A Cross-Validatory Method for Dependent Data

Article *in* Biometrika · June 1994
DOI: 10.1093/biomet/81.2.351

3 authors, including:

Prabir Burman
University of California, Davis
67 PUBLICATIONS **1,131** CITATIONS
SEE PROFILE

Deborah Nolan
University of California, Berkeley
70 PUBLICATIONS **1,112** CITATIONS
SEE PROFILE

Biometrika (1994), **81**, 2, *pp.* 351–8
Printed in Great Britain

A cross-validatory method for dependent data

BY PRABIR BURMAN
Division of Statistics, University of California, Davis, California 95616, U.S.A

EDMOND CHOW AND DEBORAH NOLAN
Department of Statistics, University of California, Berkeley, California 94720, U.S.A.

SUMMARY

In this paper we extend the technique of cross-validation to the case where observations form a general stationary sequence. We call it h-block cross-validation, because the idea is to reduce the training set by removing the h observations preceding and following the observation in the test set. We propose taking h to be a fixed fraction of the sample size, and we add a term to our h-block cross-validated estimate to compensate for the underuse of the sample. The advantages of the proposed modification over the cross-validation technique are demonstrated via simulation.

Some key words: Cross-validation; Dependence; Integrated square error; Prediction error.

1. INTRODUCTION

Cross-validation (Stone, 1974; Geisser, 1975) is acclaimed as a method for estimating prediction error in regression and classification problems, and in recent years it has received much attention for its model selection ability in the nonparametric setting. See for example Härdle & Marron (1985). Even more recently, the technique of cross-validation has been applied to prediction error estimation in the dependent setting. Work in this area includes that of Györfi et al. (1989), C. K. Chu, in a University of North Carolina Ph.D. thesis, and Burman & Nolan (1992). Here we continue this line of development, and present a modification of the traditional leave-one-out method of cross-validation for use with dependent observations. Our approach is suggested by the well-known technique of removing blocks or subseries of observations, which originates from estimation problems for dependent random variables.

To best illustrate our procedure, suppose the goal is to fit the $(k+1)$-parameter model

$$x_i = \theta_0 + \theta_1 x_{i-1} + \ldots + \theta_k x_{i-k} \tag{1}$$

to N observations $X_1, \ldots, X_N$ from a stationary process. Let $(\hat{\theta}_0, \ldots, \hat{\theta}_k)$ be the least-squares estimate, where $(X_i, \ldots, X_{i+k})$ for $i = 1, \ldots, N-k$ are the cases on which the calculation is based. Following Akaike (1970), we assess the predictive ability of the fitted model by the expectation:

$$E\{\text{PE}(\hat{\theta}_0, \ldots, \hat{\theta}_k)\} = E\{(\tilde{X}_{k+1} - \hat{\theta}_0 - \hat{\theta}_1 X_k - \ldots - \hat{\theta}_k \tilde{X}_1)^2\}, \tag{2}$$

where $\tilde{X}_1, \ldots, \tilde{X}_N$ is another process that has the same distribution as $X_1, \ldots, X_N$ but is independent of it. Leave-one-out, or ordinary, cross-validation estimates the expected

value in (2) by

$$\text{OCV}_n = \frac{1}{n}\sum_{i=1}^{n}(X_{i+k} - \hat{\theta}_{0,i} - \hat{\theta}_{1,i}X_{i+k-1} - \ldots - \hat{\theta}_{k,i}X_i)^2, \tag{3}$$

where $\hat{\theta}_{j,i}$ $(j = 0, \ldots, k)$ denotes the ith least-squares estimate of θ_j, obtained after deleting the ith case $(X_i, \ldots, X_{i+k})$, and $n = N - k$ is the number of cases.

