A METHOD OF ESTIMATING THE COEFFICIENTS IN DIFFERENTIAL EQUATIONS FROM TIME-DISCRETE OBSERVATIONS

by

Walter E. Bell and H. R. van der Vaart

Institute of Statistics
Mimeograph Series No. 845
September 1972

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

1. INTRODUCTION

Differential equations of the form

$$\frac{dy_j}{dt} = \sum_{i=1}^{k} \kappa_{ji} \, y_i, \; j = 1, \ldots, \ell, \; 1 \le \ell \le k \tag{1.1}$$

are frequently used in the mathematical models associated with biological systems. For example, differential equations of this type often appear in the models of drug disposition in the human body. Certain models of population dynamics also employ differential equations related to (1.1). Assuming the relevance of (1.1) to the true underlying biological phenomena, interest has been traditionally focused on estimates of the values of the κ_{ji} in most experimental applications of this model. For observed values $Y_i(t_m) = y_i(t_m) + \varepsilon_i(t_m)$, where $\varepsilon_i(t_m)$ is a random error and $m = 1, 2, \ldots, n$, estimation of the κ_{ji} has been accomplished in most of the existing literature by a two-step procedure:

(1) A solution function $\widehat{y_j(t)}$ of (1.1) is derived by analytical or numerical methods. Such a function is defined at least for the $t = t_m$.

(2) Values of the κ_{ji} are chosen to minimize the quantity

$$\sum_{m=1}^{n} \left[Y_j(t_m) - \widehat{Y_j(t_m)} \right]^2 \tag{1.2}$$

where $\widehat{Y_j(t_m)}$ denotes the value of $\widehat{y_j(t_m)}$ for a given set of observations.

Estimates of the κ_{ji} derived from minimization of (1.2) are usually referred to as "least-squares" estimates.

Except in conjunction with certain enzyme kinetic models (Cleland, 1967), apparently little work has been done on estimating the κ_{ji}

without solving the differential equation. The few examples of existing methods appearing in the literature involve estimating the value of the derivative $\frac{dy_j}{dt}$ at the t_m from experimental observations by such methods as difference schemes (e.g., see Rescigno and Segre (1961, p. 9)), and, by choosing the κ_{ji} to minimize

$$\sum_{m=1}^{n}\left\{\left[\sum_{i=1}^{k} \kappa_{ji}\; Y_i(t_m)\right] - \widehat{\dot{y}_j(t_m)}\right\}^2 \tag{1.3}$$

where $\widehat{\dot{y}_j(t_m)}$ denotes the estimates of $\left.\frac{dy_j}{dt}\right|_{t=t_m}$, estimating the κ_{ji} in a manner analogous to that employed in linear least-squares approaches to multiple regression. However, the estimation of the values $\left.\frac{dy_j}{dt}\right|_{t=t_m}$ from the discrete observations $Y_j(t_m)$ in the absence of an analytical solution $y_j(t)$ has traditionally required considerable care in the choice of method. (See, for example, Joksch (1966), who discusses the effects of random observational errors on the derivative estimates derived from the least-squares fitting of an approximation function to the observed values.)

The purpose of this thesis is to formalize a method of estimating the κ_{ji} in (1.1) which is based on minimization of a quantity, the expression for which is related to a discrete version of the Sobolev norm. For several differential equations of simple form, some of the properties of the derived estimators κ_{ji} are investigated by both approximate analytical and Monte Carlo simulation methods. To insure relevance to investigations of mathematical models related to biological phenomena, the number, n, of observations is limited and the distributions associated with the observational errors, ε, are chosen to permit study under both constant variance and variance proportional to the $y_i(t_m)$.

A review of some of the literature involving estimation of the coefficients of differential equations of particular biological significance is the subject of Chapter 2. Chapter 3 contains definitions and construction methods for 5-point moving-arc polynomial smoothing and derivative estimation and for a particular type of spline function composed of cubic segments. The Sobolev norm-type expression and several parameters useful for comparing estimators are also introduced in Chapter 3. In Chapter 4, the estimator of κ_1 in the simple differential equation

$$\frac{dy}{dt} = \kappa_1 y$$

is derived and investigated. Extensions of the method derived in Chapter 4 include estimation of κ_1 and κ_2 in

$$\frac{dy}{dt} = \kappa_1 y + \kappa_2 z$$

and the estimation of the κ_{ji} in the system

$$\begin{aligned} \frac{dy}{dt} &= \kappa_{11} y + \kappa_{12} z \\ \frac{dz}{dt} &= \kappa_{21} y + \kappa_{22} z , \end{aligned} \tag{1.4}$$

where there may be, but not necessarily be, restrictions on the κ_i . These extensions and examples of their use constitute Chapters 5 and 6.

2. REVIEW OF LITERATURE

2.1 Estimation of Coefficients of Differential Equations

Nearly all of the current methods of estimating the coefficients of differential equations rely on either a knowledge of the form of the analytic solution or on the existence of the values of a numerically generated solution at the points of interest, *e.g.*, the points corresponding to experimental observations. In fact, in the area of tracer kinetics, the methods of estimating the coefficients are generally described as the estimation of the λ_i ("rate constants") in the expression

$$f(t) = \sum_{i=1}^{m} A_i \exp(-\lambda_i t) .$$

For homogeneous ordinary linear differential equations with constant coefficients (whose solutions are essentially sums of exponentials), the coefficients of the differential equations appear as the λ_i or some function of these coefficients appear as the λ_i. In such cases, several methods of estimating the λ_i are available. The simplest is the common "peeling-off" technique described in detail by Perl (1960). Cook and Taylor (1971) describe a computerized peeling technique for application to radioactive tracer efflux data and report the results of simulation studies for two- and three-compartment systems for different λ_i and various relative errors. Since the peeling process assumes that the λ_i are sufficiently different to provide a linear terminal segment, the observations must be carried out far enough in time, which can be experimentally difficult.

For these models of radioactive tracer dynamics, Gardner (1963) proposes a Fourier transform method which is described by Pizer et al. (1969). Use of this method requires interpolation and extrapolation of the data to accommodate the integration schemes employed. For application to biological experiments with their large errors and short duration, this method may fail.

Other methods of estimating coefficients generally involve seeking values of the coefficients which, for a given set of data, minimize the sums of the squares of deviations

$$SS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

where the y_i are the observations and the $\hat{y}_i$ are the values of the solution (analytic or numerical). While not often specifically stated, the analog computer procedures generate curves for a given set of parameter values; each generated curve is compared with the data points in an attempt to determine the "best fit", resulting essentially in a minimization effort applied to the sums of squares. The coefficients κ_{ij} in the differential equation, say

$$\frac{dq_j}{dt} = \sum_{i=1}^{k} \kappa_{ij} q_i$$

which is common in tracer kinetics, are usually represented by variable potentiometers on the analog computer. Varying the κ_{ij} , one attempts to generate a curve which fits the observations more closely. Heinmets (1970) discusses in detail and expounds the virtues of the use of analog computers in biological model evaluation.

Analog computer notation and symbolism are used in such package programs as the IBM Continuous System Model Program, which is used by Parrish and Saila (1970) to estimate the coefficients of a system of Lotka-Volterra competition-type differential equations and a modified Volterra predator-prey equation. Evert and Randall (1970), encourage the use of the matrix approach to systems of differential equations in the context of tracer kinetics and cite the value of the Continuous System Model Program.

When an analytic solution to a differential equation (or a system of differential equations) is not available or when the solution is not linear in the parameters to be estimated, then a non-linear regression scheme is frequently employed to determine estimates of the parameters which minimizes the sums of squares of deviations. Rosenbrock and Storey (1966) and Swann (1969) present reviews of current methods, most of which are iterative and employ a linear approximation at every step. Nearly all such procedures require initial estimates of the parameters and can experience difficulty in convergence. Further, in the case of differential equations for which analytic solutions are not available, numerical integration procedures are a necessary part of the regression technique.

Cannon and Filmer (1967) describe in rigorous detail some of the mathematical properties of the estimates of the rate constants found in chemical kinetic models. The authors provide a numerical example (Cannon and Filmer, 1968) in which the initial conditions and the error rate are varied to demonstrate the behavior of their estimation method. Their original system of simultaneous differential equations is solved by finite differences while their norm

$$\int [(f(t) - C(t))]^2 dt$$

is evaluated by the trapezoidal integration procedure.

Procedures for estimating the coefficients of (a set of) ordinary differential equations without use of an integration scheme are apparently rare, with the exception of an important class of enzyme kinetic models. Since most enzyme kinetic experiments are performed by measuring initial reaction velocities at various initial substrate concentrations, the estimation of the rate constants K_i in the differential equation

$$v = \frac{dP}{dt} = f(S_1, S_2, \ldots; K_1, K_2, \ldots)$$

may be accomplished by a least squares scheme if the velocity equation is linear in the unknown rate constants or by a nonlinear technique if the unknown rate constants appear nonlinearly. In either case, the "observed" values of $v = \frac{dP}{dt}$ are made from either a continuous recording device or by fitting a function (e.g., a parabola) through the discrete observations on P, of which there are a large number (Cleland, 1967).

Whittle (1956) considers a time-dependent infection rate function, $\lambda(t)$, of wild rabbits in the integral equation

$$R(t) = \int_0^t \exp \{ \int_u^t [\lambda(v) - \beta(v)]\} \lambda(u)du$$

where $R(t)$ is the ratio of infected to healthy rabbits and $\beta(t)$ is the death rate due to infection and assumed to be 0.45. He notes that the above integral expression is a solution to the differential equation

$$\frac{dR(t)}{dt} + [\beta(t) - \lambda(t)]\, R(t) = \lambda(t) \; .$$

He proposes estimating $\frac{dR}{dt}$ by difference quotients, eye-smoothing the $R(t_i)$, and estimating $\lambda(t)$ at discrete points by

$$\lambda(t) = \frac{R'(t) + \beta(t)\, R(t)}{1 + R(t)} \; .$$

Metzler et al. (1965), in their work with non-steady state tracer kinetics, propose a similar approach to estimating the transfer rate function $B(t)$ of sodium from the plasma to rumen of sheep in the differential equation

$$\frac{dA_2}{dt} = \frac{B(t)\; [A_1(t) - A_2(t)]}{N_2(t)}$$

where A_1 and A_2 are the specific activities of sodium in the plasma and rumen, respectively, and N_2 is the concentration of sodium in the rumen. The authors suggest that each of the N_2, A_1, and A_2 be smoothed by a moving-arc polynomial scheme, for instance, from which

$$\frac{dA_2}{dt}$$

can be estimated at the times at which observations were performed. Then, $B(t)$ can be estimated at the times of observation t_i by

$$B(t_i) = \frac{N_2(t_i)\; A_2'(t_i)}{A_1(t_i) - A_2(t_i)} \; .$$

In the context of drug kinetics, Martin (1967) proposes a method of determining κ , the rate constant for elimination of drug by all routes, from measurements of unchanged drug in the urine. Assuming first-order kinetics, he derives the differential equation

$$\frac{dD_u}{dt} = K(D_{u\infty} - D_u)$$

where D_u is the cumulative amount of unchanged drug in the urine to time t and $D_{u\infty}$ is the total cumulative amount of drug in the urine <u>after</u> excretion is complete. For times t after the absorption and distributive phase, K is estimated by considering the linear relation which results from substitution of

$$\frac{\Delta D_u}{\Delta t}$$

for the derivative. According to Martin, since $\frac{\Delta D_u}{\Delta t}$ estimates $\frac{dD_u}{dt}$ at the midpoint of the urine-collection interval Δt, the values of $\frac{\Delta D_u}{\Delta t}$ would have to be interpolated to yield estimates of $\frac{dD_u}{dt}$ at the end-points of interval Δt, to coincide with the times of observations of the quantities D_u. The author shows that, when the decline of $\frac{dD_u}{dt}$ is first-order, the substitution of $\frac{\Delta D_u}{\Delta t}$ would produce an error which would not exceed two percent even if Δt were as large as one half-life of the drug.

Rescigno and Segre (1961) note that the estimation of the rate constants K_1 and K_2 in

$$\frac{dX}{dt} = K_1 Q - K_2 X$$

where Q is the constant concentration of a substance in an external medium, can be accomplished by numerically estimating the derivative $\frac{dX(t)}{dt}$ at $t=t_i$ by a certain difference scheme and writing the differential equation in the form

$$\dot{X} = M - NX(t) = (1 \vdots X) \begin{bmatrix} M \\ N \end{bmatrix}$$

which is recognized to be a special case of the usual linear expression

$$Y = \beta_0 + \beta_1 X$$

in which $Y = \dot{X}$, $\beta_0 = K_1 Q$, and $\beta_1 = K_2$. Then estimates of M and N (hence, $\hat{K}_1$ and $\hat{K}_2$) can be obtained in the customary fashion for linear least squares models. The potential value of this type of approach is obvious if fairly accurate estimates of the derivatives can be obtained. In fact, Rosenbrock and Story (1966) remark that, if we could measure the derivatives directly, then estimation of the rate constants would not involve the solutions to the differential equations; however, the authors don't pursue their remark.

2.2 Investigation of Observational Error Effect

Although many authors have dealt with the techniques of estimation of rate constants, and, hence, of the coefficients or functions of coefficients of differential equations, relatively few authors have reported the results of simulation studies in which the performance of the various estimation schemes was investigated under specific data error and rate constant magnitudes. In general, these studies involve simulated data which are constructed from hypothetical situations such as could be expected in biological applications, where the number of samples is limited and where the accuracy assumed is not normally found in non-biological situations.

In the context of tracer kinetics, Myhill (1967) reports on studies involving the sum of two exponentials with positive coefficients, i.e.,

$$f(t) = N_1 e^{-\lambda_1 t} + N_2 e^{-\lambda_2 t}$$

with $N_1, N_2, \lambda_1, \lambda_2 > 0$, for specified ratios of N_1/N_2 and λ_1/λ_2 under the limitation of 11 and 31 equally spaced points. Using a "valid least squares gaussian iterative" technique, Myhill found that for 11 points and a ratio of $\lambda_1/\lambda_2 = 2$ the technique did not converge for data sets with more than 1% error. A similar analysis of the tracer activity curve of a three-compartment steady-state open system, which is represented by a sum of three exponentials with positive coefficients, is reported by Myhill (1969). In this analysis, each point was weighted in order to reduce the differences in estimates between the cases of percentage error and constant error.

In work similar to that of Myhill, Glass and deGarreta (1967) report on error analyses in which a sum of two exponentials was fitted to generated data with error using the Marquardt method and a method based on the Newton-Raphson technique in which the weighting of data may be incorporated. Unfortunately, both the Myhill and the Glass and deGarreta results are unavoidably restricted to monotone decreasing functions.

Although Westlake (1971) did not conduct an extensive error investigation, his report appears to be the only available which has discussion on the effect of data error on a two-compartment open model which is represented by the sum of three exponential terms

$$A_i \, e^{-\lambda_i t}$$

in which the λ_i and the A_i are functions of the rate constants and the volumes of distribution. Such sums of exponentials are distinctly not monotone decreasing in the drug kinetic case in which the drug is orally administered. Westlake discusses, by means of an example, the effect of a constant error in the plasma concentrations on the estimation of the rate constants (and the functions of the rate constants) in the solution of the differential equations for the plasma concentration.

Cook and Taylor (1971) give tabular examples of the performance of their "peeling" computer routine in the estimation of rate constants for two- and three-compartment systems in a tracer efflux system.

3. THE SOBOLEV NORM-TYPE EXPRESSION, SMOOTHING AND DERIVATIVE ESTIMATION, AND ESTIMATOR COMPARISON

3.1 General Considerations Regarding Use of a Sobolev Norm-type Expression

In this chapter and the next, methods are formulated to estimate the constant coefficients of a certain class of first-order ordinary differential equations which are linear in the coefficients. Although the techniques may have more general application, the formulation will be restricted to a rather simple class of differential equations.

Consider the differential equation

$$\frac{dy(t)}{dt} = \kappa\, y(t)$$

where κ is a constant (*i.e.*, time-invariant) coefficient to be estimated. Observed values of y at times $t=t_i$ will be denoted by capital letters, *e.g.*, $Y(t_i)$.

Although most current techniques of coefficient estimation use some scheme which seeks the value of the coefficient κ which minimizes either the sums of squared deviations of a solution (analytic or numerical) from the observed values or, as in the example of Cannon and Filmer (1968), the integral of the squared deviations, the estimation procedure proposed in this discussion seeks to minimize the expression

$$\nu^2 = \sum_{i=1}^{n} \left\{ [Y(t_i) - f(t_i)]^2 + w[\dot{Y}(t_i) - g(t_i)]^2 \right\} \qquad (3.1)$$

which is somewhat analogous to the discrete version of the Sobolev norm. In equation (3.1), $\dot{Y}(t_i)$ denotes an "observed value" of the derivative $\frac{dy}{dt}$ expressed in terms of the observed values of y at the points t_i of observation by means of the model

$$\dot{Y}(t_i) = \kappa\, Y(t_i) ,$$

$w \geq 0$ is a constant (usually $w = 1$) with dimensions of time-squared, and $f(t_i)$ and $g(t_i)$ are approximations to $y(t_i)$ and $\dot{y}(t_i)$ obtained from the data by smoothing and derivative-approximation procedures. Then, in the current example,

$$\nu^2 = \sum_i \{[Y(t_i) - f(t_i)]^2 + w[\kappa\, Y(t_i) - g(t_i)]^2\}$$

and, not only an estimate of the unknown coefficient κ , but also any parameters in the functions f and g will have to be chosen so as to minimize the value of ν^2 . However, the model as written shows that the κ directly determines the values of the derivative of the function Y , so that it is natural to include in the selection criteria the minimization of the squared-deviations of derivatives of Y , as well as the values of Y itself.

Further, the incorporation of the differential equation itself in the quantity (3.1) to be minimized and the appearance of κ linearly in the differential equation permit estimation of κ instead of a function of κ , which often results when only a solution curve, in which the κ does not appear alone or linearly, is considered in the quantity to be minimized.

The functions f and g are arbitrary at this point, although several of their desirable potential properties are obvious:

(1) f and g should be defined at each time t_i , when the i^{th} observation is made.

(2) f and g should have smoothing properties, i.e., their use should reduce the effect of experimental error.

(3) f and g should be useful regardless of the "shape" of the graph of the underlying data function.

(4) g should be reasonably insensitive to experimental error.

(5) Ideally, f and g should be defined for all time, t, where $t_1 \leq t \leq t_n$.

Such properties suggest a number of potential forms of the functions f and g . For example, f could be the point function defined at the time t_i, i = 1, 2, ..., n, which results from application of a polynomial moving-average smoothing function and g could be the value of the first derivative with respect to time of the smoothing polynomial applied to the point t_i . On the other hand, g could be the point function defined at the t_i which results from some numerical differentiation technique such as a central differencing scheme. Of course, f could be the polynomial defined over the entire time domain of the data, a technique used by Lewi et al. (1970) to represent radiochemical data. These authors recognized the limitations imposed by fitting one polynomial to the entire set of discrete observations; for instance, a low degree polynomial may poorly approximate the data while a polynomial of high degree may better represent complex relationships in the data at the expense of smoothing and may also introduce excessive variation in the derivative estimates. One important group of functions which shares some of the desirable attributes of polynomials, which does not require the fitting of one polynomial over the entire range of the data but which is defined over all t such that $t_1 \leq t \leq t_n$ is a set of spline functions.

In this dissertation $f(t)$ is restricted to those functions of t which are linear in their parameters, e.g., (a) the function defined at t_i which is derived from 5-point moving-arc smoothing using either the cubic polynomial

$$p(t_i) = a_{oi} + a_{1i}t_i + a_{2i}t_i^2 + a_{3i}t_i^3 \qquad (3.2)$$

or the linear-hyperbolic function

$$q(t_i) = a_{oi}t_i + a_{1i} + a_{2i}\frac{1}{t_i} \qquad (3.3)$$

and (b) a spline-type function composed of cubic polynomial segments. Similarly, the function $g(t)$ is restricted to (a) the functions defined at t_i which are derived from the first derivative with respect to time of the two smoothing functions $p(t)$ and $q(t)$ and (b) the derivative of the segmented cubic polynomial.

3.2 5-point Moving-Arc Polynomial Smoothing and Derivative Estimation

3.2.1 Smoothing and Derivative Estimation Independent of Estimation of κ

Consider the n observations $Y(t_i)$, $i = 1, 2, \ldots, n$. Each of the $Y(t_j)$, $j = 3, 4, \ldots, n-2$, can be replaced by the value of the 5-point smoothing function (cubic or linear-hyperbolic) which is obtained by the fitting of the smoothing function to the five points $Y(t_i)$, $i = j-2, j-1, j, j+1, j+2$, by the method of linear least-squares. In the usual notation, the estimates of the $\underline{a}_i$ in equation (3.2) can be computed

$$\hat{\underline{a}}_i = (\underline{X}_j^T \underline{X}_j)^{-1} \underline{X}_j^T \underline{Y}_j$$

where, for the cubic smoothing function,

$$\underline{X}_j = \begin{bmatrix} 1 & t_{j-2} & t_{j-2}^2 & t_{j-2}^3 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & t_{j+2} & t_{j+2}^2 & t_{j+2}^3 \end{bmatrix}, \quad \underline{Y}_j = \begin{bmatrix} Y(t_{j-2}) \\ \cdot \\ \cdot \\ \cdot \\ Y(t_{j+2}) \end{bmatrix} .$$

Then, the smoothed estimates of $y(t_j)$ can be computed

$$Y^*(t_j) = [\, 1 \quad t_j \quad t_j^2 \quad t_j^3 \,] \; (\underline{X}_j^T \underline{X}_j)^{-1} \underline{X}_j^T \underline{Y}_j$$

$$= \underline{s}_j^T \underline{Y}_j$$

where $\underline{s}_j^T = [\, 1 \quad t_j \quad t_j^2 \quad t_j^3 \,] \; (\underline{X}_j^T \underline{X}_j)^{-1} \underline{X}_j^T$.

To obtain arbitrary smoothed values of the first two and last two observations, the vectors $\underline{s}_1^T$, $\underline{s}_2^T$, $\underline{s}_{n-1}^T$, and $\underline{s}_n^T$ are defined

$$\underline{s}_1^T = [\, 1 \quad t_1 \quad t_1^2 \quad t_1^3 \,] \qquad (\underline{X}_3^T \underline{X}_3)^{-1} \underline{X}_3^T$$

$$\underline{s}_2^T = [\, 1 \quad t_2 \quad t_2^2 \quad t_2^3 \,] \qquad (\underline{X}_3^T \underline{X}_3)^{-1} \underline{X}_3^T$$

$$\underline{s}_{n-1}^T = [\, 1 \quad t_{n-1} \quad t_{n-1}^2 \quad t_{n-1}^3 \,] \qquad (\underline{X}_{n-2}^T \underline{X}_{n-2})^{-1} \underline{X}_{n-2}^T$$

$$\underline{s}_n^T = [\, 1 \quad t_n \quad t_n^2 \quad t_n^3 \,] \qquad (\underline{X}_{n-2}^T \underline{X}_{n-2})^{-1} \underline{X}_{n-2}^T$$

Other possibilities, e.g., a lower degree polynomial, were employed to estimate the first two points but were rejected as no better than the make-shift method outlined here. Then, the vector of smoothed observations $\underline{Y}^*$ can be computed by $\underline{Y}^* = \underline{S}\,\underline{Y}$, where

$$\underset{(n\times n)}{\underline{S}} = \begin{bmatrix} \underline{s}_1^T & & & & & & 0 \\ \underline{s}_2^T & & & & & & \\ \underline{s}_3^T & & & & & & \\ 0 & \underline{s}_4^T & & & & & \\ & & \underline{s}_5^T & & & & \\ & & & \ddots & & & \\ & & & & \underline{s}_{n-3}^T & & 0 \\ & 0 & & & & & \underline{s}_{n-2}^T \\ & & & & & & \underline{s}_{n-1}^T \\ & & & & & & \underline{s}_n^T \end{bmatrix}$$

and where $\underline{Y}^T = \left[Y(t_1) \quad Y(t_2) \quad . \; . \; . \quad Y(t_n) \right]$.

Similarly, by defining the (1x5) vectors

$$\underline{d}_1^T = \left[0 \quad 1 \quad 2t_1 \quad 3t_1^2 \right] (\underline{X}_3^T \underline{X}_3)^{-1} \underline{X}_3^T$$

$$\underline{d}_2^T = \left[0 \quad 1 \quad 2t_2 \quad 3t_2^2 \right] (\underline{X}_3^T \underline{X}_3)^{-1} \underline{X}_3^T$$

$$\underline{d}_i^T = \left[0 \quad 1 \quad 2t_i \quad 3t_i^2 \right] (\underline{X}_i^T \underline{X}_i)^{-1} \underline{X}_i^T, \quad i=3,4,\ldots, n-2$$

$$\underline{d}_{n-1}^T = \left[0 \quad 1 \quad 2t_{n-1} \quad 3t_{n-1}^2 \right] (\underline{X}_{n-2}^T \underline{X}_{n-2})^{-1} \underline{X}_{n-2}^T$$

$$\underline{d}_n^T = \left[0 \quad 1 \quad 2t_n \quad 3t_n^2 \right] (\underline{X}_{n-2}^T \underline{X}_{n-2})^{-1} \underline{X}_{n-2}^T$$

a matrix $\underline{D}$ can be constructed so that $\dot{\underline{Y}}^* = \underline{D}\ \underline{Y}$ is a vector of estimates of $\left.\frac{dy}{dt}\right|_{t=t_i}$ derived from the 5-point moving-arc cubic smoothing.

In an obvious manner, matrices $\underline{S}$ and $\underline{D}$ can be constructed for 5-point moving averages using linear-hyperbolic functions or any power-polynomial function. Further, it is clearly not necessary to use smoothing functions of the same degree for all t_i .

The problem of choosing the $f(t_i)$, the $g(t_i)$, and κ to minimize equation (3.1) under the specification of 5-point moving-arc smoothing and subsequent derivative estimation results in minimization of

$$\nu^2 = (\underline{Y} - \underline{S}\ \underline{Y})^T (\underline{Y} - \underline{S}\ \underline{Y}) + w(\kappa\ \underline{Y} - \underline{D}\ \underline{Y})^T (\kappa\ \underline{Y} - \underline{D}\ \underline{Y}) \qquad (3.4)$$

where matrices $\underline{S}$ and $\underline{D}$ were defined above. From equation (3.4), it is evident that, for a given set of values of $Y(t_i)$ (including specific knowledge of the time-sequence $t_1, t_2, \ldots, t_n$) and for a given moving-arc smoothing polynomial, e.g., a cubic, the minimization of ν^2 in (3.4) involves choosing κ to minimize the last term,

$$w(\kappa\ \underline{Y} - \underline{D}\ \underline{Y})^T (\kappa\ \underline{Y} - \underline{D}\ \underline{Y}) \ . \qquad (3.5)$$

However, since the $Y(t_i)$ are assumed to contain an additive random error, it is possible to attempt to reduce the effect of this random error by applying the smoothing matrix $\underline{S}$ to the terms involving $\kappa\ \underline{Y}$ in (3.4) so that the expression to be minimized is

$$\nu^{*2} = w(\kappa\ \underline{S}\ \underline{Y} - \underline{D}\ \underline{Y})^T (\kappa\ \underline{S}\ \underline{Y} - \underline{D}\ \underline{Y}) \ . \qquad (3.6)$$

It will be shown in Chapter 4 that the minimization of (3.6) instead of (3.5) often does result in an estimator of κ which has greater variance and which is more subject to bias. However, for completeness, the discussion of Chapter 4 will involve equation (3.6) with references to equation (3.5) being immediate with $\underline{S} = \underline{I}$, the identity matrix.

