{\boldsymbol{\delta}} = [\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{C}]^{-1}\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{Y}$ is, as defined in equation (1.7), the least squares estimate of $\boldsymbol{\delta}$ in the fit of the full model (1.2). This last piece of information cannot be provided by the reduced model and its introduction is motivated by the need of estimating the bias component $\mathbf{C}\boldsymbol{\delta}$.

In summary, the proposed resampling scheme is to generate the starred data

$$\mathbf{Y}^* = \mathbf{X}\widehat{\boldsymbol{\beta}}_R + \mathbf{u}^* , \tag{3.4}$$

where

$$\mathbf{u}^* = \mathbf{C}\widehat{\boldsymbol{\delta}} + \sqrt{n/(n-m)}\,\boldsymbol{\varepsilon}^* . \tag{3.5}$$

(3.4)-(3.5) can be seen as a modification of the standard bootstrap based on residuals introduced by Efron (1979) and studied, among others, in Freedman (1981), Efron and Tibshirani (1993, chap. 9), and in Shao and Tu (1995, chap. 7). Put now $\widehat{\boldsymbol{\beta}}_R^* = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}^*$, $\widehat{\boldsymbol{\Psi}}_R^* = \mathbf{L}\widehat{\boldsymbol{\beta}}_R^*$ and $\widehat{\sigma}_R^{2*} = \mathbf{Y}^{*\prime}(\mathbf{I}-\mathbf{H})\mathbf{Y}^*/(n-m)$ for the bootstrap counterparts of, respectively, $\widehat{\boldsymbol{\beta}}_R$, $\widehat{\boldsymbol{\Psi}}_R = \mathbf{L}\widehat{\boldsymbol{\beta}}_R$, and $\widehat{\sigma}_R^2$. In the same fashion, write

$$T_{\mathbf{L}}^* = \frac{\sqrt{Q_{\mathbf{L}}^*/n}}{\widehat{\sigma}_R^*}\ , \tag{3.6}$$

where $Q_{\mathbf{L}}^* = (\widehat{\boldsymbol{\Psi}}_R^* - \widehat{\boldsymbol{\Psi}}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\widehat{\boldsymbol{\Psi}}_R^* - \widehat{\boldsymbol{\Psi}}_R)$ and $\widehat{\sigma}_R^* = \sqrt{\widehat{\sigma}_R^{2*}}$, for the resampling analogue to $T_{\mathbf{L}}$. The scheme (3.4)-(3.5) can be used to approximate the distribution of $T_{\mathbf{L}}$ by means of the realizable distribution of $T_{\mathbf{L}}^*$. Specifically, consider the distance between two probability distributions μ and π on a given euclidean space $\mathbb{R}^k$ such that both $\int \|\mathbf{x}\|^r \mu(d\mathbf{x})$ and $\int \|\mathbf{x}\|^r \pi(d\mathbf{x})$ are finite:

$$d_r(\mu,\pi) = \inf E^{1/r}\left[\|\mathbf{U}-\mathbf{V}\|^r\right]\ , \tag{3.7}$$

where $r \geq 1$, and the infimum is taken over all the pairs of jointly distributed k dimensional random vectors $(\mathbf{U},\mathbf{V})$ with marginals μ and π respectively. The distance above is introduced in Mallows (1972). A detailed survey of properties of $d_r(\mu,\pi)$ is given in Bickel and Freedman, (1981, sec. 8). The notation $d_r(\boldsymbol{\phi},\boldsymbol{\eta})$ is used sometimes to denote the distance between the laws of two random vectors $\boldsymbol{\phi}$ and $\boldsymbol{\eta}$. The next result is obtained in the same context of the sequence of models (2.8) and the associated regularity condition (C) introduced in subsection 2.2.

Theorem 1 *Under the sequence (2.8), if: i) the errors ε_i are i.i.d. with zero mean and constant variance; and ii) condition (C) holds, then, as $n \to \infty$:*

a) $d_1(\sqrt{Q_{\mathbf{L}}^*/n},\sqrt{Q_{\mathbf{L}}/n}) \to 0$, a.s. ;

b) $d_1(\widehat{\sigma}_R^*,\widehat{\sigma}_R) \to 0$, a.s..

Proof. See appendix B. ■

By part a) of this theorem, the conditional distribution of $\sqrt{Q_{\mathbf{L}}^*/n}$ is a strongly consistent estimate of the distribution of $\sqrt{Q_{\mathbf{L}}/n}$. Moreover, as seen in appendix B, $d_1(\widehat{\sigma}_R,\sigma) \to 0$. Therefore, by the triangle inequality and part b) above, the conditional distribution of $\widehat{\sigma}^*_R$ is close to σ. Theorem 1 justifies then approximating the law of $T_{\mathbf{L}}$ by the conditional law of $T^*_{\mathbf{L}}$ since, by identities (3.1) and (3.6), the numerators are close together and the denominators converge to the same constant σ.

3.2 Bootstrap confidence regions

As a consequence of theorem 1, the $(1-\alpha)\times 100\%$ quantile $c_n(\mathbf{L},\alpha)$ of $T_{\mathbf{L}}$ can be approximated by the corresponding $(1-\alpha)\times 100\%$ upper quantile $c^*_n(\mathbf{L},\alpha)$ of the conditional distribution of $T^*_{\mathbf{L}}$. Moreover, since the distribution of $T^*_{\mathbf{L}}$ is realizable, $c^*_n(\mathbf{L},\alpha)$ can be estimated by Monte Carlo, generating a "large" number B of independent replications $\mathbf{Y}^*_b$, $b = 1, ..., B$, from the scheme (3.4)-(3.5), computing for each generated $\mathbf{Y}^*_b$ the associated value $T^*_{\mathbf{L},b}$ and, finally, finding the constant $c^*_{n,B}(\mathbf{L},\alpha) \cong c^*_n(\mathbf{L},\alpha)$ that covers $(1-\alpha)\times 100\%$ of the B values $T^*_{\mathbf{L},b}$. Replacing, in (3.2), $c_n(\mathbf{L},\alpha)$ by $c^*_{n,B}(\mathbf{L},\alpha)$ leads to the feasible confidence region for $\mathbf{\Psi}$

$$\begin{aligned} E^*_{n,B}(\mathbf{L},\alpha) &= \{\mathbf{\Psi} : T_{\mathbf{L}} \leq c^*_{n,B}(\mathbf{L},\alpha)\} = \{\mathbf{\Psi} : Q_{\mathbf{L}}/\widehat{\sigma}^2_R \leq n c^{*2}_{n,B}(\mathbf{L},\alpha)\} \\ &= \{\mathbf{\Psi} : (\mathbf{\Psi}-\widehat{\mathbf{\Psi}}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\mathbf{\Psi}-\widehat{\mathbf{\Psi}}_R) \leq n\widehat{\sigma}^2_R c^{*2}_{n,B}(\mathbf{L},\alpha)\} \quad (3.8) \end{aligned}$$

(3.8) is, in the notation of Shao and Tu (1995, sec. 4.1), a bootstrap t-type confidence ellipsoid centered at $\widehat{\mathbf{\Psi}}_R = \mathbf{L}\widehat{\boldsymbol{\beta}}_R$ and with shape and boundaries defined, respectively, by the eigenstructure of the matrix $[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}/n\widehat{\sigma}^2_R$ and the estimated quantile $c^*_{n,B}(\mathbf{L},\alpha)$.