In classical applications of leave-one-out cross-validation, the cases are independent, therefore making the cross-validated prediction error a good approximation to the true expected prediction error. Here, however, the cases overlap, so leave-one-out cross-validation (3) may provide a very poor estimate of $E\{\text{PE}(\hat{\theta}_0, \ldots, \hat{\theta}_k)\}$. In this note we examine an alternative method of cross-validation, which we dub 'h-block cross-validation', that can handle general forms of dependence. The idea is a simple one. Rather than remove the single case $(X_i, \ldots, X_{i+k})$ when calculating the ith least-squares estimate, remove as well a block of h cases from either side of it. Now the training set contains $2h$ fewer cases, but the test set remains a singleton. Carlstein (1986), Künsch (1989) and Lele (1991) use a similar technique with jackknife variance estimates for stationary sequences, and Chu, in the thesis mentioned above, and Györfi et al. (1989) propose this modification to cross-validation when selecting the nuisance parameter in nonparametric curve estimation with dependent data. This approach is quite different from v-fold cross-validation. There, the n cases are divided into v sets roughly of size n/v. Each group of n/v cases constitutes a test set; the remaining $n - (n/v)$ cases comprise the corresponding training set. Here instead, there are n test sets, each consisting of one case, and the training sets contain roughly $n - 2h - 1$ cases. So, h-block cross-validation maintains a leave-one-out aspect.

According to the underlying structure of the data, blocking allows near independence between these two sets. It remains a question how to select the value of h. This is our main concern here. Intuitively, one should shrink the block size relative to the sample size, but to maintain independence between the test set and the training set, h should remain large. Chu, in the thesis mentioned above, and Györfi et al. (1989) require that h/n tend to 0, with the rate of decrease a complex function of the underlying structure of the data and the amount of smoothness in the model. In practice this structure is unknown and, for small samples, h will necessarily be large relative to n. Alternatively, we propose to take h as a fixed fraction of n, that is $h/n = p$ for some $0 < p < \frac{1}{2}$, and to correct for the underuse of the sample by adding a simple term to the h-block cross-validated estimate. The correction term makes possible our omnibus choice of h. It is analogous to the correction used for v-fold cross-validation (Burman, 1989). We find through simulation that, with the correction term proposed here, h-block cross-validation estimates the expected prediction error well in a wide range of settings, and we also find that, without this correction, h-block cross-validation may be as ineffective as ordinary leave-one-out cross-validation. We consider here only processes having a short range dependence. Although long-range dependence models have been proven useful in a variety of applications (Hosking, 1981; Cox, 1984; Dahlhaus, 1989), we do not know if our methodology is applicable.

The next section formally introduces h-block cross validation. Section 3 outlines a few examples where h-block cross-validation can be used. Finally in the last section, a variety of simulations are presented in support of our proposal.

2. THE TECHNIQUE

Let $Z_1, \ldots, Z_n$ be a segment of length n from a stationary process where each Z_i has distribution P on $\mathbb{R}^d$. For h a positive integer and for each i, define $P_{n,i}$, an empirical

estimate of P, as follows:

$$P_{n,i}(A) = \sum_{j=1}^{n} \omega_{i,j} I(Z_j \in A),$$

for A a Borel set in $\mathbb{R}^d$. Here the weights $\{\omega_{ij}: 1 \leqslant i, j \leqslant n\}$ form a double array of non-negative numbers such that

$$\omega_{i,j} = 0 \quad \text{if } |i-j| \leqslant h, \tag{4}$$

$$\sum_{i=1}^{n} \omega_{i,j} = 1 \quad \text{for all } j. \tag{5}$$

Many choices for the weight function satisfy these two constraints. In our simulation study we simply take:

for $1 \leqslant j \leqslant h$,

$$\omega_{i,j} = \begin{cases} 0 & (1 \leqslant i \leqslant j+h), \\ 1/(n-j-h) & \text{otherwise;} \end{cases}$$

for $h < j \leqslant n-h$

$$\omega_{i,j} = \begin{cases} 0 & (j-h \leqslant i \leqslant j+h), \\ 1/(n-2h-1) & \text{otherwise;} \end{cases}$$

for $n-h < j \leqslant n$,

$$\omega_{i,j} = \begin{cases} 0 & (j-h \leqslant i \leqslant n), \\ 1/(j-h-1) & \text{otherwise.} \end{cases}$$

