A SAS® Macro to Find the Best Fitted Model for Repeated Measures Using PROC MIXED

Jingjing Wang, U.S. Army Institute of Surgical Research, Ft. Sam Houston, TX

ABSTRACT
Specifying an appropriate covariance structure is essential for a valid and powerful repeated measure analysis using PROC MIXED. However, assessing the covariance structure is not a trivial task. To facilitate this task, the SAS® macro %RMMS was developed to find the best fitted model accurately and efficiently. The macro produces a table of various valid candidate covariance structures sorted in ascending order by AIC, AICC and BIC, respectively, with the covariance having the smallest Information Criteria (IC) specified, and a plot of estimated covariance versus duration between time points to examine the covariance structure. In addition, an optional likelihood ratio test between two user-selected nested covariance structures is provided. Moreover, once the final decision is made about the covariance structure, output of the final analysis of the best fitted model is automatically exported to a user-specified rich text format (RTF) file.

KEYWORDS: MIXED, covariance structure, repeated measures, likelihood ratio test

INTRODUCTION
Mixed model methodology in a two-stage approach using the MIXED procedure has been widely used as the modern approach for repeated measures. The first stage attempts to estimate the covariance structure and the second stage substitutes the covariance estimates into the mixed model and uses generalized least square methodology to assess the effects. This approach is appealing because 1) it directly addresses the covariance structure and attempts to accommodate it in the model; 2) it can accommodate data that are missing at random.

Specifying an appropriate covariance structure is essential to draw a valid and powerful conclusion. Ignoring important correlations by using a model that is too simple will increase the risk of the type I error of the fixed effects tests. On the other hand, using a model which is too complex will reduce the power and efficiency for the fixed effects tests. However, assessing and estimating the covariance structure requires a series of steps. It involves using IC such as AICC and BIC to compare the fit of various covariance structures by the rule “smaller is better,” and/or using likelihood ratio tests to test whether a more general model is significantly better than its special case (i.e., nested), and/or using graphics to examine the patterns of covariates over levels of repeated variables (e.g., time). Therefore, it is desirable to have access to a macro that automatically generates summaries by applying the above statistical tools so that we can find the most appropriate covariance structure quickly. It is also desirable that the macro can export the final analysis output to a specified RTF file automatically. %RMMS was developed to meet the above desirability, and to make the analysis of repeated measures less burdensome.

FEATURES OF THE MACRO
The macro automatically generates the following information:
1. A table of various valid candidate covariance structures sorted in ascending order by AIC, AICC and BIC, respectively, with the covariance having the smallest IC specified.
2. A plot of estimated covariance among time points within a subject versus duration in successive time points to examine the covariance structure.
3. A table of the likelihood ratio test between two nested covariance structures if you specify a more general structure and its special case in CovG = and CovS =, respectively, in the macro.
4. Output of the final analysis of the best fitted model to a specified RTF file once the final covariance structure is specified FinalCov = in the macro.

MACRO STRUCTURE
The macro syntax is:
%RMMS
(Dt =,
 Y =,
 Xs =,
 ClassXs =),
RepeatedVar =,
Sub =,
CovStructures =,
NumRept =,
CovG = |,
CovS = |,
FinalCov = |,
Style =,
OutFile =);

There are several macro variables used in %RMMS that need to be defined in the call. An explanation of these macro variables is provided below.

<table>
<thead>
<tr>
<th>MACRO VARIABLE</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dt</td>
<td>Input data set in the univariate format</td>
</tr>
<tr>
<td>Y</td>
<td>Dependent variable in the MIXED model</td>
</tr>
<tr>
<td>Xs</td>
<td>Fixed effects in the MIXED model, separated by blank</td>
</tr>
<tr>
<td>ClassXs</td>
<td>Class variables in the CLASS Statement in the MIXED model separated by blank</td>
</tr>
<tr>
<td>RepeatedVar</td>
<td>Repeated variable in the REPEATED statement in the MIXED model, for example, time</td>
</tr>
<tr>
<td>Sub</td>
<td>The variable for SUBJECT= option in the MIXED model</td>
</tr>
<tr>
<td>CovStructures</td>
<td>Covariance structures separated by</td>
</tr>
<tr>
<td>NumRept</td>
<td>Total number of levels in repeated variable</td>
</tr>
<tr>
<td>CovG*</td>
<td>A more general covariance structure compared to CovS as below followed by</td>
</tr>
<tr>
<td>CovS*</td>
<td>A special (nested) case of CovG followed by</td>
</tr>
<tr>
<td>FinalCov</td>
<td>Final selected covariance structure followed by</td>
</tr>
<tr>
<td>OutFile</td>
<td>File directory for the output of final model in RTF, for example, D:\stat\y.rtf</td>
</tr>
</tbody>
</table>

* CovG and CovS are for an optional likelihood ratio test between two user-selected nested covariance structures, they must be listed in the macro variable CovStructures in the call.

