

APPROXIMATE BAYES LINEAR
SMOOTHERS FOR
CONTINUOUS PROCESSES

by

J. V. Zidek
and
S. Weerahandi

Technical Report #105

Department of Statistics, University of British Columbia

February, 1991

WS/PR:ddp

APPROXIMATE BAYES LINEAR SMOOTHERS FOR CONTINUOUS PROCESSES

by

J. V. Zidek and S. Weerahandi¹
The University of British Columbia and Bellcore

ABSTRACT

This paper presents two approximations for the problem of Bayes linear estimation of continuous vector-valued stochastic processes. The problem of interpolating time series is treated as a particular application. Expressions for errors of approximations are obtained. In particular, our general conditions for zero error lead to a characterization of a generalized AR(1) process.

Key words and Phrases. Linear smoothing, Bayes linear estimator, first order autoregressive process, Taylor series expansion.

syntax error file title, between lines 28 and 28

Prepared in part under support to SIMS from the United States Environmental Protection Agency and in part from a grant provided by the Natural Science and Engineering Research Council of Canada. The work of the first author was supported by an Irack Walton Killam Memorial Fellowship; it was completed while he was on leave at the University of Bath which generously provided facilities.

1. INTRODUCTION

In this paper we derive a modification of a Bayesian locally weighted regression method proposed by Weerahandi and Zidek (1988; hereafter WZ) and show that the result is approximately the same as the result which would be obtained by any one of a large family of Bayesian linear smoothing methods. These results are found in Section 3 where a continuous vector-valued process S is supposed to be observed with noise at n scalar valued sampling points to yield a data vector, Y in a n -fold cartesian product space whose components are finite dimensional inner product spaces. The statistical objective is taken to be either interpolation or extrapolation so that the object of inference is $\beta = S(\tau)$. For simplicity only interpolation will be considered here, the extension to extrapolation being essentially formalistic. Section 3 includes a precise expression for the error incurred in the heuristic practice (c.f. Muller, 1987) of basing interpolation on a data window enclosing τ . Requiring this error to be zero leads to a characterization of a generalized AR(1) process. Bounds are obtained in Section 3 for the error incurred in using the derived modification of the WZ smoother on just the data in a window at τ instead of a Bayes linear smoother from the general family alluded to above. Finally, in Section 3 it is shown how the procedure based on just the data in the window may be extended to the entire data set and a bound on the size of the resulting error is obtained. This has the practical advantage of overcoming the need to specify a window width.

Throughout this paper, *Bayes linear estimator* will always mean the best linear estimator with respect to quadratic loss. If the prior means of β and Y are specified and hence zero (as we may assume without loss of generality to simplify our analysis), then the Bayes linear estimator of β based on Y is

$$\beta_Y \triangleq \alpha_{\beta Y} Y, \quad (1.1)$$

where in general $\alpha_{UV} \triangleq \Gamma_{UV} \Gamma^{-1}$, the covariance between U and V , Γ_{UV} , is defined as the unique linear transformation for which $E(u, U)(v, U) = (u, \Gamma_{UV} v)$ for all matrices u, v having the dimensions of those of the random matrices U and V (cf. Eaton (1983, p 74)) and $\Gamma = \Gamma_{YY}$. Here and in the sequel, Γ_{YY} is assumed to have full rank and hence be invertible.

The Bayesian approach adopted here is more appealing to us than the frequentist approach on principle. As well it has the advantage of bringing in time series and Kriging quite naturally. Moreover, it provides a natural route towards the construction of credibility sets for β . In developing confidence bands even in the frequency theory of splines, it is the Bayesian highway which has been followed (Wahba (1983)). However, we recognize that the specification of the Γ 's may present practical problems of model development with the attendant risks of misspecifying the models. Our approximations simplify the task of modeling. We will evaluate the errors incurred in these

these approximations and establish some qualitative conditions which might suggest their adoption without determining the Γ 's completely.

Section 2 is a technical prelude to Section 3 in which we treat a more general problem than that described above. The generalization is that suggested by Sacks and Ylvisaker (1978). Suppose $Y_i: 1 \times q$, $i=1, \dots, n$ are observable response vectors for which

$$Y_i = \mu_i + N_i, \quad i=1, \dots, n, \quad (1.2)$$

where the $N_i: 1 \times q$ are unobservable, uncorrelated vectors of noise. Sacks and Ylvisaker (1978), for the case $q=1$, propose an approximately linear model,

$$\mu_i = A_i \beta + r_i, \quad i=1, \dots, n, \quad (1.3)$$

where $\beta: p \times q$ and the r_i are fixed but unspecified constants. Inference is about $\Lambda\beta$, where $\Lambda: x \times p$ is specified. Their estimators are of the form cY , where $Y = (Y_1^T, \dots, Y_n^T)^T$. In a frequentist setting they find the optimum c , which minimizes the mean squared error of estimation, in the case where $|r_i| < M_i$ with M_i specified and $x=1$. In the case $x>1$, they were only able to achieve some useful bounds. Very recently, Sacks, Welch, Mitchell and Wynn (1989) published a general review of the Sacks-Ylvisaker and related theory and the reader is directed to the bibliography given with that review.

Sacks and Ylvisaker (1978) note that even in the frequency setting, their proposed optimal linear estimators tends to rely most heavily on the Y_i 's for which M_i is small. There is an obvious Bayesian counterpart determined by the degree of association between β and the individual Y_i 's. In a sense to be made precise in Section 2, we may, after suitably permuting the subscript labels of the Y_i 's, partition Y as $Y = (P^T R^T)^T$ into those Y_i 's which are proximate to β (the rows of P) and those remote from β (the rows of R). A natural approximation to β_Y is then

$$\beta_P = \alpha_P P, \quad (1.4)$$

If the number of rows of Y retained in P is large (near n), this approximation will be very good. Surprisingly there are cases like that given in the application of Section 3 where $\beta_Y = \beta_P$ even when P consists of just two Y_i 's.

This approximation is of fundamental importance in statistics and is implicitly made when selecting the responses to be observed. These will be chosen heuristically on the basis of their relevance to the issue under study, in our case to the estimation of β . In intervention analysis, for example, responses in the spatial or temporal proximity of where the intervention is to occur will be made and, for budgetary and other reasons, more remote responses will not even be taken. Robustness is a second reason why remote observations might be excluded from the study. While these observations would be of limited value in estimating β , their inclusion invites the possibility of a detrimental impact on β_Y of a big response error. Another source of potential error is

in the additional modeling required to accommodate these remote responses and the resulting possibility of model misspecification.

The second approximation presented in Section 2 is more technical and embraces ideas of Weerahandi and Zidek (1988). To specify the Γ 's it is natural to explore the model in equations (1.2) and (1.3) which relate Y and β ,

$$Y = A\beta + E, \quad (1.5)$$

where $E = R + N$ and $x = (x_1^T, \dots, x_n^T)^T$ for $x = A, E, R, N$. If we assume, as we may without loss of generality, that β and E are uncorrelated, it follows that $A = \Gamma_{Y\beta} \Gamma_{\beta\beta}^{-1}$ and $C = \Gamma_{EE} - \Gamma_{Y\beta} \Gamma_{\beta\beta}^{-1} \Gamma_{\beta Y} = \Gamma_{Y\beta} \Gamma_{\beta\beta}^{-1} \Gamma_{\beta Y}$ (see Eaton (1983, p 88)). It may be possible in some cases like that of Section 3 to find simple approximations $A_0 \approx A = \Delta(A)$ and $C_0 = C = \Delta(C)$, without finding A and C explicitly, with some a priori assurance that $\Delta(A)$ and $\Delta(C)$ will be small. It then follows that approximately

$$\Gamma_{\beta Y} \approx \Gamma_{\beta\beta} A_0 \quad (1.6)$$

and

$$\Gamma_{YY} \approx A_0 \Gamma_{\beta\beta} A_0^T + C_0. \quad (1.7)$$

Here $\Gamma_{\beta\beta}$ is the a priori covariance matrix of β . Some components of $\Gamma_{\beta\beta}$ may be allowed to approach infinity when our prior knowledge is vague. Additional levels in a hierarchical prior model may need to be added to incorporate uncertainty about elements of β as in Weerahandi and Zidek (1986, 1990). We will not address the problem of specifying $\Gamma_{\beta\beta}$ in general but will do so in the special case considered in Section 3.

The approximation in Section 3 combines those described above. We consider a continuous multivariate time series and $Y_i = Y(t_i) = (Y_1(t_i), \dots, Y_q(t_i))$ $i = 1, \dots, n$, where the t_i 's are not necessarily equally spaced. And $\beta = Y(\tau) = (Y_1(\tau), \dots, Y_q(\tau))$. So as long as the autocorrelation between $Y(t)$ and $Y(\tau) = \beta$ is decreasing as $|t - \tau|$ increases, P may be taken to be those Y 's for which $|t_i - \tau|$ is small. This gives rise to the first approximation while the second is obtained by a first order Taylor expansion of the $Y(t_i)$ around $t_i = \tau$. Details are given in Section 3. Our method and its limitations are discussed in Section 4. To complete the paper a small scale simulation study is given in Section 5.

2. APPROXIMATELY BAYES LINEAR ESTIMATION

Preliminaries. Our approximation theory will be developed in a general context where $\langle \beta, Y \rangle \in B \times Y$ and B and Y are real, finite dimensional inner product spaces with inner products, $\langle \cdot, \cdot \rangle_B$ and $\langle \cdot, \cdot \rangle_Y$, respectively. Thus $B \times Y$ is a real, finite

dimensional inner product space with inner product $(\cdot, \cdot)_R + (\cdot, \cdot)_Y$. Much of the theory underlying the developments of this section is presented by Eaton (1983) and by Stone (1987) to which the interested reader is referred for more detail. A brief sketch will now be presented for completeness.

In general, let $L(W, U)$ denote the space of all linear transformations of W into U where $(W, (\cdot, \cdot)_W)$ and $(U, (\cdot, \cdot)_U)$ are finite dimensional inner product spaces. If $\{w_i\}$ and $\{u_j\}$, respectively, are orthonormal bases for W and U , then $\{w_i \otimes u_j\}$ is an orthonormal basis for $L(W, U)$ where, in general, $w \otimes u$ denotes the exterior product of w and u defined by $(w \otimes u)v = (w, v)_W u \in U$ for every $w, v \in W$ and $u \in U$. Thus any $A \in L(W, U)$ has a unique matrix representation, $A = \sum \sum a_{ij} w_i \otimes u_j$. Furthermore, W and U induce an inner product, $\langle \cdot, \cdot \rangle_{WU}$, on $L(W, U)$ given by $\langle A, B \rangle_{WU} = \sum \sum a_{ij} b_{ij}$ when A and B are given by their matrix representation. In other words, $\langle A, B \rangle_{WU} = \text{tr}[A][B]^T$ where $[A]$ and $[B]$ are the matrices representing A and B , respectively. This determines a norm on $L(W, U)$: $\|A\|_{WU} = \langle A, A \rangle_{WU}^{1/2}$.

In this paper a different norm will be used because of our interest in quantities like the magnitude of ϵ defined below in equation (2.1) relative to $\|R\|_R$. The norm is given by $\|A\|_{WU} = \text{Sup} \left\{ \|Aw\|_U : \|w\|_W \leq 1 \right\}$ and it determines the uniform operator topology for $L(W, U)$ (cf. Taylor and Lay 1980, p 189).

If $W=U$ and $T \in L(W, W)$ is self adjoint, there exists an orthonormal basis $\{w_i\}$ for W such that $T = \sum \lambda_i (w_i \otimes w_i)$ for real scalars, λ_i , called the eigenvalues of T . Then $T^{-1} = \sum \lambda_i^{-1} (w_i \otimes w_i)$ defines the inverse of T if $\lambda_i \neq 0$ for all i . And if T is non-negative definite, i.e., $\lambda_i \geq 0$ for all i , $T^{1/2} = \sum \lambda_i^{1/2} (w_i \otimes w_i)$. These facts imply that if $A \in L(W, U)$, $\|A\|_{WU}$ is the square root of the largest eigenvalue of the self adjoint, nonnegative definite transformation $A^T A \in L(W, W)$ (cf. Eaton, 1983, exercise 29, p 67). Recall that $\langle A, A \rangle_{WW} = \text{tr}[A]^T[A]$ so $\|A\|_{WW}$ is the square root of the sum of the eigenvalues of $A^T A$. Halmos (1974, p178) shows that for every $A \in L(W, W)$, $\|A\|_{WW} = \text{Sup} \{ \|Ax\|_W : \|x\|_W \leq 1 \}$, a useful result since in this paper, A is often a covariance matrix or the difference of a pair of covariance matrices. Halmos (1974, pp. 182-183) also shows that $\|A_n - A\|_{WW} \rightarrow 0$ as $n \rightarrow \infty$ if and only if $\|(y A_n x)_W - (y A x)_W\| \rightarrow 0$ for all x and y . We note finally that $\|A^T\|_{WW} = \|A\|_{WW}$ (ibid, p 179).

