Snee, R.D. (1982). Nonadditivity in a two-way classification: Is it interaction or nonhomogeneous variance? *J. Amer. Statist. Assoc.*, **77**, 515–518.

Yochmowitz, M.G., and Cornell, R.T. (1978). Stepwise tests for multiplicative components of interaction. *Technometrics*, **20**, 79–84.

Received February 1992; Revised April 1993.

COMMUN. STATIST.—SIMULA., 22(4), 1079-1106 (1993)

COVARIANCE STRUCTURE SELECTION IN GENERAL MIXED MODELS

Russ Wolfinger

SAS Institute, Inc., SAS Campus Drive,
Cary, North Carolina 27513-2414, U.S.A.

Key Words and Phrases: random effects; repeated measures; REML; variance modeling

ABSTRACT

This article describes a unified approach to variance modeling and inference in the context of a general form of the normal-theory linear mixed model. The primary variance-modeling objects are parameterized covariance structures, examples being diagonal, compound-symmetry, unstructured, time-series, and spatial. These structures can enter in two different places in the general mixed model, and the combination of one or both of these places with the variety of structures provides a rich class of variance models. The approach is likelihood-based, and involves the use of both maximum likelihood and restricted maximum likelihood. Two examples provide illustration.

1 INTRODUCTION

1.1 The Model

The general linear mixed model (Harville, 1977; Laird and Ware, 1982) is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\upsilon + \epsilon$$

where $\mathbf{y}$ denotes observed data in a vector of length n, $\boldsymbol{\beta}$ is an unknown vector of fixed effects with known design matrix $\mathbf{X}$, υ is an unknown vector of random effects with known design matrix $\mathbf{Z}$, and ϵ is an unobserved error vector. υ and ϵ have normal (Gaussian) distributions with $\mathbf{0}$ expectations and

$$Var\begin{bmatrix} \upsilon \\ \epsilon \end{bmatrix} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}$$

where both $\mathbf{G}$ and $\mathbf{R}$ are nonsingular. As such, the variance of $\mathbf{y}$ is

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

This model can also be constructed in two stages, where in the first stage $\boldsymbol{\beta}$ and υ are fixed unknown parameters and ϵ is $N(\mathbf{0}, \mathbf{R})$, that is, is normally distributed with mean $\mathbf{0}$ and variance $\mathbf{R}$. In the second stage, υ is $N(\mathbf{0}, \mathbf{G})$ and independent of ϵ, and $\boldsymbol{\beta}$ is still fixed, or, equivalently, possesses a flat improper prior.

1.2 Covariance Structures

As in the usual linear model, the expected value of $\mathbf{y}$ is modeled through the fixed effects $\boldsymbol{\beta}$. The extension provided by the mixed model is that the variance of $\mathbf{y}$ can be modeled through $\mathbf{Z}$, $\mathbf{G}$, and $\mathbf{R}$.

The classical mixed model has $\mathbf{R} = \sigma^2\mathbf{I}$, where $\mathbf{I}$ is the $n \times n$ identity matrix, and $\mathbf{G}$ is a diagonal matrix containing variance components. This

Table 1: *Covariance structure notation and examples*

Structure	Notation	Example
Diagonal	I	$\begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \sigma^2 & \\ & & & \sigma^2 \end{bmatrix}$
Variance Components	VC(A B)	$\begin{bmatrix} \sigma_A^2 & & & \\ & \sigma_A^2 & & \\ & & \sigma_B^2 & \\ & & & \sigma_B^2 \end{bmatrix}$
Compound Symmetry	CS	$\begin{bmatrix} \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ & \sigma^2+\sigma_1 & \sigma_1 & \sigma_1 \\ & & \sigma^2+\sigma_1 & \sigma_1 \\ & & & \sigma^2+\sigma_1 \end{bmatrix}$
Unstructured	UN	$\begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ & & \sigma_{33} & \sigma_{34} \\ & & & \sigma_{44} \end{bmatrix}$
First-order Autoregressive	AR(1)	$\sigma^2\begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix}$
Banded	UN(2)	$\begin{bmatrix} \sigma_1 & \sigma_5 & & \\ & \sigma_2 & \sigma_6 & \\ & & \sigma_3 & \sigma_7 \\ & & & \sigma_4 \end{bmatrix}$
Toeplitz	TOEP	$\begin{bmatrix} \sigma_1 & \sigma_2 & \sigma_3 & \sigma_4 \\ & \sigma_1 & \sigma_2 & \sigma_3 \\ & & \sigma_1 & \sigma_2 \\ & & & \sigma_1 \end{bmatrix}$
Banded Toeplitz	TOEP(2)	$\begin{bmatrix} \sigma_1 & \sigma_2 & & \\ & \sigma_1 & \sigma_2 & \\ & & \sigma_1 & \sigma_2 \\ & & & \sigma_1 \end{bmatrix}$

All matrices are symmetric, so only the upper triangle is printed. Blanks in the upper triangle represent 0s, and Greek letters represent unknown parameters. ρ is the autoregressive parameter, and must satisfy $|\rho| < 1$ for stationarity.

Table 2: *Covariance structure notation and examples*

Structure	Notation	Example
Compound Symmetry with Heterogeneous Groups	CS*GROUP	$\begin{bmatrix} \sigma_1 & \sigma_2 & & \\ & \sigma_1 & & \\ & & \sigma_3 & \sigma_4 \\ & & & \sigma_3 \end{bmatrix}$
AR(1) plus Diagonal	AR(1)+I	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} + \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_1^2 & & \\ & & \sigma_1^2 & \\ & & & \sigma_1^2 \end{bmatrix}$
AR(1) plus Common Covariance	AR(1)+J	$\sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ & 1 & \rho & \rho^2 \\ & & 1 & \rho \\ & & & 1 \end{bmatrix} + \begin{bmatrix} \sigma_1 & \sigma_1 & \sigma_1 & \sigma_1 \\ & \sigma_1 & \sigma_1 & \sigma_1 \\ & & \sigma_1 & \sigma_1 \\ & & & \sigma_1 \end{bmatrix}$
Spatial Power Law	SP(POW)	$\sigma^2 \begin{bmatrix} 1 & \rho^{d_{12}} & \rho^{d_{13}} & \rho^{d_{14}} \\ & 1 & \rho^{d_{23}} & \rho^{d_{24}} \\ & & 1 & \rho^{d_{34}} \\ & & & 1 \end{bmatrix}$

All matrices are symmetric, so only the upper triangle is printed. Blanks in the the upper triangle represent 0s, and Greek letters represent unknown parameters. ρ is the autoregressive parameter, and must satisfy $|\rho| < 1$ for stationarity. d_{ij} is the Euclidean distance between the ith and jth spatial coordinates.

model is very useful, especially with randomized block and split-plot designs (Harville, 1977; Searle, Casella, and McCulloch, 1992). However, the classical mixed model is only a special case of the general mixed model that permits arbitrary parameterized covariance structures in both **G** and **R**.

