PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 21, Fasc. 2 (2001), pp. 265–276

ON THE BOOTSTRAPPING HETEROSCEDASTIC REGRESSION MODELS

BY

OCTAV FERCHELUC (BUCHAREST)

Abstract. The distributions of deviations of point estimators for parameters of interest are essential in the evaluation of the efficiency of point estimators. The bootstrap method suggested by B. Efron is one of the main methods directed at solving the problem of producing distributions which mimic the unobserved distributions of deviations.

The main object of this article is to study the asymptotic validity of the bootstrap in the context of heteroscedastic regression models, using the central limit resampling theorem. In the case of one-parameter linear regression, theoretical results are illustrated by an example with simulated statistical data.

Key words and phrases: Bootstrap, heteroscedastic regression, resampling, ordinary least squares estimates, central limit resampling theorem.

1. Introduction. The use of the bootstrap to estimate the sampling distribution of parameter estimates in homogeneous linear models was first proposed by Efron [4] and further developed by Freedman [7], and Wu [14]. The process involves approximating the distribution of unobserved errors with the empirical distribution of the centered residuals. Other works for bootstrapping homogeneous regression models are Navidi [12], Holm [11], Stone and Brooks [13].

When the errors are not equally distributed, and for example have very different variances, we cannot neglect this in the bootstrap resampling. Vector resampling is the first idea in the domain and this possibility has been already considered in Efron [5]. For the linear model, Wu [14] has suggested an improved idea of resampling but it seems to require rather a lot of data for good performance. The paper by Freedman and Peters [9] presents some empirical results for the bootstrap in the context of an econometric equation by considering constrained generalized least squares with an estimated covariance matrix.

We examine the accuracy of bootstrap estimates of the distribution of linear combination of heteroscedastic regression parameter estimates. In this case, the appropriate basic statistics have been found and we shall prove that the polynomial randomization gives an imitating distribution. Note also that here we drop the restriction that copies of design of regression experiments are realized from factors combined via random resampling with centered errors as in Freedman [7]. The original design of regression experiments can practically be arbitrary in the suggested approach.

This article is organized as follows: Section 2 gives a brief review of the heteroscedastic linear regression model. Section 3 gives a short review of the bootstrap procedure and the central limit resampling theorem is used to prove the accuracy of the bootstrap for heteroscedastic regression models, under the viability of some assumptions. Section 4 applies the results to the simplest case of one-parameter linear regression by proving the claimed assumptions and presents a simulation experiment to assess the validity of the bootstrap by comparing the homogeneous bootstrap with the heteroscedastic one. Some final remarks will be also drawn.

2. The heteroscedastic regression model. The model studied is defined as follows. Let $\mathcal{G}$ be a restricted set in $\boldsymbol{R}^r$ and $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ir})^T$ be factors which define conditions of the i-th experiment, $i = 1, \ldots, n$. The result of the experiment is a value of a random real variable

$$y_i = x_{i1}\beta_{10} + \ldots + x_{ir}\beta_{r0} + \omega_i, \quad i = 1, \ldots, n, \tag{1}$$

where $\beta_{j0}, j = 1, \ldots, r$, are unknown real numbers and $\omega_1, \ldots, \omega_n$ are values of unobserved random errors $W_1, \ldots, W_n$. We assume that $W_1, \ldots, W_n$ are independent r.v.'s with zero expectations and unequal finite second order moments $\sigma_1^2, \ldots, \sigma_n^2$. This statistical model is called a *heteroscedastic linear regression.* In order to rewrite (1) using vectors and matrices we will use the following notation:

$$\boldsymbol{x}_i^T = (x_{i1}, \ldots, x_{ir}), \quad \boldsymbol{X}_{nr} = (x_{ij}),\ i = 1, \ldots, n,\ j = 1, \ldots, r,$$
$$\boldsymbol{w}_n^T = (w_1, \ldots, w_n), \quad \boldsymbol{y}_n^T = (y_1, \ldots, y_n).$$

Because $X^T = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n]$, (1) takes the following form:

$$\boldsymbol{y}_n = X\beta_0 + \boldsymbol{w}_n, \tag{2}$$

where $\beta_0 = (\beta_{10}, \ldots, \beta_{r0})$,

$$\max_{1 \leqslant i \leqslant n} \|\boldsymbol{x}_i\|^2 \leqslant \sup(\|\boldsymbol{x}\|^2 \colon \boldsymbol{x} \in \mathcal{G} \in \boldsymbol{R}^r) = g_+^2 < \infty, \quad \|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x}.$$

We also assume that

$$\max_{1 \leqslant i \leqslant n} \sigma_i^2 \leqslant \sigma_+^2 < \infty. \tag{3}$$

