The Annals of Statistics
1984, Vol. 12, No. 4, 1400–1424

BOOTSTRAP AND CROSS-VALIDATION ESTIMATES OF THE PREDICTION ERROR FOR LINEAR REGRESSION MODELS

BY OLAF BUNKE AND BERND DROGE

Humboldt-University Berlin

Different estimates of the mean squared error of prediction for linear regression models are derived by the bootstrap and cross-validation approaches. A comparison is made under normal error distributions, especially by the biases and the mean square errors. The results indicate that the bias corrected bootstrap estimator is best unbiased and should be the first choice, while its simulated variant has approximately the same behaviour. On the other hand, if only a comparison between uncorrected estimators is made (with implications for nonlinear regression models in mind), then other variants of bootstrap estimates are preferable for a large or a small dimension of the model parameter. For a small dimension, the cross-validation estimate and sometimes grouped variants of it seem also to be acceptable if the model error is known to be small.

1. Introduction. One of the most frequent uses of a regression model is the prediction of future values of the dependent variable for some fixed values of the explanatory variables. The mean squared error of prediction (MSEP) describes the performance of the model and estimates of it are of interest. Such estimates have been proposed as criteria for the comparison and selection of regression models and of the explanatory variables in these models (see e.g. the surveys by Hocking, 1976, and Thompson, 1978). Model selection will usually be a more complex process than merely comparing models by a criterion like MSEP, but such comparisons are an essential tool in a (possibly stepwise) strategy, which additionally could include regression diagnostics, transformation analysis, outlier and variable elimination procedures, etc. as e.g. discussed in Weisberg (1981), Montgomery and Peck (1982) or in Bunke (1984). Nevertheless, after finally selecting a hopefully good model, a good estimate of its MSEP is needed as an assessment of its prediction performance.

Recently, the problem of estimating the prediction error has been investigated in papers of Efron (1979, 1983) in a general setup, including regression and discriminant analysis. There, the main interest was the introduction and discussion of an interesting and somewhat unorthodox approach: the bootstrap and other variants of resampling. His simulation studies indicate that for the special discriminant problems investigated there, some variants of the bootstrap approach lead to better estimators of the prediction error than other approaches, such as cross-validation.

Our paper is intended to provide additional insight into the behaviour of bootstrap error estimates by an exact comparison of the performance of different

Received October 1982; revised April 1984.
AMS 1980 *subject classifications.* Primary 62G05; secondary 62J05.
Key words and phrases. Linear regression, model selection, prediction error, bootstrap approach, cross-validation.

bootstrap and cross-validation estimators of the prediction error in a case where this is possible, namely for linear regression models with normally distributed observations. For the sake of completeness we will include in our investigation some other estimates like corrected residual sum of squares or Mallows' C_p, which have been considered elsewhere.

We choose a setup with replicated observations, which allows one to estimate the unknown variance without any restriction on the observation mean vector (as e.g. to be in a linear subspace of R^n) and, moreover, to introduce also a balanced grouped variant of cross-validation into the comparison. Results like those in Section 5 may also be performed in a similar way for other setups, e.g. without replications but with the less realistic assumption that the observation mean varies in a given linear subspace.

The model and all estimators of MSEP are introduced in Section 2, except for the bootstrap estimators, which are defined in Section 3. Section 4 presents the biases and the mean square errors of the estimators of MSEP, which are the basis for their comparison in Sections 5 and 6. The comparison and the numerical values described in Section 7 are discussed in Section 8. They show the essential advantages of the bootstrap approach and are in accordance with the findings of Efron (1983) for the discriminant rules. Almost all of the results were obtained in the thesis of Droge (1982) under the guidance of the co-author.

2. The MSEP and its estimators.

We assume a linear model

$$y = (1_h \otimes I_m)\mu + \varepsilon = A\mu + \varepsilon \tag{2.1}$$

for the random vector $y = (y_1, \cdots, y_n)^T$ of $n = mh$ observations with the standard assumptions

$$E\varepsilon = 0, \quad D\varepsilon = \sigma^2 I_n \tag{2.2}$$

and unknown parameters

$$\mu \in R^m, \quad \sigma^2 > 0. \tag{2.3}$$

In (2.1) we use the h-vector 1_h whose components are all equal to one, the $m \times m$ unit matrix I_m and the Kronecker product $\otimes$. For the integers m, h we assume $m \geq 1$, $h > 1$.

REMARK 1. (2.1) may be interpreted as h replicated observations of a dependent variable Y for each of m different fixed values x_i (design points) of a vector x of explanatory variables. The components of μ would be the values $\mu_i = g(x_i)$ of the response or regression function $g(x) = E_x Y$. Estimating these values μ_i or predicting future values, say z_i, of the dependent variable Y for the design points x_i are problems of practical interest.

The prediction problem could be formulated as predicting the values of a random vector z obeying

$$Ez = \mu, \quad Dz = \sigma^2 I_m, \tag{2.4}$$

assuming that in (2.1), (2.2), (2.4) the same parameters μ, σ^2 appear and that y

and z are uncorrelated. The performance of a predictor $\hat{z} = \hat{z}(y)$ of z depending on the observations will be described by a mean squared error (MSEP)

$$r = E\| z - \hat{z}(y) \|_W^2, \tag{2.5}$$

where the expectation is over both random vectors z and y.

In (2.5) we use the notations $\| z \|_W^2 = z^T W z$ and $W = \mathrm{Diag}[w_1, \cdots, w_m]$, where the positive weights w_i fulfil $\sum_{i=1}^m w_i = 1$. The weights measure the relative importance of prediction errors for the components of z as in Mallows (1973) or Bendel and Afifi (1977).

The prediction

$$\hat{z}(y) = X\hat{\beta}(y) \tag{2.6}$$

is assumed to be performed using a possibly inadequate linear model $X\beta$ approximating the mean vector μ. In this paper we assume that X is an $m \times p$-matrix of rank $p \leq m$.

The estimator $\hat{\beta}$ could be the ordinary least squares estimator (OLSE), but we prefer to use the linear estimator $\hat{\beta}$ minimizing the MSEP (2.5) among all linear estimators $\tilde{\beta} = Ty$ with minimum "bias"

$$\| E(z - X\tilde{\beta}) \|_W^2 = \Delta = \min_\beta \| \mu - X\beta \|_W^2. \tag{2.7}$$

This estimator is just the "weighted LSE"

$$\hat{\beta} = (X^T W X)^{-1} X^T W \hat{\mu}, \tag{2.8}$$

where $\hat{\mu}$ is the OLSE of μ in the model (2.1):

$$\hat{\mu} = h^{-1} A^T y, \quad A = 1_h \otimes I_m. \tag{2.9}$$

Its optimality is also valid in the class of all estimators $\tilde{\beta}$ with minimum bias if (y, z) follows a normal distribution (see Bunke, 1973, or Bunke and Bunke, 1984).

REMARK 2. Each component $\hat{\mu}_i$ of $\hat{\mu}$ is the mean of h components

$$y_{(k-1)m+i} (k = 1, \cdots, h)$$

of y, namely of those with expectation μ_i. The predictor $\hat{z}$ is the oblique projection $\hat{z} = P\hat{\mu}$ with

$$P = X(X^T W X)^{-1} X^T W = ((p_{ij})). \tag{2.10}$$

REMARK 3. The usual unbiased estimator of σ^2 in model (2.1) is

$$\hat{\sigma}^2 = (n - m)^{-1} \| y - A\hat{\mu} \|_I^2. \tag{2.11}$$

With the notation

$$B = WP, \quad t = h^{-1} \operatorname{tr} B \tag{2.12}$$

we may write the MSEP (2.5) as

$$r = \Delta + \sigma^2(1 + t) = \| \mu - P\mu \|_W^2 + \sigma^2(1 + t). \tag{2.13}$$

Several approaches to estimating the MSEP are possible:

1. The weighted sum of squared residuals

$$\text{RSS} = \| y - \hat{y} \|_V^2 = r_{\text{AE}} \tag{2.14}$$

is a rough empirical estimate, which corresponds to the "apparent prediction error" r_{AE} in the terminology of Efron. In (2.14) we use the vector $\hat{y} = AX\hat{\beta} = 1_h \otimes \hat{z}$ of "fitted y-values", and the diagonal matrix $V = h^{-1}I_h \otimes W$, which gives the appropriate weights corresponding to the weights w_i in (2.5). The RSS has a negative bias $-2t\sigma^2$ (see Appendix A1) as an estimator of the MSEP and a correction term yields the unbiased estimator

$$\hat{r} = \text{RSS} + 2\hat{\sigma}^2 t. \tag{2.15}$$

2. A somewhat better idea is to calculate a "plug-in" estimator, replacing in (2.13) the parameters μ, σ^2 by their estimates $\hat{\mu}$, $\hat{\sigma}^2$:

$$r_B = \| \hat{\mu} - \hat{z} \|_W^2 + \hat{\sigma}^2(1 + t). \tag{2.16}$$

In general it will be biased (see Table 1) and its corrected version would be

$$r_{UB} = \| \hat{\mu} - \hat{z} \|_W^2 + \hat{\sigma}^2(1 - h^{-1} + 2t). \tag{2.17}$$

As we will see in Section 3, it may be derived as a bootstrap estimator. In the case of equal weights, the estimators (2.15) and (2.17) coincide and are just the C_p-criterion in essence. Its generalization (2.17) with different weights has already been suggested by Mallows (1973).