In the sequel Y will be partitioned as $Y=P \times R$ for certain linear subspaces, P and R , with associated inner products, $(\cdot, \cdot)_P$ and $(\cdot, \cdot)_R$, respectively, and then we can write $y = \langle p, r \rangle$ for every $y \in Y$ where $p \in P$ and $r \in R$. Also $Y = \langle P, R \rangle$ where P and R correspond to the quantities introduced in Section 1. Finally we assume $(\cdot, \cdot)_Y = (\cdot, \cdot)_P + (\cdot, \cdot)_R$. Partition Γ_{YY} as

$$\Gamma_{YY} = \begin{bmatrix} \Gamma_{PP} & \Gamma_{PR} \\ \Gamma_{RP} & \Gamma_{RR} \end{bmatrix}$$

(see Eaton, 1983, p 87).

The First Approximation. The approximation is obtained by discarding the data subvector, R , with resulting inferential error

$$\varepsilon = \beta_Y - \beta_P.$$

where β_Y and β_P are as defined in equations (1.1) and (1.4), respectively.

THEOREM 1. Let $\alpha_{\beta R, P} = \Gamma_{\beta R, P} \Gamma_{R, P}^{-1}$ and $\Gamma_{\beta R, P} = \Gamma_{\beta R} - \Gamma_{\beta R} \Gamma_{P, P}^{-1} \Gamma_{R, P}$. Then

$$\varepsilon = \alpha_{\beta R, P} R, \quad (2.1)$$

$$\Gamma_{\beta, Y} - \Gamma_{\beta, P} = -\alpha_{\beta R, P} \Gamma_{R, P}, \quad (2.2)$$

$$\alpha_{\beta R, P} = D \alpha^* (\Gamma_{R, P} \beta + \alpha^* D \alpha^{*T})^{-1}, \quad (2.3)$$

where

$$\alpha^* = \alpha_{R, P} \alpha_{P, \beta} - \alpha_{R, \beta},$$

$$\alpha_{R, P} = \Gamma_{R, P} \Gamma_{P, P}^{-1}$$

$$\Gamma_{R, P} = \Gamma_{R, \beta} - \Gamma_{R, P} \Gamma_{P, \beta}^{-1} \Gamma_{R, \beta}.$$

and

$$D = [\Gamma_{\beta\beta}^{-1} + \alpha_{P, \beta}^{-1} \Gamma_{P, \beta}^{-1} \alpha_{P, \beta}]^{-1}.$$

PROOF. The Bartlett decomposition of Γ_{YY} is

$$\Gamma_{YY} = \begin{bmatrix} I & 0 \\ \eta & I \end{bmatrix} \begin{bmatrix} \Gamma_{PP} & 0 \\ 0 & \Gamma_{R, P} \end{bmatrix} \begin{bmatrix} I & \eta^T \\ 0 & I \end{bmatrix},$$

where $\eta = \Gamma_{R, P} \Gamma_{P, P}^{-1}$ and I denotes the identity transformation. This is easily verified by successive applications of the identity which defines the partitions of transformations like that of Γ_{YY} , introduced above.

$$Ay \triangleq \begin{bmatrix} A_{PP} & A_{PR} \\ A_{RP} & A_{RR} \end{bmatrix} \begin{bmatrix} p \\ r \end{bmatrix} \triangleq \begin{bmatrix} A_{PP} p + A_{PR} r \\ A_{RP} p + A_{RR} r \end{bmatrix}. \quad (2.4)$$

It is easily verified, by repeated use of (2.4), that

$$\Gamma_{YY}^{-1} = \begin{bmatrix} I & -\eta^T \\ 0 & I \end{bmatrix} \begin{bmatrix} \Gamma_{PP}^{-1} & 0 \\ 0 & \Gamma_{R, P}^{-1} \end{bmatrix} \begin{bmatrix} I & \eta \\ 0 & I \end{bmatrix}.$$

Now partition $\Gamma_{\beta Y}$ as $(\Gamma_{\beta P}, \Gamma_{\beta R})$, defined in an obvious way by the it operates on y . Then it is easily verified that

$$\beta_Y = \beta_P + \alpha_{\beta R, P} R$$

which proves the result in equation (2.1).

Now $\Gamma_{\beta Y} \stackrel{\Delta}{=} \Gamma_{\beta\beta} - \Gamma_{\beta Y} \Gamma_Y^{-1} \Gamma_{Y\beta}$. Again using Bartlett's decomposition and the rules of operation of the various partitioned transformations we obtain

$$\alpha_{\beta Y} \Gamma_Y^{-1} = \alpha_{\beta P} \Gamma_P^{-1} + \alpha_{\beta R, P} \Gamma_{R, P}^{-1}$$

which proves the assertion in equation (2.2).

To obtain equation (2.3), we begin with

$$\Gamma_{RR} = \Gamma_{R, \beta} + \alpha_{R\beta} \Gamma_{\beta\beta} \alpha_{R\beta}^T,$$

$$\Gamma_{RP} = \Gamma_{RP, \beta} + \alpha_{R\beta} \Gamma_{\beta\beta} \alpha_{P\beta}^T,$$

and

$$\Gamma_{PP} = \Gamma_{P, \beta} + \alpha_{P\beta} \Gamma_{\beta\beta} \alpha_{P\beta}^T,$$

all of which are true by definition. Thus

$$\Gamma_P^{-1} = \Gamma_{P, \beta}^{-1} - \Gamma_{P, \beta}^{-1} \alpha_{P\beta} \Gamma_{\beta\beta}^{-1} \alpha_{P\beta}^T [\Gamma_{\beta\beta}^{-1} + \alpha_{\beta\beta}^T \Gamma_{P, \beta}^{-1} \alpha_{P\beta}]^{-1} \alpha_{\beta\beta}^T \Gamma_{P, \beta}^{-1}$$

whose proof is formally identical to that of Lindley and Smith (1972) for the corresponding matrix identity. It follows that

$$\Gamma_{\beta\beta} \alpha_{\beta\beta}^T \Gamma_P^{-1} = D \alpha_{\beta\beta}^T \Gamma_{P, \beta}^{-1}$$

and

$$\Gamma_P^{-1} \alpha_{P\beta} \Gamma_{\beta\beta} = \Gamma_{P, \beta}^{-1} \alpha_{P\beta} D.$$

These results may be used to simplify

$$\begin{aligned} \Gamma_{R, P} &= \Gamma_{R, \beta} - \Gamma_{RP, \beta} \Gamma_P^{-1} \Gamma_{RP, \beta}^T - \Gamma_{RP, \beta} \Gamma_P^{-1} \alpha_{P\beta} \Gamma_{\beta\beta} \alpha_{R\beta}^T - \alpha_{R\beta} \Gamma_{\beta\beta} \alpha_{\beta\beta}^T \Gamma_P^{-1} \Gamma_{RP, \beta}^T \\ &\quad + \alpha_{R\beta} [\Gamma_{\beta\beta} - \Gamma_{\beta\beta} \alpha_{\beta\beta}^T \Gamma_P^{-1} \alpha_{P\beta} \Gamma_{\beta\beta}] \alpha_{R\beta}^T \end{aligned}$$

to

$$\Gamma_{R, P} = \Gamma_{R, P\beta} + \alpha^* D \alpha^{*T}.$$

Similarly,

$$\Gamma_{\beta R, P} = \Gamma_{\beta\beta} \alpha_{R\beta}^T - \Gamma_{\beta\beta} \alpha_{\beta\beta}^T \Gamma_P^{-1} \Gamma_{RP}^T$$

so

$$\Gamma_{\beta R, P} = D \alpha^*.$$

Equation (2.3) now follows. \square

Recall that $\Gamma_{\beta Y}$ is the covariance transformation of $\beta - \beta_Y$ and so it represents the residual uncertainty in β after predicting it from Y . If β_P were used in place of β_Y , the residual uncertainty in β would increase, or at least not decrease, since P carries

less information about β than Y . Equation (2.2) shows the resulting difference in the uncertainty about β , namely $\alpha_{\beta R \cdot P} \Gamma_{\beta \cdot P}$.

Now observe that $\Gamma_{\beta R \cdot P}$ represents the covariance between β and $(R - \alpha_{RP}P) \stackrel{\Delta}{=} R \cdot P$. This is because the covariance transformation representing the covariance between β and $R \cdot P$ is the unique transformation for which

$$(b, \Gamma_{\beta(R \cdot P)})_B = E(b, \beta)_B (r, R \cdot P)_R$$

for all b and r . But

$$E(b, \beta)_B (r, R \cdot P)_R = E(b, \beta)_B (r, R)_R - E(b, \beta)_B (r, \alpha_{RP}P)_R$$

and

$$\begin{aligned} E(b, \beta)_B (r, \alpha_{RP}P)_R &= E(b, \beta)_B (\alpha_{RP}^T r, P)_P \\ &= (b, \Gamma_{\beta P} \alpha_{RP}^T r)_B \end{aligned}$$

for all b and r . From this, the conclusion follows: $\Gamma_{\beta(R \cdot P)} = \Gamma_{\beta R \cdot P}$. So we see that the magnitude of ϵ depends on the degree of association between β and the R residuals after (linearly) fitting P . In fact $\epsilon=0$ is equivalent to $\Gamma_{\beta R \cdot P} = 0$. This leads to the next theorem which gives intuitive conditions under which $\epsilon=0$. By " β and R are conditionally uncorrelated given P " we will mean $Cov[(b, \beta)_B, (r, R)_R | P=p] = 0$ for all b, p , and r .

THEOREM 2. If (i), Y is normally distributed and (ii), β and R are conditionally independent given P , then $\beta_Y = \beta_P$, i.e. $\epsilon=0$.

PROOF. The conclusion holds if and only if $\Gamma_{\beta R \cdot P} = 0$ and this is true if and only if $E(b, \beta)_B (r, R \cdot P)_R = 0$ as pointed out in the preceding discussion. But this last quantity is just

$$\begin{aligned} EE[(b, \beta)_B (r, R - \alpha_{RP}P)_R | P] &= E[E[(b, \beta)_B | P] E[(r, R)_R | P] - \\ &\quad E[(b, \beta)_B | P] (r, \alpha_{RP}P)_R]. \end{aligned}$$

since β and R are conditionally uncorrelated given P . Now by definition $E[(r, R)_R | P] = (r, E(R|P))_R$ and, since Y is normally distributed, $E(R|P) = \alpha_{RP}P$ (cf. Eaton 1983, p116). The conclusion of the theorem is an immediate consequence of this observation. \square

Other conditions under which $\epsilon=0$ are given in the more specialized situation of Section 3. The conclusions of Theorem 2 and its relative in Section 3 are excessively strong in that $\epsilon=0$ whereas β_P would be considered a satisfactory approximation as long as ϵ were merely small in some sense. But unless $\Gamma_{\beta R \cdot P} = 0$, i.e. $\epsilon=0$, ϵ could be arbitrarily large depending on R . A more natural measure of the quality of the

approximation would therefore seem to be the magnitude of ϵ relative to that of R , $||R||_R$, and this leads to $|\alpha_{\beta R, P}|_{RR}$ as an index of the quality of the approximation.

Of course unless $\alpha_{\beta R, P}$ were completely specified and R observed, in which case it would be pointless to seek an approximation to β_Y , neither ϵ nor the index defined in the last paragraph of its quality could be evaluated. Theorem 1 is intended to characterize precisely the errors which result from ignoring R in predicting β and it is applied in the next section to determine when they would be negligible. By providing additional insight into the nature of these approximation errors, Theorems 1, 2, and their relative in Section 3 may help to refine the heuristics involved in deciding how to partition Y effectively and when to drop R .

Since in some cases, R will not even be observed when P is, it seems desirable to have some measure of β_P 's performance which exploits the information in P . This leads us to the following result.

THEOREM 3. The Bayesian linear predictor of ϵ based on P is

$$e_P = \alpha_{\beta R, P} \alpha_{R P} P$$

and the covariance of $\epsilon - e_P$ is

$$\Gamma_{\epsilon P} = \alpha_{\beta R, P} \Gamma_{R \beta, P} \quad (2.5)$$

PROOF. This is straightforward and omitted for brevity. \square

The results in equations (2.2) and (2.5) may be combined to give an insightful re-expression of equation (2.2):

$$\Gamma_{\beta, P} = \Gamma_{\beta, Y} + \Gamma_{\epsilon, P}.$$

So the change in residual uncertainty in predicting β which derives from dropping R is precisely that in ϵ after its linear prediction from P . In other words, β_P will be a good approximation to β_Y exactly when ϵ is well explained by P .

The Second Approximation.

A useful expression for $\alpha_{\beta\gamma}$ is now derived, starting with (1.1) and invoking model (1.5).

LEMMA 1. If E and β in equation (1.5) are uncorrelated, then

$$\alpha_{\beta\gamma} = [\Gamma_{\beta\beta}^{-1} + A^T C^{-1} A]^{-1} A^T C^{-1}, \quad (2.6)$$

where $C = \Gamma_{EE}$, and

$$\Gamma_{\beta\gamma} = (\Gamma_{\beta\beta}^{-1} + A^T C^{-1} A)^{-1}. \quad (2.7)$$

PROOF. The proof consists in showing that the linear transformation given in equation (2.6) is identical to that which defines $\alpha_{\beta\gamma}$:

$$\alpha_{\beta\gamma} = \Gamma_{\beta\beta} A^T [A \Gamma_{\beta\beta} A^T + C]^{-1},$$

since, by the definition of C , $\Gamma_{\gamma\gamma} = A \Gamma_{\beta\beta} A^T + C$. This is done using the transformation identity,

$$\Gamma_{\gamma\gamma}^{-1} = C^{-1} - C^{-1} A [\Gamma_{\beta\beta}^{-1} + A^T C^{-1} A]^{-1} A^T C^{-1}, \quad (2.8)$$

which is based on the same identity as that used to prove equation (2.3) in Theorem 1. Equation (2.7) is an immediate consequence of this last result. \square

Approximations to the model and inferential quantities are to be obtained by replacing A , C and $\Gamma_{\beta\beta}$ by A_0 , C_0 and $\Gamma_{\beta\beta}^0$, respectively, where C_0 and $\Gamma_{\beta\beta}^0$ are self-adjoint, positive definite transformations. This will induce approximation errors in the various objects of analysis. We will everywhere in the sequel use "0" as a sub- or superscript to indicate an induced approximant so, for example, β_γ^0 will denote that of β_γ . Approximation errors in objects, H , will be represented by

$$\Delta(H) \triangleq H - H_0.$$

It follows that

$$\Delta(\Gamma_{\beta\gamma}^{-1}) = \Delta(\Gamma_{\beta\beta}^{-1}) + \Delta(A^T C^{-1} A).$$

The next lemma is useful for extracting the first order terms in the approximation errors considered in this report.