Later we will use the following slightly stronger assumption on the moments of errors:

ASSUMPTION 1.

$$\max_{1 \leqslant i \leqslant n} E|W_i|^{2+\delta} \leqslant C(2+\delta) < \infty, \quad 0 < \delta \leqslant 1. \tag{4}$$

Statement (4) implies (3) for some $\sigma_+^2 < \infty$. If $\operatorname{rang}(X^T X) = r_0 < r$, then it is possible to consider generalizations of a least squares approach to statistical inference, based on notions related to pseudoinverse (or Moor–Penrose) matrices. Under the assumption that

$$\beta_0 = (X^T X)^+ (X^T X) \beta_0, \tag{5}$$

the β_0 is said to be *estimable*. In this case, it is possible to consider

$$\hat{\beta}_n = (X^T X)^+ X^T y_n, \tag{6}$$

which is often called an *ordinary least squares estimator* (OLS-estimator). If β_0 is estimable, then (5) implies that $\hat{\beta}_n$ is unbiased without assumption that all σ_i^2's are equal.

We consider the case when a number of experiments increases, i.e. $n \to \infty$. We introduce the following assumption:

ASSUMPTION 2.

$$\operatorname{tr}(X^T X)^+ \to 0 \quad \text{as } n \to \infty. \tag{7}$$

Assumption 2 is equivalent to $\max_{1 \leqslant k \leqslant r} (X^T X)_{kk}^+ \to 0$ as $n \to \infty$.

LEMMA 1. *If Assumption 2 is valid, then $\hat{\beta}_n$ given by* (6) *is an unbiased and consistent estimator for estimable β_0.*

Proof. From (2), (5) and (6) it follows that

$$\hat{\beta}_n - \beta_0 = (X^T X)^+ X^T y_n - (X^T X)^+ (X^T X) \beta_0 = \sum_{i=1}^{n} (X^T X)^+ \tilde{x}_i W_i. \tag{8}$$

We assume that $EW_i = 0$, and so

$$E\hat{\beta}_n = \beta_0.$$

We note that

$$\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T = X^T X$$

and

$$EW_n W_n^T = (EW_i W_j) = (\sigma_i^2 \delta_{ij}) = \sum_{i=1}^{n} \sigma_i^2 \delta_i \delta_i^T, \quad \text{where } \delta_i^T = (\delta_{i1}, \ldots, \delta_{in}).$$

The covariance matrix is the following:

$$(9)\qquad C_n = E(\hat{\beta}_n-\beta_0)(\hat{\beta}_n-\beta_0)^T = (X^T X)^+ X^T (EW_n W_n^T) X (X^T X)^+$$
$$= (X^T X)^+ \sum_{i=1}^{n} \sigma_i^2 \tilde{x}_i \tilde{x}_i (X^T X)^+ \leqslant \sigma_+^2 (X^T X)^+ \sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T (X^T X)^+$$
$$= \sigma_+^2 (X^T X)^+ (X^T X)(X^T X)^+ = \sigma_+^2 (X^T X)^+ .$$

The variances of components $\hat{\beta}_n-\beta_0$ are diagonal elements of C_n, and their sum is $\operatorname{tr}(C_n)$. Furthermore, (9) implies

$$\operatorname{tr}(C_n) \leqslant \sigma_+^2 \operatorname{tr}(X^T X)^+ \to 0 \quad \text{as } n \to \infty .$$

Hence, all r components of the vector $\hat{\beta}_n-\beta_0$ converge to zero in probability. ■

The rate of convergence of $\hat{\beta}_{jn}$ to β_{j0} can be different for different components, $j = 1, \ldots, r$. Therefore, it is reasonable to consider estimation of separate components of β_0 if possible or their linear combinations

$$c_0^T \beta = c_1 \beta_{10} + \ldots + c_r \beta_{r0},$$

where $c = (c_1, \ldots, c_r)^T$ is a fixed vector. The vector c is said to be a *direction vector* if $\|c\|^2 = c^T c = 1$. We can introduce

$$(10)\qquad a_{in}(c) := \frac{c^T (X^T X)^+ x_i}{\left(c^T (X^T X)^+ c\right)^{1/2}}, \quad i = 1, \ldots, n.$$

The following two assumptions will be used in the sequel. The first deals with variances of errors and the second with some coefficients based on the direction vector. Both of the following assumptions are related to the used sequence of experiments:

ASSUMPTION 3. *There exists* $0 < \sigma_-^2 \leqslant \sigma_+^2 < \infty$ *such that for any* n

$$(11)\qquad \sigma_-^2 (X^T X) \leqslant \sum_{i=1}^{n} \sigma_i^2 x_i x_i^T \leqslant \sigma_+^2 (X^T X).$$