3. Another approach to the estimation of a prediction error is cross-validation, that is, an estimate of the form

$$r_c = \| y - \hat{y}_- \|_V^2 \tag{2.18}$$

where each component $\hat{y}_{-j}$ of $\hat{y}_- = (\hat{y}_{-1}, \cdots, \hat{y}_{-n})^T$ is the jth fitted value calculated holding out the observation y_j. Analogously as in Stone (1974) we may express (2.18) by the vector $e = y - \hat{y} = (e_1, \cdots, e_n)^T$ of ordinary residuals in the form $r_c = \| e \|_{\tilde{C}}^2$ where

$$\begin{gathered} \tilde{C} = I_h \otimes C, \quad C = \text{Diag}[c_1, \cdots, c_m] \\ c_i = hw_i[h - p_{ii}]^{-2} \end{gathered} \tag{2.19}$$

(the term inside the brackets is positive because of $p_{ii} \leq 1$ (see Belsley, Kuh and Welsch, 1980)).

4. A grouped variant of cross-validation, too, may be defined in a somewhat balanced way, leaving out one of h different groups of observations, $y^{(k)} = (y_{(k-1)m+1}, \cdots, y_{(k-1)m+m})^T$, $k = 1, \cdots, m$, each with the mean $Ey^{(k)} = \mu$. This allows some interesting insights into the behaviour of grouped cross-validation estimation in comparison with others. The estimate is

$$r_c^h = h^{-1} \textstyle\sum_{k=1}^h \| y^{(k)} - \hat{y}^{(-k)} \|_W^2, \tag{2.20}$$

where for each k, $\hat{y}^{(-k)}$ is the vector of fitted values calculated without the observations in $y^{(k)}$. We have $\hat{y}^{(-k)} = P\hat{\mu}_{-k}$, where $\hat{\mu}_{-k} = (h-1)^{-1} \sum_{l \neq k} y^{(l)}$ is the OLSE of μ calculated without the observations in $y^{(k)}$.

3. Bootstrap estimates. The bootstrap approach is an idea of broad applicability, explained in Efron (1979). For its formal application we assume that the components of the standardized error vector $\tilde{\varepsilon} = \sigma^{-1}\varepsilon$ in (2.1) are i.i.d. with a p.d. Q. We use the empirical p.d. $\hat{Q}$ of the "standardized residuals" $\tilde{e}_i$, that is, of the components of $\tilde{e} = s^{-1}(y - A\hat{\mu})$, where $s^2 = n^{-1} \| y - A\hat{\mu} \|_I^2$. The p.d. $\hat{Q}$ is an estimate of the p.d. Q and both have the same mean zero and variance one, which is the reason for introducing a standardization of the residuals that is different from the usual one (see Belsley, Kuh and Welsch, 1980). We observe that in the definition (2.5) of the MSEP the expectation has to be taken under the true μ, σ^2 and Q. A variant of the bootstrap approach would consist in their replacing by the estimates $\hat{\mu}$, $\hat{\sigma}^2$, $\hat{Q}$. This leads directly to the estimate r_B in (2.16), because the MSEP depends only on μ and σ^2, namely, in the form (2.13). The Monte-Carlo variant of Efron (1979) would be defined based on N simulated bootstrap samples $\varepsilon^1, \cdots, \varepsilon^N$, where the components of the random n-vectors ε^u are i.i.d. with p.d. $\hat{Q}$.

For each "bootstrap observation"

$$y^u = A\hat{\mu} + \hat{\sigma}\varepsilon^u \quad (u = 1, \cdots, N)$$

we can calculate the corresponding estimates $\hat{\mu}^u$, $\hat{z}^u$. The simulated bootstrap estimate is

$$r_B^N = \hat{\sigma}^2 + N^{-1} \textstyle\sum_{u=1}^N \| \hat{\mu} - \hat{z}^u \|_W^2 = N^{-1} \sum_u r_B^u \tag{3.1}$$

and it may be interpreted as approximation to r_B. The simulated bootstrap estimator may also be corrected for bias (see Table 1) and provides

$$r_{UB}^N = r_B^N + \hat{\sigma}^2(t - h^{-1}). \tag{3.2}$$

If we followed the approach in Efron (1979, 1982), we would be tempted to define a bootstrap estimate in the following way: First we find the bootstrap estimate $\widehat{\text{EEE}}$ of the "expected excess error"

$$\text{EEE} = r - E \| y - \hat{y} \|_V^2 = 2\sigma^2 t$$

(see A1) and then we use

$$r_{\text{BOOT}} = \| y - \hat{y} \|_V^2 + \widehat{\text{EEE}}. \tag{3.3}$$

With $\widehat{\text{EEE}} = 2\hat{\sigma}^2 t$ we see that r_{BOOT} is identical with $\hat{r}$ given by (2.15). Intuitively it is felt and somehow apparent from numerical results of Hinkley (1978) for similar estimators in other problems that a bias correction like $\widehat{\text{EEE}}$ may considerably increase the variability of the apparent error r_{AE}, which, moreover, in many cases will be a bad estimator of r. This latter statement will be confirmed later by the mean square error (MSE) of r_{AE} in comparison with that of $\hat{r}$ and r_{UB} (see Table 2 and Droge, 1982). In Section 5 it will also be shown that $r_{\text{BOOT}} = \hat{r}$ is outdone by r_{UB}, except for cases where $\hat{r} = r_{\text{UB}}$.

Another variant of the bootstrap approach would be to use the decomposition

$$r = \sigma^2(1 - h^{-1}) + E\|\hat{\mu} - \hat{z}\|_W^2 + 2E(\hat{z} - \mu)^T W(\hat{\mu} - \mu), \tag{3.4}$$

where the second term can be estimated by its unbiased estimator $\|\hat{\mu} - \hat{z}\|_W^2$. The remaining terms can be estimated using the bootstrap approach, and it is easily seen that this yields directly the unbiased estimator (2.17). We remark incidentally that because of this property and because no bias correction seems to be necessary, a simulation variant of this approach would be convenient in the case of nonlinear regression models, where we have no simple expression for the MSEP like (2.13).

4. Mean and MSE of the estimates. As we explained in Section 1, it is our aim, if possible, to find the best of the different estimators of the MSEP introduced in Sections 2 and 3, most of which have been favoured by different authors. At least we would like to compare their biases and mean square errors under special conditions, where there is more insight into the structure of the corresponding relatively intricated MSE formulae. For this we assume in the following that the observational vector y follows a normal p.d. As all estimates of r in the previous sections were quadratic in the observations, their expectations and variances may be calculated in a straightforward manner with some algebra and using the well-known formulae

$$E\|y\|_T^2 = \|\mu\|_T^2 + \sigma^2 \operatorname{tr} T \tag{4.1}$$

$$D\|y\|_T^2 = 4\sigma^2\|\mu\|_{T^2}^2 + 2\sigma^4 \operatorname{tr} T^2 \tag{4.2}$$

for a vector y with normal p.d. $N(\mu, \sigma^2 I)$. This is done in A2 and the results are compiled in the following tables. To make the simulated bootstrap estimator r_B^N comparable with the grouped cross-validation r_c^h in the numerical effort, we use $N = h$ so that we have to calculate h LSE's in both of them.

We use the notation $\xi = P\mu$.

$$\lambda = 4\sigma^2 h^{-1}\|\mu - \xi\|_{W^2}^2, \quad \tilde{\lambda} = 4\sigma^2\|\mu - \xi\|_G^2 \tag{4.3}$$

$$\tilde{P} = h^{-1}J \otimes P = ((\tilde{p}_{ij})), \quad J = 1_h 1_h^T \tag{4.4}$$

$$K = (I_n - \tilde{P})^T \tilde{C}(I_n - \tilde{P}), \quad G = hC(I_m - P)(I_m - P)^T C. \tag{4.5}$$

The variances for the different estimators are calculated in A2, II, and together with the formulae in Table 1 they lead to the $\operatorname{MSE}(\tilde{r}) = E|\tilde{r} - r|^2$ of those estimators $\tilde{r}$.

5. Comparisons of the estimators. First we may investigate the bias for each biased estimator by looking at Table 1. This yields

THEOREM 1. *The following inequalities hold for the bias* $\operatorname{Bias}(\tilde{r}) = E\tilde{r} - r$:

1. $0 < \operatorname{Bias}(r_c^h) \le h^{-1}(h-1)^{-1}\sigma^2$
2. $0 \le \operatorname{Bias}(r_B) = \operatorname{Bias}(r_B^N) < h^{-1}\sigma^2$

TABLE 1
Means of estimators

Estimator $\tilde{r}$	Mean $E\tilde{r}$
$\hat{r}$ (2.15)	$r = \Delta + \sigma^2(1 + t)$
r_c (2.18)	$h \| \mu - \xi \|_C^2 + \sigma^2 \, tr \, K$
r_c^h (2.20)	$\Delta + \sigma^2[1 + h(h-1)^{-1}t]$
r_B (2.16)	$\Delta + \sigma^2(1 + h^{-1})$
r_{UB} (2.17)	r
r_B^N (3.1)	$\Delta + \sigma^2(1 + h^{-1})$
r_{UB}^N (3.2)	r
r_{AE} (2.14)	$\Delta + \sigma^2(1 - t)$