3.2.2 Parameter Estimation with Simultaneous Estimation of κ

The case of 5-point moving-arc cubic smoothing of n values $Y(t_i)$, i=1, ..., n, involves the estimation of n sets of the coefficients $\{a_{oi}, a_{1i}, a_{2i}, a_{3i}\}$ of equation (3.2). Similar to the situation discussed in Section 3.2.1, minimization of the Sobolev norm-type expression under the condition of simultaneous estimation of the 4n parameters a_{ji} and κ requires the representation of the $f(t_i)$ by

$$[1 \quad t_i \quad t_i^2 \quad t_i^3] \begin{bmatrix} a_{oi} \\ a_{1i} \\ a_{2i} \\ a_{3i} \end{bmatrix} .$$

However, in this case the vectors

$$\begin{bmatrix} a_{oi} \\ a_{1i} \\ a_{2i} \\ a_{3i} \end{bmatrix}$$

are not estimated by the usual regression techniques. Instead, if smoothing is not attempted for the first-two and last-two $Y(t_i)$, the specification of 5-point smoothing suggests the formulation

$$
\begin{bmatrix}
1 \; t_1 \; t_1^2 \; t_1^3 & & & \\
1 \; t_2 \; t_2^2 \; t_2^3 & & 0 & \\
\vdots & & & \\
1 \; t_5 \; t_5^2 \; t_5^3 & & & \\
& 1 \; t_2 \; t_2^2 \; t_2^3 & & \\
& \vdots & & \\
& 1 \; t_6 \; t_6^2 \; t_6^3 & & \\
& & \ddots & \\
& & & 1 \; t_{n-4} \; t_{n-4}^2 \; t_{n-4}^3 \\
& 0 & & \vdots \\
& & & 1 \; t_n \; t_n^2 \; t_n^3
\end{bmatrix}
\begin{bmatrix}
a_{03} \\ a_{13} \\ a_{23} \\ a_{33} \\ a_{04} \\ a_{14} \\ a_{24} \\ a_{34} \\ \vdots \\ a_{0,n-2} \\ a_{1,n-2} \\ a_{2,n-2} \\ a_{3,n-2}
\end{bmatrix}
\approx
\begin{bmatrix}
y(t_1) \\ y(t_2) \\ \\ y(t_5) \\ y(t_2) \\ y(t_3) \\ \\ y(t_6) \\ \vdots \\ y(t_{n-4}) \\ y(t_{n-3}) \\ y(t_{n-2}) \\ y(t_{n-1}) \\ y(t_n)
\end{bmatrix}
$$

$[5(n-4) \times (4n-4)]$ $\qquad$ $[5(n-4) \times 1]$

of the form $\underline{\Phi}\,\underline{\alpha} = \underline{Y}$, where the symbol $\approx$ indicates approximation in the sense of the minimum Sobolev norm-type expression. It will be necessary in Section 4.2 to compute the inverse of the matrix $\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}'$, where $\underline{\Phi}'$ denotes $\frac{d\underline{\Phi}'}{dt}$; the task involves the inversion of n-4 (4x4) matrices. Since we assume $t_i \neq t_j$ for $i \neq j$, these inverses exist.

3.3 A Function of Segmented Cubic Polynomials

The case involving initial 5-point moving-arc smoothing and derivative estimation (Section (3.2.1)) involves representation of the point functions $f(t_i)$ and $g(t_i)$, of equation (3.1) by linear combinations of the observations, where the coefficients are uniquely determined by the time-pattern of the observations. It will be shown in Section 4.1 that, in the case where the $f(t_i)$ and $g(t_i)$ are derived from segmented cubic polynomials, the minimization of equation (3.1) involves the simultaneous estimation of the κ and the parameters which characterize the cubic segments. In the following Section 3.3.1, we define spline functions and discuss several criteria for spline functions from the literature as motivation of our own development of the segmented cubic polynomials as a special case. In Section 3.3.2, a complete description is provided for the construction of the cubic segments. The computing procedure to simultaneously estimate the κ and the parameters which characterize the cubic segments will be developed in Chapter 4.

3.3.1 Least-Squares Spline Functions

The following definition of spline function follows that of Greville (1969). Let $t_1, t_2, \ldots, t_n$ be a strictly increasing sequence of real numbers. A spline function of degree m with knots $t_1, t_2, \ldots, t_n$ is a function $S(t)$ satisfying the following conditions:

(1) In each interval (t_i, t_{i+1}), $i = 0, 1, \ldots, n$, where $t_0 = -\infty$, and $t_{n+1} = \infty$, $S(t)$ is given by some polynomial of degree m or less.[1]

(2) $S(t)$ and its derivatives of order $1, 2, \ldots, m-1$ are

---

[1]The spline functions we construct will be of interest only for $t_1 \leq t \leq t_n$.

continuous on $(-\infty, \infty)$, i.e., $S(t)$ is a function of the class C^{m-1}.

Given the n points $(t_i, Y(t_i))$, $i = 1, 2, \ldots, n$, with each $t_i \in (a,b)$, Greville (1968) shows that for arbitrary $k < n$, the smoothest function $f(t)$ of the class C^k which "fits" the n points (i.e., $f(t_i) = Y(t_i)$) is a spline function of degree $2k-1$ having abscissa values $t_1, t_2, \ldots, t_n$ as knots, where the "smoothest" function is defined as that function which minimizes the integral

$$\int_a^b [f^{(k)}(t)]^2 \, dt \, .$$

It is noted that, in general, each of the adjoining polynomial arcs must have equal ordinates and derivatives of orders $1, 2, \ldots, m-1$ at each knot.

Reinsch's (1967) scheme for smoothing by spline functions relaxes the fitting requirement $f(t_i) = Y(t_i)$ and seeks a function $f(t)$ which minimizes the integral

$$\int_{t_1}^{t_n} [f^{(2)}(t)]^2 \, dt$$

among all functions $f(t) \in C^2(t_1, t_n)$ such that

$$\sum_{i=1}^{n} \left[\frac{f(t_i) - Y(t_i)}{\alpha_i} \right]^2 \leq S$$

where S is a given number and the α_i are weights. He shows that $f(t)$ is a cubic spline with knots $t_1, \ldots, t_n$ and, hence, that the polynomial (cubic) segments join at their endpoints and that f, f', and f'' are continuous. Therefore, the adjoining segments have equal ordinates and first derivatives at each knot.

DeBoor and Rice (1968a, 1968b) consider the case where, for the given data pairs (t_i, Y_i), a polynomial spline function of degree m is "least-square-fitted" on each of the k intervals (ξ_i, ξ_{i+1}), where the ξ_i partition the interval $[a,b]$ but do not necessarily coincide with any of the observed t-values. The authors consider both the case where the number and placement of the knots ξ_i are specified and the case where the number of knots is fixed but their placement is determined by attempts to seek a minimum (integral) least-square

$$\int_{t_1}^{t_n} [f(t) - S(t)]^2 \, dt$$

which, due to the discrete nature of the data, is evaluated by a numerical procedure, _e.g._, trapezoidal sums.

An approximation function which is closely related to the least-square splines described by deBoor and Rice (1968a, 1968b) is the "smooth" continuous function composed of segments of cubic polynomials which will be characterized by the following points:[1]

(1) The segments must join at the common knots.

(2) The first derivatives (with respect to time) of any pair of adjoining segments must be equal at their common knot.

[1]This particular characterization is derived from a more general formulation developed by D. C. Martin.

(3) The whole function of such cubic segments can be fitted to the points $(t_i, Y(t_i))$ using ordinary least-squares techniques.

(4) The choice of the domain of each of the k segments is left to the discretion of the user to encourage adaptation to the vicissitudes of the experiments originating the data.

It is natural, in the use of this function to approximate data, that the first derivative of the function be used to approximate the derivative $\frac{dy}{dt}$ at the points of observation, t_i .

It will be necessary to note, for the development of Chapter 4, that the construction method described in the next section provides that the parameters uniquely representing a segmented cubic polynomial, given a set of points and the sequence of knots, are, in fact, the values of the function and the values of the first derivative of the function at the specified knots. These parameters will be denoted by the letter B_j .

3.3.2 Construction of the Segmented Cubic Polynomials

The function $f(t)$ is composed of k cubic segments, $f_i(t)$, $i = 1, 2, \ldots, k$, and is constructed so that a pair of adjoining segments have identical ordinate values and first derivative values at the common knot or "join-point". That is, f could be defined

$$f(t) = \begin{cases} f_i(t_o^*) & \text{if} \quad t < t_o^* \\ f_i(t) & \text{if} \quad t \in [t_{i-1}^*, t_i^*] \\ f_k(t_k^*) & \text{if} \quad t > t_k^* \end{cases}$$

where $\{t_i^*\ ;\ i=0, 1, \ldots, k,\ t_o^* \le t_1,\ t_k^* \ge t_m\}$ are the user-specified knots which partition the interval $[t_1, t_m]$ for the case where the m observations

$$\{(t_i,\ Y(t_i);\ i=1, 2, \ldots, m,\ t_i \in [t_1,\ t_m],\ t_{j-1} < t_j\}$$

are given.[1] To guarantee uniqueness, we specify that $2k + 2 \le m$.

Consider, for the present development, the i^{th} segment, the cubic

$$f_i(t) = a_{io} + a_{i1}t + a_{i2}t^2 + a_{i3}t^3 \ .$$

Obviously, $\dot{f}_i = a_{i1} + 2a_{i2}t + 3a_{i3}t^2$. By the transformation

$$\tau = \frac{t - t_{i-1}^*}{t_i^* - t_{i-1}^*}$$

the subset $[t_{i-1}^*, t_i^*]$ of the domain of f is transformed onto $[0,1]$. Therefore,

$$f_i^*(\tau) = b_{io} + b_{i1}\ \tau + b_{i2}\ \tau^2 + b_{i3}\ \tau^3,\ \tau \in [0,1]$$

and

$$\frac{df_i}{dt} = \frac{df_i^*}{d\tau} \Big/ \frac{dt}{d\tau} = (b_{i1} + 2b_{i2}\ \tau + 3b_{i3}\ \tau^2)/(t_i^* - t_{i-1}^*) \ .$$

Letting $s_i = t_i^* - t_{i-1}^*$, we define B_j, $j = 2i\text{-}1, 2i, 2i\text{+}1, 2i\text{+}2$, as follows:

$$B_{2i-1} = f_i(t_{i-1}^*) = f_i^*(\tau)\big|_{\tau=0} = b_{io}$$

---

[1]For purposes of this discussion f could be left undefined for $t < t_o^*$ and for $t > t_k^*$.

$$B_{2i} = \frac{df_i}{dt}\bigg|_{t=t^*_{i-1}} = \frac{df^*_i}{d\tau}\bigg/\frac{dt}{d\tau}\bigg|_{\tau=0} = b_{i1}/s_i$$

$$B_{2i+1} = f_i(t^*_i) = f^*_i(\tau)\Big|_{\tau=1} = b_{i0} + b_{i1} + b_{i2} + b_{i3}$$

$$B_{2i+2} = \frac{df_i}{dt}\bigg|_{t=t^*_i} = \frac{df^*_i}{d\tau}\bigg/\frac{dt}{d\tau}\bigg|_{\tau=1} = (b_{i1} + 2b_{i2} + 3b_{i3})/s_i$$

In the more convenient matrix notation

$$\begin{bmatrix} B_{2i-1} \\ B_{2i} \\ B_{2i+1} \\ B_{2i+2} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/s_i & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1/s_i & 2/s_i & 3/s_i \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix}$$

so

$$\begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & s_i & 0 & 0 \\ -3 & -2s_i & 3 & -s_i \\ 2 & s_i & -2 & s_i \end{bmatrix} \begin{bmatrix} B_{2i-1} \\ B_{2i} \\ B_{2i+1} \\ B_{2i+2} \end{bmatrix}$$

Therefore, for the m_i observations in the i^{th} interval $[t^*_{i-1}, t^*_i]$,

$$\begin{bmatrix} 1 & \tau_1 & \tau_1^2 & \tau_1^3 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 1 & \tau_{m_i} & \tau_{m_i}^2 & \tau_{m_i}^3 \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} \approx \begin{bmatrix} Y_{m_{i-1}+1} \\ \cdot \\ \cdot \\ \cdot \\ Y_{m_i} \end{bmatrix}$$

so

$$\begin{bmatrix} 1 & \tau_1 & \tau_1^2 & \tau_1^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & \tau_{m_i} & \tau_{m_i}^2 & \tau_{m_i}^3 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & s_i & 0 & 0 \\ -3 & -2s_i & 3 & -s_i \\ 2 & s_i & -2 & s_i \end{bmatrix} \begin{bmatrix} B_{2i-1} \\ B_{2i} \\ B_{2i+1} \\ B_{2i+2} \end{bmatrix} \approx \begin{bmatrix} Y_{m_{i-1}+1} \\ \vdots \\ Y_{m_i} \end{bmatrix}$$

or

$$\begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ 1-3\tau_j^2+2\tau_j^3 & (\tau_j-2\tau_j^2+\tau_j^3)s_i & 3\tau_j^2-2\tau_j^3 & (-\tau_j^2+\tau_j^3)s_i \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} B_{2i-1} \\ B_{2i} \\ B_{2i+1} \\ B_{2i+2} \end{bmatrix} \approx \begin{bmatrix} Y_{m_{i-1}+1} \\ \vdots \\ Y_{m_i} \end{bmatrix}$$

which is of the form $\underline{\Phi}_i \underline{B}_i \approx \underline{Y}_{m_i}$ with the obvious notation, i.e.,

$$\underline{B}_i = \begin{bmatrix} B_{2i-1} \\ B_{2i} \\ B_{2i+1} \\ B_{2i+2} \end{bmatrix}, \qquad \underline{Y}_{m_i} = \begin{bmatrix} Y_{m_{i-1}+1} \\ \vdots \\ Y_{m_i} \end{bmatrix},$$

and $\underline{\Phi}_i$ denotes the $(m_i \times 4)$ matrix.

Now, the matrix $\underline{\Phi}$ is constructed so that to approximate the vector $\underline{Y}$

$$\underline{\Phi}\ \underline{B} = \underbrace{\begin{bmatrix} \underset{(m_1 \times 4)}{\underline{\Phi}_1} & & & 0 \\ & \underset{(m_2 \times 4)}{\underline{\Phi}_2} & & \\ & & \ddots & \\ 0 & & & \underset{(m_k \times 4)}{\underline{\Phi}_k} \end{bmatrix}}_{[m \times (2k+2)]} \begin{bmatrix} B_1 \\ B_2 \\ B_3 \\ \vdots \\ B_{2k+2} \end{bmatrix} \approx \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ \vdots \\ Y_m \end{bmatrix}$$

Remark 1. If the above equation is treated as a least-squares fit, then solving the above linear system of equations for the B_i would yield

$$\hat{\underline{B}} = (\underline{\Phi}^T\ \underline{\Phi})^{-1}\ \underline{\Phi}^T\ \underline{Y}\ .$$

Of course, the predicted values of the observations are $\hat{\underline{Y}} = \underline{\Phi}\ \hat{\underline{B}}$.

The matrix $\underline{\Phi}'$ is, then,

$$\underline{\Phi}' = \left(\frac{d\Phi}{d\tau} \Big/ \frac{dt}{d\tau}\right)_{ij}$$

i.e., the j^{th} row of the submatrix $\underline{\Phi}_i'$ is

$$\left[(-6\tau_j-6\tau_j^2)/s_i \quad 1-4\tau_j+3\tau_j^2 \quad (6\tau_j-6\tau_j^2)/s_i \quad -2\tau_j+3\tau_j^2\right]$$

Remark 2. The original coefficients a_{ij} of the i^{th} cubic

$$f_i(t) = a_{io} + a_{i1}t + a_{i2}t^2 + a_{i3}t^3$$

are easily computed. Since the relationship between the cubic coefficients before scaling, a_{ij} , and the coefficients after scaling, b_{ij} , can be expressed by

$$\begin{bmatrix} a_{i0} \\ a_{i1} \\ a_{i2} \\ a_{i3} \end{bmatrix} = \begin{bmatrix} 1 & -t^*_{i-1}/s_i & t^{*\,2}_{i-1}/s_i^{2} & -t^{*\,3}_{i-1}/s_i^{3} \\ 0 & 1/s_i & -2t^*_{i-1}/s_i^{2} & 3t^{*2}_{i-1}/s_i^{3} \\ 0 & 0 & 1/s_i^{2} & 3t^*_{i-1}/s_i^{3} \\ 0 & 0 & 0 & 1/s_i^{3} \end{bmatrix} \begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} = \underline{C}_i \underline{b}_i$$

where $s_i = t^*_i - t^*_{i-1}$, and since

$$\begin{bmatrix} b_{i0} \\ b_{i1} \\ b_{i2} \\ b_{i3} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & s_i & 0 & 0 \\ -3 & -2s_i & 3 & -s_i \\ 2 & s_i & -2 & s_i \end{bmatrix} \begin{bmatrix} B_{2i-1} \\ B_{2i} \\ B_{2i+1} \\ B_{2i+2} \end{bmatrix} = \underline{D}_i \ \underline{B}_i$$

then

$$\begin{bmatrix} a_{10} \\ a_{11} \\ \vdots \\ a_{k3} \end{bmatrix} = \underbrace{\begin{bmatrix} \underset{(4\times4)}{C_1} & & & 0 \\ & \underset{(4\times4)}{C_2} & & \\ & & \ddots & \\ 0 & & & \underset{(4\times4)}{C_k} \end{bmatrix}}_{(4k\times4k)} \underbrace{\begin{bmatrix} \underset{(4\times4)}{D_1} & & & & 0 \\ & \underset{(4\times4)}{D_2} & & & \\ & & \underset{(4\times4)}{D_3} & & \\ & & & \ddots & \\ 0 & & & & \underset{(4\times4)}{D_k} \end{bmatrix}}_{[4k\times(2k+2)]} \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_{2k+2} \end{bmatrix}$$

3.4 Comparison of Estimators

Classically, a useful comparative measure of the suitability of an estimator has been the mean-squared error, which for estimator $\hat{\kappa}$ of κ is defined to be $\mathcal{E}[(\hat{\kappa}-\kappa)^2]$ and which has the property

$$\mathcal{E}[(\hat{\kappa}-\kappa)^2] = \text{var}(\hat{\kappa}) + [\kappa-\mathcal{E}(\hat{\kappa})]^2$$

in which the quantity $\kappa-\mathcal{E}(\hat{\kappa})$ is defined to be the bias. Of several estimators $\hat{\kappa}_{(i)}$ of κ, the one which has the smallest mean-squared error if often said to be the "best".

Obviously, by definition, the variance and bias of the random variable $\hat{\kappa}$ are non-random parameters associated with the distribution of the given estimator $\hat{\kappa}$ where the random variables of which $\hat{\kappa}$ is a function have a particular, though, perhaps, unknown distribution. Frequently, the estimator $\hat{\kappa}$ cannot be expressed in a form readily amenable to application of analytic methods of determining $\mathcal{E}(\hat{\kappa})$ and $\text{var}(\hat{\kappa})$. In such cases, it is often possible to devise an approximation to $\hat{\kappa}$ from which approximations to $\mathcal{E}(\hat{\kappa})$ and $\text{var}(\hat{\kappa})$ can be analytically determined. The appropriateness of such approximations is usually verified by simulation studies for values of $\hat{\kappa}$ within the region of interest. In the event approximations to $\hat{\kappa}$ are not made, estimates of $\mathcal{E}(\hat{\kappa})$ and $\text{var}(\hat{\kappa})$ are often found by simulation studies for specific values of κ. In this context, the simulation studies frequently involve the generation of artificial observations with additive error from a known distribution. For example, the generated observations could be of the form

$$Y(t_i) = \gamma(t_i) + \varepsilon(t_i)$$

where $\varepsilon(t_i) \sim n[0,v]$ and $\mathrm{cov}[\varepsilon(t_i), \varepsilon(t_j)] = 0$ with either $v =$ constant or

$$v = \left[\frac{p}{100}\, y(t_i)\right]^2 .$$

The mean, $\bar{\hat{\kappa}}$, of the $\hat{\kappa}$ from m sets of artificial observations is used as an estimate of $\mathcal{E}(\hat{\kappa})$ for a particular v. The value of $\hat{\kappa}$ for error-free values, i.e., $y(t_i) = Y(t_i)$, will be denoted by either $\hat{\kappa}_{v=0}$ or $\hat{\kappa}_{exact}$. The total bias (bias_{total}) is

$$\kappa - \mathcal{E}(\hat{\kappa}) = (\kappa - \hat{\kappa}_{v=0}) + (\hat{\kappa}_{v=0} - \mathcal{E}(\hat{\kappa}))$$

and, from a simulation study, can be approximated by

$$\kappa - \bar{\hat{\kappa}} = (\kappa - \hat{\kappa}_{v=0}) + (\hat{\kappa}_{v=0} - \bar{\hat{\kappa}}) .$$

Hereafter, $\kappa - \hat{\kappa}_{v=0}$, which refers to the bias due-to-estimation procedure, will be denoted by bias_{method}, and $\hat{\kappa}_{v=0} - \bar{\hat{\kappa}}$, which refers to the bias due to error in the observations as bias_{error}.

While the mean-squared error may represent a useful measure for comparing several estimators, it fails to provide information concerning the relative magnitudes of the bias and variance. Since interest is often centered on the bias of an estimator, it may be worthwhile to consider the MSE and a quantity which relates the MSE and the bias. Either of the ratios

$$\xi = \frac{\text{bias}_{total}}{\sqrt{\mathrm{var}(\hat{\kappa})}}$$

or

$$\xi^* = \frac{\text{bias}_{total}^2}{\text{MSE}}$$

are such quantities. Trivially,

$$\xi^2 = \frac{MSE}{var(\hat{\kappa})} - 1$$

$$\xi^* = \frac{\xi^2}{1 + \xi^2}$$

It is obvious that ξ is a measure of $bias_{total}$ in units of standard deviation, while ξ^* is the fraction of the value of MSE which is attributable to $bias_{total}$.

It is entirely possible that, of two estimators, the one with the smaller mean-squared error might actually have the greater bias, and, hence smaller variance, than the other estimator. In fact, assuming the $\hat{\kappa}$ values are distributed normally, it is possible that the bias of the former estimator might be so large that the interval $\mathcal{E}(\hat{\kappa}) \pm 2\sqrt{var(\hat{\kappa})}$ would exclude the true κ, or, equivalently, that the interval bounded by

$$\frac{\mathcal{E}(\hat{\kappa}) \pm 2\sqrt{var(\hat{\kappa})} - \kappa}{\sqrt{var(\hat{\kappa})}}$$

would exclude zero. However,

$$\frac{\mathcal{E}(\hat{\kappa}) - \kappa \pm 2\sqrt{var(\hat{\kappa})}}{\sqrt{var(\hat{\kappa})}} = \frac{-bias_{total}}{\sqrt{var(\hat{\kappa})}} \pm 2$$

$$= -\xi \pm 2$$

indicating that the value of ξ provides information concerning the effect of the relative magnitudes of the bias_{total} and $\text{var}(\hat{k})$ in the construction of confidence intervals.

Obviously, there is interest in attempts to decrease the MSE and ξ. The value of MSE can be decreased by decreasing one of either $(\text{bias}_{total})^2$ or $\text{var}(\hat{k})$ by an amount greater than the other is increased, e.g., by accepting an increase in $(\text{bias}_{total})^2$ as the penalty for a greater decrease in the variance.

Since the bias_{total} $(\hat{k})$ is defined to be

$$\text{bias}_{total}\ (\hat{k}) = \text{bias}_{method}\ (\hat{k}) + \text{bias}_{error}\ (\hat{k})\ ,$$

it would appear naively that decreasing either $|\text{bias}_{method}\ (\hat{k})|$ or $|\text{bias}_{error}\ (\hat{k})|$ would result in decreasing $|\text{bias}_{total}\ (\hat{k})|$. However, it will be shown in Chapter 4 that, in some cases where the bias_{total} and bias_{error} are of unlike signs, minimum bias_{total} can be achieved when either $|\text{bias}_{method}|$ or $|\text{bias}_{error}|$ are judiciously increased. On the other hand, in the estimators to be discussed, an increase in $|\text{bias}_{error}|$ is generally accompanied by an increase in variance $(\hat{k})$. Individual cases will be discussed in later chapters.

4. ESTIMATING κ IN $\frac{dy}{dt} = \kappa y$

Consider the simple differential equation

$$\frac{dy}{dt} = \kappa y(t),\ y(0) = y_0, \tag{4.1}$$

in which the value of κ is unknown and is to be estimated from n observations, $\{Y(t_1), Y(t_2), \ldots, Y(t_n)\}$, where $Y(t_i) = y(t_i) + \varepsilon(t_i)$ and $\varepsilon(t_i)$ is a random error. The objective will be to minimize the Sobolev norm-type expression

$$\begin{aligned} \nu^2 &= \sum_{i=1}^{n} \{[Y(t_i)-f(t_i)]^2 + w[\dot{Y}(t_i)-g(t_i)]^2\} \\ &= \sum_{i=1}^{n} \{[Y(t_i)-f(t_i)]^2 + w[\kappa Y(t_i)-g(t_i)]^2\} \end{aligned} \tag{4.2}$$

with respect to κ, where f and g are as defined in Chapter 3. Usually f and g contain parameters which must be estimated so that the theory associated with the estimation of κ depends on the structure of the functions f and g.

In Section 4.1, the estimation of κ by minimizing (4.2) is implemented after the parameters of f and g are chosen by classical least-squares methods so that f and g are completely specified for given set of $Y(t_i)$ by knowledge of the time sequence. Of special interest is the case where the (nX1) vectors $\underline{f}$ and $\underline{g}$ are $\underline{f} = \underline{S}\,\underline{Y}$ and $\underline{g} = \underline{D}\,\underline{Y}$ with $\underline{S}$ and $\underline{D}$ defined as moving-arc smoothing and derivative estimation matrices as in Section 3.2.1. It will be shown that this procedure yields an estimator, $\hat{\kappa}$, of κ which is subject to non-negligible bias. For several true values of κ, properties of $\hat{\kappa}$ are investigated where the errors $\varepsilon(t_i)$ are distributed normally with proportional standard deviation and with constant standard deviation.

In Section 4.2, the estimation of κ is implemented simultaneously with the choice of the parameters of f and g by means of the minimization of (4.2). Two particular forms of f and g are investigated:

(i) f is constructed to be the point function resulting from 5-point moving-arc cubic smoothing and $f' = g$.

(ii) f is a continuous and differentiable function composed of cubic segments and $f' = g$.

Wherever Monte Carlo simulation is conducted, the necessary normal deviates are constructed by the built-in uniform random number generator function RAND of the PL/C computer program compiler (Blankinship, 1971). Twelve such pseudorandom numbers were summed and their theoretical mean was subtracted from the sum to yield the pseudorandom normally distributed deviates. All computer programs were compiled and executed on either the IBM 360/75 or its successor, the IBM 370/165, digital computers operated by the Triangle Universities Computation Center, Research Triangle Park, North Carolina.

