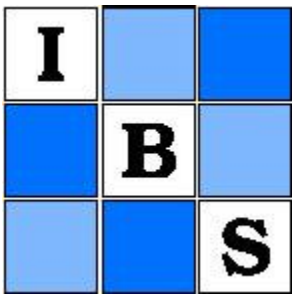


REFERENCES

BIOMETRICS 54, 401–415
June 1998

Nonparametric Estimation of Covariance Structure in Longitudinal Data

Peter J. Diggle

Department of Mathematics and Statistics,
Lancaster University, Lancaster LA1 4YF, U.K.

and

Arūnas P. Verbyla*

Department of Statistics, University of Adelaide,
Adelaide, SA 5005, Australia

SUMMARY

In longitudinal studies, the effect of various treatments over time is usually of prime interest. However, observations on the same subject are usually correlated and any analysis should account for the underlying covariance structure. A nonparametric estimate of the covariance structure is useful, either as a guide to the formulation of a parametric model or as the basis for formal inference without imposing parametric assumptions. The sample covariance matrix provides such an estimate when the data consist of a short sequence of measurements at a common set of time points on each of many subjects: but is impractical when the data are severely unbalanced or when the sequences of measurements on individual subjects are long relative to the number of subjects. The variogram of residuals from a saturated model for the mean response has previously been suggested as a nonparametric estimator for covariance structure assuming stationarity. In this paper, we consider kernel weighted local linear regression smoothing of sample variogram ordinates and of squared residuals to provide a nonparametric estimator for the covariance structure without assuming stationarity. The value of the estimator as a diagnostic tool is demonstrated in two applications, one to a set of data concerning the blood pressure of newborn babies in an intensive care unit and the other to data on the time evolution of CD4 cell numbers in HIV seroconverters. The use of the estimator in more formal statistical inferences concerning the mean profiles requires further study.

1. Introduction

A longitudinal data set consists of a collection of time series, one from each of a number of subjects in a designed experiment or observational study. As a specific example, Figure 1 shows a set of data from an investigation into the time variation in the blood pressure of newborn babies in an intensive care unit. These babies were premature and had a very low weight at birth. One aim of the study was the establishment of a standard curve for blood pressure over time. The data consist of a time sequence (in days) of blood pressure measurements (mm Hg) on each of 43 babies. The total number of observations is 1054, corresponding to an average of about 25 measurements per baby but with a range from 9 to 42. The number of distinct measurement times is 47. Most of the incompleteness stems from babies who left the intensive-care unit before completion of the study. Departure was not specifically related to a baby's blood pressure history but to their general well-being, and so we assume that the missing data are noninformative (Little and Rubin, 1987) with respect to blood pressure.

* *Corresponding author's email address:* averbyla@stats.adelaide.edu.au

Key words: Correlated errors; Kernel estimation; Kernel weighted local regression; Longitudinal data; Repeated measurements; Smoothing; Variogram.

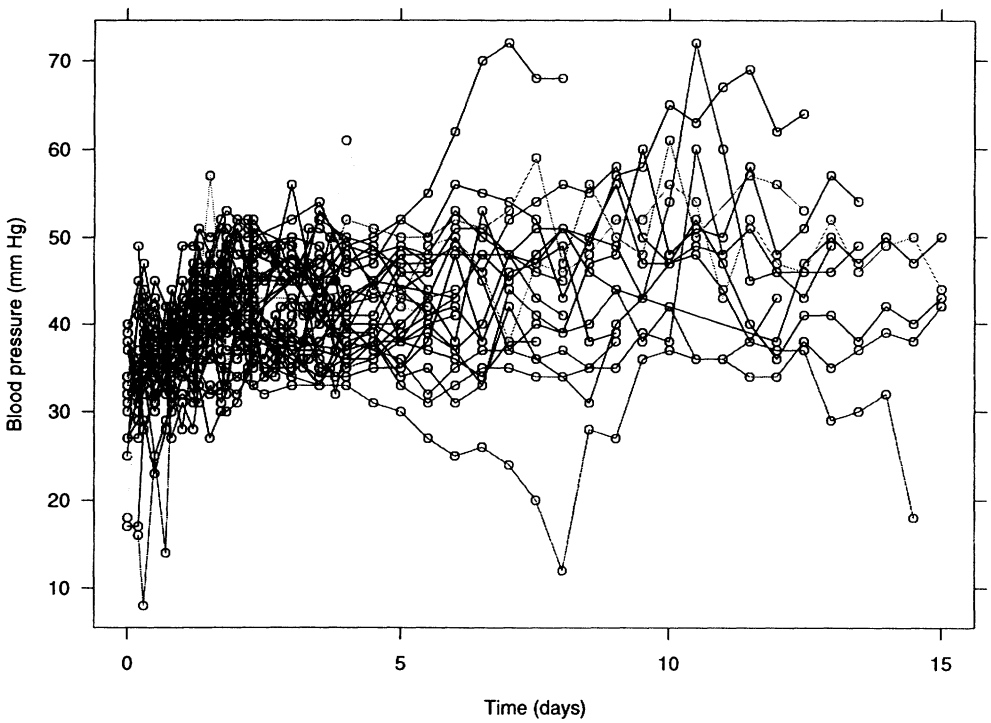


Figure 1. Blood pressure measurements over time of 43 newborn premature babies of very low birth weight.

As in many longitudinal studies, a basic objective in this example is to describe the mean response as a function of time. Any inferences about the mean response must recognise the likely correlation structure among the repeated observations on each subject.

A common approach to the analysis of this kind of data is to assume a Gaussian linear model for the mean response, together with a model for the covariance structure of the responses on each subject. We write a general longitudinal data set as

$$(y_{ij}, t_{ij}) : j = 1, \ldots, n_i;\ i = 1, \ldots, m,$$

where y_{ij} denotes the jth of n_i measurements on the ith of m subjects and t_{ij} the time at which the measurement y_{ij} is made. We assume that y_{ij} is a realisation of a random variable $Y_i(t_{ij})$, where $\{Y_i(t) : t \in \Re\} : i = 1, \ldots, m$ are a set of m mutually independent Gaussian random processes with mean value functions $\mu_i(t) = \mathrm{E}[Y_i(t)]$ and common covariance function $G(s,t) = \mathrm{cov}\{Y_i(s), Y_i(t)\}$. It is convenient to write $\sigma^2(t) = G(t,t)$ for the variance function of the processes $Y_i(t)$. Let $\mu_{ij} = \mu_i(t_{ij})$, $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{in_i})'$. If $\boldsymbol{Y}_i$ denotes the vector of responses for the ith of m subjects and $\boldsymbol{t}_i$ denotes the corresponding vector of times at which these responses are made, we have that $\boldsymbol{Y}_i : i = 1, \ldots, m$ are mutually independent Gaussian random vectors, with

$$\boldsymbol{Y}_i \sim \mathrm{MVN}\{\boldsymbol{\mu}_i, V(\boldsymbol{t}_i, \boldsymbol{\phi})\}. \tag{1}$$

In many cases, it is assumed that

$$\mu_{ij} = \boldsymbol{x}_{ij}(t_{ij})'\boldsymbol{\beta}, \tag{2}$$

where $\boldsymbol{x}_{ij}(t_{ij})$ is a vector of explanatory variables that constitutes the jth row of the n_i by p matrix X_i. In some situations, a nonparametric smoother may be used, so that $\mu_{ij} = f(t_{ij})$ for some smooth function $f(\cdot)$, or a semiparametric model

$$\mu_{ij} = \boldsymbol{x}_{ij}(t_{ij})'\boldsymbol{\beta} + f(t_{ij}) \tag{3}$$

may be used, as in Zeger and Diggle (1994).