The intent of condition (4) is to make the training set and the test set nearly independent. Our notation suppresses the dependence of h on n, of $P_{n,i}$ on h, and of $\omega_{i,j}$ on h and n. The empirical estimates $P_{n,i}$ need not be probability measures, but condition (5) ensures $\sum P_{n,i} = nP_n$, where P_n denotes the standard empirical distribution that places mass $1/n$ on each of the observations.

For some functional T on $\mathbb{R}^d \times \mathscr{P}$, where $\mathscr{P}$ is a collection of probability measures on $\mathbb{R}^d$, define the prediction error by

$$\text{PE}_n = \int T(z, P_n)\, dP(z), \tag{6}$$

and the h-block cross-validated estimate of $E(\text{PE}_n)$ to be

$$\text{CV}_n = \frac{1}{n} \sum_{i=1}^{n} T(Z_i, P_{n,i}). \tag{7}$$

Examples of functionals T are found in the next section. Also define the corrected h-block cross-validated estimate as

$$\text{CCV}_n = \text{CV}_n - \frac{1}{n} \sum_{i=1}^{n} \int T(z, P_{n,i})\, dP_n(z) + \int T(z, P_n)\, dP_n(z). \tag{8}$$

Heuristically, the extra terms in (8) follow from matching the expectation of CV_n with

that of PE_n. To see this write CV_n as:

$$\left\{\frac{1}{n}\sum_{i=1}^{n} T(Z_i, P_{n,i}) - \frac{1}{n}\sum_{i=1}^{n}\int T(z, P_{n,i})\, dP(z)\right\} + \frac{1}{n}\sum_{i=1}^{n}\int \{T(z, P_{n,i}) - T(z, P_n)\}\, dP(z) + \text{PE}_n. \tag{9}$$

If the $\{Z_i\}$ are independent then the expectation of the first term is zero. If not, this expectation is small when Z_i and $\{Z_j : |i-j| > h\}$ are nearly independent. The correction in (8) is an approximation to the second term in (9). It is needed because this term's expectation is of order $h/\{n(n-2h)\}$. If $h = np$, for $0 < p < \frac{1}{2}$, then the order of the expectation becomes $1/n$, whereas the order of the expectation of $\text{CCV}_n - \text{PE}_n$ is $1/n^2$. Note that the order terms include a constant that increases with dimension of the fitted model implicit in T.

Remark 1. Burman & Nolan (1992) consider the problem of estimating the prediction function μ for a stationary process (X_i, Y_i), where

$$Y_i = \mu(X_i) + \varepsilon_i, \quad E(\varepsilon_i | X_i) = 0.$$

It is argued there that leave-one-out cross-validation for model selection is asymptotically optimal in the sense of Shibata (1980) in many cases. It is important to point out that the test and training set need not be independent for cross-validation to work. Briefly, the prediction error can be well approximated by a quadratic form in the errors plus a bias term, provided $E(\varepsilon_i \varepsilon_j | X_1, \ldots, X_j) = 0$ $(i < j)$. A stationary Markov process with $Y_i = X_{i+1}$ satisfies this condition, which implies the AR(1) process also meets this condition. See Burman & Nolan (1992) for details of the argument.

Remark 2. Conditions for asymptotic optimality for this proposal remain open. It is noted however that in the independent case Burman (1990) has shown the closely related corrected v-fold cross-validation is asymptotically optimal for $0 < p < \frac{1}{2}$.

3. EXAMPLES

In this section we present three examples where h-block cross-validation may prove useful in estimating the prediction error. The first is a formalization of the example used in § 1 to introduce the notion of h-block cross-validation, i.e. the autoregressive model. The second example concerns additive prediction and the last is a nonparametric model.