4. SIMULATIONS

The regions (3.8) are intended to have approximate coverage probability $(1-\alpha)$. The method is asymptotic in nature and does not depend on the specific distribution for the errors. Its performance in finite sample size situations is analyzed by simulation. Comparisons with the regions (2.2) and (2.7) used by the deletion approach are also addressed.

4.1 Example 1 (continued)

In the example presented in section 2, the region of the family (3.8) corresponding to $\beta_1 = (0\ 1)\boldsymbol{\beta}$, where $\boldsymbol{\beta} = (\beta_0\ \beta_1)'$, is the interval

$$\widehat{\beta}_{1,R} \pm \widehat{\sigma}_R \sqrt{n x^{11}} c^*_{n,B}[(0\ 1),.05]\ . \tag{4.1}$$

It is interesting to compare, in this simple context, the coverage probabilities of this interval with the coverage probabilities obtained with the interval (2.5). Simulating, for each point in the same grid of values of δ in $[-.3,.3]$ considered in example 1, $N = 2000$ replications of a model of the form $\mathbf{Y} = \mathbf{X}(1\ 1)' + \mathbf{C}\delta + \boldsymbol{\varepsilon}$, where the components of $\boldsymbol{\varepsilon}$ are i.i.d. $N(0,1)$, the continuous line in figure 2 represents the empirical coverage probabilities of (4.1) for the target value $\beta_1 = 1$. The quantile $c^*_{n,B}[(0\ 1),.05]$ is obtained using $B = 1000$ bootstrap replications. The bootstrap coverage probabilities are bounded between .9150 and .9865 and, despite of the small sample size, $n = 7$, present a clear improvement over the coverage probabilities provided by the deletion approach that, in turn, are bounded between .7250 and .9475.

Figure 2

4.2 Monte Carlo experiment

To analyze the behaviour of the regions (3.8) in a more complex situation, a higher dimensional Monte Carlo experiment is performed. A random sample $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_{150}$ is generated from a given $N_{10}(\mathbf{0}, \boldsymbol{\Sigma})$ distribution, where the covariance matrix $\boldsymbol{\Sigma}$ has an structure of the form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} ,$$

being $\boldsymbol{\Sigma}_{11} = \mathbf{I}_2$, $\boldsymbol{\Sigma}_{21} = \mathbf{1}_8(1-1)'/\sqrt{2}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{\Sigma}_{21}'$, and $\boldsymbol{\Sigma}_{22} = (1/8)\mathbf{I}_8 + \mathbf{1}_8\mathbf{1}_8'$. The 150×10 data matrix

$$\mathbf{S} = \begin{pmatrix} \mathbf{s}_1' \\ \\ \\ \mathbf{s}_{150}' \end{pmatrix},$$

is formed. For values of $n = 30$, 50, 100, 150, $p = 2$, $q = 2$, 4, 8, and $M = 1+p+q = 3+q$, an $n \times M$ matrix $\mathbf{R}$ is constructed selecting the first n rows and first M columns of the matrix $(\mathbf{1}_{150}|\,\mathbf{S})$. Next, in the notation of section 1, the matrix $\mathbf{R}$ is partitioned in the form $\mathbf{R} = (\mathbf{X}|\,\mathbf{C})$, where $\mathbf{X}$ corresponds to the first $m = 1+p = 3$ columns, and $\mathbf{C}$ to the last q ones. Once a given $n \times M$ matrix $\mathbf{R} = (\mathbf{X}|\,\mathbf{C})$ has been selected, a model of the form

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{C}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \tag{4.2}$$

is considered, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$ is of 3×1, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_q)'$ is of $q \times 1$ and $\boldsymbol{\varepsilon} = (\varepsilon_i)$ is an $n \times 1$ vector of errors. Put $\widehat{\boldsymbol{\beta}}_R = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = (\widehat{\beta}_{0,R}, \widehat{\beta}_{1,R}, \widehat{\beta}_{2,R})'$. Taking $\alpha = .05$, three confidence regions of the form (3.8) are analyzed: *i)* The confidence interval

$$\widehat{\beta}_{1,R} \pm \widehat{\sigma}_R\sqrt{n\mathbf{e}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_1}\,c^*_{n,B}(\mathbf{e}_1',.05)\ , \tag{4.3}$$

for $\beta_1 = \mathbf{e}_1'\boldsymbol{\beta}$, where $\mathbf{e}_1 = (0,1,0)'$; *ii)* The confidence interval

$$(\widehat{\beta}_{1,R} - \widehat{\beta}_{2,R}) \pm \widehat{\sigma}_R\sqrt{n\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}}\,c^*_{n,B}(\boldsymbol{\lambda}',.05)\ , \tag{4.4}$$

for the difference $\beta_1 - \beta_2 = \boldsymbol{\lambda}'\boldsymbol{\beta}$, where, if $\mathbf{e}_2 = (0,0,1)'$, $\boldsymbol{\lambda} = \mathbf{e}_1 - \mathbf{e}_2 = (0,1,-1)'$; and *iii)* The confidence ellipse

$$(\boldsymbol{\Psi} - \widehat{\boldsymbol{\Psi}}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\boldsymbol{\Psi} - \widehat{\boldsymbol{\Psi}}_R) \le n\widehat{\sigma}_R^2 c^{*^2}_{n,B}(\mathbf{L},.05)\ , \tag{4.5}$$

for the parameters $\boldsymbol{\Psi} = \mathbf{L}\boldsymbol{\beta} = (\beta_1,\beta_2)'$, where $\mathbf{L} = (\mathbf{e}_1\ \mathbf{e}_2)'$ and $\widehat{\boldsymbol{\Psi}}_R = \mathbf{L}\widehat{\boldsymbol{\beta}}_R$.