DETAIL OF PROGRAM

The code for the macro is available in the Appendix. The following is a description of the logic used in the macro.

**Step 1.** Create a series of data sets named Fitdt&&p (p=1,2,3,…, p<= # of candidate structures), FitRF0&&p containing fit statistics and ParmRF&&p containing both fit statistics and dimensions from the PROC MIXED model under each valid covariance structure using DO-WHILE loop.

Some candidate covariance structures may be invalid for the PROC MIXED because the procedures under them have non-converged iterations. These invalid candidate covariance structures must be excluded from the model selection first. To do this, the macro variable named Convs containing converge status of a mixed model under each candidate structure is created.

**Step 2.** Create data set S by combining Fitdt&&p after the DO-WHILE loop and sorted by names of fit statistics and their values.

**Step 3.** Create data set Covparms containing estimated covariance by PROC MIXED under UN structure and data set named Time containing duration between time points. Then generate data set named Covplot by merging these two data sets.

**Step 4.** Create data set named User by merging ParmRF&&p and itRF0&&p and calculate likelihood ratio test for a user defined more general structure and its special structure.

**Step 5.** Automatically export the final model using the specified most appropriate structure to a specified RTF file.
EXAMPLE
The usage of the macro is illustrated through an example in this section. Assume data were from a study with one standard drug treatment (1) and one experimental drug treatment (2). The two treatments were randomly assigned to 30 subjects and pulse was measured at equally spaced time points 1, 2, 3 and 4 for each subject.

The data set example includes a dependent variable (pulse), two independent variables (drug and time). A partial listing of the data is as follows:

<table>
<thead>
<tr>
<th>id</th>
<th>drug</th>
<th>time</th>
<th>pulse</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>85</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>2</td>
<td>85</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>3</td>
<td>88</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>92</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>3</td>
<td>93</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2</td>
<td>97</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>94</td>
</tr>
</tbody>
</table>

The following macro will automatically generate fit statistics for all valid candidate covariance structures listed in ascending order (Table 1) and the plot of covariance vs. duration between time points (Figure 1).

\%
RMMS
\(\text{Dt}=\) example,
\(\text{Y}=\) pulse,
\(\text{Xs}=\) drug time drug*time,
\(\text{ClassXs}=\) drug time,
\(\text{RepeatedVar} = \) time,
\(\text{Sub} = \) id,
\(\text{CovStructures} = \) TOEP|AR(1)|ARH(1)|CS|CSH|HF|ARMA(1,1)|UN,
\(\text{NumRept} = 3\)

<table>
<thead>
<tr>
<th>Descr</th>
<th>Value</th>
<th>type</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIC (smaller is better)</td>
<td>586.6</td>
<td>HF</td>
<td>Smallest</td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>587.8</td>
<td>ARH(1)</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>589.7</td>
<td>UN</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>594.1</td>
<td>AR(1)</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>594.8</td>
<td>CS</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>595.7</td>
<td>TOEP</td>
<td></td>
</tr>
<tr>
<td>AIC (smaller is better)</td>
<td>595.7</td>
<td>ARMA(1,1)</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>587.1</td>
<td>HF</td>
<td>Smallest</td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>588.3</td>
<td>ARH(1)</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>590.9</td>
<td>UN</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>594.3</td>
<td>AR(1)</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>595.0</td>
<td>CS</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>596.0</td>
<td>TOEP</td>
<td></td>
</tr>
<tr>
<td>AICC (smaller is better)</td>
<td>596.0</td>
<td>ARMA(1,1)</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>592.2</td>
<td>HF</td>
<td>Smallest</td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>593.4</td>
<td>ARH(1)</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>596.9</td>
<td>AR(1)</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>597.6</td>
<td>CS</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>598.1</td>
<td>UN</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>599.9</td>
<td>TOEP</td>
<td></td>
</tr>
<tr>
<td>BIC (smaller is better)</td>
<td>599.9</td>
<td>ARMA(1,1)</td>
<td></td>
</tr>
</tbody>
</table>
The most promising structure is HF since it has the smallest value for AIC, AICC and BIC, respectively (Table 1). The plot further confirms the conclusion. Since HF is a special case of UN, we can use a likelihood ratio test to compare the two structures by using the following macro. Similarly, we can test the likelihood ratio between CS and HF.

