

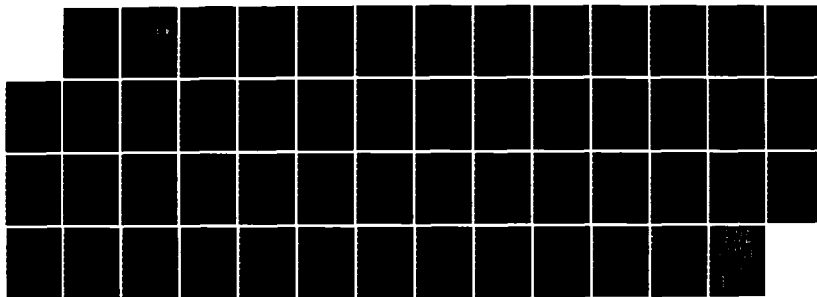
AD-A166 231

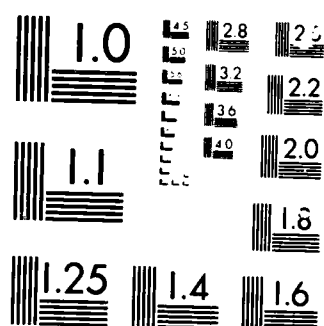
PERFORMANCE OF DISCRETE-TIME PREDICTORS OF
CONTINUOUS-TIME STATIONARY PRO. (U) NORTH CAROLINA UNIV 1/1
AT CHAPEL HILL CENTER FOR STOCHASTIC PROC..
S CAMBANIS ET AL. DEC 85 TR-123

UNCLASSIFIED

F/G 12/1

NL



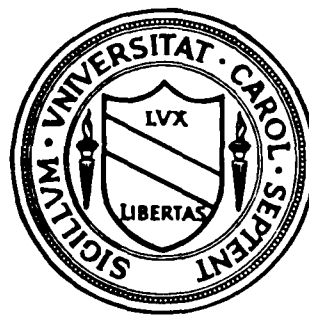


MICROCOPY RESOLUTION TEST CHART

AD-A166 231

CENTER FOR STOCHASTIC PROCESSES

Department of Statistics
University of North Carolina
Chapel Hill, North Carolina



DTIC
SELECTE
APR 01 1986
S D D

PERFORMANCE OF DISCRETE-TIME PREDICTORS
OF CONTINUOUS-TIME STATIONARY PROCESSES

by

S. Cambanis

and

E. Masry

Approved for public release;
distribution unlimited.

Technical Report No. 123

December 1985

DTIC FILE COPY

PERFORMANCE OF DISCRETE-TIME PREDICTORS
OF CONTINUOUS-TIME STATIONARY PROCESSES

by

Stamatis Cambanis
Department of Statistics
University of North Carolina
Chapel Hill, NC 27514

and

Elias Masry
Department of Electrical Engineering and Computer Science
University of California, San Diego
La Jolla, CA 92093

Abstract

We study the asymptotic performance of linear predictors of continuous-time stationary processes from observations at n sampling instants on a fixed observation interval. We consider both optimal and simpler choices of predictor coefficients; uniform sampling, as well as nonuniform sampling tailored to the statistics of the process under prediction. We concentrate on stationary processes with rational spectral densities and obtain the asymptotic performance for cases with no and with one quadratic-mean derivative. The analytical results are supplemented by numerical examples depicting small and large sample size performance.

The work of S. Cambanis was supported by the Air Force Office of Scientific Research under Grant No. F49620 85 C 0144.

The work of E. Masry was supported by the Office of Naval Research under Contract N00014 84 K 0042.

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
NOTICE OF TRANSMITTAL TO DTIC
This technical report has been reviewed and is
approved for public release IAW AFR 19D-12.
Distribution is unlimited.
MATTHEW J. KEEFER
Chief, Technical Information Division

UNCLASSIFIED

ADA 166231

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED			1b. RESTRICTIVE MARKINGS		
2a. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; Unlimited distribution unlimited.		
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR. 86-0018		
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Technical Report No. 123			7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research		
6a. NAME OF PERFORMING ORGANIZATION Center for Stochastic Processes		6b. OFFICE SYMBOL (If applicable) N/M		7b. ADDRESS (City, State and ZIP Code) Bolling Air Force Base Washington, DC 20332	
6c. ADDRESS (City, State and ZIP Code) Statistics Dept., Univ. of North Carolina Phillips Hall 039-A Chapel Hill, NC 27514		9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER F49620-85-C-0144			
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR		8b. OFFICE SYMBOL (If applicable)		10. SOURCE OF FUNDING NOS.	
8c. ADDRESS (City, State and ZIP Code) Bolling Air Force Base Washington, DC 20332		PROGRAM ELEMENT NO. G1102F		PROJECT NO. 2304	TASK NO. A5
11. TITLE (Include Security Classification) "Performance of discrete-time predictors of continuous-time stationary processes"					
12. PERSONAL AUTHOR(S) S. Cambanis and E. Masry					
13a. TYPE OF REPORT technical		13b. TIME COVERED FROM 9/85 TO 8/86		14. DATE OF REPORT (Yr., Mo., Day) technical Dec. 85	
15. PAGE COUNT 49					
16. SUPPLEMENTARY NOTATION					
17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)		
FIELD	GROUP	SUB. GR.			
19. ABSTRACT (Continue on reverse if necessary and identify by block number) We study the asymptotic performance of linear predictors of continuous-time stationary processes from observations at n sampling instants on a fixed observation interval. We consider both optimal and simpler choices of predictor coefficients; uniform sampling, as well as nonuniform sampling tailored to the statistics of the process under prediction. We concentrate on stationary processes with rational spectral densities and obtain the asymptotic performance for cases with no and with one quadratic-mean derivative. The analytical results are supplemented by numerical examples depicting small and large sample size performance.					
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐			21. ABSTRACT SECURITY CLASSIFICATION Unclassified		
22a. NAME OF RESPONSIBLE INDIVIDUAL P. 887			22b. TELEPHONE NUMBER (Include Area Code) 919-962-2507		22c. OFFICE SYMBOL

I. INTRODUCTION

A continuous-time stationary process is to be predicted linearly from observations at a finite number of sampling instants from a fixed observation interval. What is the best location of the sampling points, for fixed sample size or asymptotically as the sample size tends to infinity? At how many points should the process be sampled for the predictor to achieve a specified mean square prediction error? If the process is sampled uniformly, what should the sampling rate be for a desirable predictor performance, measured in mean square error? These questions are answered by studying the performance of discrete-time linear predictors as the sample size tends to infinity.

We consider uniform sampling as well as nonuniform sampling, e.g. at fixed quantiles of a probability density over the observation interval. We are strongly interested in the performance of uniform sampling, but we would also like to know whether appropriately chosen nonuniform sampling may result in appreciable improvement of performance. We consider optimal and certain suboptimal linear predictors. The optimal predictors require the inversion of an $n \times n$ matrix for each sample size n . The suboptimal predictors we use, require the solution of an integral equation, but then the choice of predictor coefficients for each sample size is very simple.

We concentrate on stationary processes with rational spectral density. When the process has no quadratic-mean derivative, both the optimal and suboptimal predictors have the same rate n^{-2} and the same asymptotic constant, which depends on the sampling design. This allows us to check whether the asymptotically optimal sampling design is uniform or nonuniform, and, in the latter case, to compare its performance with that of uniform sampling. The small sample performance of uniform versus nonuniform sampling, and of optimal versus suboptimal predictors is illustrated by specific numerical examples.

Availability Codes	
Dist	Avail and/or Special
A-1	

We next consider the case where the process has exactly one quadratic-mean derivative. When the suboptimal predictor is used, uniform sampling has rate n^{-1} while nonuniform sampling schemes can be designed with rate n^{-4} . This striking discrepancy in the rate of convergence is due to the presence of derivatives of delta functions in the continuous-time linear predictor filter. The asymptotic performance of the optimal predictor is an open question at present. In this case, the comparison in small sample performance and in asymptotics of optimal versus suboptimal predictors is done via specific numerical examples.

In the context of regression problems, sampling designs were considered by Sacks and Ylvisaker [3]-[5]. Sampling designs for suboptimal integral estimators and detectors were considered by Schoenfelder [6] and by Cambanis and Masry [1], respectively. The current paper considers sampling designs in the context of prediction of stationary processes and provides new analytical results, extending the earlier works [1],[6], for processes with one quadratic-mean derivative.

The organization of the paper is as follows. The formulation of the problem is given in Section II. Theoretical and numerical results for processes having no quadratic-mean derivative are presented in subsection A of Section II. The corresponding results for processes with one quadratic-mean derivative are given in subsection B of Section II. The proof of the asymptotic performance for processes with one quadratic-mean derivative is delegated to the Appendix.

II. PERFORMANCE OF DISCRETE-TIME PREDICTORS

We consider a stationary process $X = \{X(t), -\infty < t < \infty\}$ with mean zero and covariance function $R(t)$. We want to predict linearly the value of the process at time $s > T$ from n observations of the process taken at the sampling instants $D = \{t_k\}_{k=1}^n$ from the interval $I = [-T, T]$, namely from the observations $\{X(t_k)\}_{k=1}^n$. The linear predictor $X_D(s)$ has the form

$$X_D(s) = \sum_{k=1}^n c_{D,k} X(t_k) = \underline{c}_D' \underline{X}_D \quad (2.1)$$

where the row vectors $\underline{c}_D'$ and $\underline{X}_D'$ are defined by $\underline{c}_D' = (c_{D,1}, \dots, c_{D,n})$ and $\underline{X}_D' = (X(t_1), \dots, X(t_n))$. Our goal is to choose the n sampling points D and the predictor coefficients $\underline{c}_D$ so that the resulting mean-square prediction error

$$e_D^2(s) = E[X(s) - X_D(s)]^2 = R(0) - 2\underline{c}_D' \underline{R}_D(s) + \underline{c}_D' \underline{R}_D \underline{c}_D$$

should be as small as possible for fixed sample size n or asymptotically as n tends to infinity. Here $\underline{R}_D(s)$ is the row vector $(R(s-t_1), \dots, R(s-t_n))$ and $\underline{R}_D$ is the covariance matrix $[R(t_k - t_j)]_{j,k=1}^n$.

For a specified set of n sampling points D , a natural choice of predictor coefficients $\underline{c}_D$ is the optimal coefficients $\underline{c}_D' = \underline{R}_D'(s) \underline{R}_D^{-1}$, for which the linear predictor

$$\hat{X}_D(s) = \underline{R}_D'(s) \underline{R}_D^{-1} \underline{X}_D \quad (2.2)$$

is the projection of $X(s)$ on the data space generated by $\{X(\tau), \tau \in D\}$, and the corresponding minimum mean-square prediction error $e_D^2(s)$ is given by

$$\epsilon_D^2(s) = E[X^2(s)] - E[\hat{X}_D^2(s)] = R(0) - \tilde{R}_D'(s) \tilde{R}_D^{-1} \tilde{R}_D(s). \quad (2.3)$$

For a specified n point sampling design D , finding the optimal coefficients involves inverting an $n \times n$ covariance matrix, and the weight $c_{D,k}$ for the k^{th} observation $X(t_k)$ depends on *all* sampling points $\{t_k\}_{k=1}^n = D$. Simpler, non-optimal, weights are naturally suggested by the form of the continuous data predictor whenever it is known, and we discuss this next.

When the entire continuous record $\{X(\tau), \tau \in I = [-T, T]\}$ is available, we denote the optimal linear mean-square error predictor of $X(s)$ by $\hat{X}_I(s)$ and its minimum mean-square error by

$$\epsilon_I^2(s) = E[X(s) - \hat{X}_I(s)]^2 = R(0) - E[\hat{X}_I(s)X(s)]. \quad (2.4)$$

For any discrete-time predictor $X_D(s)$ (with or without optimal coefficients $\tilde{c}_D$) we have

$$\begin{aligned} e_D^2(s) &= E[X(s) - X_D(s)]^2 = E\{[X(s) - \hat{X}_I(s)] - [\hat{X}_I(s) - X_D(s)]\}^2 \\ &= \epsilon_I^2(s) + E[\hat{X}_I(s) - X_D(s)]^2 \end{aligned} \quad (2.5)$$

where the cross term vanishes since the error $X(s) - \hat{X}_I(s)$ is orthogonal to the space generated by the continuous data $\{X(\tau), \tau \in I\}$ to which both $\hat{X}_I(s)$ and $X_D(s)$ belong. Thus the excess error of any discrete-time predictor is given by

$$e_D^2(s) - \epsilon_I^2(s) = E[\hat{X}_I(s) - X_D(s)]^2 \quad (2.6)$$

and in particular the excess error of the optimal-coefficients discrete-time predictor (2.2) is

$$\epsilon_D^2(s) - \epsilon_I^2(s) = E[\hat{X}_I^2(s)] - E[\hat{X}_D^2(s)] \quad (2.7)$$

by the projection theorem.

Since the excess mean-square error (2.6) of every discrete-time predictor is measured by how well the discrete-time predictor approximates the continuous time predictor in mean-square, simple nonoptimal-coefficients discrete-time predictors can be obtained from the form of the continuous-time predictor whenever the latter is known explicitly in the time domain. In particular when the process X has a rational spectral density and has precisely k mean-square derivatives, as we assume henceforth, then the optimal linear continuous-time predictor has the form [2]

$$\hat{X}_I(s) = \sum_{j=0}^k \{a_j X^{(j)}(-T) + b_j X^{(j)}(T)\} + \int_{-T}^T c(\tau) X(\tau) d\tau \quad (2.8)$$

where the coefficients $\{a_j\}$, $\{b_j\}$ and the filter $c(t)$ all depend on the time s at which the process is predicted; and when we wish to emphasize this dependence we will write $a_j(s)$, $b_j(s)$, $c(t,s)$. The values of $\{a_j\}$, $\{b_j\}$ and $c(t)$ are obtained as the solutions of the linear integral equation

$$R(s-t) = \int_{-T}^T h(\tau) R(\tau-t) d\tau, \quad |t| \leq T, \quad (2.9)$$

with

$$h(t) = \sum_{j=0}^k (-1)^j \{a_j \delta^{(j)}(t+T) + b_j \delta^{(j)}(t-T)\} + c(t). \quad (2.10)$$

The form (2.8) for the optimal continuous-time predictor suggests generating nonoptimal-coefficients discrete-time predictors as follows. First the endpoints $\pm T$ should be included in the sampling design: $-T = t_1 < t_2 < \dots < t_{n-1} < t_n = T$. Then to obtain the discrete-data predictor $X_D(s)$ from the continuous-data predictor $\hat{X}_I(s)$, in the expression (2.8) of $\hat{X}_I(s)$:

(i) replace each quadratic mean derivative $X^{(j)}(\pm T)$, $j = 1, \dots, k$, by its natural approximation by means of the samples, e.g. replace $X^{(1)}(T)$ by $[X(T) - X(t_{n-1})]/(T - t_{n-1})$, and

(ii) replace the integral $\int_{-T}^T c(\tau)X(\tau)d\tau$ by a Riemann-type approximation in terms of its samples $\{c(t_k)X(t_k)\}_{k=1}^n$.

The appropriate quadrature rule for the approximation of the integral depends on the number k of quadratic-mean derivatives of the process and will be specified in the sequel.

Here, for simplicity, only the cases $k = 0$ and $k = 1$ will be discussed in detail as they bring out all the salient features of the problem at hand.

A. No Quadratic-mean Derivatives ($k = 0$)

In this case the optimal continuous data predictor has the form

$$\hat{X}_I(s) = a_0 X(-T) + b_0 X(T) + \int_{-T}^T c(\tau) X(\tau) d\tau \quad (2.11)$$

where the constants a, b and the filter $c(\tau)$ are the solution of (2.9) with $k = 0$.