To study the coverage probabilities of regions (4.3), (4.4) and (4.5), $N = 2000$ replications of a model of the form (4.2) are generated taking $\boldsymbol{\beta} = (1,\ 1,\ 1)'$ and $\boldsymbol{\delta} = \boldsymbol{\theta}/\sqrt{n}$, where $\boldsymbol{\theta} = \eta\mathbf{1}_q/\sqrt{q}$ is, for values of $\eta = .5$, 1., and 1.5, proportional to the unit vector $\mathbf{1}_q/\sqrt{q}$. For the i.i.d. errors ε_i with zero mean and variance $\sigma^2 = 1$, three error distributions are considered: i) $N(0,1)$; ii) uniform $U(-\sqrt{3},\sqrt{3})$ and iii) a scaled double exponential distribution with density $f(x) = \exp[-\sqrt{2}\,|x|]/\sqrt{2}$, $-\infty < x < \infty$. For each of the $N = 2000$ replications generated, the quantiles $c^*_{n,B}(\mathbf{L},.05)$ used in (4.3) through (4.5) are computed using $B = 1000$ bootstrap replications. The target values are $\beta_1 = 1$ in (4.3), $\beta_1 - \beta_2 = 0$ in (4.4), and $\mathbf{\Psi} = (1,1)'$ in (4.5). For each region, the proportion of times the corresponding target value is covered, is an empirical estimate of the exact coverage probability. For the three values $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ of $\boldsymbol{\theta}$ considered, corresponding, respectively, to the values of $\eta = .5$, 1. and 1.5, these empirical coverage probabilities are displayed in the columns labeled as E^* in tables 1, 2, and 3.

For completeness, the empirical coverage probabilities obtained with the bootstrap method (3.8) are compared with the coverage probabilities obtained using the regions for β_1, $\beta_1 - \beta_2$ and $\mathbf{\Psi} = (\beta_1,\beta_2)'$ proposed by the deletion approach. When the errors are normal, these regions are, as introduced in (2.2), the interval $\widehat{\beta}_{1,R} \pm \widehat{\sigma}_R \sqrt{\mathbf{e}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_1}\, t_{n-m,.025}$, the interval $(\widehat{\beta}_{1,R} - \widehat{\beta}_{2,R}) \pm \widehat{\sigma}_R \sqrt{\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}}\, t_{n-m,.025}$, and the ellipse $(\mathbf{\Psi} - \widehat{\mathbf{\Psi}}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\mathbf{\Psi} - \widehat{\mathbf{\Psi}}_R) \leq 2\widehat{\sigma}_R^2 F_{2,n-m,.05}$, respectively. For a nonnormal error distribution, the corresponding regions are, as in (2.7), the interval $\widehat{\beta}_{1,R} \pm \widehat{\sigma}_R \sqrt{\mathbf{e}_1'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{e}_1}\, 1.96$, the interval $(\widehat{\beta}_{1,R} - \widehat{\beta}_{2,R}) \pm \widehat{\sigma}_R \sqrt{\boldsymbol{\lambda}'(\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{\lambda}}\, 1.96$ and, finally, the ellipse $(\mathbf{\Psi} - \widehat{\mathbf{\Psi}}_R)'[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1}(\mathbf{\Psi} - \widehat{\mathbf{\Psi}}_R) \leq \widehat{\sigma}_R^2 \chi^2_{2,.05}$, where $\chi^2_{2,.05} = 5.9915$. The empirical coverage probabilities for these regions, computed using a new set of $N = 2000$ independent replications, appear in the columns labeled as D in tables 1, 2, and 3.

Table 1

Table 2

Table 3

According to the results in tables 1, 2, and 3, the regions (4.3), (4.4) and (4.5), have, for the three different error distributions considered, coverage probabilities reasonably closer to the nominal confidence coefficient. The approximation seems to deteriorate both with the size of $\boldsymbol{\theta}$ and with the number q of variables deleted from the model. However, the bootstrap coverage rates have a much better behaviour than the coverage rates obtained with the regions used in the standard deletion approach that, as seen in the columns labeled as D, lie often markedly below .95.

5. FINAL COMMENTS

This paper presents a bootstrap based technique for making inference on the parameters of interest of a linear regression model, once a set of variables considered nonimportant is deleted. The method is asymptotic in nature and does not require any specific distribution for the errors. Implementation of the resampling scheme (3.4)-(3.5), presented in subsection 3.1, depends: *i)* on summary statistics obtained in the fit of the subset model (1.3), and *ii)* on the least squares estimate $\widehat{\boldsymbol{\delta}}$ of $\boldsymbol{\delta}$ obtained in the fit of the full model. This last piece of information is needed to estimate the term $\mathbf{C}\boldsymbol{\delta}$ on the right hand side of (1.2).

The regions (3.8) could be used for making inference on $\boldsymbol{\beta}$ after applying a standard model selection procedure as the ones described in Draper and Smith (1998, chap. 15) or in Rawlings et al. (1998, chap. 7). These methods are useful for partitioning the design matrix $\mathbf{R}$ of the full model (1.1) in the form $\mathbf{R} = (\mathbf{X}\,|\,\mathbf{C})$ so that the role of the variables associated to the columns of the matrix $\mathbf{C}$ is, as compared to the role of the variables associated to the columns of $\mathbf{X}$, of minor importance. In the light of the simulation results presented in section 4, some care is needed before automatically eliminating the variables $\mathbf{C}$ since, unless very small values of $\boldsymbol{\delta}$ are considered, deletion of variables is typically bound to decrease markedly the coverage probabilities of the least squares confidence regions (2.2) or (2.7). If, nevertheless, inference on $\boldsymbol{\beta}$ is performed using the statistics $\widehat{\boldsymbol{\beta}}_R$ and $\widehat{\sigma}^2_R$ computed in the least squares fit of the, perhaps misspecified, reduced model (1.3), the appropriate region of the family (3.8) seems to offer a better behaved coverage probability.

APPENDIX A: EXACT DISTRIBUTION OF $F_{\mathbf{L}}$ UNDER (1.2) AND ASYMPTOTIC DISTRIBUTION OF $Q_{\mathbf{L}}/\widehat{\sigma}_R^2$ UNDER (2.8)

A.1 Exact distribution of $F_{\mathbf{L}}$ under (1.2)

From expressions (1.2) and (1.4), one has the representation $\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\delta} + \boldsymbol{\xi}$, where $\boldsymbol{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$ and $\mathbf{A} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{C}$ is as in (1.5). Therefore,

$$\widehat{\boldsymbol{\Psi}}_R - \boldsymbol{\Psi} = \mathbf{L}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) = \mathbf{L}\mathbf{A}\boldsymbol{\delta} + \mathbf{L}\boldsymbol{\xi} = \mathbf{L}\mathbf{A}\boldsymbol{\delta} + \mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon} \ . \tag{A.1}$$

