

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

2000 - 12th Annual Conference Proceedings

SOME STRATEGIES FOR SELECTING AND FITTING COVARIANCE STRUCTURES FOR REPEATED MEASURES

Raul E. Macchiavelli

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

Macchiavelli, Raul E. (2000). "SOME STRATEGIES FOR SELECTING AND FITTING COVARIANCE STRUCTURES FOR REPEATED MEASURES," *Conference on Applied Statistics in Agriculture*.
<https://doi.org/10.4148/2475-7772.1248>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

SOME STRATEGIES FOR SELECTING AND FITTING COVARIANCE STRUCTURES FOR REPEATED MEASURES

Raúl E. Macchiavelli, Dept. of Agronomy and Soils, Univ. of Puerto Rico, P.O. Box 9030,
Mayagüez, PR 00681-9030

Abstract

Since in longitudinal studies the covariance structure is often regarded as a nuisance parameter, the strategy has been to use a parsimonious covariance model that describes adequately the observed data and permits better inference on the parameters of interest. In this paper we present some diagnostic tools to choose an appropriate covariance structure and discuss some strategies for fitting it. The main diagnostic tool is the "residual", computed as the standardized difference between the elements of the fitted covariance (concentration or correlation) matrix and the corresponding unstructured matrix. SAS Proc Calis is a very efficient procedure that fits many covariance structures in models with no fixed effects. Based on this procedure, we discuss some strategies to choose initial values and improve convergence problems in certain commonly used structures.

1. Introduction

The analysis of repeated measurements needs to consider the dependencies among the observations taken on the same unit. In normal linear models, this is attained by assuming a particular structure for the covariance matrix (Dempster 1992, Lindsey 1993, Jennrich and Schluchter 1986, Wolfinger 1996).

The choice of structure affects both the power and the validity of test procedures about the mean parameters. Since in longitudinal studies the covariance structure is often regarded as a nuisance parameter, the ideal solution is to use a parsimonious covariance model that describes well the observed data. If an unrealistically simple covariance is selected, the tests may be invalid. On the other hand, an unnecessarily complex structure (or no structure at all) may decrease the power (Macchiavelli and Moser 1997).

Diggle (1988) mentions several steps involved in making inferences about the mean parameters in these models. The first step is to fit a mean model (overfitting if necessary to avoid creating spurious dependencies), then initial covariance structures can be studied using relevant theory and graphics. Finally, formal statistical techniques can be applied to select the final covariance structure and make inferences for the mean parameters. From a practical perspective, this strategy has two aspects that are critical to ensure a good covariance structure is chosen and fit: the selection of the structure and the likelihood optimization problem for fitting the selected structure.

Methods for choosing appropriate structures are mostly based on model selection techniques: sequence of likelihood ratio tests (Lindsey 1993, Diggle et al. 1994) or penalized likelihood criteria (Macchiavelli 1995). In the related area of linear structural equations

(Jöreskog and Sörbom 1987), indices of goodness of fit and residuals have been developed (Browne 1982, SAS 1996). In this paper we introduce residuals from the covariance and concentration structures, which, after proper standardization, permit the graphical evaluation of the goodness of the fit.

In the last five years, there has been a renewed interest in fitting new structures, due mostly to their widespread availability in SAS Proc Mixed (Wolfinger 1996, SAS 1996). One of the main difficulties that practitioners have found in the application of these more complex covariance structures is the poor convergence of the numerical routines used to fit them. This is due, in part, to the large number of parameters involved in complex structures if the number of time periods is large. The choice of appropriate initial values for the parameters is therefore very important to improve the convergence of the algorithms. We present a strategy for computing initial values for a variety of covariance structures using SAS Proc Calis.