In (1), $V(\boldsymbol{t}_i, \boldsymbol{\phi})$ is the $n_i \times n_i$ covariance matrix of $\boldsymbol{Y}_i$, which is specified by the observation times $\boldsymbol{t}_i$ and a set of q parameters $\boldsymbol{\phi}$. The elements of $V(\boldsymbol{t}_i, \boldsymbol{\phi})$ are defined in terms of the generic covariance function $G(s,t)$, which is therefore a function of $\boldsymbol{\phi}$.

There is an extensive literature on methods for the analysis of longitudinal data within the framework of (1) and (2) (see the review in Diggle, Liang, and Zeger, 1994, Chapter 5). Kernel smoothing is used in a semiparametric model by Zeger and Diggle (1994). A recurrent issue in the formulation of models for longitudinal data is the extent to which we can, or should, impose a particular parametric structure on the generic covariance matrix $G(s,t)$ and hence on V. Too tight a specification risks invalidating the inferences about the mean structure, which is usually of primary interest, whereas too loose a specification leads to a loss of efficiency since we then waste some of the information in the data by estimating redundant nuisance parameters (Altham, 1984). In practice, the balance between these two considerations depends on many things, including the shape of the data matrix. When m, the number of subjects, is large and the n_i, the numbers of observations per subject, are small, a nonparametric specification, that is an unstructured V, is feasible. This is the approach taken by Liang and Zeger (1986) and Zeger and Liang (1986) in the wider setting of generalized linear models. They estimate the parameter $\boldsymbol{\theta}$ by iteratively weighted least squares using a qualitatively sensible weighting matrix, but the standard errors for $\boldsymbol{\theta}$ are derived using the sample covariance matrix of the residuals as an estimate of V.

Conversely, when m is small and the n_i are relatively large, it becomes essential to impose structural restrictions on V either by smoothing or by fitting a parametric model. Many of the papers cited above are concerned with formulating plausible parametric models for V in which ϕ is of low dimension. Most of these parametric models assume that the variation about the mean response is stationary. When nonstationary variation is allowed, it is assumed to be of a highly structured form, as in the random effects models of Laird and Ware (1982), the ante-dependence models of Kenward (1987), or the first-difference models of Cullis and McGilchrist (1990).

In this paper, we develop smooth nonparametric estimators for components of V. Smooth estimation of V has been considered by a few authors, namely Glasbey (1988), Shapiro and Botha (1991), Sampson and Guttorp (1992), Hall, Fisher, and Hoffmann (1994), and Hall and Patil (1994). Most of these authors consider estimation in the case of stationary processes. We do not wish to impose stationarity as a prior assumption, but our methodology includes stationary variation as a special case. Our aim is to provide a diagnostic tool for parametric modelling and also to fill a niche in which parametric modelling fails to provide a convincing fit to the data, but the size and shape of the data matrix precludes a nonparametric estimation using the unsmoothed residual covariance matrix. The data in Figure 1 typify this situation. On the one hand, using the sample covariance matrix to estimate V would be equivalent to fitting a covariance model with $(1/2) \times 47 \times 48 = 1128$ parameters. On the other hand, it may be difficult to postulate a plausible parametric model for these data, and the variation about the mean response may well be nonstationary. In such cases, a smooth nonparametric estimator provides a compromise modelling strategy, as a smooth estimator will take up fewer degrees of freedom than required by the sample covariance matrix (see Hastie and Tibshirani, 1990, Section 3.5).

In Section 2, we establish more formally the ideas, terminology, and notation that we need for the remainder of the paper, with a discussion of the elements of our smoothing approach based on the variogram and squared residuals. Section 3 treats the nonparametric smoothing problem. Section 4 describes the application to the data shown in Figure 1 and to data on the time evolution of CD4 cell numbers in HIV seroconverters. Section 5 summarizes the results of the paper and identifies some areas where further work is needed.

2. Components in Covariance Modelling

2.1 *The Variogram*

For a general Gaussian process $Y(t)$ with mean value function $\mu(t)$ and covariance function $G(s,t)$, we define the residual process to be the zero-mean process $Z(t) = Y(t) - \mu(t)$. Then, the variogram of $Z(t)$ is the function

$$\gamma(s,t) = \frac{1}{2}\mathrm{E}\left[\{Z(s) - Z(t)\}^2\right] : s \neq t.$$

Note that some authors call $\gamma(\cdot)$ the semivariogram. Also,

$$\gamma(s,t) = \frac{1}{2}\{G(s,s) + G(t,t) - 2G(s,t)\}. \tag{4}$$

Thus, the variogram is closely related to the covariance function. However, from some points of view, the variogram is more useful for estimation purposes. This is especially so for applications to longitudinal data in which the observation times are not common to all subjects (Diggle, 1988).

If the process $Y(t)$ is such that $G(s,t) \to 0$ as $|s-t| \to \infty$, then for large values of $|s-t|$,

$$\gamma(s,t) \simeq \frac{1}{2}\left\{\sigma^2(s) + \sigma^2(t)\right\},$$

where $\sigma^2(t) = \text{var}\{Y(t)\}$. However, in longitudinal applications, the covariance function of $Y(t)$ typically does not have this property because random variation between subjects induces a positive correlation between any two measurements on the same subject, irrespective of the time separation between them. While (4) implies $\gamma(t,t) = 0$, it is typically the case that $\gamma(s,t) \not\to 0$ as $s \to t$ because measurement error is present. Thus, in general, separate estimation of the variance function $\sigma^2(t)$ is required.

Our definitions here and in Section 2.3 do not assume stationarity of the residual process. Most descriptions of the variogram assume that $\gamma(s,t) = \gamma(s-t)$, that is stationarity, or assume what Matheron (1971, p. 53) terms the 'intrinsic hypothesis,' in which the mean and variance of increments of the residual process depend only on the separation $s-t$.

2.2 *Smooth Components of the Covariance Function*

It is convenient at this point to indicate what we mean by smoothness of the covariance matrix or more accurately of components of the covariance structure. Consider the parametric model of Diggle (1988),

$$Y_i(t_{ij}) = \mu_{ij} + U_i + Z_{ij} + W_i(t_{ij}),$$

where μ_{ij} is the mean of the response, $U_i \sim \text{N}(0, \nu^2)$ allows for between-unit variation, $Z_{ij} \sim \text{N}(0, \tau^2)$ allows for measurement error, and $W_i(t_{ij})$ is a Gaussian process with zero mean and covariance function $\sigma^2 R(t_{ij}, t_{ik};\ \rho)$, dependent on parameters ρ. Then

$$V(t_i, \phi) = \tau^2 I_{n_i} + \nu^2 J_{n_i} + \sigma^2 R(t_i, \rho),$$

so that $\phi = (\tau^2,\ \nu^2,\ \sigma^2,\ \rho)$. This structure includes independence, split-plot, and various stationary models as special cases. The actual class of covariance functions considered by Diggle has

$$R(t_{ij}, t_{ik};\ \rho) = \exp\left\{-\rho|t_{ij} - t_{ik}|^c\right\},$$

where ρ is a scalar parameter. The variance function is constant and is

$$\sigma^2(t) = \tau^2 + \nu^2 + \sigma^2,$$

while the variogram can be shown to be

$$\gamma(s,t) = \tau^2 + \sigma^2\left\{1 - \exp\left(-\rho|s-t|^c\right)\right\}.$$

As an example, Figure 2 gives perspective plots of the variogram and the covariance matrix for $\tau^2 = 0.5$, $\nu^2 = 0.1$, $\sigma^2 = 1$, $\rho = 0.01$, and $c = 2$. The important aspects are the smoothness of the variogram (and the variances, which are constant over time) and the essential smoothness of the covariance matrix except along the diagonal, where measurement error induces a jump discontinuity. In smoothing the covariance structure, we would wish to preserve such discontinuities while providing for the natural smoothness of the separate components of the covariance matrix, namely the variance and variogram functions.