LEMMA 2 (a). Suppose that G and H are invertible transformations with associated strong operator norm, $\|\cdot\|$, that G is invertible, and that $\|HG^{-1}\| \leq 1$. Then $I + HG^{-1}$ is invertible, and

$$(G + H)^{-1} = G^{-1} \sum_{r=0}^{\infty} [HG^{-1}]^r,$$

and

$$\|(G + H)^{-1}\| \leq \|G^{-1}\| \delta_{HG^{-1}},$$

where, here and in the sequel,

$$\delta_x = x(1-x)^{-1}, 0 \leq x \leq 1.$$

(b). Suppose H_0 is an approximant to H , both H_0 and H being invertible with $\|\Delta(H^{-1})H_0\| < 1$, where, here and in the sequel,

$$\Delta(G) = G - G_0$$

for any such associated pair, G_0 and G whether or not they are invertible. Then $I - \Delta(H^{-1})H_0$ is invertible,

$$\Delta(H) = H_0 \sum_{r=0}^{\infty} [-\Delta(H^{-1})H_0]^r,$$

and

$$\|\Delta(H)\| \leq \|H_0\| (1 + \delta_{\|\Delta(H^{-1})H_0\|}).$$

PROOF: Part (a) is an immediate consequence of a standard result (cf. Taylor and Lay, 1980, Theorem 1.4, p 192) and part (b) follows from (a). \square

Note that if D is a self adjoint and positive definite transformation, $\|D\|$ is the largest eigenvalue of A (see Eaton, 1983, p 54). This fact may be helpful in verifying that $\|\Delta(H^{-1})H_0\| < 1$ in some cases.

This next result is an immediate consequence of Lemma 2(b) and it is stated without proof.

LEMMA 3. If $\|\Delta(\Gamma_{\beta}^{-1})\Gamma_{\beta, Y}\|_{BB} < 1$,

$$\Gamma_{\beta, Y}^{-1} \Delta(\Gamma_{\beta, Y}) = - \sum_{r=1}^{\infty} [\Delta(\Gamma_{\beta}^{-1})\Gamma_{\beta, Y}]^r.$$

and

$$\Gamma_{\beta, Y}^{-1} \Delta(\beta_Y) = [\Gamma_{\beta, Y}^{-1} \Delta(\Gamma_{\beta, Y}) A_0^T C_0^{-1} + \Delta(A^T C^{-1})] Y. \quad (1)$$

The leading terms in the expansions of Lemma 3 may be used in a sensitivity analysis. Of more interest than the expansions perhaps, are bounds on the magnitude of the approximation errors. In any case, it is necessary to determine how errors made

in approximating the building blocks of the model in equation (1.5) affect inferential procedures. The proof of the following lemma is straightforward and omitted.

LEMMA 4.

$$\Delta(A^T C^{-1}) = A_0^T \Delta(C^{-1}) + \Delta^T(A) C^{-1} ,$$

and

$$\Delta(A^T C^{-1} A) = A_0^T \Delta(C^{-1}) A + \Delta^T(A) C^{-1} A + A_0^T C_0^{-1} \Delta(A) .$$

Moreover, if $\|\Delta(C)C_0^{-1}\|_{YT} \leq 1$,

$$\Delta(C^{-1}) = C_0^{-1} \sum_{r=1}^{\infty} [-\Delta(C)C_0^{-1}]^r .$$

The sensitivities to moderate approximation errors of the objects of central interest in this paper are indicated by the results of the next theorem.

THEOREM 4. To first order, the following approximate equations obtain:

$$\Gamma_{\beta\gamma}^{-1} \Delta[\Gamma_{\beta\gamma}] = -\rho$$

and

$$\Gamma_{\beta\gamma}^{-1} \Delta[\beta_\gamma] = \rho A^T C^{-1} + \Delta[A^T C^{-1}] ,$$

where

$$\begin{aligned} \rho &= \Delta[\Gamma_{\beta\gamma}^{-1}] \Gamma_{\beta\gamma} , \\ \Delta[\Gamma_{\beta\gamma}^{-1}] &= \Delta[\Gamma_{\beta\beta}^{-1}] + \Delta[A^T C^{-1} A] \end{aligned}$$

and to first order,

$$\Delta[A^T C^{-1}] = \Delta^T[A] C^{-1} - A^T C^{-1} \Delta[C] C^{-1}$$

while

$$\Delta(A^T C^{-1} A) = \Delta^T[A] C^{-1} A - A^T C^{-1} \Delta[C] C^{-1} A + A^T C^{-1} \Delta[A] .$$

PROOF. These results are obtained directly from Lemma 4. \square

We would note in passing that errors made in approximating the data model, that is those in A and C enter the analysis only through $A^T C^{-1}$ and $A^T C^{-1} A$. Moreover, to first order they are given exactly the same weight as those entering the prior model through $\Gamma_{\beta\beta}$.

The following notation will be helpful in developing bounds on the errors of approximation:

$$\begin{aligned} e_{\Gamma_{\beta Y}} &= |\Delta(\Gamma_{\beta Y}^{-1})|_{BB} |\Gamma_{\beta Y}^0|_{BB} \\ e_C &= |\Delta(C)|_{YY} |C_0^{-1}|_{YY}, \\ e_A &= |\Delta(A)|_{BY} / |A_0|_{BY}. \end{aligned}$$

THEOREM 5. Suppose $\varepsilon < 1$ for $\varepsilon = e_C, e_\Gamma$. Then

$$\begin{aligned} |\Delta(\Gamma_{\beta Y})|_{BB} &\leq |\Gamma_{\beta Y}^{-1}|_{BB} \delta_{e_{\Gamma_{\beta Y}}}, \\ \frac{||\Delta(\beta_Y)||_B}{||Y||_Y} \frac{|\Gamma_{\beta Y}|_{BB} |A_0|_{YB} |C_0^{-1}|_{YY}}{\delta_{e_{\Gamma_{\beta Y}}} + e_A + e_C + e_A e_C} & \quad (2.9) \\ e_{\Gamma_{\beta Y}} &\leq |\Delta[\Gamma_{\beta Y}^{-1}]|_{BB} + |\Delta[A^T C^{-1} A]|_{BB} |\Gamma_{\beta Y}|_{BB} \end{aligned}$$

and

$$|\Delta[A^T C^{-1} A]|_{BB} \leq |A_0|_{YB}^2 |C_0^{-1}|_{YY} (\delta_{e_C} (1 + e_A) + e_A (1 + e_A) (1 + \delta_{e_C}) + e_A).$$

PROOF. Recall that in general, for any transformations, T_1 and T_2 , $|T_1 + T_2| \leq |T_1| + |T_2|$, $|T_1 T_2| \leq |T_1| |T_2|$, and $|T_i^T| = |T_i|$. Thus the conclusions follow immediately from Lemmas 2, 3, and 4. ||

We should remark that the denominator on the left hand side of equation (2.9) is a surrogate for the more natural choice $||\beta_Y||_B$ which could not be used because of insurmountable technical difficulties. The approximants A_0 and C_0 can be replaced by A and C , respectively, after inserting an appropriate multiplicative factor. But there is little difference between these alternative forms for the left hand side of that equation and our choice is simpler for technical reasons.

For the application of the next section, the force of this last result lies in its implication that the appropriately normalized versions of the approximation errors $|\Delta[\Gamma_{\beta Y}]$ and $|\Delta(\beta_Y)|$ approach zero as the various ε 's approach zero. The bounds enable rates of convergence to be established as well.

3. INTERPOLATING CONTINUOUS PROCESSES

Let $\{S(t) : t \in T\}$, be a stochastic process on T , a subinterval of $(-\infty, \infty)$, which is to be observed with additive, mean zero noise at not necessarily equi-spaced sampling points, $t = t_1, \dots, t_n$. Thus $Y = \langle Y(t_1), \dots, Y(t_n) \rangle$ where $Y(t_i) = S(t_i) + N(t_i)$ for all i . For each t , $S(t), N(t) \in S$, a finite dimensional inner product space with inner product (\cdot, \cdot) . It is supposed that $\beta = S(\tau)$ where for simplicity $t_k < \tau \leq t_{k+1}$, but our results could easily be extended to include extrapolation where $\tau < t_1$ or $\tau > t_n$. As in

Section 2 it is assumed that the specified means process has been removed so $S(t)$ has expectation 0 for all t .

In this section an approximation to the Bayes linear interpolator of β , which is itself a Bayes linear procedure, will be found by applying the approximations developed in the last section. The first approximation is made when Y is replaced by P , the observable consisting of those $S(t_i)$ for which $|t_i - \tau|$ is small. Then an order 0 Taylor approximation is made to approximate $\alpha_{\beta P}$ under the assumption of continuity to get $\alpha_{\beta P}^0$. Finally, the first approximation method in Section 2 is reapplied to obtain $\alpha_{\beta Y}^0$, an approximate Bayes linear interpolator of β based on Y .

Throughout this section it is assumed that S is continuous according to the following definition.

DEFINITION 1. $\{S(t) : t \in T\}$ will be called continuous at $t = t_0$ if

$$E \|S(t) - S(t_0)\|_S^2 \rightarrow 0 \quad \text{as } t \rightarrow t_0.$$

Define $\Gamma(s, t)$ on $T \times T$ by

$$\Gamma(s, t) = \Gamma_{S(s)S(t)};$$

this transformation will be called the covariance kernel of S .

Our analysis uses an extension of a well known result of classical time series that the continuity of S and its covariance kernel are equivalent when the continuity of Γ is appropriately defined.

DEFINITION 2. Γ will be called continuous at (s, u) if $\|\Gamma(t, v) - \Gamma(s, u)\| \rightarrow 0$ when $(t, v) \rightarrow (s, u)$ where, here and in the sequel, $\|\cdot\|$ will denote the strong operator norm on $L(S, S)$. Γ will be called continuous if it is continuous at all (s, u) .

The proof, below, of our counterpart of the classical result is similar to that of Grenander (1981, Theorem 1, p 38), who treats the case where S is a complex valued process.

THEOREM 6. S is continuous on T if and only if $\Gamma(s, t)$ is jointly continuous in (s, t) at every point $(u, u) \in T \times T$, in which case Γ is continuous on $T \times T$.

PROOF. First suppose Γ is continuous at (u, u) . Now for every (s, t) , $E(S(s), S(t))_S = \langle I, \Gamma(s, t) \rangle_{SS}$ (Eaton, 1983, p 93). Thus

$$E \|S(t) - S(u)\|_S^2 = \langle I, \Delta\Gamma \rangle_{SS},$$

where

$$\Delta\Gamma = \Gamma(t, t) - \Gamma(t, u) - \Gamma(u, t) + \Gamma(u, u).$$

But if w_i is any basis for S , then for certain coefficients, γ_{ij}

$$\Delta \Gamma = \sum \sum \gamma_{ij} w_i \square w_j, \quad \langle I, \Delta \Gamma \rangle_{SS} = \sum \gamma_{ii} = \sum (w_i, (\Delta \Gamma) w_i)_S.$$

By the assumed continuity of Γ (see Definition 2), $(w_i, (\Delta \Gamma) w_i)_S \rightarrow 0$ for each i as $t \rightarrow u$ which proves S is continuous at u for all $u \in T$ and the theorem itself in one direction. Now conversely suppose S is continuous on T . Observe that

$$\begin{aligned} D^* &\stackrel{\Delta}{=} (y, \Gamma(t, v)x)_S - (y, \Gamma(s, u)x)_S \\ &= (y, \Gamma(t, v)x)_S - (y, \Gamma(s, v)x)_S + (y, \Gamma(s, v)x)_S - (y, \Gamma(s, u)x)_S \\ &= E(y, S(t) - S(s))_S (x, S(v))_S + E(y, S(s))_S (x, S(v) - S(u))_S, \end{aligned}$$

by the definition of the covariance kernel. It follows that

$$|D^*| \leq [E(y, S(t) - S(s))_S^2]^{\frac{1}{2}} [E(x, S(v))_S^2]^{\frac{1}{2}} + [E(y, S(s))_S^2]^{\frac{1}{2}} [E(x, S(v) - S(u))_S^2]^{\frac{1}{2}}.$$

But

$$\begin{aligned} E(y, S(t) - S(s))_S^2 &\leq \|y\|_S^2 E\|S(t) - S(s)\|_S^2, \\ E(x, S(v) - S(u))_S^2 &\leq \|x\|_S^2 E\|S(v) - S(u)\|_S^2 \end{aligned}$$

and

$$\begin{aligned} [E(x, S(v))_S^2]^{\frac{1}{2}} &= [E\{(x, S(v) - S(u))_S + (x, S(u))_S\}^2]^{\frac{1}{2}} \\ &\leq [E(x, S(v) - S(u))_S^2]^{\frac{1}{2}} + [E(x, S(u))_S^2]^{\frac{1}{2}}, \end{aligned}$$

the last inequality deriving from that of Minkowski. Thus $|D^*| \rightarrow 0$ as $(t, v) \rightarrow (s, u)$ for every $x, y \in S$ and for every fixed pair, (s, u) . But this implies (Halmos, 1974, pp 182-183) that Γ is continuous on $T \times T$ in the sense of Definition 2.