2. Concentration residuals

Let $\mathbf{Y} = (Y_1, \dots, Y_T) \sim N_T(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ be a T -dimensional vector representing repeated observations on a single experimental unit. Consider n independent random vectors $(\mathbf{Y}_1, \dots, \mathbf{Y}_n)$, each with the same covariance matrix. Define

$$\mathbf{Y}_{n \times T} = \begin{pmatrix} \mathbf{Y}_1 \\ \dots \\ \mathbf{Y}_n \end{pmatrix}$$

and assume a linear model $E(\mathbf{Y}_{n \times T}) = \mathbf{X}_{n \times a} \boldsymbol{\Theta}_{a \times T}$, for some $n \times a$ matrix \mathbf{X} of full column rank. The elements of $\boldsymbol{\Theta}$ are functionally independent.

We are usually interested in making inferences about $\boldsymbol{\Theta}$, the mean parameters, while the covariance matrix, $\boldsymbol{\Sigma}$, contains the nuisance parameters. Since the multivariate normal distribution is an exponential family, the canonical parametrization for the nuisance parameter is not in terms of $\boldsymbol{\Sigma}$ but in terms of the concentration matrix, $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$. The likelihood for n independent observations using this parametrization is:

$$L(\mathbf{Y}; \boldsymbol{\Theta}, \boldsymbol{\Omega}) = (2\pi)^{-nT/2} |\boldsymbol{\Omega}|^{n/2} \exp \left\{ -\frac{1}{2} \text{tr} \left[\boldsymbol{\Omega} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\Theta})(\mathbf{Y}_i - \mathbf{X}_i \boldsymbol{\Theta})' \right] \right\}, \quad (2.1)$$

where \mathbf{X}_i is the i th row of \mathbf{X} .

As mentioned before, several structures are commonly used for $\boldsymbol{\Omega} = (\omega_{ij})$ or $\boldsymbol{\Sigma} = (\sigma_{ij})$. The compound symmetry structure assumes $\omega_{ii} = a + b$ and $\omega_{ij} = b$ if $i \neq j$. The first order autoregressive model assumes the $\sigma_{ij} = \sigma \rho^{|i-j|}$. The first order antedependence assumes that $\omega_{ij} = 0$ if $|i-j| > 1$. In the particular case of structures defined only by conditional independence (zeroes in $\boldsymbol{\Omega}$), we have a linear exponential family (Dempster 1972, Cox and Wermuth 1990). Some of the most commonly used covariance structures have structural zeroes in $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ (CS, CSH, AR(1), ARH(1), and ANTE(1)), while others have structural zeroes in $\boldsymbol{\Sigma}$, (TOEP(q), TOEPH(q), and UN(q)).

From the likelihood function, it is clear that the MLE of $\mathbf{\Omega}$ for complete data if no structure is assumed is $\hat{\mathbf{\Omega}}_U = \mathbf{S}^{-1}$, where

$$\mathbf{S} = \frac{1}{n} \sum_{i=1}^n (\mathbf{Y}_i - \mathbf{X}_i \hat{\mathbf{\theta}}) (\mathbf{Y}_i - \mathbf{X}_i \hat{\mathbf{\theta}})'. \quad (2.2)$$

Under a particular structure, the MLE of $\mathbf{\Omega}$ will be denoted as $\hat{\mathbf{\Omega}}$.

Given that the canonical parameters are the elements in $\mathbf{\Omega}$, a natural way of checking the appropriateness of a particular structure is by using the “concentration residuals.”

$$\mathbf{R} = \hat{\mathbf{\Omega}}_U - \hat{\mathbf{\Omega}}. \quad (2.3)$$

These residuals have the advantage over the “covariance residuals” (defined similarly for $\mathbf{\Sigma}$) of being in terms of canonical parameters. Furthermore, they do not have structural zeroes in many commonly used situations, like the conditionally independence models considered by Dempster (1972), Cox and Wermuth (1990), and Macchiavelli and Moser (1997).

In order to standardize these residuals, and be able to display them in a sensible way, their asymptotic covariance matrix will now be computed.