2.3 *The Variogram Cloud and Squared Residuals*

For a set of longitudinal data (y_{ij}, t_{ij}) with known mean value functions $\mu_i(t)$, the variogram cloud is the set of points in $\Re^3$,

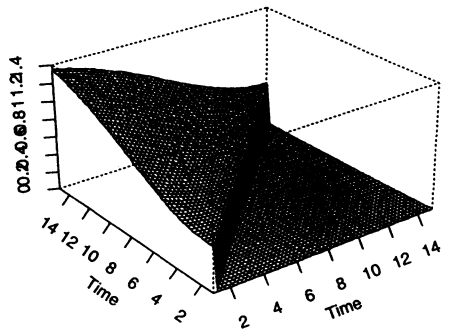
$$(t_{ij}, t_{ik}, v_{ijk}) : i = 1, \ldots, m; j = 1, \ldots, n_i; k > j,$$

where

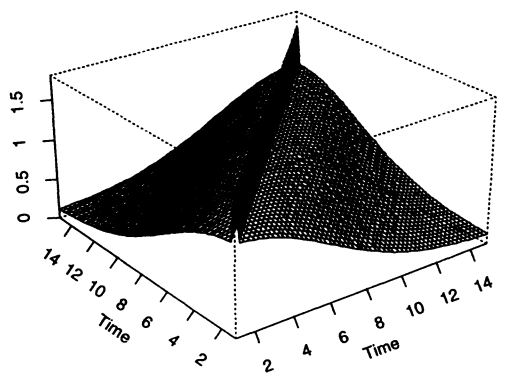
$$v_{ijk} = \frac{1}{2}[\{y_{ij} - \mu_i(t_{ij})\} - \{y_{ik} - \mu_i(t_{ik})\}]^2. \tag{5}$$

When $\mu_i(t)$ is known, and under the assumption that the process $Y_i(t)$ is Gaussian, the sampling distribution of components of the variogram cloud is

$$v_{ijk}/\gamma(t_{ij}, t_{ik}) \sim \chi^2_1,$$



(a)



(b)

Figure 2. Diggle's model with $\tau^2 = 0.5$, $\nu^2 = 0.1$, $\sigma^2 = 1$, $\rho = 0.01$, and $c = 2$. **a.** Variogram; **b.** covariance matrix.

so that v_{ijk} is unbiased for $\gamma(t_{ij}, t_{ik})$. Also, because the covariance structure in the original data arises only through the repeated measurements within subjects, v_{ijk} and $v_{i'j'k'}$ are independent whenever $i \neq i'$. It follows that, if the data contain replicated pairs (t_{ij}, t_{ik}) across subjects and we average such pairs across subjects to define the sample variogram $\bar{v}(t_{ij}, t_{ik})$,

$$\bar{v}(s,t)/\gamma(s,t) \sim \chi^2_r/r,$$

where $r = r(s,t)$ is the number of different subjects contributing to $\bar{v}(s,t)$. When all of the $r(s,t)$ are large, the sample variogram may be an adequate estimator for $\gamma(s,t)$, albeit an inefficient one because it ignores any considerations of smoothness in $\gamma(s,t)$ as s and t vary. When the $r(s,t)$ are small, a smoother estimator for $\gamma(s,t)$ is desirable. Note that, when the data are balanced, in the sense that the observation times are common to all m subjects, $r(s,t) = m$ for all (s,t).

If the mean value structure is known, then the squared residuals, $z^2_{ij} = \{y_{ij} - \mu(t_{ij})\}^2$, are unbiased for $\sigma^2(t)$ and, under Gaussian assumptions, we have the sampling distribution (just like (6) for the variogram)

$$z^2_{ij}/\sigma^2(t_{ij}) \sim \chi^2_1. \tag{7}$$

As for the variogram, if replicated values of z^2_{ij} at each value of t_{ij} are available from different subjects, the sample means of these sets of replicated values provide adequate nonparametric estimates of the variance function. In other cases, a smoother estimator for $\sigma^2(t)$ is again desirable.

The results (6) and (7) are based on the assumption of known $\mu_i(t_{ij})$. In most applications, $\mu_i(t_{ij})$ is unknown, and we then replace it by an appropriate estimate $\hat{\mu}_i(t_{ij})$ (see Diggle et al., 1994, Chapter 4, for a discussion). The modification of (6) and (7) to account for estimation of $\mu_i(t_{ij})$ is known only for special cases. There are no simple results in general, and we are forced to use unadjusted residuals in the analyses of Section 4. There is a bias in using the unadjusted residuals, and work is required to determine how to remove or reduce the bias in practice.

3. Estimation of a Nonstationary Variogram and Variance Function

Our proposal is to use the variogram cloud as the input data for a two-dimensional nonparametric estimator of the variogram, $\gamma(t_{ij}, t_{ik}) = \mathrm{E}[v_{ijk}]$, and to use the squared residuals similarly as the input data for a one-dimensional nonparametric estimator of the variance function, $\sigma^2(t_{ij}) = \mathrm{E}[z_{ij}^2]$. We begin with a review of kernel smoothing, which is required for the kernel weighted local linear regression (denoted by KWLLR) that is used in the analyses of Section 4.

3.1 *Kernel Smoothing*

A simple and natural approach to the problem is kernel smoothing. In general, the particular implementation of each kernel estimator would recognise the covariance structure of the input data. The covariance structure of the variogram ordinates depends on the true underlying covariance structure of the process $Y(t)$. This situation is in contrast to most applications of kernel estimation to correlated data, in which the covariance structure of the data corresponds to a set of unknown nuisance parameters that must be estimated separately from the regression function itself (see, for example, Diggle and Hutchinson, 1988; Hart and Wehrly, 1986; Altman, 1990; Hart, 1991; Zeger and Diggle, 1994; Moyeed and Diggle, 1994).

The kernel method is one of several generic smoothing methods used in statistics. For a general introduction to kernel smoothing, see Härdle (1990) or Hastie and Tibshirani (1990). Other generic smoothing methods include running-line smoothers, of which the best known is lowess (Cleveland, 1979), or smoothing splines (Silverman, 1985). Kernel smoothers are easy to use and interpret and can be applied readily to three-dimensional data.

We first consider the two-dimensional case. For data $(x_i, z_i) : i = 1, \ldots, n$, in which the z_i are the observed values of a response variable and the x_i the observed values of a single explanatory variable, the kernel estimator of $\mu(x) = \mathrm{E}[Z]$ is

$$\hat{\mu}(x) = \sum_{i=1}^{n} w_i^*(x) z_i, \tag{8}$$

where

$$w_i^*(x) = w_i(x) \Big/ \sum_{i=1}^{n} w_i(x)$$

and

$$w_i(x) = h^{-1} w\left[h^{-1}(x - x_i)\right]. \tag{9}$$

In (9), the bandwidth h is a positive constant that determines the amount of smoothing applied to the data z_i and $w(u)$ is a kernel function, a symmetric, nonnegative valued function that is monotone nonincreasing in $|u|$. Common choices for the kernel function include the Gaussian kernel,

$$w(u) = \exp\left(-\frac{1}{2}u^2\right),$$

or the quartic kernel,

$$w(u) = \begin{cases} \left(1 - u^2\right)^2 & \text{if } -1 \le u \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

In the three-dimensional setting, the data are $(x_i, y_i, z_i) : i = 1, \ldots, n$, in which the z_i are again the values of a response variable whose mean value is now a function of two explanatory variables x and y. The kernel estimator for $\mu(x, y) = \mathrm{E}[Z]$ is now

$$\hat{\mu}(x, y) = \sum_{i=1}^{n} w_i^*(x, y) z_i, \tag{10}$$

where

$$w_i^*(x,y) = w_i(x,y) \Big/ \sum_{i=1}^{n} w_i(x,y),$$

$$w_i(x,y) = h^{-2} w\left[h^{-1}\left\{(x-x_i)^2 + (y-y_i)^2\right\}^{\frac{1}{2}}\right], \tag{11}$$

and h and $w(\cdot)$ are as previously defined.