A.1. Optimal-Coefficients Predictors

When the optimal-coefficients discrete-data predictor $\hat{X}_D(s)$ of (2.2) is used, the excess mean-square error (2.7) can be written in the form

$$\varepsilon_D^2(s) - \varepsilon_I^2(s) = E[\hat{X}_I(s) - \hat{X}_D(s)]^2 = \|f_s - P_D f_s\|^2,$$

where $f_s(t)$, $|t| \leq T$, is the function in the reproducing kernel Hilbert space of the covariance R , restricted to $[-T, T]^2$, which corresponds to the random variable $\hat{X}_I(s)$:

$$f_s(t) = E[\hat{X}_I(s) X(t)] = a_0 R(t+T) + b_0 R(t-T) + \int_{-T}^T c(\tau) R(t-\tau) d\tau,$$

and P_D denotes the projection on the space generated by $\{R(\cdot - t_k)\}_{k=1}^n$. Thus the results in Sacks and Ylvisaker [3, Theorem 3.1 and Remark 3.3] are applicable. Specifically if $\{D_n\}_{n=2}^\infty$ is a regular sequence of sampling designs generated by a continuous, positive density $p(t)$ on $[-T, T]$, i.e., $D_n = \{t_{n,k}\}_{k=1}^n$ with $t_{n,k}$ satisfying

$$\int_{-T}^{t_{n,k}} p(t) dt = \frac{k-1}{n-1}, \quad k = 1, \dots, n, \quad n \geq 2, \quad (2.12)$$

then the excess error (2.7) of the discrete-time predictor (2.2) satisfies

$$n^2 [\varepsilon_D^2(s) - \varepsilon_I^2(s)] \rightarrow \frac{r}{12} \int_{-T}^T \frac{c^2(t,s)}{p^2(t)} dt, \quad (2.13)$$

as the sample size $n \rightarrow \infty$, where $r = R'(0-) - R'(0+) > 0$. It is seen from (2.13) that the excess error decreases precisely at the rate $1/n^2$ and that the asymptotic constant depends on the correlation function $R(t)$ of the process X via the jump r of its derivative at zero as well as via the filter $c(t,s)$ - cf. the integral equation (2.9). The result of (2.13) is valid for any continuous and positive density $p(t)$ on $[-T, T]$. In particular for the uniform density $p(t) = \frac{1}{2T} 1_{[-T, T]}(t)$, the sampling instants are equally spaced:

$$t_{n,k} = T \left[\frac{2(k-1)}{n-1} - 1 \right], \quad k = 1, \dots, n,$$

and (2.13) provides a precise result on the magnitude of the excess error:

$$n^2 [\varepsilon_D^2(s) - \varepsilon_I^2(s)] \rightarrow \frac{rT^2}{3} \int_{-T}^T c^2(t,s) dt.$$

The case of equally-spaced sampling instants is not the best possible, however. If the sampling density $p(t)$ is chosen so as to minimize the asymptotic constant in (2.13), i.e. (by Hölder's inequality),

$$p_s(t) = \frac{|c(t,s)|^{2/3}}{\int_{-T}^T |c(u,s)|^{2/3} du}, \quad |t| < T, \quad (2.14)$$

we obtain from (2.12) a sequence $\{D_n^*\}$ of sampling designs which is asymptotically optimal in the sense that [3]

$$\frac{\epsilon_{D_n^*}^2(s) - \epsilon_I^2(s)}{\inf_D \epsilon_D^2(s) - \epsilon_I^2(s)} \rightarrow 1 \quad \text{as } n \rightarrow \infty$$

where the infimum is taken over all sampling designs of size n . In this case, the excess mean-square error for the optimal-coefficients discrete-time predictor satisfies

$$n^2 [\epsilon_{D_n^*}^2(s) - \epsilon_I^2(s)] \rightarrow \frac{r}{12} \left\{ \int_{-T}^T |c(t,s)|^{2/3} dt \right\}^3. \quad (2.15)$$

It is seen from (2.14) that the sampling density $p_s(t)$ of the asymptotically optimal sequence of designs depends explicitly on the filter $c(t,s)$ and thus on the correlation function $R(t)$ of the process X via (2.9). This is, of course, in sharp contrast to equally-spaced sampling. Furthermore, in general, the optimal sampling density $p_s(t)$ depends on the time s at which prediction is required. As will be seen from the example at the end of this section, it could happen that the filter function $c(t,s)$ depends on the prediction time s in a factorable form: $c(t,s) = c_1(t)c_2(s)$, in which case the optimal sampling density $p_s(t)$ of (2.14) no longer depends on the prediction time s : $p_s(t) = p(t)$ - a desirable feature. In the general case one can derive a sampling density $p(t)$, independent of s , by averaging the asymptotic constant in (2.13) with a weight function $w(s)$ reflecting the prediction requirement as one moves away from the observation interval $[-T, T]$:

$$\frac{r}{12} \frac{\int_{-T}^T \left(\int_{-T}^T c^2(t,s) w(s) ds \right) \frac{dt}{p^2(t)},$$

and then choosing the sampling density $p(t)$ which minimizes this averaged constant. The resulting sampling density $p(t)$ is proportional to

$$\left(\int_T^\infty c^2(t,s)w(s)ds \right)^{1/3}$$

and thus does not depend on s , with corresponding asymptotic constant for the excess prediction error

$$\frac{r}{12} \int_{-T}^T \frac{c^2(t,s)}{\left(\int_T^\infty c^2(t,u)w(u)du \right)^{2/3}} dt$$

which does, of course, depend on s .

A.2. Nonoptimal Predictors With Median Sampling

In this subsection we consider the performance of predictors $X_D(s)$ whose coefficients ξ_D are not optimal. As indicated earlier, such nonoptimal-coefficients predictors can be obtained by a discrete-time approximation of the optimal continuous-time predictor which is now given by (2.11). A discrete-time approximation to (2.11) is of the form

$$X_{D_n}(s) = a_0 X(-T) + b_0 X(T) + \frac{1}{n-2} \sum_{k=1}^{n-2} \frac{c(t_{n,k})}{p(t_{n,k})} X(t_{n,k}) \quad (2.16)$$

for $n \geq 2$, where the coefficients a_0, b_0 and the function $c(t)$ are those of the optimal continuous-time predictor. Here the sampling instants are

$D_n = \{-T, t_{n,1}, \dots, t_{n,n-2}, T\}$ where the sampling points $\{t_{n,1}, \dots, t_{n,n-2}\}$ are the medians of a regular sequence of designs generated by a positive and continuous sampling density $p(t)$ on $[-T, T]$, i.e., $\{t_{n,i}\}_{i=1}^{n-2}$ satisfy

$$\int_{-T}^{t_{n,i}} p(t)dt = \frac{2i-1}{2(n-2)}, \quad i = 1, \dots, n-2, \quad n \geq 3. \quad (2.17)$$

Note that in this case the sampling points $\{t_{n,i}\}_{i=1}^{n-2}$ used for the discrete approximation of the integral $\int_{-T}^T c(\tau)X(\tau)d\tau$ are all in the interior of $[-T, T]$.

For example when $p(t) = \frac{1}{2T} 1_{[-T, T]}(t)$ we have

$$t_{n,i} = T \frac{2i-n+1}{n-2}, \quad i=1, \dots, n-2, \quad n \geq 3.$$

The mean-square prediction error is given by

$$\begin{aligned} e_{D_n}^2(s) &= E[X(s) - X_{D_n}(s)]^2 \\ &= R(0)[1 + a_0^2 + b_0^2] - 2a_0R(s+T) - 2b_0R(s-T) + 2a_0b_0R(2T) \\ &\quad - \frac{2}{n-2} \sum_{k=1}^{n-2} \frac{c(t_{n,k})}{p(t_{n,k})} [R(s - t_{n,k}) - a_0R(T + t_{n,k}) - b_0R(T - t_{n,k})] \\ &\quad + \frac{1}{(n-2)^2} \sum_{j=1}^{n-2} \sum_{k=1}^{n-2} \frac{c(t_{n,j})}{p(t_{n,j})} \frac{c(t_{n,k})}{p(t_{n,k})} R(t_{n,j} - t_{n,k}). \end{aligned} \quad (2.18)$$

The excess error of the discrete-time predictor (2.16) is then precisely the error in the integral approximation, i.e.,

$$e_{D_n}^2(s) - \varepsilon_I^2(s) = E\left[\int_{-T}^T c(t)X(t)dt - \frac{1}{n-2} \sum_{k=1}^{n-2} \frac{c(t_{n,k})}{p(t_{n,k})} X(t_{n,k})\right]^2$$

for $n \geq 3$. From the result of Schoenfelder [6] (see also Cambanis and Masry [1]) we have

$$n^2 [e_{D_n}^2(s) - \varepsilon_I^2(s)] \rightarrow \frac{r}{12} \int_{-T}^T \frac{c^2(t,s)}{p^2(t)} dt \quad (2.19)$$

as $n \rightarrow \infty$, where $r = R'(0-) - R'(0+) > 0$. The rectangular rule of integral approximation was used in (2.16) because among all Newton-Cotes formulae of order one, it results in the smallest value of the asymptotic constant (2.19) as was

shown by Schoenfelder [6].

It is seen from (2.19) that the nonoptimal-coefficients predictors (2.16) have the same asymptotic performance as the optimal-coefficients predictors (2.2). All comments made earlier about asymptotically optimal sequences of designs and optimal sampling densities $p(t)$ are applicable here verbatim. For a small sample size n one expects the optimal-coefficients discrete-time predictor (2.2) to have a smaller excess error than that of the nonoptimal-coefficients predictor (2.16). Concerning the ease of implementation, the optimal-coefficients predictors (2.2) require the inversion of the covariance matrices $[R(t_{n,k} - t_{n,j})]_{n,j=1}^n$ for each sample size n , which is not especially difficult to accomplish on a computer. On the other hand, the nonoptimal-coefficients predictors (2.16) require the solution of the prediction integral equation (2.9) with $k=0$ for the determination of a_0, b_0 and $c(t)$ in (2.16), which may be a more complex task, but is done once regardless of the sample size n .

A.3 Example

We now illustrate by an example the performance of the optimal-coefficients and nonoptimal-coefficients predictors in conjunction with uniform or asymptotically optimal sampling schemes. We focus our attention on the performance for finite sample size n , as the asymptotic performance is clear from (2.13) and (2.19). In particular we obtain the improvement in performance that the asymptotically optimal sampling (with density $p(t)$ of (2.14)) provides over the commonly used uniform sampling.

We assume the process X has spectral density

$$\Phi(\lambda) = \frac{2\beta^3}{\pi(\alpha^2 + \beta^2)} \frac{\alpha^2 + \lambda^2}{(\beta^2 + \lambda^2)^2}$$

and correlation function

$$R(t) = e^{-\beta|t|} \left[1 + \beta \frac{(\alpha^2 - \beta^2)}{\alpha^2 + \beta^2} |t| \right] \quad (2.20)$$

where $\alpha, \beta > 0$. Note that $R(0) = 1$ and that X has no quadratic-mean derivatives. When $\alpha = \beta$, X is first-order Markov so that the optimal continuous-time predictor of $X(s)$, $s > T$, from the observations $\{X(\tau), |\tau| \leq T\}$ uses only the data point $X(T)$, namely

$$\hat{X}_I(s) = e^{-\beta(s-T)} X(T),$$

and the question of sampling designs does not arise. Hence we shall assume $\alpha \neq \beta$ in the following.

The optimal continuous-time predictor is of the form (2.11) with a_0, b_0 , and $c(t)$ being the solution of the integral equation (2.9) with $k = 0$. They can be found by the method in Rozanov [2] and, after lengthy computations, we have

$$a_0(s) = -\frac{2\alpha}{d_4(T)} (\alpha^2 - \beta^2)^2 (s - T) e^{-\beta(s-T)}, \quad (2.21a)$$

$$b_0(s) = \left\{ 1 + \frac{\beta_3(T)}{d_4(T)} (\alpha^2 - \beta^2) (s - T) \right\} e^{-\beta(s-T)}, \quad (2.21b)$$

$$c(t, s) = -(\alpha^2 - \beta^2)^2 \frac{d_2(t)}{d_4(T)} (s - T) e^{-\beta(s-T)} = d_2(t) c_2(s), \quad (2.21c)$$

where

$$d_m(t) = (\alpha + \beta)^m e^{-\alpha(t+T)} - (\alpha - \beta)^m e^{-\alpha(t+T)}, \quad (2.21d)$$

$$\dot{d}_m(t) = (\alpha + \beta)^m e^{-\alpha(t+T)} + (\alpha - \beta)^m e^{-\alpha(t+T)}. \quad (2.21e)$$

The mean-square prediction error $\epsilon_I^2(s)$ for the optimal continuous-time predictor, given by (2.4), can be computed explicitly to yield

$$\epsilon_I^2(s) = 1 - a_0(s)R(s+T) - b_0(s)R(s-T) - c_2(s)A(s) \quad (2.22a)$$

where

$$\begin{aligned} A(s) = 2e^{-\beta s} \{ (\alpha + \beta)e^{\alpha T} [(1 + Bs + \frac{B}{\alpha + \beta}) \sinh(\alpha + \beta)T - BT \cosh(\alpha + \beta)T] \\ - (\alpha - \beta)e^{-\alpha T} [(1 + Bs - \frac{B}{\alpha - \beta}) \sinh(\alpha - \beta)T + BT \cosh(\alpha - \beta)T] \}, \end{aligned} \quad (2.22b)$$

and

$$B = \frac{\beta^2 - \alpha^2}{\alpha^2 + \beta^2}. \quad (2.22c)$$

For the discrete-time predictor (2.2) with optimal coefficients and sample size n , the mean-square error $\epsilon_{D_n}^2(s)$ is given by (2.3), whereas for the nonoptimal-coefficients discrete-time predictor (2.16) with sample size n the mean-square error $\epsilon_{D_n}^2(s)$ is given by (2.18).

In the following numerical results the observation interval is set to $[-1, 1]$, i.e. $T = 1$, and the parameter β of the correlation function $R(t)$ is set to $\beta = 1$. After a preliminary investigation, the value of the parameter α in the correlation function $R(t)$ was set to $\alpha = 15$. The reasons for this choice are as follows. When $\alpha = 1$ we have a first-order Markov process X for which the continuous-time predictor with optimal coefficient uses only the endpoint $X(+1)$ so that, in this case, the question of sampling designs for prediction is not interesting. Thus we are interested in choosing $\alpha \neq 1$. For $0 < \alpha < 1$ numerical computations showed that the mean-square prediction errors, $\epsilon_I^2(s)$ and $\epsilon_{D_2}^2(s)$, for the optimal continuous-time predictor and respectively for the discrete-time predictor with optimal coefficients using only $X(\pm 1)$,

are essentially identical so that the sampling design problem is again of no interest. For $\alpha > 1$ the numerical results showed that the difference between $\epsilon_I^2(s)$ and $\epsilon_{D_2}^2(s)$ becomes more pronounced as α increases. For example for $\alpha = 5.2$ we have $\epsilon_I^2(2) = .44$ and $\epsilon_{D_2}^2(2) = .48$ so that the fractional error for $n = 2$ is already too small to exhibit the performance of the design for different sample sizes n . For $\alpha = 10$ the corresponding numbers are $\epsilon_I^2(2) = .38$ and $\epsilon_{D_2}^2(2) = .458$. It is thus seen that we need to choose α much larger than $\beta = 1$ in order to deviate sufficiently from the Gauss-Markov case. We have chosen $\alpha = 15$ in order for the sampling design problem to be of interest.