4.1 Estimation of κ and Smoothing Operations Conducted Separately

4.1.1 General Development

If w is arbitrary in units of time-squared and if f and g are the point functions derived from 5-point cubic or 5-point linear hyperbolic smoothing and derivative estimation, then, as described in Section 3.2, the value of κ which minimizes ν^2 in equation (4.2) is that value of κ which minimizes the last term of equation (4.2),

$$w[\kappa \underline{Y} - \underline{D}\,\underline{Y}]^T [\kappa\, Y - \underline{D}\,\underline{Y}] . \qquad (4.3)$$

For the sake of generality, this discussion will consider the case where

$$[\kappa\, \underline{S}\,\underline{Y} - \underline{D}\,\underline{Y}]^T [\kappa\, \underline{S}\,\underline{Y} - \underline{D}\,\underline{Y}] = \underline{\delta}^T\underline{\delta} \qquad (4.4)$$

rather than expression (4.3) is to be minimized. Taking

$$\frac{\partial(\underline{\delta}^T\underline{\delta})}{\partial\kappa} = 2\hat{\kappa}\, \underline{Y}^T\underline{S}^T\underline{SY} - 2\underline{Y}^T\underline{S}^T\underline{DY} ,$$

setting

$$\frac{\partial(\underline{\delta}^T\underline{\delta})}{\partial\kappa} = 0$$

and solving for κ , we have

$$\hat{\kappa} = (\underline{Y}^T\underline{S}^T\underline{SY})^{-1}\, \underline{Y}^T\underline{S}^T\underline{DY} . \qquad (4.5)$$

Alternatively, if $\underline{Dy} = \kappa\, \underline{Sy}$ is an approximation to the differential equation $\frac{dy}{dt} = \kappa y$ where $\underline{y}^T = \{y(t_1) \ldots y(t_n)\}$ and if the observed values are $Y(t_i) = y(t_i) + \varepsilon(t_i)$, then

$$\underline{D}(\underline{Y}-\underline{\varepsilon}) = \kappa\, \underline{S}(\underline{Y}-\underline{\varepsilon})$$

or

$$\underline{DY} = \kappa\, \underline{SY} + \underline{D\varepsilon} - \kappa\underline{S\varepsilon}$$

$$= \kappa\, \underline{SY} + \underline{\delta}$$

where $\underline{\delta} = (\underline{D}-\kappa\underline{S})\underline{\varepsilon}$. Considered as a regression model with errors $\underline{\delta}$, the above approximation admits a least-squares estimate of κ , namely that κ which minimizes

$$\underline{\delta}^T\underline{\delta} = (\underline{DY}-\kappa\underline{SY})^T\ (\underline{DY}-\kappa\underline{SY})\ ,$$

the same quantity as in (4.4).

One has that, upon substitution into equation (4.5),

$$\hat{\kappa} = [(\underline{y}+\underline{\varepsilon})^T\underline{S}^T\underline{D}(\underline{y}+\underline{\varepsilon})]\ [(\underline{y}+\underline{\varepsilon})^T\underline{S}^T\underline{S}(\underline{y}+\underline{\varepsilon})]^{-1}$$

$$= \frac{\underline{y}^T\underline{S}^T\underline{Dy} + \underline{y}^T\underline{S}^T\underline{D\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Dy} + \underline{\varepsilon}^T\underline{S}^T\underline{D\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy} + \underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}} \qquad (4.6)$$

$$= \frac{\underline{y}^T\underline{S}^T\underline{Dy}}{\underline{y}^T\underline{S}^T\underline{Sy}} + \frac{\underline{y}^T\underline{S}^T\underline{Dy}}{\underline{y}^T\underline{S}^T\underline{Sy}}\left[-\frac{\underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy} + \underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}\right]$$

$$+ \frac{\underline{y}^T\underline{S}^T\underline{D\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Dy} + \underline{\varepsilon}^T\underline{S}^T\underline{D\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy} + \underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}\ .$$

Since

$$\frac{\underline{y}^T\underline{S}^T\underline{Dy}}{\underline{y}^T\underline{S}^T\underline{Sy}} = \hat{\kappa}_{exact}$$

is an estimate of κ under error-free observations, i.e., $\underline{Y} = \underline{y}$, then

$$\hat{\kappa} = \hat{\kappa}_{exact} + \hat{\kappa}_{exact}\left[-\frac{\underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy} + \underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}\right]$$

$$+ \frac{\underline{y}^T\underline{S}^T\underline{D\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Dy} + \underline{\varepsilon}^T\underline{S}^T\underline{D\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy} + \underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Sy} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}\ . \qquad (4.7)$$

For the case in which $\underline{Y}$ contains random error, i.e., $\underline{\varepsilon} \neq \underline{0}$, (-1) bias_{error} is given by the expected value of the second and third terms on the right-hand side of equation (4.7), namely

$$\hat{\kappa}_{exact}\left[-\frac{2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}+2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}\right]+\frac{\underline{y}^T\underline{S}^T\underline{D\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{Dy}+\underline{\varepsilon}^T\underline{S}^T\underline{D\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}+2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}} \tag{4.8}$$

since $\underline{y}^T\underline{S}^T\underline{S\varepsilon} = \underline{\varepsilon}^T\underline{S}^T\underline{Sy}$.

To derive an analytic expression for $\mathcal{E}(\hat{\kappa})$, it would be necessary to compute the expected value of expression (4.8). This approach is complicated by the appearance of the $\varepsilon(t_i)$ in first- and second-powers and in cross-products in the numerators and denominators of the fractions involved. However, the expression

$$\frac{1}{\underline{y}^T\underline{S}^T\underline{Sy}+2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}=\frac{1}{\underline{y}^T\underline{S}^T\underline{Sy}\left[1+\dfrac{2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}}\right]}$$

$$=\frac{1}{\underline{y}^T\underline{S}^T\underline{Sy}}\left[1-\frac{2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}}+\left(\frac{2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}}\right)^2-+\cdots\right] \tag{4.9}$$

permits an approximation to (-1) bias_{error} to be formed in which the terms involving $\varepsilon(t_i)$ appear only in the numerator. The approximation to (4.8) is of the form

$$\frac{1}{\underline{y}^T\underline{S}^T\underline{Sy}}\cdot(x)\cdot\left[-\hat{\kappa}_{exact}\,(2\underline{y}^T\underline{S}^T\underline{S\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon})+(\underline{y}^T\underline{S}^T\underline{D\varepsilon}+\underline{\varepsilon}^T\underline{S}^T\underline{Dy}+\underline{\varepsilon}^T\underline{S}^T\underline{D\varepsilon})\right]$$

where, for practical purposes, x could be either

(i) $\dfrac{1}{\underline{y}^T\underline{S}^T\underline{S}\underline{y}}$ or

(ii) $\dfrac{1}{\underline{y}^T\underline{S}^T\underline{S}\underline{y}}\left[1 - \dfrac{2\underline{y}^T\underline{S}^T\underline{S}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{S}\underline{\varepsilon}}{\underline{y}^T\underline{S}^T\underline{S}\underline{y}}\right]$ or

(iii) $\dfrac{1}{\underline{y}^T\underline{S}^T\underline{S}\underline{y}}\left[1 - \dfrac{2\underline{y}^T\underline{S}^T\underline{S}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{S}\underline{\varepsilon}}{\underline{y}^T\underline{S}^T\underline{S}\underline{y}} + \left(\dfrac{2\underline{y}^T\underline{S}^T\underline{S}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{S}\underline{\varepsilon}}{\underline{y}^T\underline{S}^T\underline{S}\underline{y}}\right)^2\right]$.

Since the expression being approximated is a factor in the complete expression (4.8), the choice of either (i), (ii) or (iii) must be investigated from two aspects:

(a) The suitability of the approximation to the factor (4.9).

(b) The suitability of the resulting approximation of $\hat{\kappa}$, as in equation (4.7).

This investigation involves computing the value of (i) and the means and standard errors of the sample means of the quantities (4.9) and (ii) and (iii) as well as those of the resulting estimates of $\hat{\kappa}$ which were made using (4.9), (i), (ii) and (iii), where $\underline{S} = \underline{I}$ and $\underline{D}$ is the matrix defined in Section 3.2.1 for initial derivative estimation by 5-point moving-arc cubics. For each of κ = -0.1, -0.2, -0.3, -0.4, -0.5, 200 sets of hypothetical observations $\{Y(t_i);\ t_i = 1, 2, \ldots, 15\}$ were constructed, where $y(t_i) = 1000^{-\kappa t_i}$, and $\varepsilon(t_i) \sim n\,[0, (\frac{10}{100}\,y(t_i))^2]$. The results shown in Table 4.1 indicate that approximation (ii) is satisfactory to approximate the fraction (4.9) and that approximation (i) is not unreasonable.

Table 4.1 Simulation results for approximations to $\frac{1}{Y^T S^T S Y}$. Entries in subcolumns 2 through 7 are means. $\overline{\mathrm{Var}(\hat{\kappa})}$, where computed, are in parentheses.

	$\frac{1}{\underline{Y}^T\underline{S}^T\underline{SY}}$		approximation (i)		approximation (ii)		approximation (iii)	
	fraction 10^{-7}	$\hat{\kappa}$	approximation 10^{-7}	$\hat{\kappa}$	approximation 10^{-7}	$\hat{\kappa}$	approximation 10^{-7}	$\hat{\kappa}$
-0.1	2.3145	-0.1020 (0.0029)	2.3300	-0.1025	2.3049	-0.1019 (0.0029)	2.3147	-0.1020 (0.0029)
-0.2	4.8707	-0.2076 (0.0048)	4.9305	-0.2092	4.8362	-0.2075 (0.0048)	4.8727	-0.2076 (0.0048)
-0.3	8.2205	-0.2994 (0.0056)	8.2222	-0.3033	8.1272	-0.2993 (0.0055)	8.2212	-0.2996 (0.0056)
-0.4	12.197	-0.4016 (0.0064)	12.255	-0.4070	11.994	-0.4010 (0.0062)	12.207	-0.4019 (0.0064)
-0.5	17.648	-0.4851 (0.0071)	17.183	-0.4914	17.343	-0.4848 (0.0069)	17.650	-0.4855 (0.0071)

Using the approximation (ii),

$$\frac{1}{\underline{y}^T\underline{C}\underline{y}}\left[1-\frac{2\underline{y}^T\underline{C}\underline{\varepsilon}+\underline{\varepsilon}^T\underline{C}\underline{\varepsilon}}{\underline{y}^T\underline{C}\underline{y}}\right],$$

of

$$\frac{1}{\underline{y}^T\underline{C}\underline{y}+2\underline{y}^T\underline{C}\underline{\varepsilon}+\underline{\varepsilon}^T\underline{C}\underline{\varepsilon}},$$

the estimator for κ (equation 4.7) can be approximated by

$$\hat{\kappa} \doteq \hat{\kappa}_{exact} + \hat{\kappa}_{exact}\left[-\frac{1}{\underline{y}^T\underline{C}\underline{y}}\right]\left[1-\frac{2\underline{y}^T\underline{C}\underline{\varepsilon}+\underline{\varepsilon}^T\underline{C}\underline{\varepsilon}}{\underline{y}^T\underline{C}\underline{y}}\right](2\underline{y}^T\underline{C}\underline{\varepsilon}+\underline{\varepsilon}^T\underline{C}\underline{\varepsilon})$$

$$+\frac{1}{\underline{y}^T\underline{C}\underline{y}}\left[1-\frac{2\underline{y}^T\underline{C}\underline{\varepsilon}+\underline{\varepsilon}^T\underline{C}\underline{\varepsilon}}{\underline{y}^T\underline{C}\underline{y}}\right](\underline{y}^T\underline{B}\underline{\varepsilon}+\underline{\varepsilon}^T\underline{B}\underline{y}+\underline{\varepsilon}^T\underline{B}\underline{\varepsilon}) \qquad (4.10)$$

where $\underline{C} = \underline{S}^T\underline{S}$ and $\underline{B} = \underline{S}^T\underline{D}$. As in equation (4.7), the second and third terms of the right-hand side of equation (4.10) are an estimate of (-1) bias_{error}. Now the expected value of the approximation (4.10) can be expressed in terms of moments of ε of the first few orders. Although the tedious details of the deviation are reserved for the appendix, the expected value of the approximation to $\hat{\kappa}$ is

$$\hat{\kappa}_{exact} + \frac{1}{\underline{y}^T\underline{C}\underline{y}}\{-\hat{\kappa}_{exact}\sum_{i=1}^{n} C_{ii}\,\text{var}(\varepsilon_i) + \sum_i B_{ii}\,\text{var}(\varepsilon_i)\}$$

$$+\frac{\hat{\kappa}_{exact}}{(\underline{y}^T\underline{C}\underline{y})^2}\{4\sum_i(\sum_j y_j C_{ji})^2\,\text{var}(\varepsilon_i)$$

$$+\sum_i\sum_j(C_{ii}B_{jj}+C_{ij}^2+C_{ji}C_{ij})\,\text{var}(\varepsilon_i)\,\text{var}(\varepsilon_j)\}$$

$$-\left(\frac{1}{\underline{y}^T\underline{Cy}}\right)^2 \{2 \sum_i (\sum_j y_j C_{ji})(\sum_j y_j B_{ji}) \operatorname{var}(\varepsilon_i)$$

$$+ 2 \sum_i (\sum_j y_j C_{ji})(\sum_j y_j B_{ij}) \operatorname{var}(\varepsilon_i) \qquad (4.11)$$

$$+ \sum_i \sum_j (C_{ii}B_{jj} + C_{ij}B_{ij} + C_{ji}B_{ij}) \operatorname{var}(\varepsilon_i) \operatorname{var}(\varepsilon_j)\}$$

An approximation of the variance of $\hat{\kappa}$, where $\hat{\kappa}$ is approximated by equation (4.10), would involve computations of $\mathcal{E}(\underline{\varepsilon}^T\underline{C\varepsilon\varepsilon}^T\underline{C\varepsilon})^2$ which arise from the form of $\mathcal{E}(\hat{\kappa}^2)$. Since the results shown in Table 4.1 indicate that $\frac{1}{\underline{y}^T\underline{S}^T\underline{Sy}}$ is not a poor approximation of

$$\frac{1}{\underline{y}^T\underline{S}^T\underline{Sy} + 2\underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}},$$

the estimator $\hat{\kappa}$ could be approximated by

$$\hat{\kappa} \doteq \hat{\kappa}_{exact} + \hat{\kappa}_{exact}\left[-\frac{2\underline{y}^T\underline{S}^T\underline{S\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{S\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}}\right] + \frac{\underline{y}^T\underline{S}^T\underline{D\varepsilon} + \underline{\varepsilon}^T\underline{S}^T\underline{Dy} + \underline{\varepsilon}^T\underline{S}^T\underline{D\varepsilon}}{\underline{y}^T\underline{S}^T\underline{Sy}}$$

which, when squared, would involve computation of $\mathcal{E}(\underline{\varepsilon}^T\underline{C\varepsilon\varepsilon}^T\underline{C\varepsilon})$, a less complicated procedure. Indeed,

$$\hat{\kappa}^2 \doteq \hat{\kappa}^2_{exact} - \frac{2\hat{\kappa}^2_{exact}}{\underline{y}^T\underline{Cy}}(2\underline{y}^T\underline{C\varepsilon} + \underline{\varepsilon}^T\underline{C\varepsilon}) + \frac{2\hat{\kappa}_{exact}}{\underline{y}^T\underline{Cy}}(\underline{y}^T\underline{B\varepsilon} + \underline{\varepsilon}^T\underline{By} + \underline{\varepsilon}^T\underline{B\varepsilon})$$

$$+ \frac{\hat{\kappa}^2_{exact}}{(\underline{y}^T\underline{Cy})^2}(2\underline{y}^T\underline{C\varepsilon} + \underline{\varepsilon}^T\underline{C\varepsilon})^2 - \frac{2\hat{\kappa}_{exact}}{(\underline{y}^T\underline{Cy})^2}(2\underline{y}^T\underline{C\varepsilon} + \underline{\varepsilon}^T\underline{C\varepsilon})$$

$$(\underline{y}^T\underline{B\varepsilon} + \underline{\varepsilon}^T\underline{By} + \underline{\varepsilon}^T\underline{B\varepsilon}) + \frac{1}{(\underline{y}^T\underline{Cy})^2}(\underline{y}^T\underline{B\varepsilon} + \underline{\varepsilon}^T\underline{By} + \underline{\varepsilon}^T\underline{B\varepsilon})^2$$

where, as before,

$$\underline{C} = \underline{S}^T\underline{S} \quad \text{and} \quad \underline{B} = \underline{S}^T\underline{D} \ .$$

Therefore,

$$\begin{aligned}
\mathcal{E}(\hat{k}^2)_{approx} = {} & \hat{k}^2_{exact} - \frac{2\hat{k}^2_{exact}}{\underline{y}^T\underline{C}\underline{y}} \sum_i C_{ii} \, var(\varepsilon_i) \\
& + \frac{2\hat{k}_{exact}}{\underline{y}^T\underline{C}\underline{y}} \sum_i B_{ii} \, var(\varepsilon_i) + \left[\frac{\hat{k}_{exact}}{\underline{y}^T\underline{C}\underline{y}}\right]^2 [4 \sum_i (\sum_j C_{ij}y_j)^2 \, var(\varepsilon_i)] \\
& + \left[\frac{\hat{k}_{exact}}{\underline{y}^T\underline{C}\underline{y}}\right]^2 [\sum_i \sum_j (C_{ii}C_{jj} + C_{ij}C_{ij} + C_{ji}C_{ij}) \, var(\varepsilon_i) \, var(\varepsilon_j)] \\
& - \frac{2\hat{k}_{exact}}{(\underline{y}^T\underline{C}\underline{y})^2} \Big\{ 2 \sum_i (\sum_j C_{ij}y_j)(\sum_m B_{mi}y_m) \, var(\varepsilon_i) \\
& + [2 \sum_i (\sum_j C_{ij}y_j)(\sum_m B_{im}y_m) \, var(\varepsilon_i)] \Big\} \\
& - \frac{2\hat{k}_{exact}}{(\underline{y}^T\underline{C}\underline{y})^2} \sum_i \sum_j (C_{ii}B_{jj} + C_{ij}B_{ij} + C_{ji}B_{ij}) \, var(\varepsilon_i) \, var(\varepsilon_j) \\
& + \frac{1}{(\underline{y}^T\underline{C}\underline{y})^2} [\sum_i (\sum_j B_{ji}y_j)^2 \, var(\varepsilon_i) + 2 \sum_i (\sum_j B_{ji}y_j)(\sum_m B_{im}y_m) \, var(\varepsilon_i) \\
& + \sum_i (\sum_j B_{ij}y_j)^2 \, var(\varepsilon_i)] \\
& + \frac{1}{(\underline{y}^T\underline{C}\underline{y})^2} \sum_i \sum_j (B_{ii}B_{jj} + B_{ij}B_{ij} + B_{ji}B_{ij}) \, var(\varepsilon_i) \, var(\varepsilon_j) \ .
\end{aligned} \tag{4.12}$$

The approximate value of $\mathcal{E}(\hat{\kappa})$ can be computed using expression (4.11). Hence, the variance of $\hat{\kappa}$ can be approximated by

$$\mathrm{var}(\hat{\kappa}) \doteq \mathcal{E}(\hat{\kappa}^2)_{\mathrm{approx}} - [\mathcal{E}(\hat{\kappa})_{\mathrm{approx}}]^2 \; . \qquad (4.12a)$$

Of course, the presence of non-normal errors would require the re-writing of equation (4.12).

4.1.2 Numerical Example

In this section, some of the properties of the estimator $\hat{\kappa}$ (equation 4.5) are investigated by means of constructing values $\{y(t_1), y(t_2), \ldots, y(t_{15}): \; t_i = i\}$ by the solution $y(t) = 1000\, e^{\kappa t}$ to $\frac{dy}{dt} = \kappa y$, $y(0) = 1000$, and generating sets of artificial data $\{Y(t_i): \; t_i = 1, 2, \ldots, 15\}$ where $Y(t_i) = y(t_i) + \varepsilon(t_i)$. In all cases $\varepsilon(t_i) \sim \mathcal{N}(0,v)$, where v is either

(i) $v = \left[\dfrac{P\, y(t_i)}{100}\right]^2$ or

(ii) $v = \text{constant} \neq 0$ or

(iii) $v = 0$,

and $\underline{S}$ and $\underline{D}$ are (15X15) matrices constructed for 5-point moving-arc cubic smoothing and derivative estimation, or $\underline{S} = \underline{I}$ and $\underline{D}$ is as defined above.

To investigate $\mathrm{bias}_{\mathrm{method}}$, $\hat{\kappa}_{\mathrm{exact}}$ is obtained by equation (4.5), the expected value of the approximate $\mathrm{bias}_{\mathrm{error}}$ is obtained by equation (4.10), and the approximate variance of $\hat{\kappa}$ by equation (4.12a). To investigate the validity of such approximations, a

Monte Carlo simulation study was conducted in which 200 sets of artificial data were generated. For a given value of κ , the same pseudorandom deviates were used to construct the $\varepsilon(t_i)$. The means and variances of the $\hat{\kappa}$ obtained by averaging over the 200 sets are included in the accompanying tables as follows:

	$v = \left[\frac{p\, y(t_i)}{100}\right]^2$, p = 10	v = 900
$\underline{S}$ as 5-point cubic smoother	expected(approx): Table 4.2 simulation: Table 4.3	expected(approx): Table 4.6 simulation: Table 4.7
$\underline{S} = \underline{I}$	expected(approx): Table 4.4 simulation: Table 4.5	expected(approx): Table 4.8 simulation: Table 4.9

These results indicate the following:

(1) There is very little difference between the $\bar{\hat{\kappa}}$ and between the $\text{var}(\hat{\kappa})$ from minimization of $(\kappa\underline{Y} - \underline{DY})^T (\kappa\underline{Y} - \underline{DY})$ and $(\kappa\underline{SY} - \underline{DY})^T (\kappa\underline{SY} - \underline{DY})$,

(2) $\text{Bias}_{\text{total}}$, although obvious, is small compared to $\sqrt{\text{var}(\hat{\kappa})}$.

(3) For each estimator $\hat{\kappa}$ as a function of κ , $|\text{bias}_{\text{method}}|$ tends to increase as $|\kappa|$ increases, indicating increasingly poor approximation of the exponential function by the cubic function.

Table 4.2 Estimates using approximations of $bias_{error}$ and $var(\hat{\kappa})$ following initial 5-point moving-arc cubic smoothing for proportional data error. p = 10. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{exact}$	$\mathcal{E}(\hat{\kappa})$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.1023	0.0000	0.0023	0.0023	0.001738	0.0417	0.0029	0.001743	0.0552	0.0030
-0.2	-0.1998	-0.2027	-0.0002	0.0029	0.0027	0.005221	0.0723	0.0051	0.005228	0.0374	0.0014
-0.3	-0.2990	-0.3012	-0.0010	0.0022	0.0012	0.009943	0.0997	0.0071	0.009944	0.0120	0.0001
-0.4	-0.3966	-0.3975	-0.0034	0.0009	-0.0025	0.016197	0.1273	0.0090	0.016203	-0.0196	0.0004
-0.5	-0.4914	-0.4908	-0.0086	-0.0006	-0.0092	0.025650	0.1602	0.0113	0.025735	-0.0574	0.0033

Table 4.3 Simulation results for the estimation of κ following initial 5-point moving-arc cubic smoothing for proportional data error. p = 10. n = 200.

κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.1030	0.0000	0.0030	0.0030	0.001766	0.0420	0.0030	0.001775	0.0714	0.0051
-0.2	-0.1998	-0.2091	-0.0002	0.0093	0.0091	0.004695	0.0685	0.0048	0.004778	0.1328	0.0173
-0.3	-0.2990	-0.3011	-0.0010	0.0021	0.0011	0.006333	0.0796	0.0056	0.006334	0.0138	0.0002
-0.4	-0.3966	-0.4034	-0.0034	0.0068	0.0034	0.008222	0.0907	0.0064	0.008234	0.0375	0.0014
-0.5	-0.4914	-0.4869	-0.0086	-0.0045	-0.0131	0.010300	0.1015	0.0072	0.010472	-0.1291	0.0164

Table 4.4 Estimates using approximations of $bias_{error}$ and $var(\hat{\kappa})$ following initial derivative estimation by 5-point moving-arc cubics for proportional data error. p = 10. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{exact}$	$\mathcal{E}(\hat{\kappa})$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.1012	0.0000	0.0012	0.0012	0.001794	0.0424	0.0030	0.001795	0.0283	0.0008
-0.2	-0.1998	-0.2012	-0.0002	0.0014	0.0012	0.005237	0.0724	0.0051	0.005238	0.0166	0.0003
-0.3	-0.2990	-0.2995	-0.0010	0.0005	-0.0005	0.009998	0.1000	0.0071	0.009998	-0.0050	0.0000
-0.4	-0.3964	-0.3957	-0.0036	-0.0007	-0.0043	0.016274	0.1276	0.0090	0.016292	-0.0337	0.0011
-0.5	-0.4912	-0.4891	-0.0088	-0.0021	-0.0109	0.025812	0.1607	0.0114	0.025931	-0.0678	0.0046

Table 4.5 Simulation results for the estimation of κ following initial derivative estimation by 5-point moving-arc cubics for proportional data error. p = 10. n = 200.

κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.1020	0.0000	0.0020	0.0020	0.001736	0.0417	0.0029	0.001740	0.0480	0.0023
-0.2	-0.1998	-0.2076	-0.0002	0.0078	0.0076	0.004627	0.0680	0.0048	0.004685	0.1117	0.0123
-0.3	-0.2990	-0.2994	-0.0010	0.0004	-0.0006	0.006268	0.0792	0.0056	0.006268	-0.0076	0.0001
-0.4	-0.3964	-0.4016	-0.0036	0.0052	0.0016	0.008107	0.0900	0.0064	0.008110	0.0178	0.0003
-0.5	-0.4912	-0.4851	-0.0088	-0.0061	-0.0149	0.010097	0.1005	0.0071	0.010319	-0.1483	0.0215

Table 4.6 Estimates using approximations of $bias_{error}$ and $var(\hat{\kappa})$ following initial 5-point moving-arc cubic smoothing and derivative estimation for the case of constant data error variance. $v = 900$. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{exact}$	$\mathcal{E}(\hat{\kappa})$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0998	0.0000	-0.0002	-0.0002	0.000224	0.0150	0.0011	0.000224	-0.0134	0.0002
-0.2	-0.1998	-0.1991	-0.0002	-0.0007	-0.0009	0.000838	0.0289	0.0020	0.000838	-0.0311	0.0010
-0.3	-0.2990	-0.2971	-0.0010	-0.0019	-0.0029	0.002389	0.0489	0.0035	0.002398	-0.0593	0.0035
-0.4	-0.3966	-0.3929	-0.0034	-0.0037	-0.0071	0.005891	0.0768	0.0054	0.005941	-0.0925	0.0085
-0.5	-0.4914	-0.4849	-0.0086	-0.0065	-0.0151	0.013516	0.1163	0.0082	0.013744	-0.1299	0.0166

Table 4.7 Simulation results for estimation of κ following initial 5-point moving-arc cubic smoothing and derivative estimation for the case of constant error variance. $v = 900$. $n = 200$.

κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0999	0.0000	-0.0001	-0.0001	0.000234	0.0153	0.0011	0.000234	-0.0065	0.0000
-0.2	-0.1998	-0.2009	-0.0002	0.0011	0.0009	0.000838	0.0289	0.0020	0.000839	0.0311	0.0010
-0.3	-0.2990	-0.2973	-0.0010	-0.0017	-0.0027	0.001517	0.0389	0.0028	0.001524	-0.0693	0.0048
-0.4	-0.3966	-0.3948	-0.0034	-0.0018	-0.0052	0.002646	0.0514	0.0036	0.002673	-0.1011	0.0101
-0.5	-0.4914	-0.4841	-0.0086	-0.0073	-0.0159	0.005312	0.0729	0.0052	0.005565	-0.2182	0.0454

Table 4.8 Estimates using approximations of $\text{bias}_{\text{error}}$ and $\text{var}(\hat{\kappa})$ following initial derivative estimation by 5-point moving-arc cubics for constant error variance. v = 900. A sample of 200 is assumed.

κ	$\hat{\kappa}_{\text{exact}}$	$\mathcal{E}(\hat{\kappa})$	bias_m	bias_e	bias_t	$\text{var}(\hat{\kappa})$	$\sqrt{\text{var}(\hat{\kappa})}$	$\sqrt{\text{var}(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0997	0.0000	-0.0003	-0.0003	0.000217	0.0147	0.0010	0.000217	-0.0204	0.0004
-0.2	-0.1998	-0.1985	-0.0002	-0.0013	-0.0015	0.000850	0.0292	0.0021	0.000852	-0.0514	0.0026
-0.3	-0.2990	-0.2957	-0.0010	-0.0033	-0.0043	0.002382	0.0488	0.0035	0.002401	-0.0881	0.0077
-0.4	-0.3964	-0.3899	-0.0036	-0.0065	-0.0101	0.006048	0.0778	0.0055	0.006150	-0.1299	0.0166
-0.5	-0.4912	-0.4800	-0.0088	-0.0112	-0.0200	0.013500	0.1162	0.0082	0.013900	-0.1721	0.0288

Table 4.9 Simulation results for estimation of κ following initial derivative estimation by 5-point moving-arc cubics for constant error variance. v = 900. n = 200.

κ	$\hat{\kappa}_{\text{exact}}$	$\bar{\hat{\kappa}}$	bias_m	bias_e	bias_t	$\text{var}(\hat{\kappa})$	$\sqrt{\text{var}(\hat{\kappa})}$	$\sqrt{\text{var}(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0997	0.0000	-0.0003	-0.0003	0.000232	0.0152	0.0011	0.000232	-0.0197	0.0004
-0.2	-0.1998	-0.2003	-0.0002	0.0005	0.0003	0.000826	0.0287	0.0020	0.000826	0.0104	0.0001
-0.3	-0.2990	-0.2959	-0.0010	-0.0031	-0.0041	0.001488	0.0386	0.0027	0.001504	-0.1063	0.0112
-0.4	-0.3964	-0.3919	-0.0036	-0.0045	-0.0081	0.002582	0.0508	0.0036	0.002648	-0.1594	0.0248
-0.5	-0.4912	-0.4791	-0.0088	-0.0121	-0.0209	0.005106	0.0715	0.0051	0.005543	-0.2925	0.0788

4.2 Simultaneous Estimation of κ and the Parameters of the Smoothing Function

4.2.1 General Theory

In Section 4.1, minimization of the Sobolev norm-type expression (equation (4.2)) for a specified set of observations resulted in the problem of minimizing

$$\sum_i w\,[\kappa y(t_i) - g(t_i)]^2$$

as a function of κ only, since the parameters associated with the function g were determined by the choice of approximating polynomial and by the time-sequence of observations. In this section, the smoothing function f and the associated derivative-estimating function g are assumed to be linear in their parameters, β_j. Therefore, for the type of polynomial functions, f, to be considered, f and g can be represented by $\underline{\Phi\beta}$ and $\underline{\Phi}'\underline{\beta}$, respectively, where $\underline{\Phi}$ and $\underline{\Phi}'$ are $(m \times p)$, $m \geq n$, matrices and $\underline{\beta}$ is the $(p \times 1)$ vector of parameters. In all cases in this discussion, $\frac{df}{dt} = g(t)$ so that $\underline{\Phi}$ and $\underline{\Phi}'$ are related, as suggested by the notation. The Sobolev norm-type expression can be written

$$\nu^2 = [\underline{Y} - \underline{\Phi\beta}]^T[\underline{Y} - \underline{\Phi\beta}] + w[\underline{\dot{Y}} - \underline{\Phi}'\underline{\beta}]^T[\underline{\dot{Y}} - \underline{\Phi}'\underline{\beta}] \qquad (4.13)$$

where, as before, $\underline{Y}^T = [Y(t_1) \ldots Y(t_n)]$ and $\underline{\dot{Y}}$ is the vector of elements, $\kappa\, Y(t_i)$, in our case evaluated with the observed values of the variables. Since (4.13) can be expanded,

$$\nu^2 = \underline{Y}^T\underline{Y} - 2\underline{\beta}^T\underline{\Phi}^T\underline{Y} + \underline{\beta}^T\underline{\Phi}^T\underline{\Phi}\,\underline{\beta} + w\underline{\dot{Y}}^T\underline{\dot{Y}} - 2w\underline{\beta}^T\underline{\Phi}'^T\underline{\dot{Y}} + w\underline{\beta}^T\underline{\Phi}'^T\underline{\Phi}'\underline{\beta}$$

so that, in the simple case $\frac{dy}{dt} = \kappa y$,

$$\nu^2 = \underline{Y}^T\underline{Y} - 2\underline{\beta}^T\underline{\Phi}^T\underline{Y} + \underline{\beta}^T\underline{\Phi}^T\underline{\Phi}\ \underline{\beta} + w\kappa^2\underline{Y}^T\underline{Y} - 2w\kappa\underline{\beta}^T\underline{\Phi}'^T\underline{Y} + w\underline{\beta}^T\underline{\Phi}'^T\underline{\Phi}'\underline{\beta} . \tag{4.14}$$

Taking $\frac{\partial(\nu^2)}{\partial\kappa}$ and $\frac{\partial(\nu^2)}{\partial\underline{\beta}}$ and setting the resulting expressions equal to zero yields

$$\frac{\partial(\nu^2)}{\partial\kappa} = 2\kappa w\underline{Y}^T\underline{Y} - 2w\underline{\beta}^T\underline{\Phi}'^T\underline{Y}$$

$$\frac{\partial(\nu^2)}{\partial\underline{\beta}} = -2\underline{\Phi}^T\underline{Y} + 2\underline{\Phi}^T\underline{\Phi}\underline{\beta} - 2w\kappa\underline{\Phi}'^T\underline{Y} + 2w\underline{\Phi}'^T\underline{\Phi}'\underline{\beta} \tag{4.15}$$

and

$$\begin{aligned} &\hat{\kappa}\underline{Y}^T\underline{Y} - 2\hat{\underline{\beta}}^T\underline{\Phi}'^T\underline{Y} = 0 \\ &\underline{\Phi}^T\underline{\Phi}\hat{\underline{\beta}} + w\underline{\Phi}'^T\underline{\Phi}'\hat{\underline{\beta}} - w\hat{\kappa}\underline{\Phi}'^T\underline{Y} = \underline{\Phi}^T\underline{Y} . \end{aligned} \tag{4.16}$$

These normal equations (4.16) can be written

$$\left[\begin{array}{c:c} \underline{\Phi}^T\underline{\Phi} + w\underline{\Phi}'^T\underline{\Phi}' & -w\underline{\Phi}'^T\underline{Y} \\ (p\times p) & (p\times 1) \\ \hdashline -w\underline{Y}^T\underline{\Phi}' & w\underline{Y}^T\underline{Y} \\ & (1\times 1) \end{array}\right] \left[\begin{array}{c} \hat{\underline{\beta}} \\ \hdashline \hat{\kappa} \end{array}\right] = \left[\begin{array}{c} \underline{\Phi}^T\underline{Y} \\ \hdashline 0 \end{array}\right] \tag{4.17}$$

$$[(p+1)\times(p+1)]$$

which is of the form $\underline{A}\ \underline{\gamma} = \underline{\Psi}$. Assuming $\underline{A}^{-1}$ exists and letting

$$\underline{A}^{-1} = \underline{G} = \left[\begin{array}{c:c} G_{11} & G_{12} \\ \hdashline G_{21} & G_{22} \end{array}\right]$$

be the partitioned inverse of $\underline{A}$, then

$$G_{21} = - [w\underline{Y}^T\underline{Y} - (-w\underline{Y}^T\underline{\Phi}')(\underline{\Phi}^T\underline{\Phi} + w\underline{\Phi}'^T\underline{\Phi}')^{-1} (-w\underline{\Phi}'^T\underline{Y})]^{-1}$$

$$(-w\underline{Y}^T\underline{\Phi}')(\underline{\Phi}^T\underline{\Phi} + w\underline{\Phi}'^T\underline{\Phi}')^{-1} ,$$

Since $\hat{\kappa} = \underline{G}_{21}\underline{\Phi}^T\underline{Y}$ from (4.17), it follows that

$$\hat{\kappa} = \frac{w\underline{Y}^T\underline{\Phi}' (\underline{\Phi}^T\underline{\Phi} + w\underline{\Phi}'^T\underline{\Phi}')^{-1}\underline{\Phi}^T\underline{Y}}{w\underline{Y}^T\underline{Y} - w\underline{Y}^T\underline{\Phi}' (\underline{\Phi}^T\underline{\Phi} + w\underline{\Phi}'^T\underline{\Phi}')^{-1} w\underline{\Phi}'^T\underline{Y}} \qquad (4.18)$$

or, for $w = 1(\text{time-unit})^2$,

$$\hat{\kappa} = \frac{\underline{Y}^T\underline{\underline{E}}\underline{Y}}{\underline{Y}^T\underline{\underline{F}}\underline{Y}} \qquad (4.19)$$

where

$$\underline{E} = \underline{\Phi}'(\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}')^{-1} \underline{\Phi}^T$$

$$\underline{F} = \underline{I} - \underline{\Phi}'(\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}')^{-1} \underline{\Phi}'^T$$

As in Section 4.1 (expressions (4.6) and (4.7)), this form of $\hat{\kappa}$ permits the construction of an approximate expression for $\text{bias}_{\text{error}}$. Specifically, if $\underline{Y} = \underline{y} + \underline{\varepsilon}$, then

$$\hat{\kappa} = \frac{(\underline{y} + \underline{\varepsilon})^T\underline{E}(\underline{y} + \underline{\varepsilon})}{(\underline{y} + \underline{\varepsilon})^T\underline{F}(\underline{y} + \underline{\varepsilon})}$$

$$= \frac{\underline{y}^T\underline{Ey}}{\underline{y}^T\underline{Fy}} + \frac{\underline{y}^T\underline{Ey}}{\underline{y}^T\underline{Fy}} \left[- \frac{\underline{y}^T\underline{F\varepsilon} + \underline{\varepsilon}^T\underline{Fy} + \underline{\varepsilon}^T\underline{F\varepsilon}}{(\underline{y} + \underline{\varepsilon})^T\underline{F}(\underline{y} + \underline{\varepsilon})} \right] + \frac{\underline{y}^T\underline{E\varepsilon} + \underline{\varepsilon}^T\underline{Ey} + \underline{\varepsilon}^T\underline{E\varepsilon}}{(\underline{y} + \underline{\varepsilon})^T\underline{F}(\underline{y} + \underline{\varepsilon})} \qquad (4.20)$$

As discussed in Section 3.4, since

$$\hat{\kappa}_{exact} = (\underline{y}^T\underline{E}\underline{y})(\underline{y}^T\underline{F}\underline{y})^{-1} ,$$

then

$$\hat{\kappa} = \hat{\kappa}_{exact} + \hat{\kappa}_{exact}\left[-\frac{\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{y} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon}}{(\underline{y}+\underline{\varepsilon})^T\underline{F}(\underline{y}+\underline{\varepsilon})}\right] + \frac{\underline{y}^T\underline{E}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{E}\underline{y} + \underline{\varepsilon}^T\underline{E}\underline{\varepsilon}}{(\underline{y}+\underline{\varepsilon})^T\underline{F}(\underline{y}+\underline{\varepsilon})} , \quad (4.21)$$

the last two terms of which are an estimate of (-1) bias_{error} .

Further, if it is assumed, as in Section 4.1, that

$$\frac{1}{\underline{y}^T\underline{F}\underline{y}}\left[1 - \frac{2\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon}}{\underline{y}^T\underline{F}\underline{y}}\right]$$

is an adequate approximation to

$$\frac{1}{\underline{y}^T\underline{F}\underline{y} + 2\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon}} ,$$

then $\hat{\kappa}$ can be approximated by

$$\hat{\kappa} \doteq \hat{\kappa}_{exact} + \hat{\kappa}_{exact}\left[-\left(\frac{1}{\underline{y}^T\underline{F}\underline{y}}\right)\left(1 - \frac{2\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon}}{\underline{y}^T\underline{F}\underline{y}}\right)\right](2\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon})$$

$$+ \left(\frac{1}{\underline{y}^T\underline{F}\underline{y}}\right)\left(1 - \frac{2\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon}}{\underline{y}^T\underline{F}\underline{y}}\right)(\underline{y}^T\underline{E}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{E}\underline{y} + \underline{\varepsilon}^T\underline{E}\underline{\varepsilon}) \quad (4.22)$$

the second and third terms of which are an approximation to (-1) bias_{error} . Since (4.22) is linear in the $\varepsilon(t_i)$ and in integer powers of the $\varepsilon(t_i)$, the expected value of the approximate

$(-1)\ \text{bias}_{\text{error}}$ (and, hence, of the approximation to $\hat{\kappa}$ by (4.22)) can be computed by equation (4.11), provided the distribution of $\underline{\varepsilon}$ is known or assumed.

If $\dfrac{1}{\underline{y}^T\underline{F}\underline{y}}$ is an adequate approximation to $\dfrac{1}{\underline{y}^T\underline{F}\underline{y} + 2\underline{y}^T\underline{F}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{F}\underline{\varepsilon}}$,

then, by expression (4.12a), an approximation to the variance of $\hat{\kappa}$ can be computed.

4.2.2 Numerical Examples

In this section, some of the properties of the estimator $\hat{\kappa}$ (equation (4.19)) are investigated from sets of artificially constructed observations $Y(t_i)$. For several values of κ , values of $y(t_i)$, $t_i = 1, 2, \ldots, 15$ are computed by

$$y(t_i) = 1000e^{-\kappa t_i}$$

and sets of observations $Y(t_i) = y(t_i) + \varepsilon(t_i)$ were generated for $\varepsilon(t_i) \sim \mathcal{N}(0,v)$, where either

(i) $v = \left[\dfrac{p\ y(t_i)}{100}\right]^2$ or

(ii) $v = \text{constant} \neq 0$ or

(iii) $v = 0$.

Two pairs of functions f and g are considered. First, f is the function composed of three cubic segments with knots at 0.5, 5.5, 10.5, and 15.5 time units. In this case, the matrices $\underline{\Phi}$ and $\underline{\Phi}'$ of Section 4.2.1 are defined in Section 3.3.

The second function f to be considered is the function derived from 5-point moving-arc cubic smoothing and derivative estimation as described in Section 3.2.2. In this case, however, the vector $\underline{Y}$ of observations in equation (4.13) is the (55x1) vector which appears in equation (3.7), so that the matrices $\underline{E}$ and $\underline{F}$ of equation (4.19) are of dimension (55x55). By matching columns of $\underline{E}$ and $\underline{F}$ with the duplicate elements of the $Y(t_i)$ which appear in the (55x1) vector Y, matrices $\underline{E}^*$ and $\underline{F}^*$ of dimension (15x15) can be formed by arithmetic addition of elements of $\underline{E}$ and $\underline{F}$ so that

$$\hat{\kappa} = \frac{\underline{Y}^T \underline{E}^* \underline{Y}}{\underline{Y}^T \underline{F}^* \underline{Y}}$$

where $\underline{Y}^T$ is the vector of 15 observations $Y(t_i)$.

It is important to note that, for a given value of κ , the same random deviates were generated to compute the $\varepsilon(t_i)$.

The tabular results are shown as follows:

	$v = \left[\frac{p\, y(t_i)}{100}\right]^2$, p = 10	v = constant = 900
5-point cubic moving-arc	expected(approx): Table 4.10 simulation: Table 4.11	expected(approx): Table 4.14 simulation: Table 4.15
3 cubic segments	expected(approx): Table 4.12 simulation: Table 4.13	expected(approx): Table 4.16 simulation: Table 4.17

Table 4.10 Estimates using approximations of $bias_{error}$ and $var(\hat{\kappa})$ for simultaneous smoothing by 5-point moving-arc cubics and estimation of κ for proportional error variance. p = 10. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{exact}$	$\mathcal{E}(\hat{\kappa})$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0997	0.0000	-0.0003	-0.0003	0.000099	0.0099	0.0007	0.000099	-0.0302	0.0009
-0.2	-0.1999	-0.1994	-0.0001	-0.0005	-0.0006	0.000264	0.0162	0.0011	0.000264	-0.0369	0.0014
-0.3	-0.2995	-0.2986	-0.0005	-0.0009	-0.0014	0.000829	0.0288	0.0020	0.000831	-0.0486	0.0024
-0.4	-0.3981	-0.3968	-0.0019	-0.0013	-0.0032	0.002468	0.0497	0.0035	0.002478	-0.0644	0.0041
-0.5	-0.4945	-0.4925	-0.0055	-0.0020	-0.0075	neg	-	-	-	-	-

Table 4.11 Simulation results for simultaneous smoothing by 5-point moving-arc cubics and estimation of κ for proportional error variance. p = 10. n = 200.

κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0987	0.0000	-0.0013	-0.0013	0.000090	0.0095	0.0007	0.000092	-0.1370	0.0184
-0.2	-0.1999	-0.2004	-0.0001	0.0005	0.0004	0.000253	0.0159	0.0011	0.000253	0.0251	0.0006
-0.3	-0.2995	-0.2990	-0.0005	-0.0005	-0.0010	0.000596	0.0244	0.0017	0.000597	-0.0410	0.0017
-0.4	-0.3981	-0.3994	-0.0019	0.0013	-0.0006	0.000811	0.0285	0.0020	0.000811	-0.0211	0.0004
-0.5	-0.4945	-0.4892	-0.0055	-0.0053	-0.0108	0.001208	0.0348	0.0025	0.001325	-0.3107	0.0880

Table 4.12 Estimates using approximations of $bias_{error}$ and $var(\hat{\kappa})$ for simultaneous smoothing by three cubic segments and estimation of κ for proportional error variance. p = 10. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{exact}$	$\mathcal{E}(\hat{\kappa})$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0997	0.0000	-0.0003	-0.0003	0.000230	0.0152	0.0011	0.000230	-0.0198	0.0004
-0.2	-0.1999	-0.1992	-0.0001	-0.0007	-0.0008	0.000793	0.0282	0.0020	0.000794	-0.0284	0.0008
-0.3	-0.2991	-0.2978	-0.0009	-0.0013	-0.0022	0.002151	0.0464	0.0035	0.002155	-0.0474	0.0022
-0.4	-0.3968	-0.3946	-0.0032	-0.0022	-0.0054	0.005323	0.0730	0.0052	0.005352	-0.0740	0.0054
-0.5	-0.4913	-0.4879	-0.0087	-0.0034	-0.0121	0.012774	0.1130	0.0080	0.012920	-0.1071	0.0113

Table 4.13 Simulation results for simultaneous smoothing by three cubic segments and estimation of κ for proportional error variance. p = 10. n = 200.

κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0988	0.0000	-0.0012	-0.0012	0.000237	0.0154	0.0011	0.000238	-0.0779	0.0060
-0.2	-0.1999	-0.2025	-0.0001	0.0026	0.0025	0.000757	0.0275	0.0019	0.000763	0.0909	0.0082
-0.3	-0.2991	-0.2974	-0.0009	-0.0017	-0.0026	0.001485	0.0385	0.0027	0.001491	-0.0675	0.0045
-0.4	-0.3968	-0.3982	-0.0032	0.0014	-0.0018	0.001874	0.0433	0.0031	0.001877	-0.0416	0.0017
-0.5	-0.4913	-0.4841	-0.0087	-0.0072	-0.0159	0.002430	0.0493	0.0035	0.002683	-0.3225	0.0942

Table 4.14 Estimates using approximations of $\text{bias}_{\text{error}}$ and $\text{var}(\hat{\kappa})$ for simultaneous smoothing by 5-point moving-arc cubics and estimation of κ for constant error variance. v = 900. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{exact}$	$\mathcal{E}(\hat{\kappa})$	bias_m	bias_e	bias_t	$\text{var}(\hat{\kappa})$	$\sqrt{\text{var}(\hat{\kappa})}$	$\sqrt{\text{var}(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0997	0.0000	-0.0003	-0.0003	0.000004	0.0020	0.0001	0.000004	-0.1487	0.0216
-0.2	-0.1999	-0.1981	-0.0001	-0.0018	-0.0019	neg	-	-	-	-	-
-0.3	-0.2995	-0.2934	-0.0005	-0.0061	-0.0066	neg	-	-	-	-	-
-0.4	-0.3981	-0.3825	-0.0019	-0.0156	-0.0175	0.000134	0.0116	0.0008	0.000441	-1.5090	0.6949
-0.5	-0.4945	-0.4615	-0.0055	-0.0330	-0.0385	0.000264	0.0162	0.0011	0.001746	-2.3707	0.8490

Table 4.15 Simulation results for simultaneous smoothing by 5-point moving-arc cubics and estimation of κ for constant error variance. v = 900. n = 200.

κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	bias_m	bias_e	bias_t	$\text{var}(\hat{\kappa})$	$\sqrt{\text{var}(\hat{\kappa})}$	$\sqrt{\text{var}(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0992	0.0000	-0.0008	-0.0008	0.000019	0.0044	0.0003	0.000020	-0.1818	0.0325
-0.2	-0.1999	-0.1979	-0.0001	-0.0020	-0.0021	0.000108	0.0104	0.0007	0.000112	-0.2019	0.0392
-0.3	-0.2995	-0.2933	-0.0005	-0.0062	-0.0067	0.000340	0.0184	0.0013	0.000385	-0.3641	0.1166
-0.4	-0.3981	-0.3837	-0.0019	-0.0144	-0.0163	0.000626	0.0250	0.0018	0.000892	-0.6520	0.2979
-0.5	-0.4945	-0.4589	-0.0055	-0.0356	-0.0411	0.001341	0.0366	0.0026	0.003030	-1.1222	0.5574

Table 4.16 Estimates using approximations of $\text{bias}_{\text{error}}$ and $\text{var}(\hat{\kappa})$ for simultaneous smoothing by three cubic segments and estimation of κ for constant error variance. v = 900. A sample size of 200 is assumed.

κ	$\hat{\kappa}_{\text{exact}}$	$\mathcal{E}(\hat{\kappa})$	bias_m	bias_e	bias_t	$\text{var}(\hat{\kappa})$	$\sqrt{\text{var}(\hat{\kappa})}$	$\sqrt{\text{var}(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0997	0.0000	-0.0003	-0.0003	0.000029	0.0054	0.0004	0.000029	-0.0559	0.0031
-0.2	-0.1999	-0.1986	-0.0001	-0.0013	-0.0014	0.000077	0.0088	0.0006	0.000079	-0.1600	0.0250
-0.3	-0.2991	-0.2952	-0.0009	-0.0039	-0.0048	0.000549	0.0234	0.0017	0.000573	-0.2048	0.0402
-0.4	-0.3968	-0.3879	-0.0032	-0.0089	-0.0121	0.002221	0.0471	0.0033	0.002367	-0.2568	0.0619
-0.5	-0.4913	-0.4737	-0.0087	-0.0176	-0.0263	neg	-	-	-	-	-

Table 4.17 Simulation results for simultaneous smoothing by three cubic segments and estimation of κ for constant error variance. v = 900. n = 200.

κ	$\hat{\kappa}_{\text{exact}}$	$\bar{\hat{\kappa}}$	bias_m	bias_e	bias_t	$\text{var}(\hat{\kappa})$	$\sqrt{\text{var}(\hat{\kappa})}$	$\sqrt{\text{var}(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
-0.1	-0.1000	-0.0993	0.0000	-0.0007	-0.0007	0.000035	0.0059	0.0004	0.000035	-0.1186	0.0138
-0.2	-0.1999	-0.1995	-0.0001	-0.0004	-0.0005	0.000164	0.0128	0.0009	0.000164	-0.0391	0.0015
-0.3	-0.2991	-0.2946	-0.0009	-0.0045	-0.0054	0.000434	0.0208	0.0015	0.000463	-0.2596	0.0630
-0.4	-0.3968	-0.3895	-0.0032	-0.0073	-0.0105	0.000731	0.0270	0.0019	0.000842	-0.3889	0.1309
-0.5	-0.4913	-0.4715	-0.0087	-0.0198	-0.0285	0.001477	0.0384	0.0027	0.002290	-0.7422	0.3548

These results suggest the following:

(1) The methods involving simultaneous smoothing, derivative estimation and estimation of κ yield estimates of κ with smaller $bias_{total}$ and smaller variance than the methods involving initial smoothing.

(2) For the case of proportional variance, the method of simultaneous 5-point moving-arc cubic smoothing and estimation of κ yields estimates of κ with smaller $bias_{total}$ and smaller variance than the method involving three cubic segments. However, for the case of constant variance, the method of cubic segments appears to yield slightly better estimates.

5. GENERALIZATIONS OF THE SIMPLE CASE

Two natural extensions of the work in Chapter 4 are obvious: (1) application of the simple case formulation to a single differential equation of more than one term and (2) application of the simple case formulation to a system of differential equations. In Section 5.1, the estimation of κ_1 and κ_2 in the model equation

$$\dot{y}(t) = \kappa_1 y(t) + \kappa_2 z(t)$$

is investigated where the function f in the Sobolev norm-type expression (equation (3.1)) is either the point function derived from initial 5-point moving-arc polynomial smoothing, or the function derived from simultaneous moving-arc polynomial smoothing and estimation of the κ_i, or the function derived from the simultaneous fitting of cubic segments and estimation of the κ_i, and where $g = f'$.

Section 5.2 includes a discussion of the estimation of the elements of the (2X2) matrix $\underline{\kappa}$ in the system

$$\begin{bmatrix} \dot{y} \\ \dot{z} \end{bmatrix} = \begin{bmatrix} \kappa_1 & \kappa_2 \\ \kappa_3 & \kappa_4 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} .$$

The formulation leading to $\hat{\kappa}_i$ is shown to closely parallel that of the simple case.

Section 5.3 is devoted to a discussion of the estimators of κ_1 and κ_2 in the system

$$\dot{y} = -\kappa_1 y + \kappa_2 z$$
$$\dot{z} = \qquad - \kappa_2 z .$$

5.1 Estimation of κ_1 and κ_2 in $\dot{y} = \kappa_1 y + \kappa_2 z$

Consider the single differential equation

$$\frac{dy(t)}{dt} = \kappa_1 y(t) + \kappa_2 z(t) \tag{5.1}$$

where κ_1 and κ_2 are constants to be estimated given observed values $Y(t_i) = y(t_i) + \varepsilon(t_i)$ and $Z(t_i) = z(t_i) + \zeta(t_i)$, $i = 1, \ldots, n$. As in the previous chapter, discussion of estimating κ_1 and κ_2 following 5-point moving-arc polynomial smoothing and derivative estimation will be followed by discussion of simultaneous smoothing and estimation of κ_1 and κ_2 for the cases involving cubic segments and 5-point moving-arc cubic smoothing.

