

A MODEL FOR MIXED CONTINUOUS AND DISCRETE RESPONSES WITH POSSIBILITY OF MISSING RESPONSES

M. Ganjali*

*Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University,
Evin, Tehran 19838, Islamic Republic of Iran*

Abstract

A model for missing data in mixed binary and continuous responses, which can be used on cross-sectional data, is presented. In this model response indicator for the binary response can be dependent on the continuous response. A closed form for the likelihood is found. For data with a complicated pattern of missing responses some new residuals are also proposed. The model of multiplicative heteroscedasticity is used to consider the problem of heteroscedasticity of the continuous response. The model is illustrated on the data of an observational study where the effect of psychological disorder of parents on both the verbal comprehension score and the presence of adverse symptoms in their children are modeled in the presence of missing responses.

Keywords: Mixed data; Missing responses; Multiplicative heteroscedasticity; Pearson residuals; Psychological disorder

1. Introduction

Some biomedical, psychological and health sciences data include both discrete and continuous outcomes. One example is the analysis of development toxicity endpoints when the relationship between fetal weight and malformation in live fetus is an important statistical issue [1]. Another example is, in the study of the maternal smoking effect on respiratory illness in children where we have a continuous measure of pulmonary function and a binary measure of chronic symptoms in children.

For the first example, separate analyses of the categorical or the continuous responses cannot properly

assess the effect of dose on fetal weight and malformations simultaneously. For the second example, separate analyses cannot assess the effect of maternal smoking on all the responses simultaneously. Furthermore, separate analyses give biased estimates for the parameters and misleading inference [9]. Consequently, we need to consider a method in which these variables can be modeled jointly.

For joint modeling of responses, one method is to use the general location model of Olkin and Tate [4], 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

* E-mail: m-ganjali@cc.sbu.ac.ir

continuous variables, *given* the categorical variables. Another method is to consider the conditional distribution of discrete variables, *given* the continuous variables and a marginal distribution for continuous variables. Cox and Wermuth [2] empirically examined the choice between these two methods. The third method which is developed here is to use simultaneous modeling of categorical and continuous variables to take into account the association and dependence between the responses by the correlation between the errors in the models for responses. For more details of this approach see, for example, Heckman [6] in which a general model for simultaneously analyzing two mixed correlated responses is introduced and Catalano and Ryan [1] who extend and used the model for a cluster of discrete and continuous outcomes.

Little and Rubin [10] made an important distinction between the various types of missing mechanism. They defined the missing mechanism as missing completely at random (MCAR) if missingness is dependent neither on the observed responses nor on the missing responses, and as random (MAR) if it is dependent on the observed responses, but not 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 missing not at random (MNAR) is non-ignorable.

For mixed data with missing outcomes, Little and Schluchter [11] and Fitzmaurice and Laird [3] used the general location model of Olkin and Tate [14] with the assumption of missingness at random (MAR) to justify ignoring the missing data mechanism [10]. This means that they used all available responses, without a model for missing mechanism, to obtain parameter estimates using the EM (Expectation Maximization) algorithm. For a good discussion of mixed normal and non-normal data with missing responses under the assumption of MAR, see [10] and [16].

In this paper a general latent variable model for simultaneously handling response and non-response in mixed discrete 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 of the response models. The aim is to use the general model of Heckman [6] for the joint modeling of the discrete and continuous responses. It is shown how we can incorporate a model for a complicated pattern of missing responses.

In Section 2, the model is described and some new residuals are introduced. In Section 3, this model is used on a subset of data available in Little and Schluchter [11] about the effects of parental psychological

disorders on various aspects of the development of their children. The results of fitting this model are also presented. In Section 4, the paper concludes with some remarks.