The proof is completed by noting that the continuity of Γ on the diagonal of $T \times T$ implies that of S on T and this in turn implies that Γ is continuous everywhere on $T \times T$ by the result of the last paragraph. \square

The process $\{S(t) : t \in T\}$ will be called wide sense stationary (WSS) if $\Gamma(s, t) = \Gamma(s - t)$. In that case $\alpha_{S(s)S(v)} \stackrel{\Delta}{=} \Gamma(s, v) \Gamma^{-1}(v, v) = \Gamma(s - v) \Gamma^{-1}(0)$. If S is WSS, it will be called a first order autoregressive process, if for all $u \geq 0$, $\Gamma(u) = \gamma \exp(-uR_0)\gamma$ where $\gamma \stackrel{\Delta}{=} \Gamma^{\frac{1}{2}}(0)$ is a self adjoint, positive definite transformation and R_0 is a fixed transformation; in general $\exp(A) \stackrel{\Delta}{=} \sum_{r=0}^{\infty} A^r/r!$ is a linear transformation on W when $A \in L(W, W)$ and $[\exp(A)]^T = \exp(A^T)$.

Observe that if $t < u$,

$$(x, \Gamma(t - u)y)_S = E(x, S(t))_S (y, S(u))_S = (y, \Gamma(u - t)x)_S = (x, \Gamma^T(u - t)y)_S$$

so that $\Gamma(t - u) = \Gamma^T(u - t)$. Thus for $u < 0$, $\Gamma(u) = \Gamma^T(-u)$. Therefore, if S is a first order autoregressive process, we must have

$$\Gamma(u) = \gamma[\exp(uR_0)]^T \gamma, \quad u < 0$$

i.e.

$$\Gamma(u) = \gamma[\exp(-|u|R_0^T)]\gamma, \quad u < 0.$$

LEMMA 5. Suppose S is WSS. Then the following statements are equivalent:

- (i) S is continuous on T and $S(t)$ is uncorrelated with $S(v) - \alpha_{S(v)|S(u)} S(u)$ for every $t \leq u \leq v, t, u, v \in T$;
- (ii) $S(t)$ is a first order autoregressive process; and
- (iii) S is continuous on T and $S(v)$ is uncorrelated with $S(t) - \alpha_{S(t)|S(u)} S(u)$ for every $t \leq u \leq v, t, u, v \in T$.

PROOF. It is first shown that (i) implies (ii). Since S is continuous on T , Γ is continuous (Theorem 8). By Definition 2, this implies the elements of any matrix representation of Γ , say $[\Gamma]$ are continuous. We also have from (i) that for every

$$\begin{aligned} x_1, x_2 \in S, \quad 0 &= E(x_1[S(v) - \alpha_{S(v)|S(u)} S(u)]_S (x_2 S(t))_S) \\ &= (x_1, \Gamma(v-t)x_2)_S - (\alpha_{S(v)|S(u)}^T x_1, \Gamma(u-t)x_2)_S \\ &= (x_1, [\Gamma(v-t) - \alpha_{S(v)|S(u)} \Gamma(u-t)]x_2)_S. \end{aligned}$$

Thus,

$$\Gamma(v-t) = \Gamma(v-u)\Gamma^{-1}(0)\Gamma(u-t), \quad (3.1)$$

for every $t \leq u \leq v, t, u, v \in T$. Let $\beta(u) = \Gamma^{-1}\Gamma(u)\Gamma^{-1}$. Then equation (3.1) is, equivalently,

$$\alpha(x+y) = \alpha(x)\alpha(y), \quad 0 \leq x \leq y < T_0$$

where $T_0 = \max\{v-t : v, t \in T\}$. In terms of the matrix representation of α , $[\alpha]$, this is

$$[\alpha(x+y)] = [\alpha(x)][\alpha(y)] \quad (3.2)$$

which is Polya's matrix equation. Since $[\alpha]$ is continuous in the neighborhood of 0, equation (3.2) has the solution

$$[\alpha(u)] = \exp(-u[R_0]), \quad u > 0$$

for some constant matrix, $[R_0]$ (C. F. Bellman, 1965, Theorem 4, p 173). Thus,

$$\alpha(u) = \exp(-uR_0), \quad u \geq 0$$

for the linear transformation R_0 determined by the matrix $[R_0]$. $[R_0]$ depends on the basis chosen to obtain the matrix representation of α so R_0 is not unique but it is not required to be and (ii) now follows.

The same sort of reasoning shows that (iii) implies (ii). To complete the proof, it is trivial to show that (ii) implies both (i) and (iii). \square

Let δ be any sufficiently large number that

$$M_\delta = \{j : |t_j - \tau| \leq \delta\} \quad (3.3)$$

is not empty and such that δ is attained for some $|t_j - \tau|$. $|M_\delta| = m$ will denote the number of elements in M_δ and the sampling times corresponding to these elements will be ordered as $t_{p_1} < \dots < t_{p_m}$ while those corresponding to the elements of the complement of M_δ are $t_{R_1} < \dots < t_{R_{n-m}}$. Finally, let $t_m = (t_{p_1}, \dots, t_{p_m})$, $P_i = S(t_{p_i})$, $R_j = S(t_{R_j})$, $P = \langle P_1, \dots, P_m \rangle$, and $R = \langle R_1, \dots, R_{n-m} \rangle$. More will be said about the choice of δ and m in the sequel. Partition $Y = \langle Y(t_1), \dots, Y(t_n) \rangle$ as $Y = \langle P, R \rangle$. Corresponding to this partition, partition Y as $Y = P \times R$ where P and R denote, respectively, the ranges of P and R . Then the Bayes linear interpolator of $\beta = S(\tau)$, $t_k < \tau \leq t_{k+1}$ based on P is that given in equation (1.4).

Theorem 1 concludes that

$$||\beta_Y - \beta_P||_S \leq |\alpha_{R,P}|_{RS} ||R||_R$$

so that the error in approximating β_Y by β_P is governed by $|\alpha_{R,P}|_{RS}$. Theorem 2 gives general conditions under which this error is identically zero. We now give other more specialized conditions under which the same result obtains.

Lemma 5 gives conditions under which there is no approximation error in reducing the data from Y to P .

THEOREM 7. Suppose S is a first order autoregressive process as defined above and that $P = \langle Y(t_k), Y(t_{k+1}) \rangle$ while $R = \langle Y(t_1), \dots, Y(t_{k-1}) \rangle$, when $t_k \leq \tau \leq t_{k+1}$. Then $\alpha_{R,P} = 0$, i.e. $\beta_Y = \beta_P$.

PROOF. Recall that $\alpha_{R,P}$ minimizes $E ||R - aP||_k^2 = \sum_{i=1}^{(n-2)} E (|R_i - a_i^{(1)}P_1 - a_i^{(2)}P_2|)^2$ as a function of a . But Lemma 5 implies that if $i \leq k-1$, $a_i^{(1)} = \alpha_{R,P_1}$ and $a_i^{(2)} = 0$ are optimal while if $i > k-1$, $a_i^{(1)} = 0$ and $a_i^{(2)} = \alpha_{R,P_2}$ are optimal, from which $\alpha_{R,P}$ is obtained. Reapplying Lemma 5, we deduce that $\beta = S(\tau)$ and $R - \alpha_{R,P}P$ are uncorrelated, i.e. $\alpha_{R,P} = 0$. \square

The one step autoregressive (AR(1)) model can only obtain when uncertainties about its governing hyperparameters are ignored by conditioning on them; unconditionally, the process cannot be AR(1). But even conditional on its hyperparameters, the AR(1) model can only be an approximate representation of the analyst's prior views. Therefore, a procedure which relies solely on P_1 and P_2 to interpolate $S(\tau)$ seems likely to be nonrobust. Instead it seems preferable to take $P = \langle Y(t_{p_1}), \dots, Y(t_{p_m}) \rangle$ where $m > 2$ is permitted. An approximation to β_P will be obtained below when S is

any member of a family of underlying processes for which S is differentiable in the sense of the next definition. But to make β_P a reasonable approximation of β_Y , we need to assume $\alpha_{\beta_R, P}$ is "small" when m , the dimension of P , is sufficiently large as an alternative to the stronger assumption that S is AR(1). This makes precise, the implicit assumption underlying any interpolator which uses only the data in a "window" enclosing the point of interest. At the same time we expand the domain of our work over that provided by the AR(1) model.

DEFINITION 3. S will be called differentiable at $s \in T$ if there exists an element, $S' \in S$ such that $E \|Q(t, s) - S'(t)\|_S^2 \rightarrow 0$ as $t \rightarrow s$, where $Q(t, s) = [S(t) - S(s)](t - s)^{-1}$.

Just as the continuity of S and Γ are linked, so are their differentiability, that of Γ being given by the following definition.

DEFINITION 4. A mapping $G : T^2 \rightarrow L(S, S)$ has the mixed derivative, G^{11} , at $(s, u) \in T^2$ if $\|Q_2 - G^{(11)}(s, u)\| \rightarrow 0$ as $(t, v) \rightarrow (s, u)$ where $Q_2 = [G(t, v) - G(s, v) - G(t, u) + G(s, u)] / [(t - s)(v - u)]$ and $\|\cdot\|$ denotes the appropriate strong operator norm. G will be said to have a mixed first derivative if it has a mixed first derivative at all $(s, u) \in T^2$. Other derivatives are defined in an analogous way.

It is easily shown that the product and addition rules of calculus hold so, for example, if

$$H(s, u) = H_1(s, u)H_2(s, u),$$

then

$$H^{(10)}(s, u) = H_1^{(10)}(s, u)H_2(s, u) + H_1(s, u)H_2^{(10)}(s, u).$$

With these definitions we have the following theorem whose proof is similar to that of Grenander (1981, Theorem 1, p. 47).

THEOREM 8. If S is differentiable on T in the sense of Definition 3, then $\Gamma^{(11)}(u, u)$ exists at every point, $(s, u) \in T \times T$ in the sense of Definition 4.

PROOF. With Q_2 given by Definition 4 on replacing G by Γ ,

$$(x, Q_2 y)_S = E(x, Q(t, s))_S (y, Q(v, u))_S$$

or

$$(x, Q_2 y)_S = E(x, D(t, s) + S'(t))_S (y, D(v, u) + S'(u))_S, \quad (3.4)$$

where $D(h, k) = Q(h, k) - S'(k)$ for all $(h, k) \in S \times S$ and Q as given in Definition 3. Then the inequality, $|E(x, W_1)_S (y, W_2)_S|^2 \leq \|x\|_S^2 \|y\|_S^2 E\|W_2\|_S^2$ in

conjunction with equation (3.4) implies that

$$(x, [Q_2 - \Gamma_{S\gamma_s, S\gamma_s}]y)_S \rightarrow 0 \text{ as } (t, \tau) \rightarrow (s, u)$$

for every x, y and this not only proves the theorem but shows, incidentally that $\Gamma^{(11)}(s, u) = \Gamma_{S\gamma_s, S\gamma_s}$.

Definitions 3 and 4 can be iterated to obtain the definitions of the higher order derivatives needed in the sequel. In particular the existence of $S^{(2)}$ implies that of $\Gamma^{(22)}$ since $S^{(2)}$ is the first derivative of $S^{(1)}$ whose covariance is $\Gamma^{(11)}(s, u) = \Gamma_{S\gamma_s, S\gamma_s}$; $\Gamma^{(22)}$ need not be continuous. Note that the existence of $S^{(2)}$ implies that of $\Gamma^{(kl)}$ for all $k, l, 0 \leq k, l \leq 2$.

Observe that Definitions 2 and 4 entail properties of Γ which hold only if they are possessed of every bivariate real valued functions of (s, u) defined by $h_{xy}(s, u) = (x, \Gamma(s, u)y)_S, (x, y) \in S \times S$. Thus, for example, the existence of $\Gamma^{(10)}(s, u)$ implies that of $h_{xy}^{(10)}(s, u)$:

$$\lim_{t \rightarrow s} [h_{xy}(t, u) - h_{xy}(s, u)]/(t-s) = (x, \Gamma^{(10)}(s, u)y)_S.$$

Recall that $(\cdot, \cdot)_P = (\cdot, \cdot)_{S^m} = (\cdot, \cdot) + \dots + (\cdot, \cdot)$, the m -fold sum of the inner products on S . Furthermore, the transformation $\alpha_{P\beta} \in L(B, P)$ with $B = S$ and $P = S^m$ can be partitioned with $\alpha_i = \alpha_{S(\beta_i), S(\tau)}$ for simplicity, as

$$\alpha_{P\beta} b = \langle \alpha_1 b, \dots, \alpha_m b \rangle.$$

Thus

$$(p, \alpha_{P\beta} b)_P = \sum_{i=1}^m (p_i, \alpha_i b)_S.$$

The approximation error, $\Delta(\alpha_{P\beta})$, is determined by $\Delta(\alpha_i)$ through $(x, \Delta(\alpha_i)b)_S$. If the process is thought to be continuous a natural approximant to $\alpha_{P\beta}$ is that for which $\alpha_i = \alpha_i^0 = I$, so that $\Delta(\alpha_i) = \alpha_i - I$, the subject of the next lemma. To make clearer the association between the results obtained below and those of the last section, $\alpha_{P\beta}$ and its approximant, $\alpha_{P\beta}^0$, will be denoted by A and A_0 , respectively. Similarly, in the sequel, C and C_0 will denote $\Gamma_{P\beta}$ and $\Gamma_{P\beta}^0$, respectively.