Tables 1 and 2 picture several useful covariance structures for **G** and **R**. The most common ones are in Table 1, and include diagonal, compound symmetry, unstructured, and AR(1). Some confusion appears to exist in the literature concerning the terms "banded" and "Toeplitz." Note that a banded matrix is really an unstructured one with 0s in higher diagonals, whereas a Toeplitz matrix has the same parameters down each diagonal.

The Toeplitz structure can be viewed as a moving average (MA) structure of order equal to the size of the matrix, and a banded-Toeplitz matrix corresponds to MA structures of lower order. The pictured parameterization for the banded-Toeplitz matrix is a simple one, and does not reflect the usual stationarity constraints (Fuller, 1976).

Table 2 pictures some more complex structures. The first one specifies heterogeneous groups for the compound symmetry structure, and note that this type of heterogeneity can be used for any of the structures pictured in Table 1. The AR(1) structure can be generalized by adding either a local error, also known as a nugget effect, or a common covariance. The spatial power law is a different generalization of AR(1) that includes unequally-spaced time series and data from 2- or higher-dimensional spaces. See Zimmerman and Harville (1991) and Cressie (1991) for several other spatial structures.

Some additional structures not pictured in Tables 1 and 2 are autoregressive moving average (ARMA; Rochon, 1992), factor-analytic (Jennrich and Schluchter, 1986), general linear (Helms and Edwards, 1991), power of the mean (Carroll and Ruppert, 1988), and log variance (Harvey, 1976; Aitkin, 1987).

1.3 Relation to the Literature

Laird and Ware (1982) are some of the earliest analysts to consider a practical application of the general formulation of **G** and **R**. They consider longitudinal data, in which time-dependent observations are taken on a number of different subjects. They write the mixed model for the ith subject as

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\boldsymbol{v}_i + \boldsymbol{\epsilon}_i$$

This can be written in the general form above by stacking $\mathbf{y}_i$, $\mathbf{X}_i$, υ_i, and ϵ_i into $\mathbf{y}$, $\mathbf{X}$, υ, and ϵ, respectively, and setting

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{Z}_s \end{bmatrix}$$

where s is the number of subjects.

Laird and Ware (1982) present an EM algorithm for computing maximum likelihood (ML) and restricted/residual maximum likelihood (REML) estimates of $\boldsymbol{\beta}$ and the unknown parameters in $\mathbf{G}$ and $\mathbf{R}$. Lindstrom and Bates (1988) provide detailed formulas for implementing a Newton-Raphson algorithm, and give reasons for preferring it to EM. Although these authors are mindful of a general formulation of $\mathbf{R}$, their applications involve setting $\mathbf{R} = \sigma^2\mathbf{I}$ and using random effects to model variability.

In contrast, Liang and Zeger (1986), in the more general setting of generalized linear models, model all variability in $\mathbf{R}$, which is assumed to be block-diagonal with blocks $\mathbf{R}_i$ corresponding to observations from the same subject. They recommend considering a collection of covariance structures like those pictured in Tables 1 and 2 and choosing one to model the $\mathbf{R}_i$. Their estimation method consists of iterating between generalized-estimating-equation (GEE) estimates of $\boldsymbol{\beta}$ and a Pearson-residual-moment estimates of the parameters in the selected structure. This procedure reduces to ML for the general Gaussian linear mixed model.

Jennrich and Schluchter (1986) elaborate on the work of Liang and Zeger for the Gaussian linear case, and apply it to the growth-curve data of Pothoff and Roy (1964). They also present Newton-Raphson and EM algorithms for

computing estimates using ML and REML. Both methods are implemented in BMDP-5V (Schluchter, 1988).

An appealing feature of a likelihood-based covariance-structure method for repeated measures is that it in a sense subsumes both the univariate and multivariate approaches traditionally used for repeated measures. This fact is seen by noting that the compound-symmetry structure corresponds to the univariate approach and the unstructured matrix corresponds to the multivariate approach. Furthermore, missing data pose no theoretical problems to a likelihood-based method, as it is easy to construct the likelihood of only those data that are observed. In the traditional methods, one must either throw away a subject's data that has missing values or devise some imputation method that is often complex. Another bonus of a covariance structure method is the possibility of using time series and heterogeneous structures in addition to the compound-symmetry and unstructured ones.

The general approach proposed in this article combines the abovementioned approaches into one unified framework. This general setup of the mixed model encompasses many common statistical methods, including those associated with random effects, repeated measures, split-plot and incomplete-block designs, random coefficients, and best linear unbiased predictors (BLUPs). One can even consider both correlated random effects and repeated measurements in the same model. ML and REML can be used in the general setting, in which case optimization is carried out over all of the unknown parameters in $\mathbf{G}$ and $\mathbf{R}$. The next section presents some of the important details of this estimation procedure, as well as methods for model selection and inference.

Section 3 illustrates the general mixed model approach with two examples; however, the literature already contains some specific analyses involving nontrivial models for $\mathbf{G}$ and $\mathbf{R}$. Several examples from Henderson (1984) contain multiple trait models with $\mathbf{G}$ and $\mathbf{R}$ known. Chi and Reinsel (1988)

and Jones and Boadi-Boateng (1991) consider models containing subject-specific random effects in v and AR(1) serial correlation in ϵ. Laird and Ware (1982) and Lindstrom and Bates (1988) also mention models of this type. Sanders (1989) provides a teacher-student example in which both $\mathbf{G}$ and $\mathbf{R}$ are unstructured. Stroup (1989) presents a line-source sprinkler irrigation model containing random block effects in v and spherically-spatial correlation in ϵ. The model of Diggle (1988) contains variance components in $\mathbf{G}$ and exponentially-spatial correlation in $\mathbf{R}$. Finally, Schaalje et al. (1991) consider a variety of structures, including the combination of compound-symmetry and AR(1) described in Pantula and Pollock (1985).

2 ANALYSIS

2.1 Estimation

Given particular covariance structures for $\mathbf{G}$ and $\mathbf{R}$, let $\boldsymbol{\theta}$ denote the vector of parameters in $\mathbf{V}$, and assume it is of length q. Usually $\boldsymbol{\theta}$ is unknown, in which case a normal-theory maximum-likelihood approach may be used to estimate it (Harville, 1977; Dempster et al., 1984; Hocking, 1985; Searle, Casella, and McCulloch, 1992). This approach is preferred to the traditional ANOVA method (Searle, 1988).

