Australian Journal of Basic and Applied Sciences, 4(8): 3815-3834, 2010
ISSN 1991-8178

Analysis of Mixed Correlated Ordinal and Continuous Responses with and Without Missing Data in R: Latent Variable Approach

E. Bahrami Samani and M. Ganjali

Department of Statistics, Shahid Beheshti University, Tehran, IRAN

Abstract: This paper describes the joint models for univariate and multivariate mixed correlated ordinal and continuous responses with and without missing data in **R**. A full likelihood based approach is used to obtain maximum likelihood estimates of parameters of each model. We describe the implementation of these models in the packages **MASS**, **Mvtnorm**, **nlme** and **Stats** and illustrate the usage of the packages on several real data examples.

Key words: Joint modelling, Latent variable, Maximum likelihood, Ordinal and continuous responses, Missing Responses, **R**.

INTRODUCTION

Outcomes related to longitudinal mixed correlated ordinal and continuous data with possibility of missing values, are pervasive and research on analyzing them needs to be promoted. For joint modelling of responses, one method is to use the general location model of Olkin and Tate (1961), where the joint distribution of the continuous and categorical variables is decomposed into a marginal multinomial distribution for the categorical variables and a conditional multivariate normal distribution for the continuous variables, given the categorical variables (for a mixed poisson and continuous responses where Olkin and Tate's method is used see Yang *et al*., 2007 and for joint modelling of mixed outcomes using latent variables see McCulloch, 2007). A second method for joint modelling is to decompose the joint distribution as a multivariate marginal distribution for the continuous responses and a conditional distribution for categorical variables given the continuous variables. Cox and Wermuth (1992) empirically examined the choice between these two methods. The third method uses simultaneous modelling of categorical and continuous variables to take into account the association between the responses by the correlation between errors in the model for responses. For more details of this approach see, for example, Heckman (1978) in which a general model for simultaneously analyzing two mixed correlated responses is introduced and Catalano and Ryan (1992) who extended and used the model for a cluster of discrete and continuous outcomes.

Rubin (1976), Little and Rubin (2002), Diggle and Kenward (1994) made important distinctions between the various types of missing mechanisms for each of the above mentioned patterns. They define the missing mechanism as missing completely at random (MCAR) if missingness is dependent neither on the observed responses nor on the missing responses, and missing at random (MAR) if, given the observed responses, it is not dependent on the missing responses. Missingness is defined as non-random if it depends on the unobserved responses. From a likelihood point of view MCAR and MAR are ignorable but not missing at random (NMAR) is non-ignorable.

For mixed data with missing outcomes, Little and Schluchter (1985) and Fitzmaurice and Laird (1997) used the general location model of Olkin and Tate (1961) with the assumption of missingness at random (MAR) to justify ignoring the missing data mechanism. This means that they used all available responses, without a model for missing mechanism, to obtain parameter estimates using the EM (Expectation Maximization) algorithm. A model for mixed continuous and discrete binary responses with possibility of missing responses is introduced by Ganjali (2003). Ganjali and Shafie (2007) present a transition model for an ordered cluster of mixed continuous and discrete binary responses with non-monotone missingness.

In this paper a general latent variable model for simultaneously handling response and non-response in mixed ordinal and continuous data with potentially nonrandom missing values in both types of responses is presented. With this model, the dependence between responses can be taken into account by the correlation between errors in the models for continuous and ordinal responses and it is shown how we can incorporate a model for a complicated pattern of missing responses.

Corresponding Author: E. Bahrami Samani, Department of Statistics, Shahid Beheshti University, Tehran, IRAN
E-mail: ehsan_bahrami_samani@yahoo.com

The remaining of the paper proceeds as follows: in section 2 we introduce briefly the model and the likelihood. In section 3 we provide a detailed description of the **R** functions. Several illustrative examples (simulated and real) including guidance concerning the fitting of the models can be found in section 4. Finally, we end up with some concluding remarks in section 5. Detailed description and presentation of the likelihood of the model is provided at the appendix.

2 Model and Likelihood:

2.1 Complete Data Model:

We use Y_{ij} to denote *j*th ordinal response for the *i*th individual with c_j levels defined as,

$$Y_{ij} = \begin{cases} 1 & Y^*_{ij} < \theta_{1,j}, \\ k+1 & \theta_{k,j} \le Y^*_{ij} < \theta_{k+1,j}, \quad k = 1,\dots,c_j - 2 \\ c_j & Y^*_{ij} \ge \theta_{c_j,j}, \end{cases}$$

where i =1,....,n , j=1,....,M_1. $\theta_{1j},\dots,\theta_{c_j-1,j}$ are the cut-point parameters and Y^*_{ij} denotes the underlying latent variable for Y_{ij} . The joint model takes the form:

$$Y^*_{ij} = \beta_j' X_i + \varepsilon^{(1)}_{ij} \qquad j = 1,\dots,M_1$$

$$Z_{ij} = \beta_{j'} X_i + \varepsilon^{(2)}_{ij} \qquad j = M_1+1,\dots,M$$

$$\varepsilon_i = (\varepsilon^{(1)}_i, \varepsilon^{(2)}_i)' \overset{iid}{\sim} MVN(0,\Sigma)$$

where $\varepsilon^{(1)}_i = (\varepsilon^{(1)}_{i1},\dots,\varepsilon^{(1)}_{iM_1})', \varepsilon^{(2)}_i = (\varepsilon^{(2)}_{i(M_1+1)},\dots,\varepsilon^{(2)}_{iM})'$, $\theta_j = (\theta_{j1},\dots,\theta_{j(c_j-1)})'$ is the vector of cut-point parameters for the *j*th ordinal response and X_i is the vector of explanatory variables for the *i*th individual and Σ is the $M \times M$ covariance matrix which for illustration, when $M_1 = 2$ and $M = 3$ has the following structure,

$$\sum = \begin{pmatrix} 1 & \rho_{12} & \sigma\rho_{13} \\ \rho_{12} & 1 & \sigma\rho_{23} \\ \sigma\rho_{12} & \sigma\rho_{23} & \sigma^2 \end{pmatrix}$$

where σ^2 is the variance of the continuous response, and $\rho_{jj'}$ for $j \ne j'$, $j = j' = 1,2,3$ is the correlation between *j*th and *i*th responses. The vector of coefficients β_j cutpoints parameters θ_j for $j = 1,\dots,M_1$ and Σ should be estimated. The parameter vector, β_j for $j = M_1+1,\dots,M$ includes an intercept parameter but β_j , for $j = 1,\dots,M_1$, due to having cutpoints parameters, are assumed not to include any intercept. In this model any multivariate distribution can be assumed for the errors in the model. Here, a multivariate normal distribution is used. The likelihood for this model, which has been given in appendix A.1 and A.3.