A.3.1. Optimal-Coefficients Predictors

Figure 1 compares the error $\epsilon_I^2(s)$ of the continuous-time predictor with the error $\epsilon_{D_n}^2(s)$ of the discrete-time predictor with optimal coefficients and equally spaced samples, for prediction lags $s - T \in [0, 3]$. It is seen that for a sample size $n = 10$, the two mean-square errors are very close. Note that for very small and for very large prediction lags, the performance of the two predictors should be very close even when the sample size is small ($n = 2$) as expected intuitively, since for zero lag, the prediction error in both cases is zero, and as the lag tends to infinity, the prediction error approaches $R(0) = 1$.

Figure 2 provides a similar comparison when the discrete-time predictor with optimal coefficients uses the asymptotically optimal sampling instants (a regular sequence (2.12) generated by the sampling density $p(t)$ of (2.14)). While in general $p(t)$ of (2.14) depends on the prediction time s , leading to different sampling instants for different values of s (cf. (2.12)), it is seen from the dependence of $c(t)$ on s in (2.21c) that for this example $p(t)$

is proportional to

$$p(t) \sim |(\alpha + \beta)^2 e^{\alpha(t+T)} - (\alpha - \beta)^2 e^{-\alpha(t+T)}|^{2/3}, \quad |t| \leq T, \quad (2.23)$$

which is independent of s , so that the sampling instants $\{t_{n,k}\}_{k=1}^n$ generated by $p(t)$ via (2.12) are the same for *all* prediction times $s > T$. It is seen from Figure 2 that with only $n=3$ samples the performance of the discrete-time predictor is already very close to that of the continuous-time predictor. The improvement that the asymptotically optimal sampling provides over uniform sampling is illustrated in Figure 3, where the fractional error

$$\gamma_n^2(s) = \frac{\varepsilon_D^2(s)}{\varepsilon_I^2(s)} - 1 \quad (2.24)$$

is plotted as a function of the sample size n , for both uniform and asymptotically optimal sampling. The dramatic improvement provided by asymptotically optimal sampling is readily apparent for moderate values of the prediction lag $s - T$. For example for lag $s - T = 1$, using $n=3$ samples, we have

$$\gamma_3^2(1) = \begin{cases} .171 & , \text{ for uniform sampling,} \\ .076 & , \text{ for asymptotically optimal sampling,} \end{cases}$$

an improvement factor of 2.25, and for $n=5$ samples,

$$\gamma_5^2(1) = \begin{cases} .087 & , \text{ for uniform sampling,} \\ .0026 & , \text{ for asymptotically optimal sampling,} \end{cases}$$

an improvement factor of 33.4.

A.3.2. Nonoptimal-Coefficients Predictors

For the discrete-time predictor (2.16) which uses nonoptimal coefficients and whose mean-square prediction error $e_{D,n}^2(s)$ is given by (2.18) it was found that its performance with *uniform sampling* is exceptionally bad. Specifically even for $n=10$ samples, the error exceeds $R(0)=1$ for all lags $s-T \in (.1,3]$, and equals 4.3 for lag $s-T=1$. Matters are much worse for $n=2$ samples where the error equals 23.06 for lag $s-T=1$. This behavior can be explained as follows: The excess error for the nonoptimal-coefficients predictor (2.16) is precisely due to the approximation of the integral $\int_{-T}^T c(t)X(t)dt$ by the sum in (2.16). Now for the selected values of α and β , $c(t)$ of (2.21c) has most of its mass concentrated near the right end point $t=T=1$. When uniform sampling points are used, only for very large values of n there will be enough samples near the right end point of the interval $[-T,T]$ to provide a reasonable approximation for the integral. This problem does not arise with predictors using uniform sampling and optimal coefficients since in this case the coefficients $\{c_{n,k}\}$ of the predictor are the best possible and each $c_{n,k}$ depends on *all* sampling points $\{t_{n,k}\}$ - indeed even for $n=2$ samples the performance is quite good as seen from Figure 1.

When asymptotically optimal sampling is used (cf. (2.23) and (2.17)) the performance of the nonoptimal-coefficients predictor is quite reasonable. This is because the mass of $p(t)$ - which is again independent of the prediction time s - is concentrated near the right end point of the interval $[-T,T]$, just like $c(t,s)$, and thus the sampling instants $\{t_{n,k}\}_{k=1}^{n-2}$ generated by $p(t)$ via (2.17) are also clustered around the end point T . Figure 4 provides a comparison with the performance of the continuous-time predictor as a function of prediction lag

$s - T$. Note that when $n = 5$ samples are used, the performance of the discrete and continuous time predictors is very close. The fractional error is plotted in Figure 5 as a function of the sample size n for various values of the prediction lag $s - T$.

If we compare the performance of the optimal-coefficients and nonoptimal-coefficients predictors when both use the (asymptotically optimal) sampling density $p(t)$ of (2.23), we find that the optimal-coefficients predictor is superior. For example for lag $s - T = 1$ we have

$\gamma_n^2(1)$	n	2	3	5	8
Optimal predictor		.25	.016	.0026	.00066
Nonoptimal predictor		63.6	.28	.0163	.0026

Thus the optimal-coefficients predictor requires considerably fewer samples to achieve a prescribed level of error.

B. Exactly One Quadratic-Mean Derivative ($k = 1$)

In this case the optimal continuous-data predictor has the form

$$\hat{X}_I(s) = a_0 X(-T) + b_0 X(T) + a_1 X'(-T) + b_1 X'(T) + \int_{-T}^T c(t) X(t) dt \quad (2.25)$$

where a_j, b_j and $c(t)$ are the solutions of the integral equation (2.9) with $k = 1$, and thus depend on the time s of prediction, which is suppressed.

B.1 Optimal-Coefficients Predictors

When the optimal-coefficients discrete-data predictor is used, the excess mean-square error can be written in the form $\|f_s - P_D f_s\|^2$, as when $k = 0$, but now the function $f_s(t)$, $|t| \leq T$, is of the form

$$f_s(t) = a_0 R(t+T) + b_0 R(t-T) - a_1 R'(t+T) - b_1 R'(t-T) + \int_{-T}^T c(\tau) R(t-\tau) d\tau$$

and the results of Sacks and Ylvisaker [3] are no longer applicable because of the presence of derivatives of the covariance. Thus no precise rates of convergence to zero of the excess error are available, and the subsequent results on non-optimal-coefficients in subsection B.2 provide upper bounds for the optimal-coefficients predictor. Some conjectures are also offered in subsection B.3.

B.2 Nonoptimal-Coefficients Predictors

Using the trapezoidal rule in approximating the integral $\int_{-T}^T c(t) X(t) dt = \int cX$ in (2.25) by

$$I_n = \frac{1}{n} \sum_{k=1}^{n-1} \frac{1}{2} \left\{ \frac{c(t_{n,k})}{p(t_{n,k})} X(t_{n,k}) + \frac{c(t_{n,k+1})}{p(t_{n,k+1})} X(t_{n,k+1}) \right\},$$

we obtain a simple nonoptimal-coefficients discrete-data predictor of the form

$$\begin{aligned}
X_{D_n}(s) = & a_0 X(-T) + b_0 X(T) + a_1 \frac{X(t_{n,2}) - X(-T)}{t_{n,2} + T} + b_1 \frac{X(T) - X(t_{n,n-1})}{T - t_{n,n-1}} \\
& + \frac{1}{n} \left\{ \frac{1}{2} \frac{c(-T)}{p(-T)} X(-T) + \sum_{k=2}^{n-1} \frac{c(t_{n,k})}{p(t_{n,k})} X(t_{n,k}) + \frac{1}{2} \frac{c(T)}{p(T)} X(T) \right\}
\end{aligned} \quad (2.26)$$

where $\{D_n\}$ is a regular sequence of designs generated by the continuous and positive sampling density $p(t)$, i.e. the sampling points in D_n are $-T = t_{n,1} < t_{n,2} < \dots < t_{n,n-1} < t_{n,n} = T$ and are specified by (2.12). Denoting by $\Delta_{-T,n}, \Delta_{T,n}$ the errors in the approximation of the quadratic-mean derivatives, i.e. $\Delta_{-T,n} = X'(-T) - [X(t_{n,2}) - X(-T)]/(t_{n,2} + T)$, $\Delta_{T,n} = X'(T) - [X(T) - X(t_{n,n-1})]/(T - t_{n,n-1})$, we can express as follows the excess mean square error in view of (2.6), (2.25) and (2.26),

$$\begin{aligned}
e_{D_n}^2(s) - \varepsilon_1^2(s) = & a_1^2 E(\Delta_{-T,n}^2) + b_1^2 E(\Delta_{T,n}^2) \\
& + 2a_1 b_1 E(\Delta_{-T,n} \Delta_{T,n}) \\
& + 2E\{(a_1 \Delta_{-T,n} + b_1 \Delta_{T,n})(\int cX - I_n)\} \\
& + E(\int cX - I_n)^2.
\end{aligned} \quad (2.27)$$

The asymptotic performance of each term is derived in the Appendix, assuming p and c/p are twice continuously differentiable, and this leads to

$$\begin{aligned}
e_{D_n}^2(s) - \varepsilon_1^2(s) = & \frac{1}{n} C_1(s) [1 + o(1)] \\
& + \frac{1}{n^2} C_2(s) [1 + o(1)] \\
& + \frac{1}{n^3} C_3(s) [1 + o(1)] \\
& + \frac{1}{n^4} C_4(s) [1 + o(1)],
\end{aligned} \quad (2.28)$$

as $n \rightarrow \infty$, where each line in (2.28) shows the asymptotic performance of the corresponding line in (2.27), and the constants $C_i(s)$ are specified in the appendix. Thus the overall slow rate n^{-1} is due to the slow convergence of the quadratic-mean derivative approximation, and we have (cf. (A.5))

$$n[e_{D_n}^2(s) - \varepsilon_I^2(s)] \rightarrow C_1(s) = \frac{1}{3}\rho \left\{ \frac{a_1^2(s)}{p(-T)} + \frac{b_1^2(s)}{p(T)} \right\} \quad (2.29)$$

where ρ is the jump of the third derivative of R at zero: $\rho = R^{(3)}(0+) - R^{(3)}(0-) > 0$. In particular, the slow rate n^{-1} is the maximum possible when uniform sampling is employed.

It is clear from (2.27) and (2.28) that the integral approximation has a much faster rate of convergence n^{-4} . This substantial loss of rate convergence can be averted by replacing in each n point sampling design D_n the points $t_{n,2}$ and $t_{n,n-1}$ by $(t_{n,2} + T)^4 - T$ and $T - (T - t_{n,n-1})^4$ respectively. The resulting modified regular sequence of designs $\{D'_n\}$ (which can no longer be uniform) has excess error whose term by term asymptotic performance is described by

$$\begin{aligned} e_{D'_n}^2(s) - \varepsilon_I^2(s) &= \frac{1}{4}C'_1(s)[1 + o(1)] \\ &+ \frac{1}{8}C'_2(s)[1 + o(1)] \\ &+ \frac{1}{6}C'_3(s)[1 + o(1)] \\ &+ \frac{1}{4}C'_4(s)[1 + o(1)]. \end{aligned} \quad (2.30)$$

In this case the rate of convergence is n^{-4} and, as $n \rightarrow \infty$,

$$n^4[e_{D'_n}^2(s) - \varepsilon_I^2(s)] \rightarrow C'_1(s) + C'_4(s) \quad (2.31)$$

where (cf. (A.8), (A.18))

$$C_1'(s) = \frac{1}{3} \left\{ \frac{a_1^2(s)}{p^4(-T)} + \frac{b_1^2(s)}{p^4(T)} \right\},$$

$$C_4(s) = \frac{1}{720} \int_{-T}^T \frac{c^2(t,s)}{p^4(t)} dt + C(s)$$

and $C(s)$ is given in (A.19) and depends on s only through $c(\pm T, s)$ and $c'(\pm T, s)$, and on p only through the boundary values $p(\pm T)$, $p'(\pm T)$. Because of the dependence of the asymptotic constant in (2.31) on the values of p and p' at the endpoints $\pm T$, its minimization with respect to $p(t)$ is messy and perhaps not feasible in view of the continuity requirements on p . On the other hand, the part of the asymptotic constant which depends on $p(t)$, $|t| < T$, i.e. $\int c^2 p^{-4}$, is minimized when $p(t)$ is proportional to $|c(t, s)|^{2/5}$.

A further small improvement can be achieved by bringing the modified sampling points even slightly closer to the endpoints so as to cancel asymptotically the effect of the approximation of the quadratic-mean derivatives. For instance if the modified sampling points are chosen by $t_{n,2}'' = (t_{n,2} + T)^{4+\epsilon} - T$, $t_{n,n-1}'' = T - (T - t_{n,n-1})^{4+\epsilon}$, $\epsilon > 0$, then the resulting modified regular sequence of designs $\{D_n''\}$ has, of course, the same rate but smaller asymptotic constant:

$$n^4 [e_{D_n''}^2(s) - \epsilon_1^2(s)] \rightarrow C_4(s). \quad (2.32)$$

B.3 Example

Here the process X has the spectral density

$$p(\lambda) = \frac{8\beta^5}{\pi(3\alpha^2 + \beta^2)} \frac{\alpha^2 + \lambda^2}{(\beta^2 + \lambda^2)^3}$$

and correlation function

$$R(t) = e^{-\beta|t|} \left[1 + \beta|t| + \frac{\beta^2(\alpha^2 - \beta^2)}{3\alpha^2 + \beta^2} t^2 \right]$$

where $\alpha, \beta > 0$. Note that $R(0) = 1$ and that X has precisely one quadratic-mean derivative. When $\alpha = \beta$, the process X is second-order Markov for which the optimal continuous-time predictor of $X(s)$, $s > T$, from the observation $\{X(t), |t| \leq T\}$ uses only the data points $X(T)$ and $X'(T)$. Since the derivative $X'(T)$ is not part of the observation, the case $\alpha = \beta$ is still of some interest. However, we shall consider the case $\alpha \neq \beta$ in order to exhibit the performance of sampling designs which provide data points *inside* the interval $[-T, T]$.

The coefficients a_0 , b_0 , a_1 , b_1 , and $c(t)$ in the optimal continuous-time predictor (2.25) can be obtained as follows. From Rozanov [2] or Yaglom [7], it can be seen that $c(t)$ is of the form $c(t) = B_1 e^{-\alpha t} + B_2 e^{\alpha t}$ and substituting $h(t)$ of (2.10) with $k=1$ in the integral equation (2.9), carrying out the integration, and equating the coefficients of $t^\ell e^{\pm \beta t}$, $\ell = 0, 1, 2$, on both sides of (2.9) leads to a system of six equations in the unknowns $a_0, b_0, a_1, b_1, B_1, B_2$ (the general expressions for a_1, b_1 given in Rozanov [2, p.137] are incorrect; hence the substitution approach taken here). After lengthy computations we find

$$a_0(s) = \frac{3\alpha\beta}{d_6(T)} (\alpha^2 - \beta^2)^3 (s-T)^2 e^{-\beta(s-T)},$$

$$a_1(s) = -a_0(s)/(3\beta),$$

$$b_0(s) = (1 + \beta(s-T) + [\frac{3}{2}\alpha\beta \frac{\delta_6(T)}{d_6(T)} - \frac{1}{2}(\alpha^2 - \beta^2)](s-T)^2) e^{-\beta(s-T)},$$

$$b_1(s) = (1 + \frac{1}{2}[\frac{\delta_6(T)}{d_6(T)} - \beta](s-T)) e^{-\beta(s-T)},$$

$$c(t, s) = \frac{1}{2}(\alpha^2 - \beta^2)^3 \frac{\delta_3(t)}{d_6(T)} (s-T)^2 e^{-\beta(s-T)} = \delta_3(t) c_3(s),$$

where d_m and δ_m are defined in (2.21d-e).