%RMMS
(Dt = example,
 Y = pulse,
 Xs = drug time drug*time,
 ClassXs = drug time,
 RepeatedVar = time,
 Sub= id,
 CovR= HF|,
 CovF= UN))

Table 2 and 3 show that HF is the better fitted model compared to UN (p=0.65) and CS (p=.0022). Finally, we can specify "FinalCov=HF|" and "outFile=D:\Pulse.rtf" in the macro to export the output of the best fitted model to a specified RTF file.

Table 2  Likelihood Ratio Test for UN and HF
H0: The Simpler Structure is Better

<table>
<thead>
<tr>
<th>Type</th>
<th>_2LLR</th>
<th>NumPara</th>
<th>DF</th>
<th>Chisq</th>
<th>p_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HF</td>
<td>578.613</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UN</td>
<td>577.736</td>
<td>6</td>
<td>2</td>
<td>0.877</td>
<td>0.6452</td>
</tr>
</tbody>
</table>

Table 3  Likelihood Ratio Test for CS and HF
H0: The Simpler Structure is Better

<table>
<thead>
<tr>
<th>Type</th>
<th>_2LLR</th>
<th>NumPara</th>
<th>DF</th>
<th>Chisq</th>
<th>p_value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS</td>
<td>590.832</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HF</td>
<td>578.613</td>
<td>4</td>
<td>2</td>
<td>12.218</td>
<td>.0022</td>
</tr>
</tbody>
</table>
CONCLUSION
%RMMS is designed to quickly find the best fitted repeated measure mixed model. If the mixed procedure fails to create a converge status data set, the macro will stop. For example, duplicate time points in the data set could cause this problem. %RMMS is frequently used in the analysis of repeated measures at this research institute, saving enormous time and resources.

REFERENCES
Tao, Jill, (2004), Mixed Models Analyses Using the SAS® System Course Notes, 4-15 to 4-45
Walker, Glen, Common Statistical Methods for Clinical Research With SAS® Example, 111-148

CONTACT INFORMATION
Your comments and questions are valued and encouraged. Contact the author at:

Jingjing Wang
3400 Rawley E. Chambers Avenue
Fort Sam Houston, Texas 78234
210.916.0345
Email: JingJing.Wang@amedd.army.mil

APPENDIX: MACRO CODE
%macro RMMS
(Dt =,
 Y =,
 Xs =,
 ClassXs =,
 RepeatedVar =,
 Sub =,
 CovStructures = TOEP|AR(1)|ARH(1)|CS|CSH|HF|ARMA(1,1)|UN|,
 NumRept=,
 CovG= |,
 CovS= |,
 FinalCov= |,
 Style=,
 OutFile=);

%Let FinalCov=%scan(&FinalCov,1,|);
%Let CovG=%scan(&CovG,1,|);
%Let CovS=%scan(&CovS,1,|);

/*I. Covariance structure selection before the final structure specified*/
%if ("&FinalCov" = "") %then %do;
/*1. Selection using IC */
/*1.1 Assign 1 to p and k and assign 1st candidate structure to CovStru*/
%let p=1;
%let k=1;
%let CovStru=%scan(&CovStructures,&k,|);
/*1.2 DO-WHILE loop for PROC MIXED under each candidate covariance structure*/
%do %while ("&CovStru" NE "");
  data temp;
    a=symget("CovStru");/*assign value of CovStru to a*/
    call symput("CovRname",compress(a,"(),"));
  run;
%end;
/*1.2.1 PROC MIXED procedure*/
proc mixed data=&dt ;
   class &ClassXs;
   model &Y=&Xs;
   repeated &repeatedVar/subject=&sub type=&CovStru;
ods output convergencestatus=cstatus&&k
      FitStatistics = Fit&&p
      FitStatistics = FitRF0&&p(rename={value=&CovRname})
      Dimensions = ParmRF0&&p(rename={value=Num& CovRname});
ods exclude ModelInfo Dimensions IterHistory ConvergenceStatus
      CovParms FitStatistics LRT ClassLevels Tests3;
/*1.2.2 Check the converge status of the model under the structure*/
data cstatus&&k;
   set cstatus&&k;
   call symput("conVS",status);
run;
/*1.2.3 To eliminate non-exist data due to non-converged PROC MIXED*/
%if &conVS=1 %then %goto RetainP;
/*1.2.4 Prepare the data Fitdt&&p contain fit statistics for valid structures*/
data Fitdt&&p;
   set Fit&&p;
      length type $16.0;
      type="&CovStru";
      FitStatName=trim(%scan(descr,1,"" ));
      if FitStatName ne "-2";
run;
/*1.3 Increase k by 1, remain p or increase p by 1 and assign the next candidate structure to CovStru*/
%let p=%eval(&p+1);
%RetainP: %let p=%eval(&p);
%let k=%eval(&k+1);
%let CovStru=%scan(&CovStructures,&k,|);
%end;
/*1.2 Create data set S containing fit statistics for all valid candidate structures and print out the results */
data S0;
   set
      %do i=1 %to &p-1;
         Fitdt&&i
      %end;
run;
proc sort data=S0;
   by FitStatName value;
Data S(drop=FitStatName);
   set S0;
      by FitStatName value ;
      if Descr ne "-2 Res Log Likelihood";
      if first.FitStatName then Note='Smallest';
proc print data=S noobs;
   Title1 "FIT STATISTICS FOR ALL VALID CANDIDATE COVARIANCE STRUCTURES
   LISTED IN ASCENDING ORDER";
run;