Minus 2 times the Gaussian log likelihood for the general linear mixed model is

$$-2ll(\boldsymbol{\beta},\boldsymbol{\theta}|\mathrm{y}) = \log|\mathrm{V}(\boldsymbol{\theta})| + (\mathrm{y}-\mathbf{X}\boldsymbol{\beta})'\mathrm{V}^{-1}(\boldsymbol{\theta})(\mathrm{y}-\mathbf{X}\boldsymbol{\beta}) + n\log 2\pi$$

Minimizing this expression analytically for $\boldsymbol{\beta}$ yields

$$\mathrm{b}(\boldsymbol{\theta}) = [\mathrm{X}'\mathrm{V}^{-1}(\boldsymbol{\theta})\mathrm{X}]^{-1}\mathrm{X}'\mathrm{V}^{-1}(\boldsymbol{\theta})\mathrm{y}$$

and substitution into the orginal equation produces minus 2 times the profile/concentrated log likelihood:

$$-2ll(\boldsymbol{\theta}|\mathrm{y}) = \log|\mathrm{V}(\boldsymbol{\theta})| + [\mathrm{y}-\mathbf{X}\mathrm{b}(\boldsymbol{\theta})]'\mathrm{V}^{-1}(\boldsymbol{\theta})[\mathrm{y}-\mathbf{X}\mathrm{b}(\boldsymbol{\theta})] + n\log 2\pi$$

The formula for minus 2 times the restricted/residual log likelihood is

$$-2ll_R(\boldsymbol{\theta}|\mathrm{y}) = \log|\mathrm{V}(\boldsymbol{\theta})| + [\mathrm{y}-\mathbf{X}\mathrm{b}(\boldsymbol{\theta})]'\mathrm{V}^{-1}(\boldsymbol{\theta})[\mathrm{y}-\mathbf{X}\mathrm{b}(\boldsymbol{\theta})] + \log|\mathrm{X}'\mathrm{V}^{-1}(\boldsymbol{\theta})\mathrm{X}| + (n-p)\log 2\pi$$

where p is the rank of $\mathbf{X}$. The final two expressions above are the objective functions for ML and REML.

A Newton-Raphson optimization algorithm is preferred to EM for reasons given by Lindstrom and Bates (1988). Additional details for carrying out a Newton-Raphson optimization of the general mixed model are in Wolfinger, Tobias, and Sall (1992).

2.2 Model Selection

The three-stage procedure of Diggle (1988) serves as a suitable foundation for model selection, and this section extends it to the general mixed model using both ML and REML.

The first stage involves an appropriate data transformation and the selection of the fixed effects. Overfitted models are preferred to underfitted ones at this stage to avoid the introduction of spurious correlations.

The second stage consists of the initial selection of covariance structures for $\mathbf{G}$ and $\mathbf{R}$ by using any relevant scientific theory as well as residual plots and semi-variograms. The statistics to consider at this stage are

$$\hat{v} = \hat{\mathbf{G}}\mathrm{Z}'\hat{\mathrm{V}}^{-1}(\mathrm{y}-\mathbf{X}\hat{\boldsymbol{\beta}})$$

$$\hat{\epsilon} = \mathrm{y}-\mathbf{X}\hat{\boldsymbol{\beta}}-\mathrm{Z}\hat{v}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$ and $\hat{\mathbf{G}}$, $\hat{\mathbf{V}}$ are REML estimates. $\hat{v}$ contains empirical BLUPs (EBLUPs), and can be used to select $\mathbf{G}$. The residuals $\hat{\boldsymbol{\epsilon}}$ can be used to select $\mathbf{R}$. Both the EBLUPs and residuals can also be used as diagnostics to check, say, the normality assumptions.

The third stage uses formal statistical techniques to compare various covariance structures. One possibility is to use twice the inverse of the final second derivative matrix, that is, the observed Fisher information matrix, to construct Wald χ^2-tests of the covariance parameters. However, the Wald tests can be unreliable in small samples. Therefore it is better, at least for nested models, to construct likelihood ratio tests (LRTs) by comparing differences of -2 log likelihoods with the χ^2-distribution (Jennrich and Schluchter, 1986; Schaalje et al., 1991). Furthermore, both analytic (Reid, 1988) and resampling (Zhang, Pantula, and Boos, 1991) adjustments of this technique are possible.

Another formal model selection procedure for covariance structures is to minimize REML information criteria such as the following:

$$\begin{aligned}
\text{AIC}_R &= -2ll_R + 2q \\
\text{HQC}_R &= -2ll_R + 2q\log\log(n-p) \\
\text{SBC}_R &= -2ll_R + q\log(n-p) \\
\text{CAIC}_R &= -2ll_R + q(\log(n-p)+1)
\end{aligned}$$

where recall q is the number of covariance parameters, n is the number of observations, and p is the rank of X. The criteria are ordered in increasing preference for parsimony.

A final step is to reduce the fixed-effects specification, but because REML eliminates the fixed effects, it cannot be used to compare them. Hence one should switch to the ML versions of LRT and information criteria, the latter being:

$$\begin{aligned}
\text{AIC} &= -2ll + 2(p+q) \\
\text{HQC} &= -2ll + 2(p+q)\log\log n \\
\text{SBC} &= -2ll + (p+q)\log n \\
\text{CAIC} &= -2ll + (p+q)(\log n + 1)
\end{aligned}$$

For a review, see Bozdogan (1987).

To summarize, the first stage selects an overfitted fixed-effects model using ML. Next, covariance structures for $\mathbf{G}$ and $\mathbf{R}$ are selected informally using EBLUPs and residuals and formally using REML LRTs and/or information criteria. Then ML is used to select the final fixed-effects model. REML can then be used to fit the final model used for inference. There are obvious overlaps among the stages, and several iterations of the entire procedure may be necessary in some cases.

2.3 Inference

Given ML or REML estimates, $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, $\boldsymbol{\beta}$ and v are estimated by solving the mixed model equations (see Henderson, 1984):

$$\begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

These solutions can be written as

$$\begin{aligned}
\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{v} \end{bmatrix} &= \begin{bmatrix} (\mathbf{X}'\hat{V}^{-1}\mathbf{X})^{-}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y} \\ \hat{\mathbf{G}}\mathbf{Z}'\hat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{bmatrix} \\
&= \hat{\mathbf{C}} \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}
\end{aligned}$$

where

$$\hat{\mathbf{C}} = \begin{bmatrix} \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\hat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\hat{\mathbf{R}}^{-1}\mathbf{Z} + \hat{\mathbf{G}}^{-1} \end{bmatrix}^{-}$$

$\hat{\boldsymbol{\beta}}$ and $\hat{v}$ can be viewed as empirical-Bayes means of relevant posterior distributions (Laird and Ware, 1982).

If $\mathbf{G}$ and $\mathbf{R}$ were known, $\hat{\boldsymbol{\beta}}$ would be the best linear unbiased estimator (BLUE) of $\boldsymbol{\beta}$ and $\hat{v}$ would be the best linear unbiased predictor (BLUP) of v (see McLean, Sanders, and Stroup, 1991). Their covariance matrix would be $\mathbf{C}$, which is any generalized inverse of the coefficient matrix in the mixed model equations. However, upon substitution of $\hat{\mathbf{G}}$ and $\hat{\mathbf{R}}$, and the resulting $\hat{\mathbf{C}}$, only approximate tests are available in general. ML LRTs are recommended here, although it is possible to construct approximate t- and F-statistics as well.