The mean square prediction error $\varepsilon_I^2(s)$ for the optimal continuous-time predictor is given by (2.4) and can be computed explicitly to yield

$$\varepsilon_I^2(s) = 1 - a_0 R(s+T) - b_0 R(s-T) + a_1 R'(s+T) + b_1 R'(s-T) - c_3(s) A(s)$$

where

$$\begin{aligned} A(s) = & 2e^{-\beta s} \{ (\alpha + \beta)^2 e^{\alpha T} \left[\left(1 + \frac{\beta}{\beta + \alpha} + \left[T^2 + \frac{2}{(\beta + \alpha)^2} \right] B + \left[\beta + \frac{2B}{\beta + \alpha} \right] s + Bs^2 \right) \sinh(\beta + \alpha) T \right. \\ & \left. - T \left(\beta + \frac{2B}{\beta + \alpha} + 2Bs \right) \cosh(\beta + \alpha) T \right] \\ & - (\alpha - \beta)^2 e^{-\alpha T} \left[\left(1 + \frac{\beta}{\beta - \alpha} + \left[T^2 + \frac{2}{(\beta - \alpha)^2} \right] B + \left[\beta + \frac{2B}{\beta - \alpha} \right] s + Bs^2 \right) \sinh(\beta - \alpha) T \right. \\ & \left. - T \left(\beta + \frac{2B}{\beta - \alpha} + 2Bs \right) \cosh(\beta - \alpha) T \right] \}, \end{aligned}$$

$$B = \frac{\beta^2(\alpha^2 - \beta^2)}{3\alpha^2 + \beta^2},$$

and

$$R'(t) = - \frac{\beta^2 t e^{-\beta t}}{3\alpha^2 + \beta^2} \{ (\alpha^2 + 3\beta^2) + \beta(\alpha^2 - \beta^2)t \}, \quad t \geq 0.$$

For the discrete-time predictor (2.2) with optimal coefficients and sample size n , the mean-square error $\varepsilon_{D_n}^2(s)$ is given by (2.3) whereas for the corresponding predictor (2.26) with nonoptimal coefficients the mean-square error $\varepsilon_{D_n}^2(s)$ is given by (2.27).

In the following numerical results the observation interval is set to $[-1, 1]$, i.e., $T=1$, and the parameter β of the correlation function $R(t)$ is set to $\beta=1$. The behavior of $\varepsilon_I^2(s)$ and $\varepsilon_{D_n}^2(s)$ (with uniform sampling) as a function of α , for a fixed s , was investigated numerically; results show that

as α increases both $\epsilon_I^2(s)$ and $\epsilon_{D_n}^2(s)$ decrease with $\epsilon_{D_n}^2(s)$ decreasing at a slower rate. For example for $s=1.2$ and $\alpha = .1, 1.2, 9.2$ the values of $\epsilon_I^2(s)$ are .64, .276, .074, respectively, whereas the values of $\epsilon_{D_n}^2(s)$ are .797, .398, .23 respectively. We have selected a moderate value of $\alpha = 2.5$.

B.3.1 Optimal-Coefficients Predictors

Figure 6 compares the error $\epsilon_I^2(s)$ of the continuous-time predictor with the error $\epsilon_{D_n}^2(s)$ of the discrete-time predictor with optimal coefficients and equally-spaced samples for prediction lags $s - T \in [0, 3]$. It is seen that $\epsilon_{D_n}^2(s)$ approaches $\epsilon_I^2(s)$, as n increases, rather slowly. In Figure 7 the fractional error (cf. (2.24)) is plotted as a function of n with the lag $s - T$ as parameter. It is clear that the fractional error is fairly large even for a sample size $n = 10$.

In view of the analysis in Sections B.1 and B.2 we are led to believe that the rate of convergence of $\epsilon_{D_n}^2(s) - \epsilon_I^2(s)$ to zero is perhaps $1/n$ due to the implicit approximation of the derivatives $X'(\pm T)$ by linear combinations of uniformly-spaced samples $\{X[2T(k-1)/(n-1) - T]\}_{k=1}^n$. It should be noted that the approximation of the integral $\int_{-T}^T c(t)X(t)dt$ by $\sum_{k=1}^n c_{n,k} X[2T(k-1)/(n-1) - T]$ has a rate of convergence of at least $1/n^4$.

The numerical results can shed light on the rate of convergence. Suppose that the asymptotic result is of the form

$$n^k [\epsilon_{D_n}^2(s) - \epsilon_I^2(s)] \rightarrow K(s) \quad (2.33)$$

is $n \rightarrow \infty$ for some (unknown) constants k and $K(s)$. Then as $n \rightarrow \infty$,

$$n^k \gamma_n^2(s) \rightarrow K(s) \epsilon_I^2(s) = Q(s).$$

In Figure 8 a plot of $n^k \gamma_n^2(s)$ is given, for lag $s - T = 1.5$, as a function of the sample size n for possible values of $k = 1, 2, 3, 4$. It is evident that when $k = 1$, $n \gamma_n^2(s)$ approaches a constant fairly quickly, supporting our belief that the rate of convergence in (2.33) is indeed $1/n$. We also conducted a mean-square fit for k and $Q(s)$ by minimizing

$$J(k) = \sum_{n=n_1}^{n_2} \{n^k \gamma_n^2(s) - Q(s)\}^2$$

with respect to k and $Q(s)$. For the range of sample sizes $n = 15, \dots, 30$ ($n_1 = 15$, $n_2 = 30$) we found that the best fit is

$$k = 1.035$$

and

$$K(s) = \begin{cases} 10.118 & \text{for } s - T = 1.0 \\ 2.35 & \text{for } s - T = 1.5 \\ .848 & \text{for } s - T = 2.0 \\ .379 & \text{for } s - T = 2.5 \end{cases}$$

Interestingly, the best fit for k turned out to be independent of the 4 values of s listed above. This result clearly supports the conjecture that the rate of convergence of $\varepsilon_{D,n}^2(s) - \varepsilon_I^2(S)$ is $1/n$.

The above slow rate of convergence can be dramatically improved if we modify the uniform sampling scheme $t_{n,i} = T(2i - n - 1)/(n - 1)$, $i = 1, \dots, n$, by appropriately shifting $t_{n,2}$ and $t_{n,n-2}$ towards the end of the data interval $[-T, T]$ so as to achieve a better approximation of the derivatives $X'(T)$ and $X'(-T)$. Suppose, for example, we let

$$t_{n,2} = \left(\frac{2T}{n-1}\right)^4 - T, \tag{2.34}$$

$$t_{n,n-2} = T - \left(\frac{2T}{n-1}\right)^4,$$

for $n > 2$, so that the approximation of the derivative $X'(-T)$ by only two data points $X(t_{n,1})$ and $X(t_{n,2})$ (and similarly for $X'(T)$ by $X(t_{n,n})$ and $X(t_{n,n-1})$) achieves a rate of convergence of $1/n^4$. We then expect the performance of the discrete-time predictor with such a modified uniform sampling to improve dramatically in comparison to uniform sampling. This turns out to be the case: In Figure 9, $\epsilon_{D,n}^2(s)$ is plotted as a function of the lag $s-T$ for sample sizes $n=2,5$ when this modified uniform sampling scheme (2.34) is used. It is seen that with $n=5$, the performance is already very close to that of the continuous-time predictor (contrast it with Figure 6 for uniform sampling). This sharp improvement for the modified uniform sampling scheme can be seen more clearly in Figure 10, where the fractional error $\gamma_n^2(s)$ is displayed as a function of n for 3 selected values of s ; it is seen that the modified sampling scheme with $n=7$ outperforms the uniform sampling scheme with $n=30$ by a factor of about 10. For example, for lag $s-T=1.5$, modified uniform sampling with only $n=7$ gives a fractional error of 5.12×10^{-3} , whereas uniform sampling with $n=7$ gives a fractional error of .109 and even when $n=30$, we have a fractional error of 2.37×10^{-2} .

With this modified uniform sampling we expect a rate of convergence of $1/n^4$, but no such analytical result is yet available. Due to numerical instability in the inversion of the covariance matrix $R_{D,n}$ when modified uniform sampling (2.34) is used with $n \geq 7$, we were unable to computationally verify this conjectured rate of convergence.

B.3.2. Nonoptimal-Coefficients Predictors

When the nonoptimal-coefficients predictor (2.26) is used with equally-spaced samples, the mean-square error $e_{D,n}^2(s)$ has precisely a rate of convergence equal to $1/n$, by (2.29). Figure 11 compares $e_{D,n}^2(s)$ for $n=2,5,10$, with $e_I^2(s)$ of the optimal continuous-time predictor. More interestingly Figure 12 provides a comparison of performance between the discrete-time predictor with optimal coefficients

and that with nonoptimal coefficients; here the fractional error is plotted for each case as a function of the sample size n for 3 representative values of the lag $s - T$. It is seen that the performance of the two predictors is fairly close in this case with equal rate of convergence ($1/n$).

From (2.31) we know that by modifying the two sampling points $t_{n,2}$ and $t_{n,n-2}$ as in (2.34) we obtain a precise rate of convergence $1/n^4$ for the discrete-time predictor with nonoptimal coefficients (2.26). The performance for finite sample size $n = 2, \dots, 30$, is displayed in Figures 13-15. Figure 13 exhibits the mean-square error $e_{D_n}^2(s)$ for $n = 2, 5, 10$, and $\epsilon_I^2(s)$ of the optimal continuous-time predictor. This should be compared with Figure 11 where uniform sampling is employed. Such a comparison is more clearly displayed in Figure 14 from which the dramatic reduction in the fractional error is evident (for a fixed lag) under the modified uniform sampling scheme.

Finally one may wish to compare the performance of the two discrete-time predictors, with optimal coefficients and with nonoptimal coefficients, (2.26), both using the modified uniform sampling scheme. Such a comparison is given in Figure 15 from which it is seen that for small sample size $n \leq 7$ the predictor with optimal coefficients significantly outperforms the one with nonoptimal coefficients. This is considerably more pronounced here under a modified uniform sampling than in Figure 12 under uniform sampling. One possible explanation is that the predictor with optimal coefficients does a much better job in estimating the derivatives $X'(\pm T)$ than the one with nonoptimal coefficients, implicitly using all data points instead of just two points as in (2.26).

APPENDIX

Here we derive the asymptotics of the terms on the right hand side of (2.27).

A. Approximation of Quadratic-mean Derivatives

For the mean-square error in the approximation of the quadratic-mean derivatives (first line in (2.27)) we have

$$\begin{aligned}
 E \{ X^{(1)}(u) - \frac{1}{h} [X(u) - X(u-h)] \}^2 \\
 &= -R^{(2)}(0) - \frac{2}{h} [R^{(1)}(0) - R^{(1)}(h)] + \frac{2}{h^2} [R(0) - R(h)] \\
 &= -R^{(2)}(0) + \frac{2}{h} [hR^{(2)}(0) + \frac{1}{2}h^2R^{(3)}(0+) + o(h^2)] \\
 &\quad - \frac{2}{h^2} [\frac{1}{2}h^2R^{(2)}(0) + \frac{1}{6}h^3R^{(3)}(0+) + o(h^3)] \\
 &= \frac{2}{3}hR^{(3)}(0+) + o(h)
 \end{aligned} \tag{A.1}$$

where the right and left derivative corresponds to h positive and negative respectively.

For the cross correlation between the approximations of the quadratic-mean derivatives (second line in (2.27)) we have with $w = u - v \neq 0$, $h > 0$, $g > 0$,

$$\begin{aligned}
 E \{ X^{(1)}(u) - \frac{1}{h} [X(u) - X(u-h)] \} \{ X^{(1)}(v) - \frac{1}{g} [X(v+g) - X(v)] \} \\
 &= -R^{(2)}(w) - \frac{1}{g} [R^{(1)}(w-g) - R^{(1)}(w)] - \frac{1}{h} [-R^{(1)}(w) + R^{(1)}(w-h)] \\
 &\quad + \frac{1}{hg} [R(w-g) - R(w) - R(w-h-g) + R(w-h)]
 \end{aligned}$$

$$\begin{aligned}
&= -R^{(2)}(w) - \frac{1}{g}[-gR^{(2)}(w) + \frac{1}{2}g^2R^{(3)}(w) - \frac{1}{6}g^3R^{(4)}(w) + O(g^4)] \\
&\quad - \frac{1}{h}[-hR^{(2)}(h) + \frac{1}{2}h^2R^{(3)}(w) - \frac{1}{6}h^3R^{(4)}(w) + O(h^4)] \\
&\quad + \frac{1}{hg}[R^{(1)}(w)\{-g + (g+h) - h\} + R^{(2)}(w)\frac{1}{2}\{g^2 - (g+h)^2 + h^2\} \\
&\quad + R^{(3)}(w)\frac{1}{6}\{-g^3 + (g+h)^3 - h^3\} + R^{(4)}(w)\frac{1}{24}\{g^4 - (g+h)^4 + h^4\} + O((g+h)^5)] \\
&= -\frac{1}{4}ghR^{(4)}(w) + O(h^3 + g^3). \tag{A.2}
\end{aligned}$$

With $u = T$, $v = -T$, $h_n = T - t_{n,n-1}$, $g_n = t_{n,2} + T$, and a regular sequence of designs, it follows from the mean value theorem that

$$\begin{aligned}
\frac{1}{n} &= \int_{-T}^{t_{n,2}} p(t)dt = p(v_n)(t_{n,2} + T) \\
&= \int_{t_{n,n-1}}^T p(t)dt = p(u_n)(T - t_{n,n-1})
\end{aligned} \tag{A.3}$$

where $-T < v_n < t_{n,2}$, $t_{n,n-1} < u_n < T$, and thus

$$ng_n \rightarrow \frac{1}{p(-T)}, \quad nh_n \rightarrow \frac{1}{p(T)}. \tag{A.4}$$

Hence from (A.1) and (A.2) we obtain

$$\begin{aligned}
nE(\Delta_{-T,n}^2) &\rightarrow -\frac{2}{3} \frac{R^{(3)}(0-)}{p(-T)} = \frac{\rho}{3p(-T)}, \\
nE(\Delta_{T,n}^2) &\rightarrow \frac{2}{3} \frac{R^{(3)}(0+)}{p(T)} = \frac{\rho}{3p(T)}, \\
n^2E(\Delta_{-T,n}\Delta_{T,n}) &\rightarrow -\frac{R^{(4)}(2T)}{4p(-T)p(T)},
\end{aligned} \tag{A.5}$$

which establishes the first two lines in (2.28). Also when $t_{n,2}$ and $t_{n,n-1}$ are modified to

$$t'_{n,2} = (t_{n,2} + T)^m - T, \quad t'_{n,n-1} = T - (T - t_{n,n-1})^m \quad (\text{A.6})$$

then

$$g_n = t'_{n,2} + T = (t_{n,2} + T)^m, \quad h_n = T - t'_{n,n-1} = (T - t_{n,n-1})^m$$

and by (A.3),

$$n^m g_n \rightarrow \frac{1}{p^m(-T)}, \quad n^m h_n \rightarrow \frac{1}{p^m(T)}, \quad (\text{A.7})$$

so that

$$n^m E(\Lambda_{+T,n}^2) \rightarrow \frac{\rho}{3p^m(+T)}, \quad (\text{A.8})$$

$$n^{2m} E(\Lambda_{-T,n} \Lambda_{T,n}) \rightarrow - \frac{R^{(4)}(2T)}{4p^m(-T)p^m(T)}.$$

When $m=4$ we obtain the first two lines in (2.30) and when $m \neq 4$ both rates are faster than n^{-4} leading eventually to (2.32).

