THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

NLINVC USER'S GUIDE

by

Marcia L. Gumpertz

Institute of Statistics Mimeograph Series No. 1991

May 1991

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

NLINVC USER'S GUIDE

Version 1a. August 1990

ABSTRACT

The NLINVC macro procedure estimates parameters for nonlinear models with variance components by the method of estimated generalized least squares.

INTRODUCTION

The NLINVC procedure estimates parameters for models of the form:

$$y = f(X,\beta) + \sum_{i=1}^{k} U_i e_i$$

where $e_1,\ldots,e_k$ are independent random vectors with $E(e_i)=0$ and $Var(e_i)=\sigma_i^2 I$. The model for the mean, $f(X,\beta)$ should be nonlinear in the parameters and the parameters, β, are all assumed to be fixed. The matrices, U_i, correspond to random effects and are usually class variables such as replication or treatment x block. Each random effect, U_i, is associated with a variance component, σ_i^2.

The parameters of the nonlinear mean model are estimated using the method of estimated generalized least squares (EGLS). The procedure follows a three part algorithm:

1) Obtain an initial estimate of β by the ordinary least squares (OLS) approach using a modified Gauss-Newton algorithm.
2) Fix β at the value obtained in step 1) and estimate the variance components using an approximate maximum likelihood method or an approximate restricted maximum likelihood method.
3) Using the estimated variance-covariance matrix from step 2), compute the estimated generalized least squares estimate of β, again using a modified Gauss-Newton algorithm.

The user must specify the function for the mean model, initial values for the fixed parameters, partial derivatives of the mean function, a list of random factors, and the number of levels for each factor. Variance components can be estimated by approximate maximum likelihood or approximate restricted maximum likelihood. The user can also specify the number of iterations and adjust the convergence criterion for each step of the procedure. In addition to the printed output, NLINVC produces SAS data sets containing parameter estimates, variance components, predicted values and residuals.

NLINVC is written in the SAS macro facility. The syntax is similar, but not identical, to that of a regular SAS procedure. This program uses the SAS procedures NLIN, MIXMOD and IML; consequently the user has all the flexibility of NLIN for defining the mean function and doing auxiliary data manipulations. The required input follows that of PROCs NLIN and MIXMOD closely but not all options of those two procedures have been implemented in NLINVC.

PRELIMINARIES

In order to run the program NLINVC, you must have a copy of the procedure MIXMOD and the data library NGUMP.NLINVC.MACAUTOS. At North Carolina State University the JCL for use with the program NLINVC is as follows:

```
//jobcard
// EXEC SAST,REGION=2000K,OPTIONS='MAUTOSOURCE IMPLMAC
//  NOMRECALL'
//STEPLIB DD DSN=NFGG.MIXMOD.TTTT,DISP=SHR
//SASAUTOS DD DISP=SHR,DSN=NGUMP.NLINVC.MACAUTOS
//           DD DISP=SHR,DSN=SYSSAS.MACAUTOS.VERCUR
//any additional ddcards for input or output data sets
//SYSIN DD *
```

Please note that the SAS autocall library is called SYSSAS.MACAUTOS.VERCUR on the system at NCSU but may be called something else on your system. Explanation of the JCL can be found in the SAS Guide to Macro Processing, Version 5 Edition, expecially page 144 and Appendix 2.

Two other preliminary steps must be taken before running NLINVC:

1) If the input data set has any missing values, delete those observations from the data set. NLINVC uses PROC IML heavily, which does not handle missing values.
2) Issue a RUN command before issuing any NLINVC statements.

SPECIFICATIONS

The following statements are part of the NLINVC macro procedure. All statements are required.

NLINVC options;
PARMINIT parameter=value parameter=value ...;
AUXIL programming statement, programming statement,...;
FUNCTION dependent=expression;
DERIV DERparameter=expression, DERparameter=expression, ...;
VOPTIONS options;
RANDOM random effects;
LVLS values;
PRIORVC values;
DATAOUT options;
RNLINVC;

NLINVC statement

NLINVC options;

The NLINVC statement is required. The options below can appear in the NLINVC statement:

DATA=SASdataset

names the SAS data set containing the data to be analyzed by macro procedure NLINVC. If DATA= is omitted, the most recently created SAS data set is used.

OUTEST=SASdataset

names the SAS data set to contain the parameter estimates produced by NLINVC. If OUTEST is omitted, the data set _BETA is used.

OUTVC=SASdataset

names the SAS data set to contain the variance component estimates produced by NLINVC. If OUTVC is omitted, the data set _VC is used.

MAXITER=i

places a limit on the number of iterations NLINVC performs in estimation of the fixed parameters. The i value must be a positive integer. The default is 30.

CONVERGE=c

specifies the relative convergence criterion. The iterations are said to have converged if (LASTSSE-SSE)/(SSE+10E-6) < c. The default is c=10E-8. The constant c should be a small positive number.

PARMINIT statement

PARMINIT parameter=value ...;

A PARMINIT statement must follow the NLINVC statement. A parameter name and value must appear for every fixed parameter to be estimated. Specify only one value for each parameter. The parameter names must all be valid SAS names no longer than 5 characters and must not duplicate the names of any variables in the input data set. The values specify the starting values of the parameters.