Consider the estimable linear hypothesis (Searle, 1971) of the form $H: \mathbf{L}\boldsymbol{\beta} = 0$. To carry out the ML LRT, use ML to fit both the full and constrained models and compare the difference of their $-2ll$ values with the χ^2-distribution, with degrees of freedom equal to the rank of $\mathbf{L}$. Rank 1 tests can be inverted to form likelihood-ratio confidence intervals. REML should not be used for these fixed-effects tests because the REML optimization is carried out over a space orthogonal to the particular fixed-effects specification.

The approximate t- and F-statistics are

$$t = \mathbf{L}\hat{\boldsymbol{\beta}} / \sqrt{\mathbf{L}\hat{\mathbf{C}}\mathbf{L}'}$$

$$F = \hat{\boldsymbol{\beta}}'\mathbf{L}'(\mathbf{L}\hat{\mathbf{C}}\mathbf{L}')^{-1}\mathbf{L}\hat{\boldsymbol{\beta}} / rank(\mathbf{L})$$

One advantage these statistics have over the LRT is that they can be computed from the full model fit alone. Satterthwaite-type degrees-of-freedom adjustments can improve the approximations (McLean and Sanders, 1988).

Table 3: *Line-source sprinkler irrigation data: yields of 3 cultivars of winter wheat*

		(North)				Irrigation Level					(South)		
Block	Cultivar	1	2	3	4	5	6	6	5	4	3	2	1
1	Luke	2.4	2.7	5.6	7.5	7.9	7.1	6.1	7.3	7.4	6.7	3.8	1.8
1	Nugaines	2.2	2.5	4.3	6.3	7.9	7.1	6.2	5.3	5.3	5.2	5.4	2.9
1	Bridger	2.9	3.2	5.1	6.9	6.1	7.5	5.6	6.5	6.6	5.3	4.1	3.1
2	Nugaines	2.4	2.5	4.0	5.8	6.1	6.2	7.0	6.4	6.7	6.4	3.7	2.2
2	Bridger	2.6	3.1	5.7	6.4	7.7	6.8	6.3	6.2	6.6	6.5	4.2	2.7
2	Luke	2.2	2.7	4.3	6.9	6.8	8.0	6.5	7.3	5.9	6.6	3.0	2.0
3	Nugaines	1.8	1.9	3.7	4.9	5.4	5.1	5.7	5.0	5.6	5.1	4.2	2.5
3	Luke	2.1	2.3	3.7	5.8	6.3	6.3	6.5	5.7	5.8	4.5	2.7	2.3
3	Bridger	2.7	2.8	4.0	5.0	5.2	5.2	5.9	6.1	6.0	4.3	3.1	3.1

A line-source sprinkler is placed east-to-west between the 2 columns with irrigation level 6 (Hanks et al., 1980).

3 EXAMPLES

The following two examples illustrate covariance structure selection in general mixed models.

3.1 Line-Source Sprinkler Irrigation

Line-source sprinkler irrigation data, presented spatially in Table 3, are from Hanks et al. (1980), and are analyzed by Johnson, Chaudhuri, and Kanemasu (1983) and Stroup (1989). Three cultivars of winter wheat are randomly assigned to rectangular plots within each of three blocks. The nine plots are located side-by-side, and a line-source sprinkler is placed across the middle of them. Each plot is subdivided into twelve subplots, six to the north of the line-source, six to the south. The subplots closest to the line-source represent the maximum irrigation level (6), the two next-closest the next-highest level (5), and so forth.

Taking the 3-stage approach outlined in Section 2.2, the first stage is to select fixed effects while favoring less parsimonious models. A reasonable

starting point is the full model which includes cultivar, direction, and irrigation main effects and all of their interactions. Using traditional dummy variables, the $\mathbf{X}$ matrix for this case is 108×84, and can be subdivided as follows:

$$\mathbf{X} = [1|\mathbf{X}_C|\mathbf{X}_D|\mathbf{X}_E|\mathbf{X}_{CD}|\mathbf{X}_{CE}|\mathbf{X}_{DE}\mathbf{X}_{CDE}]$$

where the subscripts C, D, and E stand for cultivar, direction, and irrigation, respectively. The number of columns equals $1+3+2+6+6+18+12+36=84$, although the rank of $\mathbf{X}$ is only 36.

As an initial variance model, consider the one consisting of the 8 variance components resulting from the interaction between block and the full set of fixed effects. This is a type of strip-plot design. The interaction of blocks with the 3-factor interaction corresponds to the residual variance modeled in $\mathbf{R} = \sigma^2\mathbf{I}$, and the remaining 7 variance components are modeled in the main diagonal of $\mathbf{G}$. The corresponding 108×126 $\mathbf{Z}$ matrix contains dummy variables, and can be subdivided as

$$\mathbf{Z} = [\mathbf{Z}_B|\mathbf{Z}_{BC}|\mathbf{Z}_{BD}|\mathbf{Z}_{BE}|\mathbf{Z}_{BCD}|\mathbf{Z}_{BCE}|\mathbf{Z}_{BDE}]$$

where the B subscript stands for blocks, and the other subsripts are the same as for $\mathbf{X}$. The number of columns in $\mathbf{Z}$ equals $3+9+6+18+18+54+36=126$, and is greater than the number of observations, 108. This leads one to suspect possible over-parameterization problems, and in fact, upon fitting the mixed model using ML, the 3 variance components involving cultivar are estimated to be 0. The REML estimates are 0 as well, and therefore it appears sensible to drop them from the model. The model is thereby reduced to 5 variance components, in which $\mathbf{Z}$ is subdivided as

$$\mathbf{Z} = [\mathbf{Z}_B|\mathbf{Z}_{BD}|\mathbf{Z}_{BE}|\mathbf{Z}_{BDE}]$$

with $3+6+18+36=63$ columns.

Table 4: *REML likelihood ratio tests of covariance structures for sprinkler data*

Model	G	R	R Subject	q	$-2ll_R$	CM	df	χ^2
1	none	I	BCDE	1	201.2			
2	none	AR(1)	BCD	2	176.8	1	1	24.4***
3	none	TOEP	BCD	6	172.7	2	4	4.1
4	none	AR(1)	BC	2	179.1	1	1	22.1***
5	none	TOEP(5)	BC	5	169.2	1	4	22.1***
6	none	SP(POW)	D	2	168.1	1	1	33.1***
7	none	SP(POW)	1	2	170.4	1	1	30.8***
8	VC(B)	I	BCDE	2	170.7	1	1	30.5***
9	VC(B)	AR(1)	BCD	3	166.7	2	1	10.1**
10	VC(B)	TOEP	BCD	7	160.4	9	4	6.3
11	VC(B)	AR(1)	BC	3	168.6	4	1	10.5**
12	VC(B)	TOEP(5)	BC	6	156.0	5	1	13.2***
13	VC(B)	SP(POW)	D	3	165.3	6	1	2.8
14	VC(B)	SP(POW)	1	3	167.3	7	1	3.1
15	VC(B BD)	TOEP(5)	BC	7	155.2	12	1	0.8
16	VC(B BE)	TOEP(5)	BC	7	149.5	12	1	6.5*
17	VC(B BD BE)	TOEP(5)	BC	8	147.9	16	1	1.6
18	VC(B BD BE BDE)	I	BCDE	5	167.0	8	3	3.7

Notation for $\mathbf{G}$ and $\mathbf{R}$ is from Tables 1 and 2.
The $\mathbf{R}$ subject determines the blocks of $\mathbf{R}$ via distinct combinations of the levels of the factors listed. B=block; C=cultivar; D=direction; E=irrigation.
q is the total number of parameters in $\mathbf{G}$ and $\mathbf{R}$.
CM is the comparison model for the likelihood ratio test.
* $p < .05$; ** $p < .005$; *** $p < .0005$.