B. Approximation of Integral

The integral approximation mean-square error can be written in the form

$$E \left[\int_{-T}^T c(t)X(t)dt - I_n \right]^2 = \sum_{k,j=1}^n \int_{t_k}^{t_{k+1}} \int_{t_j}^{t_{j+1}} M_{k,j}(t,\tau) p(t)p(\tau) dt d\tau$$

$$\triangleq \sum_{k,j=1}^n J_{k,j} \quad (A.9)$$

where

$$M_{k,j}(t,\tau) = f(t)f(\tau)R(t-\tau) - \frac{1}{2}f(t)[f(t_j)R(t-t_j) + f(t_{j+1})R(t-t_{j+1})]$$

$$- \frac{1}{2}f(\tau)[f(t_k)R(\tau-t_k) + f(t_{k+1})R(\tau-t_{k+1})]$$

$$+ \frac{1}{4}[f(t_k)f(t_j)R(t_k-t_j) + f(t_k)f(t_{j+1})R(t_k-t_{j+1})]$$

$$+ f(t_{k+1})f(t_j)R(t_{k+1}-t_j) + f(t_{k+1})f(t_{j+1})R(t_{k+1}-t_{j+1})],$$

and $f(t) = c(t)/p(t)$, and where n is dropped from $t_{n,k}$ for ease of notation.

From (2.12) we have by the mean value theorem

$$\frac{1}{n} = \int_{t_k}^{t_{k+1}} p(t)dt = p(u_k)\Delta t_k \quad (A.11)$$

where $t_k < u_k < t_{k+1}$. In the following we will make use of the quantities

$$F_{0,k} = \int_{t_k}^{t_{k+1}} \left\{ f(t) - \frac{1}{2}[f(t_k) + f(t_{k+1})] \right\} p(t)dt,$$

$$F_{m,k} = \int_{t_k}^{t_{k+1}} \left\{ f(t)(t-t_k)^m - \frac{1}{2}f(t_{k+1})\Delta t_k^m \right\} p(t)dt,$$

$m = 1, 2, 3, 4, 5$. Their asymptotics are found by Taylor expanding $(fp)(t)$ and $p(t)$ about t_k . Using $\approx$ to denote equality up to higher order terms in Δt_k , we find

$$F_{m,k} \approx G_m \Delta t_k^3 \quad \text{for } m = 0, 1, 2, \quad (\text{A.12})$$

$$F_{m,k} \approx G_m \Delta t_k^{m+1} \quad \text{for } m = 3, 4, 5,$$

where

$$\begin{aligned} G_0 &= \frac{1}{12}(f'p' - f''p), \\ G_1 &= \frac{1}{12}(fp' - 2f'p), \\ G_2 &= -\frac{1}{6}fp, \end{aligned} \quad (\text{A.13})$$

and in these expressions all functions are evaluated at some, possibly distinct, point in $[t_k, t_{k+1})$.

For the diagonal terms $J_{k,k}$ in (A.9) we use the Taylor expansion of R about zero,

$$R(\tau) = R(0) + \frac{1}{2}\tau^2 R^{(2)}(0) + \frac{1}{6}\tau^3 R^{(3)}(\xi)$$

where ξ is in between 0 and τ . Since R does not have a third derivative at 0, in fact $R^{(3)}(0+) = -R^{(3)}(0-) = \rho/2 > 0$, we need to keep track of whether the intermediate point ξ is positive or negative. After considerable algebra we find that

$$J_{k,k} = R(0)F_{0,k}^2 + R^{(2)}(0)[F_{2,k}F_{0,k} - F_{1,k}^2] + R_{3,k} \quad (\text{A.14})$$

where the first two terms are of the order Δt_k^6 , by (A.12), and the third term involves the third order derivatives of R :

$$R_{3,k} = \int_{t_k}^{t_{k+1}} \int dt d\tau p(t) p(\tau) \frac{1}{12} \{ R^{(3)}(\xi_1) f(t) f(\tau) 2(t-\tau)^3 - R^{(3)}(\xi_2) f(t) f(t_k) (t-t_k)^3 \\ - R^{(3)}(\xi_3) f(t) f(t_{k+1}) (t_{k+1}-t)^3 - R^{(3)}(\xi_4) f(\tau) f(t_k) (\tau-t_k)^3 \\ - R^{(3)}(\xi_5) f(\tau) f(t_{k+1}) (t_{k+1}-\tau)^3 + R^{(3)}(\xi_6) f(t_k) f(t_{k+1}) \Delta t_k^3 \}.$$

The intermediate points ξ_2 to ξ_6 are between 0 and $t-t_k$, $t_{k+1}-t$, ..., Δt_k respectively, hence, all positive, while ξ_1 is between 0 and $t-\tau$. Thus the integral of the first term should be done separately above and below the diagonal of the square $(t_k, t_{k+1}) \times (t_k, t_{k+1})$, while all other integrals can be evaluated directly on the entire square. Using mean value theorem to pull the part of the integrand involving $R^{(3)}$, f and p out, and evaluating the resulting integral we find

$$R_{3,k} = \frac{1}{12} \{ R^{(3)}(\text{pos}) f p f p \frac{1}{10} + R^{(3)}(\text{neg}) f p f p (-\frac{1}{10}) \\ - R^{(3)}(\text{pos}) f p f p \frac{1}{4} - R^{(3)}(\text{pos}) f p f p \frac{1}{4} - R^{(3)}(\text{pos}) f p f p \frac{1}{4} \\ - R^{(3)}(\text{pos}) f p f p \frac{1}{4} + R^{(3)} f f p p \} \Delta t_k^5$$

where f and p are evaluated at points in (t_k, t_{k+1}) and $R^{(3)}$ at points tending to 0 with n whose sign is indicated. Hence $R_{3,k}$ is the dominant term in (A.14) and using its expression above along with (A.11) we obtain

$$\begin{aligned}
n^4 \sum_k J_{k,k} &\rightarrow \left\{ \frac{R^{(3)}(0+)}{120} - \frac{R^{(3)}(0-)}{120} - \frac{R^{(3)}(0+)}{12} + \frac{R^{(3)}(0+)}{12} \right\} \int_{-T}^T \frac{f^2(t) p^2(t)}{p^4(t)} dt \\
&= \frac{1}{120} [R^{(3)}(0+) - R^{(3)}(0-)] \int_{-T}^T \frac{f^2(t)}{p^2(t)} dt \\
&= \frac{1}{120} \rho \int_{-T}^T \frac{f^2(t)}{p^2(t)} dt.
\end{aligned} \tag{A.15}$$

For the off-diagonal terms $J_{k,j}$, $k \neq j$, in (A.9), we can Taylor expand R about $t_k - t_j \neq 0$ as far as it is necessary, as it is infinitely differentiable away from 0. We find that terms involving fifth and higher order derivatives of R are of higher order in $(\Delta t_k)(\Delta t_j)$, and after considerable algebra we have

$$\begin{aligned}
J_{k,j} &= R(t_k - t_j) [F_{0,k} F_{0,j}] + R^{(1)}(t_k - t_j) [F_{1,k} F_{0,j} - F_{0,k} F_{1,j}] \\
&\quad + \frac{1}{2} R^{(2)}(t_k - t_j) [F_{0,k} F_{2,j} + F_{2,k} F_{0,j} - 2F_{0,k} F_{0,j}] \\
&\quad + \frac{1}{6} R^{(3)}(t_k - t_j) [F_{3,k} F_{0,j} - F_{0,k} F_{3,j} + 3F_{1,k} F_{2,j} - 3F_{2,k} F_{1,j}] \\
&\quad + \frac{1}{24} R^{(4)}(t_k - t_j) [F_{0,k} F_{4,j} + F_{4,k} F_{0,j} - 4F_{1,k} F_{3,j} - 4F_{3,k} F_{1,j} + 6F_{2,k} F_{2,j}] \\
&\quad + \text{higher order terms.}
\end{aligned}$$

In fact, in view of (A.12), the terms involving $F_{3,k}$ and $F_{4,k}$ are also of higher order, and all remaining terms are of the order $\Delta t_k^3 \Delta t_j^3$. Using (A.11) and (A.12), (A.13), we find

$$n^4 \sum_{k \neq j} J_{k,j} \rightarrow \frac{1}{12^2} \iint_{t \neq \tau} \left[\sum_{m=0}^4 r_m(t, \tau) R^{(m)}(t - \tau) \right] dt d\tau \tag{A.16}$$

where

$$r_0(t, \tau) = \left(\frac{f'}{p}\right)'(t) \left(\frac{f'}{p}\right)'(\tau),$$

$$r_1(t, \tau) = \left[\frac{f'}{p}(t) + \left(\frac{f}{p}\right)'(t)\right] \left(\frac{f'}{p}\right)'(\tau) - \left(\frac{f'}{p}\right)'(t) \left[\frac{f'}{p}(\tau) + \left(\frac{f}{p}\right)'(\tau)\right],$$

$$r_2(t, \tau) = \frac{f}{p}(t) \left(\frac{f'}{p}\right)'(\tau) + \left(\frac{f'}{p}\right)'(t) \frac{f}{p}(\tau) - \left[\frac{f'}{p}(t) + \left(\frac{f}{p}\right)'(t)\right] \left[\frac{f'}{p}(\tau) + \left(\frac{f}{p}\right)'(\tau)\right],$$

$$r_3(t, \tau) = \left[\frac{f'}{p}(t) + \left(\frac{f}{p}\right)'(t)\right] \frac{f}{p}(\tau) - \frac{f}{p}(t) \left[\frac{f'}{p}(\tau) + \left(\frac{f}{p}\right)'(\tau)\right],$$

$$r_4(t, \tau) = \frac{f}{p}(t) \frac{f}{p}(\tau).$$

Putting together (A.15) and (A.16) we have

$$n^4 E(\int_0^T X - I_n)^2 \rightarrow \frac{0}{120} \int \frac{f^2}{p^2} + \frac{1}{12^2} \iint_{t \neq s} \left(\sum_{m=0}^4 r_m R^{(m)} \right) = C_4. \quad (A.17)$$

The expression of the asymptotic constant $C_4 = C_4(s)$ can be simplified considerably by integration by parts. For instance

$$\begin{aligned} \iint_{t \neq s} r_4 R^{(4)} &= \int_{-T}^T dt \frac{f}{p}(t) \left(\int_{-T}^t + \int_t^T \right) d\tau R^{(4)}(t - \tau) \frac{f}{p}(\tau) \\ &= \int_{-T}^T dt \frac{f}{p}(t) \left\{ \left[-R^{(3)}(t - \tau) \frac{f}{p}(\tau) \right]_{\tau=-T}^t + \left[R^{(3)}(t - \tau) \frac{f}{p}(\tau) \right]_{\tau=t}^T + \left(\int_{-T}^t + \int_t^T \right) d\tau R^{(3)}(t - \tau) \left(\frac{f}{p} \right)'(\tau) \right\} \\ &= \int_{-T}^T dt \frac{f}{p}(t) \left\{ -R^{(3)}(0+) \frac{f}{p}(t) + R^{(3)}(t+T) \frac{f}{p}(-T) - R^{(3)}(t-T) \frac{f}{p}(T) + R^{(3)}(0-) \frac{f}{p}(t) \right\} \\ &\quad + \int_{-T}^T \int_{-T}^T R^{(3)}(t - \tau) \frac{f}{p}(t) \left(\frac{f}{p} \right)'(\tau) dt d\tau \\ &= - \int_{-T}^T \frac{f^2}{p^2} + \int_{-T}^T \frac{f}{p}(t) \left[R^{(3)}(t+T) \frac{f}{p}(-T) - R^{(3)}(t-T) \frac{f}{p}(T) \right] dt \\ &\quad + \frac{1}{2} \int_{-T}^T \int_{-T}^T R^{(3)}(t - \tau) \left\{ \frac{f}{p}(t) \left(\frac{f}{p} \right)'(\tau) - \left(\frac{f}{p} \right)'(t) \frac{f}{p}(\tau) \right\} dt d\tau. \end{aligned}$$

The first term is of the same form as that coming from the diagonal terms.

Using repeated integration by parts on the remaining terms we obtain finally

$$C_4 = \frac{\rho}{720} \int \frac{f^2}{p} + C \quad (A.18)$$

where

$$\begin{aligned} 12^2 C = & R(0) \left[\left(\frac{f'}{p} \right)^2 (-T) + \left(\frac{f'}{p} \right)^2 (T) \right] - 2R(2T) \frac{f'}{p}(-T) \frac{f'}{p}(T) \\ & + R^{(1)}(-2T) \left[\frac{f'}{p}(-T) \frac{f}{p}(T) - \frac{f}{p}(-T) \frac{f'}{p}(T) \right] \\ & + 2R^{(2)}(2T) \frac{f}{p}(-T) \frac{f}{p}(T) - R^2(0) \left[\left(\frac{f}{p} \right)^2 (-T) + \left(\frac{f}{p} \right)^2 (T) \right] \end{aligned} \quad (A.19)$$

and $f = c/p$.

It should be noted that as follows from the work of Sacks and Ylvisaker [4,5] the first term on the right hand side of (A.18) is the asymptotic constant of the regular sequence of designs using optimal coefficients. Thus our estimator I_n of the integral is not asymptotically optimal, its asymptotic constant exceeding the least possible value by the amount C determined by (A.19).

C. Cross Correlation Between Quadratic-mean Derivative Approximation And Integral Approximation

Putting $\Delta_{T,n} = X^{(1)}(T) - [X(T) - X(T - h_n)]/h_n$ we have

$$E[\Delta_{T,n}(\int cX - I_n)] = \sum_{k=1}^n \int_{t_k}^{t_{k+1}} M_k(t) p(t) dt = \sum_{k=1}^n J_k$$

where

$$\begin{aligned} M_k(t) &= f(t) \{ -R^{(1)}(t - T) - \frac{1}{h_n} [R(t - T) - R(t - T + h_n)] \} \\ &\quad - \frac{1}{2} f(t_k) \{ -R^{(1)}(t_k - T) - \frac{1}{h_n} [R(t_k - T) - R(t_k - T + h_n)] \} \\ &\quad - \frac{1}{2} f(t_{k+1}) \{ -R^{(1)}(t_{k+1} - T) - \frac{1}{h_n} [R(t_{k+1} - T) - R(t_{k+1} - T + h_n)] \}, \end{aligned}$$

and $f = c/p$. As in A of Appendix, h_n satisfies (A.7) with $m=1$ or 4 or >4 . Since for each k , the argument of R and $R^{(1)}$ in the expression of M_k never vanishes in the interior of the interval, we can Taylor expand about $t_k - T$ and regrouping terms we have

$$\begin{aligned} J_k &= R^{(2)}(t_k - T) \frac{1}{2} h_n F_{0,k} + R^{(3)}(t_k - T) \left[\frac{1}{6} h_n^2 F_{0,k} + \frac{1}{2} h_n F_{1,k} \right] \\ &\quad + R^{(4)}(t_k - T) \left[\frac{1}{24} h_n^3 F_{0,k} + \frac{1}{6} h_n^2 F_{1,k} + \frac{1}{4} h_n F_{2,k} \right] \\ &\quad + \text{higher order terms,} \end{aligned}$$

i.e., the coefficients of $R(t_k - T)$, $R^{(1)}(t_k - T)$ vanish and those of $R^{(5)}$, etc., are of higher order in Δt_k . Then using (A.7), (A.12) and (A.13) we obtain

$$n^{2+m}E[\Delta_{T,n}(\int cX - I_n)] \rightarrow - \frac{D}{24p^m(T)}$$

where

$$D = \int_{-T}^T R^{(2)}(t-T) \left(\frac{f'}{p}\right)'(t) dt + \int_{-T}^T R^{(3)}(t-T) \left[\frac{f'}{p}(t) + \left(\frac{f}{p}\right)'(t)\right] dt + \int_{-T}^T R^{(4)}(t-T) \frac{f}{p}(t) dt$$

and can be simplified by integration by parts to

$$D = R^{(2)}(0) \left(\frac{f'}{p}\right)(T) - R^{(2)}(-2T) \left(\frac{f'}{p}\right)(-T) + R^{(3)}(0-) \frac{f}{p}(T) - R^{(3)}(-2T) \frac{f}{p}(-T).$$

When $m=1$ we obtain the third line in (2.28), when $m=4$ we get the third line in (2.30), and $m>4$ helps lead to (2.32).