LEMMA 6. Assume (i) S is differentiable on T in the sense of Definition 3, (ii) for any given constant, $K_0 > 1$, $\delta > 0$ is chosen sufficiently small that

$$[E \|D(t, \tau)\|^2]^{1/2} \leq (K_0 - 1)^2 \|\Gamma^{-1}(\tau, \tau)\|_{SS}$$

for every $t \in [\tau - \delta, \tau + \delta]$, where D is as defined in the proof of Theorem 8, and (iii) M_δ defined in equation (3.3), is not empty. Then

$$e_A = [\Delta(A)]_{AP} / [A_0]_{AP} \leq e_A^0$$

where

$$\varepsilon_A^0 = m^3 \delta K_0 | \Gamma^{-1}(\tau, \tau) |_{SS}.$$

PROOF. By definition and since the noise process is independent of the S - process,

$$(p_i, A_i b)_S = E (p_i, S(t_{p_i}))_S (\Gamma^{-1}(\tau, \tau) b, S(\tau))_S,$$

and $(p_i, b)_S = E (p_i, S(\tau))_S (\Gamma^{-1}(\tau, \tau) b, S(\tau))_S$. Thus,

$$(p_i, \Delta(A_i) b)_S = (t_{p_i} - \tau) E (p_i, D(t_{p_i}, \tau) + S'(\tau))_S (\Gamma^{-1}(\tau, \tau) b, S(\tau))_S.$$

The Cauchy-Schwarz inequality implies that

$$| (p_i, \Delta(A_i) b)_S | \leq | t_{p_i} - \tau | h_i h$$

where

$$h_i^2 \triangleq E (p_i, D(t_{p_i}, \tau) + S'(\tau))_S^2$$

and

$$h^2 \triangleq (\Gamma^{-1}(\tau, \tau) b, b)_S.$$

But

$$h \leq | \Gamma^{-1}(\tau, \tau) |_{SS}^{1/2} | b |_{FS}.$$

At the same time, Minkowski's inequality implies

$$h_i \leq [E (p_i, D(t_{p_i}, \tau))_S^2]^{1/2} + [E (p_i, S'(\tau))_S^2]^{1/2}.$$

and hence

$$h_i \leq \{ [E | D(t_{p_i}, \tau) |^2]^{1/2} + | \Gamma^{(11)}(\tau, \tau) |_{SS}^{1/2} \} | p_i |_S,$$

which by hypothesis (ii) implies

$$h_i \leq K_0 | \Gamma^{-1}(\tau, \tau) |_{SS}^{1/2} | p_i |_S.$$

It follows that

$$(p_i, \Delta(A_i) b)_S \leq K_0 | t_{p_i} - \tau | | \Gamma^{-1}(\tau, \tau) |_{SS}^{1/2} | p_i |_S | b |_{FS}.$$

Thus

$$(p, \Delta(A) b)_P \leq K_0 | \Gamma^{-1}(\tau, \tau) |_{SS}^{1/2} | b |_S \sum | t_{p_i} - \tau | | p_i |_S,$$

which implies

$$(p, \Delta(A) b)_P \leq m \delta K_0 | \Gamma^{-1}(\tau, \tau) |_{SS}^{1/2} | b |_S | p |_P.$$

It is easily shown that $| A_0 | \leq m^{-3/2}$ by, for example, choosing a basis for S and maximizing $(p, A_0 b)_P$ using Lagrange multipliers. The conclusion follows. \square

We turn now to the analysis of the residual covariance transformation, $\Gamma_{P, \beta}$ which plays the role of C of Section 2. To simplify the statement of the next result, the following notation will be adopted for $k, l = 0, 1, \dots, l = 1, \dots, n$:

$$\begin{aligned}\Gamma^{(kl)} &= \Gamma^{(kl)}(\tau, \tau), \\ \Gamma_{M, 0} &= \Gamma^{(kl)} - \Gamma^{(kl)} \Gamma^{-1} \Gamma^{(kl)}, \\ \Delta_i &= t_i - \tau.\end{aligned}$$

Note that $\Gamma^{(00)} = \Gamma^{(00)T}$.

Our approximation, C_0 , to C will be defined as a partitioned transformation through $(p, C_0 q)_P = \sum_{i,j \in M_k} (p_i, C_{P_i, P_j}^0 q_j)_S$ for all $p, q \in P$, where

$$C_{P_i, P_j}^0 = \Sigma_{P_i, P_j} + \eta_{P_i, P_j} + \zeta_{P_i, P_j} \quad (3.5)$$

where

$$\begin{aligned}\eta_{P_i, P_j} &= \Delta_{P_i, P_j} \Gamma_{11, 0} \\ \zeta_{P_i, P_j} &= |\Delta_{P_i}| |\Delta_{P_j}| (|\Delta_{P_i}| + |\Delta_{P_j}|) \sigma_{P_i, P_j}, \\ \Delta_{P_i, P_j} &= \Delta_{P_i} \Delta_{P_j} \exp[-(|\Delta_{P_i}| + |\Delta_{P_j}|) \gamma \Delta_0].\end{aligned}$$

$\Delta_0 > 0$ is a specified constant, $\Sigma_{P_i, P_j} = \sigma_{P_i, P_j} = 0$, $i \neq j$, and $\Sigma_{P_i, P_i}, \sigma_{P_i, P_i}$ are positive definite transformations in $L(S, S)$, $i, j \in M_k$. Let Σ_{PP}, η_{PP} and ζ_{PP} denote the matrices corresponding to the partition components defined in equation (3.5).

ASSUMPTION 1. $\Gamma^{(kl)}$ exists and is continuous for all $k, l \leq 3$.

ASSUMPTION 2. $\varepsilon_0 < 1$ where

$$\varepsilon_0 = m \delta^2 \max_{i \in M_k} (|\Gamma_{11, 0}|_{SS} / \Delta_0 + \delta |\sigma_{P_i, P_i}|_{SS}) |\Sigma_{P_i, P_i}^{-1}|_{SS}.$$

Let

$$G(s, \mu) = \Gamma(s, \mu) - \Gamma(s, \tau) \Gamma^{-1}(\tau, \tau) \Gamma^T(\mu, \tau). \quad (3.6)$$

Now if $g: T^2 \rightarrow L(S, S)$ is continuous at (s, μ) , then so is $\|g\|$, where $\|\cdot\|$ denotes the appropriate operator norm, since

$$\| \|g(t, v)\| - \|g(s, \mu)\| \| \leq \|g(t, v) - g(s, \mu)\|.$$

Assumption 1 implies therefore that

$$K_C = \sup \|G^{(r, 3-r)}(s, \mu)\|_{SS}, \quad (3.7)$$

is finite where the supremum is over all $(s, \mu) \in [\tau - \delta, \tau + \delta]^2$ and $r = 0, 1, 2, 3$.

LEMMA 7. If Assumptions 1 and 2 hold then

$$e_C = |\Delta(C)|_{PP} |C_0^{-1}|_{PP} \leq e_C^0$$

where with the constant K_C defined in equation (3.7),

$$e_C^0 = 2m \delta^3 (K_C + |\Gamma_{110}|_{SS}/\Delta_0 + \max |\sigma_{P_i P_i}|_{SS}) \max |\sigma_{P_i P_i}^{-1}|_{SS} (1 + \delta_{e_0}) ,$$

$\delta_x = x(1-x)^{-1}, 0 \leq x < 1$, and e_0 is given in Assumption 2.

PROOF. By definition

$$\begin{aligned} \langle p, Cq \rangle_{PP} &= E \langle p, P \cdot \beta \rangle_P \langle q, P \cdot \beta \rangle_P \\ &= \sum \sum E \langle p_i, P_i \cdot \beta \rangle_S \langle q_j, P_j \cdot \beta \rangle_S \end{aligned}$$

and it is straightforward then to show that

$$\langle p, Cq \rangle_P = \sum \sum \langle p_i, C_{P_i P_j} q_j \rangle_S$$

where for all i, j ,

$$C_{P_i P_j} = G(t_{P_i}, t_{P_j}) + \Sigma_{P_i P_j}$$

and G is defined in equation (3.6). Our assumptions imply that $G^{(kl)}$ exists and is continuous at (τ, τ) , for $k, l = 0, 1, 2, 3$. Thus the real valued function,

$$h(x, u) = h_{xy}(x, u) = \langle x, G(x, u)y \rangle_S ,$$

has the same continuous derivatives by the definition of differentiation and hence it has a Taylor series expansion. It follows that

$$\langle p_i, G(t_{P_i}, t_{P_j}) q_j \rangle_S = \sum_{k=0}^2 \frac{1}{k!} \sum_{r=0}^k \binom{k}{r} (\Delta_{P_i})^r (\Delta_{P_j})^{k-r} h^{(r, k-r)} + r_{P_i P_j} ,$$

where

$$\begin{aligned} h^{(r, k-r)} &= \langle p_i, G^{(r, k-r)} q_j \rangle_S , \\ G^{(r, k-r)} &= G^{(r, k-r)}(\tau, \tau) , \\ r_{P_i P_j} &= \frac{1}{3!} \sum_{r=0}^3 \binom{3}{r} \Delta_{P_i}^r \Delta_{P_j}^{3-r} h^{(r, 3-r)}(\tau_i^*, \tau_j^*) , \end{aligned}$$

for some τ_i^*, τ_j^* for which $|\tau_i^* - \tau| \leq |\Delta_{P_i}| \leq \delta$, $|\tau_j^* - \tau| \leq |\Delta_{P_j}|$ and $h^{(r, k-r)}(x, u) = \langle p_i, G^{(r, k-r)}(x, u) q_j \rangle_S$. Now it is straightforward to show that $G(\tau, \tau) = G^{(10)}(\tau, \tau) = G^{(01)}(\tau, \tau)$ and hence that the Taylor expansion for h reduces to

$$\langle p_i, G(t_{P_i}, t_{P_j}) q_j \rangle_S = \Delta_{P_i} \Delta_{P_j} h^{(11)} + r_{P_i P_j} ,$$

where

$$h^{(11)} = \langle p_i, \Gamma_{110} q_j \rangle_S$$

since by definition,

$$G^{(11)}(\tau, \tau) = \Gamma_{110}.$$

Define $R_{p_i p_j}$ and $\Delta(C_{p_i p_j})$, respectively, by

$$R_{p_i p_j} = G(t_{p_i} t_{p_j}) - \Delta_{p_i} \Delta_{p_j} \Gamma_{110} \quad (3.8)$$

and

$$\Delta(C_{p_i p_j}) = C_{p_i p_j} - C_{p_i p_j}^0.$$

Then $|(p_i R_{p_i p_j} q_j)_S| = |r_{p_i p_j}| \leq K_C ||p_i||_S ||q_j||_S \delta^3$ uniformly in p_i and q_j where K_C is given in equation (3.7). It is easily verified that

$$\Delta(C_{p_i p_j}) = \Delta_{p_i} \Delta_{p_j} \Gamma_{110} - \Sigma_{p_i p_j}^*,$$

where for expository convenience $\Sigma_{p_i p_j}^* = \eta_{p_i p_j} + \zeta_{p_i p_j}$, η and ζ being defined in equation (3.5). Since in general, $0 \leq 1 - \exp(-x) \leq x$ for all $x \geq 0$ it follows that

$$|\Delta(C_{p_i p_j})|_{SS} \leq |\Delta_{p_i}| |\Delta_{p_j}| (|\Delta_{p_i}| + |\Delta_{p_j}|)(|\Gamma_{110}|_{SS}/\Delta_0 + |\sigma_{p_i p_j}|_{SS}).$$

Thus

$$|(p_i \Delta(C_{p_i p_j}) q_j)_S| \leq 2\delta^3 (K_C + |\Gamma_{110}|_{SS}/\Delta_0 + |\sigma_{p_i p_j}|_{SS}) ||p_i||_S ||q_j||_S,$$

and

$$|(p \Delta(C) q)_P| \leq 2\delta^3 (K_C + |\Gamma_{110}|_{SS}/\Delta_0 + \max |\sigma_{p_i p_j}|_{SS}) \sum \sum ||p_i||_S ||q_j||_S.$$

Since

$$\sum \sum ||p_i||_S ||q_j||_S \leq m ||p||_P ||q||_P,$$

we conclude that

$$|\Delta(C)|_{PP} \leq 2m \delta^3 (K_C + |\Gamma_{110}|_{SS}/\Delta_0 + \max |\sigma_{p_i p_j}|_{SS})$$

Next define $R_{p_i p_j}^*$ by

$$G(t_{p_i} t_{p_j}) = C_{p_i p_j}^0 + R_{p_i p_j}^*, \quad i, j \in M_\delta.$$

From equations (3.5) and (3.8) we obtain $|(p_i R_{p_i p_j}^* q_j)_S| \leq K ||p_i||_S ||q_j||_S \delta^3$, uniformly in $||p_i||_S$ and $||q_j||_S$.