2.2 Incomplete Data Model:

The ordinal response vector for the *i*th individual and the continuous response vector for the *i*th individual are denoted by $Y_i = (Y_{i1},...,Y_{iM_1})'$ and $Z_i = (Z_{i(M_1+1)},...,Z_{iM})'$. Typically, when missing data occur in an outcome, assume $R_{y_i} = (R_{y_{i1}},...,R_{y_{iM_1}})'$ as the indicator vector of responding to Y_i and $R_{y_{ij}}$ is defined as

$$R_{y_{ij}} = \begin{cases} 1 & R^*_{y_{ij}} > 0 \\ 0 & Otherwise \end{cases}$$

$R_{z_i} = (R_{z_{i(M_1+1)}},...,R_{z_{iM}})'$ is the indicator vector for responding to Z_i and $R_{z_{ij}}$ is defined as,

$$R_{z_{ij}} = \begin{cases} 1 & R^*_{z_{ij}} > 0 \\ 0 & Otherwise \end{cases}$$

where $R^*_{y_{ij}}$ and $R^*_{z_{ij}}$ denote the underlying latent variable of the non-response mechanism respectively, for the ordinal and continuous variables.
The joint model takes the form:

$$Y^*_{ij} = \beta_j' X_i + \varepsilon^{(1)}_{ij} \qquad j = 1,...,M_1$$

$$Z_{ij} = \beta_j' X_i + \varepsilon^{(2)}_{ij} \qquad j = M_1 + 1,...,M$$

$$R^*_{y_{ij}} = \alpha_j' X_i + \varepsilon^{(3)}_{ij} \qquad j = 1,...,M_1$$

$$R^*_{z_{ij}} = \alpha_j' X_i + \varepsilon^{(4)}_{ij} \qquad j = M_1 + 1,...,M$$

where X_i is the design matrix for the *i*th individual. In the above presented model, the vector of parameters β_j for $j = 1,...,M$, the parameters $\theta_{1j},...,\theta_{c_j-1,j}$ for $j = 1,...,M$, should be estimated. The vector, β_j for $j = M_1 + 1,...,M$ includes an intercept parameter but β_j, for $j = 1,...,M_1$, due to having cutpoint parameters, are assumed not to include any intercept.

Let

$$\varepsilon_i = (\varepsilon_i^{(1)'}, \varepsilon_i^{(2)'}, \varepsilon_i^{(3)'}, \varepsilon_i^{(4)'})' \overset{iid}{\sim} MVN(0, \Sigma_\varepsilon),$$

where $\varepsilon_i^{(u)} = (\varepsilon_{i1}^{(u)}, ..., \varepsilon_{iM_1}^{(u)})'$, for $u = 1,3$, $\varepsilon_i^{(u)} = (\varepsilon_{i(M_1+1)}^{(u)}, ..., \varepsilon_{iM}^{(u)})'$, for $u = 2,4$ and

$$\Sigma_\varepsilon = \begin{pmatrix} \Sigma_{11}^\varepsilon & \Sigma_{12}^\varepsilon & \Sigma_{13}^\varepsilon & \Sigma_{14}^\varepsilon \\ \Sigma_{21}^\varepsilon & \Sigma_{22}^\varepsilon & \Sigma_{23}^\varepsilon & \Sigma_{24}^\varepsilon \\ \Sigma_{31}^\varepsilon & \Sigma_{32}^\varepsilon & \Sigma_{33}^\varepsilon & \Sigma_{34}^\varepsilon \\ \Sigma_{41}^\varepsilon & \Sigma_{42}^\varepsilon & \Sigma_{43}^\varepsilon & \Sigma_{44}^\varepsilon \end{pmatrix}$$

where $\Sigma_{uu}^\varepsilon = Var(\varepsilon_i^{(u)})$, for $u = 1,2,3,4$ and $\Sigma_{uv}^\varepsilon = Cov(\varepsilon_i^{(u)}, \varepsilon_i^{(v)})$, $u < v,\ u,v = 1,2,3,4$ and $\Sigma_{uv}^\varepsilon = \Sigma_{vu}^{\varepsilon}{}'$. Because of identifiability problem we have to assume $Var(Y_{ij}^*) = Var(R_{y_{ij}}^*) = Var(R_{y_{ij}}^*) = 1$, so $\Sigma_{jj}^\varepsilon = I$ for $j = 1,3,4$. Note, if one of the matrixes $\Sigma_{13}^\varepsilon, \Sigma_{14}^\varepsilon, \Sigma_{23}^\varepsilon, \Sigma_{24}^\varepsilon$ is not zero, then the missing mechanism of response is not at random. The likelihood for this model, which has been given in appendix **A.2** and **A4**.

3 R Function for Model:

3.1 Short Description of Functions and Installation:

In order to run the maximum likelihood estimation for the A multivariate latent variable model for mixed continuous and ordinal responses, you have to install to MASS, Mvtnorm, nlme and Stats packages which are available from comprehensive R archive network (CRAN) at http:// CARN. R-project.org/. The following functions are available for direct use in R:

read.table(utils) Data Input
dmvnorm(mvtnorm) Multivariate Normal Density
pmvnorm(mvtnorm) Multivariate Normal Distribution
nlminb(stats) Optimization using PORT routines (The PORT documentation is at http://netlib.bell-labs.com/cm/cs/cstr/153.pdf.)
fdHess(nlme) Finite difference Hessian
lm(stats) Fitting Linear Models
class.ind(nnet) Generates Class Indicator Matrix from a Factor
ginv(MASS) Generalized Inverse of a Matrix

3.2 The function ``dnorm "and ``pnorm":

Density and distribution function for the normal distribution with mean equal to mean and standard deviation equal to sd. The function can be called using the following system:

dnorm(x, mean = 0, sd = 1, log = FALSE) pnorm(q, mean = 0,sd = 1, lower.tail = TRUE, log.p = FALSE)

Required Arguments
x,q : vector of quantiles.
p: vector of probabilities.
n: number of observations.

3.3 The function ``dmvnorm " and ``pmvnorm":

Function dmvnorm evaluates density function for the multivariate normal distribution with mean equal to mean and standard deviation equal to sigma. The function can be called using the following system:

dmvnorm(x, mean, sigma, log=FALSE)

Required Arguments

x : Vector or matrix of quantiles. If x is a matrix, each row is taken to be a quantile.

n : Number of observations.

mean: Mean vector, default is rep(0,length =ncol(x)).

sigma : Covariance matrix, default is diag(ncol(x)). log Logical; if TRUE, densities d are given as log(d). method Matrix decomposition used to determine the matrix root of sigma, possible methods are eigenvalue decomposition (eigen, default), singular value decomposition (svd), and Cholesky decomposition (chol).

Function pmvnorm computes the distribution function of the multivariate normal distribution. The function can be called using the following system:

pmvnorm(lower=-Inf, upper=Inf, mean,corr=NULL, sigma=NULL)

Required Arguments

lower: the vector of lower limits of length n.

upper: the vector of upper limits of length n.

mean :the mean vector of length n

corr : the correlation matrix of dimension n.

sigma: the covariance matrix of dimension n.

3.4 The function ``nlminb":

Function nlminb unconstrained and constrained optimization using PORT routines. The function can be called using the following system:

nlminb(start, objective, gradient = NULL, hessian = NULL, scale = 1, control = list(), lower = -Inf, upper =Inf)

Required Arguments

start: numeric vector, initial values for the parameters to be optimized.

objective: Function to be minimized. Must return a scalar value (possibly NA/Inf). The first argument to objective is the vector of parameters to be optimized, whose initial values are supplied through start. Further arguments (fixed during the course of the optimization) to objective may be specified as well.

gradient: Optional function that takes the same arguments as objective and evaluates the gradient of objective at its first argument. Must return a vector as long as start.

hessian: Optional function that takes the same arguments as objective and evaluates the hessian of objective at its first argument. Must return a square matrix of order length(start). Only the lower triangle is used.

scale: See PORT documentation

control :A list of control parameters.

lower, upper: vectors of lower and upper bounds, replicated to be as long as start. If unspecified, all parameters are assumed to be unconstrained.