REFERENCES

- [1] S. Cambanis and E. Masry, "Sampling designs for the detection of signals in noise," IEEE Trans. Inform. Theory, vol. IT-29, pp. 83-104, Jan. 1984.
- [2] Yu.A. Rozanov, Stationary Random Processes. San Francisco: Holden-Day, 1967.
- [3] J. Sacks and D. Ylvisaker, "Designs for regression problems with correlated errors," Ann. Math. Statist., vol 37, pp. 66-89, 1966.
- [4] J. Sacks and D. Ylvisaker, "Designs for regression problems with correlated errors III," Ann. Math. Statist., vol. 41, 2057-2074, 1970.
- [5] J. Sacks and D. Ylvisaker, "Statistical designs and integral approximation," in Proc. Twelfth Biennial Seminar of the Canadian Math. Congress, pp. 115-136, Canadian Math. Congress, Montreal, 1970.
- [6] C. Schoenfelder, "Random designs for estimating integrals of stochastic processes," Institute of Statistics Mimeo Series 1201, Univ. of North Carolina, 1978.
- [7] A.M. Yaglom, "Effective solutions of linear approximation problems for multivariate stationary processes with a rational spectrum," Theor. Probability Appl., pp. 239-264, 1960.

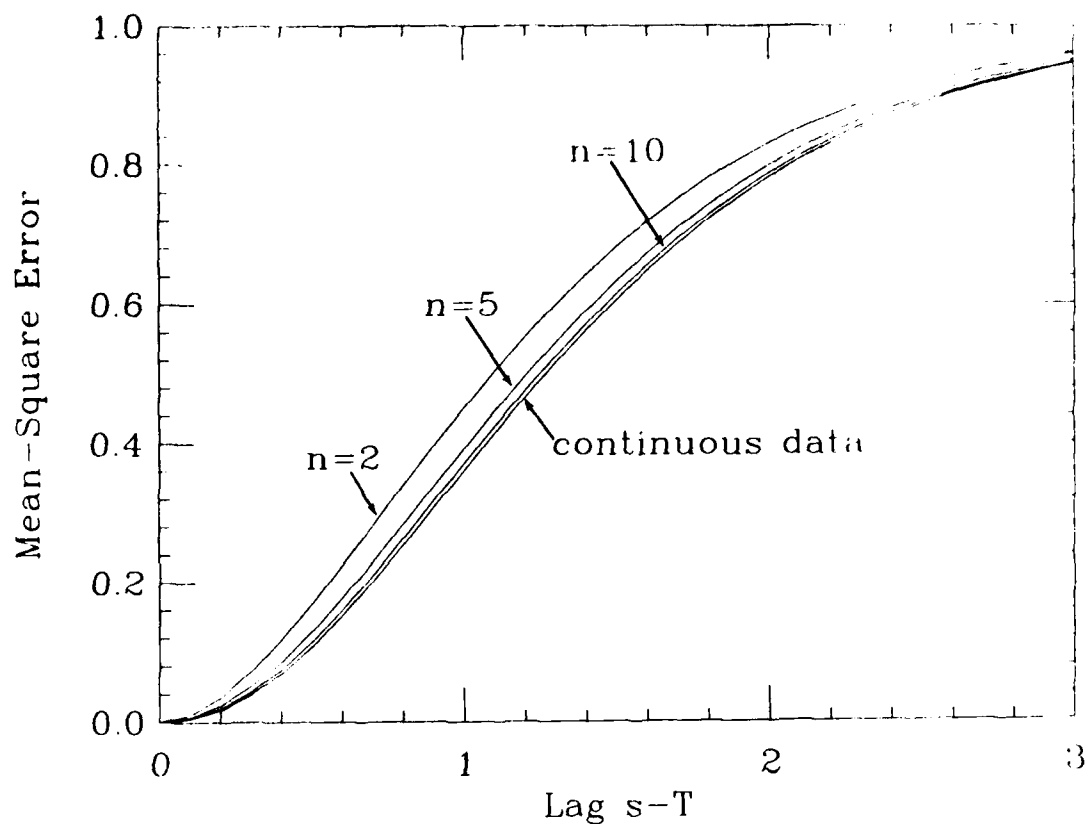


Figure 1. Case $k = 0$. Optimal coefficients, uniform sampling.

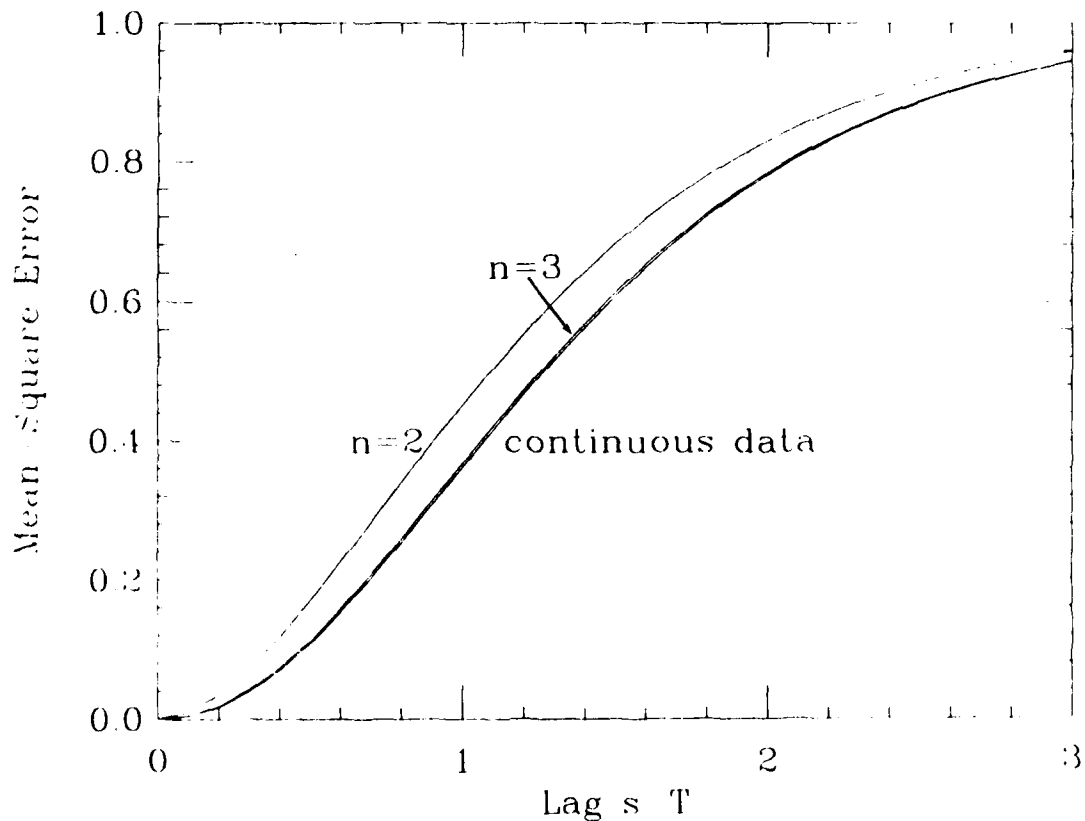


Figure 2. Case $k = 0$. Optimal coefficients, asymptotically optimal sampling.

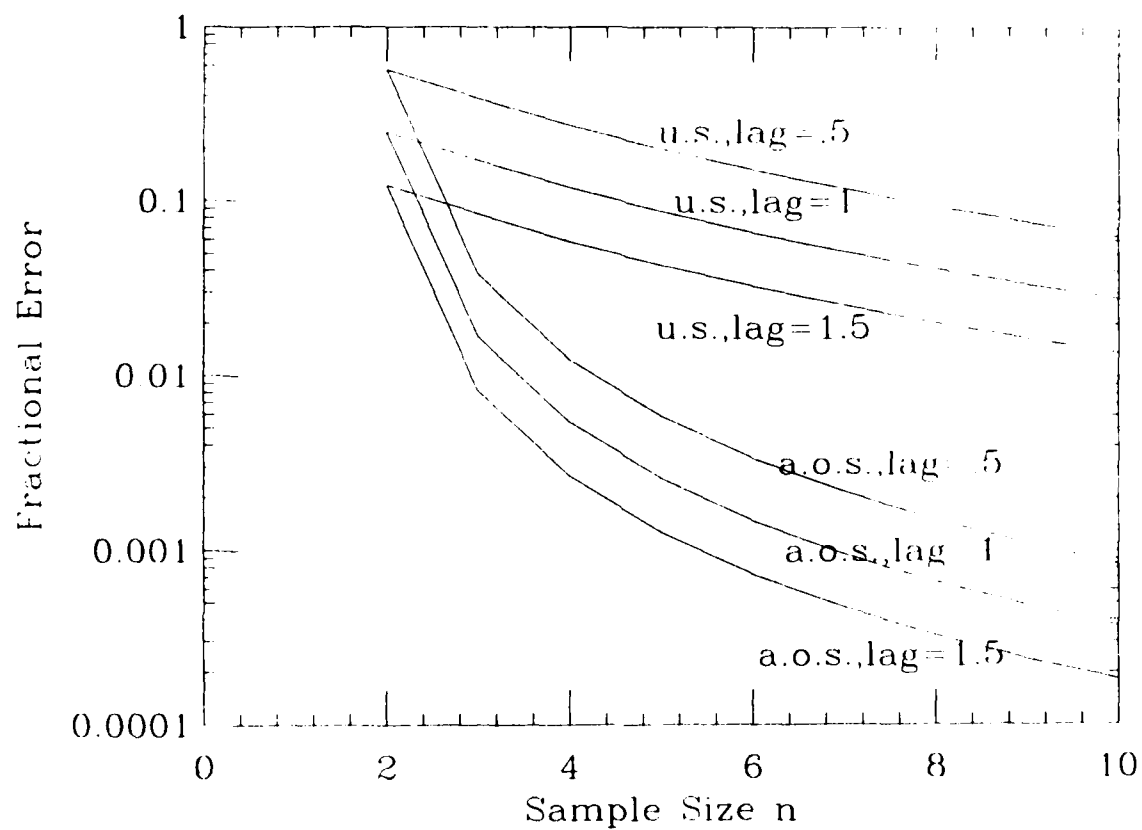


Figure 3. Case $k = 0$. Optimal coefficients.
Performance under uniform and asymptotically optimal sampling.

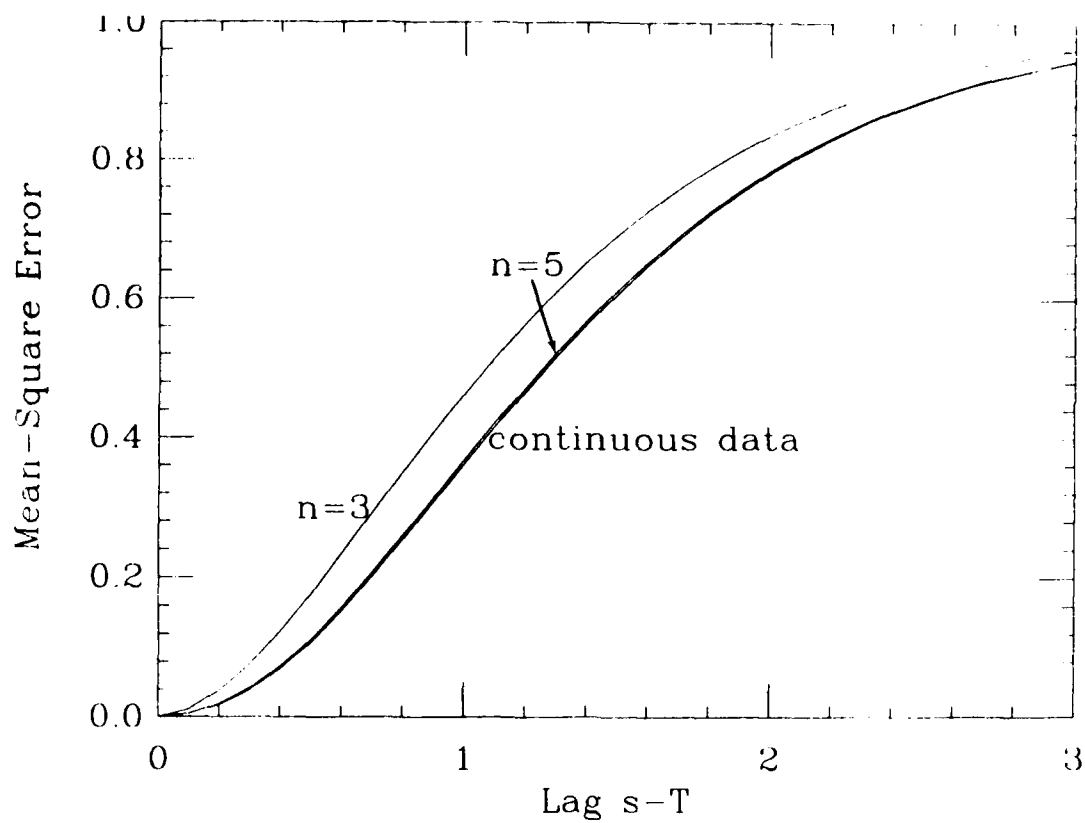


Figure 4. Case $k = 0$.
Nonoptimal coefficients, asymptotically optimal median sampling.

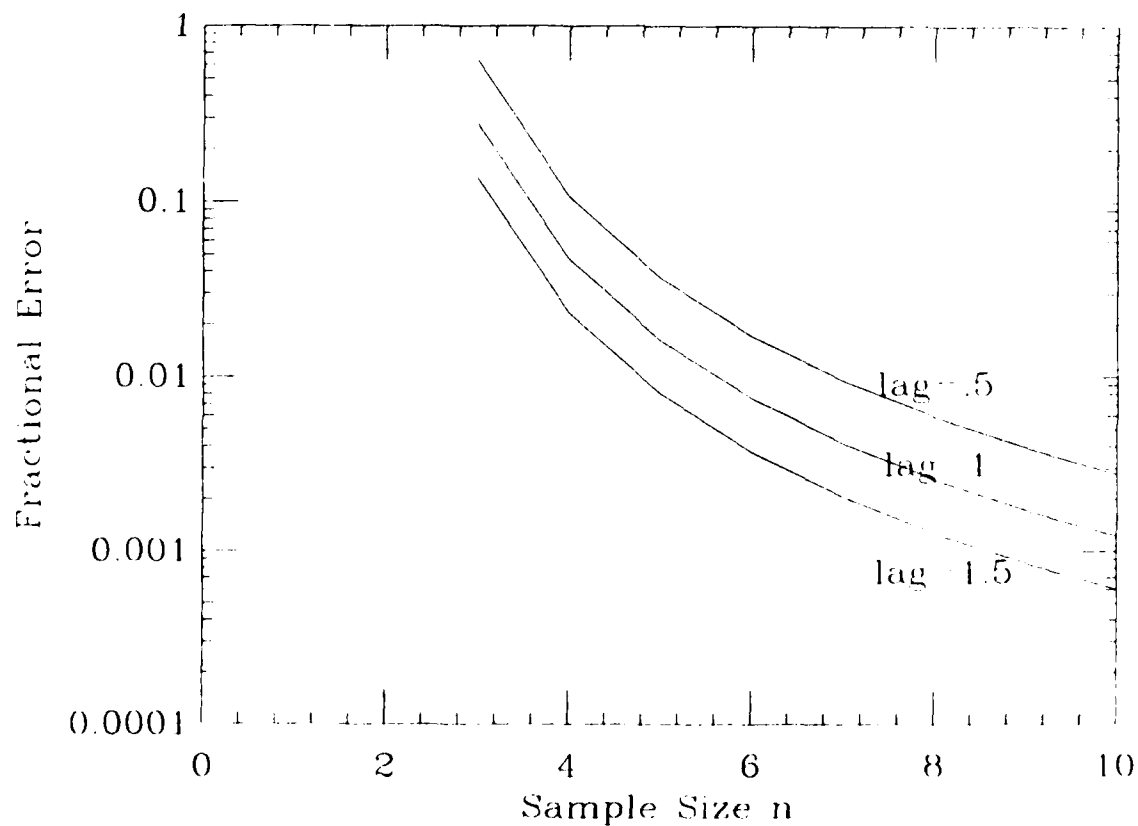


Figure 5. Case $k = 0$.
Nonoptimal coefficients, asymptotically optimal median sampling.