Now referring to the objects defined in equations (3.5), for every $p, q \in P$, with $||p||_P = ||q||_P = 1$

$$\begin{aligned} (p \Sigma_{PP}^* \Sigma_P^{-1} q)_P &= \sum \sum (p_i \Sigma_{p_i p_j}^* \Sigma_{p_j}^{-1} q_j)_S \\ &\leq \sum \sum ||p_i||_S ||q_j||_S |\Sigma_{p_i p_j}^* \Sigma_{p_j}^{-1}|_{SS} \\ &\leq m \max_{i,j} |\Sigma_{p_i p_j}^*|_{SS} |\Sigma_{p_j}^{-1}|_{SS}. \end{aligned}$$

But,

$$|\Sigma_{P_i, P_j}^*|_{SS} \leq \delta^2 |\Gamma_{11-0}|_{SS} + \delta^3 |\sigma_{P_i, P_j}|_{SS}.$$

Thus

$$|\Sigma_{PP}^* \Sigma_{PP}^{-1}|_{PP} \leq e_0, \quad (3.9)$$

with e_0 defined in Assumption 2. By that assumption, the quantity on the left of inequality (3.9) is less than 1. By Lemma 2(a),

$$|C_0^{-1}|_{PP} \leq |\Sigma_{PP}^{-1}|_{PP}(1 + \delta_{e_0}).$$

Since $|\Sigma_{PP}^{-1}|_{PP} = \max |\Sigma_{P_i, P_j}^{-1}|_{SS}$, the conclusion is obtained. \square

We may now obtain bounds for the local approximation errors.

THEOREM 9. Suppose the hypotheses of Lemmas 6 and 7 hold. Let $e_T = |(\Gamma_{\beta\beta}^0)^{-1}|_{SS} |\Gamma_{\beta P}|_{SS}$. Then

$$|\Delta(\Gamma_{\beta P})|_{SS} / |\Gamma_{\beta P}|_{SS} \leq \delta_{eT_{\beta P}},$$

and

$$||\Delta(\beta_P)||_{SS} / [|\Gamma_{\beta P}|_{SS} |A_0|_{PS} |C_0^{-1}|_{PP}] \leq \delta_{eT_{\beta P}} + e_A^0 + e_C^0 + e_A^0 e_C^0,$$

where e_A^0 and e_C^0 are defined in Lemmas 6 and 7, respectively, and

$$e_{T_{\beta P}}^0 = e_T + m \max |\sigma_{P_i, P_j}^{-1}|_{SS} [\delta_{e_0}(1 + e_A^0) + e_A^0(1 + e_A^0)(1 + \delta_{e_0}) + e_A^0].$$

PROOF. This conclusion is an immediate consequence of Theorem 5 and the preceding analysis. \square

To simplify the application of our proposed approximate interpolator/smoothen we will now extend the approximation β_P^0 to β_Y^0 and show that the resulting error $\Delta(\beta_Y^0) = \beta_Y^0 - \beta_P^0$ is "small" under certain circumstances. To this end let $\alpha_{Y\beta}^0 \in L(FS, Y)$ be the obvious extension of $\alpha_{P\beta}^0 = A_0$, i.e.

$$\alpha_{Y\beta}^0 b = \langle b, \dots, b \rangle \in Y$$

for every $b \in S$. By the definition of $\alpha_{Y\beta}^0$, it follows that for the i th partition, $\Gamma_{k,\beta}^0$ of $\Gamma_{k\beta}^0$, we must take

$$\Gamma_{k,\beta}^0 = \Gamma_{\beta\beta}^0.$$

To obtain $\Gamma_{Y,\beta}^0$ we extend the approximation for $\Gamma_{P,\beta}^0 = C_0$ given in equation (3.5), in the obvious way so, for example, the i - j th component of $\Gamma_{kP,\beta}^0$ is

$$\Gamma_{kP,\beta}^0 = \Delta_{kP} \Gamma_{11-0}.$$

With this notation, we may state

ASSUMPTION 3. $\epsilon_i < 1$, $i = 1, 2, 3$ where

$$\epsilon_1 = \Delta^2 |\Gamma_{11-0}|_{SS} |\Gamma_{PP}^{-1}|_{PP} \max |\Sigma_{R_i R_i} + \eta_{R_i R_i}|_{SS}^{-1},$$

$$\Delta = \min |\Delta_{R_i}|,$$

$$\epsilon_2 = \delta^2 \max |\Sigma_{PP}^{-1}|_{SS} [|\Gamma_{11-0}|_{SS} + \delta \max |\sigma_{PP}|_{SS}] |(\sum (\Sigma_{PP}^{-1}))^{-1}|_{SS} (1 + \delta_{\omega})$$

and

$$\epsilon_3 = 1/(\delta/\Delta) \exp[-2(\delta + \Delta/\Delta_0)] |\Gamma_{11-0}|_{SS}^2 |\Gamma_{PP}^{-1}|_{PP} |\sigma_{RR}^{-1}|_{PP},$$

where δ_x is defined in Lemma 7 for all x .

THEOREM 10. If $\delta \leq \Delta_0 \leq \Delta$ assumptions 2 and 3 imply

$$||\Delta(\beta_{\beta}^0)||_{\beta\beta} ||R||_{RR} \leq |D|_{SS} |\alpha^*|_{SR} |(\Gamma_{R,P}^0)^{-1}|_{RR} \quad (3.10)$$

and

$$|\Delta(\Gamma_{\beta\gamma}^0)|_{SS} \leq |\Delta|_{SS}^2 |\alpha^*|_{SR} |(\Gamma_{R,P}^0)^{-1}|_{RR} \quad (3.11)$$

where D and α^* are defined in Theorem 1 and

$$|D|_{SS} \leq |(\sum (\Sigma_{PP}^{-1}))^{-1}|_{SS} (1 + \delta_{\omega}), \quad (3.12)$$

$$|\alpha^*|_{SR} \leq (n-m)^k + m^{1/2} \delta \exp[-(\delta + \Delta/\Delta_0)] |\Gamma_{11-0}|_{SS} |\Gamma_{PP}^{-1}|_{PP} (1 + \delta_{\omega}), \quad (3.13)$$

and

$$|(\Gamma_{R,P}^0)^{-1}|_{RR} \leq 1/\Delta^{-1} |\sigma_{RR}^{-1}|_{RR} (1 + \delta_{\epsilon}). \quad (3.14)$$

PROOF. It is easily shown that if G and H are any self adjoint transformation with G positive and H nonnegative definite, the $|(G+H)^{-1}| \leq |G^{-1}|$ in the appropriate strong operator norm. Thus

$$|D|_{SS} \leq |[(\alpha_{\beta\beta}^0)^T (\Gamma_{\beta\beta}^0)^{-1} (\alpha_{\beta\beta}^0)]^{-1}|_{SS}$$

It is shown in the proof of Lemma 7 that

$$|\eta_{PP}|_{PP} \leq \delta^2 |\Gamma_{11-0}|_{SS} + \delta^3 \max |\sigma_{PP}|_{SS}$$

and also that

$$|(\eta_{PP} + \zeta_{PP}) \Sigma_{PP}^{-1}|_{PP} \leq \epsilon_0$$

where by Assumption 2, $\epsilon_0 < 1$. So by Lemma 2(a)

$$(\Gamma_{\beta\beta}^0)^{-1} = \Sigma_{PP}^{-1} - \Sigma_{PP}^{-1} \Sigma_{PP}^* (\sum_{r=0}^{\infty} [-\Sigma_{PP}^* \Sigma_{PP}^{-1}]^r) \Sigma_{PP}^{-1},$$

where $\Sigma_{PP}^* = \eta_{PP} + \zeta_{PP}$. Thus

$$(\alpha_{\beta\beta}^0)^T (\Gamma_{\beta\beta}^0)^{-1} \alpha_{\beta\beta}^0 = H_1 - H_2 G_1 H_2^T$$

where

$$H_1 = (\alpha_{\beta\beta}^0)^T \Sigma_{\beta\beta}^{-1} \alpha_{\beta\beta}^0 , \\ H_2 = (\alpha_{\beta\beta}^0)^T \Sigma_{\beta\beta}^{-1} .$$

and

$$G_1 = \Sigma_{\beta\beta}^* \sum_{r=0}^{\infty} [- \Sigma_{\beta\beta}^* \Sigma_{\beta\beta}^{-1}]^r .$$

It follows that

$$[(\alpha_{\beta\beta}^0)^T (\Gamma_{\beta\beta}^0)^{-1} (\alpha_{\beta\beta}^0)]^{-1} = H_1^{-1} [I - H_2 G_1 H_2^T H_1^{-1}]^{-1} .$$

But it is easily seen that

$$|H_2|_{\beta\beta} \leq m^k \max |\Sigma_{\beta\beta}^{-1}|_{\beta\beta}$$

and that

$$|G_1|_{\beta\beta} \leq [\delta^2 |\Gamma_{11,0}|_{\beta\beta} + \delta^3 \max |\sigma_{\beta\beta}|_{\beta\beta}](1 + \delta_{\epsilon_0}) .$$

Consequently

$$|H_2 G_1 H_2^T H_1^{-1}|_{\beta\beta} \leq \epsilon_2$$

and by Assumption 3, $\epsilon_2 < 1$. So

$$|[(\alpha_{\beta\beta}^0)^T (\Gamma_{\beta\beta}^0)^{-1} (\alpha_{\beta\beta}^0)]^{-1}|_{\beta\beta} \leq |H_1^{-1}|_{\beta\beta} (1 + \delta_{\epsilon_0}) ,$$

where

$$H_1^{-1} = [\sum (\Sigma_{\beta\beta}^{-1})]^{-1} .$$

Inequality (3.12) now follows

Now

$$|\alpha^*|_{\beta\beta} \leq |\alpha_{\beta\beta}^0|_{\beta\beta} + |\Gamma_{\beta\beta}^0|_{\beta\beta} |(\Gamma_{\beta\beta}^0)^{-1}|_{\beta\beta} |\alpha_{\beta\beta}^0|_{\beta\beta} .$$

But it is easily shown that $|\alpha_{\beta\beta}^0|_{\beta\beta} = (n - m)^k$ and that $|\alpha_{\beta\beta}^0|_{\beta\beta} \leq m^k$. Furthermore, in the proof of Lemma 7, it is shown that

$$|(\Gamma_{\beta\beta}^0)^{-1}|_{\beta\beta} \leq |\Sigma_{\beta\beta}^{-1}|_{\beta\beta} (1 + \delta_{\epsilon_0}) .$$

Finally, since the function $x \exp[-x]$, $x \geq 0$ has a unique maximal value at $x = 1$,

$$|\Gamma_{\beta\beta}^0|_{\beta\beta} = |\eta_{\beta\beta}|_{\beta\beta} \\ \leq \delta \Delta \exp[-(\delta + \Delta/\Delta_0)] |\Gamma_{11,0}|_{\beta\beta} .$$

Inequality (3.13) obtains.

To complete the proof, observe that

$$(\Gamma_{\beta\beta}^0)^{-1} = (\Gamma_{\beta\beta}^0)^{-1} (I - G)^{-1} ,$$

where

$$G = \Gamma_{RP}^0 (\Gamma_{P\beta}^0)^{-1} (\Gamma_{RP}^0)^T (\Gamma_{R\beta}^0)^{-1}$$

so

$$|G|_{RR} \leq |\eta_{RP}|_{PR}^2 |\Sigma_{PP}^{-1}|_{PP} |\zeta_{RR}^{-1}|_{RR}.$$

But

$$|\eta_{RP}|_{PR} \leq \delta \Delta \exp[-(\delta + \Delta/\Delta_0)] |\Gamma_{110}|_{SS},$$

$$|\Sigma_{PP}^{-1}|_{PP} = \max |\Sigma_{Pj}^{-1}|_{SS},$$

and

$$|\zeta_{RR}^{-1}|_{RR} \leq \frac{1}{\Delta} \Delta^{-3} \max |\sigma_{Ri}^{-1}|_{SS}.$$

Thus

$$|G|_{RR} \leq c_3$$

and

$$|(\Gamma_{RP}^0)^{-1}|_{PR} \leq \frac{1}{\Delta} \Delta^{-3} |\sigma_{RR}^{-1}|_{RR} (1 + \delta_{c_3}),$$

which gives inequality (3.14). Inequalities (3.10) and (3.11) are immediate consequences of Theorem 1 and the proof is complete. \square

We now summarize our results. Suppose the proposed first order Taylor approximation has been adopted together with the covariance approximation in equation (3.5) and its extension described just above Theorem 10. Add assumptions 1, 2 and 3 together with the condition $\delta \leq \Delta_0 \leq \Delta$ where Δ_0 is the scaling parameter in the covariance approximation (see (3.5)) and Δ is the smallest value of $|\epsilon_{Ri} - \tau|$, $i \in M_3^*$. Then

$$||\beta_T - \beta_P^0||_{PS} \leq ||\beta_T - \beta_P||_{PS} + ||\beta_P - \beta_P^0||_{PS} + ||\beta_P^0 - \beta_P^0||_{PS},$$

and

$$|\Gamma_{\beta T} - \Gamma_{\beta Y}^0|_{SS} \leq |\Gamma_{\beta T} - \Gamma_{\beta P}|_{SS} + |\Gamma_{\beta P} - \Gamma_{\beta P}^0|_{SS} + |\Gamma_{\beta P}^0 - \Gamma_{\beta Y}^0|_{SS}.$$

Uniform bounds for the first, second, and third terms, respectively, on the right hand sides of each of these equations are obtained from Theorems 1, 8 and 9.