Control Parameters

Possible names in the control list and their default values are:

eval.max : Maximum number of evaluations of the objective function allowed. Defaults to 200.

iter.max: Maximum number of iterations allowed. Defaults to 150.

trace: The value of the objective function and the parameters is printed every trace'th iteration. Defaults to 0 which indicates no trace information is to be printed.

abs.tol: Absolute tolerance. Defaults to 1e-20.

rel.tol: Relative tolerance. Defaults to 1e-10. x.tol X tolerance. Defaults to 1.5e-8.

step.min: Minimum step size. Defaults to 2.2e-14

Output Component

Value A list with components:

par: The best set of parameters found.

objective : The value of objective corresponding to par.

convergence: An integer code. 0 indicates successful convergence.

message: A character string giving any additional information returned by the optimizer, or NULL. For details, see PORT documentation.

iterations : Number of iterations performed.

evaluations: Number of objective function and gradient function evaluations.

3.5 The function ``fdHess":

Function fdHess Evaluate an approximate Hessian and gradient of a scalar function using finite differences. The function can be called using the following system:

fdHess(pars, fun)

Required Arguments

pars: the numeric values of the parameters at which to evaluate the function fun and its derivatives.

fun : a function depending on the parameters pars that returns a numeric scalar. *Output Component*

mean: the value of function fun evaluated at the parameter values

pars gradient an approximate gradient

Hessian: a matrix whose upper triangle containst an approximate Hessian.

3.6 The function ``class.ind":

Function class.ind Evaluate Generates a class indicator function from a given factor. The function can be called using the following system:

class.ind(cl)

Required Arguments

c1: factor or vector of classes for cases

4 Examples:

4.1 Example 1: Simulated data:

The aim of this simulation study is to compare ordinary Pearson correlation and the correlation obtained by the model in system (1) (vide also complete data model). We consider two continuous variables *Z* and *Y**. The ordinal variable *Y* with three levels is defined as

$$Y = \begin{cases} 1 & Y^* < \theta_1, \\ 2 & \theta_1 \le Y^* < \theta_2, \\ 3 & Y^* \ge \theta_2, \end{cases}$$

the variables *Z* and *Y** are generated by a bivariate normal distribution with zero mean and covariance matrix

$$\Sigma = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$$

We consider two sets of cut points. In the first set we let $\theta_1 = -1$ and $\theta_2 = 1$ In the second set we let $\theta_1 = -2$ and $\theta_2 = -1$. In the first set of cut points, not having any covariate in the model for latent variable of Y, one expect to have, roughly, 16 percents of *Y* values to be equal to 1, 16 percent to be equal to 3 and 68 percent to be equal to 2. So, the low and high values have nearly the same frequency but the middle value have the highest frequency. In the second set of cut points one expect to have more frequency for high values (2 percent to be equal to 1, 14 percent to be equal to 2 and 84 percent to be equal to 3). By these two sets of cut-points we may reach to different values for two correlations. We want to compare methods in small, reasonable and large sample sizes. For this we consider 3 values for n (50, 100, and 1000). In this analysis we use 1000 sets of simulation. In each simulation we analyze the following simple model

$$Z = \mu_z + \varepsilon_1$$
$$Y^* = \varepsilon_2$$
$$\sigma = exp(\psi_0).$$

A Latent Variable Model for Mixed Ordinal and Continuous responses.

Here we illustrate how we can obtain maximum likelihood estimator of model. All commands used in the illustration of example one. We firstly load the MASS, Stats and Mvtnorm packages the following commands:

```
library(MASS)
library(Stats)
library(Mvtnorm)
library(nlme)
```

and likelihood function for generating our data by:

```
 f<-function(X){
muz<-X[1]
muy<-0
sez<-exp(X[2])
 muygivenzx[,t]<-muy+(X[3]*(z[,t]-muz))/sez
 seygivenzx<-(1-(X[3])^2)
 q1[,t]<-(X[4]-muygivenzx[,t])/sqrt(seygivenzx)
q2[,t]<-(X[5]-muygivenzx[,t])/sqrt(seygivenzx)
 # Likelihood function for model
 l1[,t]<-log(pnorm(q1[,t]))+log(dnorm(z[,t],muz,sez))
 l2[,t]<-log(pnorm(q2[,t])-pnorm(q1[,t]))+log(dnorm(z[,t],muz,sez))
 l3[,t]<-log(1-pnorm(q2[,t]))+log(dnorm(z[,t],muz,sez))
 data0[,,t]<-cbind(Y[,t],l1[,t])
 data1[,,t]<-cbind(Y[,t],l2[,t])
 data2[,,t]<-cbind(Y[,t],l3[,t])
data0[data0[,1,t]==1,2,t]<-0
 data0[data0[,1,t]==2,2,t]<-0
 data1[data1[,1,t]==0,2,t]<-0
data1[data1[,1,t]==2,2,t]<-0
 data2[data2[,1,t]==0,2,t]<-0
data2[data2[,1,t]==1,2,t]<-0
 t0[t]<-sum(data0[,2,t])
 t1[t]<-sum(data1[,2,t])
 t2[t]<-sum(data2[,2,t])
 #-log(likelihood)
Tfinal[t]<-sum(t0[t])+sum(t1[t])+sum(t2[t])
 return(-Tfinal[t]) }
```

and the maximum likelihood estimation for model of data obtained by:

```
 #The maximum likelihood estimation
n=nlminb(X,f,lower=c(-Inf,-Inf,-0.999,-Inf,-Inf),upper=c(Inf,Inf,0.999,Inf,Inf))
 #Standard Deviation(se)
 h=fdHess(par, Li) h=h$Hessian;ih=ginv(h);se=sqrt(diag(ih))
```

where ginv(.) is Calculates the Moore-Penrose generalized inverse of a matrix pnorm is distribution function for the normal distribution and
X<-c(0,0,0.5,-1,1) is initial values.

Table I contains the average estimated values of μ_z , ψ_0 , ρ (estimated by the Model), ρ^* (estimated by the Pearson measure), θ_1 and θ_2 for n=50, n=100 and n=1000. The parameter estimates by the model for μ_z , ψ_0 , θ_1 and θ_2 (for *n*=50, n=100 and n=1000) are close to the true values of the parameters for both set of cut-points. Of course, the more the value of n the better the estimates. It can be seen from table I that for the first set of cut points ρ (Model)- ρ^* (Pearson) $\simeq$ 0.08 and for the second set of cut points ρ (Model)- ρ^* (Pearson) $\simeq$ 0.18 for any value of n. So, for a data set with high frequency for high values of ordinal response the Pearson correlation underestimates the true correlation between two errors. Hence, this measure should not be used for measuring association between two continuous and ordinal responses.

Table I: Results of the simulation study for both set of cut points.

The first set of cut points $(\theta_1 = -1, \theta_2 = 1)$

Parameter	True value	n=50 Est.	n=50 S.D	n=100 Est.	n=100 S.D.	n=1000 Est.	n=1000 S.D.
μ_z	0.000	0.036	0.138	0.020	0.100	0.002	0.033
ψ_0	0.000	-0.021	0.103	-0.002	0.071	-0.001	0.022
ρ (Model)	0.500	0.469	0.136	0.492	0.091	0.515	0.028
ρ^* (Pearson)	0.500	0.390	0.097	0.419	0.091	0.443	0.019
θ_1	-1.000	-1.152	0.233	-0.973	0.150	-0.995	0.047
θ_2	1.000	1.071	0.222	1.020	0.153	0.986	0.047