Let $\mathbf{A}_{T \times T}$ be a symmetric matrix. Define the operator *vech* as the $T(T-1)/2$ -vector of nonredundant elements of \mathbf{A} , arranged by rows (Henderson and Searle, 1979). Let $\hat{\omega}_U = \text{vech}(\hat{\mathbf{\Omega}}_U)$ and $\hat{\omega} = \text{vech}(\hat{\mathbf{\Omega}})$. Then $\mathbf{r} = \text{vech}(\mathbf{R}) = \hat{\omega}_U - \hat{\omega}$.

The asymptotic distribution of $\sqrt{n} \text{cov}(\hat{\omega}_U - \omega)$ is $N(0, \text{Iss}(\mathbf{\Omega}))$ (Cox and Wermuth, 1990), where $\text{Iss}(\mathbf{\Omega})$ has elements $n \text{cov}(\hat{\omega}_{ij}, \hat{\omega}_{kl}) = \omega_{ik} \omega_{jl} + \omega_{il} \omega_{jk}$.

If a structure is present in $\mathbf{\Omega}$, fewer parameters will be necessary. Let γ be the vector containing such parameters. We will denote the structured concentration as $\mathbf{\Omega}(\gamma)$ to emphasize its dependence on γ . Since the usual regularity conditions hold, $\hat{\gamma}$, the MLE of γ , is asymptotically normal:

$$\sqrt{n}(\hat{\gamma} - \gamma) \sim N(0, \mathbf{C}(\gamma)), \quad (2.4)$$

where $\mathbf{C}(\gamma) = n \mathbf{I}^{-1}(\gamma)$. The asymptotic distribution of $\sqrt{n}(\hat{\omega} - \omega)$ can be deduced from this one using standard results on functions of MLEs (see, for example, Seber 1984, p. 532):

$$\sqrt{n}(\hat{\omega} - \omega) \sim N(0, \mathbf{J}(\gamma) \mathbf{C}(\gamma) \mathbf{J}(\gamma)'), \quad (2.5)$$

where $\mathbf{J} = \partial \text{vech} \mathbf{\Omega} / \partial \gamma$.

Since $\sqrt{n}(\hat{\omega} - \omega) = \mathbf{J}(\gamma)(\hat{\gamma} - \gamma) + o_p(1)$, then $\text{cov}(\sqrt{n} \hat{\omega}_U, \sqrt{n} \hat{\omega}) = \mathbf{J}(\gamma) \mathbf{C}(\gamma) \mathbf{J}(\gamma)' + o_p(1)$.

Therefore,

$$\text{Cov}(\sqrt{n}(\hat{\omega}_U - \hat{\omega})) \approx \text{Iss}(\mathbf{\Omega}) - \mathbf{J}(\gamma) \mathbf{C}(\gamma) \mathbf{J}(\gamma)'. \quad (2.6)$$

From these results we can obtain the asymptotic distribution of $\sqrt{n} \mathbf{r}$, which will be normal with expectation 0 and covariance matrix $n \mathbf{K} = \text{Iss}(\mathbf{\Omega}) - \mathbf{J}(\gamma) \mathbf{C}(\gamma) \mathbf{J}(\gamma)'$. In practice, an estimate of \mathbf{K} can be obtained by substituting $\mathbf{\Omega}$ by $\hat{\mathbf{\Omega}}_U$ (or $\hat{\mathbf{\Omega}}$), and γ by $\hat{\gamma}$.

In order to display the concentration residuals, two standardizations can be done. The first one consists in dividing each residual by its standard deviation (the square root of each diagonal element of \mathbf{K}), and plot these standardized residuals in a boxplot or a stem-and-leaf display. Large absolute values or extreme asymmetry in the plot are indications of bad fit. The second alternative is to plot $\mathbf{K}^{-1/2}\mathbf{r}$ in a normal probability plot: we should observe a straight line if the structure fits well.

3. Some strategies for finding initial values of covariance parameters

One of the main difficulties encountered when trying to fit covariance structures to large data sets with missing data and complex covariances is the optimization of the likelihood function. This is a particularly difficult task given the number of covariance parameters found when there are many time points and heterogeneous models are used. In cases like this, it is critical to use “good” initial estimates to improve (or even attain) convergence.