ASSUMPTION 4. *For any* $n = 1, 2, \ldots$ *and* $i = 1, \ldots, n$, *there exists* $A_0(c) < \infty$, $A(c) < \infty$ *such that*

$$(12)\qquad \frac{\operatorname{tr}(X^T X)^+}{c^T (X^T X)^+ c} \leqslant A_0(c), \qquad \max_{1 \leqslant i \leqslant n} |a_{in}(c)| \leqslant \frac{A(c)}{\sqrt{n}}.$$

The variance of $c^T(\hat{\beta}_n-\beta_0)$ is

$$(13)\quad \sigma_n^2(c) = Ec^T(\hat{\beta}_n-\beta_0)(\hat{\beta}_n-\beta_0)^T c = c^T (X^T X)^+ \Big(\sum_{i=1}^{n} \sigma_i^2 x_i x_i^T\Big)(X^T X)^+ c.$$

Assumption 3, (11), implies the inequality

$$(14)\qquad \sigma_-^2 c^T (X^T X)^+ c \leqslant \sigma_n^2(c) \leqslant \sigma_+^2 c^T (X^T X) c.$$

Hence it is reasonable to consider the following normed deviations:

$$U_n(c) = \frac{c^T(\hat{\beta}_n - \beta_0)}{\left(c^T(X^T X)^+ c\right)^{1/2}}. \tag{15}$$

From (8) and (10) it follows that

$$U_n(c) = \sum_{i=1}^{n} a_{in}(c) W_i. \tag{16}$$

THEOREM 1. *If Assumptions* 1, 3 *and* 4 *are valid, then the distribution law of normed deviations* $U_n(c)$ *can be imitated by the normal distribution of mean and standard deviation* $\sigma_n^2(c)/c^T(X^T X)^+ c$ *when the number of observations, n, increases to* ∞, *i.e.*

$$L(U_n(c)) - N_1\left(0, \frac{\sigma_n^2(c)}{c^T(X^T X)^+ c}\right) \xrightarrow{wa} 0 \quad \textit{as } n \to \infty. \tag{17}$$

Proof. The variance of summands in (16) is

$$\tilde{\sigma}_n^2(c) = EU_{in}^2(c) = \frac{\sigma_i^2 c^T (X^T X)^+ \tilde{x}_i \tilde{x}_i^T (X^T X)^+ c}{c^T (X^T X)^+ c}$$

and

$$\tilde{\sigma}_n^2(c) = EU_n^2(c) = \sum_{i=1}^{n} \tilde{\sigma}_{in}^2(c).$$

From (9), (10), and (12)–(14) we obtain

$$\tilde{\sigma}_{in}^2(c) \leqslant \frac{1}{n}\sigma_+^2 A^2(c), \qquad \sigma_-^2 \leqslant \tilde{\sigma}_n^2(c) = \frac{c^T C_n c}{c^T(X^T X)^+ c} \leqslant \sigma_+^2.$$

Assumptions 1 and 4 imply

$$\begin{aligned} E|U_{in}(c)\sqrt{n}|^{2+\delta} &= n^{1+\delta/2} E|U_{in}(c)|^{2+\delta} \\ &= |\sqrt{n} a_{in}(c)|^{2+\delta} E|W_i|^{2+\delta} \leqslant A(c)^{2+\delta} C(2+\delta). \end{aligned}$$

Hence the moments $E|U_{in}(c)\sqrt{n}|^{2+\delta}$ are bounded uniformly in n, and the desired result is obvious now. ∎

3. The bootstrap procedure. We consider the following general scheme of data collecting. Let $(\Omega, \mathscr{B}(\Omega), P_0(\cdot))$ be a basic probability space. Data collecting can be understood as a sequence of experiments $\mathscr{E}_1, \mathscr{E}_2, \ldots$ The plan (design) of the i-th experiment and its result can be represented as a point $x_i \in \mathscr{X}_i$, where $\mathscr{X}_i$ is the set of all possible values for x_i. We consider x_i as a value of the $\mathscr{X}_i$-valued r.v. $X_i(\cdot)$ which is a measurable mapping from Ω into $\mathscr{X}_i$. The probability that $X_i \in B$ is given by $P_{\theta_0,i}(B) = P_0(X_i(w) \in B)$ under the assumption that the *true parameter* $\theta_0 \in \Theta$. Throughout this paper we use the assumption that *random* X_i*-valued variables* X_i, $i = 1, 2, \ldots,$ *are independent.*