AUXIL statement

AUXIL programming statement, programming statement, ...;

Any number of programming statements can be included in the AUXIL statement. Note that the programming statements are separated by commas, not semicolons. Use a semicolon to mark the end of the last programming statement. NLINVC can execute any statement that is acceptable to PROC NLIN, including assignment statements, explicitly subscripted ARRAY statements and references, IF statements and program control statements. For further details see the PROC NLIN description in the *SAS User's Guide: Statistics* (SAS Institute Inc. 1985).

FUNCTION statement

FUNCTION dependent=expression;

The FUNCTION statement declares the dependent variable and defines the function for the mean model. The expression can be any valid SAS expression and can include parameter names, variables in the input data set and variables created in the AUXIL statement. A FUNCTION statement must appear.

DERIV statement

DERIV DERparameter=expression, ...;

The DERIV statement is mandatory and a derivative definition, DERparameter=expression, for each fixed parameter to be estimated must appear in the DERIV statement. In each derivative definition the expression must be the algebraic representation of the partial derivative of the mean model given in the FUNCTION statement with respect to the parameter named in DERparameter. Separate the different derivative definitions with commas, and note that DERparameter has no period between DER and the parameter name, unlike PROC NLIN syntax.

VOPTIONS statement

VOPTIONS options;

The options below can appear in the VOPTIONS statement:

NORMALEQ=SASdataset

names the SAS data set which will contain the matrices involved in the normal equations for estimating β. If NORMALEQ= is omitted, the normal equations are stored in the data set _NEQ.

VARIT=i

puts a limit on the number of iterations NLINVC performs to estimate the variance components. The i value must be a positive integer. The default is 3.

The following two options specify which method of estimation for variance components is desired. Only one of these options should be specified. If neither is specified, approximate maximum likelihood will be used for estimating the variance components.

VAREQMML=SASdataset

specifies that the method of estimating the variance components is approximate modified maximum likelihood and names the SAS dataset to contain the matrices used in the variance estimating equations, $\left(\left(\mathrm{tr}(\hat{Q}V_i\hat{Q}V_j)\right)\right)$ and $\left(\left(y'\hat{Q}V_i\hat{Q}y\right)\right)$. If VAREQMML= is not specified, the matrices will be stored in the SAS data set _VEQMML.

VAREQML=SASdataset

specifies that the method of estimating the variance components is approximate maximum likelihood and names the SAS dataset to contain the matrices used in the variance estimating equations, $\left(\left(\mathrm{tr}(\hat{V}^{-1}V_i\hat{V}^{-1}V_j)\right)\right)$ and $\left(\left(y'\hat{Q}V_i\hat{Q}y\right)\right)$. If VAREQML= is not specified, the matrices will be stored in the SAS data set _VEQML.

RANDOM statement

RANDOM random effects;

The RANDOM statement lists the random effects in the model. The random effects are constructed from numeric variables in the input data set. Crossed and nested effects are both indicated by sets of variables joined by asterisks. Effects may be either discrete or continuous, but it is not possible to mix discrete and continuous variables in one effect, as in descrete*continuous. The rules for specifying effects follow the syntax for the MIXMOD MODEL statement. The RANDOM statement is required.

LVLS statement

LVLS values;

The LVLS statement specifies whether effects are continuous or discrete and specifies the maximum number of levels for each discrete effect. The model is stored with the fixed parameters of the nonlinear mean model first and the random effects following. The value list should include, in the order used in the PARMINIT and RANDOM statements, the number of levels for each effect in the model. Each fixed parameter in the nonlinear mean function is considered to be a continuous effect. Continuous effects are indicated by a 1 in the value list. For discrete effects the value should be the number of levels of the effect. See the MIXMOD LEVELS statement description for more details.

PRIORVC statement

PRIORVC values;

The PRIORVC statement allows the user to give initial values for the variance components. The value list should include a value for every random effect in the model including the residual error.

DATAOUT statement

DATAOUT options;

The DATAOUT statement allows the user to specify a SAS data set name for the output data and variable names for the residuals and predicted values. The available options are:

OUT=SASdataset

names the SAS data set to be created by NLINVC. If OUT= is not specified, the output is stored in data set _OUTDATA. The new data set includes all variables in the input data set plus residuals, predicted values, auxiliary variables created with the AUXIL statement and derivatives.

R=variable name

names a variable in the output data set to contain the residuals. The variable name must be a valid SAS name and must not match any variable name in the input data set. If R= is not specified, the residuals are stored under the name _RESID.

P=variable name

names a variable in the output data set to contain the predicted values. The variable name must be a valid SAS name and must not match any variable name in the input data set. If P= is not specified, the predicted values are stored under the name _YHAT.

RNLINVC statement

RNLINVC;

The RNLINVC statement runs the macros that estimate the variance components and the fixed parameters. This statement is required.

STATISTICAL METHOD

In the nonlinear model with random components, $y = f(X,\beta) + \sum_{i=1}^{k} U_i e_i$, where $e_1,\ldots,e_k$ are independent random vectors with $E(e_i)=0$ and Var $(e_i)=\sigma_i^2 I$, y has the special covariance structure, $V=Var(y)=\sum_{i=1}^{k} V_i\sigma_i^2$, where $V_i = U_iU_i'$. An estimated generalized least squares estimator, $\hat{\beta}_{GLS}$, is defined as any vector that minimizes the sum of squares, $(y - f(X,\beta))'\hat{V}^{-1}(y - f(X,\beta))$ with respect to β, where $\hat{V}$ is some previously computed estimate of V. If $\hat{V}$ is a strongly consistent estimator for V, then this estimator is strongly consistent for β and asymptotically normal and efficient under certain regularity conditions. It has the same asymptotic distribution as the generalized least squares estimator, which minimizes $(y - f(X,\beta))'V^{-1}(y - f(X,\beta))$ if V is known.