Also, recalling that $\mathbf{I}-\mathbf{H}$ is an $n \times n$ idempotent matrix,

$$(n-m)(\widehat{\sigma}_R^2/\sigma^2) = \boldsymbol{\zeta}'(\mathbf{I}-\mathbf{H})\boldsymbol{\zeta} = \|(\mathbf{I}-\mathbf{H})\boldsymbol{\zeta}\|^2 \ , \tag{A.2}$$

where $\boldsymbol{\zeta} = (\mathbf{C}\boldsymbol{\delta} + \boldsymbol{\varepsilon})/\sigma$. If $\boldsymbol{\varepsilon} \sim N_n(0\,,\sigma^2\mathbf{I})$, it follows immediately that $\widehat{\boldsymbol{\Psi}}_R - \boldsymbol{\Psi} \sim N_s[\mathbf{L}\mathbf{A}\boldsymbol{\delta}\ ,\ \sigma^2\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']$ and $\boldsymbol{\zeta} \sim N_n[\mathbf{C}(\boldsymbol{\delta}/\sigma)\ ,\ \mathbf{I}]$. Since $\mathbf{I}-\mathbf{H}$ has rank $n-m$, one has, by theorem 3.12 in Arnold (1981, sec. 3.5, p. 50), $(n-m)(\widehat{\sigma}_R^2/\sigma^2) \sim \chi^2_{n-m}(\lambda)$, where $\lambda = [\boldsymbol{\delta}'\mathbf{C}'(\mathbf{I}-\mathbf{H})\mathbf{C}\boldsymbol{\delta}]/\sigma^2$ is a noncentrality parameter. Notice now that $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{I}-\mathbf{H}) = 0$. By theorem 3.15 in Arnold (1981, sec. 3.5, p. 51), $\boldsymbol{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\varepsilon}$ and $(\mathbf{I}-\mathbf{H})\boldsymbol{\varepsilon}$ are independent and, therefore, $\widehat{\boldsymbol{\Psi}}_R$ and $\widehat{\sigma}_R^2$ are, by (A.1) and (A.2) above, independent as well. Finally, writing

$$F_{\mathbf{L}} = \frac{Q_{\mathbf{L}}}{s\widehat{\sigma}_R^2} = \frac{(Q_{\mathbf{L}}/\sigma^2)/s}{\widehat{\sigma}_R^2/\sigma^2} \ ,$$

and taking into account the representation $Q_{\mathbf{L}}/\sigma^2 = \boldsymbol{\eta}'\boldsymbol{\eta}$, where $\boldsymbol{\eta} = [\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1/2}(\widehat{\boldsymbol{\Psi}}_R - \boldsymbol{\Psi})/\sigma \sim N_s[\boldsymbol{\nu}\ ,\ \mathbf{I}]$ and $\boldsymbol{\nu} = [\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1/2}\mathbf{L}\mathbf{A}(\boldsymbol{\delta}/\sigma)$, the statement (2.3) about the distribution of $F_{\mathbf{L}}$ follows. ■

A.2 Asymptotic distribution of $Q_{\mathbf{L}}/\widehat{\sigma}_R^2$ under (2.8)

Suppose the regularity condition *(C)* introduced in subsection 2.2. Under a model of the form (1.2), the asymptotic behaviour of $Q_{\mathbf{L}}$ depends on the limit properties of $\sqrt{n}\,(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})$. From the representation $\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta} = \mathbf{A}\boldsymbol{\delta} + \boldsymbol{\xi}$, if not all the coefficients $\boldsymbol{\delta}$ are simultaneously zero, $\mathbf{A}\boldsymbol{\delta}$ has a nonnull finite limit so, in principle, $\sqrt{n}\,(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})$

does not have a proper asymptotic distribution. However, if the parameters $\boldsymbol{\delta}$ are "small", a possible remedial action is to think of (1.2) as the *nth* term of the sequence of contiguous models (2.8).

Proposition A.1 *Under the sequence (2.8), if: i) the errors ε_i are i.i.d. with zero mean and constant variance; and ii) condition (C) holds, then, as $n \to \infty$:*

a) $\sqrt{n}\ (\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) \xrightarrow{D} N_m[\mathbf{V}^{-1}\mathbf{Z}\boldsymbol{\theta}\ ;\ \sigma^2\mathbf{V}^{-1}]$;

b) $\widehat{\sigma}_R^2 \to \sigma^2$, *a.s.* .

Proof. a) Write $\sqrt{n}\ (\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) = \sqrt{n}\,\boldsymbol{\alpha} + \sqrt{n}\,\boldsymbol{\xi}$, where $\sqrt{n}\,\boldsymbol{\alpha} = \mathbf{A}\boldsymbol{\theta}$. Under condition *(C)*,

$$\sqrt{n}\boldsymbol{\alpha} = \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{C}}{n}\right)\boldsymbol{\theta} \to \mathbf{V}^{-1}\mathbf{Z}\boldsymbol{\theta}\ ,$$

as $n \to \infty$. To finish the proof is enough then to establish $\sqrt{n}\,\boldsymbol{\xi} \xrightarrow{D} N_m[0\ ;\ \sigma^2\mathbf{V}^{-1}]$. This follows from application of the Lindeberg condition and the Cramér-Wold device as in lemma 3.1 in Miller (1974, p. 883). b) Let $\mathbf{e}_R = (\mathbf{I} - \mathbf{H})\mathbf{Y}$ be, as in (3.3), the least squares residuals of the fit of the reduced model (1.3). Under the sequence (2.8), $\mathbf{e}_R - \boldsymbol{\epsilon} = (\mathbf{I} - \mathbf{H})\mathbf{Y} - \boldsymbol{\epsilon} = (\mathbf{I} - \mathbf{H})\mathbf{C}\boldsymbol{\delta}_n - \mathbf{H}\boldsymbol{\epsilon}$. The two summands on the right hand side of this latter identity are orthogonal and, therefore,

$$\begin{aligned} \frac{1}{n}\ \|\mathbf{e}_R - \boldsymbol{\varepsilon}\|^2 &= \frac{1}{n}\,[\frac{1}{n}\,\boldsymbol{\theta}'\mathbf{C}'(\mathbf{I} - \mathbf{H})\mathbf{C}\boldsymbol{\theta} + \boldsymbol{\varepsilon}'\mathbf{H}\boldsymbol{\varepsilon}']\ . \\ &\leq \frac{1}{n}\,[\boldsymbol{\theta}'(\frac{\mathbf{C}'\mathbf{C}}{n})\boldsymbol{\theta}] + (\frac{\boldsymbol{\varepsilon}'\mathbf{X}}{n})(\frac{\mathbf{X}'\mathbf{X}}{n})^{-1}(\frac{\mathbf{X}'\boldsymbol{\varepsilon}}{n})\ . \end{aligned} \tag{A.3}$$