TABLE 2
MSE *of estimators*

$\tilde{r}$	MSE($\tilde{r}$)
$\hat{r}$	$\lambda + 2\sigma^4\{h^{-2}\mathrm{tr}[hW^2 - 2WB + B^2] + 4n^{-1}t + 4(n-m)^{-1}t^2\}$
r_c	$\tilde{\lambda} + 2\sigma^4\mathrm{tr}\, K^2 + [\| \mu - \xi \|_{hC-W}^2 + \sigma^2(\mathrm{tr}\, K - 1 - t)]^2$
r_c^h	$\lambda + \sigma^4\{2h^{-1}\mathrm{tr}[W^2 + 2(h-1)^{-1}WB + (h^2 + h - 1)(h-1)^{-3}B^2] + (h-1)^{-2}t^2\}$
r_B	$\lambda + \sigma^4 2h^{-2}\mathrm{tr}(W - B)^2 + 2(n-m)^{-1}(1+t)^2 + (h^{-1} - t)^2$
r_{UB}	$\lambda + 2\sigma^4\{h^{-2}\mathrm{tr}(W - B)^2 + (n-m)^{-1}(1 - h^{-1} + 2t)^2\}$
r_B^h	$\mathrm{MSE}(r_B) + 2h^{-3}(n-m)^{-1}\sigma^4[(n - m + 2)\mathrm{tr}\, B^2 - 3h^{-1}\sum_{i=1}^m w_i^2 p_{ii}^2]$
r_{UB}^h	$\mathrm{MSE}(r_{UB}) + 2h^{-3}(n-m)^{-1}\sigma^4[(n - m + 2)\mathrm{tr}\, B^2 - 3h^{-1}\sum_{i=1}^m w_i^2 p_{ii}^2]$
r_{AE}	$\lambda + 2\sigma^4\{h^{-2}\mathrm{tr}[hW^2 - 2WB + B^2] + 2t^2\}$

3. $\mathrm{Bias}(r_B) \geq \mathrm{Bias}(r_c^h)$ *iff*

$$t \leq h^{-1} - h^{-2} \tag{5.1}$$

4. $\mathrm{Bias}(r_c) > 0$ *if* $\sum_{j=1}^m p_{ij}^2 \geq p_{ii} \quad (i = 1, \cdots, m)$

5. $\mathrm{Bias}(r_c) = O(h^{-1})$.

The relations 3, 4 and 5 are proved in A3 while 1 and 2 follow from the inequality

$$0 < t = h^{-1} \sum_{i=1}^m p_{ii} w_i \leq h^{-1} \sum_{i=1}^m w_i = h^{-1}. \tag{5.2}$$

We see that the grouped cross-validation estimate r_c^h overestimates r in the average and that in most cases (if condition (5.1) is fulfilled) it has a smaller bias magnitude than the also nonnegatively biased bootstrap estimator r_B. With infinitely increasing number h of replications, the bias of all the estimators r_B, r_B^N, r_c tends to zero with order h^{-1}, that of the grouped cross-validation estimator r_c^h even with order h^{-2}.

The unbiased estimator r_{UB} depends only on the sufficient and complete statistics $(\hat{\mu}, \hat{\sigma}^2)$ and therefore

THEOREM 2. *r_{UB} is a best unbiased estimator of r.*

Comparing the MSE for the different estimators, we first state that obviously the corrected bootstrap estimator r_{UB} is uniformly better than (or equivalent to) other unbiased estimators as e.g. $\hat{r}$ and r_{UB}^h. This is also true for the biased bootstrap estimator r_B, which is also uniformly better than the simulation bootstrap r_B^h in essence. Together with the other interesting relations this is stated in:

THEOREM 3. (Proof in A4)

1. $2n^{-1}(n-m)^{-1}(n-m-1)t^2\sigma^4 \le \mathrm{MSE}(r_{\mathrm{UB}}^h) - \mathrm{MSE}(r_{\mathrm{UB}})$
$$= \mathrm{MSE}(r_B^h) - \mathrm{MSE}(r_B)$$
$$\le 2h^{-1}(n-m)^{-1}(n-m+2-3n^{-1})t^2\sigma^4$$
2. $\mathrm{MSE}(r_{\mathrm{UB}}^h) < \mathrm{MSE}(\hat{r})$ *if* $\mathrm{tr}\ W^2 \ge m^{-2}(m+1)$ *and* $h \ge m+1 \ge 3$
3. $\mathrm{MSE}(r_{\mathrm{UB}}) \le \mathrm{MSE}(r_B)$ *and* $\mathrm{MSE}(r_{\mathrm{UB}}^h) \le \mathrm{MSE}(r_B^h)$, *where equality holds iff* $t = h^{-1}$, *which is fulfilled for* $p = m$.

Concerning the condition in 2 of Theorem 3, it should be remarked that $\mathrm{tr}\ W^2$ fulfils $m^{-1} \le \mathrm{tr}\ W^2 \le 1$ as a consequence of Jensen's inequality.

It is perhaps interesting that for large sample sizes (large h, m fixed) the estimators r_{UB}, r_{UB}^N and r_B^N are equivalent in their MSE and better than the equivalent estimators r_{AE}, $\hat{r}$, r_c, r_c^h, that is, for large samples the bootstrap approach leads to better estimates than cross-validation. More precisely, with the notation

(5.3) $$\tau = 4\sigma^2 \|\mu - \xi\|_{W^2}^2 + 2\sigma^4 m^{-1}$$

(5.4) $$\kappa = \tau + 2\sigma^4(\mathrm{tr}\ W^2 - m^{-1}),$$

we prove in A5:

THEOREM 4. *For the asymptotic MSE*

(5.5) $$M(\tilde{r}) = \lim_{h\to\infty} h\ \mathrm{MSE}(\tilde{r}) = \lim_{h\to\infty} hD\tilde{r}$$

of the following estimators $\tilde{r}$ it holds

(5.6) $$M(r_{\mathrm{UB}}) = M(r_B) = M(r_B^N) = \tau \le \kappa$$

(5.7) $$M(r_{\mathrm{AE}}) = M(\hat{r}) = M(r_c) = M(r_c^h) = \kappa.$$

6. The case of equal weights. A deeper insight into the behaviour of the estimators is obtained in the case of equal weights

(6.1) $$w_1 = w_2 = \cdots = w_m = m^{-1},$$

because there is a considerable simplification in the formulae. In this case we have

$$W = m^{-1}I_m, \quad t = h^{-1}\mathrm{tr}[m^{-1}X(X^TX)^{-1}X^T] = n^{-1}p$$

and the corrected bootstrap estimator r_{UB} is identical with

$$\hat{r} = r_{\mathrm{BOOT}} = r_{\mathrm{UB}} = n^{-1}\{\|y - \hat{y}\|_I^2 + 2p\hat{\sigma}^2\}. \tag{6.2}$$

The derivations in A6 provide

THEOREM 5. *Assuming* (6.1) *we obtain*

1. $\mathrm{MSE}(\hat{r}) \le \mathrm{MSE}(r_{\mathrm{UB}}^h) < \mathrm{MSE}(r_c^h)$
2. $\mathrm{MSE}(r_c^h) < \mathrm{MSE}(r_B)$ *iff* $p < a_h$, *where* a_h *is given in* (A9) *and* $a_h \to m$ *for fixed* m *and* $h \to \infty$.
3. $\mathrm{MSE}(r_c^h) \le \mathrm{MSE}(r_B^h)$ *if* $p \le b_h$ *and* $\mathrm{MSE}(r_c^h) \ge \mathrm{MSE}(r_B^h)$ *if* $p \ge d_h$, *where* b_h, d_h *are given by* (A10), (A11) *and* $b_h \to m$, $d_h \to m$ *for fixed* m *and* $h \to \infty$.

From this theorem we learn that the corrected exact and simulated bootstrap estimators are always to be preferred to the grouped cross-validation r_c^h, while for $p < m$ and a sufficiently large number h of replications, r_c^h is better than the uncorrected bootstrap estimator r_B.

Assuming (6.1), $\hat{r}$ is essentially better than r_{UB}^h. Only under violation of (6.1) r_{UB}^h may be better than $\hat{r}$, namely, if for instance the "unsymmetry" between the weights w_i is high ($\mathrm{tr}\, W^2 \ge m^{-2}[m+1]$) and if $h \ge m + 1 \ge 3$. Then the corrected simulated bootstrap estimator (and obviously also r_{UB}) will be better than $\hat{r}$.

As it can be noticed, a comparison of the cross-validation estimate with the other estimates is relatively cumbersome, even in the case of equal weights. In the following theorem (proved in A7), some bounds are given for its bias and MSE. In particular, we state that r_c overestimates r in the average (see 4 in Theorem 1, where we may use $\sum_{j=1}^m p_{ij}^2 = p_{ii}$, which is fulfilled with (6.1) because of $P^TP = P$).

THEOREM 6. *With* (6.1) *it holds that*

1. $n^{-2}p^2\sigma^2 \le \mathrm{Bias}(r_c) \le (h-1)^{-2}(2h-1)\Delta + n^{-1}(h-1)^{-1}p\sigma^2$
2. $\underline{M} < \mathrm{MSE}(r_c) \le \bar{M}$, *where*

$$\underline{M} = \lambda + 2n^{-2}(n-p)\sigma^4 + n^{-4}p^4\sigma^4,$$

$$\bar{M} = (h-1)^{-4}h^4\lambda + 2n^{-2}(h-1)^{-4}h^4(n-p)\sigma^4 + [(h-1)^{-2}(2h-1)\Delta + n^{-1}(h-1)^{-1}p\sigma^2]^2.$$

For $p = m$ *we have* $\mathrm{MSE}(r_c) = \bar{M}$.

Under the assumption that $h \ge 6(\max_{i=1}^m p_{ii})$ it is possible to obtain a better lower bound (see A8)

$$\mathrm{MSE}(r_c) \ge \underline{M}^* = \lambda + n^{-2}\sigma^4[2p + 2(n-p)^{-4}n^4(n-2p) + n^{-2}p^4] \tag{6.3}$$

which is relatively sharp, since in the special case

(6.4) $$p_{ii} = \text{const}(= m^{-1}p), \quad i = 1, \cdots, m,$$

the following relation holds (see also A8) with $\Delta = 0$:

(6.5) $$\begin{aligned}\text{MSE}(r_c) - \underline{M}^* \\ = n^{-2}\sigma^4[2((n-p)^{-4}n^4 - 1)p + n^{-2}(n-p)^{-2}(2n-p)p^5].\end{aligned}$$

For the special case (6.4) a comparison of r_c with other estimators is possible (see A9) and places cross-validation behind the others:

THEOREM 7. *With* (6.1) *and* (6.4) *it holds that*

1. $\text{MSE}(\hat{r}) < \text{MSE}(r_c)$ *if* $h \geq 3$
2. $\text{MSE}(r_c^h) = \text{MSE}(r_c)$ *if* $p = m$.