5.1.1 Estimation of $\underline{\kappa}$ Following Initial 5-point Polynomial Smoothing and Derivative Estimation

For the differential equation (5.1) above, estimation of κ_1 and κ_2 by minimization of the Sobolev norm-type expression (3.1) following initial 5-point polynomial moving-arc smoothing and derivative estimation reduces to the problem of minimizing the term

$$w \sum_{i=1}^{n} (\dot{Y}(t_i) - g(t_i))^2$$

of equation (3.1) or minimizing the equivalent expression

$$w\,\underline{\delta}^T\underline{\delta} = w(\kappa_1\underline{SY} + \kappa_2\underline{SZ} - \underline{DY})^T\,(\kappa_1\underline{SY} + \kappa_2\underline{SZ} - \underline{DY}) \tag{5.2}$$

where, as defined in Section 3.2, $\underline{S}$ is either the smoothing matrix associated with initial 5-point moving-arc polynomial smoothing or $\underline{S} = \underline{I}$. Expression (5.2) could have arisen independently of the

Sobolev norm-type expression if the differential equation (5.1) is written in the approximation form

$$\underline{D}\,\underline{y} = \kappa_1 \underline{S}\,\underline{y} + \kappa_2 \underline{S}\,\underline{z}$$

$$= \underline{S}\,[\underline{y} \quad \underline{z}] \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix}$$

and, subsequently, in terms of the observed $\underline{Y}$ and $\underline{Z}$, as

$$\underline{D}\,\underline{Y} = \underline{S}\,[\underline{Y} \quad \underline{Z}] \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} + \underline{\delta}^* \qquad (5.3)$$

where $\underline{Y} = \underline{y} + \underline{\varepsilon}$, and $\underline{Z} = \underline{z} + \underline{\varepsilon}$, and $\underline{\delta}^* = \underline{D}\,\underline{\varepsilon} - (\kappa_1 \underline{S\varepsilon} + \kappa_2 \underline{S\zeta})$. We find that, since $w\,\underline{\delta}^{*T}\underline{\delta}^*$ is, in fact, expression (5.2), estimates of κ_1 and κ_2 derived from minimizing $w\,\underline{\delta}^{*T}\underline{\delta}^*$ are identical to those derived from minimizing $w\,\underline{\delta}^T\underline{\delta}$.

To minimize expression (5.2), we take

$$\frac{\partial(\underline{\delta}^T\underline{\delta})}{\partial\underline{\kappa}} = -2(\underline{S}[\underline{Y} \quad \underline{Z}])^T\,\underline{D}\,\underline{Y} + 2(\underline{S}[\underline{Y} \quad \underline{Z}])^T\,(\underline{S}[\underline{Y} \quad \underline{Z}])\,\underline{\kappa}\,.$$

Setting $\frac{\partial(\underline{\delta}^T\underline{\delta})}{\partial\underline{\kappa}} = 0$,

$$(\underline{S}[\underline{Y} \quad \underline{Z}])^T\,(\underline{S}[\underline{Y} \quad \underline{Z}])\,\hat{\underline{\kappa}} = (\underline{S}[\underline{Y} \quad \underline{Z}])^T\,\underline{D}\,\underline{Y}$$

we find

$$\hat{\underline{\kappa}} = ([\underline{Y} \quad \underline{Z}]^T\,\underline{S}^T\underline{S}\,[\underline{Y} \quad \underline{Z}])^{-1}[\underline{Y} \quad \underline{Z}]^T\,\underline{S}^T\underline{D}\,\underline{Y}$$

or

$$\begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \end{bmatrix} = \left[\begin{bmatrix} \underline{Y}^T \\ \underline{Z}^T \end{bmatrix} \underline{S}^T\underline{S} \, [\underline{Y} \quad \underline{Z}] \right]^{-1} \begin{bmatrix} \underline{Y}^T \\ \underline{Z}^T \end{bmatrix} \underline{S}^T\underline{D} \, \underline{Y} \tag{5.4a}$$

$$= \begin{bmatrix} \underline{Y}^T\underline{S}^T\underline{SY} & \underline{Y}^T\underline{S}^T\underline{SZ} \\ \underline{Z}^T\underline{S}^T\underline{SY} & \underline{Z}^T\underline{S}^T\underline{SZ} \end{bmatrix}^{-1} \begin{bmatrix} \underline{Y}^T\underline{S}^T\underline{DY} \\ \underline{Z}^T\underline{S}^T\underline{DY} \end{bmatrix} . \tag{5.4b}$$

Since $\underline{S}^T\underline{S}$ is symmetric, it follows that $\underline{Z}^T\underline{S}^T\underline{SY} = \underline{Y}^T\underline{S}^T\underline{SZ}$

so

$$d = \det \begin{bmatrix} \underline{Y}^T\underline{S}^T\underline{SY} & \underline{Y}^T\underline{S}^T\underline{SZ} \\ \underline{Z}^T\underline{S}^T\underline{SY} & \underline{Z}^T\underline{S}^T\underline{SZ} \end{bmatrix} = \underline{Z}^T\underline{S}^T\underline{SZY}^T\underline{S}^T\underline{SY} - \underline{Z}^T\underline{S}^T\underline{SYZ}^T\underline{S}^T\underline{SY}$$

$$= \underline{Z}^T\underline{S}^T\underline{S}(\underline{ZY}^T - \underline{YZ}^T)\underline{S}^T\underline{SY} .$$

Therefore,

$$\begin{bmatrix} \hat{K}_1 \\ \hat{K}_2 \end{bmatrix} = \frac{1}{d} \begin{bmatrix} \underline{Z}^T\underline{S}^T\underline{SZ} & -\underline{Z}^T\underline{S}^T\underline{SY} \\ -\underline{Y}^T\underline{S}^T\underline{SZ} & \underline{Y}^T\underline{S}^T\underline{SY} \end{bmatrix} \begin{bmatrix} \underline{Y}^T\underline{S}^T\underline{DY} \\ \underline{Z}^T\underline{S}^T\underline{DY} \end{bmatrix}$$

$$= \begin{bmatrix} \underline{Z}^T\underline{S}^T\underline{SZY}^T\underline{S}^T\underline{DY} - \underline{Z}^T\underline{S}^T\underline{SYZ}^T\underline{S}^T\underline{DY} \\ -\underline{Y}^T\underline{S}^T\underline{SZY}^T\underline{S}^T\underline{DY} + \underline{Y}^T\underline{S}^T\underline{SYZ}^T\underline{S}^T\underline{DY} \end{bmatrix} \frac{1}{d} . \tag{5.5}$$

However, $Y(t_i) = y(t_i) + \varepsilon(t_i)$ and $Z(t_i) = z(t_i) + \zeta(t_i)$, which leads to the expression

$$\hat{\kappa}_1 = \frac{\underline{z}^T\underline{S}^T\underline{S}(\underline{zy}^T - \underline{yz}^T)\ \underline{S}^T\underline{Dy}}{\underline{z}^T\underline{S}^T\underline{S}(\underline{zy}^T - \underline{yz}^T)\ \underline{S}^T\underline{Sy}} + \frac{a}{b}$$

$$= \hat{\kappa}_{1\,exact} + \frac{a}{b} \tag{5.6}$$

where, with $\underline{C} = \underline{S}^T\underline{S}$

$$a = \left\{\underline{z}^T\underline{C}(\underline{zy}^T-\underline{yz}^T)\underline{Cy}\right\}\left\{(\underline{z}+\underline{\zeta})^T\underline{C}\left[(\underline{z}+\underline{\zeta})(\underline{y}+\underline{\varepsilon})^T - (\underline{y}+\underline{\varepsilon})(\underline{z}+\underline{\zeta})^T\right]\underline{B}(\underline{y}+\underline{\varepsilon})\right\}$$
$$- \left\{\underline{z}^T\underline{C}(\underline{zy}^T- \underline{yz}^T)\underline{By}\right\}\left\{(\underline{z}+\underline{\zeta})^T\underline{C}\left[(\underline{z}+\underline{\zeta})(\underline{y}+\underline{\varepsilon})^T - (\underline{y}+\underline{\varepsilon})(\underline{z}+\underline{\zeta})^T\right]\underline{C}(\underline{y}+\underline{\varepsilon})\right\}$$

and

$$b = \left\{\underline{z}^T\underline{C}(\underline{zy}^T-\underline{yz}^T)\underline{Cy}\right\}\left\{(\underline{z}+\underline{\zeta})^T\underline{C}\left[(\underline{z}+\underline{\zeta})(\underline{y}+\underline{\varepsilon})^T - (\underline{y}+\underline{\varepsilon})(\underline{z}+\underline{\zeta})^T\right]\underline{C}(\underline{y}+\underline{\varepsilon})\right\}$$

A similar expression for $\hat{\kappa}_2$ is obvious. The complexity and length of the expressions for the quantities a and b preclude efficient investigation of analytical forms of the expected value of the $\text{bias}_{\text{error}}$, as we did in Chapter 4, necessitating investigation of bias by Monte Carlo simulation with repeated estimation of κ_1 and κ_2 from artificially constructed data.

5.1.2 Estimation of κ_1 and κ_2 with Simultaneous Smoothing

Estimation of κ_1 and κ_2 in the differential equation

$$\dot{y}(t) = \kappa_1 y(t) + \kappa_2 z(t)$$

by minimization of the Sobolev norm-type expression (equation (3.1)) where f is either the function composed of three cubic segments or the function derived from simultaneous smoothing and derivative estimation and concomitant estimation of the κ_i , and where $g = f'$ can be accomplished by setting the appropriate partial derivatives equal to zero and solving the resulting normal equations for the unknown parameters. The normal equations in matrix form are

$$\left[\begin{array}{c|cc} \underline{\Phi}^T\underline{\Phi} + w\underline{\Phi}'^T\underline{\Phi}' & -w\underline{\Phi}'^T\underline{Y} & -w\underline{\Phi}'^T\underline{Z} \\ \hline -w\underline{Y}^T\underline{\Phi}' & w\underline{Y}^T\underline{Y} & w\underline{Y}^T\underline{Z} \\ -w\underline{Z}^T\underline{\Phi}' & w\underline{Y}^T\underline{Z} & w\underline{Z}^T\underline{Z} \end{array}\right] \left[\begin{array}{c} \hat{B}_1 \\ \hat{B}_2 \\ \vdots \\ \hat{B}_8 \\ \hline \hat{\kappa}_1 \\ \hat{\kappa}_2 \end{array}\right] = \left[\begin{array}{c} \underline{\Phi}^T\underline{Y} \\ \hline 0 \\ 0 \end{array}\right]$$

where $\underline{\Phi}$ is the matrix associated with the particular function of segmented cubics or 5-point moving-arc cubic smoothing as defined in Sections 3.2.2 and 3.3. The partitioned form of the matrix in the normal equations leads to a natural partitioning of the usual inverse, so that the vector $\underline{\hat{\kappa}}^T = (\hat{\kappa}_1, \hat{\kappa}_2)$ can be expressed as (for $w = 1$ time-unit-squared)

$$\begin{bmatrix} \hat{\kappa}_1 \\ \hat{\kappa}_2 \end{bmatrix} = \left[\begin{bmatrix} \underline{Y}^T \\ \underline{Z}^T \end{bmatrix} \underline{E} \, [\underline{Y} \; \underline{Z}] \right]^{-1} \begin{bmatrix} \underline{Y}^T \\ \underline{Z}^T \end{bmatrix} \underline{F} \, \underline{Y} \tag{5.8}$$

$$= \begin{bmatrix} \underline{Y}^T \underline{E} \, \underline{Y} & \underline{Y}^T \underline{E} \, \underline{Z} \\ \underline{Z}^T \underline{E} \, \underline{Y} & \underline{Z}^T \underline{E} \, \underline{Z} \end{bmatrix}^{-1} \begin{bmatrix} \underline{Y}^T \underline{F} \, \underline{Y} \\ \underline{Z}^T \underline{F} \, \underline{Y} \end{bmatrix}$$

where

$$\underline{F} = \underline{\Phi}'(\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}')^{-1}\underline{\Phi}^T$$

$$\underline{E} = \underline{I} - \underline{\Phi}'(\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}')^{-1}\underline{\Phi}'^T \quad .$$

Paralleling the argument in Section 5.1.1,

$$\hat{k}_1 = \frac{\underline{Z}^T\underline{E}(\underline{Z}\underline{Y}^T - \underline{Y}\underline{Z}^T)\ \underline{F}\ \underline{Y}}{\underline{Z}^T\underline{E}(\underline{Z}\underline{Y}^T - \underline{Y}\underline{Z}^T)\ \underline{E}\ \underline{Y}}$$

$$\hat{k}_2 = -\frac{\underline{Y}^T\underline{E}(\underline{Z}\underline{Y}^T - \underline{Y}\underline{Z}^T)\ \underline{F}\ \underline{Y}}{\underline{Z}^T\underline{E}(\underline{Z}\underline{Y}^T - \underline{Y}\underline{Z}^T)\ \underline{E}\ \underline{Y}} \quad . \tag{5.9}$$

and, so

$$\hat{k}_1 = \frac{\underline{z}^T\underline{E}(\underline{z}\underline{y}^T - \underline{y}\underline{z}^T)\ \underline{F}\ \underline{y}}{\underline{z}^T\underline{E}(\underline{z}\underline{y}^T - \underline{y}\underline{z}^T)\ \underline{E}\ \underline{y}} + \frac{a}{b}$$

$$= \hat{k}_{1\,exact} + \frac{a}{b} \tag{5.10}$$

where a and b are as in expressions (5.7) upon substituting $\underline{E}$ for $\underline{C}$ and $\underline{F}$ for $\underline{B}$. A similar expression is obvious for $\hat{k}_2$.

5.2 Generalization to a System of Differential Equations

Consider the system of differential equations

$$\frac{dy}{dt} = k_1 y(t) + k_2 z(t)$$

$$\frac{dz}{dt} = k_3 y(t) + k_4 z(t) \tag{5.11}$$

and assume that observations $\{Y(t_i), Z(t_i):\ i = 1, 2, \ldots, n\}$ are given where $Y(t_i) = y(t_i) + \varepsilon(t_i)$ and $Z(t_i) = z(t_i) + \zeta(t_i)$.

It follows from (5.11) that

$$\begin{bmatrix} \underline{\dot{y}} \\ \\ \underline{\dot{z}} \end{bmatrix} = \begin{bmatrix} \underline{y} & \underline{z} & \underline{0} & \underline{0} \\ \\ \underline{0} & \underline{0} & \underline{y} & \underline{z} \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \end{bmatrix} \qquad (5.12)$$

where

$$\underline{\dot{y}}^T = (\dot{y}(t_1)\ \dot{y}(t_2)\ \ldots\ \dot{y}(t_n))$$

$$\underline{y}^T = (y(t_1)\ y(t_2)\ \ldots\ y(t_n))$$

$$\underline{0}^T = (\ 0 \quad 0 \quad \ldots \quad 0\ )$$

and similarly for $\underline{\dot{z}}$ and $\underline{z}$.

For the system (5.11), we define the Sobolev norm-type expression

$$W^{*2} = \Sigma\ [(Y(t_i) - f_1(t_i))^2 + w(\dot{Y}(t_i) - g_1(t_i))^2$$

$$+ (Z(t_i) - f_2(t_i))^2 + w(\dot{Z}(t_i) - g_2(t_i))^2] \qquad (5.13)$$

where $\dot{Y}(t_i) = \kappa_1 Y(t_i) + \kappa_2 Z(t_i)$ and $\dot{Z}(t_i) = \kappa_3 Y(t_i) + \kappa_4 Z(t_i)$.

If the values of $f_1(t_i)$ and $f_2(t_i)$ are the elements of the vectors $\underline{SY}$ and $\underline{SZ}$, respectively, where $\underline{S}$ is the matrix associated with initial 5-point moving-arc polynomial smoothing, and if $g_1(t_i)$ and $g_2(t_i)$ are the derivative estimates $\underline{DY}$ and $\underline{DZ}$, respectively, where $\underline{D}$ is the matrix associated with $\underline{S}$ as defined in Section 3.2.1, then minimization with respect to the κ_i of

$$\ell^{*2} = (\underline{Y} - \underline{SY})^T(\underline{Y} - \underline{SY}) + w(\kappa_1\underline{Y} + \kappa_2\underline{Z} - \underline{DY})^T(\kappa_1\underline{Y} + \kappa_2\underline{Z} - \underline{DY})$$

$$+ (\underline{Z} - \underline{SZ})^T(\underline{Z} - \underline{SZ}) + w(\kappa_3\underline{Y} + \kappa_4\underline{Z} - \underline{DZ})^T(\kappa_3\underline{Y} + \kappa_4\underline{Z} - \underline{DZ})$$

can be considered as a problem involving minimization of

$$w(\kappa_1\underline{Y} + \kappa_2\underline{Z} - \underline{DY})^T(\kappa_1\underline{Y} + \kappa_2\underline{Z} - \underline{DY})$$

$$+ w(\kappa_3\underline{Y} + \kappa_4\underline{Z} - \underline{DZ})^T(\kappa_3\underline{Y} + \kappa_4\underline{Z} - \underline{DZ}) \qquad (5.14)$$

since the terms of ℓ^{*2} involving $\underline{S}$ are independent of the κ_i. As in the univariate case, the details of constructing the estimators $\hat{\kappa}_i$ will actually be derived from minimizing

$$w(\kappa_1\underline{SY} + \kappa_2\underline{SZ} - \underline{DY})^T(\kappa_1\underline{SY} + \kappa_2\underline{SZ} - \underline{DY})$$

$$+ w(\kappa_3\underline{SY} + \kappa_4\underline{SZ} - \underline{DZ})^T(\kappa_3\underline{SY} + \kappa_4\underline{SZ} - \underline{DZ}) \qquad (5.15)$$

with respect to the κ_i rather than minimization of expression (5.14). On the other hand, we can write equation (5.12) in the approximation form

$$\begin{bmatrix} \underline{D}\,\underline{y} \\ \underline{D}\,\underline{z} \end{bmatrix} = \begin{bmatrix} \underline{S}\,\underline{y} & \underline{S}\,\underline{z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{S}\,\underline{y} & \underline{S}\,\underline{z} \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \end{bmatrix}$$

or, in terms of $\underline{Y}$ and $\underline{Z}$,

$$\begin{bmatrix} \underline{D}\,\underline{Y} \\ \underline{D}\,\underline{Z} \end{bmatrix} = \begin{bmatrix} \underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \end{bmatrix} + \underline{\delta}$$

where $\underline{S}$ and $\underline{D}$ are the smoothing and derivative estimation matrices, respectively, associated with initial 5-point moving-arc polynomial smoothing and where

$$\underline{\delta} = \begin{bmatrix} \underline{D\varepsilon} \\ \underline{D\zeta} \end{bmatrix} - \begin{bmatrix} \underline{S\varepsilon} & \underline{S\zeta} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{S\varepsilon} & \underline{S\zeta} \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_3 \\ \kappa_4 \end{bmatrix} .$$

One can obtain least-squares-type estimates of the κ_i by minimizing, with respect to the κ_i , the expression

$$\underline{\delta}^T\underline{\delta} = \left[\begin{bmatrix} \underline{DY} \\ \underline{DZ} \end{bmatrix} - \begin{bmatrix} \underline{SY} & \underline{SZ} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{SY} & \underline{SZ} \end{bmatrix} \begin{bmatrix} \vdots \\ \kappa_i \\ \vdots \end{bmatrix}\right]^T \left[\begin{bmatrix} \underline{DY} \\ \underline{DZ} \end{bmatrix} - \begin{bmatrix} \underline{SY} & \underline{SZ} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{SY} & \underline{SZ} \end{bmatrix} \begin{bmatrix} \vdots \\ \kappa_i \\ \vdots \end{bmatrix}\right]$$

which is formally equivalent to expression (5.15) for $w = 1$ time-unit-squared. Therefore, estimators derived from minimization of (5.15) will be identical to those obtained from minimization of $\underline{\delta}^T\underline{\delta}$ and will be referred to as "least-squares" estimators in Section 5.3. Letting

$$\underline{\mathcal{D}} = \begin{bmatrix} \underline{D} & \underline{0} \\ \underline{0} & \underline{D} \end{bmatrix} \quad \text{and} \quad \underline{\mathcal{S}} = \begin{bmatrix} \underline{S} & \underline{0} \\ \underline{0} & \underline{S} \end{bmatrix} ,$$

we have

$$\underline{\delta}^T\underline{\delta} = \left[\underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} - \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{Y} & \underline{Z} \end{bmatrix} \underline{\kappa} \right]^T \left[\underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} - \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{Y} & \underline{Z} \end{bmatrix} \underline{\kappa} \right]$$

$$= \left[[\underline{Y}^T \ \underline{Z}^T]\, \underline{\mathcal{D}}^T - \underline{\kappa}^T \begin{bmatrix} \underline{Y}^T & \underline{0}^T \\ \underline{Z}^T & \underline{0}^T \\ \underline{0}^T & \underline{Y}^T \\ \underline{0}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \right] \left[\underline{\mathcal{D}} \begin{matrix} \underline{Y} \\ \underline{Z} \end{matrix} - \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{Y} & \underline{Z} \end{bmatrix} \underline{\kappa} \right]$$

$$= [\underline{Y}^T \ \underline{Z}^T]\, \underline{\mathcal{D}}^T \underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} - [\underline{Y}^T \ \underline{Z}^T]\, \underline{\mathcal{D}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{Y} & \underline{Z} \end{bmatrix} \underline{\kappa}$$

$$- \underline{\kappa}^T \begin{bmatrix} \underline{Y}^T & \underline{0}^T \\ \underline{Z}^T & \underline{0}^T \\ \underline{0}^T & \underline{Y}^T \\ \underline{0}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} + \underline{\kappa}^T \begin{bmatrix} \underline{Y}^T & \underline{0}^T \\ \underline{Z}^T & \underline{0}^T \\ \underline{0}^T & \underline{Y}^T \\ \underline{0}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{Y} & \underline{Z} \end{bmatrix} \underline{\kappa}$$

Taking

$$\frac{\partial(\underline{\delta}^T\underline{\delta})}{\partial\underline{\kappa}} = -2 \begin{bmatrix} \underline{Y}^T & \underline{0}^T \\ \underline{Z}^T & \underline{0}^T \\ \underline{0}^T & \underline{Y}^T \\ \underline{0}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix}$$

$$+ 2 \begin{bmatrix} \underline{Y}^T & \underline{0}^T \\ \underline{Z}^T & \underline{0}^T \\ \underline{0}^T & \underline{Y}^T \\ \underline{0}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{Y} & \underline{Z} \end{bmatrix} \underline{\hat{\kappa}} \ ,$$

setting $\dfrac{\partial(\underline{\delta}^T\underline{\delta})}{\partial\underline{\kappa}} = 0$, and solving for $\underline{\hat{\kappa}}$, one obtains

$$\underline{\hat{K}} = \left(\begin{bmatrix} \underline{Y}^T & \underline{O}^T \\ \underline{Z}^T & \underline{O}^T \\ \underline{O}^T & \underline{Y}^T \\ \underline{O}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{O} & \underline{O} \\ \underline{O} & \underline{O} & \underline{Y} & \underline{Z} \end{bmatrix} \right)^{-1} \begin{bmatrix} \underline{Y}^T & \underline{O}^T \\ \underline{Z}^T & \underline{O}^T \\ \underline{O}^T & \underline{Y}^T \\ \underline{O}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} \tag{5.16}$$

which is similar to equation (4.5) for the simple differential equation $\dot{y} = Ky$. However,

$$\left(\begin{bmatrix} \underline{Y}^T & \underline{O}^T \\ \underline{Z}^T & \underline{O}^T \\ \underline{O}^T & \underline{Y}^T \\ \underline{O}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Y} & \underline{Z} & \underline{O} & \underline{O} \\ \underline{O} & \underline{O} & \underline{Y} & \underline{Z} \end{bmatrix} \right)^{-1}$$

$$= \begin{bmatrix} \underline{Z}^T & \underline{O}^T \\ -\underline{Y}^T & \underline{O}^T \\ \underline{O}^T & \underline{Z}^T \\ \underline{O}^T & -\underline{Y}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Z} & -\underline{Y} & \underline{O} & \underline{O} \\ \underline{O} & \underline{O} & \underline{Z} & -\underline{Y} \end{bmatrix} \cdot \frac{1}{d}$$

where $d = \underline{Y}^T \underline{\mathcal{S}}^T \underline{\mathcal{S}} \underline{Y} \underline{Z}^T \underline{\mathcal{S}}^T \underline{\mathcal{S}} \underline{Z} - \underline{Y}^T \underline{\mathcal{S}}^T \underline{\mathcal{S}} \underline{Z} \underline{Y}^T \underline{\mathcal{S}}^T \underline{\mathcal{S}} \underline{Z}$

so

$$\underline{\hat{K}} = \frac{1}{d} \begin{bmatrix} \underline{Z}^T & \underline{O}^T \\ -\underline{Y}^T & \underline{O}^T \\ \underline{O}^T & \underline{Z}^T \\ \underline{O}^T & -\underline{Y}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{S}} \begin{bmatrix} \underline{Z} & -\underline{Y} & \underline{O} & \underline{O} \\ \underline{O} & \underline{O} & \underline{Z} & -\underline{Y} \end{bmatrix} \begin{bmatrix} \underline{Y}^T & \underline{O}^T \\ \underline{Z}^T & \underline{O}^T \\ \underline{O}^T & \underline{Y}^T \\ \underline{O}^T & \underline{Z}^T \end{bmatrix} \underline{\mathcal{S}}^T \underline{\mathcal{D}} \begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} . \tag{5.17}$$

After multiplication,

$$\hat{K}_1 = \frac{1}{d} \cdot \left[\underline{Z}^T \underline{S}^T \underline{SZY}^T \underline{S}^T \underline{DY} - \underline{Z}^T \underline{S}^T \underline{SYZ}^T \underline{S}^T \underline{DY} \right]$$

$$\hat{K}_2 = \frac{1}{d} \cdot \left[\underline{Y}^T \underline{S}^T \underline{SYZ}^T \underline{S}^T \underline{DY} - \underline{Y}^T \underline{S}^T \underline{SZY}^T \underline{S}^T \underline{DY} \right]$$

$$\hat{K}_3 = \frac{1}{d} \cdot \left[\underline{Z}^T \underline{S}^T \underline{SZY}^T \underline{S}^T \underline{DZ} - \underline{Z}^T \underline{S}^T \underline{SYZ}^T \underline{S}^T \underline{DZ} \right] \tag{5.18}$$

$$\hat{K}_4 = \frac{1}{d} \cdot \left[\underline{Y}^T \underline{S}^T \underline{SYZ}^T \underline{S}^T \underline{DZ} - \underline{Y}^T \underline{S}^T \underline{SZY}^T \underline{S}^T \underline{DZ} \right]$$

where

$$d = \underline{Y}^T \underline{S}^T \underline{SYZ}^T \underline{S}^T \underline{SZ} - \underline{Y}^T \underline{S}^T \underline{SZY}^T \underline{S}^T \underline{SZ} \quad .$$

Alternatively, the approximation to the system (5.11) can be written

$$\begin{pmatrix} (\underline{Dy})^T \\ (\underline{Dz})^T \end{pmatrix} = \begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix} \begin{pmatrix} (\underline{Sy})^T \\ (\underline{Sz})^T \end{pmatrix} \tag{5.19}$$

or, in terms of the observed $\underline{Y}$ and $\underline{Z}$,

$$\begin{pmatrix} (\underline{DY})^T \\ (\underline{DZ})^T \end{pmatrix} = \begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix} \begin{pmatrix} (\underline{SY})^T \\ (\underline{SZ})^T \end{pmatrix} + \begin{pmatrix} (\underline{D\varepsilon})^T \\ (\underline{D\zeta})^T \end{pmatrix} - \begin{pmatrix} K_1 & K_2 \\ K_3 & K_4 \end{pmatrix} \begin{pmatrix} (\underline{S\varepsilon})^T \\ (\underline{S\zeta})^T \end{pmatrix} \quad .$$

For purposes of this development, a more convenient form is

$$(\underline{DY} \;\vdots\; \underline{DZ}) = (\underline{SY} \;\vdots\; \underline{SZ}) \begin{pmatrix} K_1 & K_3 \\ K_2 & K_4 \end{pmatrix} + \underline{\delta}^*$$

where

$$\underline{\delta}* = (\underline{D\varepsilon} \quad \underline{D\zeta}) - (\underline{S\varepsilon} \quad \underline{S\zeta}) \begin{pmatrix} \kappa_1 & \kappa_3 \\ \kappa_2 & \kappa_4 \end{pmatrix} .$$

In the preceding discussion, the estimates of the κ_i as elements of a vector were derived from minimization of the sum of the squares of elements of the error vector, viz., $\underline{\delta}^T\underline{\delta} = \Sigma\ \delta_i^2$. In an analogous manner, estimates of the κ_i in the matrix formulation above can be derived by considering minimization of the euclidean (or Frobenius) matrix norm $\sum_{i,j} \delta^{*\,2}_{ij}$ where

$$(\delta^*_{ij}) = \underline{\delta}* = (\underline{DY} \,\vdots\, \underline{DZ}) - (\underline{SY} \,\vdots\, \underline{SZ}) \begin{bmatrix} \kappa_1 & \kappa_3 \\ \kappa_2 & \kappa_4 \end{bmatrix} .$$

However,

$$\underline{\delta}* = [\underline{DY} - (\underline{SY}\kappa_1 + \underline{SZ}\kappa_2) \;\vdots\; \underline{DZ} - (\underline{SY}\kappa_3 + \underline{SZ}\kappa_4)]$$

$$= [\quad \underline{\delta}_1^* \quad \vdots \quad \underline{\delta}_2^* \quad]$$

so

$$\Sigma\ \delta_{ij}^{*2} = \underline{\delta}_1^* \underline{\delta}_1^* + \underline{\delta}_2^{*T} \underline{\delta}_2^* \qquad (5.20)$$

which is formally equivalent to (5.15) for $w = 1$ time-unit-squared. Therefore, the estimators $\hat{\kappa}_i$ derived from minimization of (5.20) are identical to those derived from minimization of the Sobolev norm-type expression (5.15).