The second set of cut points $(\theta_1 = -2, \theta_2 = -1)$

Parameter	True value	n=50 Est.	n=50 S.D.	n=100 Est.	n=100 S.D.	n=1000 Est.	n=1000 S.D.
μ_z	0.000	0.046	0.134	0.061	0.095	0.020	0.031
ψ_0	0.000	-0.051	0.099	-0.041	0.071	-0.001	0.022
ρ (Model)	0.500	0.503	0.165	0.502	0.116	0.501	0.037
ρ^* (Pearson)	0.500	0.327	0.119	0.329	0.098	0.341	0.023
θ_1	-2.000	-2.078	0.280	-2.047	0.138	-2.009	0.086
θ_2	-1.000	-1.019	0.216	-0.963	0.148	-1.045	0.048

5 Example 2: Foreign language Data:

In this section, we present an application of the methodology to a foreign language achievement study (Raymond and Roberts, 1983, vide also Schafer, 1997, ch 9). The main purpose of the foreign language achievement study was to investigate the psychometric properties of the Foreign Language Attitude Scale (FLAS), an instrument developed for predicting proficiency in a variety of foreign language learning settings. The data we used in our analysis are based on N=236 students enrolled in foreign language courses at the Pennsylvani State University in the 1980s with complete records on the following variables: (1) Gender (0=male and 1=female); (2) Age, age group (0= less than 20 and 1=20^+); (3) PRI, number of prior foreign language courses (1=none, 2=1^+); (4) LAN, the foreign language studied by the student (1=French, 2=Spanish, 3=Other); (5) CGPA, current college grade point average; (6) FLAS, a continuous variable representing the student's score on the Foreign Language Attitude Scale; and incomplete records for (7) GRD, an ordinal variable with four levels representing the student's final letter grade in the foreign language course (1 if *D*, 2 if *C*, 3 if B and 4 if *A*). The vector of explanatory variables is X=(Gender, Age, PRI, LAN_1, LAN_2, CGPA) where LAN_1 and LAN_2 are dummy variables for languages French and Spanish, respectively.

The two correlated responses are FLAS and GRD. The mean of FLAS is 82.406 and its standard deviation is 13.655. A Kolmogorov Smrinov test of assumption of normality for FLAS is not rejected (P-value=0.776). A frequency table for student grade shows that 14.8% percent of GRD values are missing. The highest level of GRD is for *A* level (50.4%) and the lowest is for *D* level (3.6%), the other two levels contribute (22.4%) for B level and (8.8%) for C level. We also consider the interaction effect between Gender and LAN, due to its significant effect in pervious works of Schafer (1997) and De-Leon *et al*., (2007), who assume an ignorable missing mechanism, for these data.

A Latent Variable Model for Mixed Ordinal and Continuous responses with Possibility of Missing Responses.

The model takes into account the correlation between two responses (ρ_{12}) with missing mechanism, with the following form.

$$FLAS = \beta_0 + \beta_1 Gender + \beta_2 Age + \beta_3 PRI + \beta_4 LAN_1 + \beta_5 LAN_2 + \beta_6 CGPA + \beta_7 Gender \times LAN_1 + \beta_8 Gender \times LAN_2 + \sigma\varepsilon_1 \qquad (5a)$$

$$GRD^* = \gamma_1 Gender + \gamma_2 Age + \gamma_3 PRI + \gamma_4 LAN_1 + \gamma_5 LAN_2 + \gamma_6 CGPA + \gamma_7 Gender \times LAN_1 + \gamma_8 Gender \times LAN_2 + \varepsilon_2 \qquad (5b)$$

$$R^*_{GRD} = \alpha_0 + \alpha_1 Gender + \alpha_2 Age + \alpha_3 PRI + \alpha_4 LAN_1 + \alpha_5 LAN_2 + \alpha_6 CGPA + \alpha_7 Gender \times LAN_1 + \alpha_8 Gender \times LAN_2 + \varepsilon_3 \qquad (5c).$$

The correlation structure of the errors in this model is

$$\Sigma = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{pmatrix}$$

We firstly load the MASS, Stats, Mvtnorm, nlme and nnet packages the following commands:

```
library(MASS)
library(Stats)
library(Mvtnorm)
library(nlme)
library(nnet)
```

The variables are are following:

```
>nanmes(D12)
[1] lan sex hgpa age pri
[6] Cgpa flas grd Ry
```

Variables flas, grd and Ry prescribe the three responses (flas is continuous response and grd is ordinal response). The likelihood function for Foreign Language Data obtained by :

```
#making indicator variable
lan1<-class.ind(lan)[,1]
lan2<-class.ind(lan)[,2]
#Likelihood function for model
Li<-function(X){
```

```
 #The elements of matrix covariance
s12=X[1]
s13=X[2]
s23=X[3]
 Se23<-matrix(c(1,s23,s23,1),2,2)
Se3.1<-sqrt(1-(s13^2))
 Se2.1<-sqrt(1-(s12^2))
Se23.1<-Se23-matrix(c(s12^2,s12*s13,s12*s13,s13^2),2,2)
#cutpoints
 theta<-c(-Inf,X[4],X[4]+X[5],X[4]+X[5]+X[6],Inf)
 Se1<-X[7]
 for(i in 1:236){
   mu1[i]<-X[8]+X[9]*sex[i]+X[10]*age[i]+X[11]*pri[i]+X[12]*lan1[i]+
X[13]*lan2[i]+X[14]*cgpa[i]+X[15]*lan1[i]*sex[i]+X[16]*lan2[i]*sex[i]

   mu2[i]<-X[17]*sex[i]+X[18]*age[i]+X[19]*pri[i]+X[20]*lan1[i]+
X[21]*lan2[i]+X[22]*cgpa[i]+X[23]*lan1[i]*sex[i]+X[24]*lan2[i]*sex[i]

mu3[i]<-X[25]+X[26]*sex[i]+X[27]*age[i]+X[28]*pri[i]+X[29]*lan1[i]+
X[30]*lan2[i]+X[31]*cgpa[i]+X[32]*lan1[i]*sex[i]+X[33]*lan2[i]*sex[i] }

 for(i in 1:35){
 #For individual who observe only flas the likelihood(L[i]) is
 mu3.1[i]<-s13*((flas[i]-mu1[i])/Se1)
L[i]<-dnorm(flas[i],mu1[i],Se1)*pnorm(-mu3[i],mu3.1[i],Se3.1) }
 for(i in 36:236){
 #For individual with both flas and grd observed the likelihood (L[i]) is
mu2.1[i]<-s12*((flas[i]-mu1[i])/Se1)
 mu23.1[,,i]<-((flas[i]-mu1[i])/Se1)*c(s12,s13)
 L[i]<-(pnorm(theta[grd[i]+1]-mu2[i],mu2.1[i],Se2.1)
 -pnorm(theta[grd[i]]-mu2[i],mu2.1[i],Se2.1)
 -pmvnorm(lower=c(-Inf,-Inf),upper=c(theta[grd[i]+1]-mu2[i],-mu3[i]) ,mean=mu23.1[,,i],sigma=Se23.1)
 +pmvnorm(lower=c(-Inf,-Inf),upper=c(theta[grd[i]]-mu2[i],-mu3[i])
,mean=mu23.1[,,i],sigma=Se23.1) )
*dnorm(flas[i],mu1[i],Se1) }
 #-log(Likelihood)
 ;-sum(log(L)) }
```

and the maximum likelihood estimation for the parameter of model and standard deviation of the maximum likelihood estimation obtained by:

```
#The maximum likelihood estimation
 n=nlminb(X,f,lower=c(-Inf,-Inf,-0.999,-Inf,-Inf),upper=c(Inf,Inf,0.999,Inf,Inf))
 #Standard Deviation(se)
 h=fdHess(par, Li) h=h$Hessian;ih=ginv(h2);se=sqrt(diag(ih))
```

where ginv is Calculates the Moore-Penrose generalized inverse of a matrix.
X<-c(ro12,ro13,ro23,co1, co2,co3) is initial value which the following commands:

```
mo1<-lm(flas~sex+age+pri+lan1+lan2+cgpa+lan1*sex+lan2*sex)
 co1<-as.numeric(coef(mo1))
 mo2<-polr(factor(grd)~sex+age+pri+lan1+lan2+cgpa+lan1*sex+lan2*sex ,method=probit)
 co2<-as.numeric(coef(mo2))
 mo3<-glm(Ry~sex+age+pri+lan1+lan2+cgpa+lan1*sex+lan2*sex ,binomial(link=probit))
 co3<-as.numeric(coef(mo3))
 ro12=cor(grd[36:236],flas[36:236])
```

ro23=cor(Ry,flas) ro13=(grd[36:236],Ry[36:236])

where lm is used to fit linear models. It can be used to carry out regression, single stratum analysis of variance and analysis of covariance.