In the case (6.4) it is also possible to derive the sufficient and necessary condition

(6.6) $$(n-p)(m-p)\sigma^2 \leq m(2n-p)(n-1)\Delta \quad \text{(see A10)}$$

for $\text{Bias}(r_c^h) \leq \text{Bias}(r_c)$, which is e.g. fulfilled for $p = m$.

7. Numerical results. To get an impression of the differences between the estimators, we have calculated, under (6.1), some quantities characterizing the MSE itself. A selection of these numerical results is contained in Table 3. We calculated:

$$\rho(\tilde{r}) = \sigma^{-4}[\text{MSE}(\tilde{r}) - \lambda]$$

for estimators $\tilde{r}$ with the exception of the simulated bootstrap estimators r_B^h and r_{UB}^h, in the case of r_c only with the assumptions (6.4) and $\Delta = 0$, denoting

$$\rho_0(r_c) = \rho(r_c) = \sigma^{-4}\text{MSE}(r_c) \quad (\Delta = 0).$$

Additionally we calculated for

$$\tilde{\rho}(r_c) = \sigma^{-4}[\text{MSE}(r_c) - \tilde{\lambda}]$$

the upper and lower bounds $\bar{\rho}(r_c)$, $\underline{\rho}(r_c)$ according to 2 in Theorem 6, and

$$\rho^*(r_c) = \sigma^{-4}(\underline{M}^* - \lambda)$$

with $\underline{M}^*$ from (6.3), which is a lower bound for $\rho(r_c)$ if $h \geq 6\ (\max_{i=1}^m p_{ii})$.

Because of $\tilde{\lambda} \geq \lambda$, implying $\tilde{\rho}(r_c) \leq \rho(r_c)$, it is obvious that $\underline{\rho}(r_c)$ is also a lower bound for $\rho(r_c)$, while, in general, $\bar{\rho}(r_c)$ is only an upper bound for $\rho(r_c)$ if $\Delta = 0$.

Further we should note that, with (6.4) and without the assumption that $\Delta = 0$, it holds that $\rho_0(r_c) = \tilde{\rho}(r_c) \leq \rho(r_c)$.

Concerning the estimators r_B^h and r_{UB}^h we calculated upper bounds for ρ:

$$\bar{\rho}(r_B^h)[\bar{\rho}(r_{\text{UB}}^h)] = \rho(r_B)[\rho(r_{\text{UB}})] + 2h^{-1}n^{-2}(n-m)^{-1}(n-m+2)p.$$

We observe that the difference between these upper bounds and the values of

TABLE 3
Some values characterizing the MSE of different estimators of the MSEP

m	h	p	$\rho(r_{AE})$	$\rho(r_{UB}) = \rho(\hat{r})$	$\rho(r_B)$	$\bar{\rho}(r^h_{UB})$	$\bar{\rho}(r^h_B)$	$\rho(r^h_C)$	$\rho_0(r_C)$	$\underline{\rho}(r_C)$	$\bar{\rho}(r_C)$	$\rho^*(r_C)$
5	2	1	0.2200	0.2760	0.7240	0.2900	0.7380	0.4900	0.2745	0.1801	0.2890	0.2640
		3	0.5000	0.5240	0.7560	0.5660	0.7980	1.1300	0.5996	0.1481	2.3300	0.4013
		5	1.1000	0.9000	0.9000	0.9700	0.9700	1.8500	1.8500	1.6250	1.8500	1.6250
	10	1	0.0408	0.0425	0.0558	0.0426	0.0559	0.0430	0.0425	0.0392	0.0597	0.0424
		3	0.0520	0.0478	0.0531	0.0481	0.0534	0.0490	0.0482	0.0376	0.0574	0.0475
		5	0.0760	0.0538	0.0538	0.0542	0.0542	0.0550	0.0550	0.0361	0.0550	0.0529
10	2	1	0.1050	0.1170	0.4680	0.1200	0.4710	0.1725	0.1166	0.0950	1.5225	0.1155
		4	0.2400	0.1920	0.4080	0.2040	0.4200	0.4200	0.1978	0.0816	1.3200	0.1681
		7	0.5550	0.3030	0.4020	0.3240	0.4230	0.7125	0.3997	0.0800	1.1625	0.2181
		10	1.0500	0.4500	0.4500	0.4800	0.4800	1.0500	1.0500	0.1125	1.0500	0.1125
	5	1	0.0408	0.0425	0.0916	0.0426	0.0918	0.0438	0.0425	0.0392	0.0957	0.0424
		4	0.0624	0.0509	0.0775	0.0516	0.0782	0.0556	0.0514	0.0368	0.0902	0.0501
		7	0.1128	0.0607	0.0710	0.0619	0.0722	0.0679	0.0634	0.0348	0.0852	0.0586
		10	0.1920	0.0720	0.0720	0.0737	0.0737	0.0806	0.0806	0.0336	0.0806	0.0682

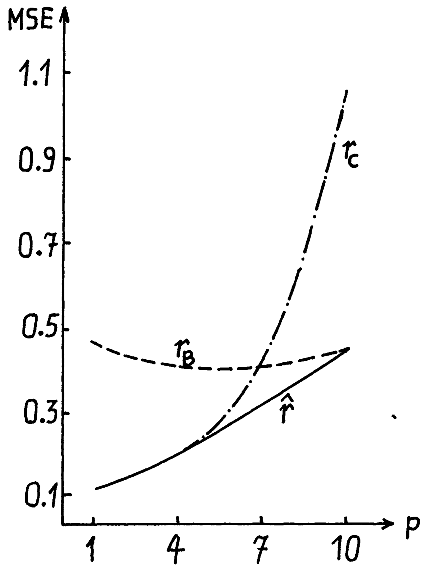

FIG. 1. *MSE vs. p for some estimators of the MSEP in the case* (6.4), $\Delta = 0$, $\sigma = 1$, $m = 10$, $h = 2$.

$\rho(r_B^h)$ and $\rho(r_{\mathrm{UB}}^h)$ is at most $6h^{-1}n^{-3}(h-1)^{-1}p$, since

$$\sum_{i=1}^m w_i^2 p_{ii}^2 = m^{-2} \sum_{i=1}^m p_{ii}^2 \le m^{-2} \sum_{i=1}^m p_{ii} = m^{-2} p$$

(see Table 2) if (6.1) holds.

Figure 1 illustrates the results showing the form of the MSE of $\hat{r}$, r_B, and r_c as functions of p in the case of $\Delta = 0$, $\sigma = 1$, $m = 10$ and $h = 2$.

8. Discussion of the results.

1. From the exact comparisons and the numerical results, it is obvious that the corrected bootstrap estimator r_{UB} should be the first choice among the estimators considered in this paper. The corrected simulated variant r_{UB}^N is approximately as good as r_{UB} if N is large.

2. If we only choose among the estimators r_B, r_c, r_c^h, and r_{BOOT} to investigate an interesting situation like that occuring in nonlinear regression, where it is hard to get reliable unbiased estimators, we will derive the following recommendations, which of course are only preliminary and require further investigation:

(A) For large $p(=m$ or somewhat smaller) use the bootstrap estimator r_B. If the weights are approximately equal (see (6.1)) and $p < m$, one can also use the grouped cross-validation r_c^h, assuming a sufficiently large number h of replications (2 in Theorem 5).

(B) For smaller p use the bootstrap estimate r_{BOOT}.

(C) For small p the cross-validation estimate r_c is not as good as r_{BOOT}, but may be acceptable if the model error Δ is known to be small. The grouped cross-validation r_c^h is also acceptable if h is very large. This could be important in nonlinear cases in which approximate calculations of $\hat{y}_{-i}$ (see Fox, Hinkley and Larntz, 1980) and therefore of r_c are easier than those of r_{BOOT} or of its simulated version.

With respect to the other estimates, the reliability of r_c^h increases with h. Under (6.1) we see from Table 3 that for $h = 5$, r_c^h is better than r_B for $p \lessapprox 0.6m$, while for $h = 10$ this is already the case for $p \lessapprox 0.8m$.

3. With the exception of r_c the difference between the MSE's of estimation does not depend on Δ. Increasing model errors $\Delta > 0$ are increasingly disadvantageous for r_c, and even for moderate h the MSE of the cross-validation estimator r_c could be very large (see Table 2). On the other hand, under (6.1), (6.4), it is obvious by some rough bounds for the MSE (increasing with Δ), that the cross-validation remains a relatively reliable estimator for small dimension p and small Δ. If (6.4) is not fulfilled, the behaviour of r_c will not be essentially better than in the case (6.4), because of the smallness of (6.5).