In summary, the estimators $\hat{\kappa}_1$, $\hat{\kappa}_2$, $\hat{\kappa}_3$, and $\hat{\kappa}_4$ in the system (5.11) are identical for each of the following formulations in the case of initial 5-point moving-arc polynomial smoothing and derivative estimation:

(1) Consideration of each equation of the system separately as in Section 5.1.

(2) Minimization of the Sobolev norm-type expression (5.13).

(3) Consideration of the system in vector formulation (5.12).

(4) Consideration of the system in matrix formulation (5.19).

For the case involving simultaneous smoothing, derivative estimation and concomitant estimation of the κ_i , the values of the functions f_1 , f_2 , g_1 , and g_2 at the t_i can be represented by the vectors $\underline{\Phi\beta}$, $\underline{\Phi\gamma}$, $\underline{\Phi}'\underline{\beta}$, and $\underline{\Phi}'\underline{\gamma}$, respectively, where the matrix $\underline{\Phi}$, whose elements are functions of the t_i , and $\underline{\beta}$ and $\underline{\gamma}$, which are vectors whose elements are parameters characterizing the smoothing function, are defined as in Sections 3.2.2 and 3.3. For 5-point moving-arc cubic smoothing, the $\underline{\beta}$ and $\underline{\gamma}$ are the coefficients of cubic polynomials; for cubic-segment smoothing, the $\underline{\beta}$ and $\underline{\gamma}$ are the values of the function and the first derivative at the chosen knots. Equation (5.13) is, then, of the form

$$\nu^{*2} = (\underline{Y} - \underline{\Phi\beta})^T (\underline{Y} - \underline{\Phi\beta}) + w(\dot{\underline{Y}} - \underline{\Phi}'\underline{\beta})^T (\dot{\underline{Y}} - \underline{\Phi}'\underline{\beta})$$

$$+ (\underline{Z} - \underline{\Phi\gamma})^T (\underline{Z} - \underline{\Phi\gamma}) + w(\dot{\underline{Z}} - \underline{\Phi}'\underline{\gamma})^T (\dot{\underline{Z}} - \underline{\Phi}'\underline{\gamma}) \quad .$$

Taking $\frac{\partial(v^2)}{\partial \underline{\beta}}$, $\frac{\partial(v^2)}{\partial \underline{\gamma}}$, and $\frac{\partial(v^2)}{\partial \kappa_i}$, letting w = 1 time-unit-squared, and setting the partial derivatives equal to zero yields normal equations of the form

$$\left(\begin{array}{cc|cccc} \underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}' & \underline{0} & -\underline{\Phi}'^T\underline{Y} & -\underline{\Phi}'^T\underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}' & \underline{0} & \underline{0} & -\underline{\Phi}'^T\underline{Y} & -\underline{\Phi}'^T\underline{Z} \\ \hline -\underline{Y}^T\underline{\Phi}' & \underline{0}^T & \underline{Y}^T\underline{Y} & \underline{Y}^T\underline{Z} & 0 & 0 \\ -\underline{Z}^T\underline{\Phi}' & \underline{0}^T & \underline{Y}^T\underline{Z} & \underline{Z}^T\underline{Z} & 0 & 0 \\ \underline{0}^T & -\underline{Y}^T\underline{\Phi}' & 0 & 0 & \underline{Y}^T\underline{Y} & \underline{Y}^T\underline{Z} \\ \underline{0}^T & -\underline{Z}^T\underline{\Phi}' & 0 & 0 & \underline{Y}^T\underline{Z} & \underline{Z}^T\underline{Z} \end{array}\right) \begin{pmatrix} \hat{\underline{\beta}} \\ \hat{\underline{\gamma}} \\ \hat{\kappa}_1 \\ \hat{\kappa}_2 \\ \hat{\kappa}_3 \\ \hat{\kappa}_4 \end{pmatrix} = \begin{pmatrix} \underline{\Phi}^T\underline{Y} \\ \underline{\Phi}^T\underline{Z} \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

Clearly, from the blocked symmetry of the submatrices of the partitioned matrix, the estimates of the κ_i from this formulation are identical to those derived from consideration of the two differential equations separately, as in Section 5.1.2.

5.3 Generalization to a System of Differential Equations in Which Some of the Coefficients Are Related or Assume Known Values

As a particular case of the general system of differential equations (5.11), consider the system

$$\dot{y} = -\kappa_1 y + \kappa_2 z \ , \ y(0) = 0$$

$$\dot{z} = \quad - \kappa_2 z \ , \ z(0) = z_0$$

in the general form

$$\begin{pmatrix} \dot{y} \\ \\ \dot{z} \end{pmatrix} = \begin{pmatrix} y & z & 0 & 0 \\ \\ 0 & 0 & y & z \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} \qquad (5.21)$$

$$\text{where} \quad \beta_1 = -\kappa_1 \ , \quad \beta_2 = \kappa_2 \ , \quad \beta_3 = 0 \ , \quad \beta_4 = -\kappa_2 \ . \qquad (5.22)$$

For n observed values $\{[Y(t_i), Z(t_i)]: \ Y(t_i) = y(t_i) + \varepsilon(t_i),$ $Z(t_i) = z(t_i) + \zeta(t_i), \ i = 1, 2, \ldots, n\}$ and for initial 5-point moving-arc polynomial smoothing and derivative approximation represented by matrices $\underline{S}$ and $\underline{D}$, equation (5.21) can be written in the approximation form

$$\begin{pmatrix} \underline{D}\,\underline{Y} \\ \\ \underline{D}\,\underline{Z} \end{pmatrix} = \begin{pmatrix} \underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} & \underline{0} & \underline{0} \\ \\ \underline{0} & \underline{0} & \underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{pmatrix} + \underline{\delta} \qquad (5.23)$$

where

$$\underline{\delta} = \begin{pmatrix} \underline{D}\,\underline{\varepsilon} \\ \\ \underline{D}\,\underline{\zeta} \end{pmatrix} - \begin{pmatrix} \underline{S}\,\underline{\varepsilon} & \underline{S\zeta} & \underline{0} & \underline{0} \\ \\ \underline{0} & \underline{0} & \underline{S}\,\underline{\varepsilon} & \underline{S\zeta} \end{pmatrix} \ .$$

In this section, it will be shown that estimates of $\kappa_1 = -\beta_1$ and $\kappa_2 = \beta_2$ by the least-squares method as defined in Section 5.2 after adjustment for the linear restrictions (5.22) are identical to estimates of κ_1 and κ_2 from application of the same least-squares methods to

$$\begin{bmatrix} \underline{D}\,\underline{Y} \\ \underline{D}\,\underline{Z} \end{bmatrix} = \begin{bmatrix} -\underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} \\ 0 & -\underline{S}\,\underline{Z} \end{bmatrix} \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} + \underline{\delta}^* \qquad (5.24)$$

where

$$\underline{\delta}^* = \begin{bmatrix} \underline{D}\,\underline{\varepsilon} \\ \underline{D}\,\underline{\zeta} \end{bmatrix} - \begin{bmatrix} -\underline{S}\,\underline{\varepsilon} & \underline{S}\,\underline{\zeta} \\ \underline{0} & -\underline{S}\,\underline{\zeta} \end{bmatrix} .$$

Goldberger (1964) proves that, if $\underline{b}$ is the unrestricted least-squares estimator of $\underline{\beta}$ in the general linear model, of which our equation (5.23) is a special case, then $\underline{b}^*$, the restricted least-squares estimator of $\underline{\beta}$, is

$$\underline{b}^* = \underline{b} + (\underline{X}^T\,\underline{X})^{-1}\,\underline{R}^T\,[\underline{R}(\underline{X}^T\,\underline{X})^{-1}\,\underline{R}^T]^{-1}\,(\underline{r} - \underline{R}\,\underline{b}) \qquad (5.25)$$

where, in our case,

$$\underline{X} = \begin{bmatrix} \underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{S}\,\underline{Y} & \underline{S}\,\underline{Z} \end{bmatrix}$$

and where $\underline{r} = \underline{R}\,\underline{b}$ defines the exact restriction on the κ_i. In the case under consideration, the restrictions (equations (5.22)) $\underline{r} = \underline{R}\,\underline{\beta}$ have the form

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix}$$

It is shown in the appendix that the second term in equation (5.25) is a 4x1 vector, the first two elements of which are

$$\Big((\underline{Y}^T\underline{CZ})^2\underline{Z}^T\underline{CZY}^T\underline{BZ} - (\underline{Y}^T\underline{CZ})^2\underline{Y}^T\underline{CZZ}^T\underline{BZ} + \underline{Y}^T\underline{CZZ}^T\underline{CZY}^T\underline{CYZ}^T\underline{BY}$$

$$- \underline{Y}^T\underline{CZZ}^T\underline{CZZ}^T\underline{CYY}^T\underline{BY} + \underline{Y}^T\underline{CZZ}^T\underline{CZY}^T\underline{CYZ}^T\underline{BZ} - \underline{Y}^T\underline{CZZ}^T\underline{CZZ}^T\underline{CYY}^T\underline{BZ}\Big)\cdot\frac{1}{d_1 d_2} \tag{5.26}$$

and

$$\Big(-\underline{Y}^T\underline{CYZ}^T\underline{CYZ}^T\underline{CZY}^T\underline{BZ} + \underline{Y}^T\underline{CYZ}^T\underline{CYY}^T\underline{CZZ}^T\underline{BZ} - \underline{Y}^T\underline{CYZ}^T\underline{CZY}^T\underline{CYZ}^T\underline{BY}$$

$$+ \underline{Y}^T\underline{CYZ}^T\underline{CZZ}^T\underline{CYY}^T\underline{BY} - \underline{Y}^T\underline{CYZ}^T\underline{CZY}^T\underline{CYZ}^T\underline{BZ} + \underline{Y}^T\underline{CYZ}^T\underline{CZZ}^T\underline{CYY}^T\underline{BZ}\Big)\cdot\frac{1}{d_1 d_2}$$

where $\underline{C} = \underline{S}^T\underline{S}$ and $\underline{B} = \underline{S}^T\underline{D}$ and where

$$d_1 = 2\underline{Y}^T\underline{CYZ}^T\underline{CZ} - \underline{Y}^T\underline{CZZ}^T\underline{CY}$$

$$d_2 = \underline{Y}^T\underline{CYZ}^T\underline{CZ} - \underline{Y}^T\underline{CZY}^T\underline{CZ} \quad .$$

From equation (5.18), the unrestricted estimators b_1 and b_2 are

$$b_1 = [-\underline{Z}^T\underline{CZY}^T\underline{BY} + \underline{Z}^T\underline{CYZ}^T\underline{BY}]\cdot\frac{1}{d_2}$$

$$b_2 = [-\underline{Y}^T\underline{CYZ}^T\underline{BY} + \underline{Y}^T\underline{CZY}^T\underline{BY}]\cdot\frac{1}{d_2} \tag{5.27}$$

The sum of equations (5.26) and (5.27) yield the restricted estimators

$$b_1^* = (-2\underline{Z}^T\underline{CZY}^T\underline{BY} + \underline{Y}^T\underline{CZZ}^T\underline{BY} - \underline{Y}^T\underline{CZZ}^T\underline{BZ}) \cdot \frac{1}{d_1}$$

$$b_2^* = (-\underline{Z}^T\underline{CYY}^T\underline{BY} + \underline{Y}^T\underline{CYZ}^T\underline{BY} - \underline{Y}^T\underline{CYZ}^T\underline{BZ}) \cdot \frac{1}{d_1} \quad . \tag{5.28}$$

On the other hand, the least-squares estimators $\hat{\kappa}_1$ and $\hat{\kappa}_2$ in equation (5.24) can be computed by minimizing $\underline{\delta}^{*T}\,\underline{\delta}^*$. It is easily verified that, writing $\underline{C} = \underline{S}^T\underline{S}$ and $\underline{B} = \underline{S}^T\underline{D}$,

$$\begin{matrix} \hat{\kappa}_1 \\ \\ \hat{\kappa}_2 \end{matrix} = \begin{pmatrix} -2\underline{Z}^T\underline{CZY}^T\underline{BY} + \underline{Z}^T\underline{CYZ}^T\underline{BY} - \underline{Z}^T\underline{CYZ}^T\underline{BZ} \\ \\ -\underline{Y}^T\underline{CZY}^T\underline{BY} + \underline{Y}^T\underline{CYZ}^T\underline{BY} - \underline{Y}^T\underline{CYZ}^T\underline{BZ} \end{pmatrix} \cdot \frac{1}{d} \tag{5.29}$$

where $d = 2\underline{Z}^T\underline{CZY}^T\underline{CY} - (\underline{Y}^T\underline{CZ})^2$. The estimators of equation (5.29) are identical to those derived by restricted least squares, equation (5.28).

The estimation of the κ_1 and κ_2 in this case by minimization of the Sobolev norm-type expression (5.13), which for simultaneous smoothing, derivative estimation, and concomitant estimation of the κ_i , is

$$\begin{aligned} \nu^{*2} = {} & (\underline{Y} - \underline{\Phi\beta})^T (\underline{Y} - \underline{\Phi\beta}) + w(-\kappa_1 \underline{Y} + \kappa_2 \underline{Z} - \underline{\Phi}'\underline{\beta})^T \\ & (-\kappa_1 \underline{Y} + \kappa_2 \underline{Z} - \underline{\Phi}'\underline{\beta}) + (\underline{Z} - \underline{\Phi}\,\underline{\gamma})^T (\underline{Z} - \underline{\Phi}\,\underline{\gamma}) \\ & + w(-\kappa_2 \underline{Z} - \underline{\Phi}'\underline{\gamma})^T (-\kappa_2 \underline{Z} - \underline{\Phi}'\underline{\gamma}) \end{aligned}$$

involves normal equations of the form (for $w = 1$ time-unit-squared)

$$\begin{bmatrix} \underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}' & \underline{0} & \underline{\Phi}'^T\underline{Y} & -\underline{\Phi}'^T\underline{Z} \\ \underline{0} & \underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}' & \underline{0} & \underline{\Phi}'^T\underline{Z} \\ \hline \underline{Y}^T\underline{\Phi}' & \underline{0}^T & \underline{Y}^T\underline{Y} & -\underline{Y}^T\underline{Z} \\ -\underline{Z}^T\underline{\Phi}' & \underline{Z}^T\underline{\Phi}' & -\underline{Y}^T\underline{Z} & 2\underline{Z}^T\underline{Z} \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}} \\ \hat{\underline{\gamma}} \\ \hline \hat{\kappa}_1 \\ \hat{\kappa}_2 \end{bmatrix} = \begin{bmatrix} \underline{\Phi}^T\underline{Y} \\ \underline{\Phi}^T\underline{Z} \\ \hline 0 \\ 0 \end{bmatrix}$$

The solution of these normal equations yields

$$\begin{matrix} \hat{\kappa}_1 \\ \hat{\kappa}_2 \end{matrix} = - \begin{pmatrix} 2\underline{Z}^T\underline{EZ} & \underline{Y}^T\underline{EZ} \\ \underline{Y}^T\underline{EZ} & \underline{Y}^T\underline{EY} \end{pmatrix} \begin{pmatrix} \underline{Y}^T\underline{FY} \\ -\underline{Z}^T\underline{FY} + \underline{Z}^T\underline{FZ} \end{pmatrix} \cdot \frac{1}{d}$$

$$= \begin{pmatrix} -2\underline{Z}^T\underline{EZY}^T\underline{FY} + \underline{Y}^T\underline{EZZ}^T\underline{FY} - \underline{Y}^T\underline{EZZ}^T\underline{FZ} \\ -\underline{Y}^T\underline{EZY}^T\underline{FY} + \underline{Y}^T\underline{EYZ}^T\underline{FY} - \underline{Y}^T\underline{EYZ}^T\underline{FZ} \end{pmatrix} \cdot \frac{1}{d}$$

where

$$d = 2\underline{Y}^T\underline{EYZ}^T\underline{EZ} - (\underline{Y}^T\underline{EZ})^2$$

$$\underline{E} = \underline{I} - \underline{\Phi}'(\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}')^{-1}\ \underline{\Phi}'^T$$

$$\underline{F} = \underline{\Phi}'\ (\underline{\Phi}^T\underline{\Phi} + \underline{\Phi}'^T\underline{\Phi}')^{-1}\ \underline{\Phi}^T \quad .$$

Except for the definition of the matrices $\underline{E}$ and $\underline{F}$, these estimators are identical in form to those derived from initial smoothing in the preceding discussion (equation (5.2[illegible])).

6. BIAS REDUCTION

6.1 General Considerations

From the discussion in Chapter 4 on the estimation of κ in the simple differential equation $\dot{y}(t) = \kappa y(t)$, the presence of non-negligible $bias_{total}$ in the estimates is obvious. Attempts to reduce $bias_{total}$ in this case naturally involve attempts to decrease $|bias_{method}|$ and/or $|bias_{error}|$. Although these two components of $bias_{total}$ have been treated separately computationally in previous chapters, the fact that they are related is obvious from inspection of the simulation sampling results of Sections 4.1.2 and 4.2.2. It is not surprising, then that reductions in $bias_{method}$ are accompanied by changes in $bias_{error}$.

In general, three approaches to reduction of $bias_{total}$ will be considered in this chapter. First, a function f (as in the Sobolev norm-type expression (3.1)) can be chosen which better approximates the underlying data-generating function and its derivatives at the points of observation. For instance, f might be chosen to be the simultaneous 5-point moving-arc cubic rather than the initial 5-point cubic smoother. This approach also includes such procedures as eliminating the points of known or suspected poor fit or poor derivative approximation, such as eliminating the first two and/or the last two points in the case of initial 5-point moving-arc polynomial smoothing and derivative approximation. For example, the omission of the smoothing (or approximation in the case of error-free values $y(t_i)$) and the derivative approximation of the first and second points in the case of initial 5-point moving-arc cubic smoothing for $\kappa = -0.5$ (see the fifth row of Table 4.5) yields the following estimates of κ :

points omitted	$\hat{\kappa}_{exact}$	$bias_{method}$
none	-0.4912	-0.0088
$y(t_1)$	-0.5057	0.0057
$y(t_1)$, $y(t_2)$	-0.5017	0.0017

The second approach involves the utilization of auxiliary information. For example, such information can appear in the form of knowledge of a second differential equation which contains coefficients in common with the first. An example will appear later in this chapter.

The third approach depends on a reduction in the error associated with the observations. It is to be noted, however, that, in the situations where $bias_{method}$ and $bias_{error}$ are of unlike signs (as in initial 5-point moving-arc cubic smoothing for proportional error variances, results of which appear in Tables 4.4 and 4.5), a minimum $bias_{total}$ can occur for nonzero p .

Remark 1. Although the estimator of κ resulting from minimizing the expression (4.4)

$$(\kappa \underline{S}\underline{Y} - \underline{D}\underline{Y})^T (\kappa \underline{S}\underline{Y} - \underline{D}\underline{Y}) ,$$

where $\underline{S}$ and $\underline{D}$ are the smoothing and derivative estimation matrices associated with initial 5-point moving-arc cubic smoothing, would intuitively yield less bias than the estimator derived from minimizing

$$(\kappa \underline{Y} - \underline{D}\underline{Y})^T (\kappa \underline{Y} - \underline{D}\underline{Y}) ,$$

the approximation and simulation investigations of Section 4.1.2 suggest that this conjecture is not generally true.

Remark 2. In the above example of decreasing $\text{bias}_{\text{method}}$ by the omission of the smoothing and derivative estimation at the first two points, the assumed underlying function is $y(t) = 1000\ e^{-0.5t}$. Although $y(1) = 606.53$ and $y(2) = 367.88$ are used to smooth the third point, the omission of these values from the formula for the estimation of κ results in disregarding part of the information in the sample. In this particular case, the disregarded information corresponds to an important part of the graph of the underlying function, namely, that part where the values are greatest, where the function is changing most rapidly, and where the probable errors are greatest.

6.2 Examples of Reducing $\text{Bias}_{\text{total}}$

In this section, the estimation of κ_1 and κ_2 is considered for the model differential equation

$$\dot{y}(t) = -\kappa_1 y(t) + \kappa_2 z(t) \tag{6.1}$$

as a single equation from the underlying system of differential equations

$$\begin{aligned} \dot{y}(t) &= -\kappa_1 y(t) + \kappa_2 z(t) \\ \dot{z}(t) &= \qquad\qquad\ -\kappa_2 z(t)\ . \end{aligned} \tag{6.2}$$

For purposes of simulation, first, values of $y(t_i)$ and $z(t_i)$ are computed for $t_i = 1, 2, \ldots, 15$ from the solution of the system (6.2)

$$y(t) = z_0 \left[\frac{K_2}{K_2 - K_1}\right] \left[e^{-K_1 t} - e^{-K_2 t}\right]$$

$$z(t) = z_0 \, e^{-K_2 t}$$

where $z_0 = 1000$, $K_2 = 0.21$, and $K_1 = 0.20$. Using the procedures of Section 5.1, estimates of K_1 and K_2 are obtained for the following smoothing and derivative approximation schemes:

(i) Initial 5-point moving-arc cubic smoothing and derivative approximation followed by estimation of K_1 and K_2 .

(ii) Initial 5-point moving-arc linear-hyperbolic smoothing and derivative approximation followed by estimation of K_1 and K_2 .

(iii) Simultaneous smoothing and derivative approximation using three cubic segments with knots at 0.5 , 5.5 , 10.5 , and 15.5 time-units and concomitant estimation of K_1 and K_2 .

(iv) Simultaneous smoothing and derivative approximation using 5-point moving-arc cubics and concomitant estimation of the K_1 and K_2 .

To investigate the properties of

$$\underline{\hat{K}}^T = [\hat{K}_1 \quad \hat{K}_2]$$

in the presence of proportional data error, Monte Carlo simulation was conducted in which the artificial observations,

$$Y(t_i) = y(t_i) + \varepsilon(t_i)$$

and

$$Z(t_i) = z(t_i) + \zeta(t_i)$$

were constructed where

$$\varepsilon(t_i) \sim n\left(0, \left[\frac{p}{100}\, y(t_i)\right]^2\right)$$

and

$$\zeta(t_i) \sim n\left(0, \left[\frac{p}{100}\, z(t_i)\right]^2\right) \quad .$$

At the risk of the possible inclusion of peculiar (i.e., unlikely) samples, the same sets of random deviates were used for both $p = 5$ and $p = 10$ and for all four estimation procedures. It is hoped that such duplicity assists in elucidating the effects of increased error and modifications in the estimators. Table 6.1 displays the results of estimation from error-free values, $y(t_i)$ and $z(t_i)$, and from the sampling investigation.

From the $\hat{\kappa}_{exact}$ values (Table 6.1), it is apparent that the estimator associated with the initial 5-point moving-arc linear-hyperbolic function is much poorer than those associated with the other smoothing and derivative estimation functions. Comparison of the original values $y(t_i)$ and $z(t_i)$ and their approximations and of the true derivative values and their approximations reveals that approximations of $y(t_i)$ and $z(t_i)$ are relatively poor for $t_i = 1, 2,$ and 3. The results of estimation of κ_1 and κ_2 using smoothed values and derivative approximations for $t_i = 4, 5, \ldots, 15$ are shown in Table 6.2. These results indicate that the omission of

Table 6.1 Simulation results for estimation of κ_1 and κ_2 in $\dot{y} = -\kappa_1 y + \kappa_2 z$ treated as a single equation from system 6.2 .