For each fixed n we consider $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ as given original statistical data. Let $\hat{\theta}_n = S_n(x_1, \ldots, x_n)$ be a consistent point estimator of θ_0, i.e., in some sense, $\hat{\theta}_n \to \theta_0$ as $n \to \infty$. If $\Theta \subset \boldsymbol{R}^r$, we consider the *distribution law of normed deviations*:

$$\mathscr{L}_{\theta_0} := \mathscr{L}(\sqrt{n}(\hat{\theta}_n - \theta_0)). \tag{18}$$

If a parameter of interest $\gamma_0 = \gamma(\theta_0)$, where $\gamma\colon \Theta \to \Upsilon$, $\Upsilon = \{\gamma(\theta)\colon \theta \in \Theta\} \subset \boldsymbol{R}^s$, then we can consider

$$\mathscr{L}_{\gamma(\theta_0),n} := \mathscr{L}(\sqrt{n}(\hat{\gamma}_n - \gamma_0)), \tag{19}$$

where $\hat{\gamma}_n = T_n(x_1, \ldots, x_n)$ is a point estimator of γ_0. The estimator $\tilde{\gamma}_n := \gamma(\hat{\theta}_n)$ is called a *plug-in estimator*. If we consider $\hat{\theta}_n$, a true parameter, then $\tilde{\gamma}_n$ should be meant as the true value for the parameter of interest.

How then can we find or at least to know more about the distributions (18) or (19) when θ_0 is unknown?

It is natural to think that the order of experiments is not essential to a statistical inference. One has to have the same inference when the experiments $\mathscr{E}_1, \ldots, \mathscr{E}_n$ have been fulfilled in another order, say, $\mathscr{E}_{i_1}, \mathscr{E}_{i_2}, \ldots, \mathscr{E}_{i_n}$. Then the original data are $(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$. We want to know more about the distribution of normed deviations of the point estimator $\hat{\gamma}_n = T_n(x_1, \ldots, x_n)$ for the parameter $\gamma_0 = \gamma(\theta_0)$.

We will call the *bootstrap* the approach which is described by the following sequence of actions:

(i) Find $\hat{\gamma}_n = T(\boldsymbol{x}_n)$ (a consistent estimator).

(ii) Take B random samples with replacement from the original data $\boldsymbol{x}_n = (x_1, \ldots, x_n)$ with the same size n:

$$\boldsymbol{x}_n^{*1} = (x_1^{*1}, x_n^{*1}, \ldots, x_n^{*1}),$$
$$\boldsymbol{x}_n^{*2} = (x_1^{*2}, x_n^{*2}, \ldots, x_n^{*2}), \ldots, \boldsymbol{x}_n^{*B} = (x_1^{*B}, x_n^{*B}, \ldots, x_n^{*B});$$

we call $\boldsymbol{x}_n^{*b}$ the *bootstrap copy* of the original data.

(iii) If $\gamma_0 = \gamma(\theta_0)$ is a parameter of interest and $\hat{\gamma}_n = T(\boldsymbol{x}_n)$ is its consistent estimator, then for each bootstrap copy of the original data find

$$\hat{\gamma}_n^{*1} = T(\boldsymbol{x}_n^{*1}),\ \hat{\gamma}_n^{*2} = T(\boldsymbol{x}_n^{*2}), \ldots, \hat{\gamma}_n^{*B} = T(\boldsymbol{x}_n^{*B}). \tag{20}$$

(iv) Find the deviations of estimated values (20) from $\hat{\gamma}_n$, i.e.

$$(\hat{\gamma}_n^{*1} - \hat{\gamma}_n, \hat{\gamma}_n^{*2} - \hat{\gamma}_n, \ldots, \hat{\gamma}_n^{*B} - \hat{\gamma}_n);$$

compute the bootstrap version for the conditional law of normed deviations from $\hat{\gamma}_n$ given $x_1, \ldots, x_n$:

$$\mathscr{L}_{n,B}^* := \mathscr{L}(\sqrt{n}(\hat{\gamma}_n^* - \hat{\gamma}_n) \mid x_1, \ldots, x_n) = \frac{1}{B}\sum_{b=1}^{B} I(\sqrt{n}(\hat{\gamma}_n^{*b} - \hat{\gamma}_n) \in \cdot). \tag{21}$$

Under some assumptions, the central limit bootstrap theorem states that $\mathscr{L}^*_{n,B}$ will mimic $\mathscr{L}_{\gamma(\theta_0),n}$ for large values of n.