In practice, the choice of kernel function $w(\cdot)$ is much less important than the choice of the value of h. Note also that (10) defines a special case of the general kernel smoother for three-dimensional data, but one that is sensible in our context where both explanatory variables represent the same physical quantity, time. For further comments, see Section 5 below.

3.2 *Kernel Weighted Local Linear Regression*

One problem with kernel smoothing is increased bias near the boundaries of the range of the explanatory variables. Jones (1993) gives an overview of bias correction methods for kernel density estimation and makes important connections with the regression situation. In particular, a linear combination of two kernels, $K(u)$ and $uK(u)$, produces a simple and effective bias correction near the boundaries. In the regression context, this approach corresponds to local linear regression (see Hastie and Loader, 1993). Note that the decrease in bias is accompanied by an increase in variance. However, this also applies to other approaches, often to a greater degree.

In general terms, local regression methods involve fitting of polynomials in a neighbourhood of the point of interest. Introducing kernel weights (thus kernel weighted local linear regression or KWLLR) ties the simple local regression to kernel methods and gives rise to the desirable boundary properties.

In the two-dimensional setting, if $b(x)' = [1\ x]$, $W(x)$ is a diagonal matrix of kernel weights $w_i(x)$ given by (9) and

$$B = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix},$$

the KWLLR estimator of $\mu(x)$ is

$$\hat{\mu}(x) = b(x)'(B'W(x)B)^{-1}B'Wz,$$

where z is the vector of responses. Note that if $b(x) = 1$ and B is the $n \times 1$ vector of ones, we obtain the kernel estimator (8).

The three-dimensional setting gives rise to the same form, but we have $b(x,y)' = [1\ x\ y]$ in place of $b(x)'$, $w_i(x)$ is replaced by $w_i(x,y)$ given by (11), and the matrix B becomes

$$B = \begin{bmatrix} 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & y_n \end{bmatrix}.$$

3.3 *Mean Square Error and Bandwidth Choice*

The critical factor that determines the performance of a kernel-related estimator is the value of the bandwidth h. In some applications, it may be appropriate to choose h subjectively, but in general, it is desirable to have available methods for choosing h automatically from the data. There is an extensive literature on methods of bandwidth selection. For our particular application, we have considered two approaches. Both operate from the premise that a reasonable measure of the performance of an estimator is its total mean square error, defined for the variogram estimator by

$$\mathrm{TMSE}(h) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{j-1} \mathrm{E}\left[\{\hat{\gamma}(t_{ij},t_{ik}) - \gamma(t_{ij},t_{ik})\}^2\right].$$

Note that, for fixed h, the KWLLR estimator is a linear function of the response data. Thus, writing v for the collection of all values of v_{ijk}, γ for the corresponding vector of values of $\gamma(t_{ij}, t_{ik})$, and $\hat{\gamma}$ for the KWLLR estimator of γ, we can write

$$\hat{\gamma} = Wv,$$

where W is a square matrix of weights with ith row $b'(x_i, y_i)(B'W(x_i, y_i)B)^{-1}B'W(x_i, y_i)$. It follows that $\mathrm{E}[\hat{\gamma}] = W\gamma$ and that $\mathrm{var}(\hat{\gamma}) = D\Gamma D'$, where $\Gamma = \mathrm{var}(v)$. Combining these results gives

$$\mathrm{TMSE}(h) = \mathrm{trace}(W\Gamma W') + \gamma'(I - W)'(I - W)\gamma. \tag{12}$$

Evidently, TMSE(h) depends on the unknown variogram γ. Two ways around this difficulty are the following.

First, because the elements of the covariance matrix Γ are explicit functions of γ, we could substitute $\hat{\gamma}$ for γ in (12) and minimise the resulting explicit function of h. However, two problems with this approach are that $\hat{\gamma}$ is a biased estimator for γ and that the explicit form of Γ depends on the assumption that $Y(t)$ is a Gaussian process.

A second approach is to use the cross-validation criterion developed by Rice and Silverman (1991) in the context of spline regression and adapted by Zeger and Diggle (1994) to one-dimensional kernel regression estimation. The cross-validation involves leaving out the information on each subject in turn, and the approach of Zeger and Diggle (1994) can be implemented directly for smoothing the squared residuals using KWLLR. In the two-dimensional context, the cross-validation criterion, CV(h), for unbiased estimation of TMSE(h) up to an undetermined additive constant, is as follows. Write the KWLLR estimator for $\gamma(t_{ij}, t_{ik})$ in the form

$$\hat{\gamma}(t_{ij}, t_{ik}) = \sum_{p=1}^{m}\sum_{q=1}^{n_p}\sum_{r=1}^{q-1} w_{pqr}v_{pqr},$$

where, for each target value (t_{ij}, t_{ik}), the weights w_{pqr} satisfy

$$\sum_{p=1}^{m}\sum_{q=1}^{n_p}\sum_{r=1}^{q-1} w_{pqr} = 1.$$

For each p, define $w_p = \Sigma_{q=1}^{n_p}\Sigma_{r=1}^{q-1} w_{pqr}$. Then

$$\mathrm{CV}(h) = \sum_{i=1}^{m}\sum_{j=1}^{n_i}\sum_{k=1}^{j-1} a_{ijk}^2,$$

where

$$a_{ijk} = \{v_{ijk} - \hat{\gamma}(t_{ij}, t_{ik})\} + \{w_i/(1 - w_i)\}\left\{\sum_{q=1}^{n_i}\sum_{r=1}^{q-1} w_{iqr}v_{iqr}/w_i - \hat{\gamma}(t_{ij}, t_{ik})\right\}.$$

Note that this criterion does not involve the covariance matrix Γ of $\hat{\gamma}$. In most applications, this is an advantage since Γ is unknown and a nuisance. In our context, Γ is known under Gaussian assumptions, and the failure of the Rice and Silverman method to exploit this knowledge might be seen as a disadvantage. However, in our view, this is counterbalanced by its avoidance of Gaussian assumptions since the unbiasedness of v_{ijk} for $\gamma(t_{ij}, t_{ik})$ does not rely on Gaussianity.

3.4 *Summary*

Our approach is to choose the bandwidths for variance and variogram smoothing separately by minimizing the cross-validation criterion. Using the chosen bandwidths, we smooth the squared residual and the variogram cloud of the squared residuals. These two smooth components are combined using (4), that is,

$$G(s,t) = \begin{cases} \frac{1}{2}\left\{\sigma^2(s) + \sigma^2(t)\right\} - \gamma(s,t), & s \neq t \\ \sigma^2(t), & s = t, \end{cases}$$

to form the smoothed estimate of the covariance function.

4. Examples

4.1 *Babies' Blood Pressure Data*

For these data, there are no experimental treatments, as the aim of the study was to describe the change in blood pressure over time. Therefore, all 43 babies are treated as a single group. The mean response was modelled by a KWLLR, using a quartic kernel with bandwidth $h = 1.58$ chosen according to Rice and Silverman's (1991) cross-validation method.