From lemma 2.3 in Freedman (1981, p. 1222), $\mathbf{X}'\boldsymbol{\varepsilon}/n \to 0$, *a.s.*, so, by *(C)*, the left hand side of (A.3) converges to zero, *a.s.* Also, since $|\|\mathbf{e}_R\| - \|\boldsymbol{\epsilon}\|| \leq \|\mathbf{e}_R - \boldsymbol{\epsilon}\|$, one has, using the representation $\widehat{\sigma}_R = \|\mathbf{e}_R\|\,/\sqrt{n-m}$,

$$|\widehat{\sigma}_R - \sigma| \leq \sqrt{\frac{n}{n-m}}\frac{1}{\sqrt{n}}\ \|\mathbf{e}_R - \boldsymbol{\varepsilon}\| + \left|\frac{\|\boldsymbol{\varepsilon}\|}{\sqrt{n-m}} - \sigma\right|\ . \tag{A.4}$$

By inequality (A.3) and the law of the large numbers, (A.4) shows that $\widehat{\sigma}_R \to \sigma$, *a.s.* ∎

As a corollary, $[\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1/2}(\widehat{\boldsymbol{\Psi}}_R - \boldsymbol{\Psi})/\sigma = [\mathbf{L}(\mathbf{X}'\mathbf{X}/n)^{-1}\mathbf{L}']^{-1/2}\mathbf{L}\sqrt{n}(\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})/\sigma \xrightarrow{D} \varsigma \sim N_s(\boldsymbol{\upsilon}\ ,\ \mathbf{I})$, where $\boldsymbol{\upsilon} = [\mathbf{L}\mathbf{V}^{-1}\mathbf{L}']^{-1/2}\mathbf{L}\mathbf{V}^{-1}\mathbf{Z}(\boldsymbol{\theta}/\sigma)$. Therefore, $Q_{\mathbf{L}}/\widehat{\sigma}_R^2 =$

$(Q_{\mathbf{L}}/\sigma^2)/(\widehat{\sigma}_R^2/\sigma^2) \xrightarrow{D} \boldsymbol{\varsigma}'\boldsymbol{\varsigma} \sim\chi_s^2(\boldsymbol{v}'\boldsymbol{v})$. Notice, finally, that $\boldsymbol{v} = [\mathbf{LV}^{-1}\mathbf{L}']^{-1/2}\mathbf{LV}^{-1}\mathbf{Z}(\boldsymbol{\theta}/\sigma)$ $\cong [\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1/2}\mathbf{LA}(\boldsymbol{\delta}_n/\sigma)$.

APPENDIX B: PROOF OF THEOREM 1

The properties of the Mallows distance (3.7) listed below, taken from Bickel and Freedman (1981, p. 1211 and ff.), are needed:

i) If $E(\|\boldsymbol{\phi}\|^2)$ and $E(\|\boldsymbol{\psi}\|^2)$ are finite, the additive decomposition holds:

$$d_2^2(\boldsymbol{\phi},\boldsymbol{\psi}) = \|E(\boldsymbol{\phi}) - E(\boldsymbol{\psi})\|^2 + d_2^2[\boldsymbol{\phi} - E(\boldsymbol{\phi}), \boldsymbol{\psi} - E(\boldsymbol{\psi})] \ ; \qquad \text{(B.1)}$$

ii) For any real number a,

$$d_r(a\boldsymbol{\phi},a\boldsymbol{\psi}) = |a|\, d_r(\boldsymbol{\phi},\boldsymbol{\psi}) \ ; \qquad \text{(B.2)}$$

and

iii) A necessary and sufficient condition for a sequence of random vectors $\{\boldsymbol{\phi}_n\}$ to satisfy $d_r(\boldsymbol{\phi}_n,\boldsymbol{\phi}) \to 0$ is that both $\boldsymbol{\phi}_n \xrightarrow{D} \boldsymbol{\phi}$ and $E[\|\boldsymbol{\phi}_n\|^r] \to E\,[\|\boldsymbol{\phi}\|^r]$;

The proof of theorem 1 requires also two auxiliary lemmas first. Let H_n and F_n be the empirical distribution functions of, respectively, the scaled residuals $\sqrt{n/(n-m)}\mathbf{e}_R = \sqrt{n/(n-m)}(\mathbf{Y} - \mathbf{X}\widehat{\boldsymbol{\beta}}_R) = (\sqrt{n/(n-m)}e_{i,R})$, and the true errors ε_1, ..., ε_n.

Lemma B.1 *Under the sequence (2.8) if: i) the errors ε_i are i.i.d. with zero mean and constant variance; and ii) condition (C) holds, then $d_2^2(H_n, F_n) \to 0$, a.s.*

Proof. Let (U, V) be jointly distributed with mass $1/n$ at $(\sqrt{n/(n-m)}e_{i,R}, \varepsilon_i)$, $i = 1, ..., n$. By definition (3.7) and the inequality $\|\mathbf{x}+\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 + 2\,\|\mathbf{x}\|\,\|\mathbf{y}\|$, valid for $\mathbf{x}$, $\mathbf{y} \in R^n$,

$$\begin{aligned} d_2^2(H_n, F_n) &\leq E[(U-V)^2] = \frac{1}{n}\sum_{i=1}^{n}\left(\sqrt{n/(n-m)}e_{i,R} - \epsilon_i\right)^2 \\ &= \frac{1}{n}\left\|\sqrt{n/(n-m)}\mathbf{e}_R - \boldsymbol{\varepsilon}\right\|^2 \end{aligned}$$

$$= \frac{1}{n}\left\|\sqrt{n/(n-m)}(\mathbf{e}_R-\boldsymbol{\varepsilon})+\left(\sqrt{\frac{n}{n-m}}-1\right)\boldsymbol{\varepsilon}\right\|^2$$

$$\leq \left(\frac{n}{n-m}\right)\frac{1}{n}\left\|\mathbf{e}_R-\boldsymbol{\varepsilon}\right\|^2+\left(\sqrt{\frac{n}{n-m}}-1\right)^2\frac{1}{n}\left\|\boldsymbol{\varepsilon}\right\|^2 \tag{B.3}$$

$$+2\sqrt{\frac{n}{n-m}}\left(\sqrt{\frac{n}{n-m}}-1\right)\left(\frac{1}{\sqrt{n}}\left\|\mathbf{e}_R-\boldsymbol{\varepsilon}\right\|\right)\left(\frac{1}{\sqrt{n}}\left\|\boldsymbol{\varepsilon}\right\|\right) \tag{B.4}$$

By (A.3) and the law of the large numbers, all the terms in both (B.3) and (B.4) tend to zero, *a.s.*. ■

Lemma B.2 *Under the sequence (2.8) if: i) the errors ε_i are i.i.d. with zero mean and constant variance; and ii) condition (C) holds, then, as $n \to \infty$:*

$$d_2(\widehat{\boldsymbol{\beta}}_R^* - \widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) \to 0,\ a.s.\ . \tag{B.5}$$