Smoothing methods are initial 5-point moving-arc cubic = I-C, initial 5-point moving-arc linear-hyperbolic = I-L-H , simultaneous 5-point moving arc cubic = S-C, and simultaneous three cubic segments = S-3-C. $\kappa_1 = 0.20$, $\kappa_2 = 0.21$. n = 25.

method	p	κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
I-L-H	5	κ_1	0.2612	0.2598	-0.0612	0.0014	-0.0598	0.000751	0.0274	0.0055	0.004327	-2.1825	0.8265
		κ_2	0.2773	0.2752	-0.0673	0.0021	-0.0652	0.001087	0.0330	0.0066	0.005338	-1.9758	0.7964
	10	κ_1	0.2612	0.2585	-0.0612	0.0027	-0.0587	0.002862	0.0535	0.0107	0.006307	-1.0972	0.5463
		κ_2	0.2773	0.2737	-0.0673	0.0036	-0.0637	0.004180	0.0647	0.0129	0.008238	-0.9845	0.4926
I-C	5	κ_1	0.1986	0.1893	0.0014	0.0093	0.0107	0.000814	0.0285	0.0057	0.000928	0.3754	0.1233
		κ_2	0.2085	0.1979	0.0015	0.0106	0.0121	0.001104	0.0332	0.0066	0.001249	0.3645	0.1172
	10	κ_1	0.1986	0.1795	0.0014	0.0191	0.0205	0.003217	0.0567	0.0113	0.003637	0.3616	0.1155
		κ_2	0.2085	0.1874	0.0015	0.0311	0.0226	0.004362	0.0660	0.0132	0.004873	0.3424	0.1048
S-3-C	5	κ_1	0.1988	0.1944	0.0002	0.0044	0.0046	0.000084	0.0092	0.0018	0.000106	0.5000	0.2005
		κ_2	0.2085	0.2038	0.0015	0.0047	0.0062	0.000136	0.0117	0.0023	0.000175	0.5299	0.2200
	10	κ_1	0.1988	0.1854	0.0002	0.0134	0.0136	0.000310	0.0176	0.0035	0.000495	0.7727	0.3734
		κ_2	0.2085	0.1950	0.0015	0.0135	0.0150	0.000494	0.0222	0.0044	0.000719	0.6757	0.3129
S-C	5	κ_1	0.1991	0.1941	0.0009	0.0050	0.0059	0.000099	0.0100	0.0020	0.000134	0.5900	0.2593
		κ_2	0.2089	0.2036	0.0011	0.0053	0.0064	0.000178	0.0133	0.0027	0.000219	0.4812	0.1870
	10	κ_1	0.1991	0.1831	0.0009	0.0160	0.0169	0.000381	0.0195	0.0039	0.000667	0.8667	0.4283
		κ_2	0.2089	0.1921	0.0011	0.0168	0.0179	0.000680	0.0261	0.0052	0.001000	0.6858	0.3204

Table 6.2 Simulation results of estimating κ_1 and κ_2 in $\dot{y} = \kappa_1 y + \kappa_2 z$ following initial 5-point linear-hyperbolic smoothing and derivative estimation for restricted sample. $t_i = 4, 5, \ldots, 15$. $\kappa_1 = 0.20$, $\kappa_2 = 0.21$. $n = 25$.

p	κ	$\hat{\kappa}_{exact}$	$\hat{\kappa}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
5	κ_1	0.1896	0.1936	0.0104	-0.0040	0.0064	0.000273	0.0165	0.0033	0.000314	0.3879	0.1303
	κ_2	0.1921	0.1987	0.0179	-0.0066	0.0113	0.000645	0.0253	0.0051	0.000772	0.4466	0.1654
10	κ_1	0.1896	0.1965	0.0104	-0.0069	0.0035	0.001086	0.0330	0.0066	0.001099	0.1061	0.0111
	κ_2	0.1921	0.2046	0.0179	-0.0125	0.0054	0.002625	0.0512	0.0102	0.002654	0.1055	0.0110

smoothing and derivative approximation at those points of poor fit does result in reduced $\text{bias}_{\text{method}}$ and reduced $\text{var}(\hat{K})$ but in increased $\text{bias}_{\text{error}}$ of opposite sign. These results also show that, since the $\text{bias}_{\text{method}}$ and $\text{bias}_{\text{error}}$ are of unlike signs, the $\text{bias}_{\text{total}}$ is reduced as the error rate, p , increases. As expected, the $\text{var}(\hat{K})$ increases by approximately a factor of four when the error variance is doubled. This increase is also reflected in the increase in MSE, although the marked decrease in the values of ξ^* indicate that the proportion of MSE which is due to $\text{bias}_{\text{total}}$ decreases.

Although in a real experimental situation it might be inappropriate to consider the incomplete model consisting of one equation (6.1) when, in fact, the complete model is known to be the system (6.2), the example presently under consideration (equations (6.1) and (6.2)) is appropriate for demonstrating use of the method of $\text{bias}_{\text{total}}$ reduction incorporating a second differential equation with coefficients in common with the original model equation. For purposes of this example, estimation of K_1 and K_2 in the system (6.2) is conducted utilizing the first procedure discussed in Section 5.3, which involves the use of initial 5-point moving-arc cubic smoothing and initial 5-point moving-arc linear-hyperbolic smoothing. For each of the estimation methods, and for each $p = 5, 10$, the same sets of random deviates are used to construct the errors $\varepsilon(t_i)$ and $\zeta(t_i)$. The sampling results for both methods are shown in Table 6.3. These results indicate that, in the case of both smoothing functions, the $\text{bias}_{\text{method}}$ is reduced in comparison with the single equation estimation (see Table 6.1), although the reduction is most drastic

Table 6.3 Simulation results for estimation of κ_1 and κ_2 in system 6.2 following initial smoothing and derivative estimation. Smoothing methods are 5-point moving-arc linear-hyperbolic = I-L-H and 5-point moving-arc cubic = I-C. $\kappa_1 = 0.20$, $\kappa_2 = 0.21$. $n = 25$.

method	p	κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
I-L-H	5	κ_1	0.2097	0.2356	-0.0097	-0.0259	-0.0356	0.000731	0.0270	0.0054	0.001998	-1.3185	0.6343
		κ_2	0.2184	0.2504	-0.0084	-0.0320	-0.0404	0.000904	0.0310	0.0060	0.002537	-1.3422	0.6434
	10	κ_1	0.2097	0.2431	-0.0097	-0.0334	-0.0431	0.002910	0.0539	0.0108	0.004768	-0.7996	0.3896
		κ_2	0.2184	0.2579	-0.0084	-0.0395	-0.0479	0.003695	0.0608	0.0122	0.005989	-0.7878	0.3831
I-C	5	κ_1	0.1995	0.2036	0.0005	-0.0041	-0.0036	0.000502	0.0224	0.0045	0.000515	-0.1607	0.0252
		κ_2	0.2094	0.2125	0.0006	-0.0031	-0.0025	0.000605	0.0246	0.0049	0.000612	-0.1016	0.0102
	10	κ_1	0.1995	0.2078	0.0005	-0.0083	-0.0078	0.002027	0.0450	0.0090	0.002088	-0.1733	0.0291
		κ_2	0.2094	0.2160	0.0006	-0.0066	-0.0060	0.002484	0.0498	0.0100	0.002520	-0.1205	0.0143

in the linear-hyperbolic case. It is apparent, however, that the $bias_{error}$ increases in the linear-hyperbolic case while it decreases for the cubic smoothing function.

Estimation of K_1 and K_2 in the system 6.2 by simultaneous smoothing and concomitant estimation of the K_i as discussed in Section 5.3 results in estimates with insignificant $bias_{total}$ and greatly reduced variance compared to single equation estimation. The results of Monte Carlo simulation using the same sets of random deviates as for the case involving initial smoothing (Table 6.3) appear in Table 6.4.

Table 6.4 Simulation results for simultaneous smoothing and estimation of κ_1 and κ_2 in system 6.2 Smoothing methods are 5-point moving-arc cubics = S-C and three cubic segments = S-3-C . $\kappa_1 = 0.20$, $\kappa_2 = 0.21$. n = 25.

method	p	κ	$\hat{\kappa}_{exact}$	$\bar{\hat{\kappa}}$	$bias_m$	$bias_e$	$bias_t$	$var(\hat{\kappa})$	$\sqrt{var(\hat{\kappa})}$	$\sqrt{var(\bar{\hat{\kappa}})}$	MSE	ξ	ξ^*
S-3-C	5	κ_1	0.1997	0.2016	0.0003	-0.0019	-0.0016	0.000059	0.0077	0.0015	0.000062	-0.2078	0.0413
		κ_2	0.2095	0.2117	0.0005	-0.0022	-0.0017	0.000080	0.0089	0.0018	0.000083	-0.1910	0.0350
	10	κ_1	0.1997	0.2011	0.0003	-0.0014	-0.0011	0.000221	0.0149	0.0030	0.000222	-0.0738	0.0054
		κ_2	0.2095	0.2123	0.0005	-0.0028	-0.0023	0.000304	0.0174	0.0035	0.000309	-0.1322	0.0171
S-C	5	κ_1	0.1998	0.2000	0.0002	-0.0002	0.0000	0.000030	0.0055	0.0011	0.000030	0.0000	0.0000
		κ_2	0.2097	0.2107	0.0003	-0.0010	-0.0007	0.000047	0.0069	0.0014	0.000048	-0.1014	0.0102
	10	κ_1	0.1998	0.1979	0.0002	0.0019	0.0021	0.000124	0.0112	0.0022	0.000129	0.1875	0.0342
		κ_2	0.2097	0.2100	0.0003	-0.0003	0.0000	0.000191	0.0138	0.0028	0.000191	0.0000	0.0000

7. SUMMARY AND OVERVIEW OF OPEN PROBLEMS

7.1 Summary

Consider a differential equation of the form

$$\frac{dy}{dt} = \kappa_1 y + \kappa_2 z \quad .$$

Given observed values $Y(t_i)$ and $Z(t_i)$ of $y(t_i)$ and $z(t_i)$ at discrete times t_1, t_2, ..., t_n, one is tempted to smooth the data by fitting an approximating polynomial in t to the

$$\underline{Y}^T = [Y(t_1), \ldots, Y(t_n)]$$

$$\underline{Z}^T = [Z(t_1), \ldots, Z(t_n)]$$

by a linear regression scheme, to estimate the derivatives

$$\left. \frac{dy}{dt} \right|_{t=t_i}$$

by computing the derivative of the approximating polynomial at the t_i, and to proceed to estimate the κ_i by the usual regression procedure applied to the model

$$\dot{\underline{Y}} = [\underline{Y}^* \,\vdots\, \underline{Z}^*] \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} + \underline{\delta}$$

by minimizing $\underline{\delta}^T\underline{\delta}$, where $\underline{Y}^*$ and $\underline{Z}^*$ are the smoothed $\underline{Y}$ and $\underline{Z}$. This method often gives unsatisfactory estimates of the κ_i due to the possibly poor approximation of the $y(t_i)$ from the observed values and the subsequent relatively poorer estimates of the derivatives.

In this thesis, a method is proposed which is based on the minimization of a quantity, the expression for which is related to the discrete version of the Sobolev norm, namely

$$\nu^2 = (\underline{Y} - \underline{f})^T (\underline{Y} - \underline{f}) + w(\dot{\underline{Y}} - \underline{g})^T (\dot{\underline{Y}} - \underline{g})$$

where f and g are vectors, the elements of which represent the values of the approximating and derivative estimating functions at the t_i and where

$$\dot{\underline{Y}} = [\underline{Y} \ \ \underline{Z}] \begin{bmatrix} \kappa_1 \\ \kappa_2 \end{bmatrix} .$$

Obviously, this procedure incorporates simultaneously the approximation of the $y(t_i)$ (or smoothing of the $Y(t_i)$ in the case of data with error), the estimation of the derivatives

$$\left.\frac{dy}{dt}\right|_{t=t_i}$$

and the estimation of the κ_i . By suitable definition of $\underline{f}$ and $\underline{g}$, the naive regression-type approach is a special case of this more general method. In each case discussed in this thesis, we have $g = f'$ and f is a function which is linear in its parameters. Due to the linearity of the parameters of f , and so, of g , and of the coefficients κ_i , the problem reduces to solving a linear system of equations.

In this thesis, three forms of differential equations are considered in detail:

(1) $\frac{dy}{dt} = \kappa y$

(2) $\frac{dy}{dt} = \kappa_1 y + \kappa_2 z$

(3) $\frac{dy}{dt} = \kappa_1 y + \kappa_2 z$

$$\frac{dz}{dt} = \kappa_3 y + \kappa_4 z \quad .$$

The general theory of estimating the coefficients is developed for each and examples are included for certain representative values of the κ_i . The approximation function f is either a 5-point moving-arc cubic polynomial, a 5-point moving-arc linear-hyperbolic function, or a recently-developed function composed of joined cubic segments.

The numerical results, which include Monte Carlo simulation of each of the above differential equations with additive normally distributed error, the standard deviation of which is either constant or proportional to the $y(t_i)$, indicate that the minimization of ν^2 yields substantially improved estimates of the κ_i in terms of less bias and less variance than the minimization of the above $\underline{\delta}^T\underline{\delta}$.

The estimation of the coefficients of a differential equation by minimization of the Sobolev norm-type expression has several advantages over other current methods. It is not necessary to know either the analytic solution, if it exists, or the values of a numerical solution. Initial estimates of the parameters and coefficients are not required, as they are in most iterative schemes.

Further, use of this method avoids, for the above described type, the frequent convergence problems associated with iterative schemes. The procedure does not require a specific time-interval-sequence of observations; in particular, it does not require equally-spaced observations as do some transform methods.

On the other hand, disadvantages include the need for observations on each variable in the differential equation at identical times. Although the procedure is computationally uncomplicated, the choice of smoothing and derivative estimating functions is a potential source of poor estimation.

7.2 Overview of Open Problems

This work suggests several general areas in which further research is possible, including the following:

(1) Investigation of other choices of smoothing and derivative estimation functions.

(2) Extension of the basic method to different forms of the model differential equation,

(3) Investigation of minimizing a weighted Sobolev norm-type expression.

Each of the above areas will be considered in more detail, although their inter-relationships prevent separate and exclusive treatments.

In this thesis, we have considered two of the most common types of smoothing and derivative estimation functions, (1) the moving-arc polynomial function and (2) the continuous function defined over the entire time-span of the data, an example of which is the function composed of joined cubic segments. Obviously, neither type of function completely smoothed the observations and neither type provided exact estimates of the derivatives. This was hardly unexpected, since polynomial approximations to exponential functions are not exact. However, the use of approximating polynomials does involve the assumption that the polynomial is a "suitable" approximating function to the underlying data-generating function at least over the span of fit. Intuitively, the term "suitable" is associated with such equally-undefined notions as "not distorting the data", which, for example, can be interpreted as yielding values "very close" to those of the underlying function. Further, when derivative estimation is accomplished by means of computing the first time-derivative of the approximating function, the scope of meaning of "suitable" is enlarged to include such features as preservation of monotonicity, the presence of maxima (or minima), and the existence of asymptotes. This type of derivative estimation has been investigated by many authors (for example, see Carnahan, et al., 1969), nearly all of whom warn that small errors in the approximating function tend to be magnified in differentiation.

One should, then, strive for a balance on one hand of faithfulness to the underlying function and, on the other hand, of reduction, by smoothing, of peculiarities in the observations which are present due to error. It is with these goals in mind that the minimization of a Sobolev norm-type expression (equation (3.1)) was proposed.

Specifically, in the case of moving-arc polynomial approximations, consideration must be given to the degree of the approximating polynomial, the span of the arc, and the treatment of the first and last groups of observations. Most literature dealing with choosing the appropriate degree for the approximating polynomial is based on applications to physical experiments with many more observations than we have assumed to be available from biological experimentation (e.g., see Savitzky and Golay, 1964). Unfortunately, even the number of points considered in one span is frequently larger than the total number available from one biological experiment. Such is the case in the paper by Luers and Wenning (1971) who discuss derivative estimation by least-squares polynomial fitting of either cubics or straight lines as a function of the number of points in one span. Apparently, when relatively few observations are available, the choice of both degree and number of points in a span is as much art as science and is based mostly on the intuition and experience of the experimenter. In some biological applications, knowledge of initial values suggests a particular smoothing and/or derivative estimation function.

For the case of approximation by segmented polynomials, choices involving the number and degree of segments and placement of the knots must be made. Some work has been done on choosing knots by least-squares criteria, as in the papers by deBoor and Rice (1968) and Hudson (1966), although in most cases it is still necessary for the user to choose the knots, as Fuller (1969) did, using the data on which he reports.

Two potentially profitable directions for further research include (1) considerations regarding choice of norm and approximating functions guided by graphical display and interactive computing devices, which enable the user to make instantaneous decisions (see LaFata and Rosen, 1970) and (2) diversification of the form of the approximating function, an example of which is the exponential spline function developed by Späth (1969) to avoid undesirable inflection points.

Another problem which would be worthy of more research involves the parameter w in the Sobolev norm-type expression (3.1). For the case where the κ_i are estimated following initial moving-arc smoothing and derivative estimation, the value of w is arbitrary. However, in the cases involving simultaneous smoothing, derivative estimation, and estimation of the κ_i , w has been set equal to one time-unit-squared in this thesis without explanation. This particular choice was made both on the basis of intuition and on the basis of several numerical studies which suggest that this particular value is, in fact, reasonable. The first such study was one in which the single equation model (6.1) from system (6.2) was investigated for $\kappa_1 = 0.20$ and $\kappa_2 = 0.21$. The estimation of the κ_i was accomplished by simultaneous smoothing, derivative estimation and estimation of the κ_i using three cubic segments with knots at $t = 0.5$, 5.5 , 10.5 , and 15.5 , as described in Section 5.1.2. In this study, w in the Sobolev norm-type expression was allowed to vary. The results are shown in Table 7.1 and include $\hat{\kappa}_{1_{exact}}$, $\hat{\kappa}_{2_{exact}}$, $\nu_s^2 = \sum_i [Y(t_i) - f(t_i)]^2$,

$\nu_d^2 = w \sum_i [\dot{Y}(t_i) - g(t_i)]^2$, and $\nu^2 = \nu_s^2 + \nu_d^2$. These results suggest that the minimum value of ν^2 occurs for w equals one time-unit-squared, at least for the values of w considered.

Table 7.1 Values of Sobolev norm-type expression as a function of w . $K_1 = 0.20$ and $K_2 = 0.21$.

w	$\hat{K}_{1exact}$	$\hat{K}_{2exact}$	ν_s^2	ν_d^2	ν^2
0.25	0.1981	0.2076	2.586	13.226	15.812
0.50	0.1984	0.2081	2.936	12.243	15.179
0.75	0.1986	0.2083	3.197	11.816	15.013
0.90	0.1987	0.2084	3.319	11.666	14.985
1.00	0.1988	0.2085	3.390	11.591	14.981
1.10	0.1988	0.2085	3.454	11.530	14.984
1.25	0.1989	0.2086	3.539	11.458	14.997
1.50	0.1989	0.2086	3.655	11.373	15.028
2.00	0.1989	0.2086	3.655	11.373	15.101
3.00	0.1991	0.2088	4.042	11.184	15.226

In a second study, error-free values of $y(t)$ and $z(t)$ were computed for the same model equation as in the above example but for the values of K_1 and K_2 and the time-sequence of observations associated with the example included by Metzler (1969) in the descriptive manual for the computer program NONLIN (see Table 7.2). Estimates of K_1 and K_2 were computed simultaneously with smoothing and derivative estimation by three cubic segments with knots at $t = 0.5$, 5.5 , 20.5 , and 30.5 time-units. The values of $\hat{K}_{1exact}$, $\hat{K}_{2exact}$, ν_s^2 , ν_d^2 , and ν^2 , which are displayed in Table 7.3 for several values of w , suggest that the minimum value of ν^2 does occur at w equals one time-unit-squared. It would clearly be worthwhile to investigate the influence of w on the quality of the estimators $\hat{K}_i$ for data with errors.

Table 7.2 Values of $y(t_i)$ and $z(t_i)$ from solution of system (6.2) for $z_0 = 100$, $\kappa_1 = 0.05775$, $\kappa_2 = 0.1155$. (From Metzler, 1969)

t_i	$y(t_i)$	$z(t_i)$
1.0	10.59	89.09
2.0	19.44	79.37
3.0	26.75	70.72
4.0	32.74	63.00
5.0	37.58	56.13
10.0	49.25	31.51
15.0	48.74	17.68
20.0	43.16	9.93
25.0	36.07	5.57
30.0	29.11	3.13

Table 7.3 Values of Sobolev norm-type expression as a function of w . $\kappa_1 = 0.05775$, $\kappa_2 = 0.1155$.

w	$\hat{\kappa}_{1exact}$	$\hat{\kappa}_{2exact}$	ν_s^2	ν_d^2	ν^2
0.25	0.05661	0.1147	0.007516	0.081135	0.088651
0.5	0.05677	0.1147	0.012562	0.067005	0.079567
0.75	0.05687	0.1147	0.016662	0.060308	0.076970
0.9	0.05691	0.1148	0.018746	0.057770	0.076516
1.0	0.05693	0.1148	0.020015	0.056432	0.076447
1.1	0.05695	0.1148	0.021204	0.055298	0.076502
1.25	0.05697	0.1148	0.022864	0.053882	0.076746
1.5	0.05700	0.1148	0.025374	0.052049	0.077423
2.0	0.05703	0.1147	0.029758	0.049517	0.079275
3.0	0.05710	0.1147	0.037194	0.046485	0.083679

One generalization of the role of w as a weighting coefficient leads to the following Sobolev norm-type expression

$$\nu_w^2 = [(\underline{Y} - \underline{f})^T \mid (\dot{\underline{Y}} - \underline{g})^T] \left[\begin{array}{c|c} \underline{W}_{11} & \underline{W}_{12} \\ \hline \underline{W}_{21} & \underline{W}_{22} \end{array}\right] \left[\begin{array}{c} (\underline{Y} - \underline{f}) \\ \hline (\dot{\underline{Y}} - \underline{g}) \end{array}\right] .$$

If we let the submatrices $\underline{W}_{ij}$ equal $\underline{V}_{ij}^{-1}$ where the $\underline{V}_{ij}$ are the submatrices of the variance-covariance matrix

$$\underline{V} = \left[\begin{array}{c|c} \underline{V}_{11} & \underline{V}_{12} \\ \hline \underline{V}_{21} & \underline{V}_{22} \end{array}\right]$$

associated with the vector

$$\left[\begin{array}{c} (\underline{Y} - \underline{f}) \\ (\dot{\underline{Y}} - \underline{g}) \end{array}\right]$$

then, for the case involving the model equation $\dot{y} = Ky$ with observations $Y(t_i) = y(t_i) + \varepsilon(t_i)$, $i = 1, 2, \ldots, n$, where the $\varepsilon(t_i)$ are independently distributed, and where $\underline{Y}$ is of dimension $(n \times 1)$, the matrix $\underline{W}_{11}$ is diagonal with elements

$$(\underline{W}_{11})_{ii} = \text{var}(Y(t_i))^{-1} .$$

Therefore,

$$\nu_w^2 = [(\underline{Y} - \underline{f})^T \mid (\dot{\underline{Y}} - \underline{g})^T] \left[\begin{array}{c|c} \underline{W}_{11} & K^{-1}\underline{W}_{11} \\ \hline K^{-1}\underline{W}_{11} & K^{-2}\underline{W}_{11} \end{array}\right] \left[\begin{array}{c} (\underline{Y} - \underline{f}) \\ \hline (\dot{\underline{Y}} - \underline{g}) \end{array}\right] .$$

If we assume that $\text{var}(Y(t_i)) = \sigma^2$, $i = 1, 2, \ldots, n$, then $\underline{W}_{11} = \sigma^2 \underline{I}$, so

$$\nu_w^2 = \sigma^2 \left[(\underline{Y} - \underline{f})^T \;\vdots\; (\dot{\underline{Y}} - \underline{g})^T \right] \left[\begin{array}{c|c} \underline{I} & \kappa^{-1}\underline{I} \\ \hline \kappa^{-1}\underline{I} & \kappa^{-2}\underline{I} \end{array} \right] \left[\begin{array}{c} (\underline{Y} - \underline{f}) \\ \hline (\dot{\underline{Y}} - \underline{g}) \end{array} \right] .$$

The Sobolev norm-type expression minimized in this thesis is, in the above form for the case $\dot{y} = \kappa y$,

$$\nu^2 = \left[(\underline{Y} - \underline{f})^T \;\vdots\; (\dot{\underline{Y}} - \underline{g})^T \right] \left[\begin{array}{c|c} \underline{I} & \underline{0} \\ \hline \underline{0} & w\underline{I} \end{array} \right] \left[\begin{array}{c} (\underline{Y} - \underline{f}) \\ \hline (\dot{\underline{Y}} - \underline{g}) \end{array} \right]$$

which suggests that our expression, in the sense of this particular case of variance-covariance weighting, ignores correlations between the $\underline{Y}$ and the $\kappa\underline{Y}$, and which suggests the role assumed by the w , with dimensions identical to those of κ^{-2} . In fact, setting w equal to one time-unit-squared, as we have done, is tantamount to designating $\text{var}(\underline{Y}) = \underline{I}$ and $\text{var}(\kappa\underline{Y}) = \underline{I}$.

Obviously, the development of estimators of κ by minimizing a Sobolev norm-type expression with a weighting scheme is deserving of future study. One important aspect of such a study must be the choosing of the weighting matrices. Although the inverses of the covariance matrices are used in the usual linear least-squares techniques to insure minimum variance of the estimators for the parameters of the model, it is not clear that use of the covariance matrices in the Sobolev norm-type expression for the estimation of parameters in differential systems will result in minimum variance estimates of the κ_i . In fact, in the above example, $\underline{V}^{-1}$ does not even exist. Even if the inverses of the partitioned covariance matrices are used as weighting matrices, the values of $var(Y(t_i))$ and of κ must be known or estimated. Such estimation, if required, would possibly be accomplished by an iterative scheme. Further, for cases where the $Y(t_i)$ are normally distributed and where the vector $\underline{Y}$ in the Sobolev norm-type expression is of dimension (m×1), where $m > n$ as in simultaneous 5-point moving-arc smoothing, derivative estimation, and estimation of κ , the computation of the inverse of the covariance matrices would probably require use of a form of the generalized inverse.

A second major area of potential future interest involves the extension of the basic method, i.e., minimization of a Sobolev norm-type expression, to other forms of the model differential equations. It is our intention to include in this category both the way the variables occur in the differential equation and the form of the coefficients. Perhaps, the simplest example is the differential equation of the type

$$\frac{dy(t)}{dt} = \kappa\, z(t)$$

which is often incorporated in models of the excretion phase of pharmacokinetic studies (Cummings and Martin, 1964), where $y(t)$ is the amount of a drug metabolite in the body and $z(t)$ is the amount of drug in the body. Assuming that observed values of each of the variables y and z are available at discrete times t_i , then the Sobolev norm-type expression to be minimized is of the form

$$\nu^2 = (\underline{Y} - \underline{\Phi\beta})^T(\underline{Y} - \underline{\Phi\beta}) + w(\kappa\underline{Z} - \underline{\Phi}'\underline{\beta})^T(\kappa\underline{Z} - \underline{\Phi}'\underline{\beta})$$

from which the following normal equations are derived

$$\left(\begin{array}{c|c} \underline{\Phi}^T\underline{\Phi} + w\,\underline{\Phi}'^T\underline{\Phi}' & -w\,\underline{\Phi}'^T\underline{Z} \\ \hline -w\underline{Z}^T\underline{\Phi}' & w\underline{Z}^T\underline{Z} \end{array}\right) \left(\begin{array}{c} \hat{\underline{\beta}} \\ \hline \hat{\kappa} \end{array}\right) = \left(\begin{array}{c} \underline{\Phi}^T\underline{Y} \\ \hline 0 \end{array}\right) .$$

Solving for $\hat{\kappa}$ yields

$$\hat{\kappa} = \frac{\underline{Z}^T\,\underline{E}\,\underline{Y}}{\underline{Z}^T\,\underline{F}\,\underline{Z}} \tag{7.1}$$

an expression which is similar to equations (4.5) and (4.18) for $\hat{\kappa}$ in the case $\frac{dy}{dt} = \kappa y$. Obviously, the derivation of the estimate of the κ (equation 7.1) is straightforward; the investigation of the properties of $\hat{\kappa}$ is made easier by the assumption of additive and independent errors $\varepsilon(t_i)$ and $\zeta(t_i)$ associated with $Y(t_i)$ and $Z(t_i)$, respectively.

