

---

**Online Appendix to accompany S.K. Sterba and N.C. Gottfredson (2015).  
Diagnosing global case influence on MAR versus MNAR model comparisons.  
*Structural Equation Modeling*. 22, 294-307.**

---

Part I. Fitting MAR and MNAR models and exporting casewise loglikelihoods from each model  
Syntax.....p. 2-3  
Documentation.....p. 4

Part II. Using exported casewise loglikelihoods to calculate and plot model selection influence  
diagnostics  $\Delta ind_{BIC_i}$  and  $\Delta ind_{AIC_i}$  .....p. 5-6

**Part 1.** Different software packages, including *Mplus*, *Mx*, and *Open Mx*, could be used to fit the MAR and MNAR models and export casewise loglikelihoods from each fitted model, as discussed in the article. Below, example syntax is provided for doing so using one popular software package, *Mplus* (Muthén & Muthén, 1998-2013).

***Mplus* 7 syntax to fit the Example 1 MAR model and export casewise loglikelihoods:**

Here shown where dropout occurs only at times 4 and 5 and where dropout indicators *d4* and *d5* were constructed outside of *Mplus*.

```
DATA: file is mydata.dat;

VARIABLE:
names are caseid y1-y5 d4 d5 treat;
usevariables are y1-y5 d4 d5 treat;
idvariable is caseid;
categorical are d4 d5;
missing is . ;

ANALYSIS:
estimator = ml; integration = montecarlo;
starts = 100 10;

MODEL:
!submodel for outcome generating process
int linear | y1@0 y2@1 y3@2 y4@3 y5@4;
int linear; int with linear;
int linear on treat;
[int linear]; y1-y5;
!submodel for missingness mechanism
[d4$1-d5$1];
d4 on treat (a1);
d5 on treat (a1);
d4 on y3 (a2);
d5 on y4 (a2);

SAVEDATA:
file is ll_mar_example1.dat;
save= loglikelihood;
```

***Mplus* 7 syntax to fit the Example 1 (Diggle-Kenward) MNAR model and export casewise loglikelihoods:**

Here shown where dropout occurs only at times 4 and 5 and where dropout indicators *d4* and *d5* were constructed outside of *Mplus*.

```
DATA: file is mydata.dat;

VARIABLE:
names are caseid y1-y5 d4 d5 treat;
usevariables are y1-y5 d4 d5 treat;
idvariable is caseid;
categorical are d4 d5;
missing is . ;

ANALYSIS:
estimator = ml; integration = montecarlo;
starts = 100 10;

MODEL:
!submodel for outcome generating process
int linear | y1@0 y2@1 y3@2 y4@3 y5@4;
int linear; int with linear;
int linear on treat;
[int linear]; y1-y5;
!submodel for missingness mechanism
[d4$1-d5$1];
d4 on treat (a1);
d5 on treat (a1);
d4 on y3 (a2);
d5 on y4 (a2);
d4 on y4 (a3);
d5 on y5 (a3);

SAVEDATA:
file is ll_mnar_example1.dat;
save= loglikelihood;
```

Here dropout indicators were constructed outside of *Mplus* in the following manner. Let *k* be the timepoint of dropout on repeated measures for person *i*. Let *t* indicate a timepoint at which dropout could occur in the dataset; here *t*=4-5. Indicators of dropout, *dt*, (here *d4-d5*) for person *i* are constructed as: *dt*=0 if *t*<*k*; *dt*=1 if *t*=*k*; *dt*=. if *t*>*k*. Dropout indicators can also be constructed in *Mplus*, as shown below.

**Mplus 7 syntax to fit the Example 2 MAR model and export casewise loglikelihoods:**

Here shown where dropout occurs at times 2, 3, 4, and 5 and where dropout indicators  $d_2$ ,  $d_3$ ,  $d_4$  and  $d_5$  are constructed *inside* this Mplus code using the DATA MISSING commands.