Proof. Using property (B.1), the squared d_2 distance between the laws of $\widehat{\boldsymbol{\beta}}_R^* - \widehat{\boldsymbol{\beta}}_R$ and $\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}$ is

$$d_2^2(\widehat{\boldsymbol{\beta}}_R^* - \widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) = \left\|\mathbf{A}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_n)\right\|^2 + d_2^2(\boldsymbol{\xi}^*, \boldsymbol{\xi}) , \tag{B.6}$$

where $\boldsymbol{\xi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon$ and $\boldsymbol{\xi}^* = \sqrt{n/(n-m)}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon^*$. From the regularity condition *(C)* and results in Lai, Robbins and Wei (1979) on the strong consistency of the least squares estimators in multiple regression, $\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_n$ goes to zero almost surely and the same is true for $\mathbf{A}(\widehat{\boldsymbol{\delta}} - \boldsymbol{\delta}_n)$. Let now F be the common distribution function of ϵ_1, ..., ϵ_n. From lemma 8.9 in Bickel and Friedman (1981, p. 1214), $d_2^2(\boldsymbol{\xi}^*, \boldsymbol{\xi}) \leq n^{-1}tr[(\mathbf{X}'\mathbf{X}/n)^{-1}]\, d_2^2(H_n,F)$ and, since d_2 is a metric,

$$d_2^2(H_n,F) \leq 2\,[d_2^2(H_n,F_n) + d_2^2(F_n,F)] , \tag{B.7}$$

where F_n is the empirical distribution of the errors ϵ_1, ..., ϵ_n. Both summands in (B.7) converge to zero almost surely. The first by lemma B.1 above and the second by lemma 8.4 of Bickel and Freedman (1981, p. 1212). ■

Proof of theorem 1. a) For any given vectors $\mathbf{x}$, $\mathbf{y}$ in $\mathbb{R}^m$ and for any rank $s \leq m$ matrix $\boldsymbol{\Gamma}$ of $s \times m$, the chain of inequalities below holds:

$$\left|\left\|\mathbf{x}\right\|_{\boldsymbol{\Gamma}} - \left\|\mathbf{y}\right\|_{\boldsymbol{\Gamma}}\right| \leq \left\|\mathbf{x}-\mathbf{y}\right\|_{\boldsymbol{\Gamma}} \leq \left\|\boldsymbol{\Gamma}\right\|_E \left\|\mathbf{x}-\mathbf{y}\right\| , \tag{B.8}$$

where $\|\mathbf{x}\|_{\boldsymbol{\Gamma}} = \|\boldsymbol{\Gamma}\mathbf{x}\| = (\mathbf{x}'\boldsymbol{\Gamma}'\boldsymbol{\Gamma}\mathbf{x})^{1/2}$, $\|\boldsymbol{\Gamma}\|_E = \lambda_{\max}^{1/2}(\boldsymbol{\Gamma}'\boldsymbol{\Gamma})$ is the spectral norm of $\boldsymbol{\Gamma}$, and $\|\mathbf{x}\|$ is the usual euclidean norm of $\mathbf{x}$. Observe also that

$$\sqrt{Q_{\mathbf{L}}^*/n} = \left\|\widehat{\boldsymbol{\beta}}_R^* - \widehat{\boldsymbol{\beta}}_R\right\|_{\boldsymbol{\Gamma}_n} \quad , \quad \sqrt{Q_{\mathbf{L}}/n} = \left\|\widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}\right\|_{\boldsymbol{\Gamma}_n} . \tag{B.9}$$

where $\boldsymbol{\Gamma}_n = [\mathbf{L}(\mathbf{X}'\mathbf{X}/n)^{-1}\mathbf{L}']^{-1/2}\mathbf{L}$ is a rank s matrix of $s \times m$. From (B.8), (B.9) and definition (3.7) of the Mallows distance, it can be seen

$$d_1(\sqrt{Q_{\mathbf{L}}^*/n}, \sqrt{Q_{\mathbf{L}}/n}) \leq \|\boldsymbol{\Gamma}_n\|_E \, d_2(\widehat{\boldsymbol{\beta}}_R^* - \widehat{\boldsymbol{\beta}}_R, \widehat{\boldsymbol{\beta}}_R - \boldsymbol{\beta}) \to 0 \ , \ a.s. \tag{B.10}$$

by regularity condition *(C)* and convergence (B.5) obtained in lemma B.2; b) From the scaling property (B.2),

$$d_1(\widehat{\sigma}_R^*, \widehat{\sigma}_R) = \sqrt{\frac{n}{n-m}}\, d_1\Big(\frac{\|\mathbf{e}_R^*\|}{\sqrt{n}}, \frac{\|\mathbf{e}_R\|}{\sqrt{n}}\Big) , \tag{B.11}$$

where $\mathbf{e}_R^* = (\mathbf{I} - \mathbf{H})\mathbf{Y}^*$. After applying the triangle inequality twice, the second factor in (B.11) is bounded above by the sum

$$d_1\Big(\frac{\|\mathbf{e}_R^*\|}{\sqrt{n}}, \frac{\|\boldsymbol{\varepsilon}^*\|}{\sqrt{n}}\Big) + d_1\Big(\frac{\|\boldsymbol{\varepsilon}^*\|}{\sqrt{n}}, \frac{\|\boldsymbol{\varepsilon}\|}{\sqrt{n}}\Big) + d_1\Big(\frac{\|\boldsymbol{\varepsilon}\|}{\sqrt{n}}, \frac{\|\mathbf{e}_R\|}{\sqrt{n}}\Big) . \tag{B.12}$$

To finish the proof it is then enough to proof that the three summands of (B.12) go to zero. The first and the third summands have a similar treatment. One has

$$d_1^2\Big(\frac{\|\mathbf{e}_R^*\|}{\sqrt{n}}, \frac{\|\boldsymbol{\varepsilon}^*\|}{\sqrt{n}}\Big) \leq \frac{1}{n} E^*(\|\mathbf{e}_R^* - \boldsymbol{\varepsilon}^*\|^2) \leq \frac{1}{n} \left\|\mathbf{C}\widehat{\boldsymbol{\delta}}\right\|^2 + \widehat{\sigma}_R^2 q_n , \tag{B.13}$$

where $q_n = 2[1 - (m/n)][1 - \sqrt{1 - (m/n)}]$, and

$$d_1^2\Big(\frac{\|\mathbf{e}_R\|}{\sqrt{n}}, \frac{\|\boldsymbol{\varepsilon}\|}{\sqrt{n}}\Big) \leq \frac{1}{n} E(\|\mathbf{e}_R - \boldsymbol{\varepsilon}\|^2) \leq \frac{1}{n} \|\mathbf{C}\boldsymbol{\delta}_n\|^2 + \sigma^2(m/n) . \tag{B.14}$$