There are many possible methods for obtaining an EGLS estimator, each differing in the way the covariance matrix, V, is estimated. However all follow a three part algorithm.

1) Obtain an initial estimate, $\hat{\beta}_0$, of the fixed parameters.

2) Estimate the variance-covariance matrix, V.

3) Minimize $SSE=(y - f(X,\beta))'\hat{V}^{-1}(y - f(X,\beta))$ to obtain a final estimate, $\hat{\beta}_{EGLS}$.

In the macro procedure NLINVC the initial estimate, $\hat{\beta}_0$, of the fixed parameter vector, β, is computed using ordinary nonlinear least squares. For this first step the error structure represented by the random effects is ignored and the unweighted residual sums of squares, $(y - f(X,\beta))'(y - f(X,\beta))$, is minimized. This initial fitting is carried out in PROC NLIN with the option METHOD=GAUSS.

The variance components, σ_i^2, are estimated based on the residuals, $r_0 = y - f(X,\hat{\beta}_0)$, from the initial estimate of β. Under the assumption of normally distributed random effects, the log of the likelihood function is

$$L = \text{constant} - \tfrac{1}{2}\ln|V| - \tfrac{1}{2}(y - f(X,\beta))'V^{-1}(y - f(X,\beta)).$$

Consider the linear approximation of the mean function obtained by Taylor series expansion,

$$f(X,\beta) \doteq f(X,\hat{\beta}_0) + \frac{\partial f(X,\hat{\beta}_0)}{\partial \beta'}(\beta - \hat{\beta}_0).$$

Define $D_0 = \dfrac{\partial f(X,\hat{\beta}_0)}{\partial \beta'}$ and $\delta_0 = \beta - \hat{\beta}_0$. Then an approximate log likelihood function which can be maximized with respect to β and σ_i^2, i=1,...,k is

$$L \doteq \text{constant} - \tfrac{1}{2}\ln|V| - \tfrac{1}{2}(r_0 - D_0\delta)'V^{-1}(r_0 - D_0\delta).$$

Using the special structure of the covariance matrix, $V=\sum_{i=1}^{k} V_i\sigma_i^2$, this equation can be maximized iteratively by a method adapted from Hemmerle and Hartley's (1973) method for linear variance components models, which yields the following estimating equations:

$$\Big(\Big(\text{trace}(V_i V_{(h)}^{-1} V_j V_{(h)}^{-1})\Big)\Big)\Big(\Big(\hat{\sigma}_{j(h+1)}^2\Big)\Big) = \Big(\Big(r_0' Q_{(h)} V_i Q_{(h)} r_0\Big)\Big),$$

where h indicates the iteration number, oversize double parentheses indicate a matrix with elements of the enclosed form, and

$$Q_{(h)} = V_{(h)}^{-1}(I - D_0(D_0'V_{(h)}^{-1}D_0)^{+}D_0'V_{(h)}^{-1}),$$

where $(\ )^{+}$ indicates a generalized inverse. Solutions to these equations, achieved when further iteration does not increase the log likelihood, may not be positive, in which case they are reset to zero before the next iteration. These equations are iteratively produced by PROC MIXMOD and solved in PROC IML until either the convergence criterion is satisfied or the maximum number of iterations specified is reached. The approximate covariance matrix of the variance components, based on the variance of a quadratic form of a normally distributed random vector with mean zero, is computed as:

$$\hat{V}\left(\left(\hat{\sigma}_i^2\right)\right) = 2\left(\left(\text{trace}(\hat{V}^{-1}V_i\hat{V}^{-1}V_j)\right)\right)\left(\left(\text{trace}(\hat{Q}V_i\hat{Q}V_j)\right)\right)\left(\left(\text{trace}(\hat{V}^{-1}V_i\hat{V}^{-1}V_j)\right)\right),$$

where $\hat{V}$ is the final estimate of V.

Variance components can also be estimated by approximate modified maximum likelihood, in which the likelihood function is based on vectors in the error space, that is, on linear combinations of y which have expectation zero, rather than on y itself. The macro procedure NLINVC uses the linear approximation $r_0 \doteq D_0\delta + e$ to obtain vectors in the error space. Vectors of the form $k'y$, where k is chosen so that $k'D_0 = 0$, then fall in this linear approximation to the error space. The log likelihood function of $K'y$, where K is a full rank matrix of vectors k, is:

$$L = \text{constant} - \tfrac{1}{2}\ln|K'VK| - \tfrac{1}{2}(K'y - K'f(X,\beta))'(K'VK)^{-1}(K'y - K'f(X,\beta)),$$

which is approximated by:

$$L \doteq \text{constant} - \tfrac{1}{2}\ln|K'VK| - \tfrac{1}{2}r_0{}'K(K'VK)^{-1}K'r_0.$$