DATA: file is mydata.dat;

VARIABLE:

names are caseid y1-y5 treat;  
usevariables are y1-y5 treat d2-d5;  
idvariable is caseid;  
categorical are d2-d5;  
missing is . ;

DATA MISSING:

names = y1 y2 y3 y4 y5;  
type = sdropout;  
binary = d2 d3 d4 d5;

ANALYSIS:

estimator = ml;  
integration = montecarlo;  
starts = 100 10;

MODEL:

**!submodel for outcome generating process**

int shape | y1@0 y2\* y3\* y4\* y5@4;  
int shape; int with shape;  
int shape on treat;  
[int shape]; y1-y5;

**!submodel for missingness mechanism**

[d2\$1-d5\$1];  
d2-d5 on treat (a1);

SAVEDATA:

file is ll\_mar\_example2.dat;  
save= loglikelihood;

**Mplus 7 syntax to fit the Example 2 (Wu-Carroll) MNAR model and export casewise loglikelihoods:**

Here shown where dropout occurs at times 2, 3, 4, and 5 and where dropout indicators  $d_2$ ,  $d_3$ ,  $d_4$  and  $d_5$  are constructed *inside* this Mplus code using the DATA MISSING commands.

DATA: file is mydata.dat;

VARIABLE:

names are caseid y1-y5 treat;  
usevariables are y1-y5 treat d2-d5;  
idvariable is caseid;  
categorical are d2-d5;  
missing is . ;

DATA MISSING:

names = y1 y2 y3 y4 y5;  
type = sdropout;  
binary = d2 d3 d4 d5;

ANALYSIS:

estimator = ml;  
integration = montecarlo;  
starts = 100 10;

MODEL:

**!submodel for outcome generating process**

int shape | y1@0 y2\* y3\* y4\* y5@4;  
int shape; int with shape;  
int shape on treat;  
[int shape]; y1-y5;

**!submodel for missingness mechanism**

[d2\$1-d5\$1];  
d2-d5 on treat (a1);  
d2-d5 on int (a2);  
d2-d5 on shape (a3);

SAVEDATA:

file is ll\_mnar\_example2.dat;  
save= loglikelihood;

Here dropout indicators were constructed inside Mplus using the DATA MISSING command together with the *type=sdropout* option. The above syntax under the DATA MISSING command constructs dropout indicators in the following manner. Let  $k$  be the timepoint of dropout for person  $i$ . Let  $t$  indicate a timepoint at which dropout could occur in the dataset; here  $t=2-5$ . The indicator of dropout,  $dt$ , (here  $d_2-d_5$ ) on repeated measure  $yt$  is constructed for person  $i$  as:  $dt=0$  if  $t < k$ ;  $dt=1$  if  $t=k$ ;  $dt=.$  if  $t > k$ .

**Notes regarding *Mplus* syntax.**

If higher dimensional Monte Carlo integration is required for fitting a researcher's MNAR model, they may want to increase the total number of integration points beyond the *Mplus* default. This can be done, for instance, by putting the following code under the ANALYSIS: command.

```
integration = montecarlo(10000);
```

This requests a total of 10,000 integration points to be used. The sum of the saved, outputted casewise loglikelihoods should match the sample loglikelihood that *Mplus* prints in the output file "Loglikelihood H0 Value"; if they are discrepant (more than rounding error), increase the number of integration points until they match closely.

If a potentially influential case is flagged, and a researcher desires to confirm its influence by refitting the MAR and MNAR models without including that case, the following code can be added under the VARIABLE: command:

```
useobservations are caseid ne 11;
```

This requests that the subject #11 be excluded from the analysis. Note that *caseid* is not a *Mplus* key word, it is the name of the subject identifier in the example dataset.