Now we are ready to apply the central limit resampling theorem (Hall and Mammen [10]). Let $M_1, \ldots, M_n$ be random variables obtained via n independent polynomial experiments with n outcomes and probability n for each outcome. From the above-mentioned theorem it follows that the distribution laws of the sequence of r.v.'s given in (16), $n = 1, 2, \ldots$, approach weakly the distribution laws of the sequence of r.v.'s:

$$\sum_{i=1}^{n} \frac{c^T (X^T X)^+ \tilde{x}_i}{\left(c^T (X^T X)^+ c\right)^{1/2}} (M_i - 1) W_i = \sum_{i=1}^{n} \frac{c^T (X^T X)^+ \tilde{x}_i}{\left(c^T (X^T X)^+ c\right)^{1/2}} (M_i - 1)(Y_i - \tilde{x}_i^T \beta_0). \tag{22}$$

There is no possibility of applying this nice fact directly because β_0 is not known. But we can do that if we replace β_0 by its OLS-estimator $\hat{\beta}_n$ given by (6). This is what we are doing below. We define the following *basic statistics*:

$$U_n^*(c) = \sum_{i=1}^{n} \frac{c^T (X^T X)^+ \tilde{x}_i}{\left(c^T (X^T X)^+ c\right)^{1/2}} (M_i - 1)(Y_i - \hat{Y}_{in}), \tag{23}$$

where $\hat{Y}_{in} = \tilde{x}_i^T \hat{\beta}_n = \tilde{x}_i^T (X^T X)^+ X^T Y_n$ are components of the predictor vector and $Y_i - \hat{Y}_{in}$ are estimated errors, $i = 1, \ldots, n$. We note that we can find, for each n, the distribution law $L\left(U_n^*(c) \mid Y_n = y_n, X\right)$ via simulation M_i, $i = 1, \ldots, n$.

THEOREM 2. *Let Assumptions* 1–4 *be valid. Then the distribution law of normed deviations $U_n(c)$ can be imitated by the conditional distribution law of $U_n^*(c)$ given $Y_n = y_n$ and X, when the number of observations, n, increases to ∞.*

Proof. If we add (22) to (23) and subtract one from the other, then we have

$$\begin{aligned} U_n^*(c) = & \sum_{i=1}^{n} \frac{c^T (X^T X)^+ \tilde{x}_i}{\left(c^T (X^T X)^+ c\right)^{1/2}} (M_i - 1)(Y_i - \tilde{x}_i^T \beta_0) \\ & - \sum_{i=1}^{n} (M_i - 1) \frac{c^T (X^T X)^+ \tilde{x}_i}{\left(c^T (X^T X)^+ c\right)^{1/2}} (\hat{Y}_{in} - \tilde{x}_i^T \beta_0), \end{aligned} \tag{24}$$

where $Y_i - \tilde{x}_i^T \beta_0 = W_i$, and hence the distribution of the first sum has the desired property, as shown above. We have to prove that the second sum in (24) converges to 0 in probability. The predicted response is $\hat{Y}_{in} = \tilde{x}_i^T \hat{\beta}_n$. Hence we can write

$$Q_n(c) = V_n^T(c) H_n(c), \tag{25}$$

where

$$V_n^T(c) = \sum_{i=1}^{n} (M_i - 1) c^T (X^T X)^+ \tilde{x}_i \tilde{x}_i^T, \quad H_n(c) = \frac{\hat{\beta}_n - \beta_0}{\left(c^T (X^T X)^+ c\right)^{1/2}}.$$

From (25) we have

$$Q_n^2(c) \leqslant \|V_n(c)\|^2 \, \|H_n(c)\|^2.$$

Relations (9) and (12) imply

$$E\|H_n(c)\|^2 = \frac{E\|\hat{\beta}_n - \beta_0\|^2}{c^T (X^T X)^+ c} \leqslant \frac{\sigma_+^2 \operatorname{tr}(X^T X)^+}{c^T (X^T X)^+ c} \leqslant \sigma_+^2 A_0(c) < \infty.$$

Therefore, it is sufficient to prove that the first factor $V_n^T(c)$ in (25) tends to zero in probability when $n \to \infty$. The randomness of this factor is due to the randomness of $(M_1, \ldots, M_n)$. We note that $EM_i = 1$, $E(M_i - 1)^2 = 1 - 1/n$ and $E(M_{i_1} - 1)(M_{i_2} - 1) = -1/n$, $i_1 \neq i_2$. Consequently, we have $EV_n(c) = 0$, and