This problem is compounded by the fact that many specifications of the covariance structure in SAS Proc Mixed are done by parametrizing Σ instead of $\Omega = \Sigma^{-1}$. This is important when the relationship between the elements in both matrices is not trivial (of course, there is always a one-to-one correspondence between them). In the case of conditional independence models Ω has a very simple structure (because it has many structural zeroes), while Σ is much more complex. One way of obtaining initial values for the parameters in Σ is by using Proc Calis, which has a very flexible way of modeling covariances. In this procedure one can specify, in the COSAN option, any structure of the form:

$$\Sigma = \mathbf{F}_1\mathbf{F}_2\mathbf{L}\mathbf{F}_n\mathbf{P}\mathbf{F}_n'\mathbf{L}\mathbf{F}_1' \quad (3.1)$$

where each \mathbf{F} can be a general matrix \mathbf{G} , its inverse \mathbf{G}^{-1} , or $(\mathbf{I}-\mathbf{G})^{-1}$; and \mathbf{P} can be a symmetric matrix \mathbf{Q} or its inverse \mathbf{Q}^{-1} .

If the data were complete, the maximum likelihood estimator under any structure could be obtained from $\hat{\Sigma}_U$, the MLE of Σ assuming no structure. This can be done in Proc Calis very efficiently (Macchiavelli and Moser 1996). If the data are not complete, the strategy is to obtain an approximate estimate of Σ based on $\hat{\Sigma}_U$ in Proc Calis and then extract from this the initial values for the parameters in Proc Mixed. This may require some processing in Proc IML in order to have a smooth transition between the procedures.

The first order antedependence structure, ANTE(1), can be specified in Proc Calis as $\Sigma = (\mathbf{I}-\mathbf{C})^{-1}\mathbf{D}(\mathbf{I}-\mathbf{C})^{-1}$, where \mathbf{C} is a lower triangular matrix whose only $(T-1)$ nonzero elements are located along its secondary diagonal and \mathbf{D} is a diagonal matrix with T different elements. The first order autoregressive structure, AR(1), can be specified similarly but with all nonzero elements in \mathbf{C} equal, \mathbf{D} as a diagonal matrix of σ^2 and boundary constraints to assure $|\rho| \leq 1$. If the structure is the first order autoregressive structure with heterogeneous variances, ARH(1), the specification is similar, except that the elements of \mathbf{D} are all different.

For the covariances with structural zeroes in Σ , the specifications are similar but use $\Sigma = (\mathbf{I} - \mathbf{C})\mathbf{D}(\mathbf{I} - \mathbf{C})'$ (i.e., do not invert the first factor).

4. Examples

In order to check the feasibility of the proposed methods, 100 random normal vectors were generated using SAS. Each vector had 8 elements, and was transformed so that the dependencies correspond to a heteroscedastic autoregressive model of order 1. The covariance matrix $\Sigma = (\sigma_{ij})$ has elements $\sigma_{ij} = ij\rho^{|i-j|}$, with $\rho = 0.8$.

Standardized concentration residuals were obtained assuming the following structures: compound symmetry, first order autoregressive, first order heteroscedastic autoregressive (the correct one) and first order antedependence (also correct but not parsimonious). Computations were carried out in Maple (Redfern 1994). The residuals are presented in Figure 1. We can see that the two invalid structures result in values of standardized residuals well beyond (-3,3). On the other hand, for the two valid structures the residuals are all less than 3 in absolute value. Had covariance residuals been used instead of concentration residuals (using, for example, Proc Calis), structural zeroes would have appeared in 17 out of the 36 residuals in the antedependence structure.