The SAVEDATA: command in the above syntax examples is used to export casewise loglikelihood values for a given fitted into an ASCII file. The name for the exported file is your choice (above, as an example, one file is called *ll\_mar\_example1.dat*). This file will be saved in the same location as your input file. At the very end of the regular *Mplus* output file, the order of variables in this exported file will be listed. Each name in the list refers to a column in the exported dataset, reading from left to right. It is crucial to inspect this list since the exported dataset will not have variable names included at the tops of the columns. Often, this list will begin with each variable used in the analysis, followed by your subject identifier (from the IDVARIABLE input statement), followed by the casewise loglikelihood values for the fitted model (which *Mplus* calls OUTLOGL). *Note:* sometimes factor scores and their associated standard errors will be provided in this exported dataset automatically, even if they were not requested under the SAVEDATA command.

A researcher would want to output casewise loglikelihood values for both competing models fitted in *Mplus*, as shown in the above syntax examples. Once ASCII files containing casewise loglikelihood values for both fitted models have been obtained, the researcher can move on to Part II.

**Part II.** Once the casewise loglikelihood values have been obtained from the fitted MAR and MNAR models under consideration, the model selection diagnostics can be computed and plotted in any standard software package allowing data manipulation and graphics (e.g., SAS, R, Stata, SPSS, etc.). Here, this is demonstrated using SAS 9.2.

After fitting the two competing models in *Mplus* in Part I, a researcher should have two ASCII files containing the casewise loglikelihood values from each fitted model (e.g., *ll\_mar\_example1.dat* and *ll\_mnar\_example1.dat*). The researcher will need to read their two ASCII datasets into SAS. Dataset variables must be read in the order that they were listed at the end of the *Mplus* output file, but they can be given different names.

```
DATA model_mar;
INFILE "C:\ll_mar_example1.dat";
INPUT d4 d5 y1-y5 treat caseid LL_mar; run;

DATA model_mnar;
INFILE "C:\ll_mnar_example1.dat";
INPUT d4 d5 y1-y5 treat caseid LL_mnar; run;
```

The researcher will then need to modify the input values in red before running the following code to compute and plot model selection influence diagnostics.

- $N$  is the number of cases (highest level units in the analysis, here, persons).
- $k^A$  is the number of estimated (free) parameters in the MAR model.
- $k^B$  is the number of estimated (free) parameters in the MNAR model.
- *case* should be set equal to whatever the researcher is calling their case identifier in the SAS datasets (e.g., *caseid*).
- Here, index plots will be saved to a file on the C drive called *analyses* in a document called *myindexplots.rtf*.

```
%let N=150;
%let case=id;
%let k_A=16;
%let k_B=17;

proc sort data=model_mar; by &case.; run;
proc sort data=model_mnar; by &case.; run;
data compare; merge model_mar model_mnar; by &case.; run;

data compare; set compare;
deltaindbic= -2*(LL_mar-LL_mnar)+ ((&k_A.-&k_B.)*log(&N./(&N.-1)));
deltaindaic= -2*(LL_mar-LL_mnar); run;

options helpbrowser=sas;
ods rtf file="C:\analyses\myindexplots.rtf"; options orientation=portrait;

goptions reset=all hsize=5 vsize=4 ftext='times new roman' htext=1.15 ;
axis1 label=(a=90 h=1.15 f='times new roman' "deltaindBIC") minor=none;
axis2 label=(a=0 h=1.15 f='times new roman' "ID") minor=none value=(h=1.15
f='times new roman');

proc gplot data=compare;
symbol v=dot pointlabel=("##&case.") w=.4 h=.4 c=black;
```

## ONLINE APPENDIX

```
plot deltaindbic*&case. / vaxis=axis1 haxis=axis2 ; run;

axis1 label=(a=90 h=1.15 f='times new roman' "deltaindAIC") minor=none;

proc gplot data=compare;
symbol v=dot pointlabel=("#&case.") w=.4 h=.4 c=black;
plot deltaindaic*&case. / vaxis=axis1 haxis=axis2 ; run;

ods rtf close;
```

The .RTF file containing two index plots (case ID vs.  $\Delta\text{ind}_{\text{BIC}_i}$  and case ID vs.  $\Delta\text{ind}_{\text{AIC}_i}$ ) which will be saved to the directory requested by the researcher. Each point on the index plot is labeled with its associated case ID.