This restricted likelihood function can be maximized by iteratively solving the system of equations:

$$\left(\left(\text{trace}(Q_{(h)}V_iQ_{(h)}V_j)\right)\right)\left(\left(\hat{\sigma}^2_{j(h+1)}\right)\right) = \left(\left(y'Q_{(h)}V_iQ_{(h)}y\right)\right).$$

As in the approximate maximum likelihood algorithm, if this procedure results in any negative variance estimate, that variance component is set to zero and then iterations are resumed. For modified maximum likelihood the covariance matrix of the variance components is approximately

$$\hat{V}\left(\left(\hat{\sigma}_i^2\right)\right) = 2\left(\left(\text{trace}(\hat{Q}V_i\hat{Q}V_j)\right)\right).$$

Finally, $\hat{\beta}_{EGLS}$ is computed using a modified Gauss-Newton algorithm to minimize $SSE=(y - f(X,\beta))'\hat{V}^{-1}(y - f(X,\beta))$ with respect to β. The variance components, $\hat{\sigma}_i^2$, obtained in the previous step are used to form $\hat{V}=\sum_{i=1}^{k} V_i\hat{\sigma}_i^2$. The Gauss-Newton algorithm for generalized least squares is similar to that for ordinary least squares, essentially approximating the sum of squares surface with the ellipsoid $(r_0 - D_0\hat{\delta})'\hat{V}^{-1}(r_0 - D_0\hat{\delta})$. The approximate normal equations, $D_0{}'\hat{V}^{-1}D_0\hat{\delta} = D_0{}'\hat{V}^{-1}r_0$, are solved, then the program checks to ensure that the new sum of squares, $SSE_{(1)} = (y - f(X,\hat{\beta}_{(1)}))'\hat{V}^{-1}(y - f(X,\hat{\beta}_{(1)}))$ is less than $SSE_{(0)}$. If it is not, the update vector $\hat{\delta}$ is halved (up to ten times) until SSE is smaller than $SSE_{(0)}$. These approximate normal equations are produced by PROC MIXMOD, solved in PROC IML and the process is iterated until convergence. If the convergence criterion is satisfied or if halving δ does not provide any improvement, so that SSE is assumed to be at a minimum, the standard errors and Wald-type confidence intervals, $\hat{\beta} \pm t(.975,df)se(\hat{\beta})$, of the fixed parameter estimates are printed out. The variance-covariance matrix of $\hat{\beta}_{EGLS}$ is estimated as

$$\hat{V}(\hat{\beta}_{EGLS}) = (\hat{D}'\hat{V}^{-1}\hat{D})^{-1},$$

where $\hat{D}$ indicates the matrix of partial derivatives evaluated at $\hat{\beta}_{EGLS}$.

The estimated standard errors for the ordinary least squares estimates are also printed. If the estimates of the variance components have high variance, for instance if they are computed based on a small number of replicates, then $\hat{\beta}_{OLS}$ may be preferred over $\hat{\beta}_{EGLS}$. In this case the variance-covariance matrix of $\hat{\beta}_{OLS}$ is estimated as

$$\hat{V}(\hat{\beta}_{OLS}) = (D_0{'}D_0)^{-1}D_0{'}\hat{V}D_0(D_0{'}D_0)^{-1},$$

where the partial derivatives are computed using $\hat{\beta}_{OLS}$. If, on the other hand, the variance components are estimated well, then $\hat{\beta}_{EGLS}$ has smaller variance than the ordinary least squares estimator. In this case the variance-covariance matrix of $\hat{\beta}_{OLS}$ can be estimated as

$$\hat{V}(\hat{\beta}_{OLS}) = (\hat{D}'\hat{D})^{-1}\hat{D}'\hat{V}\hat{D}(\hat{D}'\hat{D})^{-1},$$

where the matrix of partial derivatives is evaluated at $\hat{\beta}_{EGLS}$. Then the estimated standard error of $\hat{\beta}_{OLS}$ based on this covariance estimate can be compared to the estimated standard error of $\hat{\beta}_{EGLS}$ to get an indication of how much precision is gained by using estimated generalized least squares over ordinary nonlinear least squares.

DETAILS

NLINVC Output

The three steps of the EGLS algorithm each produce output. The first step prints the title 'INITIAL ESTIMATE OF FIXED PARAMETERS USING OLS' at the top of each page. The output from this step is produced by PROC NLIN and the user should be aware that the estimated standard errors and correlation matrix produced by PROC NLIN are not correct for the nonlinear model with variance components. The correct asymptotic standard errors for $\hat{\beta}_{OLS}$ appear in the output for step 3. The second step has pages entitled 'ESTIMATE VARIANCE COMPONENTS' and the third step of the procedure has pages entitled 'ESTIMATE FIXED PARAMETERS'.

Troubleshooting

NLINVC contains macro invocations, so in general it will be necessary to print the SAS statements generated by the macros in order to pinpoint where an error has occurred. This is accomplished by using the options MPRINT and SYMBOLGEN either in an OPTIONS statement or on the EXEC SAS card. Use of these options generates several pages of log output. Even without these options, the log produced by NLINVC may be longer than desired. It can be reduced somewhat by using the NONOTES option.

NLINVC may take more time than the default time limit. Experience indicates that TIME=1 will often be a good starting value.

Global Macro Variables

The following global macro variables are created by NLINVC commands. If any user-defined macro variables are given names from this list prior to running NLINVC unexpected results may occur. The contents of these macro variables are available to the user after running NLINVC.