The ML LRT for eliminating the 3-factor interaction from this model has $\chi^2 = 19.6$ on 10 degrees of freedom, with a p-value of 0.033. On these grounds the full fixed-effects model is retained throughout covariance structure selection.

For Stage 2, a graphical analysis of the residuals and EBLUPs of Model 1 did not reveal any outliers or serious departures from normality. These graphs are omitted here.

Stage 3 entails a more rigorous comparison of covariance structures using REML. Using the notation from Tables 1 and 2, a sequence of model fits is described in Table 4.

In all of the models in Table 4, the **G** matrix is either omitted from the model or is diagonal containing variance components. The next example illustrates more complex covariance structures for **G**. The preceding initial variance model is Model 18, and a REML LRT with the simpler Model 8 consisting of a single variance component for blocks reveals no significant difference. For all of the variance component models, the accompanying **Z** matrix consists of a subset of the columns of the **Z** matrix of Model 18.

Concerning the **R** matrix, better-fitting models are obtained by considering covariance structures more complex than $\mathbf{R} = \sigma^2\mathbf{I}$ assumed by Models 1, 8, and 18. Several time-series and spatial structures are considered here, and besides selecting one of them, one must also select the blocks of **R**. Table 4 uses a simple notation that lists the factors whose levels distinguish the blocks of **R**. To avoid confusion with the blocking variable in the experiment, the list of these factors is called the subject. For example, the BC subject specifies that observations that have the same level of both block and cultivar are correlated according to the selected covariance structure. Therefore, Models 4 and 11 specify an AR(1)-type correlation for the 12 observations in each row of Table 3. Models 6 and 13 specify a spatial correlation across the 54 observations in each direction, and Models 7 and 14 specify a spatial correlation across all 108 observations.

The LRTs carried out in Table 4 favor Models 6, 7, 9, 11, and 16; however, because they are not nested, the LRT provides no further guidance. It is therefore natural to look at the values of $-2ll_R$ in order to rank these models, and the information criteria described in Section 2.2 provide adjustments to $-2ll_R$ that enable a comparison. Figure 1 plots SBC_R versus AIC_R

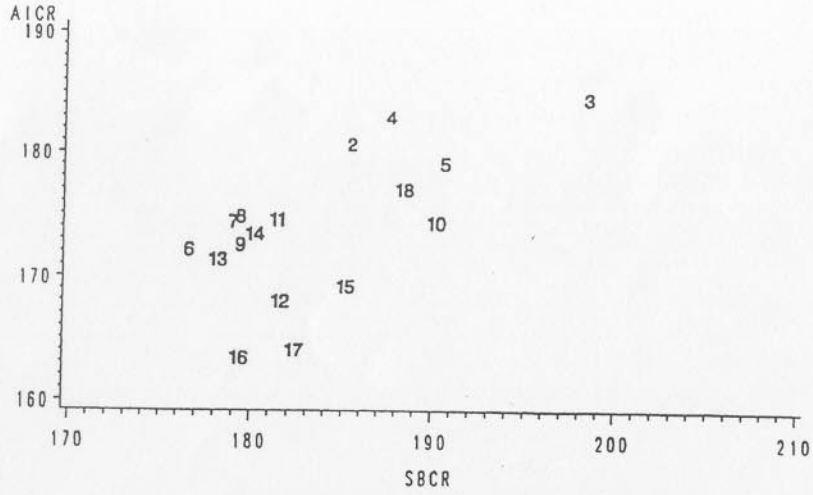

Figure 1. AICR vs. SBCR for sprinkler–data models from Table 4. Model 1 is off of the graph to the upper right.

for these and other models from Table 4. Note that AIC_R favors Model 16, whereas SBC_R favors Model 6.

The **G** matrix for Model 16 specifies variance components for block and the block by irrigation interation, and the **R** matrix is banded-Toeplitz with 5 bands for the 12 observations in each row of Table 3. This is akin to an MA(4) assumption on the errors. The 5 estimated parameters of **R** are plotted in Figure 2. An interesting extension of this analysis is to try fitting more parsimonious structures that actually model the decline shown in Figure 2. The spatial structures in Zimmerman and Harville (1991), along with a nugget effect, can be useful in this regard.

Concerning the final fixed effects, the ML LRT in Model 16 for dropping the 3-term interaction yielded $\chi^2 = 145.0 - 121.1 = 23.9$ on 10 degrees of freedom, with a p-value less than 0.01. The same test in Model 6 yielded $\chi^2 = 173.7 - 149.1 = 24.6$. Therefore, both models appear to indicate that

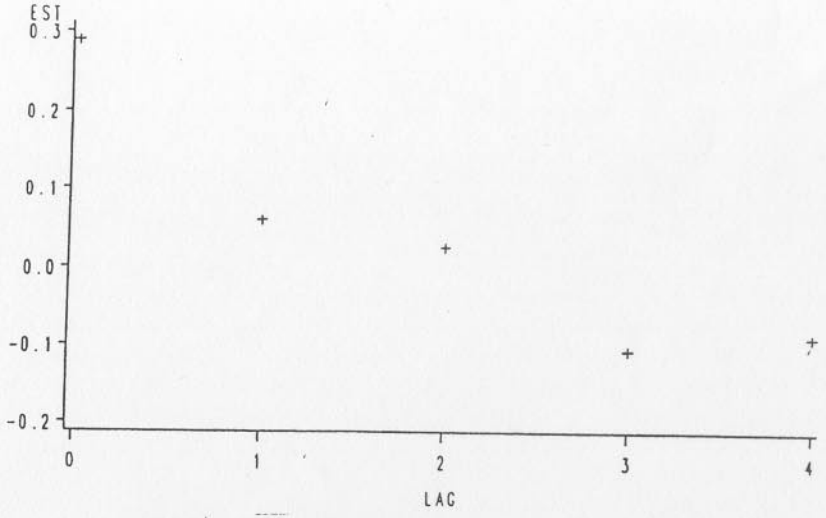

Figure 2: REML banded–Toeplitz values from Model 16 in Table 4.

no reduction of the full fixed-effects model is necessary. Irrigation, cultivar, and their interaction are the most significant effects.

Computations for this example and the next were performed on a Hewlett-Packard series 700 workstation using the MIXED procedure (SAS Institute Inc., 1992). All except the most complex models required less than a minute of real time.