$$\begin{aligned} E\|V_n(c)\|^2 = EV_n^T(c)\, V_n(c) &= \sum_{i=1}^{n} E(M_i - 1)^2 c^T (X^T X)^+ \tilde{x}_i \tilde{x}_i^T \tilde{x}_i \tilde{x}_i^T (X^T X)^+ c \\ &\quad + \sum_{i_1 \neq i_2} E(M_{i_1} - 1) E(M_{i_2} - 1) c^T (X^T X)^+ \tilde{x}_{i_1} \tilde{x}_{i_1}^T \tilde{x}_{i_2} \tilde{x}_{i_2}^T (X^T X)^+ c \\ &= \sum_{i=1}^{n} c^T (X^T X)^+ \tilde{x}_i \|\tilde{x}_i\|^2 \tilde{x}_i^T (X^T X)^+ c - \frac{1}{n} \Big\| \sum_{i=1}^{n} c^T (X^T X)^+ \tilde{x}_i \tilde{x}_i^T \Big\|^2 \\ &\leqslant g_+^2 \sum_{i=1}^{n} \big(c^T (X^T X)^+ \tilde{x}_i\big)^2 \leqslant g_+^2 \frac{A^2(c)}{n} \sum_{i=1}^{n} c^T (X^T X)^+ c \\ &\leqslant g_+^2 A^2(c) \operatorname{tr}(X^T X)^+ \end{aligned}$$

because we can use $\|c\|^2 = 1$ and the inequality $c^T (X^T X)^+ c \leqslant \operatorname{tr}(X^T X)^+$.

Assumption 2 implies $E\|V_n(c)\| \to 0$ as $n \to \infty$. Hence we obtain $Q_n(c) = V_n^T(c) H_n(c) \to 0$ in probability, as $n \to \infty$. ∎

4. One-parameter linear regression. We consider the simplest case of the one-parameter linear regression:

$$y_i = g(t_i)\beta_0 + w_i, \tag{26}$$

where $w_i = h(t_i) z_i$, $z_1, \ldots, z_n$ are values of i.i.d. normally distributed r.v.'s $Z_1, \ldots, Z_n$, $\mathscr{L}(Z_i) = N_1(0, 1)$. The OLS-estimator is

$$\hat{\beta}_n = \frac{\sum_{i=1}^n g(t_i) h(t_i)}{\sum_{i=1}^n g^2(t_i)},$$

the predicted responses are

$$\hat{y}_{in} = g(t_i)\hat{\beta}_n,$$

and the corresponding residuals are

$$\hat{w}_{in} = y_{in} - \hat{y}_{in}.$$

In this case we have

$$X = \big(g(t_1), \ldots, g(t_n)\big)^T, \quad X^T X = \sum_{i=1}^{n} g^2(t_i) \quad \text{and} \quad (X^T X)^+ = \big(\sum_{i=1}^{n} g^2(t_i)\big)^{-1}.$$

Hence, it will be easy to check the validity of the above assumptions. We assume that

$$\max\big(|g(t)|\colon t \in \mathscr{G} = [0, 1]\big) = g_+ < \infty \quad \text{and} \quad \max\big(|h(t)|\colon t \in [0, 1]\big) = h_+ < \infty.$$

Assumption 1 follows directly from the normality of the distribution of $(Z_i)_{i \geqslant 1}$ (e.g. $E\big(h(t_i)Z_i\big)^4 \leqslant 3h_+^4$). Assume also that the experiments are planned in such a way that for some $h_0 > 0$ and $g_0 > 0$ the number of the experiments $n(g_0, h_0)$, with $|g(t_i)| \geqslant g_0 > 0$ and $|h(t_i)| \geqslant h_0 > 0$, increases proportionally with n, i.e. for all sufficiently large n

$$n(g_0, h_0) \geqslant q_0 n \text{ for some } q_0, \quad 0 < q_0 \leqslant 1.$$

In this case, Assumptions 2–4 are valid. For example, we have

$$|\tilde{x}_i| = |g(t_i)| \leqslant g_+, \quad (ng_+)^{-1} \leqslant (X^T X)^+ \leqslant (nq_0 g_0^2)^{-1},$$

and for a_{in} given by (10) the following inequality holds true:

$$|a_{in}| \leqslant \frac{1}{\sqrt{n}} \frac{g_+^2}{|c|(q_0 g_0)^2}.$$

Hence Assumption 4 is fulfilled.

A SIMULATION EXPERIMENT. Let the statistical data (t_i, y_i), presented in Fig. 1, be chosen in such a way that y_i is the response and t_i is the parameter defining the conditions of the i-th experiments. The number of observations is $n = 25$. These data were simulated with $g(t) = t$, $h(t) = 1 - t$ and $t_i = i/n$, $i = 1, \ldots, n$.