For model (NMAR) we have the following estimates:
$par
[1] 0.23898082 0.20928379 -0.93883391 2.17438695 0.82940910 1.00077366
[7] 12.67528408 81.62705194 4.47565667 -1.74855430 -1.06644859 -0.42114818
[13] -7.28932079 0.28053541 2.10675565 10.62660566 1.01985590 0.47907464
[19] 0.35397494 0.09009515 -0.59831440 1.17463488 -1.56052760 -0.72604704
[25] -2.91300783 -0.56122184 0.93157654 -0.03058144 0.78643054 0.01385496
[31] 1.10343949 0.26730806 0.97182207
$objective
[1] 1175.676
$convergence
[1] 0
$message
[1] relative convergence (4)
$iterations [1] 2
$evaluations function gradient
7 157

and
se:
[1] 0.07988473 0.09543372 0.17612701 0.65564968 0.16733498 0.12190140
[7] 0.58343433 5.31311319 2.43480451 1.72328072 1.89403708 2.90042147
[13] 2.72620124 1.64024407 4.26004067 3.95115241 0.28835382 0.17801354
[19] 0.19645500 0.27498248 0.24895405 0.19555585 0.44014148 0.40614075
[25] 0.87229845 0.25083515 0.23979358 0.29102204 0.39891892 0.68383212
[31] 0.25752219 0.71308748 0.40813403

In this model there is no constrain for the missing mechanism of GRD (NMAR). We use the likelihood ratio to test for NMAR. By these results we can conclude that the two responses are correlated and also the missing indicator for GRD is related to both responses. This leads to have a not at random missing mechanism (Deviance= 10.086 with 1 d.f., P-value=0.002). The mean of latent variable for grade is less for Spanish language learners than that for other language learners which gives an expectation of low value for GRD (low levels of GRD). On the other hand, the mean of *GRD** in increasing with high values for CGPA, which leads to high level for GRD. *GRD** is also less for French female learners than other learners. Some covariates are also effective in responding to GRD. For example, male are more likely to respond. French people are more likely to respond and people with higher values of CGPA are more likely to respond.

6 Conclusion:

In this paper a univariate and multivariate latent variable model is presented for simultaneously modelling of ordinal and continuous correlated responses with and without potentially non-random missing values in both types of responses is proposed. We assume a multivariate normal distribution for errors in the model. However, any other multivariate distribution such as t or logistic can be also used. Binary responses are a special case of ordinal responses. So, our model can also be used for mixed binary and continuous responses. For correlated nominal, ordinal and continuous responses Deleon and Carrière (2007) have developed a model by extending general location model. Generalization of our model for nominal, ordinal and continuous responses is an ongoing research on our part.

REFERENCES

Ash, R.B. and C.A. Dolens-Dade, 2000, *Probability and measure thoery*: Academic Press.

Catalano, P. and L.M. Ryan, 1992. Bivariate latent variable models for clustered discrete and continuous outcoms. Journal of the American Statistical Association, 87(419): 651-658.

Cox, D.R. and N. Wermuth, 1992. Response models for mixed binary and quantitative variables. *Biometrika*, 79(3): 441-461.

De Leon, A.R. and K.C. Carrière, 2007. General mixed-data model: Extension of general location and grouped continuous models. Canadian Journal of Statistics, 35(4): 533-548.

Diggle, P.J. and M.G. Kenward, 1994. Informative Drop-out in longitudinal data analysis. Applied Statistics, 43(1): 49-93.

Fitzmaurice, G.M. and N.M. Laird, 1997, Regression models for mix discrete and continuous responses with potentially missing values. Biometrics, 53: 110-122.

Ganjali, M., 2003. A model for mixed continuous and discrete responses with possibilty of missing responses. Journal of Sciences, Islamic Republic of Iran, 14(1): 53-60.

Ganjali, M. and K. Shafie, 2006. A transition model for an ordered cluster of mixed continuous and discrete responses with non- monotone missingness Journal of Applied Statistical sciences, 15 Issue 3.

Heckman, J.J.D., 1978. Endogenous variable in a simutaneous Equation system. Econometrica, 46(6): 931-959.

Little, R.J. and M. Schluchter, 1987. Maximum likelihood estimation for mixed continuous and categorical data with missing values. Biometrika, 72: 497-512.

Little, R.J. and D. Rubin, 2002. Statistical analysis with missing data. Second edition. New york, Wiley.

McCulloch, C., 2007. Joint modelling of mixed outcome type using latend variables. statistical methods in Medical Research, 1-27.

Rubin, D.B., 1976. Inference and missing data. Biometrica, 82: 669-710.

Olkin, L. and R.F. Tate, 1961. Multivariate correlation models with mixed discrete and continuous variables. Biometrka, 63: 581-592.

Schafer, J.L., 1997. Analysis of incomplete multivariate data, Chapman and Hall.

Yang, Y., J. Kang, K. Mao and J. Zhang, 2007. Regression models for mixed Poisson and continuous longitudinal data. Statistics in Medicine, 26: 3782-3800.

A. Likelihood:

6.1 A.1. Likelihood for Univariate latent mixed model for Mixed Continuous and Ordinal Responses:

Let $\eta = (\beta_1', \beta_2', \psi', \rho, \theta_1, ..., \theta_{c-1})'$. The likelihood for this model is

$$L(\eta \mid z, y, x_1, x_2) = \prod_{i=1}^{n} f(z_i, Y_i = y_i \mid x_{i1}, x_{i2}, \eta)$$

$$= \prod_{i=1}^{n} f(Y_i = y_i \left| z_i, x_{i2}, \beta_2, \theta_1, ..., \theta_{c-1}, \rho) f(z_i \mid x_{i1}, \beta_1, \psi)\right.$$

$$= \prod_{i=1}^{n} f(\theta_{y_{i-1}} \le Y_i^* \le \theta_{y_i} \left| z_i, x_{i2}, \beta_2, \theta_1, ..., \theta_{c-1}) f(z_i \mid x_{i1}, \beta_1, \psi)\right.$$

$$= \prod_{i=1}^{n} [\Phi(\frac{\theta_{y_i} - \mu_{Y_i^*|z_i, x_{i2}}}{(Var(Y_i^* | z_i))^{1/2}}) - \Phi(\frac{\theta_{y_{i-1}} - \mu_{Y_i^*|z_i, x_{i2}}}{(Var(Y_i^* | z_i))^{1/2}})] f(z_i \mid x_{i1}, \beta_1, \psi)$$

where $\theta = -\infty$, $\theta = +\infty$, Φ is the cumulative standard normal distribution, $\mu_{Y_i^*|z_i, x_{i2}} = \beta_2 ' X_{i2} + \frac{\rho}{\sigma}(z_i - \beta_1 ' X_{i1})$, $Var(Y_i^* | z_i) = 1 - \rho^2$. Likelihood shows the simplification obtained by using the assumption of normality for errors of model in system of equations.