4. DISCUSSION

The analysis of Section 3 leads to a modification of the Bayes linear smoother proposed by Weerahandi and Zidek (1988; hereafter, WZ for simplicity). The residual covariance of the data with respect to the parameter of inferential interest, β is given locally by equation (3.5) and extended in the discussion following Theorem 9. The

setting of Section 3 is more specialized than that of WZ who include any available derivatives in their structural model for the data. On the other hand, their smoother is restricted to the case of scalar valued processes.

To describe the modified Bayes linear smoother, suppose for expository simplicity, that

$Y(t)$ is a scalar valued process. Where only continuity is assumed, WZ use the structural, $Y(t) = \beta + E$, where $\beta = S(\tau)$, $E = E(t, \tau)$, and E is the noise plus the remainder from the zero-th order Taylor expansion of $S(t)$ about $t = \tau$.

The Taylor remainder would be small when $t \approx \tau$. Given data, $Y = (Y_1, \dots, Y_n)$, where $Y_i = Y(t_i)$, and a prior distribution on β , an estimator of β is easily obtained once $\Gamma_{Y\beta}$ has been specified, if β and the vector of residuals, E_i , $i = 1, \dots, n$ are uncorrelated as assumed by WZ.

Our analysis leads us to conclude that the $\Gamma_{Y\beta}$ of WZ, while appropriate in certain cases may be unsatisfactory in general since it does not incorporate the correlation among the $E_i = E(t_i, \tau)$, a potentially important component of the model when the variance of the noise is small. Our analysis points to an appropriate modification of $\Gamma_{Y\beta}$ and shows that the resulting Bayes linear estimator of β based on the data, P , in a window about τ , approximates well, every member of a large class of Bayes linear estimators acting in the same window.

This last result supposes, however, that the prior distributions for members of this class have the same value of Γ_{110} , i.e. first term in the expansion of their $\Gamma_{P\beta}$'s, where

$$\Gamma_{110} = E[S^{(1)}(\tau) - \alpha_{10}S(\tau)]^2,$$

and α_{10} gives the best linear predictor of $S^{(1)}(\tau)$ based on $S(\tau)$. In the case of a weakly stationary S - process, Γ_{110} is just the variance of $S^{(1)}(\tau)$; this will be zero if and only if one believes the process is locally constant, in fact zero, at $t = \tau$. The implication of this belief would then be that the residuals from the Taylor expansion and hence the E_i are uncorrelated (to first order). The implications of this observation for our modification of the WZ smoother will be described precisely below. Qualitatively, it means that $\Gamma_{P\beta}$ is "small" intuitively speaking; so that the data corresponding to the t_i near τ are heavily weighted in any of the Bayes linear smoothers sharing this lack of first order correlation, as well as in the modified WZ approximant to these smoothers; more remote data are not "pulled in" as they would be when the first order correlation terms were nonzero.

When S is thought to have more than three derivatives, higher order terms in the residual correlation structure can be obtained and these depend on high order analogues of Γ_{110} . These quantities appear to jointly index the "curviness" of S in some way, but we do not fully understand their role even in the simple case under discussion where S is scalar valued.

To examine the modified WZ smoother in more detail we assume that the noise process has constant variance so for all i , $\Sigma_{\alpha_i \alpha_i} = \sigma_N^2$ and that $\sigma_{\alpha_i \alpha_i} = \sigma_N^2$ for $x = P, R$ (see equation (3.5) and the discussion immediately following Theorem 9 for the definition of these objects). With these simplifications, it is readily shown that

$$\beta_T^0 = (\Gamma_{\beta\beta}^0)^{-1} + \sum w_i)^{-1} (\sum w_i Y_i) \quad (4.1)$$

which is a locally weighted moving average with weights which capture the effects of the bandwidth or "window" parameter, σ_N^2 , the noise parameter, σ_N^2 , and so on. To be precise,

$$w_i = [1 - \zeta_c \Delta_i \exp(-|\Delta_i|/\Delta_0)]/\Lambda_i,$$

where

$$\zeta_c = [\sum \Delta_j \exp(-|\Delta_j|/\Delta_0)/\Lambda_j][\gamma_{11,0}^2 + \sum \Delta_j^2 \exp(-2|\Delta_j|/\Delta_0)\Lambda_j^{-1}]^{-1},$$

$$\Lambda_j = \sigma_N^2 + 2\sigma_N^2 |\Delta_j|^3.$$

$\Delta_j = t_j - \tau$, and $\Delta_0 > 0$ is a parameter which is to be chosen by the analyst. A large value of Δ_0 means that the correlation of the residuals from the Taylor remainder will tend to persist over a broader range.

Observe that when $\Gamma_{11,0} = 0$, $\zeta_c = 0$ so that $w_i = \Lambda_i^{-1} = (\sigma_N^2 + 2|\Delta_i|^3\sigma_N^2)^{-1}$ which leads to a simple locally weighted moving average. This emphasizes the comparative complexity of the smoother when $\Gamma_{11,0} \neq 0$.

In the extreme case, when σ_N^2 is comparatively large, $w_i = \sigma_N^{-2}$. If in addition, prior knowledge about β is vague and hence, $(\Gamma_{\beta\beta}^0)^{-1} = 0$, then not surprisingly, $\beta_T^0 = \bar{Y}$, the sample average. Here the WZ smoother is essentially the moving average.

At the other extreme, when $\sigma_N^2 = 0$, noise is non-existent. In this case, if $\tau = t_i$ for some i , it is readily seen that $\beta_T^0 = Y_i$ so that β_T^0 switches from being a smoother to being an interpolator as would be expected on intuitive grounds.

It is obvious that β_T^0 shrinks toward 0 because the data were centered a priori. To return to the natural origin, we should express this smoother

$$\beta_T^0 = [m(\tau) - \sum w_i^* m(t_i)] + \sum w_i^* Y_i,$$

where $w_i^* = (\Gamma_{\beta\beta}^0)^{-1} + \sum w_i)^{-1} w_i$. It should be emphasized that the process mean must have been specified, in any case.

The analysis in Section 3 shows that the modified WZ smoother approximates well, each of the members of a class of Bayes linear smoothers when the observations are sufficiently dense around the point at which smoothing is to be carried out. But its potential merit to a Bayesian investigator can be assessed quite independently of any merit which may be ascribed as a result of the analysis of Section 3. That is the spirit in which the methods of Weerahandi and Zidek (1986, 1988) were proposed. If, however, the observations are deemed to have been made on a continuous process,

coherence dictates that this feature of the process be embraced as we have tried to do here.

The assumption that the process is wide sense stationary with uncorrelated residuals has the dramatic implications shown in Lemma 5 and Theorem 7. All the measurements may be discarded except the two (or one in the case of forecasting) made at the times (time) adjacent to τ may be discarded without any approximation error in going from Y down to P . Moreover, the resulting Bayes linear smoother is very simple, in fact, that originally developed in classical time series. The first order spline is obtained as a limiting case. There would seem to be no point to using the zeroth order Taylor approximation in this case, but if one did, the local residual covariance structure would be a block diagonal matrix, the blocks corresponding to the t_i 's below and above τ , respectively. The elements in the diagonal blocks would be close to zero, away the diagonal and the $i-j$ th elements near the diagonal would be approximately a constant multiple of $\min(|t_i - \tau|, |t_j - \tau|)$ as is easily shown by expanding the elements of the covariance transformation of the $AR(1)$ process. We have not investigated the quality of that approximation.

The obvious nonrobustness of the procedure derived from the $AR(1)$ process in the last paragraph to recording errors and model misspecification leads to the main body of work in Section 3. If a continuous process is also differentiable, a possibility which would be hard to rule out on the basis of a priori knowledge, then as the analysis of Section 3 (Lemma 5, in particular) shows, the local residuals from fitting $S(t_i)$ to $\beta = S(\tau)$, for the t_i 's near τ , are correlated and the Bayes linear smoother will depend on more than merely the one or two measurements at points nearest τ . The results of Section 3 show that the proposed approximation will be good if the observation points, t_i , near τ are sufficiently dense and S is thrice differentiable or rather, if the slightly stronger condition that $\Gamma^{(3,3)}$ be continuous holds. This justification for the use of the approximation is analogous to that provided by large sample optimality for the use of classical procedures in samples of moderate size. The hyperparameters would need to be specified, to determine precisely how dense the observation points t_i would need to be. Usually the needed hyperparameter values would be difficult to assess. In fact, if they could be determined, there would be no point to the proposed approximation; the Bayes linear procedure itself would presumably be then available. The point of the analysis is to show that for a fairly large class of Bayes linear smoothers, our proposed procedure will serve as an approximation whose error diminishes in a precisely determined manner as the data's observation grow increasingly dense.

A desirable feature of our approximation is that it is not fully automatic; it does allow prior information about $S(\tau)$ at $\tau = \tau$ to be incorporated and this would be very important especially when the data are sparse. A fully automatic, so-called "objective" procedure could be at a considerable disadvantage under these circumstances.

By pasting together approximations for $\alpha_{R,P}R$ and $\alpha_{P,P}P$ we obtain an approximation, β_P^0 , based on all of Y thereby eliminating the need, in practice, to decide what of Y to put in each of P and R . But the distinction is of fundamental importance in establishing the asserted quality of β_P^0 .

If S is supposed to be differentiable, say p times, then it would seem natural to incorporate these derivatives into the approximation data regression model explicitly to reduce the correlation in the residuals. This can be done directly if the values of $S^{(j)}(t_i)$, $j=1,\dots,p$ can be observed without noise for all i by taking $Y(t)$ as $(S^{(0)}(t), \dots, S^{(p)}(t))$. The theory developed in this paper is sufficiently general to encompass this case directly. Theorem 7 will apply if the newly defined Y is an $AR(1)$ process. Alternatively if for this new Y , $\Gamma^{(3,3)}$ is thought to exist, a zeroth order Taylor approximation might be used instead.

Alternatively when S 's derivatives cannot be observed, $S^{(j)}$, $j=1,\dots,p$ can nevertheless be incorporated through higher order Taylor expansions. This approach is taken by Weerahandi and Zidek (1986, 1988, 1990) and will not be discussed further here.

Although the discrete case where the t_i are equi-spaced is not discussed here, it should be emphasized that the proposed approximation can be used there as well, provided that the process which generates the discrete data may be viewed as at least continuous.

The proposed method is, roughly speaking, a Bayesian counterpart to the locally weighted method of Cleveland (1979) and a locally constant version of LOWESS which is locally linear (although here the weights are fixed). It turns, locally weighted regression methods are equivalent in a certain sense to kernel smoothing methods (Muller, 1987). But unlike automatic methods such as the kernel and locally weighted regression methods, the proposed procedure, does allow prior knowledge about $\beta_w^\Delta S(\tau)$ to be input very simply.

The context of Section 2 is much more general than that of Section 3 although, the results presented in the former are regarded as technical preliminaries to those of the latter. The extension involved in going from Section 3 to Section 2 is essentially that recognized by Sacks and Ylvisaker (1978) which has given way to the study of semi-parameter models (c.f. Heckman, 1988).

Section 2 presents two general approximations to Bayes linear procedures. The effect of the first, which results from not collecting or discarding certain data, is stated in Theorem 1 where the approximation error is specified. Since this error depends on hyperparameters which could not usually be specified, this theorem is mainly descriptive. It says that if a subvector, R , of the data vector, Y , is discarded so that only P is left, the error will be "large" if the residuals, $\beta \cdot P$, and $R \cdot P$ are highly correlated. The result is a precise index of what is usually an undefined product of an heuristic process. The Theorem does lead to Theorems 2 and 6 which may assist and refine this

heuristic process since they yield ideal conditions under which there is no approximation error.

The second and more technical approximation of Section 2 and its application in Section 3, derive from deliberate or accidental, misspecifications of the basic model hyperparameters. Our bounds on the resulting inferential errors hold only for hyperparameter misspecification errors of small or moderate size as in Section 2. Nothing can be said in general about the impact of gross errors which are ignored. The lowest order terms in Lemma 3 reveal the sensitivity of the Bayes linear rule to misspecifications of the hyperparameters and these are given in Theorem 4. Furthermore, Theorem 5 gives upper bounds on the size of the errors in β_Y and $\Gamma_{\beta,Y}$ induced by moderate misspecifications of hyperparameters like $\Gamma_{\beta\beta}$.

As is well known, if the posterior distribution of β given Y were normal its posterior mean, $\mu_{\beta|Y}$, which is often taken to be the Bayes estimator of β , is linear in Y ; and the posterior covariance, $\Gamma_{\beta,Y}$, is the matrix of residual errors of estimation. Moreover, under the hypothesized normal posterior distribution, this paper would yield approximation error bounds for estimators of attributes derived from $\mu_{\beta|Y}$ and $\Gamma_{\beta,Y}$, like quantiles in the univariate case. However, we have not investigated the normal posterior in detail, partly because of our conviction that it will seldom obtain in practice because of uncertainty about parameters in the normal covariance which must be "marginalized out" to obtain the posterior distribution of β given Y . This marginal distribution, on which the analysis of this paper is based, will not typically be normal and the Bayes estimator of β will not then be linear in Y . In this paper we investigate Bayes linear estimators even though may not be globally Bayesian, because of their simplicity and consequent general appeal.