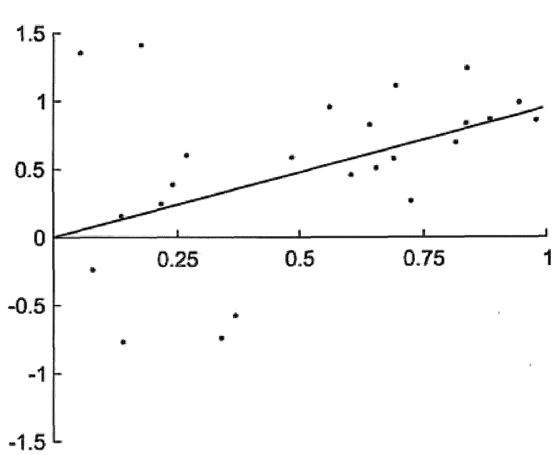

Fig. 1

The parameter β_0, which has to be estimated, is taken to be 1. The OLS--estimator for simulated β_0's based on the data is $\hat{\beta}_n = 0.949$. Even if we do not know $h(t)$, it is easy to see that the regression data obviously look heteroscedastic. We want to know the distribution law $\mathscr{L}(\sqrt{n}(\hat{\beta}_n - \beta_0))$.

Let us ignore the heteroscedasticity and try to use the bootstrap technics for homogeneous regression (Freedman [7]). In the case of homogeneous regression, all variances of errors are the same. We center the residuals and consider them as errors:

$$w_{in}^0 = y_i - \hat{y}_{in} - \frac{1}{n} \sum_{i'=1}^{n} (y_{i'} - \hat{y}_{i'n}), \quad i = 1, \ldots, n.$$

We take the j-th random sample with replacement of volume n from $(w_{1n}^0, \ldots, w_{nn}^0)$. Let $(w_{1n}^{*j}, \ldots, w_{nn}^{*j})$ be the sampled values. Then we find the j-th bootstrap copy of responses

$$y_{in}^{*j} = g(t_i)\,\hat{\beta}_n + w_{in}^{*j}, \quad i = 1, \ldots, n, \tag{27}$$

and calculate the value $\hat{\beta}_n^{*j}$, which is an OLS-estimator based on the data $(y_1^{*j}, g(t_1); \ldots; y_n^{*j}, g(t_n))$, i.e.

$$\hat{\beta}_n^{*j} = \frac{\sum_{i=1}^{n} g(t_i)\, y_i^{*j}}{\sum_{i=1}^{n} g^2(t_i)}. \tag{28}$$

We repeat these calculations for B bootstrap copies, i.e. for $j = 1, \ldots, B$. This approach is called a *homogeneous bootstrap*. It is known (Freedman [7]) that the empirical function based on the normed deviations,

$$\{\sqrt{n}(\hat{\beta}_n^{*1} - \hat{\beta}_n), \sqrt{n}(\hat{\beta}_n^{*2} - \hat{\beta}_n), \ldots, \sqrt{n}(\hat{\beta}_n^{*B} - \hat{\beta}_n)\}, \tag{29}$$

mimics the unknown (true) distribution function of the normed deviations, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ as $n \to \infty$, in the homogeneous case when all errors W_i have the same distribution. Here, we have the heteroscedastic case and there is no guarantee for such a similarity.

By (29), the homogeneous bootstrap distribution law for normed deviations is

$$\mathscr{L}_n^* = \mathscr{L}(\sqrt{n}(\hat{\beta}_n^* - \beta_0)) \approx \frac{1}{B} \sum_{j=1}^{n} 1(\sqrt{n}(\hat{\beta}_n^{*j} - \hat{\beta}_n) \leqslant u), \quad u \in \boldsymbol{R}.$$

From the above suggested theory it follows that there is another way to mimic true distribution of the normed deviation $\mathscr{L}_n^* = \mathscr{L}(\sqrt{n}(\hat{\beta}_n^* - \beta_0))$ via the randomizing statistic (23), which in our example with $c = 1$ is

$$U_n^*(1) = \sum_{i=1}^{n} (M_i - 1) \frac{\left(\sum_{l=1}^{n} g^2(t_l)\right)^{-1} g(t_i)}{\left(\sum_{l=1}^{n} g^2(t_l)\right)^{-1/2}} (y_i - \hat{y}_{in}), \tag{30}$$

and where $M_1, \ldots, M_n$ are numbers of outcomes in n simulated polynomial experiments. We simulate the data, and hence we know that the true distribution law $\mathscr{L}_n$ is normal and due to (8) we have here

$$\sqrt{n}(\beta_n - \beta_0) = \left(\frac{1}{n}\sum_{i=1}^{n} g^2(t_i)\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(t_i)\,h(t_i)\,Z_i.$$

Hence the true distribution law $\mathscr{L}_n$ is

$$\mathscr{L}_n = N_1\left(0, \frac{(1/n)\sum_{i=1}^{n} g^2(t_i)\,h^2(t_i)}{\left((1/n)\sum_{i=1}^{n} g^2(t_i)\right)^2}\right).$$

In Fig. 2, three distribution functions are shown: a — theoretical $\mathscr{L}_n$; b — homogeneous bootstrap $\mathscr{L}_n^*$; and c — $\mathscr{L}_n^*$ based on randomized statistics $V^*(1)$.