Example 1. Let $X_1, \ldots, X_N$ be N observations from a stationary process. Use least squares to fit an autoregressive model of order k to the data as in (1). Call the fitted parameters $\hat{\theta}_0, \ldots, \hat{\theta}_k$. Then the expected quadratic loss (2) in predicting a new observation is

$$E(\text{PE}_n) = E\left\{\int (x_{k+1} - \hat{\theta}_0 - \hat{\theta}_1 x_k - \ldots - \hat{\theta}_k x_1)^2\, dP(x_1, \ldots, x_{k+1})\right\}. \tag{10}$$

Here

$$Z_i = (X_i, \ldots, X_{i+k}), \quad T(z, P_n) = (x_{k+1} - \hat{\theta}_0 - \hat{\theta}_1 x_k - \ldots - \hat{\theta}_k x_1)^2,$$

with $z = (x_1, \ldots, x_{k+1})$ and $n = N - k$. This is the model that is fitted in the simulations of the next section.

To explain further, take $k = 1$. Then the least squares estimates $\hat{\theta}_0, \hat{\theta}_1$ in (10) are the

minimizers of

$$\sum_{i=1}^{n} (X_{i+1} - \theta_0 - \theta_1 X_i)^2,$$

P is the joint distribution of (X_1, X_2), and P_n puts mass $1/n$, or $1/(N-1)$, on each of the pairs (X_i, X_{i+1}) for $i = 1, \ldots, N-1$. The leave-one-out cross-validated estimate (3) in this case is

$$\text{OCV}_n = \frac{1}{n} \sum_{i=1}^{n} (X_{i+1} - \hat{\theta}_{0,i} - \hat{\theta}_{1,i} X_i)^2,$$

where $\hat{\theta}_{0,i}$ and $\hat{\theta}_{1,i}$ minimize

$$\textstyle\sum_{j \neq i} (X_{j+1} - \theta_0 - \theta_1 X_j)^2.$$

To h-block cross-validate, minimize for each i the following quadratic:

$$\sum_{j=1}^{n} (X_{j+1} - \theta_0 - \theta_1 X_j)^2 \omega_{i,j},$$

where the weights $\omega_{i,j}$ satisfy (4) and (5). Call the minimizers $\hat{\theta}_{0,i,w}$ and $\hat{\theta}_{1,i,w}$. Finally, the corrected h-block cross-validated estimate of (10) is

$$\text{CCV}_n = \frac{1}{n} \sum_{i=1}^{n} (X_{i+1} - \hat{\theta}_{0,i,w} - \hat{\theta}_{1,i,w} X_i)^2 - \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_{j+1} - \hat{\theta}_{0,i,w} - \hat{\theta}_{1,i,w} X_j)^2$$
$$+ \frac{1}{n} \sum_{i=1}^{n} (X_{i+1} - \hat{\theta}_0 - \hat{\theta}_1 X_i)^2.$$

Example 2. As in the previous example, take $X_1, \ldots, X_N$ to be N observations from a stationary process. Generalize the ideas of Hastie & Tibshirani (1987) and Stone (1985) by considering the problem of predicting X_i by an additive nonparametric model: $\mu_0 + \mu_1(X_{i-1}) + \ldots + \mu_k(X_{i-k})$. If we employ least-squares to fit splines or polynomials then, as in the above example, the prediction error (6) is given by

$$\text{PE}_n = \int \{x_{k+1} - \hat{\mu}_0 - \hat{\mu}_1(x_k) - \ldots - \hat{\mu}_k(x_1)\}^2 \, dP(x_1, \ldots, x_{k+1}).$$

Again, P represents the distribution of $Z_i = (X_i, \ldots, X_{i+k})$, the function T is the quadratic integrand above, $z = (x_1, \ldots, x_{k+1})$ and $n = N - k$.