4. The estimators considered in this paper may also be defined in the more general case of a model

$$Ey = F\alpha, \quad Dy = \sigma^2 I_n, \quad Ez = H\alpha, \quad Dz = \sigma^2 I_m, \tag{8.1}$$

instead of (2.1) and (2.4), where F is an $n \times q$-matrix and H an $m \times q$-matrix, both of rank q, and where $\alpha \in R^q$, $\sigma^2 > 0$ are unknown parameters. (8.1) covers the special cases

(a) $$F = 1_h \otimes I_m, \quad H = I_m,$$

which corresponds to model (2.1) ($\mu = H\alpha = \alpha$),

(b) $$F = I_h \otimes X, \quad H = X,$$

where the projection $\hat{z} = X\hat{\beta}$ is performed with an adequate model,

(c) $$F = \begin{bmatrix} 1_{h_1} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & & \\ 0 & \cdots & 0 & 1_{h_q} \end{bmatrix}, \quad q = m, \quad H = I_m,$$

which describes an experiment with h_i replications at each of $m = q$ different design points (see Remark 1 in Section 2).

As before, the predictor of z will be of the form $\hat{z} = X\hat{\beta}$, assuming $X = HT$ for some matrix T. In (2.8) $\hat{\mu}$ is now defined to be the OLSE

$$\hat{\mu} = H\hat{\alpha} \text{ of } \mu = H\alpha \ (\hat{\alpha} = (F^TF)^{-1}F^Ty).$$

A best unbiased estimator of the MSEP may again be obtained by correcting a plug-in-estimator or, alternatively, by the bootstrap approach and is given by

$$r_{\mathrm{UB}} = \| \hat{\mu} - \hat{z} \|_W^2 + \hat{\sigma}^2\{1 + 2t - \mathrm{tr}[WH(F^TF)^{-1}H^T]\}, \tag{8.2}$$

with

$$\hat{\sigma}^2 = (n-q)^{-1} \| y - F\hat{\alpha} \|_I^2, \quad t = \mathrm{tr}[WPH(F^TF)^{-1}H^T].$$

APPENDIX

(Proof of the statements in Sections 2 to 6)

A1. With (4.1), $\tilde{\mu} = A\mu$ and $\tilde{B} = h^{-2}J \otimes B = \tilde{P}^TV$ we obtain

$$E\| y - \hat{y} \|_V^2 = E\| (I - \tilde{P})y \|_V^2 = E\| y \|_{V-\tilde{B}}^2 = \| \hat{\mu} \|_{V-\tilde{B}}^2 + \sigma^2\mathrm{tr}(V - \tilde{B})$$
$$= \Delta + \sigma^2(1 - t).$$

A2. I. Calculation of means (using (4.1)).

I.1 With K given in (4.4) and $\eta = \tilde{P}\tilde{\mu} = A\xi$ it holds that

$$Er_c = E\| y \|_K^2 = \| \tilde{\mu} - \eta \|_{\tilde{C}}^2 + \sigma^2\mathrm{tr}\, K = h\| \mu - \xi \|_C^2 + \sigma^2\mathrm{tr}\, K \quad \text{(see (2.19))}.$$

I.2 For calculating $E(r_c^h)$ we rewrite (2.20) in the form

$$r_c^h = \| y - \hat{y}^h \|_V^2,$$

where

$$\hat{y}^h = (I_h \otimes P)\begin{pmatrix} \hat{\mu}_{-1} \\ \vdots \\ \hat{\mu}_{-k} \end{pmatrix}.$$

Observing that, because of $\hat{\mu}_{-i} = h(h-1)^{-1}\hat{\mu} - (h-1)^{-1}y^{(i)}$, we have $y - \hat{y}^h = Sy$, where

$$S = \{I_h \otimes [I_m + (h-1)^{-1}P]\} - h(h-1)^{-1}\tilde{P},$$

we may also write

$$\| y - \hat{y}^h \|_V^2 = \| y \|_R^2, \quad \text{where} \quad R = S^TVS.$$

A straightforward algebra provides

$$hR = \{I_h \otimes [W + (2h-1)(h-1)^{-2}B]\} - \{h(h-1)^{-2}(J \otimes B)\}$$

and

$$\mathrm{tr}\, R = 1 + h(h-1)^{-1}t.$$

Then

$$Er_c^h = \| E(y - \hat{y}^h) \|_V^2 + \sigma^2\mathrm{tr}\, R = \Delta + \sigma^2[1 + h(h-1)^{-1}t].$$

I.3 We have

$$E\,\|\hat{\mu} - \hat{z}\|_W^2 = E\,\|\hat{\mu}\|_{W-B}^2 = \|\mu\|_{W-B}^2 + \sigma^2 h^{-1}(1 - \operatorname{tr} B)$$
$$= \Delta + \sigma^2(h^{-1} - t),$$

and therefore the formula for Er_B in Table 1.

I.4 Now and in the following we will work with the useful short notations E^* and D^* for the conditional mean and variance over the bootstrap observations y^u under the condition of a fixed y. Then,

$$Er_B^N = E[E^*(r_B^N)] = E[N^{-1} \textstyle\sum_u E(r_B^u/y)] = Er_B$$

and analogously $Er_{\mathrm{UB}}^N = Er_{\mathrm{UB}} = r \cdot E\hat{r}$ and Er_{AE} follow from A1.

II. Calculation of variances. In the following, the variances of the estimators are calculated basing on (4.2) and with the notation of I.

II.1 With the notation

$$U = V - \tilde{B} + 2(n - m)^{-1}t(I_n - h^{-1}J \otimes I_m)$$

it holds that

$$D\hat{r} = D\,\|y\|_U^2 = 4\sigma^2\|\tilde{\mu}\|_{U^2}^2 + 2\sigma^4 \operatorname{tr} U^2.$$

Now, because of $(I_n - h^{-1}J \otimes I_m)\tilde{\mu} = 0$:

$$\|\tilde{\mu}\|_{U^2}^2 = \|\tilde{\mu}\|_{(V-\tilde{B})^2}^2 = \|\tilde{\mu} - \eta\|_{V^2}^2 = h^{-1}\|\mu - \xi\|_{W^2}^2.$$

A straightforward algebra provides

$$\operatorname{tr} U^2 = h^{-2}[h \operatorname{tr} W^2 - 2\operatorname{tr}(WB) + \operatorname{tr} B^2] + 4n^{-1}t + 4(n - m)^{-1}t^2.$$

These relations show that $D\hat{r} = \mathrm{MSE}(\hat{r})$ has the form given in Table 2. Similarly it can be verified that

$$Dr_{\mathrm{AE}} = \lambda + 2\sigma^4 h^{-2}\operatorname{tr}[hW^2 - 2WB + B^2].$$

II.2 With $L = \tilde{C}(I - \tilde{P})(I - \tilde{P})^T\tilde{C}$ it follows that

$$K^2 = (I - \tilde{P})^T L(I - \tilde{P})$$

(see (4.4)) and

(A1) $$Dr_c = D\,\|y\|_K^2 = 4\sigma^2\|\tilde{\mu} - \eta\|_L^2 + 2\sigma^4 \operatorname{tr} K^2.$$

II.3 Using the notation from I.2 we obtain

$$Dr_c^h = 4\sigma^2\|\tilde{\mu}\|_{R^2}^2 + 2\sigma^4 \operatorname{tr} R^2.$$

Because of $S\tilde{\mu} = \tilde{\mu} - \eta$, $VS = S^TV$, $S(\tilde{\mu} - \eta) = \tilde{\mu} - \eta$:

$$\|\tilde{\mu}\|_{R^2}^2 = \|\tilde{\mu} - \eta\|_{S^TV^2S}^2 = h^{-1}\|\mu - \xi\|_{W^2}^2.$$

Writing R in the form

$$R = I_h \otimes G_1 - J \otimes G_2$$

with

$$hG_1 = W + (2h - 1)(h - 1)^{-2}B, \quad G_2 = (h - 1)^{-2}B$$

we easily obtain

$$\text{(A2)} \qquad \begin{aligned} \operatorname{tr} R^2 &= h \operatorname{tr}(G_1^2 - 2G_1G_2 + hG_2^2) \\ &= h^{-1}\operatorname{tr}[W^2 + 2(h - 1)^{-1}WB + (h^2 + h - 1)^{-3}B^2]. \end{aligned}$$

Finally we have reached

$$\text{(A3)} \qquad Dr_c^h = \lambda + 2\sigma^4 \operatorname{tr} R^2.$$

II.4 The derivation for the variance of bootstrap estimators will be carried out for the more general statistics

$$r_B(\gamma) = \| \hat{\mu} - \hat{z} \|_W^2 + \gamma\hat{\sigma}^2,$$

which provides r_B for $\gamma = 1 + t$ and r_{UB} for $\gamma = 1 + 2t - h^{-1}$. Similarly, if

$$r_B^u(\gamma) = \| \hat{\mu} - \hat{z}^u \|_W^2 + \gamma\hat{\sigma}^2,$$

r_B^N and r_{UB}^N are given by

$$r_B^N(\gamma) = N^{-1} \textstyle\sum_u r_B^u(\gamma)$$

for $\gamma = 1$ and $\gamma = 1 + t - h^{-1}$, respectively. $\hat{\mu}$ and $\hat{\sigma}^2$ are independent and therefore

$$Dr_B(\gamma) = D \| \hat{\mu} - \hat{z} \|_W^2 + \gamma^2 D\hat{\sigma}^2 = D \| \hat{\mu} \|_{W-B}^2 + \gamma^2 D\hat{\sigma}^2.$$

Now, because of $\| \mu \|_{(W-B)^2}^2 = \| \mu - \xi \|_{W^2}^2$,

$$D \| \hat{\mu} \|_{W-B}^2 = \lambda + 2\sigma^4 h^{-2} \operatorname{tr}(W - B)^2$$

and as a consequence

$$Dr_B(\gamma) = \lambda + 2\sigma^4[h^{-2}\operatorname{tr}(W - B)^2 + \gamma^2(n - m)^{-1}].$$