The methodology described in Section 3 was applied to obtain smooth, nonparametric estimates of the variance and variogram functions. The quartic kernel is used in this and the following example.

The cross-validation criterion for the variances was minimized at $h = 8.9$; that for the variogram was minimized at $h = 7.75$. These bandwidths were used to obtain the results discussed below. All figures in this and the following section were prepared by estimation on a regular grid of values of the time variable.

Figure 3 is a plot of squared residuals (the ordinate is truncated to provide a reasonable plot) together with the KWLLR estimate (solid line). The smooth change in variation is evident. The ratio of largest to smallest variance is approximately four.

Figure 4 is the smoothed variogram using KWLLR. The nonstationarity is evident, for the surface changes for lags of equal size. Thus, the smoothing has enabled this aspect to be checked. Our method constrains this nonstationarity to be smoothly changing, which we feel should be a reasonable reflection of underlying changes in blood pressure in the absence of abrupt clinical interventions.

The implied estimate of the covariance function is given in Figure 5. Notice that the covariance structure is clearly nonstationary, as was evident from the smoothed variogram. Furthermore, this implies the correlation structure is nonstationary. Note also that there is considerable measurement error (the nugget effect in geostatistics). Thus, the nature of the underlying structure is revealed.

In a parametric modelling approach, the estimated covariance structure would now be used to calculate confidence or probability bands for the mean profile, which is of prime interest in this application. This can be done using our smoothed estimate, and approximate pointwise 95% prediction or probability bands are included in Figure 6. These bands are calculated ignoring the bias in estimation of the mean curve and the effect of estimation of the covariance structure.

4.2 *CD4 Data*

Our second example is a set of data previously analysed in Zeger and Diggle (1994, henceforth ZD). The data consist of a total of 2376 measurements of the CD4 cell count taken at approximately 6-month intervals on a group of 369 gay and bisexual men taken from the Multicentre Aids Cohort Study. The subjects were those whose HIV status changed from seronegative to seropositive during the study, and time is measured in months relative to the estimated date of seroconversion. The objective is to investigate the population-averaged and subject-specific time evolution of CD4 cell depletion, after adjusting for the effects of known covariates. As in our first example, the data are unbalanced, with the number of measurements per subject varying from 1 to 12. Furthermore, in this example, the precise sequences of measurement times are different for the 369 subjects.

Following ZD, we first subtracted from the square-root-transformed counts a set of mean responses that consist of a nonparametric time trend estimated by a kernel method plus linear covariate adjustments for cigarette smoking (number of packs per day), recreational drug use (yes/no), number of sexual partners, and a measure of depressive symptoms (see equation (3)). ZD proceeded to fit a stationary, parametric model to the covariance structure of the residual series.

We were forced to use an approximation for the calculation of the cross-validation criterion for the variogram because of the size of the data set. Time to seroconversion was rounded to the nearest 1/10 of a year, reducing the computation to a manageable level. Figures 7 and 8 show our smoothed estimates of the variance and variogram functions, respectively; bandwidths here were 3 for the variances and 2.45 for the variogram.

The smoothed variances range from 27 to 50 so that, while the variances show some increasing trend, it is not severe. The smoothed variogram shows stability for time pairs of equal lags, so stationarity might be a reasonable approximation. The anomalous behaviour at the extreme time pairs appears to be an artifact of sparseness of data and occurs with kernel smoothing as well as KWLLR. The smooth estimate of the covariance structure is given in Figure 9. Substantial measurement error is evident, and there is some indication that the variance increases over time. However, a stationary covariance structure as in ZD is likely to provide a reasonable model. Thus, the smoothing approach provides a diagnostic for parametric modelling of the covariance structure.

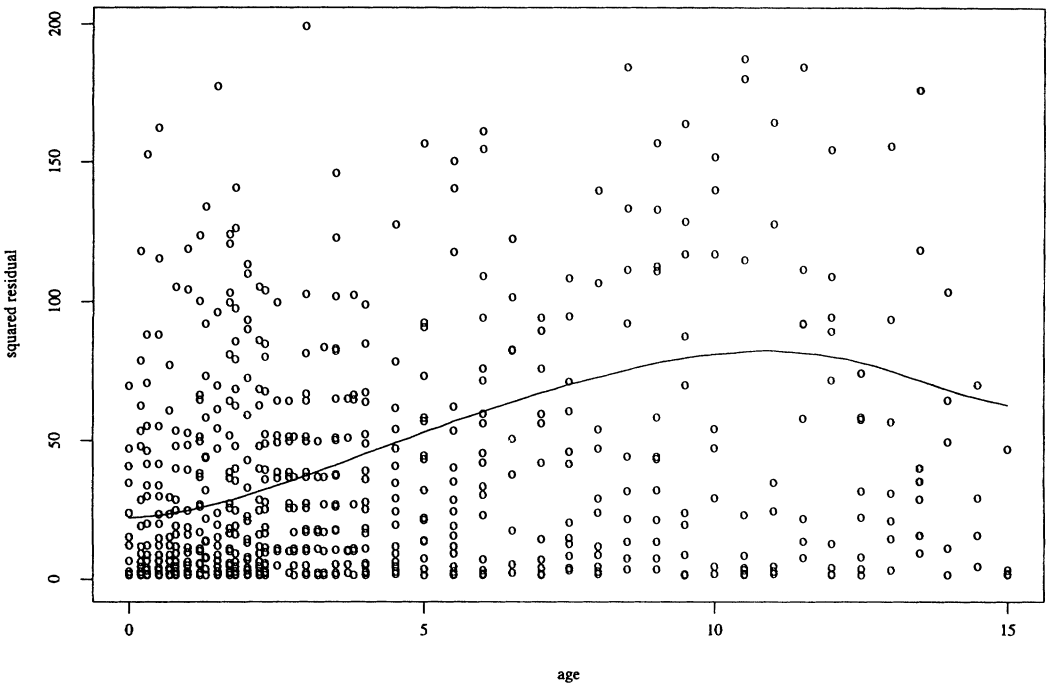


Figure 3. Squared residuals (the extreme values have been omitted for clarity of presentation) and the smooth estimate of the variance over time for the babies' blood pressure data.