6.2 A.2. Likelihood for Univariate latent mixed model for Mixed Continuous and Ordinal Responses With Missing Responses:

For individual who observe neither y nor z the likelihood is

$$L = Pr(R_y = 0, R_z = 0)$$

$$= \Phi_{12}(-\alpha_1 ' X_3, -\alpha_2 ' X_4; \rho_{34})$$

where $\Phi_{12}(.;.)$ is the cumulative bivariate normal distribution with mean vector of 0 and

$$\Sigma_{34} = \begin{pmatrix} 1 & \rho_{34} \\ \rho_{34} & 1 \end{pmatrix}$$

For individual who observe only y the likelihood is

$$L = Pr(Y = y, R_y = 1, R_z = 0)$$

$$= Pr(Y = y, R_z = 0) - Pr(Y = y, R_y = 0, R_z = 0)$$

$$= \Pr(\eta_{y-1} \le Y^* \le \eta_y, R_z^* < 0) - \Pr(\eta_{y-1} \le Y^* \le \eta_y, R_y^* < 0, R_z^* < 0)$$

$$= \Phi_{12}(\eta_y - \beta_1 ' X_1, -\alpha_2 ' X_4; \rho_{14}) - \Phi_{12}(\eta_{y-1} - \beta_1 ' X_1, -\alpha_2 ' X_4; \rho_{14})$$

$$-\Phi_{123}(\eta_y - \beta_1 ' X_1, -\alpha_1 ' X_3, -\alpha_2 ' X_4; \Sigma_{134}) + \Phi_{123}(\eta_{y-1} - \beta_1 ' X_1, -\alpha_1 ' X_3, -\alpha_2 ' X_4; \Sigma_{134})$$

where

$$\Sigma_{134} = \begin{pmatrix} 1 & \rho_{13} & \rho_{14} \\ \rho_{13} & 1 & \rho_{34} \\ \rho_{14} & \rho_{34} & 1 \end{pmatrix}$$

and $\Phi_{123}(.,.;.)$ is cumulative three-variate normal distribution with mean vector of 0. For individual who observe only z the likelihood is

$$L = f(z)\Pr(R_y = 0, R_z = 1 | z)$$

$$= f(z)[\Pr(R_y = 0 | z) - \Pr(R_y = 0, R_z = 0 | z)]$$

$$= f(z)[\Pr(R_y^* < 0 | z) - \Pr(R_y^* < 0, R_z^* < 0 | z)]$$

$$= f(z)[\Phi(\frac{-\alpha_1{}'X_3 - \frac{\rho_{23}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{23}^2}})$$

$$-\Phi_{12}(\frac{-\alpha_1{}'X_3 - \frac{\rho_{23}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{23}^2}}, \frac{-\alpha_2{}'X_4 - \frac{\rho_{24}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{24}^2}}; \frac{\rho_{34}-\rho_{24}\rho_{23}}{\sqrt{1-\rho_{23}^2}\sqrt{1-\rho_{24}^2}})]$$

For individual with both z and y observed the likelihood is

$$L = f(z)\Pr(Y=y, R_y=1, R_z=1|z)$$

$$= f(z)[\Pr(Y=y|z) - \Pr(Y=y, R_y=0|z)$$

$$-\Pr(Y=y, R_z=0|z) + \Pr(Y=y, R_y=0, R_z=0|z)]$$

$$= f(z)[\Pr(\eta_{y-1} \le Y^* \le \eta_y|z) - \Pr(\eta_{y-1} \le Y^* \le \eta_y, R_y^* < 0|z)$$

$$-\Pr(\eta_{y-1} \le Y^* \le \eta_y, R_z^* < 0|z) + \Pr(\eta_{y-1} \le Y^* \le \eta_y, R_y^* < 0, R_z^* < 0|z)]$$

$$= f(z)[\Phi(\frac{\eta_y - \beta_1{}'X_1 - \frac{\rho_{12}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{12}^2}}) - \Phi(\frac{\eta_{y-1} - \beta_1{}'X_1 - \frac{\rho_{12}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{12}^2}})$$

$$-\Phi_{12}(\frac{\eta_y - \beta_1{}'X_1 - \frac{\rho_{12}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{12}^2}}, \frac{-\alpha_1{}'X_3 - \frac{\rho_{23}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{23}^2}}; \frac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{23}^2}})$$

$$+\Phi_{12}(\frac{\eta_{y-1} - \beta_1{}'X_1 - \frac{\rho_{12}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{12}^2}}, \frac{-\alpha_1{}'X_3 - \frac{\rho_{23}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{23}^2}}; \frac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{23}^2}})$$

$$-\Phi_{12}(\frac{\eta_y - \beta_1{}'X_1 - \frac{\rho_{12}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{12}^2}}, \frac{-\alpha_2{}'X_4 - \frac{\rho_{24}}{\sigma}(z-\beta_2{}'X_2)}{\sqrt{1-\rho_{24}^2}}; \frac{\rho_{14}-\rho_{12}\rho_{24}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{24}^2}})$$

$$+\Phi_{12}(\frac{\eta_{y-1}-\beta_1'X_1-\frac{\rho_{12}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{12}^2}},\frac{-\alpha_2'X_4-\frac{\rho_{24}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{24}^2}};\frac{\rho_{14}-\rho_{12}\rho_{24}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{24}^2}})$$

$$+\Phi_{123}(\frac{\eta_{y}-\beta_1'X_1-\frac{\rho_{12}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{12}^2}},\frac{-\alpha_1'X_3-\frac{\rho_{23}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{23}^2}}$$

$$,\frac{-\alpha_2'X_4-\frac{\rho_{24}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{24}^2}};\Sigma_{134|2})$$

$$-\Phi_{123}(\frac{\eta_{y-1}-\beta_1'X_1-\frac{\rho_{12}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{12}^2}},\frac{-\alpha_1'X_3-\frac{\rho_{23}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{23}^2}}$$

$$,\frac{-\alpha_2'X_4-\frac{\rho_{24}}{\sigma}(z-\beta_2'X_2)}{\sqrt{1-\rho_{24}^2}};\Sigma_{134|2})]$$

where

$$\Sigma_{134|2}=\begin{pmatrix} 1 & \frac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{23}^2}} & \frac{\rho_{14}-\rho_{12}\rho_{24}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{24}^2}} \\ \frac{\rho_{13}-\rho_{12}\rho_{23}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{23}^2}} & 1 & \frac{\rho_{34}-\rho_{24}\rho_{23}}{\sqrt{1-\rho_{24}^2}\sqrt{1-\rho_{23}^2}} \\ \frac{\rho_{14}-\rho_{12}\rho_{24}}{\sqrt{1-\rho_{12}^2}\sqrt{1-\rho_{24}^2}} & \frac{\rho_{34}-\rho_{24}\rho_{23}}{\sqrt{1-\rho_{24}^2}\sqrt{1-\rho_{23}^2}} & 1 \end{pmatrix}$$

where $\eta_0=-\infty$, $\eta_c=+\infty$, (in this model, η_c is maximum cut-point and .

6.3 A.3. Likelihood for Multivariate Latent Mixed Model for Mixed Continuous and Ordinal Responses:

We apply the following definition and theorem for multivariate distributions to obtain the form of the likelihood.