Example 3. Consider the nonparametric regression model

$$E(Y \mid X = x) = \mu(x).$$

Take $\{Z_i = (X_i, Y_i) : i = 1, \ldots, N\}$ to be N observations from a strictly stationary process. Once again if a polynomial of order k, or a spline with k knots, is fitted to the data by the method of least squares then the prediction error using quadratic loss is

$$\text{PE}_n = \int \{y - \hat{\mu}_k(x)\}^2 \, dP(x, y),$$

where $\hat{\mu}_k$ is the estimate of μ. In this example $n = N$.

In each of these examples, quadratic loss was used both to estimate the unknown

prediction function and to evaluate the prediction error. Our method is not restricted to the use of quadratic loss; in fact, the functional T can be any reasonably smooth function of z and P. We do not investigate the performance of other loss functions here.

4. SIMULATION STUDY

Six simulations are presented here. The simulations demonstrate a variety in autocorrelation, sample size, and fitted model. Each simulation is based on 10 000 replications. For one replication, N observations are generated from a stationary zero-mean Gaussian sequence with standard deviation 3 and specified autocorrelation. A model is fitted to the generated data using least squares; the exact one-step prediction error is calculated based on the fitted parameters and pre-specified correlation structure; and, finally, the expected prediction error is estimated according to leave-one-out cross-validation, that is $h = 0$, and according to corrected (8) and uncorrected (7) h-block cross-validation for h as the nearest integer to each of the following fractions of n: 0·05, 0·10, 0·15, 0·20, 0·25, 0·30, where n is the effective number of cases. Example 1 in § 3 provides details for computing these quantities. Table 1 reports estimates of $E(\text{PE}_n)$, $E(\text{CV}_n)$, $E(\text{CCV}_n)$ based on the 10 000 repetitions.

The simulations show that classical leave-one-out cross-validation can be very misleading, and caution should be exercised in using the leave-one-out technique when observations are not independent. On the other hand, these simulations also demonstrate that blocking effectively adapts cross-validation to the dependent setting. In each of the simulations, at least one block size produces good results for uncorrected h-block cross-validation. However, as discussed earlier, the best block size to use is determined by the autocorrelation and the appropriateness of the fitted model, both of which are presumed unknown. Therefore it is reassuring to see positive results for corrected h-block cross-validation over a wide range of block sizes. Whether $\frac{1}{5}$ or $\frac{1}{2}$ of the cases are removed, blocking with the corrective term yields good cross-validated estimates of the expected prediction error. Our simulations suggest the rule-of-thumb of removing $\frac{1}{3}$ of the data; that is the fourth column in each simulation of Table 1 shows that $h = n/6$ appears to be a sensible choice in a variety of settings. Finally, it is noted that it can be difficult for the corrective term to compensate for large data loss, where h exceeds $0{\cdot}25n$, because it makes only a first order correction.

Specifically, the first simulation fits the linear model

$$x_i = \theta_0 + \theta_1 x_{i-1}.$$

The sample is of size $N = 25$ with autocorrelations $C(X_i, X_{i+j}) = 0{\cdot}3, 0{\cdot}4, 0{\cdot}3, 0{\cdot}3, 0{\cdot}3, 0{\cdot}4^6, 0{\cdot}4^7, 0{\cdot}4^8, \ldots, 0{\cdot}4^{24}$, for $j = 1, \ldots, 24$. The results appear at the top of Table 1. Notice ordinary leave-one-out cross-validation greatly underestimates the expected prediction error in this example.