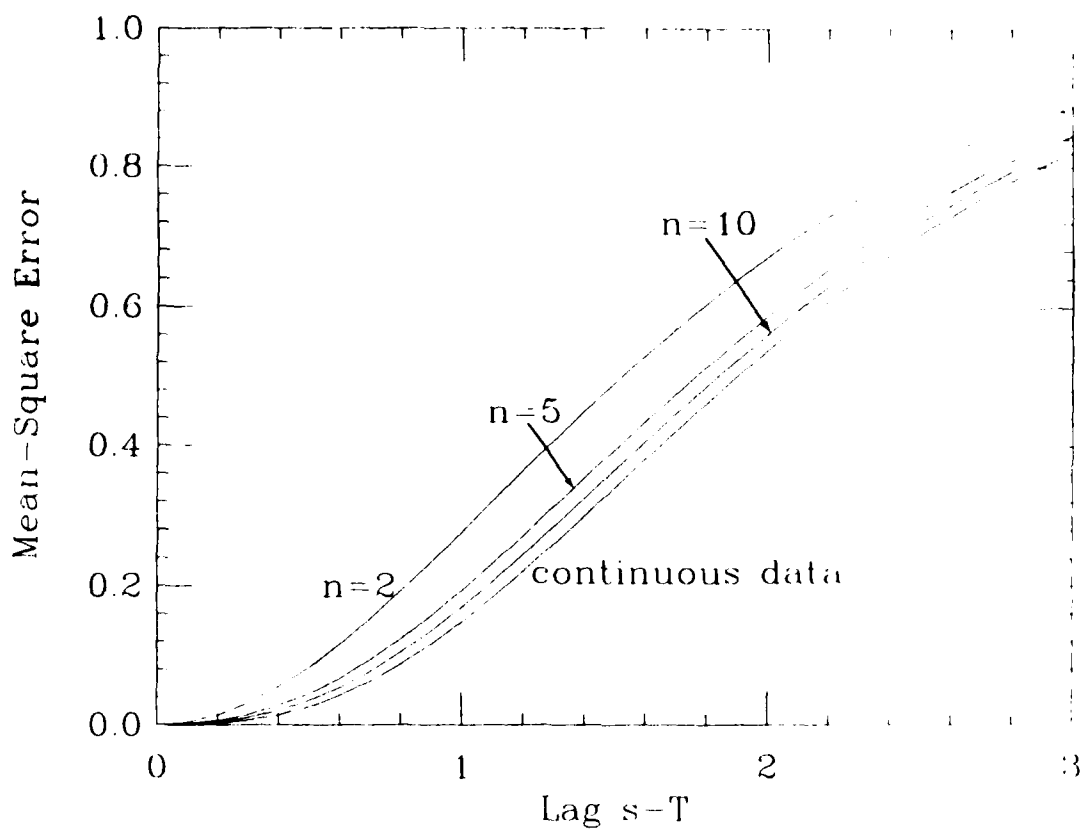


Figure 6. Case $k = 1$. Optimal coefficients, uniform sampling.

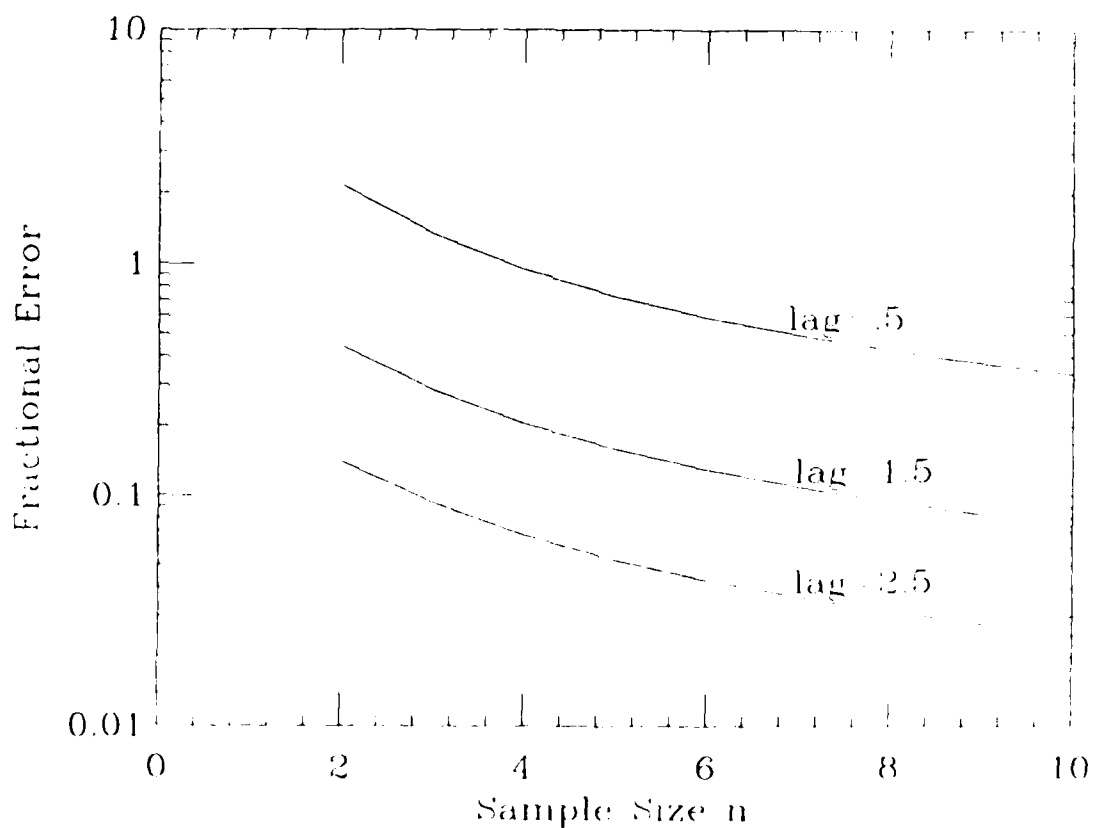


Figure 7. Case $k = 1$. Optimal coefficients, uniform sampling.

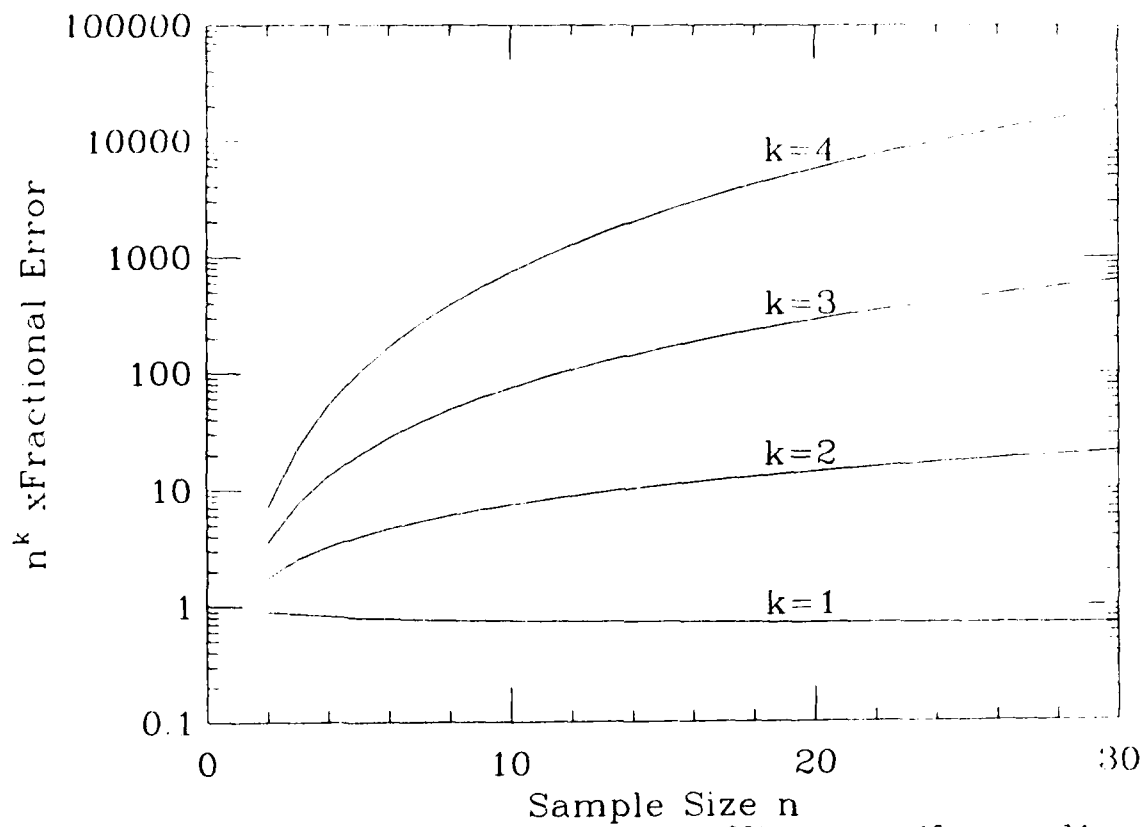


Figure 8. Case $k = 1$. Optimal coefficients, uniform sampling.

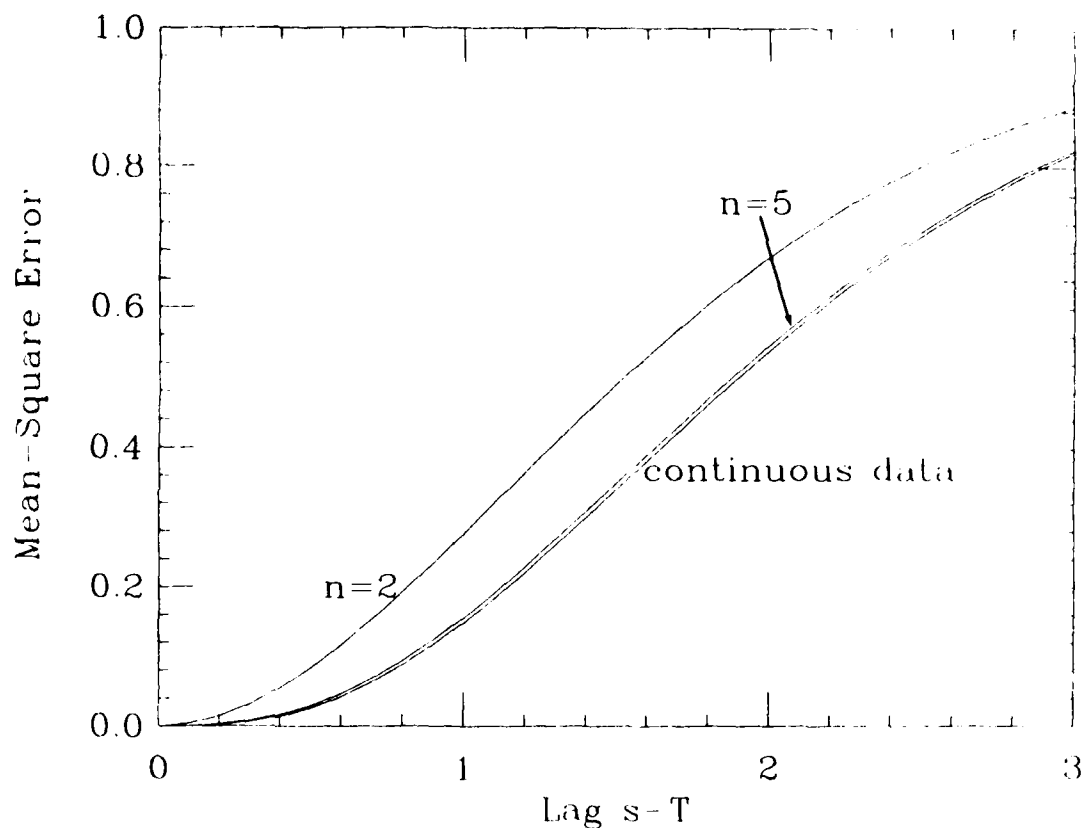


Figure 9. Case $k = 1$. Optimal coefficients, modified uniform sampling.

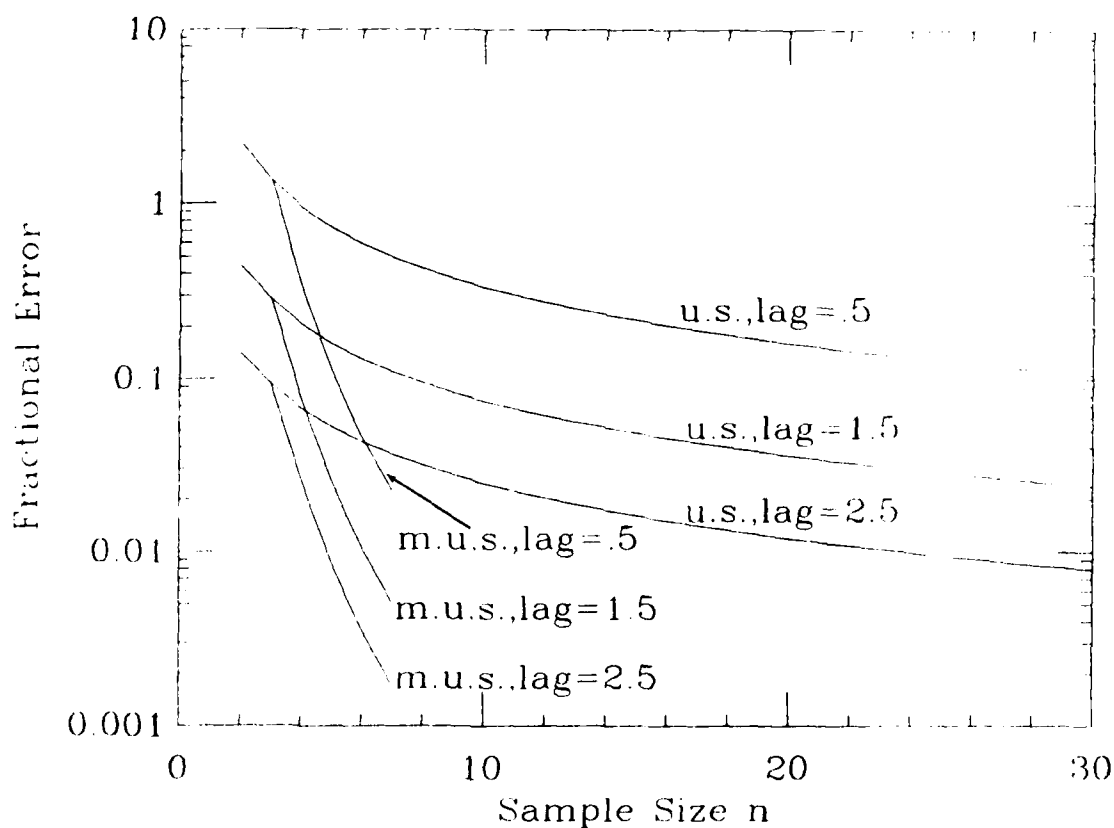


Figure 10. Case $k = 1$. Optimal coefficients.
Performance under uniform and modified uniform sampling.