Taking into account the conditional independence of $r_B^u(\gamma)$ and $r_B^v(\gamma)$ (for $u \neq v$) we have

$$\text{(A4)} \qquad \begin{aligned} Dr_B^N(\gamma) &= D\{E^*[N^{-1} \textstyle\sum_u r_B^u(\gamma)]\} + E\{D^*[N^{-1} \textstyle\sum_u r_B^u(\gamma)]\} \\ &= Dr_B(\gamma + t) + E[N^{-1}D^* r_B^u(\gamma)] \quad (u \text{ fixed}), \end{aligned}$$

since

$$E^* r_B^u(\gamma) = r_B(\gamma + t).$$

Because of

$$(\hat{\mu} - \hat{z})^T WP = 0 \quad \text{and} \quad \tilde{P}(h^{-1}J \otimes I_m) = \tilde{P}$$

it holds

$$\|\hat{\mu} - \hat{z}^u\|_W^2 = \|\hat{\mu} - \hat{z}\|_W^2 + \|\hat{z} - \hat{z}^u\|_W^2 = \|\hat{\mu} - \hat{z}\|_W^2 + \hat{\sigma}^2\|\varepsilon^u\|_{\tilde{B}}^2.$$

Thus

$$D^* r_B^u(\gamma) = D^*\|\hat{\mu} - \hat{z}^u\|_W^2 = \hat{\sigma}^4 D^*\|\varepsilon^u\|_{\tilde{B}}^2.$$

Recalling that $E^*\varepsilon^u = 0$ and $D^*\varepsilon^u = I_n$, we derive, because of the well-known formula for the variance of a quadratic form (see Anderson, 1971),

$$D\|\varepsilon\|_T^2 = 2\sigma^4 \text{tr } T^2 + (E\varepsilon_i^4 - 3\sigma^4) \textstyle\sum_i t_{ii}^2,$$

which holds for a vector $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_n)^T$ with $\varepsilon_i \sim (0, \sigma^2)$ i.i.d. and a symmetric matrix $T = ((t_{ij}))$:

$$D^*\|\varepsilon^u\|_{\tilde{B}}^2 = 2 \text{ tr } \tilde{B}^2 + \textstyle\sum_i v_i^2 \tilde{p}_{ii}^2 [E^*(\varepsilon_i^u)^4 - 3].$$

Noting

$$E^*(\varepsilon_i^u)^4 = n(n-m)^{-2}\hat{\sigma}^{-4} \textstyle\sum_{i=1}^m \sum_{k=1}^h (y_{(k-1)m+i} - \hat{\mu}_i)^4,$$

$$E\hat{\sigma}^4 E^*(\varepsilon_i^u)^4 = 3\sigma^4$$

and

$$E\hat{\sigma}^4 = D\hat{\sigma}^2 + \sigma^4 = (n - m + 2)(n - m)^{-1}\sigma^4$$

we come to

$$\begin{aligned} E[D^* r_B^u(\gamma)] &= E\hat{\sigma}^4 D^*\|\varepsilon^u\|_{\tilde{B}}^2 \\ &= 2(n-m)^{-1}\sigma^4[(n-m+2)\text{tr } \tilde{B}^2 - 3 \textstyle\sum_i v_i^2 \tilde{p}_{ii}^2]. \end{aligned}$$

Together with (A4) and tr $\tilde{B}^2 = h^{-2}$tr B^2 the above equation implies

$$\begin{aligned} Dr_B^N(\gamma) &= Dr_B(\gamma + t) \\ &\quad + 2N^{-1}(n-m)^{-1}h^{-2}\sigma^4[(n-m+2)\text{tr } B^2 - 3h^{-1} \textstyle\sum_i w_i^2 p_{ii}^2]. \end{aligned}$$

A3. Statement 3 of Theorem 1 follows immediately from Table 1:

$$\text{Bias}(r_B) \geq \text{Bias}(r_c^h) \quad \text{iff}$$

$$h^{-1} - t \geq (h-1)^{-1}t \quad \text{iff} \quad t \leq h^{-1} - h^{-2}.$$

Concerning statement 4, we note first that the diagonal matrix

$$hC - W = \text{Diag}[w_1\{h^2(h - p_{11})^{-2} - 1\}, \cdots, w_m\{h^2(h - p_{mm})^{-2} - 1\}]$$

is always nonnegative-definite. Thus

$$h\|\mu - \xi\|_C^2 \geq \|\mu - \xi\|_W^2 = \Delta. \tag{A5}$$

By using the assumption $\sum_j p_{ij}^2 \geq p_{ii}$ $(i = 1, \cdots, m)$ we obtain further

$$\begin{aligned} \operatorname{tr} K = \operatorname{tr}[P^TCP + hC - 2CP] &= \textstyle\sum_i w_i h(h - p_{ii})^{-2}(\sum_j p_{ij}^2 + h - 2p_{ii}) \\ &\geq \textstyle\sum_i w_i h(h - p_{ii})^{-1} \\ &= \textstyle\sum_i w_i h^{-1}(h + p_{ii}) + \sum_i w_i h^{-1}(h - p_{ii})^{-1} p_{ii}^2 \\ &= 1 + t + \textstyle\sum_i w_i h^{-1}(h - p_{ii})^{-1} p_{ii}^2, \end{aligned}$$

since $\sum_i w_i = 1$ and $h^{-1} \sum_i w_i p_{ii} = t$. From this and from (A5) it finally results that

$$Er_c \geq \Delta + \sigma^2[1 + t + \textstyle\sum_i w_i h^{-1}(h - p_{ii})^{-1} p_{ii}^2] > r.$$

Statement 5 is obvious from

$$\begin{aligned} \operatorname{Bias}(r_c) = {} & \| \mu - \xi \|_{hC-W}^2 \\ & + [\textstyle\sum_i w_i h^{-1}(h - p_{ii})^{-1} p_{ii}^2 + \sum_i w_i h(h - p_{ii})^{-2}(\sum_j p_{ij}^2 - p_{ii})]\sigma^2, \end{aligned}$$

since $hc_i - w_i = O(h^{-1})$.

A4.

1. Clearly,

$$\textstyle\sum_{i=1}^m w_i^2 p_{ii}^2 \leq \sum_{i,j=1}^m w_i w_j p_{ij} p_{ji} = \operatorname{tr} B^2 \leq (\operatorname{tr} B)^2 = h^2 t^2,$$

and Jensen's inequality yields

$$\textstyle\sum_{i=1}^m w_i^2 p_{ii}^2 \geq m^{-1} (\sum_{i=1}^m w_i p_{ii})^2 = m^{-1} h^2 t^2.$$

Altogether we obtain

$$\begin{aligned} & 2n^{-1}(n - m)^{-1}(n - m - 1)t^2\sigma^4 \\ & \qquad \leq 2h^{-3}(n - m)^{-1}(n - m - 1)\operatorname{tr} B^2 \sigma^4 \\ & \qquad \leq 2h^{-3}(n - m)^{-1}[(n - m + 2)\operatorname{tr} B^2 - 3h^{-1} \textstyle\sum_i w_i^2 p_{ii}^2]\sigma^4 \\ \text{(A6)} \quad & \qquad = \operatorname{MSE}(r_B^h) - \operatorname{MSE}(r_B) = \operatorname{MSE}(r_{\mathrm{UB}}^h) - \operatorname{MSE}(r_{\mathrm{UB}}) \\ & \qquad \leq 2h^{-3}(n - m)^{-1}[(n - m + 2)\operatorname{tr} B^2 - 3m^{-1}ht^2]\sigma^4 \\ & \qquad \leq 2h^{-1}(n - m)^{-1}(n - m + 2 - 3n^{-1})t^2\sigma^4. \end{aligned}$$

2. From Table 2 we see that, with $\operatorname{tr} W^2 \geq m^{-2}(m + 1)$:

$$\operatorname{MSE}(\hat{r}) - \operatorname{MSE}(r_{\mathrm{UB}}) = 2h^{-2}(h - 1)(\operatorname{tr} W^2 - m^{-1})\sigma^4 \geq 2n^{-2}(h - 1)\sigma^4.$$

Using $m \geq 2$ and $t \leq h^{-1}$, (A6) implies

$$\operatorname{MSE}(r_{\mathrm{UB}}^h) - \operatorname{MSE}(r_{\mathrm{UB}}) \leq 2(h - 1)^{-1}h^{-2}\sigma^4,$$

which together with the first inequality leads to

$$\operatorname{MSE}(\hat{r}) - \operatorname{MSE}(r_{\mathrm{UB}}^h) \geq 2n^{-2}(h - 1)^{-1}[(h - 1)^2 - m^2]\sigma^4 \geq 0,$$

since we have assumed that $h \geq m + 1$.

3. Table 2 and (5.2) yield

$$\begin{aligned}\mathrm{MSE}(r_B) &- \mathrm{MSE}(r_{\mathrm{UB}})\\ &= \mathrm{MSE}(r_B^h) - \mathrm{MSE}(r_{\mathrm{UB}}^h)\\ &= (n-m)^{-1}(h^{-1}-t)[(n-m-2)h^{-1}+4-(n-m-6)t]\sigma^4\\ &\geq 0,\end{aligned}$$

since it can be easily verified that

$$(n-m-2)h^{-1}+4 > (n-m-6)t.$$

Therefore, $\mathrm{MSE}(r_B) = \mathrm{MSE}(r_{\mathrm{UB}})$ and $\mathrm{MSE}(r_B^h) = \mathrm{MSE}(r_{\mathrm{UB}}^h)$ hold iff $h^{-1} = t$, which is fulfilled if $p_{ii} = 1$, for $i = 1, \cdots, m$, i.e. if $p = m$ (see (5.2)).

