Appendix A

EVS Data Set

Data sampled from the European Value Study (EVS) are used to illustrate the usage of the MDLV toolbox. Ten items (cf. Table 1) from the section of studying the attitude of the important criteria for a successful marriage are analyzed. We sampled 50 participants from each of the 31 countries in Wave 3 (1999) and Wave 4 (2008). Responses to the 10 items of the sampled participants are analyzed separately for Wave 3 and Wave 4 using a LCM and a MLCM.

A synthetic longitudinal data set was constructed for fitting LMM and MLMM with the MDLV toolbox. The background information available in the EVS database was used to link each respondent in Wave 3 to a demographically similar one in Wave 4 to yield a matched pair. Responses to the items of each pair were then treated as repeated responses over the two waves by a same individual. Country and household income (variable x047) were used as the matching criteria. Operationally, each participant within a country was ranked by household income for each wave. Then within each country, individuals of the same rank in the two waves were paired. The resulting data set is referred to as 'synthetic longitudinal data' hereafter.

The EVS data were available in the SPSS format. For the purpose of sampling data to be used in the present analysis, the SPSS data were saved in SAS format for later processing. A 3-point rating scale was used for each item in the survey: very important, rather important, and not very important. As the rating distributions for the items were skewed with most responses indicating 'very important', the other two ratings were combined into one for subsequent analysis. The 'very important' rating was recoded to '1' while 'rather important' and 'not very important' ratings were recoded to '0'. For further preparation of data for use in the MDLV toolbox, the SAS program was employed to sample participants, link data between waves, and recode item responses. Appendix A1 presents the SAS codes for creating the base data sets. Appendix A2 provides the details of the MATLAB syntax for preparing the data sets to fit LCM, LMM, MLCM, and MLMM using the MDLV toolbox.

Appendix A1: SAS Codes for data sampling, linking, and recoding

```
options pagesize=10000 linesize=256 nodate nocenter nonumber;
libname WVS 'C:\Documents and Settings\Administrator\Desktop\WVS2005';

Data Job1; set WVS.WVS_include3;RUN;
PROC CONTENTS; RUN;
/*
d027     important in marriage: faithfulness
d028     important in marriage: adequate income
d029     important in marriage: same social background
d031     important in marriage: shared religious beliefs
d032     important in marriage: good housing
d033     important in marriage: agreement on politics
```

```
d035     important in marriage: live apart from in-laws
d036     important in marriage: happy sexual relationship
d037     important in marriage: share household chores
d038     important in marriage: children
 2    s003        Num       8    country code
 3    s006        Num       8    original respondent number
 4    s007        Num       8    unified respondent number
 5    s009        Char      9    country abbreviation
 6    s020        Num       8    survey year
 1    s002evs     Num       8    EVS-wave
44    x047        Num       8    income household respondent
*/

/*  Sampling */
Data KeepWave34; SET Job1; *Delete country without complete data for
both Wave 3 and Wave 4;
IF s003 in(8,31,51,70,196,197,268,498,499,688,756,807,915,124,578,840) then
delete;
IF s002evs in(1,2) then delete; *Delete wave 1 and wave 2;
array Marriage[10] d027 d028 d029 d031 d032 d033 d035 d036 d037 d038;
Do i= 1 to 10; if Marriage(i)<0 then delete;END;
RUN;

PROC SORT data=KeepWave34;
    BY s003 s002evs;
RUN;

* sample 50 subjects for each country and each wave.;
PROC SURVEYSELECT data=KeepWave34
        method=srs n=50
        seed=123 out=SampleN50;
          strata s003 s002evs;
RUN;

Data SampleN50_Marriage(keep=s003 s002evs x047 d027 d028 d029 d031 d032 d033
d035 d036 d037 d038);
set SampleN50;
* recode the data;
Data SampleN50_MarriageV2; set SampleN50_Marriage;
array Marriage[10] d027 d028 d029 d031 d032 d033 d035 d036 d037 d038;
array M[10] M1-M10;
Do i= 1 to 10; if Marriage(i) in(2,3) then M(i)=1; else M(i)=0;END;
RUN;

Data SampleN50_MarriageV3; set SampleN50_MarriageV2;
IF s003 in(191,792) then delete; *Delete wave 1 and wave 2;
RUN;

* Link both waves through x047: household income;
PROC SORT data=SampleN50_MarriageV3;
    BY s003 s002evs x047;
RUN;
```

```
/*PROC PRINT data=SampleN50_MarriageV3; RUN;*/

Data IDindex;
    Do Country=1 to 31;
        Do Wave=1 to 2;
            Do Subject=1 to 50;
                output;
            End;
        End;
    End;
RUN;
PROC PRINT data=IDindex; RUN;

Data SampleN50_MarriageV4; merge IDindex SampleN50_MarriageV3;RUN;

PROC PRINT data= SampleN50_MarriageV4;
var Country Subject Wave M1-M10;
RUN;
```