/*2. Selection by visualizing the covariance structure*/

proc mixed data=&dt ;
  class &ClassXs;
  model &Y=&Xs;
  repeated &repeatedVar/subject=&sub type=un r rcorr;;
  ods output covparms=cov rcorr=corr;
  ods exclude ModelInfo Dimensions IterHistory ConvergenceStatus
       CovParms  FitStatistics LRT ClassLevels Tests3 Rcorr CovParms R;
run;

data time;
  %do time1=1 %to &NumRept;
    Time=&time1;
    %do time2=1 %to &time1;
      duration=&time1-&time2;
      output;
    %end;
  %end;
run;

data covplot;
  merge time cov(rename=(estimate=Covariance));
run;

goptions reset=all ftext=swiss ftitle=swissb htitle=1.5 htext=1.5;

proc gplot data=covplot;
  label duration='Duration Between Time Points';
  plot Covariance*Duration=Time / haxis=axis1;
  symbol1 v=circle color=black i=none ;
  %do s=2 %to &NumRept;
    %let s1=%eval(&s-1);
    symbol&&s v=circle i=j l=%eval(&s-1);
  %end;
  axis1 offset=(2,2) minor=none;
  title1 'Covariance vs. Duration Between Time Points';
run;
quit;

/*3. Compare user defined nested structures using likelihood ratio test;*/

%if "&CovG" ne "" or "&CovS" ne "" %then  
  %do;
    data _null_; 
    call symput("CovGV",compress("&CovG","(),"));
    call symput("CovSV",compress("&CovS","(),"));
    data User;
    merge 
      %do  i=1 %to &p-1;
        FitRF0&&i 
        ParmRF0&&i 
      %end;
    if _n_ = 1 then 
      do;
        Chisq=(&CovSV-&CovGV);
        DF=(Num&&CovGV-Num&&CovSV);
        p_value=1-probchi(Chisq,DF);
        output;
        stop;
  %end;
data UserP(rename=(Chisqr=Chisq p_valuer=p_value));
length type $16 _2LLR 4.3 NumPara 4.0 DF 4.0 Chisq 4.3 p_value 4.3;
set User;
  Type = "&CovS";
  _2LLR = &CovSV;
  NumPara = Num&&CovSV;
  Chisqr = .;
  p_valuer = .;
output;
  Type="&CovG";
  _2LLR = &CovGV;
  NumPara = Num&&CovGV;
  Chisqr = Chisq;
  p_valuer = p_value;
output;
keep Type _2LLR NumPara DF Chisqr p_valuer ;
proc print data= UserP noobs;
title1 "Likelihood Ratio Test for &CovS and &CovG";
title2 "H0: The Simpler Structure is Better";
run;
%end;
%end;

/*II. Selected final model: PROC MIXED should be modified as needed*/
%if "&FinalCov" ne "" %then
  %do;
    ods rtf body="&outFile" style=&style startpage=yes;
    ods noproctitle;
    proc mixed data=&dt ;
      class &ClassXs;
      model &Y=&Xs;
      repeated /subject=&sub type=&FinalCov;
      title1 "Repeated Measure Mixed Model Using &FinalCov";
    run;
    ods rtf close;
  %end;
%mend RMMS;