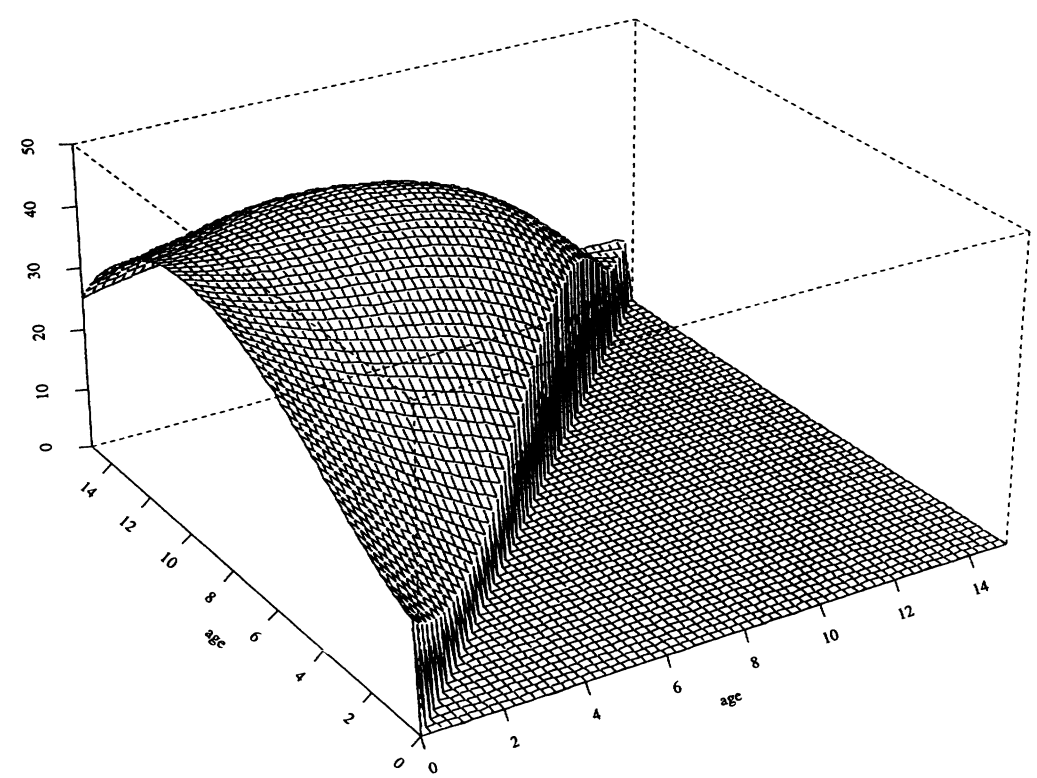


Figure 4. The kernel weighted local linear regression estimate of the variogram for the babies' blood pressure data.

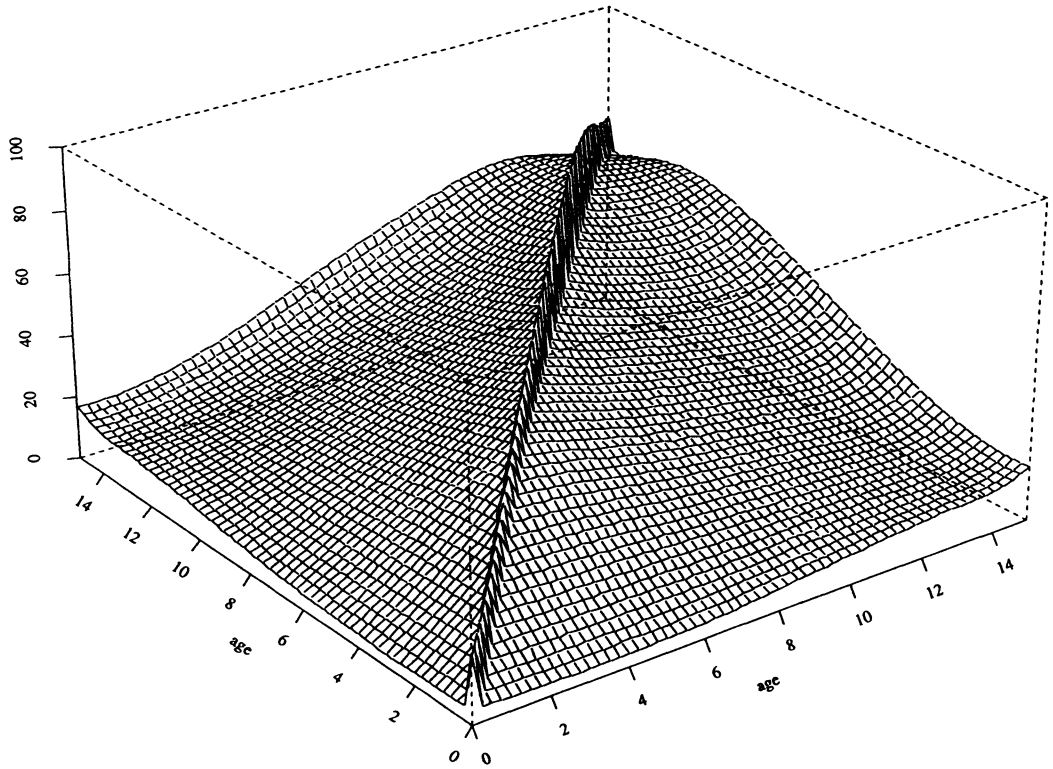


Figure 5. The estimate of covariance function for the babies' blood pressure data, found by combining the smooth variance and the smooth variogram estimates.

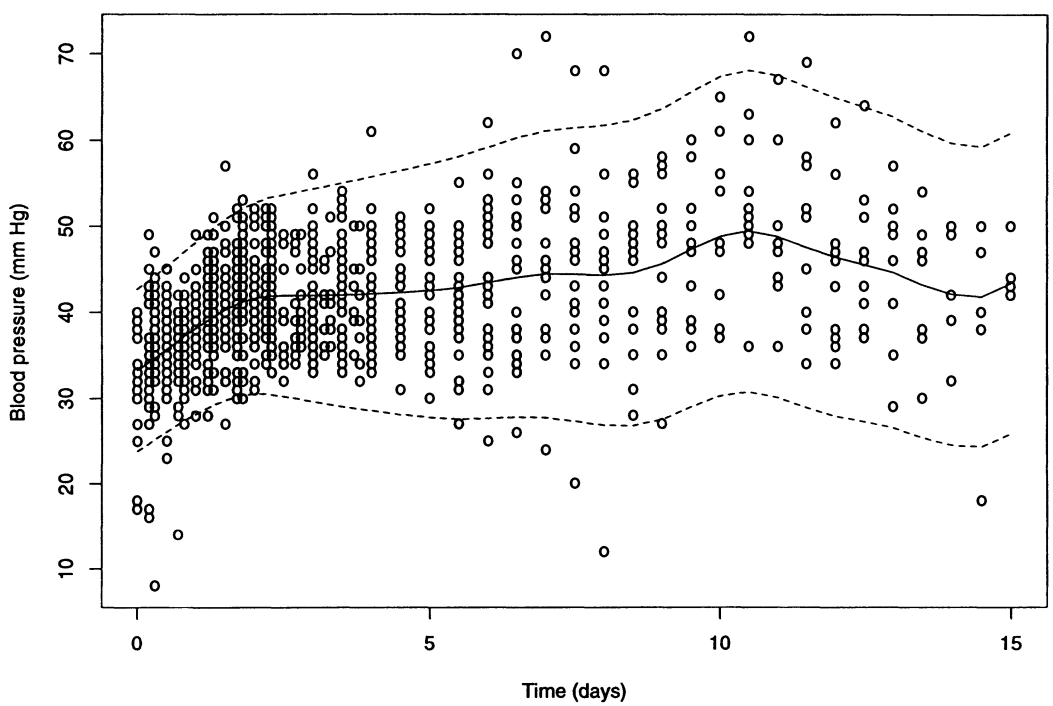


Figure 6. Smooth estimate of mean blood pressure over time and approximate 95% confidence bands using the smoothed estimate of the covariance function.