Variable Name	Contents	Default
INDATA	name of input data set	current
EST	name of data set to store $\hat{\beta}$	_BETA
VC	name of data set to store $\hat{\sigma}_i^2$	_VC
MAXIT	number of iterations for estimating β	30
EPSBETA	convergence criterion for β	10E-8

Variable Name	Contents	Default
PARMNAME	list of parameter names	
DERLIST	list of parameter names with prefix DER	
NF	number of fixed parameters	
PARMDEF	list of initial parameter definitions (parameter=value parameter=value ...)	
AUXILDEF	programming statements to perform auxiliary data manipulations (statement 1; statement 2; ...)	
FNDEF	mean function definition (dependent=expression)	
DERDEF	derivative definitions (DERparameter=expression; ...)	
OUTDATA	name of output data set for data, residuals and predicted values	_OUTDATA
RESID	name of variable to store residuals	_RESID
PRED	name of variable to store pred. values	_YHAT
NORMEQ	data set name to store normal equations	_NEQ
VAREQ	data set name for var. est. equations	_VEQML
VMETHOD	var. component est. method (ML or MML)	ML
VARITER	number of iterations for estimating $\hat{\sigma}_i^2$	3
VEQMML	data set name for MML var. est. equations	_VEQMML
VEQML	data set name for ML var. est. equations	_VEQML
ULIST	list of random effects	
NR	number of random effects (incl. residual)	
ULISTQ	ULIST with each term enclosed in quotes	
UVARLIST	list of variables included in random effects	

Data Sets Created by NLINVC

The NLINVC statement options create the following data sets. Examples of these data sets are provided in the next section. In all cases the data sets are created regardless of whether or not the user specifies names for them. The default data set names are given in parentheses. For more information about the NORMALEQ=, VAREQMML= and VAREQML= data sets see the MIXMOD user guide (Giesbrecht 1985).

statement	option (default dsname)	contents
NLINVC	OUTEST=(_BETA)	parameter estimates, update vector, δ, and SSE from all iterations
NLINVC	OUTVC=(_VC)	variance component estimates from all iterations
DATAOUT	OUT=(_OUTDATA)	input variables, auxiliary variables, derivatives evaluated at $\hat{\beta}_{EGLS}$, predicted values and residuals
VOPTIONS	NORMALEQ=(_NEQ)	columns of $\hat{D}'\hat{V}^{-1}\hat{D}$ (labeled XVIX) and $\hat{D}'\hat{V}\hat{r}$ (labeled XVIY), n (last entry of XVIX_001) and $\hat{r}'\hat{V}^{-1}\hat{r}$ (last entry in XVIY)
VOPTIONS	VAREQMML=(_VEQMML)	columns of $\left(\left(\text{tr }\hat{Q}V_i\hat{Q}V_j\right)\right)$ (labeled QVQV) and $\left(\left(y'\hat{Q}V_i\hat{Q}y\right)\right)$ (labeled YQVQY); values of the variance components and the ln(likelihood) are stored in that order in the row labeled PRIORS

statement	option (default dsname)	contents
VOPTIONS	VAREQML=(_VEQML)	columns of $\left(\left(\text{tr } \hat{V}^{-1}V_i\hat{V}^{-1}V_j\right)\right)$ (labeled MLEQ) and $\left(\left(y'\hat{Q}V_i\hat{Q}y\right)\right)$ (labeled YQVQY); values of variance components and the ln(likelihood) are stored in that order in the row labeled PRIORS

The following data sets are created by the RNLINVC statement and the user does not have the option of naming them. Except for the data set _OUT0, they store results of the last iteration only.

data set	contents
_OUT0	input variables, auxiliary variables, derivatives, predicted values, and residuals evaluated at $\hat{\beta}_{OLS}$
_SIG2	variance component estimates
_VCI	variance components and ln(likelihood)
_SSEI	SSE
_BETAI	parameter estimates and update values, δ

EXAMPLE

Weibull Curve with Random Year and Block Effects

This example demonstrates use of NLINVC for combining data from different experiments. Studies of the effect of ozone exposure on soybean yield were conducted in 1982, 1984 and 1986. In two of the years randomized block designs were used, but in the third year the experiment was completely randomized. The experiments differed further in that only one moisture treatment was used in 1982 (well watered); in 1984 a second moisture treatment was added (water stressed); and in 1986 a third moisture treatment involving rain exclusion caps was added. In this example the effect of ozone exposure (x) on soybean yield (y) is modelled with a Weibull curve with separate intercept parameters (α) for different watering regimes. Year and block(year) are included as random effects. The model for the i^{th} year, the j^{th} block within year and the k^{th} plot within a block is:

$$y_{ijk}=(\alpha+ \alpha_2 mdum_{2ijk}+\alpha_3 mdum_{3ijk})\exp\{-(x_{ijk}/w)^{\lambda}\}+e_{1i}+e_{2ij}+e_{3ijk}$$

where $mdum_2$ and $mdum_3$ are dummy variables for the second and third moisture treatments, e_{1i} is the year effect, e_{2ij} is the block(year) effect and e_{3ijk} is the within-block error.

```
OPTIONS NONOTES;
DATA SOYBEAN;
  INFILE DATA;
  INPUT ID $ CULTIVAR $ BLOCK OZONE $ SULFUR $ MOISTURE $ O7HR O12HR
  SO2 KG_HA SEEDWT COV;
  Y=SQRT(KG_HA);
  IF ID='R82SO' THEN DO;
    YEAR=2; BLK=BLOCK; END;
  IF ID='R84SO' THEN DO;
    YEAR=4; BLK=BLOCK; END;
  IF ID='R86SO' THEN DO;
    YEAR=6; BLK=1; END;
  MDUM1=(MOISTURE='-');
  MDUM2=(MOISTURE='+');
  MDUM3=(MOISTURE='T');
  RENAME O12HR=X;
  DROP ID CULTIVAR BLOCK OZONE SULFUR O7HR SO2 KG_HA SEEDWT COV;
PROC PRINT DATA=SOYBEAN;
```