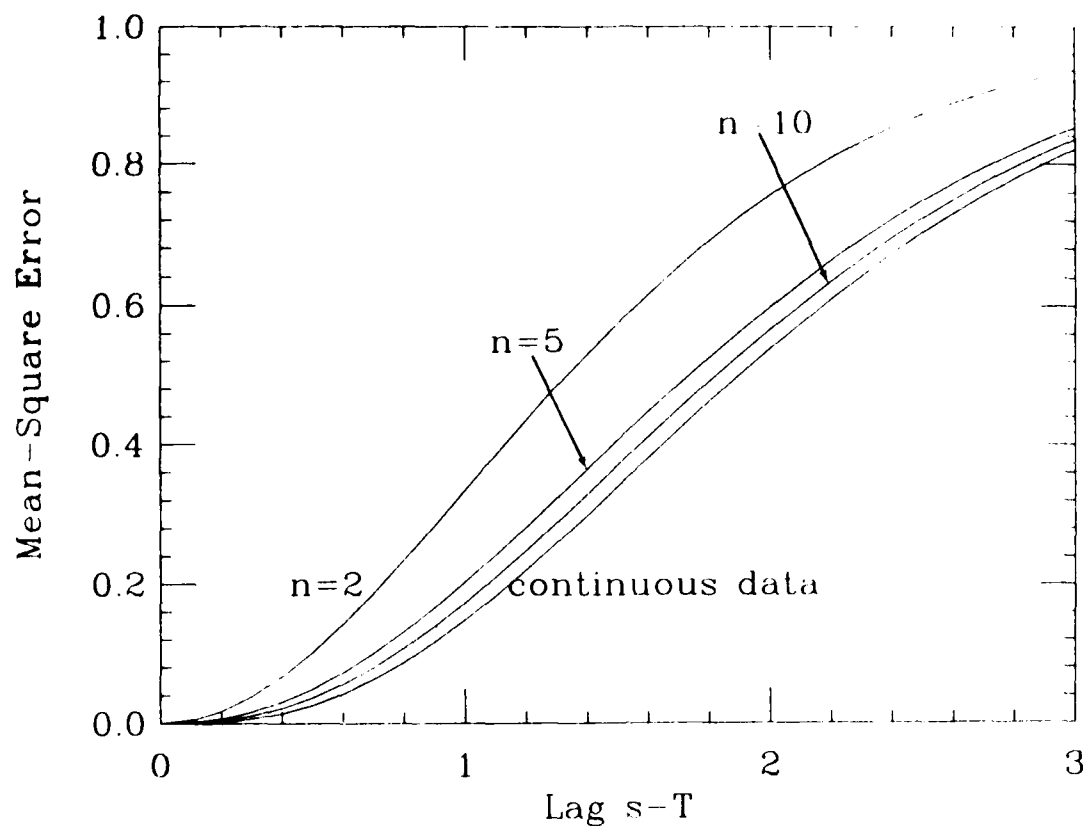


Figure 11. Case $k = 1$. Nonoptimal coefficients, uniform sampling.

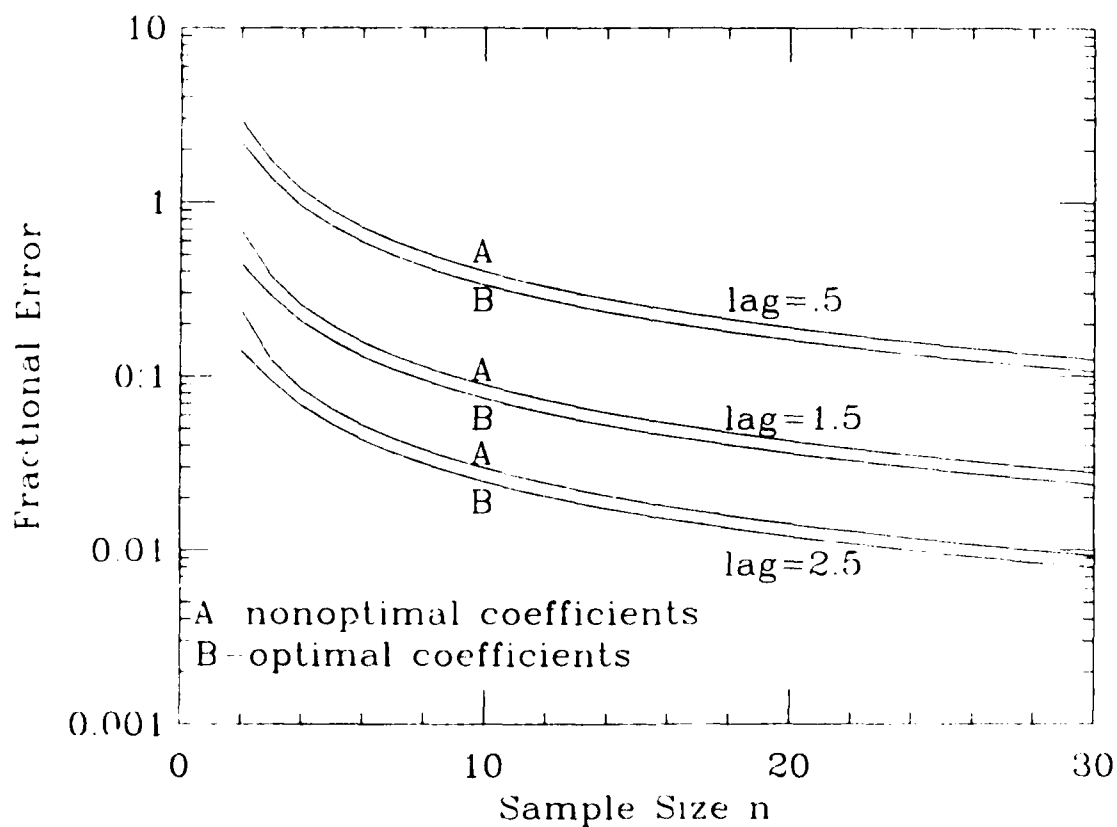


Figure 12. Case $k = 1$. Uniform sampling.
Performance of predictors with optimal and nonoptimal coefficients.

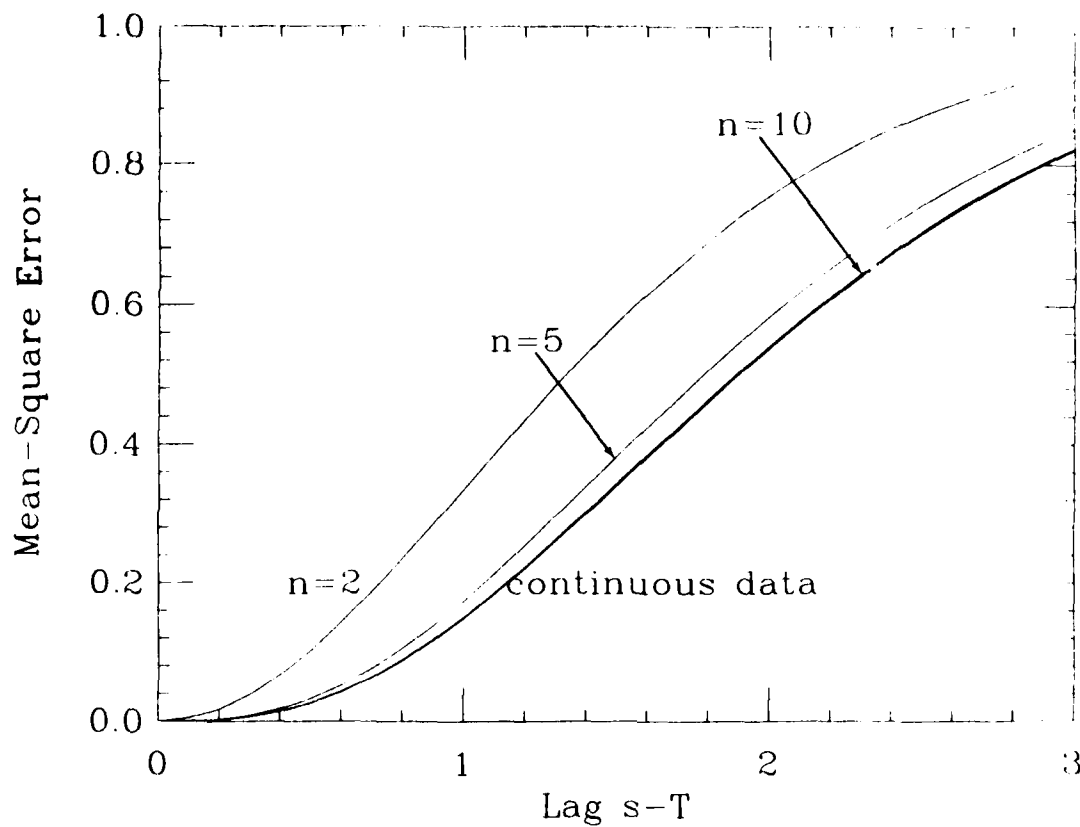


Figure 13. Case $k=1$. Nonoptimal coefficients, modified uniform sampling.

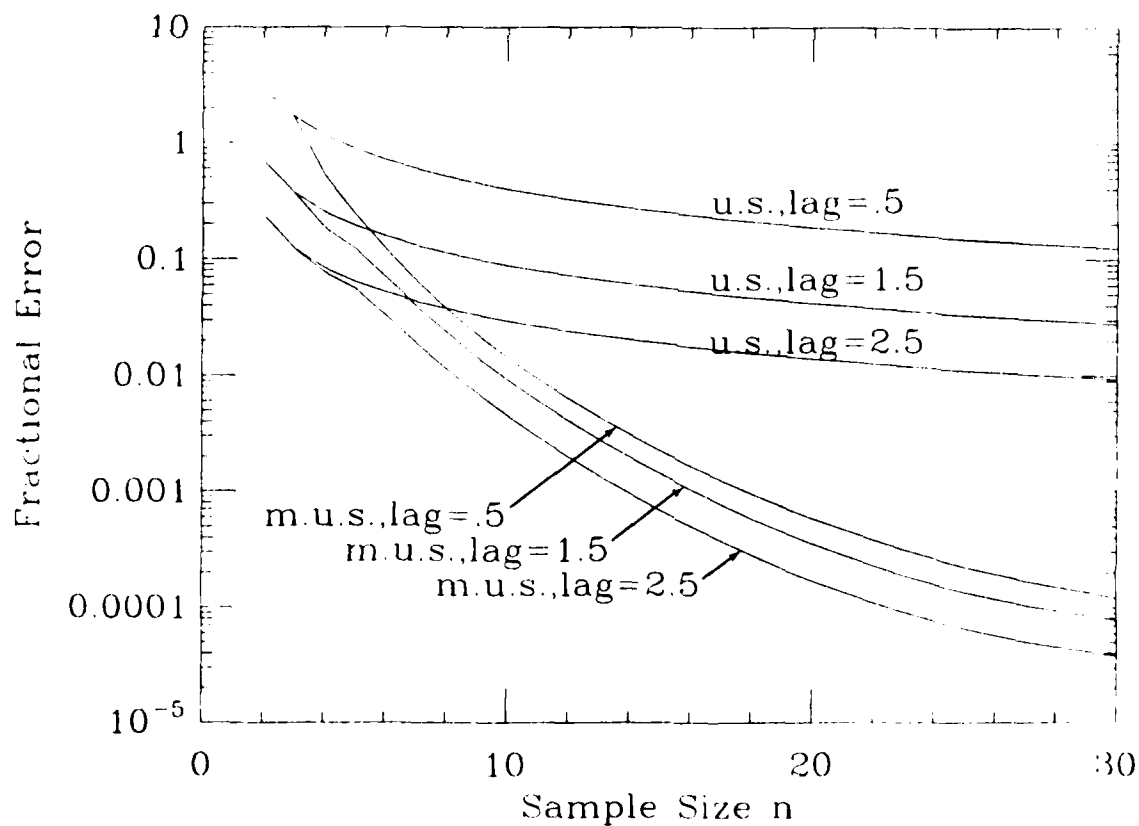


Figure 14. Case $k=1$. Nonoptimal coefficients. Performance under uniform and modified uniform sampling.

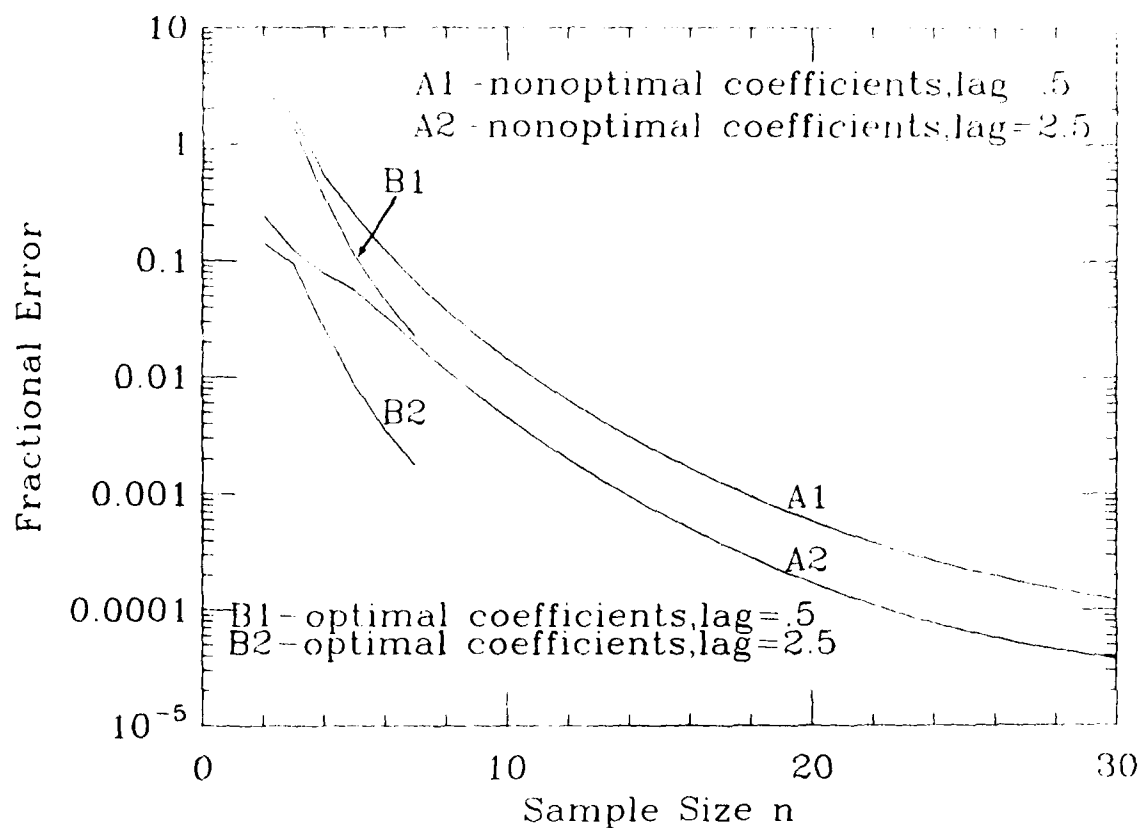


Figure 15. Case $k = 1$. Modified uniform sampling.
Performance of predictors with optimal and nonoptimal coefficients.

END
FILMED

5-86

DTIC