The two inequalities above are obtained writing first $\|\mathbf{e}_R^* - \boldsymbol{\varepsilon}^*\|^2 = Q^{*\prime}Q^*$, where $Q^* = \mathbf{e}_R^* - \boldsymbol{\varepsilon}^* = (\mathbf{I} - \mathbf{H})\mathbf{C}\widehat{\boldsymbol{\delta}} + (\mathbf{I} - \mathbf{H})\sqrt{n/(n-m)}\boldsymbol{\varepsilon}^* - \boldsymbol{\varepsilon}^*$, and $\|\mathbf{e}_R - \boldsymbol{\varepsilon}\|^2 = Q'Q$, where $Q = \mathbf{e}_R - \boldsymbol{\varepsilon} = (\mathbf{I} - \mathbf{H})\mathbf{C}\boldsymbol{\delta}_n - \mathbf{H}\boldsymbol{\varepsilon}$, noting then that $E^*[Q^*] = (\mathbf{I} - \mathbf{H})\mathbf{C}\widehat{\boldsymbol{\delta}}$, $tr(Var[Q^*]) = \widehat{\sigma}_R^2 q_n$, $E[Q] = (\mathbf{I} - \mathbf{H})\mathbf{C}\boldsymbol{\delta}_n$, and $tr(Var[Q]) = \sigma(m/n)$, and, finally, applying the well-know formula that gives the expected value of a quadratic form in a random vector as a function of its first two moments. The right hand sides of (B.13) and (B.14) go to zero, the former *a.s.* and the latter deterministically. For treatment of the second summand, observe that

$$d_1^2\left(\frac{\|\varepsilon^*\|}{\sqrt{n}},\frac{\|\varepsilon\|}{\sqrt{n}}\right) \le \frac{1}{n} d_2^2(\varepsilon^*,\varepsilon) \le d_2^2(G_n,F) \ , \tag{B.15}$$

where G_n is, as introduced in subsection 3.1, the empirical distribution function of the residuals $\mathbf{e}_R = (e_{i,R})$. The second inequality in (B.15) follows from lemma 8.9 in Bickel and Freedman (1981, p. 1214). Using a similar argument as the one given in the proof of lemma B.1, it can be seen $d_2^2(G_n,F_n) \to 0$, *a.s.*. Using inequality (B.7), with H_n replaced by G_n, gives *a.s.* convergence to zero of the upper bound in (B.15). ■

Observe, finally, that by property *iii)* above, expresion (1.6) and part b) of proposition A.1, one has $d_1(\widehat{\sigma}_R^2,\sigma^2) \to 0$. Therefore, by lemma 8.5 in Bickel and Freedman (1981, p. 1213), $d_1(\widehat{\sigma}_R,\sigma) \to 0$.

REFERENCES

[1] Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis*, New York: John Wiley.

[2] Bickel, P. J. and Freedman, D. A. (1981). "Some Asymptotic Theory for the Bootstrap", *The Annals of Statistics*, **9**, 1196-1217.

[3] Draper, N. and Smith, H. (1998). *Applied Regression Analysis*, 3rd Ed., New York: John Wiley.

[4] Efron, B. (1979). "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, **7**, 1-26.

[5] Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, New York: Chapman and Hall.

[6] Freedman, D. A. (1981). "Bootstrapping Regression Models," *The Annals of Statistics*, **9**, 1218-1228.

[7] Hocking, R. R. (1976). "The Analysis and Selection of Variables in Linear Regression", *Biometrics*, **32**, 1-49.

[8] Lai, T. L., Robbins, H., and Wei, C. Z. (1979). "Strong Consistency of Least Squares Estimates in Multiple Regression II", *Journal of Multivariate Analysis*, **9**, 343-361.

[9] Mallows, C.L. (1972). "A Note on Asymptotic Joint Normality", *Annals of Mathematical Statistics*, **43**, 508-515.

[10] 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.

[11] Miller, R. G. Jr. (1974). "An Unbalanced Jackknife", *The Annals of Statistics*, **2**, 880-891.

[12] Rawlings, J. O., Pantula, S. G. and Dickey, D. A. (1998). *Applied Regression Analysis: A Research Tool*, 2nd Ed., New York: Springer Verlag.

[13] Seber, G. A. F. (1977). *Linear Regression Analysis*. New York: John Wiley.

[14] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap*. New York: Springer Verlag.

CAPTIONS FOR FIGURES AND TABLES

Figure 1. Estimated coverage probabilities (ec) of interval (2.5) ($\cdot\, -$) as a function of δ.

Figure 2. Estimated coverage probabilities (ec) of intervals (2.5) ($\cdot\, -$) and (4.1) ($-$) as a function of δ.

Table 1. Estimated coverage probabilities for normal errors of the regions studied in the simulation study of subsection 4.2.

Table 2. Estimated coverage probabilities for uniform errors of the regions studied in the simulation study of subsection 4.2.

Table 3. Estimated coverage probabilities for double exponential errors of the regions studied in the simulation study of subsection 4.2.