OBS	MOISTURE	X	Y	YEAR	BLK	MDUM1	MDUM2	MDUM3
1	+	0.054	60.0833	2	1	0	1	0
2	+	0.065	56.6701	2	1	0	1	0
3	+	0.081	50.5791	2	1	0	1	0
⋮								
69	T	0.088	67.3976	6	1	0	0	1

```
RUN;
NLINVC MAXITER=10;
  PARMINIT ALPHA=75 AM2=50 AM3=50 OMEGA=0.10 LAMBD=2;
  AUXIL
    E=EXP(-(X/OMEGA)**LAMBD),
    F=(ALPHA+AM2*MDUM2+AM3*MDUM3)*E;
  DERIV
    DERALPHA=E, DERAM2=MDUM2*E, DERAM3=MDUM3*E,
    DEROMEGA=F*(LAMBD/OMEGA)*((X/OMEGA)**LAMBD),
    DERLAMBD=-F*((X/OMEGA)**LAMBD)*LOG(X/OMEGA);
  FUNCTION Y=F;
  DATAOUT OUT=OUTDATA R=RESID P=YHAT;
  VOPTIONS NORMALEQ=NEQ VAREQML=VEQ VARIT=5;
  RANDOM YEAR YEAR*BLK;
  LVLS 1 1 1 1 1 3 5;
  PRIORVC 1 1 1;
  RNLINVC;
PROC PRINT DATA=_BETA; TITLE 'LIST DATA SET _BETA';
PROC PRINT DATA=_VC; TITLE 'LIST DATA SET _VC';
PROC PRINT DATA=OUTDATA; TITLE 'LIST DATA SET OUTDATA';
PROC PRINT DATA=NEQ; TITLE 'LIST DATA SET NEQ';
PROC PRINT DATA=VEQ; TITLE 'LIST DATA SET VEQ';
PROC PRINT DATA=_VEQMML; TITLE 'LIST DATA SET _VEQMML';
```

INITIAL ESTIMATE OF FIXED PARAMETERS USING OLS

NON-LINEAR LEAST SQUARES ITERATIVE PHASE

DEPENDENT VARIABLE: Y METHOD: GAUSS-NEWTON

ITERATION	ALPHA	AM2	AM3	OMEGA	LAMBD	RESIDUAL SS
0	75.0000	50.0000	50.0000	0.1000	2.0000	49628.335919008
1	77.0177	1.3382	19.9920	0.1261	1.3341	8666.514323055
2	77.9718	3.2105	21.3487	0.1756	0.9175	7292.033800502
3	78.7180	3.0706	21.4851	0.2214	0.8762	6041.735933834
4	78.8261	3.0514	21.4986	0.2344	0.8694	5992.522121252
5	78.8841	3.0466	21.5098	0.2356	0.8663	5992.391782962
6	78.8750	3.0458	21.5076	0.2355	0.8668	5992.391738169

NOTE: CONVERGENCE CRITERION MET.

NON-LINEAR LEAST SQUARES SUMMARY STATISTICS DEP VARIABLE Y

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE
REGRESSION	5	286259.49326	57251.89865
RESIDUAL	64	5992.39174	93.63112
UNCORR TOTAL	69	292251.88500	
(CORR TOTAL)	68	10802.76591	

PARAMETER	ESTIMATE	ASYMPTOTIC STD. ERROR	ASYMPTOTIC 95 % CONFIDENCE INTERVAL LOWER	UPPER
ALPHA	78.87495533	16.441819370	46.028610878	111.72129978
AM2	3.04582683	3.467822112	-3.881951484	9.97360513
AM3	21.50755168	5.779685044	9.961291409	33.05381195
OMEGA	0.23549555	0.132140182	-0.028485083	0.49947619
LAMBD	0.86677640	0.825411199	-0.782173747	2.51572656

INITIAL ESTIMATE OF FIXED PARAMETERS USING OLS

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

CORR	ALPHA	AM2	AM3	OMEGA	LAMBD
ALPHA	1.0000	-.0092	.4588	.8052	-.9591
AM2	-.0092	1.0000	.4509	.1110	-.1263
AM3	.4588	.4509	1.0000	.4462	-.5320
OMEGA	.8052	.1110	.4462	1.0000	-.9279
LAMBD	-.9591	-.1263	-.5320	-.9279	1.0000

ESTIMATE VARIANCE COMPONENTS

VCRECORD	ITERATION	YEAR	YEAR*BLK	ERROR	LNLR
	0	1	1	1	-522.775
	1	66.08578	9.849002	11.76968	-192.098
	2	85.66606	0.8973977	11.79198	-191.069
	3	89.84488	0.478754	11.77929	-191.047
	4	89.98436	0.5181502	11.77074	-191.046
CONVERGENCE	5	89.98599	0.5133806	11.77174	-191.046
CRITERION MET					