A5. At first we remark that $\lim_{h\to\infty} h\,\mathrm{MSE}(\tilde{r}) = \lim_{h\to\infty} hD\tilde{r}$ for our estimators because of $(E\tilde{r} - r)^2 = O(h^{-2})$. Also,

$$\lim_{h\to\infty} h\tilde{\lambda} = 4\sigma^2 \| \mu - \xi \|_{W^2}^2 = h\lambda,$$

because of

$$\lim_{h\to\infty} hG = W(I_m - P)(I_m - P)^T W = (I_m - P)^T W^2 (I_m - P)$$

and $(I_m - P)(\mu - \xi) = \mu - \xi$ (see (4.3), (4.5)). Thus we have to examine only the asymptotic behaviour of the factor σ^4 in the MSE-formulae of Table 2. For the estimators $\hat{r}$, r_{AE}, r_B, r_{UB}, r_B^N, r_{UB}^N the propositions of the theorem are obvious from the MSE-formulae. In A2, II.3 we see from the formulae (A2) and (A3) that $\lim_{h\to\infty} hDr_c^h = h\lambda + 2\sigma^4 \mathrm{tr}\ W^2$, providing $M(r_c^h) = \kappa$. The assertion for r_c follows from (A1),

$$K = I_h \otimes C - h^{-1}(J \otimes P^T C) - h^{-1}(J \otimes CP) + h^{-1}(J \otimes P^T CP)$$

and

$$\lim_{h\to\infty} h \ \mathrm{tr}\ K^2 = \lim_{h\to\infty} h \ \mathrm{tr}(I_h \otimes C)^2 = \lim_{h\to\infty} h^2 \mathrm{tr}\ C^2 = \mathrm{tr}\ W^2.$$

A6.

1. Because of

$$\mathrm{tr}\ W^2 = m^{-1}, \quad t = n^{-1}p, \quad \mathrm{tr}\ B^2 = \mathrm{tr}\ BW = m^{-2}p, \quad \mathrm{tr}(W-B)^2 = m^{-2}(m-p) \tag{A7}$$

the MSE-formulae in Table 2 lead to

$$\mathrm{MSE}(r_c^h) - \mathrm{MSE}(\hat{r}) = n^{-2}m^{-1}(h-1)^{-3}p[4h^2m + 8(h^2-2h)(m-p) + 6(m-p) + p(n-m-2)]\sigma^4.$$

Thus, under consideration of (A6)

$$\begin{aligned}&\mathrm{MSE}(r_c^h) - \mathrm{MSE}(r_{\mathrm{UB}}^h)\\ &> n^{-2}m^{-1}(h-1)^{-3}p\sigma^4[2hn + 8h(h-2)(m-p)\\ &\qquad + 6(m-p) + (6m-4)(h-1) + 4(1-h^{-1}) + 2mh^{-1}],\end{aligned}$$

which is always positive since $p \le m$ and $h \ge 2$. The inequality $\mathrm{MSE}(\hat{r}) \ge \mathrm{MSE}(r_{\mathrm{UB}}^h)$ follows immediately from $\hat{r} = r_{\mathrm{UB}}$ with (6.1).

2. After some straightforward calculations we obtain from Table 2 by using (A7)

(A8) $$\mathrm{MSE}(r_B) - \mathrm{MSE}(r_c^h) = n^{-2}\sigma^4(a - 2bp + cp^2)$$

where

$$a = m[4 + m + 2(h-1)],$$
$$b = (h-1)^{-1}[n - m + 2h + 2 + 4(h-1)^{-1} + (h-1)^{-2}],$$
$$c = m^{-1}(h-1)^{-2}[nh - 2n + 2h - 2].$$

Obviously a, b, and c are positive. With the notations

$$\psi = (nh - 2n + 2h - 2)^{-1}(n - m),$$
$$\delta = n - m + 2h + 2 + 4(h-1)^{-1} + (h-1)^{-2}$$

and

$$\begin{aligned}\omega = {} & 6n + 6m + m^2 + 4h^2 + 16 + (h-1)^{-1}(4m + 20)\\ & +24(h-1)^{-2} + 8(h-1)^{-3} + (h-1)^{-4}\end{aligned}$$

it follows that

$$b^2 - ac = (h-1)^{-2}\omega > 0,$$

so that the roots of the quadratic function in p

$$a - 2bp + cp^2$$

are given by

(A9) $$a_h[a_h^*] = \psi(\delta - [+]\omega^{1/2}).$$

Clearly,

$$\omega^{1/2} > m, \quad \psi > (n+2)^{-1}m \quad \text{and} \quad \delta > n - m + 2$$

and therefore $a_h^* > m$, from which it follows that

$$\mathrm{MSE}(r_c^h) < \mathrm{MSE}(r_B) \quad \text{iff} \quad p < a_h,$$

recalling that $p \le m$ and $c > 0$. From (A9) and the definition of ψ, δ and ω it can be easily seen that a_h tends to m for fixed m and $h \to \infty$.

3. From the derivations in A4, 1. and using (A7) we obtain

$$2n^{-2}h^{-1}(n-m)^{-1}(n-m-1)p\sigma^4 \le \mathrm{MSE}(r_B^h) - \mathrm{MSE}(r_B) \le 2n^{-2}h^{-1}(n-m)^{-1}(n-m+2)p\sigma^4.$$

Thus, with (A8)

$$n^{-2}\sigma^4(a - 2b_{(1)}p + cp^2) \le \mathrm{MSE}(r_B^h) - \mathrm{MSE}(r_c^h) \le n^{-2}\sigma^4(a - 2b_{(2)}p + cp^2),$$

where a, b and c are defined as in A6, 2. and

$$b_{(1)} = b - h^{-1}(n-m)^{-1}(n-m-1), \quad b_{(2)} = b - h^{-1}(n-m)^{-1}(n-m+2).$$

With the notations

$$\begin{aligned}\omega_0 &= 4n + 10m + m^2 + 4h^2 - 2h + 4(h-1)^{-1}(m+3)\\ &\quad + 22(h-1)^{-2} + 8(h-1)^{-3}\\ &\quad + (h-1)^{-4} + h^{-2} + 8h^{-1}(h-1)^{-1} + 2h^{-1}(h-1)^{-2},\\ \omega_1 &= \omega_0 + 19 + n^{-2} + 2n^{-1}h^{-1} - 2mh^{-1} + 4m^{-1} + 2n^{-1}\\ &\quad + 8n^{-1}(h-1)^{-1} + 2n^{-1}(h-1)^{-2},\\ \omega_2 &= \omega_0 + 13 + 2h^{-1}(3-m) - 4n^{-1} + 4n^{-2} - 2n^{-1}h^{-1}\\ &\quad - 16n^{-1}(h-1)^{-1} - 4n^{-1}(h-1)^{-2},\\ \delta_1 &= n - m + 2h + 1 + 4(h-1)^{-1} + h^{-1} + (h-1)^{-2} + n^{-1}\end{aligned}$$

and

$$\delta_2 = \delta_1 - 3n^{-1}$$

we find

$$b_{(i)}^2 - ac = (h-1)^{-2}\omega_i > 0 \quad (i = 1, 2).$$

The roots b_h, b_h^* and d_h, d_h^* of the quadratic functions in p, $a - 2b_{(1)}p + cp^2$ and $a - 2b_{(2)}p + cp^2$, respectively, are given by

$$b_h[b_h^*] = \psi(\delta_1 - [+]\omega_1^{1/2}) \tag{A10}$$

and

$$d_h[d_h^*] = \psi(\delta_2 - [+]\omega_2^{1/2}), \tag{A11}$$

where ψ is defined as in A6, 2. Similarly as there we have

$$\omega_i^{1/2} > m, \quad \psi > (n+2)^{-1}m \quad \text{and} \quad \delta_i > n - m + 2 \quad (i = 1, 2),$$

providing $b_h^* > m$ and $d_h^* > m$. Taking into account that $p \le m$ and $c > 0$, this shows:

$$p \ge d_h \text{ implies } \mathrm{MSE}(r_B^h) \le \mathrm{MSE}(r_c^h),$$

whereas

$$p \leq b_h \text{ is sufficient for } \mathrm{MSE}(r_B^h) \geq \mathrm{MSE}(r_c^h).$$

From the form of b_h and d_h it is obvious that both terms tend to m for fixed m and $h \to \infty$.

A7.

1. With (6.1) we have by using (2.19) and $PP^T = P = P^T$

$$\operatorname{tr} K = \operatorname{tr}[P^TCP + hC - 2CP] = \operatorname{tr}[hC - CP] = n^{-1}h^2 \textstyle\sum_i (h - p_{ii})^{-1}$$

and therefore (see Table 1)

$$\mathrm{Bias}(r_c) = \| \mu - \xi \|_{hC-W}^2 + \sigma^2 n^{-1} \textstyle\sum_i (h - p_{ii})^{-1} p_{ii}^2.$$

(A5) together with $p_{ii} \leq 1$ provides

$$0 \leq \| \mu - \xi \|_{hC-W}^2 \leq (h-1)^{-2}(2h-1)\Delta,$$

since

$$hc_i - w_i = m^{-1}[h^2(h - p_{ii})^{-2} - 1] \leq m^{-1}(h-1)^{-2}(2h-1) \quad (i = 1, \cdots, m).$$

From

(A12) $$0 \leq p_{ii} \leq 1, \quad p_{ii}^2 \leq p_{ii}, \quad \textstyle\sum_i p_{ii} = p$$

it is obvious that

$$n^{-1} \textstyle\sum_i (h - p_{ii})^{-1} p_{ii}^2 \leq n^{-1}(h-1)^{-1} \textstyle\sum_i p_{ii}^2 \leq n^{-1}(h-1)^{-1} p,$$

and Jensen's inequality provides

$$n^{-1} \textstyle\sum_i (h - p_{ii})^{-1} p_{ii}^2 \geq n^{-1}h^{-1} \textstyle\sum_i p_{ii}^2 \geq n^{-2}p^2.$$

Summarizing the above estimations, we obtain statement 1 of Theorem 6.