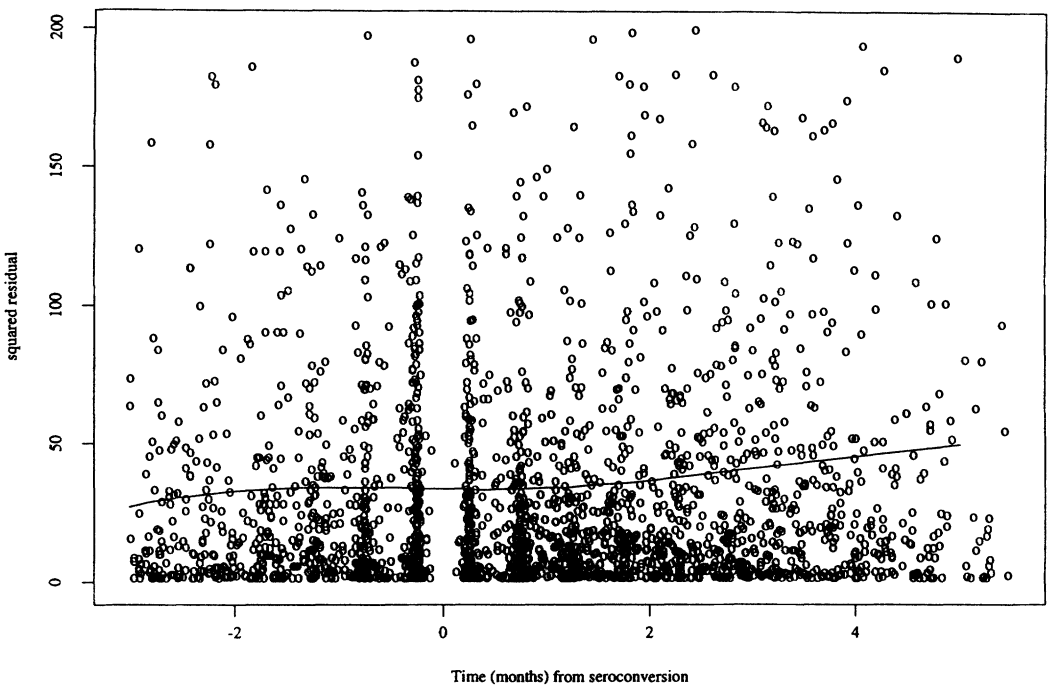


Figure 7. The squared residuals and the smooth estimate of the variance over time for the square-root-transformed CD4 data.

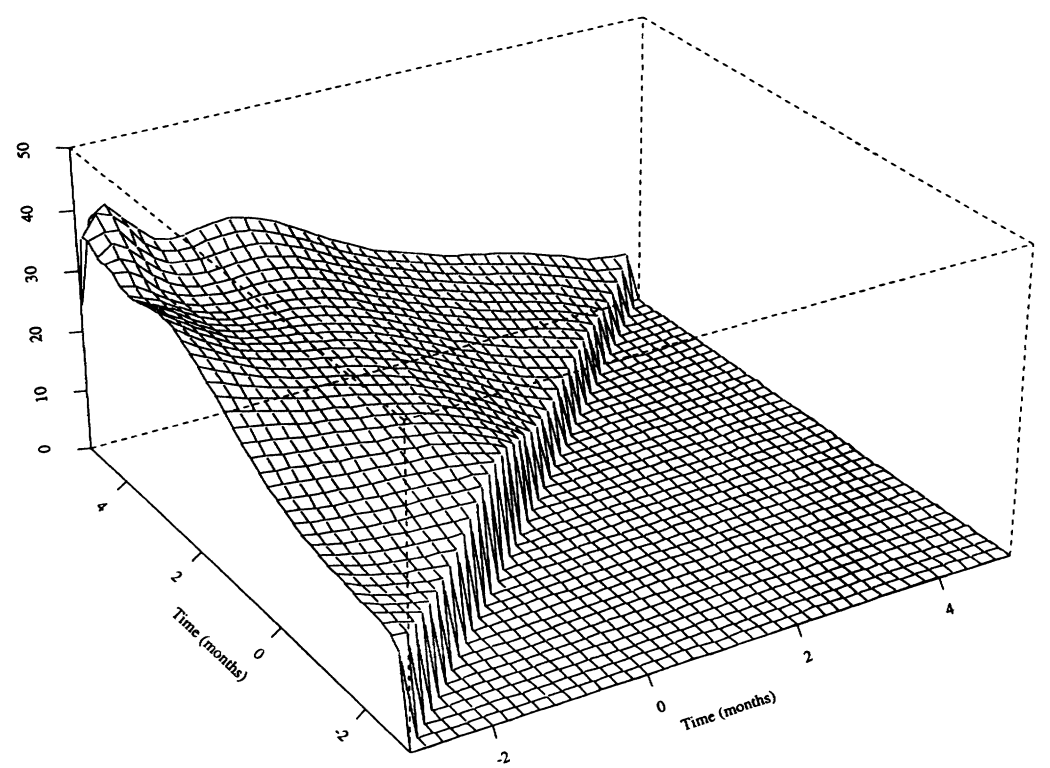


Figure 8. The estimate of the variogram found by kernel weighted local linear regression for the square-root-transformed CD4 data.

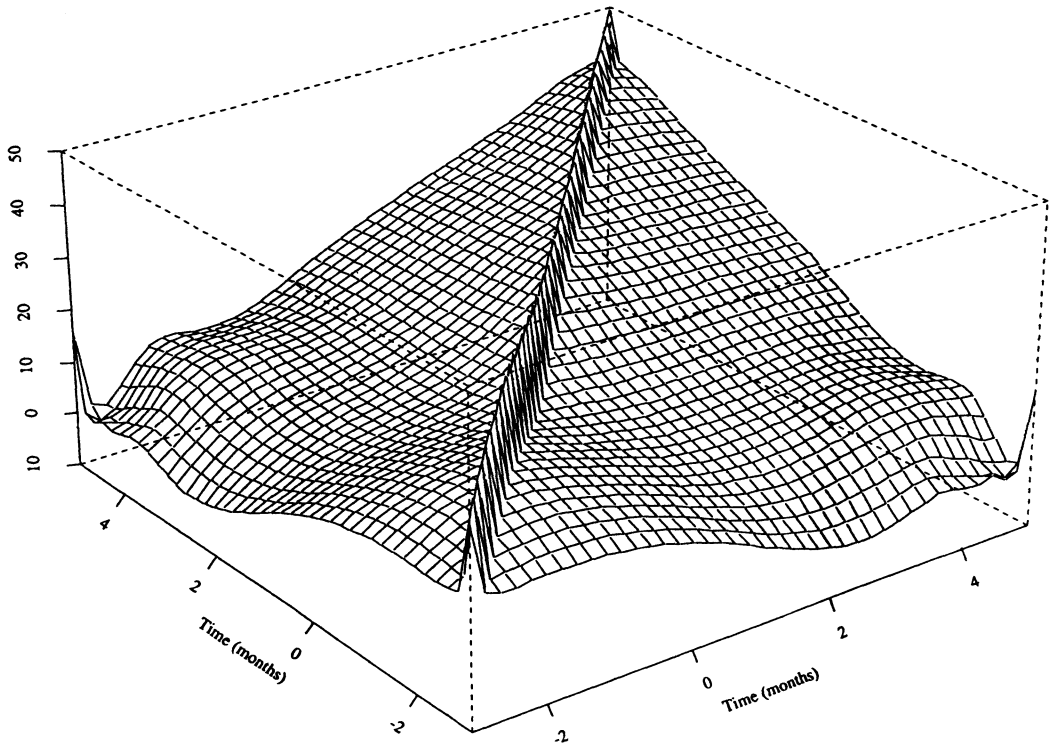


Figure 9. The estimate of the covariance function for the square-root-transformed CD4 data found by combining the smooth estimates of the variance function and the variogram.

5. Discussion

The examples indicate that the approach presented is useful as a diagnostic tool. There are, however, a number of aspects that warrant further discussion.

With regard to the choice of kernel function, we could replace the Euclidean metric in (11) by a general positive-definite quadratic form, but we would then need to specify the values of three bandwidth parameters, equivalent to specifying the covariance matrix of a bivariate distribution. A possible compromise would be to use a kernel whose contours are ellipses with principal axes oriented at 45 degrees to the horizontal to acknowledge that both axes represent the same physical quantity while permitting more smoothing along subdiagonals of the (s,t)-space. This would be a sensible strategy for approximately stationary covariance structures.