In the following example we simulated a real situation with 4 treatments, 8 times and 8 replicates. A scenario like this could occur, for example, if we are comparing the production (dry matter yield) of a pasture under 4 types of fertilization regimes. The plots would be arranged completely at random, with 8 replicates of each treatment. Each plot is cut every 3 months during 2 years (8 cuts in total) and its dry matter yield recorded. The assumed covariance structure was a first order antedependence. (This particular structure can be used for equidistant and not equidistant intervals.) The true matrices \mathbf{C} and \mathbf{D} (defining the regression coefficients between lagged observations and the conditional variances, respectively) were:

$$\mathbf{I} - \mathbf{C} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.8 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -0.6 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.4 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -0.6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -0.8 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -0.9 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}, \quad \text{diag}(\mathbf{D}) = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 10 \\ 10 \\ 10 \\ 10 \end{pmatrix}.$$

The resulting true covariance matrix was computed as $\Sigma = (\mathbf{I} - \mathbf{C})^{-1} \mathbf{D} (\mathbf{I} - \mathbf{C})^{-1'}$. The values were the following (covariances in the lower triangle, variances on the diagonal and correlations in the upper triangle):

$$\begin{pmatrix} 1.00 & 0.62 & 0.38 & 0.17 & 0.04 & 0.02 & 0.02 & 0.01 \\ 0.80 & 1.64 & 0.61 & 0.27 & 0.06 & 0.04 & 0.03 & 0.02 \\ 0.48 & 0.98 & 1.59 & 0.45 & 0.09 & 0.06 & 0.04 & 0.04 \\ 0.19 & 0.39 & 0.64 & 1.25 & 0.21 & 0.13 & 0.10 & 0.08 \\ 0.12 & 0.24 & 0.38 & 0.75 & 10.5 & 0.63 & 0.48 & 0.40 \\ 0.09 & 0.19 & 0.31 & 0.60 & 8.36 & 16.7 & 0.76 & 0.64 \\ 0.08 & 0.17 & 0.27 & 0.54 & 7.53 & 15.0 & 23.5 & 0.84 \\ 0.08 & 0.17 & 0.27 & 0.54 & 7.53 & 15.0 & 23.5 & 33.5 \end{pmatrix}$$

Using this mean and covariance structure, random normal observations were generated using SAS. In order to simulate missing observations, data were dropped randomly using Bernoulli random numbers with $\pi=0.25$. The SAS code used follows:

```
proc mixed method=ml data=a;
  class treat time rep;
  model y=treat|time;
  repeated time /subject=rep(treat) type=un;
  make 'r' out=unsig;

data unsig (type=cov);
  set unsig;
  _type_='COV';
  _name_='col' || left(trim(row));
  drop row;

proc calis data=unsig method=ml cov all
  outstat=hat dfe=28 gconv=.00001
  maxiter=200;
  cosan b(8,low,imi)*d(8,dia);
  matrix b
    [1,]= 0,
    [2,]= ga1 0 ,
    [3,]= 0 ga2 0,
    [4,]= 2*0 ga3 0,
    [5,]= 3*0 ga4 0,
    [6,]= 4*0 ga5 0,
    [7,]= 5*0 ga6 0,
    [8,]= 6*0 ga7 0;
  matrix d
    [1,1]=delta1-delta8 ;

data hat;
  set hat;
  if _type_='MAXPRED';
  keep col1-col8;

proc iml;
  use hat;
  read all into sigmahat;
  sigma=vecdiag(sigmahat);
  rho=j(1,7,1);
  do i=1 to 7;
    rho[i] = sigmahat[i+1,i] /
      sqrt(sigmahat[i+1,i+1]*sigmahat[i,i]);
  end;
  names='covp1':'covp15';
  param=sigma`||rho;
  create hat2 from param [colname=names];
  append from param;

proc mixed data=a;
  class treat time rep;
  model y=treat|time;
  parms / parmsdata=hat2;
  repeated time / subject=rep(treat)
    type=ante(1);

run;
```

For this particular small example the algorithm with no initial values converged to a solution in 7 iterations, while the one using the initial values did it in 4 iterations. In more complex data sets the use of “good” initial values improved noticeably the convergence.