The next two simulation results presented in Table 1 are based on the same stationary process, an autoregressive model with single coefficient 0·7. Burman & Nolan (1992) show that ordinary cross-validation is asymptotically equivalent to PE_n for the AR(1) process. The first of these two simulations fits a model that is linear in the first lag, and the second fits a model that is quadratic in the first lag:

$$x_i = \theta_0 + \theta_1 x_{i-1} + \theta_2 x_{i-1}^2.$$

The fourth and fifth simulations generate observations from a sequence of Gaussians

Table 1. *Mean and standard deviations of h-block and corrected h-block estimates of the prediction error for six simulations*

	$N = 25$, $n = 24$, $E(PE_n) = 10{\cdot}38$, $SD(PE_n) = 2{\cdot}74$						
	$h = 0$	$h = 1$	$h = 2$	$h = 4$	$h = 5$	$h = 6$	$h = 7$
$E(CV_n)$	7·96	8·22	8·89	10·58	11·82	12·32	13·01
$E(CCV_n)$	7·93	8·12	8·64	9·81	10·68	10·73	10·83
$SD(CV_n)$	2·83	3·00	3·52	4·85	5·92	6·75	7·86
$SD(CCV_n)$	2·82	2·96	3·39	4·32	5·05	5·43	5·96
	$N = 36$, $n = 35$, $E(PE_n) = 5{\cdot}09$, $SD(PE_n) = 0{\cdot}75$						
	$h = 0$	$h = 2$	$h = 4$	$h = 5$	$h = 7$	$h = 9$	$h = 11$
$E(CV_n)$	4·84	5·03	5·20	5·30	5·52	5·84	6·32
$E(CCV_n)$	4·83	4·97	5·07	5·12	5·20	5·30	5·42
$SD(CV_n)$	1·19	1·28	1·43	1·52	1·79	2·19	2·82
$SD(CCV_n)$	1·19	1·26	1·36	1·42	1·57	1·78	2·09
	$N = 36$, $n = 35$, $E(PE_n) = 5{\cdot}50$, $SD(PE_n) = 0{\cdot}75$						
	$h = 0$	$h = 2$	$h = 4$	$h = 5$	$h = 7$	$h = 9$	$h = 11$
$E(CV_n)$	5·12	5·43	5·82	6·04	6·69	7·79	10·34
$E(CCV_n)$	5·11	5·32	5·55	5·65	5·95	6·36	7·25
$SD(CV_n)$	1·37	1·68	2·34	2·73	4·42	7·72	18·24
$SD(CCV_n)$	1·36	1·61	2·10	2·36	3·47	5·68	14·26
	$N = 25$, $n = 24$, $E(PE_n) = 11{\cdot}59$, $SD(PE_n) = 2{\cdot}94$						
	$h = 0$	$h = 1$	$h = 2$	$h = 4$	$h = 5$	$h = 6$	$h = 7$
$E(CV_n)$	8·01	8·62	10·08	12·17	12·70	13·32	13·92
$E(CCV_n)$	7·99	8·48	9·74	11·21	11·34	11·46	11·47
$SD(CV_n)$	3·31	3·72	4·85	6·73	7·42	8·40	9·25
$SD(CCV_n)$	3·30	3·65	4·63	6·00	6·31	6·82	7·14
	$N = 64$, $n = 62$, $E(PE_n) = 6{\cdot}18$, $SD(PE_n) = 0{\cdot}56$						
	$h = 0$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 16$	$h = 19$
$E(CV_n)$	6·06	6·21	6·30	6·40	6·51	6·76	7·07
$E(CCV_n)$	6·06	6·16	6·20	6·23	6·25	6·31	6·38
$SD(CV_n)$	1·23	1·29	1·37	1·47	1·58	1·81	2·08
$SD(CCV_n)$	1·23	1·28	1·34	1·40	1·44	1·54	1·66
	$N = 60$, $n = 58$, $E(PE_n) = 10{\cdot}17$, $SD(PE_n) = 1{\cdot}78$						
	$h = 0$	$h = 3$	$h = 6$	$h = 9$	$h = 12$	$h = 15$	$h = 17$
$E(CV_n)$	8·19	8·54	9·64	10·36	11·03	11·53	12·24
$E(CCV_n)$	8·18	8·34	9·34	9·79	10·10	10·09	10·34
$SD(CV_n)$	2·18	2·29	2·99	3·62	4·34	4·92	5·73
$SD(CCV_n)$	2·18	2·25	2·85	3·30	3·74	3·93	4·39