A second example is the well-known system

$$\frac{dy}{dt} = \kappa_1 y - \kappa_2 yz$$

$$\frac{dz}{dt} = -\kappa_3 z + \kappa_4 yz$$

proposed by Volterra in his study of fish populations (see Goel et al. (1971)). Depending on whether or not relationships exist between the κ_i , estimates of the κ_i can be derived from the methods discussed in Sections 5.1 or 5.3 . It is the presence of non-linear errors from the product $\underline{Y}^T\underline{Z} = (\underline{y} + \underline{\varepsilon})^T(\underline{z} + \underline{\zeta})$ which calls for further investigation.

A third related example which involves both the presence of variables non-linearly and a non-linear relationship between the κ_i is the differential equation

$$\frac{dy}{dt} = \kappa_1 y \left[\frac{\kappa_2 - y}{\kappa_2} \right]$$

$$= \kappa_1 y - \kappa_3 y^2 \quad , \tag{7.2}$$

where $\kappa_3 = \frac{\kappa_1}{\kappa_2}$, which Gause (1934) applied in his famous work on yeast growth, where y is the size (in volume units) of a pure culture of <u>Saccharomyces</u> <u>cerevisiae</u>. From his observations, which are shown in Table 7.4, Gause estimated κ_2 , the maximum population size, by visual methods to be 13.0 volume-units and κ_1 , the "coefficient of geometric increase", by linear regression of $\log \left[\frac{\kappa_2 - y}{y} \right]$ on t to be 0.21827 . Since equation (7.2) is linear in κ_1 and and $\frac{\kappa_1}{\kappa_2}$, it is of interest to obtain estimates of κ_1 and $\frac{\kappa_1}{\kappa_2}$, and hence, of κ_1 and κ_2 , by minimization of the Sobolev norm-type expression (3.1) . For simultaneous smoothing, derivative estimation, and estimation of the κ_i by one cubic segment with knots at $t = 6$ and 40 time-units, the estimates of $\kappa_1 = 0.184$ and $\kappa_2 = 13.28$ are obtained. This particular choice of knots, although it eliminates consideration of the observations at $t = 48$ and 53 time-units, is justified by knowledge of the behavior of the cubic polynomial in approximating a group of observations near the horizontal asymptote. It might be worthwhile to compare the quality of $\hat{\kappa}_2$ by regressing

$$\log\left[\frac{\kappa_2 - y}{y}\right]$$ on t with our method.

Table 7.4 Observations reported by Gause (1934) for the growth of Saccharomyces cerevisiae.

time, hrs.	population size (volume units)
6	0.37
16	8.87
24	10.66
29	12.50
40	13.27
48	12.87
53	12.70

As a preliminary demonstration inspired by Gause's data, values of $y(t)$ were computed from equation (7.2) for $\kappa_1 = 0.25$, $\kappa_2 = 13.0$, $y(6) = 0.4$ by a fourth-order Runge-Kutta scheme with step-size two. For the eight constructed values of $y(t)$ shown in Table 7.5, estimates of κ_1 and κ_2 were obtained by the above simultaneous procedure with two cubic segments with knots at $t = 5.5$, 27.0, and 48.5 time-units. From artificially constructed observations $Y(t_i) = y(t_i) + \varepsilon(t_i)$, where $\varepsilon(t_i) \sim n\left(0, \left[\frac{p}{100}\, y(t_i)\right]^2\right)$, $p = 3$ and 5, estimates of κ_1 and κ_2 were computed by the same method. The results are displayed in Table 7.6 and exhibit a strong suggestion of significant bias in $\hat{\kappa}_1$, and, therefore, in $\hat{\kappa}_2$. Interest in models of the type (7.2) should probably initially center on the effect of squared error terms, that is, on comparing the influence of either squaring the smoothed $Y(t_i)$ or smoothing the $Y(t_i)^2$.

Table 7.5 Values of $y(t_i)$ computed by numerical integration of $\frac{dy}{dt} = \kappa_1 y \left[\frac{\kappa_2 - y}{\kappa_2}\right]$. $\kappa_1 = 0.4$, $\kappa_2 = 13.0$.

time	y(t)
6	0.4
12	1.62
18	5.06
24	9.63
30	12.06
36	12.78
42	12.95
48	12.99

Table 7.6 Estimates of κ_1 and κ_2 in $\frac{dy}{dt} = \kappa_1 y \left[\frac{\kappa_2 - y}{\kappa_2}\right]$.

Entries are means and standard deviations (in parentheses) of samples of size n . $\kappa_1 = 0.25$ and $\kappa_2 = 13.0$.

p	n	$\hat{\kappa}_1$	$\hat{\kappa}_2$
0	-	0.2417	13.087
3	20	0.2358	13.160
		(0.0143)	(0.3350)
5	25	0.2208	13.390
		(0.0271)	(0.5053)

The final example of the types of differential equations suggested for future investigation is of the form

$$\frac{dy}{dt} = K_1 y^{K_2}$$

which was suggested from empirical considerations in a study of water vapor sorption in wood (Kelly and Hart, 1970). One possible approach in the estimation of the K_i involves the use of transformations to yield an equation of the form

$$\log \frac{dy}{dt} = \log K_1 + K_2 \log y \quad .$$

Another suggested area of further research is motivated by two facts:

(1) Sigmoid curves are often found in studies relating drug dose and physiological response.

(2) Certain differential equations are known to have solutions, the graphs of which are sigmoid curves.

In many experimental situations, an estimate is sought either for the dose which is lethal to 50 percent of the experimental subjects, *e.g.*, the so-called LD_{50}, or for the dose which induces a 50 percent-of-maximum response, *e.g.*, the so-called ED_{50}. Most current methods of estimating these parameters rely on some scheme which purports to linearize the dose-response relationship to permit the application of simple linear regression techniques. Besides often lacking theoretical foundations, the basis of these methods frequently puts restrictions on the inclusion of observations associated with 0 percent and 100 percent responses. The estimation techniques advanced in this thesis have potential value in deriving estimators of the dose-response parameters, since, in many interesting cases, these parameters are functions of the coefficients of the differential equations whose solutions are sigmoid curves.

The final recommended area of further research is that which would compare the methods of this thesis with those presently available under the general title of non-linear regression procedures, which involve minimizing the sum of squared deviations between either analytical or numerical solutions to the differential equations and the observed values.

8. LIST OF REFERENCES

Blankinship, S. (ed.) 1971. The User's Guide to PL/C. Document Number LSR-095-1, Triangle Universities Computation Center, Research Triangle Park, North Carolina.

deBoor, C. and J. Rice. 1968. Least Square Cubic Spline Approximation I: Fixed Knots. Report CSD TR 20, Computer Science Department, Purdue University, Lafayette, Indiana.

deBoor, C. and J. Rice. 1968. Least Square Cubic Spline Approximation II: Variable Knots. Report CSD TR 21, Computer Science Department, Purdue University, Lafayette, Indiana.

Cannon, J. R. and D. L. Filmer. 1967. The Determination of Unknown Parameters in Analytic Systems of Ordinary Differential Equations. SIAM Journal in Applied Mathematics 15:799-809.

Cannon, J. R. and D. L. Filmer. 1968. A Numerical Experiment on the Determination of Unknown Parameters in an Analytic System of Ordinary Differential Equations. Mathematical Biosciences 3:267-274.

Carnahan, B., H. A. Luther, and J. O. Wilkes. 1969. Applied Numerical Methods. John Wiley and Sons, Inc., New York, New York.

Cleland, W. W. 1967. The Statistical Analysis of Enzyme Kinetic Data, pp. 1-32. In F. F. Nord (ed.), Advances in Enzymology, Volume 29.

Cook, D. A. and G. S. Taylor. 1971. The Use of the APL/360 System in Pharmacology: A Computer Assisted Analysis of Efflux Data. Computers and Biomedical Research 4:157-166.

Cummings, A. J. and B. K. Martin. 1964. Drug Elimination by Simultaneous First-Order and Zero-Order Processes. Nature 202:779-780.

Evert, C. F. and M. J. Randall. 1970. Formulation and Computation of Compartment Models. Journal of Pharmaceutical Sciences 59:403-409.

Fuller, W. A. 1969. Grafted Polynomials as Approximating Functions. The Australian Journal of Agricultural Economics 13:35-46. (Reprint Series Number 255, Statistical Laboratory, Iowa State University of Science and Technology, Ames, Iowa.)

Gardner, D. G. 1963. Resolution of Multi-Component Exponential Decay Curves Using Fourier Transforms, pp. 195-203. In Multi-Compartment Analysis of Tracer Experiments, Annals of the New York Academy of Sciences, Volume 108, Article 1.

Gause, G. F. 1934. The Struggle for Existence. The Williams and Wilkins Company, Baltimore, Maryland.

Glass, H. I. and A. C. deGarreta. 1967. Quantitative Analysis of Exponential Curve Fitting for Biological Applications. Physics in Medicine and Biology 12:379-388.

Goel, N. S., S. C. Maitra, and E. W. Montroll. 1971. Nonlinear Models of Interacting Populations. Academic Press, Inc., New York, New York.

Goldberger, A. S. 1964. Econometric Theory. John Wiley and Sons, Inc., New York, New York.

Greville, T. N. E. 1968. Data Fitting by Spline Functions. Mathematics Research Center Technical Summary Report Number 893, The University of Wisconsin, Madison, Wisconsin.

Greville, T. N. E. 1969. Introduction to Spline Functions, pp. 1-35. In T. N. E. Greville (ed.), Theory and Applications of Spline Functions. Academic Press, New York, New York.

Heinmets, F. 1970. Quantitative Cellular Biology: An Approach to the Quantitative Analysis of Life Processes. Marcel Dekker, New York, New York.

Hudson, D. J. 1966. Fitting Segmented Curves Whose Join Points Have To Be Estimated. Journal of the American Statistical Association 61:1097-1129.

Joksch, H. C. 1966. Random Errors of Derivatives Obtained from Least Squares Approximations to Empirical Functions. SIAM Review 8:47-56.

Kelly, M. W. and C. A. Hart. 1970. Water Vapor Sorption Rates by Wood Cell Walls. Wood and Fiber 1:270-282.

LaFata, P. and J. B. Rosen. 1970. An Interactive Display for Approximation by Linear Programming. Communications of the ACM 13:651-659.

Lewi, P. J., A. E. F. Chandler, D. G. VanRiel, P. J. Somers, F. T. N. Allewijn, J. G. H. Dony, and P. A. J. Janssen. 1970. Computer Procedures for Evaluating the Results of Studies on the Distribution and Metabolism of Drugs in the Body. Arzneimittel-Forschung 20:684-693.

Luers, J. K. and R. H. Wenning. 1971. Polynomial Smoothing - Linear VS Cubic. Technometrics 13:589-600.

Martin, B. K. 1967. Drug Urinary Excretion Data: Some Aspects Concerning the Interpretation. British Journal of Pharmacology and Chemotherapy 29:181-193.

Metzler, C. M. 1969. A Users Manual for NONLIN. Technical Report Number 7292/69/7292/005, The Upjohn Company, Kalamazoo, Michigan.

Metzler, C. M., G. Matrone, and H. L. Lucas, Jr. 1965. Estimation of Transport Rates by Radioisotope Studies of Non-Steady-State Systems. Institute of Statistics Mimeo Series Number 446, North Carolina State University, Raleigh, North Carolina.

Myhill, J. 1967. Investigation of the Effect of Data Error in the Analysis of Biological Tracer Data. Biophysical Journal 7:903-911.

Myhill, J. 1969. Investigation of the Effect of Data Error in the Analysis of Biological Tracer Data from a Three-Compartment System. Journal of Theoretical Biology 23:218-231.

Parrish, J. D. and S. B. Saila. 1970. Interspecific Competition, Predation and Species Diversity. Journal of Theoretical Biology 27:207-220.

Perl, W. 1960. A Method for Curve-Fitting by Exponential Functions. International Journal of Applied Radiation and Isotopes 8:211-222.

Pizer, S. M., A. B. Ashare, A. B. Callahan, and G. L. Brownell. 1969. Fourier Transform Analysis of Tracer Data, pp. 105-129. In F. Heinmets (ed.), Concepts and Models in Biomathematics: Simulation Techniques and Methods. Marcel Dekker, New York, New York.

Reinsch, C. H. 1967. Smoothing by Spline Functions. Numerische Mathematik 10:177-183.

Rescigno, A. and G. Segre. 1961. La Cinetica dei Farmaci e dei Traccianti Radioattivi. Editore Boringhieri Societa per Azioni, Torino, via Brofferio 3. English translation, 1966, P. Ariotti (translator), Drug and Tracer Kinetics. Blaisdell Publishing Company, Waltham, Massachusetts.

Rosenbrock, H. H. and C. Storey. 1966. Computational Techniques for Chemical Engineers. Pergamon Press, Oxford, England.

Savitzky, A. and M. J. E. Golay. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 36:1627-1639.

Späth, H. 1969. Exponential Spline Interpolation. Computing 4:225-233.

Swann, W. H. 1969. A Survey of Non-Linear Optimization Techniques. FEBS Letters, Volume 2 Supplement:S39-S55.

Westlake, W. J. 1971. Problems Associated with Analysis of Pharmacokinetic Models. Journal of Pharmaceutical Sciences 60:882-885.

Whittle, P. 1956. The Estimation of Age-Specific Infection Rates from a Curve of Relative Infection. Biometrics 12:154-162.

9. APPENDICES

9.1 Derivation of Equation (4.11)

The right-hand side of equation (4.10) is of the form

$$\hat{K}_{exact} + \frac{1}{\underline{Y}^T\underline{C}\underline{Y}}\left[-\hat{K}_{exact}\,(2\underline{Y}^T\underline{C}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{C}\underline{\varepsilon}) + (\underline{Y}^T\underline{B}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{B}\underline{Y} + \underline{\varepsilon}^T\underline{B}\underline{\varepsilon})\right]$$

$$+ \left[\frac{1}{\underline{Y}^T\underline{C}\underline{Y}}\right]^2 \hat{K}_{exact}\,(2\underline{Y}^T\underline{C}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{C}\underline{\varepsilon})^2 \qquad (9.1)$$

$$- \left[\frac{1}{\underline{Y}^T\underline{C}\underline{Y}}\right]^2 (2\underline{Y}^T\underline{C}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{C}\underline{\varepsilon})(\underline{Y}^T\underline{B}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{B}\underline{Y} + \underline{\varepsilon}^T\underline{B}\underline{\varepsilon}) \quad .$$

Since the $\varepsilon(t_i) = \varepsilon_i$ are assumed to be independent, it follows that the expected value of the second term is

$$- \left[\frac{\hat{K}_{exact}}{\underline{Y}^T\underline{C}\underline{Y}}\right] \mathcal{E}\,(\underline{\varepsilon}^T\underline{C}\underline{\varepsilon}) + \frac{1}{\underline{Y}^T\underline{C}\underline{Y}}\,\mathcal{E}\,(\underline{\varepsilon}^T\underline{B}\underline{\varepsilon})$$

$$= - \left[\frac{\hat{K}_{exact}}{\underline{Y}^T\underline{C}\underline{Y}}\right] \mathcal{E}\,(\sum_{ij}\sum C_{ij}\,\varepsilon_i\varepsilon_j) + \frac{1}{\underline{Y}^T\underline{C}\underline{Y}}\,\mathcal{E}\,(\sum_{ij}\sum B_{ij}\,\varepsilon_i\varepsilon_j)$$

$$= - \left[\frac{\hat{K}_{exact}}{\underline{Y}^T\underline{C}\underline{Y}}\right] \sum_i C_{ii}\,\mathrm{var}(\varepsilon_i) + \frac{1}{\underline{Y}^T\underline{C}\underline{Y}} \sum_i B_{ii}\,\mathrm{var}(\varepsilon_i) \quad . \qquad (9.2)$$

Now,

$$(2\underline{Y}^T\underline{C}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{C}\underline{\varepsilon})^2 = 4\underline{Y}^T\underline{C}\underline{\varepsilon}\underline{Y}^T\underline{C}\underline{\varepsilon} + 4\underline{Y}^T\underline{C}\underline{\varepsilon}\underline{\varepsilon}^T\underline{C}\underline{\varepsilon} + \underline{\varepsilon}^T\underline{C}\underline{\varepsilon}\underline{\varepsilon}^T\underline{C}\underline{\varepsilon} \quad .$$

But $4\underline{Y}^T\underline{C\varepsilon Y}^T\underline{C\varepsilon} = 4[\sum \varepsilon_j^2 (\sum_k Y_k C_{kj})^2]$ + terms in $\varepsilon_i \varepsilon_j$, $i \neq j$,

and $\underline{\varepsilon}^T\underline{C\varepsilon\varepsilon}^T\underline{C\varepsilon} = \sum_i \sum_j \sum_m \sum_k \varepsilon_k \varepsilon_m \varepsilon_j \varepsilon_i C_{km} C_{ij}$

$$= \sum_i \sum_k \varepsilon_i^2 \varepsilon_k^2 C_{ii} C_{kk} + \sum_i \sum_j \varepsilon_i^2 \varepsilon_j^2 C_{ij} C_{ij}$$

$$+ \sum_i \sum_k \varepsilon_i^2 \varepsilon_k^2 C_{ki} C_{ik} - 2 \sum \varepsilon_i^4 C_{ii}^2$$

+ terms in odd powers of the ε_i ,

from which it follows that

$$\mathcal{E} (4\underline{Y}^T\underline{C\varepsilon Y}^T\underline{C\varepsilon}) = 4 \sum_j (\sum Y_k C_{kj})^2 \operatorname{var}(\varepsilon_j) \tag{9.3}$$

and

$$\mathcal{E} (\underline{\varepsilon}^T\underline{C\varepsilon\varepsilon}^T\underline{C\varepsilon}) = \sum_i \sum_k C_{ii}C_{kk} \mathcal{E} (\varepsilon_i^2 \varepsilon_k^2) + \sum \sum (C_{ij})^2 \mathcal{E} (\varepsilon_i^2 \varepsilon_j^2)$$

$$+ \sum_i \sum_k (C_{ki}C_{ik}) \mathcal{E} (\varepsilon_i^2 \varepsilon_j^2) - 2 \sum_i C_{ii}^2 \mathcal{E} (\varepsilon_i^4)$$

$$= \sum_i \sum_k C_{ii}C_{kk} \operatorname{var}(\varepsilon_i)\operatorname{var}(\varepsilon_k) + \sum_i \sum_j C_{ij}^2 \operatorname{var}(\varepsilon_i)\operatorname{var}(\varepsilon_j)$$

$$+ \sum_i \sum_k C_{ki}C_{ik} \operatorname{var}(\varepsilon_i)\operatorname{var}(\varepsilon_k) \quad . \tag{9.4}$$

The expected value of the fourth term of (9.1) is computed in a similar manner. First,

$$(2\underline{Y}^T\underline{C\varepsilon} + \underline{\varepsilon}^T\underline{C\varepsilon})(\underline{Y}^T\underline{B\varepsilon} + \underline{\varepsilon}^T\underline{BY} + \underline{\varepsilon}^T\underline{B\varepsilon}) = 2\underline{Y}^T\underline{C\varepsilon Y}^T\underline{B\varepsilon} + 2\underline{Y}^T\underline{C\varepsilon\varepsilon}^T\underline{BY}$$

$$+ 2\underline{Y}^T\underline{C\varepsilon\varepsilon}^T\underline{B\varepsilon} + \underline{\varepsilon}^T\underline{C\varepsilon Y}^T\underline{B\varepsilon} + \underline{\varepsilon}^T\underline{C\varepsilon\varepsilon}^T\underline{BY} + \underline{\varepsilon}^T\underline{C\varepsilon\varepsilon}^T\underline{B\varepsilon} \quad .$$

Now,

$$\mathcal{E}\ (2\underline{Y}^T\underline{C\varepsilon}\underline{Y}^T\underline{B\varepsilon}) = 2 \sum_j (\sum_k C_{kj}Y_k)(\sum_i B_{ij}Y_i)\ \mathrm{var}(\varepsilon_j) \quad , \qquad (9.5)$$

$$\mathcal{E}\ (2\underline{Y}^T\underline{C\varepsilon}\underline{\varepsilon}^T\underline{BY}) = 2 \sum_i (\sum_k C_{ki}Y_k)(\sum_j B_{ij}Y_j)\ \mathrm{var}(\varepsilon_i) \quad , \qquad (9.6)$$

and

$$\mathcal{E}\ (\underline{\varepsilon}^T\underline{C\varepsilon}\underline{\varepsilon}^T\underline{B\varepsilon}) = \sum_i \sum_k C_{ii}B_{kk}\ \mathrm{var}(\varepsilon_i)\mathrm{var}(\varepsilon_k)$$

$$+ \sum_i \sum_j C_{ij}B_{ij}\ \mathrm{var}(\varepsilon_i)\mathrm{var}(\varepsilon_j) \qquad (9.7)$$

$$+ \sum_i \sum_k C_{ki}B_{ik}\ \mathrm{var}(\varepsilon_i)\mathrm{var}(\varepsilon_k) \quad .$$

The sum of expressions (9.2) through (9.7), where each is weighted as in expression (9.1), yields the desired relationship (4.11) .

9.2 Derivation of Equation (5.26)

Since

$$\begin{bmatrix} \underline{Y}^T\underline{CY} & \underline{Y}^T\underline{CZ} \\ \underline{Z}^T\underline{CY} & \underline{Z}^T\underline{CZ} \end{bmatrix}^{-1} = \frac{1}{d_2} \begin{bmatrix} \underline{Z}^T\underline{CZ} & -\underline{Z}^T\underline{CY} \\ -\underline{Y}^T\underline{CZ} & \underline{Y}^T\underline{CY} \end{bmatrix}$$

where $d_2 = \underline{Y}^T\underline{CY}\underline{Z}^T\underline{CZ} - (\underline{Z}^T\underline{CY})^2$ and since

$$\underline{X} = \begin{bmatrix} \underline{SY} & \underline{SZ} & \underline{0} & \underline{0} \\ \underline{0} & \underline{0} & \underline{SY} & \underline{SZ} \end{bmatrix} \quad ,$$

it follows that

$$(\underline{X}^T\underline{X})^{-1} = \begin{bmatrix} \underline{Z}^T\underline{CZ} & -\underline{Z}^T\underline{CY} & 0 & 0 \\ -\underline{Y}^T\underline{CZ} & \underline{Y}^T\underline{CY} & 0 & 0 \\ 0 & 0 & \underline{Z}^T\underline{CZ} & -\underline{Z}^T\underline{CY} \\ 0 & 0 & -\underline{Y}^T\underline{CZ} & \underline{Y}^T\underline{CY} \end{bmatrix} \cdot \frac{1}{d_2}$$

Therefore,

$$\underline{R}(\underline{X}^T\underline{X})^{-1}\underline{R}^T = \frac{1}{d_2} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} (\underline{X}^T\underline{X})^{-1} \begin{bmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \underline{Z}^T\underline{CZ} & -\underline{Z}^T\underline{CY} \\ -\underline{Y}^T\underline{CZ} & 2\underline{Y}^T\underline{CY} \end{bmatrix} \cdot \frac{1}{d_2}$$

and

$$[\underline{R}(\underline{X}^T\underline{X})^{-1}\underline{R}^T]^{-1} = \begin{bmatrix} 2\underline{Y}^T\underline{CY} & \underline{Y}^T\underline{CZ} \\ \underline{Z}^T\underline{CY} & \underline{Z}^T\underline{CZ} \end{bmatrix} \cdot \frac{d_2}{d_1}$$

where $d_1 = 2\underline{Y}^T\underline{CYZ}^T\underline{CZ} - (\underline{Y}^T\underline{CZ})^2$, from which it follows that

$$(\underline{X}^T\underline{X})^{-1}\underline{R}^T[\underline{R}(\underline{X}^T\underline{X})^{-1}\underline{R}^T]^{-1} = \frac{1}{d_2} \begin{bmatrix} 0 & -\underline{Y}^T\underline{CZ} \\ 0 & \underline{Y}^T\underline{CY} \\ \underline{Z}^T\underline{CZ} & -\underline{Y}^T\underline{CZ} \\ \underline{Z}^T\underline{CY} & \underline{Y}^T\underline{CY} \end{bmatrix} \begin{bmatrix} 2\underline{Y}^T\underline{CY} & \underline{Y}^T\underline{CZ} \\ \underline{Z}^T\underline{CY} & \underline{Z}^T\underline{CZ} \end{bmatrix} \cdot \frac{d_2}{d_1}$$

$$= \begin{bmatrix} -\underline{Y}^T\underline{CZZ}^T\underline{CY} & -\underline{Y}^T\underline{CZZ}^T\underline{CZ} \\ \underline{Y}^T\underline{CYZ}^T\underline{CY} & \underline{Y}^T\underline{CYZ}^T\underline{CZ} \\ \vdots & \vdots \end{bmatrix} \frac{1}{d_1} .$$

However,

$$\underline{r} - \underline{Rb} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{Z}^T\mathbf{CZY}^T\mathbf{BY} - \mathbf{Z}^T\mathbf{CYZ}^T\mathbf{BY} \\ \mathbf{Y}^T\mathbf{CYZ}^T\mathbf{BY} - \mathbf{Y}^T\mathbf{CZY}^T\mathbf{BY} \\ \mathbf{Z}^T\mathbf{CZY}^T\mathbf{BZ} - \mathbf{Z}^T\mathbf{CYZ}^T\mathbf{BZ} \\ \mathbf{Y}^T\mathbf{CYZ}^T\mathbf{BZ} - \mathbf{Y}^T\mathbf{CZY}^T\mathbf{BZ} \end{bmatrix} \cdot \frac{1}{d_2}$$

$$= \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} \mathbf{Z}^T\mathbf{CZY}^T\mathbf{BZ} - \mathbf{Z}^T\mathbf{CYZ}^T\mathbf{BZ} \\ \hline \mathbf{Y}^T\mathbf{CYZ}^T\mathbf{BY} - \mathbf{Y}^T\mathbf{CZY}^T\mathbf{BY} + \mathbf{Y}^T\mathbf{CYZ}^T\mathbf{BZ} \\ - \mathbf{Y}^T\mathbf{CZY}^T\mathbf{BZ} \end{bmatrix} \cdot \frac{1}{d_2}$$

so

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T[\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{R}^T]^{-1} \quad (\underline{r} - \underline{Rb})$$

$$= \frac{1}{d_1 d_2} \begin{bmatrix} -\mathbf{Y}^T\mathbf{CZZ}^T\mathbf{CY} & -\mathbf{Y}^T\mathbf{CZZ}^T\mathbf{CZ} \\ \mathbf{Y}^T\mathbf{CYZ}^T\mathbf{CY} & \mathbf{Y}^T\mathbf{CYZ}^T\mathbf{CZ} \\ \vdots & \vdots \end{bmatrix} \begin{bmatrix} -\mathbf{Z}^T\mathbf{CZY}^T\mathbf{BZ} + \mathbf{Z}^T\mathbf{CYZ}^T\mathbf{BZ} \\ \hline -\mathbf{Y}^T\mathbf{CYZ}^T\mathbf{BY} + \mathbf{Y}^T\mathbf{CZY}^T\mathbf{BY} \\ -\mathbf{Y}^T\mathbf{CYZ}^T\mathbf{BZ} + \mathbf{Y}^T\mathbf{CZY}^T\mathbf{BZ} \end{bmatrix}$$

and the indicated product is exactly equation (5.26).