There appears to be a logical inconsistency in how the variances are smoothed in comparison to the variogram. There are two issues. First, the amount of smoothing for each component may differ substantially. We do not see this as a large problem unless either component is undersmoothed. The second point concerns how the variance smoothing is carried out. In the standard approach (Diggle, 1988), the process variance is estimated using squared residual differences

$$v^*_{iji'k} = \frac{1}{2}\{z_{ij} - z_{i'k}\}^2, \tag{13}$$

where $i \neq i'$, so that a more self-consistent approach might be to define a function $\sigma^*(s,t) = (1/2)\{\sigma^2(s) + \sigma^2(t)\}$ and to estimate $\sigma^*(s,t)$ by smoothing (13). It would then be reasonable to minimise a single cross-validation criterion incorporating both the v_{ijk} and the $v^*_{iji'k}$. This is computationally expensive, and an alternative might be to smooth $\sigma^*(s,t)$ using the bandwidth found by minimizing the cross-validation criterion for the variogram. However, in a limited comparison, the estimate of the variance of the stationary process using $v_{iji'k}$ was very similar to the estimate found using squared residuals.

The estimate of covariance structure implied by the relationship $G(s,t) = \sigma^*(s,t) - \gamma(s,t)$ is not guaranteed to be positive-definite. In fact, for both our examples, a single eigenvalue of the estimated covariance matrix was negative; it appears that this is a result of a boundary problem, as any submatrix is positive-definite in both examples. This may not matter if we are using the

variogram simply as an exploratory tool, but if we want to construct an estimated covariance matrix for making inferences about the mean response profiles as an alternative to the sample covariance matrix used in Liang and Zeger (1986) and Zeger and Liang (1986), positive-definiteness (or at least nonnegative-definiteness) is essential. Hall et al. (1994) use Fourier transform methods to construct a positive-semidefinite estimate, while Glasbey (1988) uses a local smoothing of products of residuals, which ensures a positive-semidefinite estimate; this needs to be examined further.

There is a need to study the impact of these methods on inference for the mean profiles. It is known that the impact of parametric models of covariance structure on point estimates of the mean profiles can be small if the parametric model is a reasonable fit, but that the standard errors are generally sensitive to the fitted model. We would conjecture that the same is true with nonparametric models of the covariance structure; thus, oversmoothing may give misleading standard errors. We have not investigated these or other implications, for example, in testing hypotheses concerning the mean profiles.

Acknowledgements

This work was supported by grants from the Australian Research Council and the University of Adelaide Research Scheme. The data on babies' blood pressure were kindly provided by Sally Hollis, Lancaster University. The CD4 data were collected by the Multicenter AIDS Cohort Study (MACS) with centers at the Johns Hopkins School of Public Health (Alfred Saah, Alvaro Munoz, Joseph Margolick), Howard Brown Health Center and Northwestern University Medical School (John Phair), University of California, Los Angeles (Roger Detels, Janis V. Giorgi), and the University of Pittsburgh (Charles Rinaldo). The MACS is funded by the National Institute of Allergy and Infectious Diseases, with additional supplemental funding from National Cancer Institute grants UO1-AI-35042, 5-M01-RR-00722 (GCRC), UO1-AI-35043, UO1-AI-37984, UO1-AI-35039, UO1-AI-35040, UO1-AI-37613, UO1-AI-35041. We thank David Smith for incorporating some suggestions into the Oswald software, thereby simplifying the computational aspects of the analyses.

Résumé

Dans les études longitudinales, en général, on s'intéresse à l'effet au cours du temps des différents traitements. Cependant, les observations effectuées sur un même sujet sont le plus souvent corrélées et l'analyse doit prendre en compte la structure de covariance sous-jacente. Un estimateur non paramétrique de la structure de covariance peut servir, soit comme aide à la formulation d'un modèle paramétrique, soit comme support à des inférences formelles sans présupposés paramétriques. Lorsque les données proviennent d'une brève séquence de mesures, effectuées aux mêmes temps, sur tous les sujets d'un échantillon de grande taille, on peut utiliser la matrice de covariance empirique, mais ce n'est pas le cas si les données sont très déséquilibrées ou si les séries de mesures sont longues par rapport au nombre de sujets. On a proposé d'utiliser comme estimateur non paramétrique de la structure de covariance, le variogramme des résidus, supposés stationnaires, d'un modèle saturé de l'espérance des observations. Pour obtenir un estimateur non paramétrique de la structure de covariance sans supposer la stationarité, nous considérons une régression linéaire, lissée par noyau localement pondéré, sur les ordonnées du variogramme et sur le carré des résidus. Deux applications illustrent l'intérêt de cet estimateur en tant qu'outil diagnostic; l'une concerne la mesure de la tension artérielle de bébés en unité de soins intensifs; l'autre l'évolution du nombre de cellules CD4 chez des séropositifs VIH. L'utilisation de cet estimateur pour des inférences statistiques plus formelles sur l'évolution des moyennes reste à étudier.

References

Altham, P. M. E. (1984). Improving the precision of estimation by fitting a model. *Journal of the Royal Statistical Society, Series B* **46,** 118–119.

Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *Journal of the American Statistical Association* **85,** 749–759.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association* **74,** 829–836.

Cullis, B. R. and McGilchrist, C. A. (1990). A model for the analysis of growth data from designed experiments. *Biometrics* **46,** 131–142.

Diggle, P. J. (1988). An approach to the analysis of repeated measures. *Biometrics* **44,** 959–971.

Diggle, P. J. and Hutchinson, M. J. (1988). On spline smoothing with autocorrelated errors. *Australian Journal of Statistics* **31,** 194–199.

Diggle, P. J., Liang, K.-Y., and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Oxford University Press.

Glasbey, C. A. (1988). Standard errors resilient to error variance misspecification. *Biometrika* **75,** 201–206.

Hall, P. and Patil, P. (1994). Properties of nonparametric estimators of autocovariance for stationary random fields. *Probability Theory and Related Fields* **99,** 399–424.

Hall, P., Fisher, N. I., and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *Annals of Statistics* **22,** 2115–2134.

Härdle, W. (1990). *Applied Nonparametric Regression.* New York: Cambridge University Press.

Hart, J. D. (1991). Kernel regression estimation with time series errors. *Journal of the Royal Statistical Society, Series B* **53,** 173–187.

Hart, J. D. and Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association* **81,** 1080–1088.

Hastie, T. J. and Loader, C. (1993). Locally weighted regression (with discussion). *Statistical Science* **8,** 120–143.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized Additive Models.* New York: Chapman and Hall.

Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing* **3,** 135–146.

Kenward, M. G. (1987). A method for comparing profiles of repeated measurements. *Applied Statistics* **36,** 296–308.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38,** 963–974.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73,** 13–22.

Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Matheron, G. (1971). *The Theory of Regionalised Variables and Its Applications.* Paris: École Nationale Supérieure des Mines de Paris.

Moyeed, R. A. and Diggle, P. J. (1994). Rates of convergence in semi-parametric modelling of longitudinal data. *Australian Journal of Statistics* **36,** 75–93.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B* **53,** 233–243.

Sampson, P. D. and Guttorp, P. (1992). Nonparametric representation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* **87,** 108–119.

Shapiro, A. and Botha, J. D. (1991). Variogram fitting with a general class of conditionally nonnegative definite functions. *Computational Statistics and Data Analysis* **11,** 87–96.

Silverman, B. W. (1985). Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *Journal of the Royal Statistical Society, Series B* **47,** 1–52.

Zeger, S. L. and Diggle, P. J. (1994). Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. *Biometrics* **50,** 689–699.

Zeger, S. L. and Liang, K.-Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* **42,** 121–130.

Received April 1995; *revised February and August* 1997; *accepted August* 1997.