For some other covariance structures, the specification of the COSAN model in Proc Calis needs to be changed accordingly. For example, the following is the code needed for an ARH(1) structure:

```
proc calis data=unsig method=ml cov all
  outest=hat dfe=28 gconv=.00001 maxiter=200;
  cosan b(8,low,imi)*d(8,dia);
  matrix b
    [1,]= 0 ,
    [2,]= covp9 0,
    [3,]= 0 x2 0,
    [4,]= 2*0 x3 0,
    [5,]= 3*0 x4 0,
    [6,]= 4*0 x5 0,
    [7,]= 5*0 x6 0,
    [8,]= 6*0 x7 0;
  matrix d
    [1,1]=covp1-covp8;
  bounds -1< covp9 <1;
  x2=covp9;
  x3=covp9;
  x4=covp9;
  x5=covp9;
  x6=covp9;
  x7=covp9;
run;
```

5. Summary

In this paper we discussed how the covariance and concentration matrices have parameters that can be interpreted as covariances and conditional covariances. In particular a corner of zeroes in the concentration matrix yields a more realistic structure for repeated measures, since it implies conditional independence. In order to assess the fit of a given structure, concentration residuals are introduced, together with two possibilities for their standardization. Finally, the issue of obtaining reasonable initial values for the iterative procedures is approached using SAS Proc Calis parametrizations, which permit very general structures to be fit.

Acknowledgments

The author thanks Dr. E. Barry Moser and an anonymous referee for very helpful suggestions that improved the quality of this paper. This research was supported by Univ. of Puerto Rico, Agric. Exp. Station Project SP-340.

References

- Browne, M.W. (1982) Covariance Structures. In *Topics in Multivariate Analysis*, D.M. Hawkins, ed. New York: Cambridge University Press.
- Cox, D.R. and Wermuth, N. (1990) An approximation to maximum likelihood estimates in reduced models. *Biometrika* 77: 747-761.
- Dempster, A.P. (1972) Covariance Selection. *Biometrics* 28: 157-175.
- Diggle, P. (1988) An approach to the analysis of repeated measurements. *Biometrics* 44: 959-71.
- Diggle, P., Liang, K and Zeger, S. (1979) *Analysis of Longitudinal Data*. London: Oxford Univ. Press.
- Henderson, H. and Searle, S. (1979) Vec and vech operators for matrices, with some uses in jacobians and multivariate statistics. *Canad. J. of Statistics* 7: 65-81.
- Jennrich, R. and Schluchter, M. (1986) Unbalanced repeated measures models with structured covariance matrices. *Biometrics* 42: 805-820.
- Jöreskog, K. and Sörbom, D. (1988) *LISREL 7: A guide to the program and applications*. Chicago: SPSS, Inc.
- Lindsey, J. (1993) *Models for repeated measurements*. London: Oxford Univ. Press.
- Littell, R., Henry, P. and Ammerman, C. (1998) Statistical analysis of repeated measures data using SAS Procedures. *Journal of Animal Science* 76:1216-1231.
- Macchiavelli, R. and S. Arnold (1995) Difficulties with the use of penalized likelihood criteria in antedependence and polynomial models. *Communications in Statistics, Theory and Methods*, 24 (2): 501-522.
- Macchiavelli, R. and E.B. Moser (1996) Tests of hypotheses in repeated measures models under general ante-dependence covariance structures in SAS software. *Proceedings of the 21st. SAS Users Group International Conference*, 1265-68.
- Macchiavelli, R. and E.B. Moser (1997) Analysis of repeated measures with ante-dependence models. *Biometrical Journal* 39 (3) :339-350.
- Redfern, Darren (1994) *The Maple Handbook: Maple V Release 3*. New York: Springer-Verlag.
- SAS Institute, Inc. (1996) *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary, NC: SAS Institute Inc.

Seber, G. (1984) *Multivariate Observations*. New York: Wiley.

Wolfinger, R. (1996) Heterogeneous variance-covariance structures for repeated measures. *Journal of Agricultural, Biological, and Environmental Statistics*, 1:205-230.

Figure 1: Standardized Concentration Residuals