2. Let T be an arbitrary idempotent $n \times n$-matrix and $D = \mathrm{Diag}[d_1, \cdots, d_n]$ a diagonal matrix with $d_i > 0$ $(i = 1, \cdots, n)$. Then, for any eigenvector f of T corresponding to the eigenvalue 1 the following relations can be easily verified:

(A13) $$d_i \genfrac{}{}{0pt}{}{\geq}{(\leq)} 1 (i = 1, \cdots, n) \text{ implies } \| f \|_{DTD}^2 \genfrac{}{}{0pt}{}{\geq}{(\leq)} \| f \|_T^2 = \| f \|_I^2.$$

Now, with (6.1) (see (2.18) and A2, II.2) it holds

$$L = \tilde{C}(I - \tilde{P})\tilde{C}, \quad \tilde{C} = n^{-1}h^2\{I_h \otimes \mathrm{Diag}[(h - p_{11})^{-2}, \cdots, (h - p_{mm})^{-2}]\}.$$

Thus, applying (A13):

(A14) $$\begin{aligned} \lambda = 4n^{-1}\Delta\sigma^2 &\leq 4 \| \tilde{\mu} - \eta \|_L^2 \sigma^2 \\ &= \tilde{\lambda} \leq 4n^{-1}(h-1)^{-4}h^4\Delta\sigma^2 = (h-1)^{-4}h^4\lambda, \end{aligned}$$

since

$$(h - p_{ii})^{-2}h^2 \geq 1$$

and

$$[(h - p_{ii})^{-2}h^2][(h - 1)^2h^{-2}] \le 1 \ (i = 1, \cdots, m).$$

We find

$$\text{(A15)} \quad \begin{aligned} \operatorname{tr} K^2 &= \operatorname{tr}[hC^2 + (PC)^2 - 2PC^2] \\ &= n^{-2}[\textstyle\sum_{i,j} p_{ij}^2 h^4(h-p_{ii})^{-2}(h-p_{jj})^{-2} + \sum_i (h-2p_{ii})h^4(h-p_{ii})^{-4}], \end{aligned}$$

and therefore, recalling (A12) and $\sum_j p_{ij}^2 = p_{ii}$,

$$n^{-2}(n - p) < \operatorname{tr} K^2 \le n^{-2}(h - 1)^{-4}h^4(n - p),$$

which proves $\underline{M} < \text{MSE}(r_c) \le \bar{M}$ (see (A14) and 1 of Theorem 6). In the case of $p = m$, the above derivations imply $\text{MSE}(r_c) = \bar{M}$, since then

$$p_{ii} = 1 \quad \text{for} \quad i = 1, \cdots, m.$$

A8.

1. With the notation

$$\varphi(q) = q^{-4}(2q - h), \quad q_i = h - p_{ii} \ (i = 1, \cdots, m)$$

we derive from (A16)

$$n^2 \operatorname{tr} K^2 > p + \textstyle\sum_i (h - p_{ii})^{-4}h^4(h - 2p_{ii}) = p + h^4 \sum_i \varphi(q_i).$$

Straightforward calculations show that φ is a convex function of q iff $6q \ge 5h$. Hence Jensen's inequality yields

$$\begin{aligned} n^2 \operatorname{tr} K^2 > p + h^4 \textstyle\sum_i \varphi(q_i) &\ge p + h^4 m \varphi(m^{-1} \textstyle\sum_i q_i) \\ &= p + (n - p)^{-4}n^4(n - 2p) \end{aligned}$$

for $h \ge 6(\max_i p_{ii})$, and we obtain the inequality (6.3) analogous to 2 in A7.

2. In the special case (6.4) we have

$$\tilde{C} = (n - p)^{-2}nI_n, \quad L = (n - p)^{-4}n^2(I - \tilde{P}),$$
$$K = (n - p)^{-2}n(I - \tilde{P})$$

and therefore

$$\text{(A16)} \quad \begin{aligned} \text{Bias}(r_c) &= (n - p)^{-2}(2np - p^2)\Delta + \sigma^2p^2(n - p)^{-1}n^{-1}, \\ \tilde{\lambda} &= (n - p)^{-4}n^4\lambda, \quad \operatorname{tr} K^2 = (n - p)^{-4}n^2\operatorname{tr}(I - \tilde{P}) = (n - p)^{-3}n^2. \end{aligned}$$

Consequently,

$$\text{(A17)} \quad \text{MSE}(r_c) = (n - p)^{-4}n^4\lambda + 2(n - p)^{-3}n^2\sigma^4 + [\text{Bias}(r_c)]^2$$

leading to Proposition (6.5) for $\Delta = 0$.

A9.

1. With (6.1) and (6.4) we derive from (A17), $\Delta \geq 0$ and from the MSE-formula for $\hat{r}$

$$\begin{aligned}\text{MSE}(r_c) - \text{MSE}(\hat{r}) &\\ &\geq n^{-2}\sigma^4[2(n-p)^{-3}n^4 + (n-p)^{-2}p^4 - 2n - 6p - 8(n-m)^{-1}p^2] \\ &= n^{-2}(n-p)^{-2}\sigma^4[12np^2 - 6p^3 + 2(n-p)^{-1}np^3 \\ &\qquad + p^4 - 8(n-m)^{-1}p^2(n-p)^2].\end{aligned}$$

Clearly, for $h \geq 3$

$$n - m \geq 2n/3$$

and therefore

$$8(n-m)^{-1}p^2(n-p)^2 \leq 12n^{-1}p^2(n-p)^2 \leq 12p^2(n-p).$$

Using this, the MSE difference can be estimated from below by

$$\text{MSE}(r_c) - \text{MSE}(\hat{r}) \geq n^{-2}(n-p)^{-2}\sigma^4[6p^3 + 2(n-p)^{-1}np^3 + p^4] > 0.$$

2. For $m = p$ and thus $\lambda = \Delta = 0$ we see from (A7), (A17) and Table 2:

$$\begin{aligned}\text{MSE}(r_c) - \text{MSE}(r_c^h) &\\ &= 2(n-m)^{-1}(h-1)^{-2}h^2\sigma^4 + [(n-m)^{-1}n^{-1}m^2\sigma^2]^2 \\ &\quad - \sigma^4[2n^{-1} + 2n^{-1}(h-1)^{-3}(3h^2 - 3h + 1) + (h-1)^{-2}h^{-2}] = 0.\end{aligned}$$

A10. With (6.4), the bias formulae (see (A7), (A16) and Table 1) yield

$$\begin{aligned}\text{Bias}(r_c) - \text{Bias}(r_c^h) &= (n-p)^{-2}(2n-p)p\Delta \\ &\quad - \sigma^2(n-p)^{-1}(h-1)^{-1}n^{-1}(n-hp)p,\end{aligned}$$

which is obviously nonnegative iff (6.6) holds.

Acknowledgement. The authors are very grateful for valuable remarks of the referees, which led to an improved presentation of the results.

REFERENCES

ANDERSON, T. W. (1971). *The Statistical Analysis of Time Series.* Wiley, New York.

BELSLEY, D. A., KUH, E. and WELSCH, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity.* Wiley, New York.

BENDEL, R. B. and AFIFI, A. A. (1977). Comparison of stopping rules in forward "stepwise" regression. *J. Amer. Statist. Assoc.* **72** 46–53.

BUNKE, H. (1973). Approximation of regression functions. *Math. Oper. Statist.* **4** 315–325.

BUNKE, H. and BUNKE, O. (Eds.) (1984). *Statistical Inference in Linear Models.* Wiley, London (to appear).

BUNKE, O. (1984). Selecting variables and models in regression analysis: some tools and suggestions for a strategy. Preprint No. 68, Sektion Mathematik, Humboldt-Universität, Berlin.

DROGE, B. (1982). Criteria for the selection of regression models. Doctoral thesis, Humboldt-University, Berlin (in German).

EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26.

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Planes.* SIAM, Philadelphia.

EFRON, B. (1983). Estimating the error rate of a prediction rule: improvement on cross-validation. *J. Amer. Statist. Assoc.* **78** 316–331.

FOX, T., HINKLEY, D. V. and LARNTZ, K. (1980). Jackknifing in nonlinear regression. *Technometrics* **C22** 29–33.

HINKLEY, D. V. (1978). Improving the jackknife with special reference to correlation estimation. *Biometrika* **65** 13–21.

HOCKING, R. R. (1976). The analysis and selection of variables in linear regression. *Biometrics* **32** 1–49.

MALLOWS, C. L. (1973). Some comments on C_p. *Technometrics* **15** 661–675.

MONTGOMERY, D. C. and PECK, E. A. (1982). *Introduction to Linear Regression Models.* Wiley, New York.

STONE, M. (1974). Cross-validatory choice and assessment of statistical predictions. *J. Roy. Statist. Soc. B.* **36** 111–147.

THOMPSON, M. L. (1978). Selection of variables in multiple regression: Part I. A review and evaluation. *Internat. Statist. Rev.* **46** 1–19. Part II. Chosen procedures, computations and examples. Ibid. **46** 129–146.

WEISBERG, S. (1980). *Applied Linear Regression.* Wiley, New York.

DEPARTMENT OF STATISTICS
SEKTION MATHEMATIK
HUMBOLDT UNIVERSITÄT ZU BERLIN
1086 BERLIN
PSF 1297
GERMAN DEMOCRATIC REPUBLIC