3.2 Effect of Lead on Bird Growth

Hoffman et al. (1985) study the effect of different lead treatments on the growths of American Kestrel nestlings. 40 nestlings, 4 in each of 10 nests, are orally dosed with one of 4 lead treatments every day for the first 10 days of their life. The 4 treatments consist of a control and 3 increasing levels of lead content. Each nest is considered a subject, and within each subject, measurements are taken on the body weights of 4 birds for 10 consecutive

Table 5: *Lead data: weights of birds*

	Treatment/Bird 1					Treatment/Bird 2					Treatment/Bird 3				
	Days					Days					Days				
Nest	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	13	14	20	25	25	13	15	20	24	25	11	13	17	19	19
2	13	15	21	26	34	15	17	22	28	34	16	16	22	25	29
3	11	13	19	23	29	11	13	18	22	29	11	13	18	21	25
4	12	17	21	27	32	10	11	15	19	21	10	12	15	18	16
5	12	18	21	26	31	14	16	20	24	29	11	11	12	14	16
6	8	10	13	16	24	9	12	14	18	25	11	14	16	20	27
7	14	19	25	36	40	14	18	24	32	35	13	16	21	28	29
8	13	13	18	23	30	11	11	13	17	22	16	17	22	25	30
9	16	21	26	30	39	12	17	22	28	34	19	21	26	30	33
10	12	17	20	25	29	17	20	28	34	36	15	23	24	32	36

Each bird is measured for 5 days under one of 3 different lead treatments (Hoffman et al., 1985).

days. Some birds receiving the highest lead treatment died before the end of the experiment, resulting in a total of 387 observations.

Schaalje et al. (1991) analyze the logarithms of a balanced subset of these data consisting of the first 5 measurements on all birds except those receiving the highest lead treatment; these 150 observations are presented in Table 5. For comparison purposes, their mixed-model setup is used:

$$\mathbf{y}_{ijk} = \mathbf{X}_{ijk}\boldsymbol{\beta} + v_{ik} + \epsilon_{ijk}$$

where $\mathbf{y}_{ijk}$ is the log(body weight) of the jth bird ($j = 1,2,3$) of the ith nest ($i = 1,\ldots,10$) on the kth day ($k = 1,\ldots,5$). The fixed effects consist of a different cubic trend for each treatment. In particular, $\mathbf{X}_{ijk} = (\mathbf{X}_{ijk1},\ldots,\mathbf{X}_{ijk12})$: $\mathbf{X}_{ijk1} = 1$; $\mathbf{X}_{ijk2}$, and $\mathbf{X}_{ijk3}$, $\mathbf{X}_{ijk4}$ correspond to linear, quadratic, and cubic orthogonal polynomials over time; $\mathbf{X}_{ijk5} = 2$ if $j = 1$ and -1 if $j = 2$ or 3; $\mathbf{X}_{ijk6} = 0$ if $j = 1$, 1 if $j = 2$, and -1

Table 6: *REML likelihood ratio tests of covariance structures for lead data*

Model	G	R	q	$-2ll_R$	CM	df	χ^2
1	none	I	1	8.8			
2	I	I	2	-48.2	1	1	57.0***
3	CS	I	3	-112.7	2	1	64.5***
4	CS	CS	4	-219.6	3	1	106.9***
5	CS	AR(1)	4	-222.7	3	1	110.0***
6	AR(1)	I	3	-105.8	2	1	57.6***
7	AR(1)	CS	4	-225.2	6	1	119.4***
8	AR(1)	AR(1)	4	-224.3	6	1	118.5***
9	AR(1)	AR(1)+I	5	-230.9	8	1	6.6*
10	AR(1)	AR(1)+J	5	-228.4	8	1	4.1*
11	AR(1)	AR(1)*TRT	8	-227.5	8	4	3.2
12	AR(1)	AR(1)*TRT+I	9	-232.4	9	4	1.5
13	AR(1)	TOEP	7	-232.3	9	2	1.4
14	AR(1)	UN	17	-246.2	9	12	15.3
15	AR(1)+I	AR(1)+I	6	-231.0	9	1	0.1
16	AR(1)+J	AR(1)+J	6	-228.4	8	2	4.1
17	TOEP	TOEP	10	-232.7	9	5	1.8
18	UN	UN	30	-256.6	17	20	23.9

Notation for **G** and **R** is from Tables 1 and 2.
G corresponds to nests and **R** to birds within nests.
q is the total number of parameters in **G** and **R**.
CM is the comparison model for the likelihood ratio test.
* $p < .05$; ** $p < .005$; *** $p < .0005$.

if $j = 3$; $\mathbf{X}_{ijk7} = \mathbf{X}_{ijk2}\mathbf{X}_{ijk5}$; $\mathbf{X}_{ijk8} = \mathbf{X}_{ijk2}\mathbf{X}_{ijk6}$; $\mathbf{X}_{ijk9} = \mathbf{X}_{ijk3}\mathbf{X}_{ijk5}$; $\mathbf{X}_{ijk10} = \mathbf{X}_{ijk3}\mathbf{X}_{ijk6}$; $\mathbf{X}_{ijk11} = \mathbf{X}_{ijk4}\mathbf{X}_{ijk5}$; $\mathbf{X}_{ijk12} = \mathbf{X}_{ijk4}\mathbf{X}_{ijk6}$. The full **X** matrix is 150×12 and the fixed-effects vector $\boldsymbol{\beta}$ is 12×1.

The **Z** matrix is 150×50, and its columns consist of dummy variables corresponding to days within nests. The random-effects vector v is of length 50 and contains 5 random effects for each nest, one for each day. **G** is therefore 50×50 and consists of 10 5×5 blocks corresponding to each nest. **R** is 150×150 and is block-diagonal with 30 5×5 blocks corresponding to each bird.

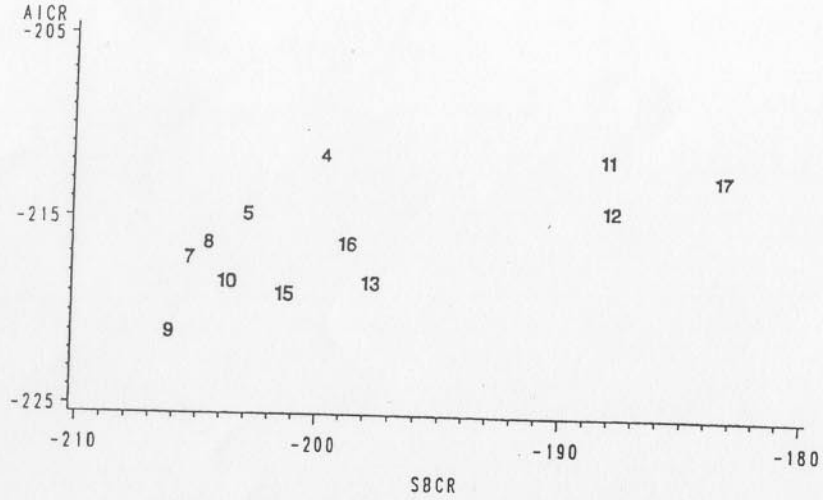

Figure 3. AICR vs. SBCR for lead–data models from Table 6. The models not pictured are off of the graph to the upper right.