It should be emphasized that the topology on which our analysis are based gives rise to an extremely severe test of the modification of the WZ procedure in that convergence of point estimators to be uniform in a certain sense. This seems desirable in some respects and the topology is convenient in that it allows us to treat the convergence of operators and point estimators simultaneously. But extremely strong requirements must be met in order to obtain the conclusions of our theorems giving bounds on the errors of approximation. This can be seen very clearly in Theorem 10 where the conditions essentially demand that R and P must be well separated to insure that the bounds obtain (alternately, σ_{RR} must be extremely large). In future work, the effect of choosing weaker topologies will be explored.

5. SIMULATION

To obtain a numerical assessment of the performance of the WZ approximation and the exact Bayes linear estimator we have carried out a limited simulation study with a particular scalar valued process. For simplicity, suppose

$$S(t) = e^{tZ} = e^{(\delta^2/2)t^2} \quad (5.1)$$

is observed at $t = (t_1, \dots, t_n)$ with normally distributed noise

$$N(t) \sim N(0, \sigma^2) \quad , \quad (5.2)$$

for all t . Suppose $Z \sim N(0, \delta^2)$ so that the process has zero mean. It follows from the moment generating function of the normal distribution that

$$\text{Cov}(t_i, t_j) = e^{\delta^2(t_i^2 + t_j^2)/2} \left[e^{\delta^2 t_i t_j} - 1 \right] \quad . \quad (5.3)$$

The observed process is $Y(t) = S(t) + N(t)$. For the purpose of the simulation all the parameters of these processes are specified below.

Let τ be a point of inference. The objective of the simulation is to observe how the exact Bayes linear estimator and the WZ approximation converge to the true value of $\beta = S(\tau)$ as the density of sampling points around τ increases. The exact Bayes linear estimator given by (1.3) can be computed as

$$\beta_Y = \Gamma_{SS}(\Gamma_{SS} + \sigma^2 I_n)^{-1} Y \quad ,$$

where $Y = (Y(t_1), \dots, Y(t_n))$, $S = (S(t_1), \dots, S(t_n))$, and the covariance matrices are computed using (5.3). The WZ approximation β_Y^0 can be easily computed using (4.1).

Since δ is a scale parameter for t , it was fixed at 1. A reasonably dispersed sample of the S process is obtained by first generating 10 equally spaced values over the $[0,1]$ interval. The points of inference, τ , were chosen to be .25, .50 and 1.05. The density of data points around $\tau=.25$ is increased in two steps by doubling n twice; the density of sampling points around .50 and 1.05 were not increased to enable us to see how the increased density at a remote sampling point can affect the WZ smoother. Specifically, the S process was generated at

$$t_1 = (.1, .2, \dots, 1.0), \quad t_2 = t_1 \cup \{.2 + (t_1/10)\}, \quad \text{and}$$

$$t_3 = t_2 \cup \{.21 + (t_1/10)\} \cup \{.22 + (t_1/10)\} \quad ,$$

respectively. A large noise generated from (5.2) with $\sigma=0.25$ is added to each of the unobservable S values. Shown in Table 1 is a particular set of data generated when $n=10$ and $Z=-.88$. Notice that the noise is large enough to mask the monotonicity in S values.

-- PLACE TABLE 1 ABOUT HERE --

For each value of t the simulation is repeated with five Z values, namely with $Z = -.88, .80, -.29, -.43$, and 1.14 generated from the $N(0,1)$ distribution. The values of $\beta(t)$ for these values of Z and for $\tau = .25, .5$ and 1.05 are shown in Table 2. Shown in the same table are the exact Bayes linear estimates, β_Y , and the WZ approximation, β_Y^0 ; in computing the latter using (4.1), Δ_0 and σ_{Δ_0} were each set at $.5$. The values of $||\Delta\Gamma_{\beta_Y}||^2 = (\Gamma_{\beta_Y} - \Gamma_{\beta_Y^0})^2$ are also displayed for each value of τ and t , where Γ_{β_Y} and $\Gamma_{\beta_Y^0}$ serve to measure the reliability of the exact Bayes linear estimator and the WZ approximation respectively.

-- PLACE TABLE 2 ABOUT HERE --

It can be observed that the exact Bayes linear estimator performs fairly well even for sample sizes as small as 10. Of course in this simulation study, for the purpose of obtaining insight of what is happening, we have assumed all the parameters of the underlying processes to be known. In practical applications this will not really be the case and therefore additional stages of the hierarchical Bayesian specifications will be required. Notice that because most of the points are remote from $\tau = 1.05$, the relative accuracy of the WZ approximation is least at this point of inference. When the density of data around $\tau = .25$ is increased both the exact Bayes linear estimator and the WZ approximation tend to improve and converge towards the true value $\beta(.25)$. The exact Bayes linear estimate at $\tau = 1.05$ is almost unaffected by increased density around $\tau = .25$. On the other hand, the increased density of remote sampling points has the expected adverse effect on the approximation β_Y^0 for $\beta(1.05)$. These effects are also reflected in $||\Delta\Gamma_{\beta_Y}||^2$. At the point $\tau = .25$ this quantity is dramatically reduced almost by ten fold at each doubling of the sample size.

REFERENCES

- Bellman, R. (1960). *Introduction to Matrix Analysis*. Toronto: McGraw-Hill Book Co., Inc.

- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatter-plots. *Journal of the American Statistical Association*, 74, 829-836.
- Eaton, M.L. (1983). *Multivariate Statistics: A Vector Space Approach*. Wiley, New York.
- Grenander, U. (1981). *Abstract Inference*. Toronto: John Wiley and Sons.
- Halmos, P. R. (1974). *Finite dimensional vector spaces*. New York: Springer Verlag.
- Heckman, N.E. (1988). Minimax estimates in a semiparametric model. *Journal of the American Statistical Association*, 83, 1090-1096.
- Lindley, D. V. and Smith, A.F.M. (1972). Bayes Estimates for the linear model(with Discussion). *Journal of the Royal Statistical Society, B*, 34, 1-41.
- Müller, Hans-Georg (1987). Weighted local regression and kernel methods for non-parametric curve fitting. *Journal of the American Statistical Association*, 82, 231-238.
- Sacks, J. and Ylvisaker, D. (1978). Linear estimation for approximately linear models. *Annals of Statistics*, 6, 1122-1137.
- Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H. (1989) Design and analysis of computer experiments. *Statistical Science*, 4, 409-435.
- Stone, M. (1987). *Coordinate-free multivariable statistics*. Oxford: Oxford University Press.
- Taylor, A. E. and Lay, D. C. (1980). *Introduction to Functional Analysis*. New York: John Wiley and Sons.
- Wahba, G. (1983). Bayesian "confidence intervals" for the cross validated smoothing spline. *Journal of the Royal Statistical Society, B*, 45, 133-150.
- Weerahandi, S. and Zidek, J. V. (1986). Analyses of multiple time series by Bayesian and empirical Bayesian nonparametric methods. SIMS Tech. Rep. No. 96, Department of Statistics, University of British Columbia.
- Weerahandi, S. and Zidek, J.V. (1988). Bayesian nonparametric smoothers for regular processes. *The Canadian Journal of Statistics*, 16, 1, 61-73.

TABLE 1: Observed Values of S process, Noise N, and $Y = S + N$										
t	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
S(t)	-0.09	-0.18	-0.28	-0.38	-0.49	-0.61	-0.74	-0.88	-1.04	-1.23
N(t)	0.33	-0.22	-0.48	0.09	0.01	0.03	-0.74	0.17	0.02	-0.20
Y(t)	0.24	-0.40	-0.76	-0.29	-0.48	-0.58	-1.48	-1.05	-1.02	-1.43

TABLE 2

Approximate (β_Y^0) and Exact (β_Y) Bayes Linear Estimates with $t=t_1$									
	$\tau = .25$			$\tau = .50$			$\tau = 1.05$		
	$ \Delta\Gamma_{\beta_Y} ^2 = .00040$			$ \Delta\Gamma_{\beta_Y} ^2 = .000012$			$ \Delta\Gamma_{\beta_Y} ^2 = .029$		
Z	$\beta(\tau)$	β_Y	β_Y^0	$\beta(\tau)$	β_Y	β_Y^0	$\beta(\tau)$	β_Y	β_Y^0
-.88	-0.23	-0.15	-0.46	-0.49	-0.36	-0.51	-1.34	-1.36	-0.61
0.80	0.19	0.13	0.37	0.36	0.28	0.39	0.59	0.71	0.43
-.29	-0.10	-0.10	-0.33	-0.27	-0.25	-0.36	-1.00	-1.02	-0.44
-.43	-0.13	-0.12	-0.35	-0.33	-0.29	-0.40	-1.10	-1.15	-0.49
1.14	0.30	0.22	0.64	0.64	0.50	0.70	1.58	1.79	0.83

Approximate (β_Y^0) and Exact (β_Y) Bayes Linear Estimates with $t=t_2$									
	$\tau = .25$			$\tau = .50$			$\tau = 1.05$		
	$ \Delta\Gamma_{\beta_Y} ^2 = .000047$			$ \Delta\Gamma_{\beta_Y} ^2 = .000018$			$ \Delta\Gamma_{\beta_Y} ^2 = .032$		
Z	$\beta(\tau)$	β_Y	β_Y^0	$\beta(\tau)$	β_Y	β_Y^0	$\beta(\tau)$	β_Y	β_Y^0
-.88	-0.23	-0.17	-0.33	-0.49	-0.39	-0.37	-1.34	-1.34	-0.46
0.80	0.19	0.14	0.26	0.36	0.29	0.27	0.59	0.70	0.32
-.29	-0.10	-0.10	-0.20	-0.27	-0.25	-0.23	-1.00	-1.03	-0.30
-.43	-0.13	-0.11	-0.21	-0.33	-0.27	-0.24	-1.10	-1.16	-0.32
1.14	0.30	0.24	0.47	0.64	0.54	0.51	1.58	1.76	0.62

Approximate (β_Y^0) and Exact (β_Y) Bayes Linear Estimates with $t=t_3$									
	$\tau = .25$			$\tau = .50$			$\tau = 1.05$		
	$ \Delta\Gamma_{\beta_Y} ^2 = .0000056$			$ \Delta\Gamma_{\beta_Y} ^2 = .000024$			$ \Delta\Gamma_{\beta_Y} ^2 = .034$		
Z	$\beta(\tau)$	β_Y	β_Y^0	$\beta(\tau)$	β_Y	β_Y^0	$\beta(\tau)$	β_Y	β_Y^0
-.88	-0.23	-0.20	-0.29	-0.49	-0.43	-0.31	-1.34	-1.32	-0.36
0.80	0.19	0.14	0.21	0.36	0.29	0.22	0.59	0.70	0.25
-.29	-0.10	-0.10	-0.16	-0.27	-0.25	-0.17	-1.00	-1.03	-0.22
-.43	-0.13	-0.11	-0.16	-0.33	-0.27	-0.17	-1.10	-1.17	-0.23
1.14	0.30	0.26	0.39	0.64	0.57	0.42	1.58	1.74	0.49

REFERENCES

- Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, 74, 829-836.
- Cleveland, W. S. and Devlin, S.J. (1988). Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American Statistical Association*, 83, 596-610.
- Harrison, J. and West, M. (1989)
- Heckman, N.E. (1988). Minimax estimates in a semiparametric model. *Journal of the American Statistical Association*, 83, 1090-1096.
- Joe, H., Ma, Wilson, and Zidek, J.V. (1986). A Bayesian nonparametric univariate smoothing method with applications to acid rain data analysis. SIMS Technical Report # 47, Department of Statistics, University of British Columbia.
- Lindley, D. V. and Smith, A.F.M. (1972). Bayes Estimates for the linear model (with Discussion). *Journal of the Royal Statistical Society, B*, 34, 1-41.
- Muller, Hans-Georg (1987). Weighted local regression and kernel methods for non-parametric curve fitting. *Journal of the American Statistical Association*, 82, 231-238.
- Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *Journal of the American Statistical Association*, 83, 1134-1143.
- O'Hagen, A. (1978). Curve fitting and optimal design for prediction (with discussion). *Journal of the Royal Statistical Society, Ser. B*, 40, 1-42.
- Sacks, J. and Ylvisaker, D. (1978). Linear estimation for approximately linear models. *Annals of Statistics*, 6, 1122-1137.
- Sacks, J., Welch, W.J., Mitchell, T.J., and Wynn, H. (1989). Design and analysis of computer experiments. *Statistical Science*, 4, 409-435.
- Wahba, G. (1983). Bayesian "confidence intervals" for the cross validated smoothing spline. *Journal of the Royal Statistical Society, B*, 45, 133-150.
- Weerahandi, S. and Zidek, J. V. (1986). Analyses of multiple time series by Bayesian and empirical Bayesian nonparametric methods. SIMS Tech. Rep. No. 96, Department of Statistics, University of British Columbia.
- Weerahandi, S. and Zidek, J.V. (1988). Bayesian nonparametric smoothers for regular processes. *The Canadian Journal of Statistics*, 16, 1, 61-73.
- Weerahandi, S. and Zidek, J. V. (1990). Approximate Bayes linear smoothers for continuous processes. Submitted