with autocorrelations: 0, 0·6, 0, 0·4, 0, $0{\cdot}4^6$, 0, $0{\cdot}4^8$, ... In the fourth simulation a linear fit in one lag is made to the data, for $N = 25$. The results show a large expected prediction error, and the h-block method does a very good job estimating it. In the fifth simulation, $N = 64$ and the fitted model is linear in two lags:

$$x_i = \theta_0 + \theta_1 x_{i-1} + \theta_2 x_{i-2}.$$

Here, both the h-block method and leave-one-out cross-validation perform well, which is not surprising given the results of Burman & Nolan (1992).

Finally, the last simulation further exemplifies the advantages of h-blocking. There the autocorrelations are: 0·2, 0·2, 0·2, 0·6, 0·1, 0·1, 0, $0{\cdot}6^2$, 0, 0, 0, $0{\cdot}6^3$, 0, 0, 0, $0{\cdot}6^4$, . . . ; and the fitted model is

$$x_i = \theta_0 + \theta_1 x_{i-1} + \theta_2 x_{i-1}^2 + \theta_3 x_{i-2}.$$

ACKNOWLEDGEMENT

The authors thank the referees for helpful comments. This research was partially supported by the United States National Science Foundation.

REFERENCES

AKAIKE, H. (1970). Statistical predictor identification. *Ann. Inst. Statist. Math.* **22**, 203–17.

BURMAN, P. (1989). A comparative study of ordinary cross-validation, v-fold cross-validation and repeated learning-testing methods. *Biometrika* **76**, 503–14.

BURMAN, P. (1990). Estimation of optimal transformations using v-fold cross-validation and repeating learning-testing methods. *Sankhyā* A **76**, 503–14.

BURMAN, P. & NOLAN, D. (1992). Data dependent estimation of prediction functions. *J. Time Ser. Anal.* **13**, 189–207.

CARLSTEIN, E. (1986). The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann. Statist.* **14**, 1171–9.

COX, D. R. (1984). Long-range dependence: A review. In *Statistics: An Appraisal*, Ed. H. A. David and H. T. David, pp. 55–74. Ames: Iowa State University.

DAHLHAUS, R. (1989). Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749–66.

GEISSER, S. (1975). A predictive sample reuse method with applications. *J. Am. Statist. Assoc.* **70**, 320–8.

GYÖRFI, L., HÄRDLE, W., SARDA, P. & VIEU, P. (1989). *Nonparametric Curve Estimation from Time Series.* New York: Springer-Verlag.

HÄRDLE, W. & MARRON, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **13**, 1465–81.

HASTIE, T. & TIBSHIRANI, R. (1987). A comparative study of ordinary cross-validation, v-fold cross-validation and repeated learning-testing methods. *Biometrika* **76**, 503–14.

HOSKING, J. R. M. (1981). Fractional differencing. *Biometrika* **68**, 165–76.

KÜNSCH, H. R. (1989). The jackknife and the bootstrap for general stationary observations. *Ann. Statist.* **17**, 1217–41.

LELE, S. (1991). Jackknifing linear estimating equations: Asymptotic theory and applications in stochastic processes. *J. R. Statist. Soc.* B **53**, 253–67.

SHIBATA, R. (1980). Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Ann. Statist.* **13**, 230–40.

STONE, C. (1985). Additive regression and other nonparametric models. *Ann. Statist.* **13**, 689–705.

STONE, M. (1974). Cross-validatory choice and the assessment of statistical predictions (with discussion). *J. R. Statist. Soc.* B **36**, 111–33.

[*Received April* 1991. *Revised October* 1993]