APPROXIMATE COVARIANCE MATRIX OF THE VARIANCE COMPONENTS

VARVC	YEAR	YEAR*BLK	ERROR
YEAR	3658.475	-1.99244	0.2788725
YEAR*BLK	-1.99244	2.98767	-0.409736
ERROR	0.2788725	-0.409736	4.056032

ESTIMATE FIXED PARAMETERS

SUM-MARY	ITER	SUBIT	ALPHA	AM2	AM3	OMEGA	LAMBD	SSE
ROW1	0	0	78.8750	3.0458	21.5076	0.2355	0.8668	182.17503
ROW2	1	0	46.6501	5.6031	1.0128	0.00206	2.4086	474.56634
ROW3	1	1	62.7625	4.3244	11.2602	0.1188	1.6377	110.83753
ROW4	2	0	63.6058	6.0229	6.7815	0.1362	2.2739	75.742882
ROW5	3	0	66.8697	5.7957	7.3275	0.1300	2.3219	72.667594
ROW6	4	0	66.7780	5.8108	7.3489	0.1306	2.3140	72.649527
ROW7	5	0	66.7854	5.8088	7.3470	0.1306	2.3153	72.649506
ROW8	6	0	66.7874	5.8088	7.3469	0.1306	2.3148	72.649504

CONVERGENCE CRITERION MET

PARAM	ESTIMATE	STD ERROR	LCL95	UCL95
ALPHA	66.78744680	5.77652861	55.24750365	78.32738994
AM2	5.80878848	1.20617598	3.39917482	8.21840215
AM3	7.34685956	1.66245165	4.02573041	10.66798872
OMEGA	0.13064281	0.01195526	0.10675943	0.15452618
LAMBD	2.31479635	0.39104709	1.53358994	3.09600276

ASYMPTOTIC CORRELATION MATRIX OF THE PARAMETERS

CORRB	ALPHA	AM2	AM3	OMEGA	LAMBD
ALPHA	1.0000	-.1866	-.1280	.5892	-.2562
AM2	-.1866	1.0000	.3696	-.02156	-.01085
AM3	-.1280	.3696	1.0000	-.03725	.01458
OMEGA	.5892	-.02156	-.03725	1.0000	-.8498
LAMBD	-.2562	-.01085	.01458	-.8498	1.0000

ESTIMATE FIXED PARAMETERS

ASYMPTOTIC STANDARD ERRORS AND CORRELATION MATRIX
OF ORDINARY LEAST SQUARES ESTIMATE, BOLS

1) COMPUTED USING DERIVATIVES EVALUATED AT BOLS

PARAM	EST	STD ERROR	CORR	ALPHA	AM2	AM3	OMEGA	LAMBD
ALPHA	78.8750	20.3897	ALPHA	1.000	-.404	.820	.921	.920
AM2	3.0458	6.3358	AM2	-.404	1.000	.016	-.089	.140
AM3	21.5076	12.4844	AM3	.820	.016	1.000	.932	-.952
OMEGA	0.2355	0.1450	OMEGA	.921	-.089	.932	1.000	-.980
LAMBD	0.8668	0.9448	LAMBD	-0.920	.140	-.952	-.980	1.000

2) COMPUTED USING DERIVATIVES EVALUATED AT BEGLS

PARAM	EST	STD ERROR	CORR	ALPHA	AM2	AM3	OMEGA	LAMBD
ALPHA	78.8750	8.5446	ALPHA	1.000	-.683	.227	.572	-.529
AM2	3.0458	5.5764	AM2	-.683	1.000	.087	-.251	.299
AM3	21.5076	7.6719	AM3	.227	.087	1.000	.870	-.862
OMEGA	0.2355	0.0384	OMEGA	.572	-.251	.870	1.000	-.985
LAMBD	0.8668	1.3880	LAMBD	-.529	.299	-.862	-.985	1.000

LIST DATA SET _BETA

OBS	ALPHA	AM2	AM3	OMEGA	LAMBD	OLD1	DELTA1	OLD2	DELTA2
1	78.875	3.0458	21.508	0.236	0.8668	.	.	.	.
2	46.650	5.6031	1.013	0.002	2.4086	78.8750	-32.225	3.04583	2.5572
3	62.763	4.3244	11.260	0.119	1.6377	78.8750	-16.112	3.04583	1.2786
4	63.606	6.0229	6.782	0.136	2.2740	62.7625	0.843	4.32444	1.6984
5	66.870	5.7957	7.328	0.130	2.3219	63.6058	3.264	6.02285	-0.2272
6	66.778	5.8108	7.349	0.131	2.3140	66.8697	-0.092	5.79569	0..0151
7	66.785	5.8088	7.347	0.131	2.3153	66.7780	0.007	5.81082	-0.0020
8	66.787	5.8088	7.347	0.131	2.3148	66.7854	0.002	5.80883	-0.0000