Definition 2.2.1: If $F(w_1, \cdots, w_{M_1}) = P(W_1^* \le w_1, ..., W_{M_1}^* \le w_{M_1})$ is a distribution function, operator $\Delta_{b_j a_j} F(w_1, \cdots, w_{M_1})$ is defined as, $(b_j \ge a_j)$

$$F(w_1, \cdots, w_{(j-1)}, b_j, w_{(j+1)}, ..., w_{M_1}) - F(w_1, \cdots, w_{(j-1)}, a_j, w_{(j+1)}, ..., w_{M_1}).$$

Theorem 2.2.2: If for $j = 1, ..., M_1$, $a_j \le b_j$, then

$$P(a_1 < W_1^* \le b_1, ..., a_{M_1} < W_{M_1}^* \le b_{M_1}) = \Delta_{b_1 a_1} \cdots \Delta_{b_{M_1} a_{M_1}} F(w_1, \cdots, w_{M_1})$$

where

$$\Delta_{b_1 a_1} \cdots \Delta_{b_{M_1} a_{M_1}} F(w_1, \cdots, w_{M_1}) = F_0 - F_1 + F_2 - ... + (-1)^{M_1} F_{M_1};$$

And F_j is the sum of all $\binom{M_1}{j}$ terms of the from $F(g_1, ..., g_{M_1})$ with $g_k = a_k$ for exactly *j* integers in $\{1, ..., M_1\}$, and $g_k = b_k$ for the remaining $M_1 - j$ integers.

Proof. See (Ash, 2000, page 27.)

The likelihood of the model for $z_i = (z_{i(M_1+1)}, ..., z_{iM})'$, $y_i = (y_{i1}, ..., y_{iM_1})'$, $x_i = (x_{i1}, ..., x_{ip})'$, where *p* is the number of explanatory variables for the *ith* individual (the number of components in this vector may also be dependent on the chosen variable, i.e x_i be x_{im} and p be p_m, here, we ignore this for simplicity), $z = (z_1', ..., z_n')'$, $y = (y_1', ..., y_n')'$ and $x = (x_1', ..., x_n')'$ takes the form

$$L(\eta, \Sigma \mid z, y, x) = \prod_{i=1}^{n} f(z_i, y_i \mid x_i, \eta, \Sigma)$$

$$= \prod_{i=1}^{n} [P(Y_{i1} = y_{i1}, ..., Y_{iM_1} = y_{iM_1} \left| z_i, x_i) \; f(z_i \mid x_i)\right]$$

$$= \prod_{i=1}^{n} [P(\theta_{1,(y_{i1-1})} < Y_{i1}^* \le \theta_{1,y_{i1}}, ..., \theta_{M_1(y_{i,M_1-1})} < Y_{iM_1}^* \le \theta_{M_1, y_{iM_1}} \left| z_i, x_i) f(z_i \mid x_i)\right]$$

where $\eta = (\beta_1, ..., \beta_M, \theta_1, ..., \theta_{M_1})'$, $\theta_{m0} = -\infty$ and $\theta_{mc_{im}} = +\infty$

Using the above Theorem, the likelihood could be summarized as,

$$L(\eta,\Sigma \mid z,y,x)=\prod_{i=1}^{n}[\Delta_{b_{i1}a_{i1}}\cdots\Delta_{b_{iM_1}a_{iM_1}}F(w_{i1},\cdots,w_{iM_1}\mid z_i,x_i)f(z_i\mid x_i)]$$

$$=\prod_{i=1}^{n}[(F_{i0}-F_{i1}+F_{i2}-...+(-1)^{M_1}F_{iM_1})f(z_i\mid x_i)]$$

where $b_{ij}=\theta_{jy_{ij}}$ and $a_{ij}=\theta_{j(y_{ij}-1)}$ and F_{ij} is the sum of all $\binom{M_1}{j}$ terms of the from $F(g_{i1},...,g_{iM_1}\mid z_i,x_i)$ with $g_{ik}=a_{ik}$ for exactly *j* integers in $\{1,...,M_1\}$, and $g_{ik}=b_{ik}$ for the remaining M_1-j integers. This likelihood can be maximized by function ``**nlminb**'' in software **R**. For example, if $M_1=2$, $M=3$, $c_1=c_2=3$, $(a_{i1},b_{i1})=(\theta_{1,y_{i1}-1},\theta_{1,y_{i1}})$ and $(a_{i2},b_{i2})=(\theta_{2,y_{i2}-1},\theta_{2,y_{i2}})$ for an individual with $y_{i1}=y_{i2}=2$, and $Z_i=z_i$ the likelihood is:

$$L_i(\eta,\Sigma\mid z_i,y_i,x_i)=[\Delta_{b_{i1}a_{i1}}\Delta_{b_{i2}a_{i2}}F(w_{i1},w_{i2}\mid z_i,x_i)]f(z_i\mid x_i)$$

$$=f(z_i\mid x_i)[F(\theta_{1,y_{i1}},\theta_{2,y_{i2}}\mid z_i,x_i)+F(\theta_{1,y_{i1}-1},\theta_{2,y_{i2}-1}\mid z_i,x_i)$$

$$-F(\theta_{1,y_{i1}-1},\theta_{2,y_{i2}}\mid z_i,x_i)-F(\theta_{1,y_{i1}},\theta_{2,y_{i2}-1}\mid z_i,x_i)].$$

where

$$F(w_{i1},w_{i2}\,|z_i,x_i)=\Phi(w_{i1}-\mu_{y_{i1}|z_i},w_{i2}-\mu_{y_{i2}|z_i};\Sigma_{12|3})$$

and $\Phi(.,.;.)$ is cumulative bivariate normal distribution with mean vector of 0 and covariance matrix,

$$\Sigma_{12|3}=\begin{pmatrix}1-\rho_{13}^2 & \rho_{12}-\rho_{13}\rho_{23}\\ \rho_{12}-\rho_{13}\rho_{23} & 1-\rho_{23}^2\end{pmatrix}$$

where ρ_{12}, ρ_{13} and ρ_{23} are correlations between (y_{i1}^*,y_{i2}^*), (y_{i1}^*,z_{i3}) and (y_{i2}^*,z_{i3}), respectively and

$$\mu_{y_{i1}|z_i,x_i}=X_i'\beta_1+\rho_{13}(z_i-X_i'\beta_3)/\sigma,\ \mu_{y_{i2}|z_i,x_i}=X_i'\beta_2+\rho_{23}(z_i-X_i'\beta_3)/\sigma$$

and $Var(Z_i)=\sigma^2$.

6.4 A.4. Likelihood for Multivariate Latent Mixed Model for Mixed Continuous and Ordinal Responses with Missing Responses:

Likelihood for this model is as follows,the likelihood of the model for $y_i=(y_{i1},...,y_{iM_1})'$,