Schaalje et al. (1991) consider mixed models with the same covariance structure for the blocks of both **G** and **R**. Using the notation from Tables 1 and 2, their structures are I, CS, AR(1), AR(1)+J, TOEP, and UN. They fit these models using a weighted-least-squares procedure on three quantities: subject means, deviations from these means, and a combination. By sequentially applying approximate likelihood ratio tests, they favor the AR(1)+J structure (Pantula and Pollock, 1985).

REML is used to fit a variety of covariance structures for both **G** and **R**. Table 6 describes the fits using the notation from Tables 1 and 2. Model 9, which consists of an AR(1) specification for **G** and an AR(1)+I specification for **R**, is the best according to LRT. It also is the best according to both AIC_R and SBC_R, as is seen in Figure 3. Note that the local error term of Model 9 contributes significantly to the model fit as evidenced by the LRT with Model 8.

Table 7: *REML parameter estimates for lead-data Model 9*

Parameter	Description	Estimate	Approx. SE	Ratio
β_1	Intercept	2.9833	0.0492	60.71
β_2	Linear	0.2304	0.0096	23.91
β_3	Quadratic	-0.0052	0.0047	-1.09
β_4	Cubic	-0.0084	0.0045	-1.90
β_5	Trt vs. Control	0.0257	0.0139	1.85
β_6	Trt 1 vs. Trt 2	0.0215	0.0241	0.89
β_7	Lin*(T vs. C)	0.0051	0.0025	2.00
β_8	Lin*(T1 vs. T2)	0.0097	0.0044	2.20
β_9	Qua*(T vs. C)	-0.0012	0.0016	-0.75
β_{10}	Qua*(T1 vs. T2)	-0.0022	0.0027	-0.81
β_{11}	Cub*(T vs. C)	0.0020	0.0017	1.18
β_{12}	Cub*(T1 vs. T2)	0.0041	0.0029	1.41
σ^2_D	Variance in **G**	0.0229	0.0112	2.05
ρ_D	AR(1) in **G**	0.9252	0.0421	21.96
σ^2_R	Variance in **R**	0.0122	0.0039	3.16
ρ_R	AR(1) in **R**	0.9555	0.0287	33.24
σ^2	Local error	0.0012	0.0004	2.72

Models 11 and 12 have different AR(1) parameters for each treatment in **R**, and these heterogeneous models do not fit significantly better than their homogeneous counterparts. Model 16 is the model chosen by Schaalje et al. (1991); however, in this analysis, it does not fit significantly better than the simpler Model 8 which they also consider. Note that the extra parameters in Models 14, 17, and 18 appear unnecessary, which is appealing from the point-of-view of parsimony.

The parameter estimates for Model 9 are in Table 7. The fixed-effects estimates $\hat{\boldsymbol{\beta}}$ are listed first, along with their approximate standard errors and t-statistics from Section 2.3. These estimates are close to those obtained by Schaalje et al. (1991), with a possible exception being that for the linear orthogonal polynomial. They estimate its value to be 0.7286, while this REML analysis yields 0.2329.

For the covariance parameter estimates in Table 7, the blocks of **G** are denoted as

$$\sigma^2_D \begin{bmatrix} 1 & \rho_D & \rho_D^2 & \rho_D^3 \\ & 1 & \rho_D & \rho_D^2 \\ & & 1 & \rho_D \\ & & & 1 \end{bmatrix}$$

and the blocks of **R** are denoted as

$$\sigma^2_R \begin{bmatrix} 1 & \rho_R & \rho_R^2 & \rho_R^3 \\ & 1 & \rho_R & \rho_R^2 \\ & & 1 & \rho_R \\ & & & 1 \end{bmatrix} + \begin{bmatrix} \sigma^2 & & & \\ & \sigma^2 & & \\ & & \sigma^2 & \\ & & & \sigma^2 \end{bmatrix}$$

The approximate standard errors of the covariance parameter estimates are from the observed inverse Fisher information matrix, and a comparison of the ratio with the standard normal distribution provides a Wald test.

Regarding computing, most models took less than a minute of real time on an HP-700. The UN-UN model took 1 hour, with most of the time spent calculating the first and second derivatives of the likelihood for each Newton-Raphson step.

4 DISCUSSION

There is a bit of redundancy in using the general linear mixed model, as specifying a simple random effect for each subject along with $\mathbf{R} = \sigma^2\mathbf{I}$ is equivalent to ignoring **Z** and specifying a compound-symmetry structure in the blocks of **R**. Besides this, though, variance modeling in **G** and **R** take on different forms. As **Z** will tend to have many more rows than columns, random effects are suited to modeling correlation among a large number of observations, whereas covariance modeling in **R** is more local.

An interesting challenge for the approach outlined in this paper is time-series cross-sectional data (Dielman, 1989), also known as panel data or multivariate repeated measures. Some current methods for such data are special cases of the mixed model. For example, the crossed-error method of Fuller and Battese (1974) involves specifying variance components for both the time-series and cross-sectional units. However, more complex covariance structures involve the Kronecker product of different structures. The structure of Parks (1967), for instance, is the Kronecker product of an AR(1) and unstructured matrix. Estimation is typically carried out by method-of-moments, although ML or REML appears to be feasible.

Two structures not considered in this article are the power-of-the-mean (POM, Carroll and Ruppert, 1988) and the log-variance (Harvey, 1976; Aitkin, 1987). Both are heterogeneous-diagonal, and can be used to weight other structures such as compound symmetry, AR(1), and Toeplitz. The latter two of these produce structures akin to autoregressive conditionally heterogeneous (ARCH) models in econometrics (Engle and Bollerslev, 1986; Bollerslev, Chou, and Kroner, 1992). The parameters of the log-variance model, $\boldsymbol{\delta}$, are called dispersion effects, and are modeled with a design matrix, $\mathbf{U}$, that may be the same or different from $\mathbf{X}$. Log-variance weighting thereby offers a series of structures that bridges the gap between compound symmetry and unstructured in a way different from factor-analytic structures.

BIBLIOGRAPHY

Aitkin, M. (1987). Modelling variance heterogeneity in normal regressing using GLIM. *Applied Statistics* **36**, 332–339.

Bollerslev, T., Chou, R.Y., and Kroner, K.F. (1992). ARCH modeling in finance. *Journal of Econometrics* **52**, 5-59.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* **52**, 345-370.

Carroll, R.J. and Ruppert, D. (1988). *Transformation and Weighting in Regression.* London: Chapman & Hall.

Chi, E.M. and Reinsel, G.C. (1988). Models for longitudinal data with random effects and AR(1) errors. *Journal of the American Statistical Association,* **84**, 452-459.

Cressie, N. (1991). *Statistics for Spatial Data.* New York: John Wiley and Sons.

Dempster, A.P., Selwyn, M.R., Patel, C.M., and Roth, A.J. (1984). Statistical and computational aspects of mixed model analysis. *Applied Statistics* **33**, 203-214.