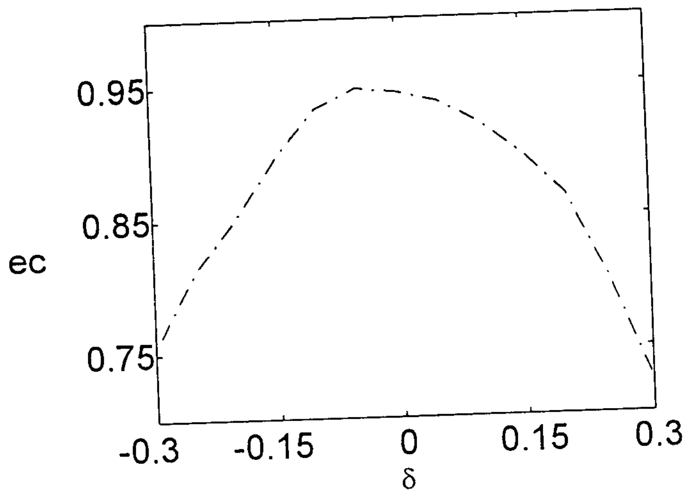

Figure 1

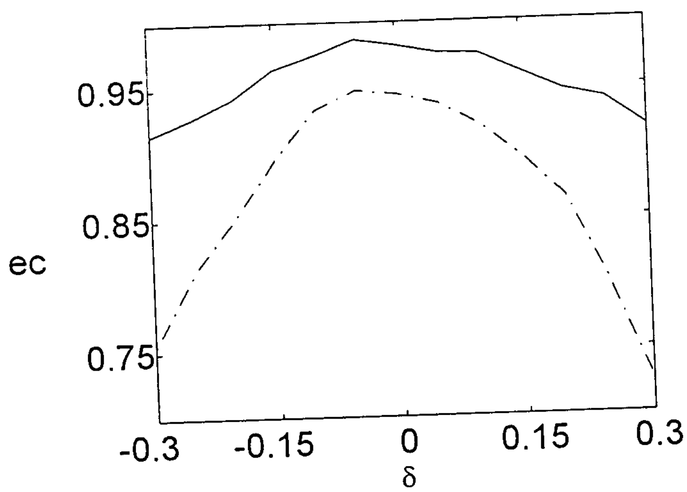

Figure 2

Normal Errors

	$q=2$						$q=4$						$q=8$					
	β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$		β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$		β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$	
	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D
									$n=30$									
θ_1	.991	.942	.984	.923	.982	.936	.986	.922	.988	.883	.976	.902	.973	.853	.970	.804	.976	.864
θ_2	.973	.880	.971	.846	.968	.873	.955	.780	.943	.687	.961	.789	.928	.548	.912	.433	.920	.552
θ_3	.937	.747	.933	.705	.936	.771	.906	.557	.877	.415	.899	.551	.820	.235	.805	.117	.826	.202
									$n=50$									
θ_1	.979	.937	.983	.926	.979	.939	.984	.923	.983	.898	.987	.921	.973	.864	.981	.840	.979	.874
θ_2	.964	.879	.956	.846	.968	.893	.956	.809	.957	.725	.961	.794	.946	.624	.916	.491	.932	.593
θ_3	.943	.782	.940	.715	.947	.804	.920	.605	.898	.474	.904	.604	.868	.308	.846	.150	.848	.262
									$n=100$									
θ_1	.982	.931	.983	.908	.982	.933	.986	.908	.988	.879	.980	.904	.981	.864	.976	.775	.981	.838
θ_2	.971	.853	.957	.788	.954	.844	.951	.759	.943	.629	.954	.705	.939	.556	.906	.325	.922	.440
θ_3	.944	.744	.909	.590	.922	.671	.914	.523	.873	.307	.895	.408	.862	.207	.811	.040	.826	.088
									$n=150$									
θ_1	.982	.934	.983	.904	.987	.914	.982	.907	.981	.740	.982	.880	.981	.850	.975	.745	.981	.819
θ_2	.979	.847	.962	.762	.966	.820	.960	.781	.939	.569	.948	.644	.939	.549	.903	.229	.894	.349
θ_3	.935	.738	.904	.532	.919	.605	.909	.525	.860	.223	.851	.305	.853	.213	.815	.022	.821	.052

Table 1

Uniform Errors

	$q=2$						$q=4$						$q=8$					
	β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$		β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$		β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$	
	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D
									$n=30$									
θ_1	.987	.915	.983	.906	.984	.993	.978	.899	.981	.875	.994	.985	.975	.828	.971	.789	.980	.970
θ_2	.967	.824	.967	.814	.979	.980	.960	.755	.940	.667	.953	.946	.930	.530	.911	.420	.911	.847
θ_3	.940	.716	.941	.655	.945	.950	.902	.525	.884	.412	.909	.831	.843	.201	.812	.088	.814	.521
									$n=50$									
θ_1	.987	.915	.984	.919	.982	.995	.987	.906	.986	.894	.988	.993	.977	.859	.971	.827	.983	.985
θ_2	.968	.863	.965	.851	.967	.987	.961	.793	.964	.721	.967	.964	.939	.589	.915	.475	.938	.884
θ_3	.934	.774	.924	.712	.931	.964	.916	.605	.906	.460	.921	.890	.863	.298	.832	.143	.845	.619
									$n=100$									
θ_1	.985	.928	.984	.901	.984	.993	.988	.891	.989	.854	.988	.990	.986	.848	.983	.768	.981	.967
θ_2	.970	.851	.953	.779	.968	.975	.954	.755	.949	.613	.950	.936	.940	.556	.906	.330	.918	.792
θ_3	.941	.734	.914	.567	.922	.922	.911	.531	.869	.307	.875	.789	.851	.204	.811	.038	.824	.392
									$n=150$									
θ_1	.987	.920	.985	.901	.988	.994	.984	.893	.986	.860	.986	.987	.980	.840	.975	.736	.982	.964
θ_2	.962	.840	.954	.753	.965	.970	.960	.766	.943	.556	.953	.921	.928	.553	.913	.232	.924	.760
θ_3	.930	.734	.905	.527	.920	.907	.903	.530	.859	.230	.865	.729	.860	.201	.800	.021	.805	.249

Table 2

Double Exponential Errors

	$q=2$						$q=4$						$q=8$					
	β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$		β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$		β_1		$\beta_1-\beta_2$		$(\beta_1,\beta_2)'$	
	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D	E^*	D
	$n=30$																	
θ_1	.981	.931	.982	.892	.981	.982	.982	.883	.982	.868	.976	.976	.974	.833	.972	.792	.977	.959
θ_2	.966	.833	.968	.798	.969	.970	.951	.731	.943	.644	.956	.932	.915	.498	.900	.383	.915	.806
θ_3	.928	.706	.929	.640	.939	.934	.908	.510	.883	.372	.896	.795	.825	.213	.814	.108	.824	.474
	$n=50$																	
θ_1	.984	.927	.987	.924	.979	.993	.985	.892	.990	.883	.982	.986	.985	.839	.979	.812	.987	.977
θ_2	.958	.862	.966	.833	.972	.980	.959	.784	.954	.693	.963	.951	.944	.560	.924	.458	.936	.864
θ_3	.941	.768	.929	.701	.939	.942	.912	.594	.904	.452	.914	.874	.852	.287	.820	.144	.848	.581
	$n=100$																	
θ_1	.984	.932	.986	.894	.983	.989	.987	.897	.987	.865	.983	.988	.986	.847	.982	.780	.982	.967
θ_2	.967	.838	.953	.760	.967	.974	.961	.744	.945	.597	.956	.934	.940	.516	.905	.279	.917	.789
θ_3	.936	.725	.911	.576	.931	.923	.918	.521	.871	.284	.875	.760	.851	.200	.837	.050	.819	.356
	$n=150$																	
θ_1	.985	.928	.983	.914	.986	.995	.988	.906	.983	.835	.988	.987	.985	.845	.974	.746	.981	.972
θ_2	.969	.858	.959	.750	.957	.968	.962	.746	.944	.551	.946	.907	.939	.530	.900	.244	.911	.732
θ_3	.931	.712	.911	.520	.920	.912	.907	.508	.856	.207	.875	.697	.863	.198	.821	.022	.824	.251

Table 3