Appendix A2: Data Preparation MATLAB Codes

There are 13 variables in the data set (Data), the names and order the variables are: "Country" "Subject" "Wave" "M1" "M2" "M3" "M4" "M5" "M6" "M7" "M8" "M9" "M10." The first three variables indicate the group ID, participant ID, and indication of waves for the responses of the 10 items in column 4 to 13. This information was used to reshape the responses of the 10 items into appropriate dimensions for specific models. The matrix (`Data2D`) includes all responses of the J (=10) items as columns, and rows are combinations of group ID, participant ID, and indication of waves. This matrix has the dimension of ($G \times Ng \times T$) by J. In this particular example, there are 3100 ($31 \times 50 \times 2$) rows and 10 columns. Note that G (=31) is the number of countries, N_g (=50) is the number of subjects, and T (=2) is the number of time points (survey waves).

This 2D data (`Data2D`) can be restructured into the required 4D data using the MATLAB function of *reshape* and *permute* with the dimension $G \times Ng \times J \times T$ for fitting MLMM (`Data4D`). The data used for fitting MLCM were obtained by separating the synthetic longitudinal data into two time waves (`ItemsT1_3D` and `ItemsT2_3D`), each is a $G \times Ng \times J$ matrix. The LCM and LMM can be applied when the nested data structure is ignored in the analyses. Thus a $N \times J \times T$ matrix (Data3D) is required for fitting LMM, where N (= $G \times Ng$) is the total sample size (=1550 in this example). For each time point, a $N \times J$ data matrix is required to fit with LCM (`ItemsT1_2D` and `ItemsT2_2D`). The EVSdata.m script contains the syntax to create data sets used in the analyses with these models. Executing this file will produce all the required data sets for the illustrative examples in this paper.

EVSdata.m MATLAB codes:

```
clc
clear

Data=[
```

```
    1         1       1     1     1     0     0     1     0     0     1     0     1 ;
    1         1       2     1     1     0     0     1     0     1     0     0     0 ;
    1         2       1     0     0     0     1     0     0     1     1     0     0 ;

...

   31        49       2     1     1     0     1     0     1     0     1     0     0 ;
   31        50       1     1     0     0     1     1     0     0     1     0     0 ;
   31        50       2     1     1     0     0     0     0     1     1     1     0 ];

%% Data matrix (4D) for MLMM (G X Ng X J X T)
Data2D=Data(:,4:13); %G*ng*t J keep only 10 items
A1=reshape(Data2D',[10,2,50,31]); % transform into 4D data matrix.
Data4D=permute(A1,[4,3,1,2]);  %G,Ng,J,T
%% reshape data matrix into 3D for LMM (N X J X T);
%Data4D : G,Ng,J,T
Data3D=reshape(Data4D,[50*31,10,2]); %(N X J X T)

%% Data matrix (2D) for LCM (G X Ng , J)
% data for time 1
idxT1 = ( Data(:,3)==1 );
ItemsT1_2D = Data(idxT1,:); %(G X Ng , J);
ItemsT1_2D = ItemsT1_2D(:,4:13);
% data for time 2
idxT2 = ( Data(:,3)==2 );
ItemsT2_2D = Data(idxT2,:); %(G X Ng , J);
ItemsT2_2D = ItemsT2_2D(:,4:13);

%% reshape data matrix into 3D for MLCM (G X Ng X J)
% wave=3;
ItemsT1=ItemsT1_2D; % Keep items only;
ItemsT13D=reshape(ItemsT1',[10,50,31]); %J X Ng X G;
ItemsT1_3D=permute(ItemsT13D,[3,2,1]); %G X Ng X J;
% wave=4;
ItemsT2=ItemsT2_2D; % Keep items only;
ItemsT23D=reshape(ItemsT2',[10,50,31]); %J X Ng X G;
ItemsT2_3D=permute(ItemsT23D,[3,2,1]); %G X Ng X J;

%% Summary: Data for
%LCM:  Wave3:ItemsT1_2D
%      Wave4:ItemsT2_2D
%LMM:  Data3D
%MLCM: Wave3:ItemsT1_3D
%      Wave4:ItemsT2_3D
%MLMM: Data4D
```