$z_i = (z_{i(M_1+1)},...,z_{iM})'$ and $x_i = (x_{i1},...,x_{ip})'$, where p is the number of explanatory variables for the *ith* individual (the number of components in this vector may also be dependent on the chosen variable, i.e x_i be x_{ij} and p be p_j , here, we ignore this for simplicity), $z = (z_1',...,z_n')'$, $y = (y_1',...,y_n')'$, $x = (x_1',...,x_n')'$, $J^y_{obs} = \{j : y_{ij}\ is\ observed\}$, $J^y_{Mis} = (J^y_{obs})^C$ and $J^z_{obs} = \{j : z_{ij}\ is\ observed\}$, $J^z_{Mis} = (J^z_{obs})^C$. The Likelihood for this model:

$$L = \prod_{i=1}^{n} f(Y_{i,obs}, Z_{i,obs}, R_{y_i}, R_{z_i} \mid X_i)$$

$$= \prod_{i=1}^{n} P(Y_{i,obs}, R_{y_{i,obs}}, R_{z_{i,obs}}, C^*_{Mis} \mid Z_{i,obs}, X_i) f(Z_{i,obs} \mid X_i)$$

$$= \prod_{i=1}^{n} P(Y^*_{i,obs}, R^*_{y_{i,obs}}, R^*_{z_{i,obs}}, C^*_{Mis} \mid Z_{i,obs}, X_i) f(Z_{i,obs} \mid X_i)$$

where $P(Y^*_{i,obs}, R^*_{y_{i,obs}}, R^*_{z_{i,obs}}, C^*_{Mis} \mid Z_{i,obs}, X_i)$ is conditional distribution of $(Y^*_{i,obs}, R^*_{y_{i,obs}}, R^*_{z_{i,obs}}, C^*_{Mis})'$ given $(Z_{i,obs}, X_i)'$ and

$$Z_{i,obs} = \{Z_{ij}, \forall j \in J^z_{obs}\}$$

$$Y_{i,obs} = \{Y_{ij}, \forall j \in J^y_{obs}\}$$

$$Y^*_{i,obs} = \{\theta_{y_{ij}-1} \le Y^*_{ij} \le \theta_{y_{ij}}, \forall j \in J^y_{obs}\}$$

$$R_{z_{i,obs}} = \{R_{z_{ij}} = 1, \forall j \in J^z_{obs}\}$$

$$R_{y_{i,obs}} = \{R_{y_{ij}} = 1, \forall j \in J^y_{obs}\}$$

$$R^*_{z_{i,obs}} = \{R^*_{z_{ij}} > 0, \forall j \in J^z_{obs}\}, \quad R^*_{z_{i,Mis}} = (R^*_{z_{i,obs}})^C$$

$$R^*_{y_{i,obs}} = \{R^*_{y_{ij}} > 0, \forall j \in J^y_{obs}\}, \quad R^*_{y_{i,Mis}} = (R^*_{y_{i,obs}})^C$$

$$R_{y_i} = (R_{y_{i1}},....,R_{y_{iM_1}})'$$

$$R_{z_i} = (R_{z_{i(M_1+1)}}, ..., R_{z_{iM}})'$$

and

$$C^*_{Mis} = \{R_{y_{ij}} = 0; \forall j \in J^y_{Mis}, R_{z_{ij}} = 0; \forall j \in J^z_{Mis}\}$$

and

$$P(Y^*_{i,obs}, R^*_{y_{i,obs}}, R^*_{z_{i,obs}}, C^*_{Mis} \mid Z_{i,obs}, X_i) = \gamma^*_i - \Gamma^*_{y_i} - \Gamma^*_{z_i} + \Gamma^*_{y_i z_i}$$

where

$$\gamma^*_i = P(Y^*_{i,obs}, C^*_{Mis} \mid Z_{i,obs}, X_i)$$

$$\Gamma^*_{z_i} = P(Y^*_{i,obs}, C^*_{Mis}, R^*_{z_{i,Mis}} \mid Z_{i,obs}, X_i)$$

$$\Gamma^*_{y_i} = P(Y^*_{i,obs}, C^*_{Mis}, R^*_{y_{i,Mis}} \mid Z_{i,obs}, X_i)$$

$$\Gamma^*_{y_i z_i} = P(Y^*_{i,obs}, C^*_{Mis}, R^*_{y_{i,Mis}}, R^*_{z_{i,Mis}} \mid Z_{i,obs}, X_i).$$

Supposed the $g_1 -$ elements of Y_1 and $g_2 -$ elements of Z_i are observed, so $J^y_{obs} = \{o_1, ..., o_{g_1}\}$ and $J^z_{obs} = \{o_1, ..., o_{g_2}\}$. Using the Theorem 2.2.2 :

$$\gamma_i^* = \Delta_{b_{o_1} a_{o_1}} \ldots \Delta_{b_{og_1} a_{og_1}} P(Y_{io_1}^* \le \omega_{io_1}, \ldots, Y_{iog_1}^* \le \omega_{iog_1}, C_{Mis}^* \mid Z_{i,obs}, X_i)$$
$$= F_{i0}^{(1)} - F_{i1}^{(1)} + F_{i2}^{(1)} - \ldots + (-1)^{g_1} F_{ig_1}^{(1)}$$

$$\Gamma_{z_i}^* = \Delta_{b_{o_1} a_{o_1}} \ldots \Delta_{b_{og_1} a_{og_1}} P(Y_{io_1}^* \le \omega_{io_1}, \ldots, Y_{iog_1}^* \le \omega_{iog_1}, C_{Mis}^*, R_{z_{i,Mis}}^* \mid Z_{i,obs}, X_i)$$
$$= F_{i0}^{(2)} - F_{i1}^{(2)} + F_{i2}^{(2)} - \ldots + (-1)^{g_1} F_{ig_1}^{(2)}$$

$$\Gamma_{y_i}^* = \Delta_{b_{o_1} a_{o_1}} \ldots \Delta_{b_{og_1} a_{og_1}} P(Y_{io_1}^* \le \omega_{io_1}, \ldots, Y_{iog_1}^* \le \omega_{iog_1}, C_{Mis}^*, R_{y_{i,Mis}}^* \mid Z_{i,obs}, X_i)$$
$$= F_{i0}^{(3)} - F_{i1}^{(3)} + F_{i2}^{(3)} - \ldots + (-1)^{g_1} F_{ig_1}^{(3)}$$

$$\Gamma_{y_i z_i}^* = \Delta_{b_{o_1} a_{o_1}} \ldots \Delta_{b_{og_1} a_{og_1}} P(Y_{io_1}^* \le \omega_{io_1}, \ldots, Y_{iog_1}^* \le \omega_{iog_1}, C_{Mis}^*, R_{y_{i,Mis}}^*, R_{z_{i,Mis}}^* \mid Z_{i,obs}, X_i)$$
$$= F_{i0}^{(4)} - F_{i1}^{(4)} + F_{i2}^{(4)} - \ldots + (-1)^{g_1} F_{ig_1}^{(4)}$$

where $b_{o_j} = \theta_{o_j y_{io_j}}$ and $a_{o_j} = \theta_{o_j (y_{io_j} - 1)}$ for $j = 1, \ldots, g_1$ and $F_{ij}^{(1)}, F_{ij}^{(2)}, F_{ij}^{(3)}$ and $F_{ij}^{(4)}$ are the sum of all $\binom{g_1}{j}$ terms of the from:

$$P(Y_{ic_1}^* \le c_1, \ldots, Y_{ic_{g_1}}^* \le c_{g_1}, C_{Mis}^* \mid Z_{i,obs}, X_i),$$

$$P(Y_{ic_1}^* \le c_1, \ldots, Y_{ic_{g_1}}^* \le c_{g_1}, C_{Mis}^*, R_{z_{i,Mis}}^* \mid Z_{i,obs}, X_i),$$

$$P(Y_{ic_1}^* \le c_1, \ldots, Y_{ic_{g_1}}^* \le c_{g_1}, C_{Mis}^*, R_{y_{i,Mis}}^* \mid Z_{i,obs}, X_i)$$

$$P(Y_{ic_1}^* \le c_1, \ldots, Y_{ic_{g_1}}^* \le c_{g_1}, C_{Mis}^*, R_{y_{i,Mis}}^*, R_{z_{i,Mis}}^* \mid Z_{i,obs}, X_i)$$

with $c_k = a_k$ for exactly j integers in $\{1, \ldots, g_1\}$, and $c_k = b_k$ for the remaining $g_1 - j$ integers,