Dielman, T.E. (1989). *Pooled Cross-Sectional and Time Series Data Analysis.* Marcel Dekker, Inc.

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

Engle, R.F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances (with discussion). *Econometric Reviews* **5**, 1-50.

Fuller, W.A. (1976). *Introduction to Statistical Time Series.* New York: John Wiley and Sons.

Fuller, W.A. and Battese, G.E. (1974). Estimation of linear models with crossed-error structure. *Journal of Econometrics* **2**, 67-78.

Hanks, R.J., Sisson, D.V., Hurst, R.L., and Hubbard, K.G. (1980). Statistical analysis of results from irrigation experiments using the line-source sprinkler system. *Journal of the American Soil Science Society* **44**, 886-888.

Harvey, A.C. (1976). Estimating regression models with multiplicative heteroscedasticity. *Econometrica* **44**, 461-465.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320-338.

Helms, Ronald W., and Edwards, Lloyd (1991). Analysis of irregular and/or incomplete data from crossover designs using mixed models with linear

covariance structure. *Proceedings of the Biopharmaceutical Section,* Atlanta meetings, American Statistical Association.

Henderson, C.R. (1984). *Applications of Linear Models in Animal Breeding.* University of Guelph, Canada.

Hocking, R.R. (1985). *The Analysis of Linear Models.* Monterey, CA: Brooks/Cole.

Hoffman, D.J., Franson, J.C., Pattee, O.H., Bunck, C.M., and Anderson, A. (1985). Survival, growth, and accumulation of ingested lead in nestling American Kestrels (*Falco spaverius*). *Archives of Environmental Contamination and Toxicology* **14**, 89-94.

Jennrich, R.I., and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805-820.

Johnson, D.E., Chaudhuri, U.N., and Kanemasu, E.T. (1983). Statistical analysis of line-source sprinkler irrigation experiments and other non-randomized experiments using multivariate methods. *Journal of the American Soil Science Society* **47**, 309-312.

Jones, R.H., and Boadi-Boateng, F. (1991). Unequally spaced longitudinal data with AR(1) serial correlation. *Biometrics* **47**, 161-175.

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

Lindstrom, M.J. and Bates, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association* **83**, 1014-1022.

McLean, Robert A. and Sanders, W.L. (1988). Approximating degrees of freedom for standard errors in mixed linear models. *Proceedings of the Statistical Computing Section,* American Statistical Association, New Orleans, pp. 50-59.

McLean, Robert A., Sanders, William L., and Stroup, Walter W. (1991). A unified approach to mixed linear models. *The American Statistician* **45**, 54-64.

Pantula, S.G. and Pollock, K.H. (1985). Nested analysis of variance with autocorrelated errors. *Biometrics* **41**, 909-920.

Parks, R.W. (1967). Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association* **62**, 500-509.

Reid, N. (1988). Saddlepoint methods and statistical inference (with discussion). *Statistical Science* **3**, 213-38.

Robinson, G.K. (1991). That BLUP is a good thing: the estimation of random effects (with discussion). *Statistical Science* **6**, 15-51.

Rochon, J. (1992). ARMA covariance structures with time heteroscedasticiy for repeated measures experiments. *Journal of the American Statistical Association* **87**, 777-784.

SAS Institute Inc. (1992). SAS Technical Report P-229. *SAS/STAT Software: Changes and Enhancements, Release 6.07.* Chapter 16: The MIXED Procedure. Cary, NC: SAS Institute Inc.

Sanders, W.L. (1989). A multivariate mixed model. *Applications of Mixed Models in Agriculture and Related Disciplines,* Southern Cooperative Series Bulletin No. 343. Louisiana Agricultural Experiment Station, Baton Rouge, 138-144.

Schaalje, B., Zhang, J., Pantula, S.G., and Pollock, K.H. (1991). Analysis of repeated-measurements data from randomized block experiments. *Biometrics* **47**, 813-824.

Schluchter, M. (1988). Unbalanced repeated measures models with structured covariance matrices. BMDP Statistical Sofware Manual, Vol. 2, University of California Press, Berkeley, 1081-1187.

Searle, S.R. (1971). *Linear Models.* New York: John Wiley and Sons.

Searle, S.R. (1988). Mixed models and unbalanced data: wherefrom, whereat, and whereto? *Communications in Statistics - Theory and Methods* **17(4)**, 935-968.

Searle, S.R., Casella, G., and McCulloch, C.E. (1992). *Variance Components.* New York: John Wiley and Sons.

Stroup, W.W. (1989). Use of mixed model procedure to analyze spatially correlated data: an example applied to a line-source sprinkler irrigation experiment. *Applications of Mixed Models in Agriculture and Related Disciplines,* Southern Cooperative Series Bulletin No. 343. Louisiana Agricultural Experiment Station, Baton Rouge, 104-122.

Wolfinger, R.D., Tobias, R.D., and Sall, J. (1992). Computing Gaussian likelihoods and their derivatives for general linear mixed models. in review.

Zhang, J., Pantula, S.G., and Boos, D.D. (1991). Robust methods for testing the pattern of a single covariance matrix. *Biometrika* **78**, 787-798.

Zimmerman, D.L. and Harville, D.A. (1991). A random field approach to the analysis of field-plot experiments and other spatial experiments. *Biometrics* **47**, 223-239.

Received April 1993.

COMMUN. STATIST.—SIMULA., 22(4), 1107-1114 (1993)

AN EFFICIENT CROSS-VALIDATION ALGORITHM FOR WINDOW WIDTH SELECTION FOR NONPARAMETRIC KERNEL REGRESSION

Jeff Racine

Visiting Assistant Professor
Department of Economics, 0508
9500 Gilman Drive
University of California, San Diego
La Jolla, CA 92093

Key Words and Phrases: kernel regression; window width selection; cross-validation; computational efficiency.

ABSTRACT

This paper presents an approach to cross-validated window width choice which greatly reduces computation time, which can be used regardless of the nature of the kernel function, and which avoids the use of the Fast Fourier Transform. This approach is developed for window width selection in the context of kernel estimation of an unknown conditional mean.

1. INTRODUCTION.

Nonparametric density estimation procedures have provided many exciting new techniques for statistical analysis and data exploration (for an excellent survey of density estimation techniques, see Izenman [2]). However it is well known that these techniques tend to be very computationally intensive. Therefore, it is extremely important to use algorithms which are computationally efficient to avoid excessively long calculation times.

Algorithms for nonparametric estimation can be decomposed into two parts – one for the calculation of the kernel and another for window width choice. Silverman [6] (page 88) states that "One important factor in reducing the computer time is the choice of a kernel that can be calculated very quickly.". Having chosen a kernel that is efficient to compute, one must then choose the window widths.

There is no generally accepted method for choosing the window widths. Methods currently available include 'subjective choice' and automatic methods such as the 'plug-in', 'cross-validation' (CV), and 'penalizing function' approaches (See Marron [3] for an excellent survey). Härdle [1] (page 173) compared various automatic methods and found that "The best overall performance, though, showed GCV (generalized cross-validation).".