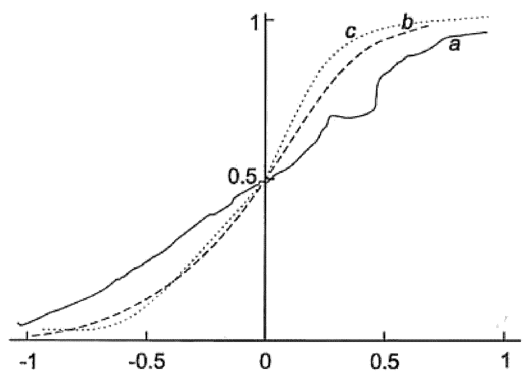

Fig. 2

It is seen that the formal usage of the homogeneous bootstrap gives erroneous results. The basic statistics (30) gives rather good approximation to the true distribution of normed deviations. Of course, such a good coincidence of $\mathscr{L}_n^*$ with the true $\mathscr{L}_n$ for $n = 25$ can also be random due to "too good" data. But, in any case, the larger n is better if the heteroscedastic law $\mathscr{L}_n^*$ is as an approximation to

$$N_1\left(0, \left((1/n)\sum_{i=1}^{n} g^2(t_i)\,h^2(t_i)\right) / \left(\left((1/n)\sum_{i=1}^{n} g^2(t_i)\right)^2\right)\right)$$

whereas this will not be the case for the homogeneous bootstrap.

Concluding remarks. The history of intensive theoretical research in this area especially related to the bootstrap has now been going on for more than 18 years. In this paper, the bootstrap method is suggested to heteroscedastic regression models, via the central limit resampling theorem. The problems related to regression models are of wide interest, as shown in a paper by Wu [14].

The bootstrap can be also used in other regression contexts. For instance, Delaney and Chatterjee [3] proposed a bootstrap method for choosing the ridge parameter in a ridge regression, although their bootstrap unit is a vector of observations (Y_i, x_i) rather than a residual. Freedman [8] and Freedman and Peters [9] applied the bootstrap in more complex regression models such as a dynamic linear model estimated by two-stage least squares estimates. They showed that it is asymptotically valid, like classical methods, but that the bootstrap solution sometimes outperforms the classical methods. Efron and Tibshirani [6] also applied the bootstrap to Cox's proportional-hazards model and to the projection pursuit.

REFERENCES

[1] P. Bickel and D. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist. 9 (1981), pp. 1196–1217.

[2] P. Bickel and D. Freedman, *Bootstrapping Regression Models with Many Parameters. A Festschrift for Erich L. Lehman*, Wadsworth, Belmont, C.A., 1983, pp. 28–48.

[3] N. J. Delaney and S. Chatterjee, *Use of the bootstrap and cross-validation in ridge regression*, J. Business Economic Statist. 4 (1986), pp. 255–262.

[4] B. Efron, *Bootstrap methods: Another look at the jackknife*, Ann. Statist. 7 (1979), pp. 1–26.

[5] B. Efron, *The Jackknife, the Bootstrap and Other Resampling Plans*, Regional Conference Series in Applied Mathematics, Monograph 38, SIAM, Philadelphia 1982.

[6] B. Efron and R. J. Tibshirani, *Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy*, Statist. Sci. 1 (1986), pp. 54–77.

[7] D. Freedman, *Bootstrapping regression models*, Ann. Statist. 9 (1981), pp. 1218–1228.

[8] D. Freedman, *On bootstrapping two-stage least squares estimates in stationary linear models*, Ann. Statist. 12 (1984), pp. 827–842.

[9] D. Freedman and S. C. Peters, *Bootstrapping an econometric model: some empirical results*, J. Business Economic Statist. 2 (1984), pp. 150–158.

[10] P. Hall and E. Mammen, *On general resampling algorithms and their performance in distribution estimation*, Ann. Statist. 22 (1994), pp. 2011–2030.

[11] S. Holm, *Abstract bootstrap confidence intervals in linear models*, Scand. J. Statist. 20 (1990), pp. 157–170.

[12] W. Navidi, *Edgeworth expansions for bootstrapping regression models*, Ann. Statist. 17 (1989), pp. 1472–1478.

[13] M. Stone and R. J. Brooks, *Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression*, J. Roy. Statist. Soc. Ser. B 52 (1990), pp. 237–269.

[14] C. F. J. Wu, *Jackknife, bootstrap and other resampling methods in regression analysis*, Ann. Statist. 14 (1986), pp. 1261–1295.

14 Academici st., Faculty of Mathematics
Bucharest University, Romania

Address for correspondence:

str. Eftimiu Mache, Nr. 25A
Bucharest Sector 5, Romania

Received on 25.8.1998