OBS	OLD3	DELTA3	OLD4	DELTA4	OLD5	DELTA5	ITER	SUBIT	SSE
1	.	.	.	.	.	.	0	0	182.175
2	21.5076	-20.495	0.2355	-0.2334	0.8668	1.5418	1	0	474.566
3	21.5076	-10.247	0.2355	-0.1167	0.8668	0.7709	1	1	110.838
4	11.2602	-4.479	0.1188	0.0175	1.6377	0.6363	2	0	75.743
5	6.7815	0.546	0.1362	-0.0063	2.2740	0.0480	3	0	72.668
6	7.3275	0.021	0.1300	0.0007	2.3219	-0.0080	4	0	72.650
7	7.3489	-0.002	0.1306	-0.0000	2.3140	0.0014	5	0	72.650
8	7.3470	-0.000	0.1306	0.0000	2.3153	-0.0005	6	0	72.650

LIST DATA SET _VC

OBS	ITER	U1	U2	U3	LNLR	LASTLNLR	NOTES
1	0	1.0000	1.00000	1.0000	-522.78	.	
2	1	66.0858	9.84900	11.7697	-192.10	-522.78	
3	2	85.6661	0.89740	11.7920	-191.07	-192.10	
4	3	89.8449	0.47875	11.7793	-191.05	-191.07	
5	4	89.9844	0.51815	11.7707	-191.05	-191.05	
6	5	89.9860	0.51338	11.7717	-191.05	-191.05	

LIST DATA SET OUTDATA

OBS	ALPHA	AM2	AM3	OMEGA	LAMBD	OLD1	DELTA1
1	66.7874	5.80879	7.34686	0.130643	2.3148	66.7854	0.00205844
2	66.7874	5.80879	7.34686	0.130643	2.3148	66.7854	0.00205844
3	66.7874	5.80879	7.34686	0.130643	2.3148	66.7854	0.00205844
⋮							
69	66.7874	5.80879	7.34686	0.130643	2.3148	66.7854	0.00205844

OBS	OLD2	DELTA2	OLD3	DELTA3	OLD4	DELTA4
1	5.80883	-0.000039929	7.347	-0.00013741	0.130629	0.0000137185
2	5.80883	-0.000039929	7.347	-0.00013741	0.130629	0.0000137185
3	5.80883	-0.000039929	7.347	-0.00013741	0.130629	0.0000137185
⋮						
69	5.80883	-0.000039929	7.347	-0.00013741	0.130629	0.0000137185

OBS	OLD5	DELTA5	ITER	SUBIT	MOIST	X	Y	YEAR	BLK	MDUM(	2	3)	E
1	2.31533	-0.00053346	6	0	+	0.054	60.0833	2	1	0	1	0	0.878649
2	2.31533	-0.00053346	6	0	+	0.065	56.6701	2	1	0	1	0	0.819788
3	2.31533	-0.00053346	6.	0	+	0.081	50.5791	2	1	0	1	0	0.718414
⋮													
69	2.31533	-0.00053346	6	0	T	0.088	67.3976	6	1	0	0	1	0.669879

OBS	F	DERALPHA	DERAM2	DERAM3	OMEGA	LAMBD	YHAT	RESID
1	63.7866	0.878649	0.878649	0	146.214	7.29054	63.7866	-3.703
2	59.5135	0.819788	0.819788	0	209.537	8.25542	59.5135	-2.843
3	52.1541	0.718414	0.718414	0	305.607	8.24479	52.1541	-1.575
⋮								
69	49.6610	0.669879	0	0.669879	352.547	7.86194	49.6610	17.7366

LIST DATA SET NEQ

OBS	VAR_NAME	ID	XVIX_001	XVIX_002	XVIX_003	XVIX_004	XVIX_005	XVIY
1	SET_001	1	0.072	0.046	0.0106	-46.1	-0.92	-0.0000
2	SET_002	1	0.046	0.827	-0.2062	-23.1	-0.39	-0.0000
3	SET_003	1	0.011	-0.206	0.4211	-3.9	-0.09	0.0000
4	SET_004	1	-46.066	-23.103	-3.9330	54619.9	1244.10	0.0104
5	SET_005	1	-0.922	-0.385	-0.0949	1244.1	35.36	0.0015
6	SUMMARY		69.000	.	.	.	.	72.6495

LIST DATA SET VEQ

OBS	MLEQ_01	MLEQ_02	MLEQ_03	YQVQY	CMP_NAME
1	3.63274E-04	2.42211E-04	1.68657E-05	3.30126E-02	SIGMA_01
2	2.42211E-04	6.76251E-01	6.85245E-02	1.17600E+00	SIGMA_02
3	1.68657E-05	6.85245E-02	4.69019E-01	5.55785E+00	ERROR
4	8.99860E+01	5.13381E-01	1.17717E+01	-1.91046E+02	PRIORS
5	3.00000E+00	5.00000E+00	6.90000E+01	.	LEVELS

LIST DATA SET _VEQMML

OBS	QVQV_01	QVQV_02	QVQV_03	YQVQY	CMP_NAME
1	2.41313E-04	1.60921E-04	2.74879E-05	3.30126E-02	SIGMA_01
2	1.60921E-04	6.73475E-01	6.84576E-02	1.17600E+00	SIGMA_02
3	2.74879E-05	6.84576E-02	4.39967E-01	5.55785E+00	ERROR
4	8.99860E+01	5.13381E-01	1.17717E+01	-1.88879E+02	PRIORS
5	3.00000E+00	5.00000E+00	6.90000E+01	.	LEVELS