2. Model and Residuals for Mixed Data with Missing Responses

2.1. The Model

Suppose y is a binary discrete variable and z is a continuous variable. Variables y and z are correlated and must be modeled simultaneously. Let y^* , R_y^* and R_z^* denote the underlying latent variables of the binary response, the non-response mechanism for the binary variable and the non-response for the continuous variable, respectively, and define

$$R_y = \begin{cases} 1 & \text{if } R_y^* > 0 \\ 0 & \text{Otherwise} \end{cases} ,$$

as the response indicator for y ,

$$R_z = \begin{cases} 1 & \text{if } R_z^* > 0 \\ 0 & \text{Otherwise} \end{cases} ,$$

as the response indicator for z and

$$y = \begin{cases} 1 & \text{if } y^* > 0 \\ 0 & \text{Otherwise} \end{cases} ,$$

as the binary response. The model takes the form:

$$y^* = \beta_1' X_1 + \varepsilon_1, \tag{1a}$$

$$z = \beta_2' X_2 + \sigma \varepsilon_2, \tag{1b}$$

$$R_y^* = \alpha_1' X_3 + \varepsilon_3, \tag{1c}$$

$$R_z^* = \alpha_2' X_4 + \varepsilon_4, \tag{1d}$$

where X_j for $j=1,2,3,4$ are vectors of explanatory variables. It is assumed that $E(\varepsilon_j)=0$ for $j=1,\dots,4$. The unstructured (*us*) variance covariance matrix of the vector of errors $(\varepsilon_1 \ \varepsilon_2 \ \varepsilon_3 \ \varepsilon_4)'$ is

$$\Sigma_{us} = \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \rho_{14} \\ \rho_{12} & 1 & \rho_{23} & \rho_{24} \\ \rho_{13} & \rho_{23} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{pmatrix}$$

The vector parameters $\beta_1, \beta_2, \alpha_1, \alpha_2$ and the correlation parameters $\rho_{jj'}$ for $j < j'$, $j = 1,2,3,4$ and

$j' = 2, \dots, 4$ should be estimated. In this model any multivariate distribution can be assumed for the errors in the model. Here a multivariate normal distribution is assumed.

To consider the problem of heteroscedasticity of the continuous response we can use multiplicative model of heteroscedasticity [5] for σ which is

$$\sigma = \exp(\gamma_0 + \gamma'X) \tag{1e}$$

where γ_0 and the vector of parameter γ should be estimated and X is the vector of explanatory variables.

The non-response of y (R_y) in this model depends on both responses, the non-response for z (R_z) also depends on both responses and z depends on y . If, one of the correlation parameters $\rho_{jj'}$ for $j=1,2$ and $j'=3,4$ is found to be significant, then we have a nonrandom missing process. Parameter ρ_{14} (ρ_{23}) can tell us whether or not non-response in the continuous variable (binary variable) is dependent on response in the discrete variable (continuous variable). Parameter ρ_{34} can tell us whether or not non-response in the continuous variable is dependent on the non-response in the discrete variable.

The likelihood for this model, which has been given in appendix, shows the simplification obtained by using the assumption of normality in the models for non-response. The simplification arises from the fact that, instead of resorting to a numerical approximation of the integral needed to integrate out the missing values of the continuous response, we just need to invoke a point on the cumulative three-variate normal distribution, which although approximate, can be computed in many standard statistical packages with sufficiently high precision.

The NAG [13] routine EO4UCF is used in this paper to minimize the mines logarithm of the likelihood given in the appendix. EO4UCF is a Fortran routine to minimize a smooth function subject to constraints (for example simple bounds on the variables) using a sequential quadratic programming (SQP, [4]) method. In this routine all unspecified derivatives are approximated by finite differences.

2.2. Residuals

The missing values of responses create problems for the usual residual diagnostics, see *e.g.* Hirano *et al.* [8]. For example usual Pearson residuals of the z variable are of the form

$$r_z = \frac{z - E(z)}{\sqrt{\text{var}(Z)}}$$

i.e. they assume MCAR. These unconditional residuals will be misleading if missingness is MAR or MNAR as in these cases the z 's are from a conditional or truncated conditional distribution. We can examine the residuals of the responses conditional on being observed. Let start with using theoretical form, involving $E(Z|R_z = 1)$, $E(Y|R_y = 1)$ and $\text{var}(Z|R_z = 1)$ rather than their predicted values. The Pearson residuals for continuous response can take the form

$$r_z = \frac{z - E(Z|R_z = 1)}{\sqrt{\text{var}(Z|R_z = 1)}} \tag{2a}$$

and for the discrete response can take the form

$$r_y = \frac{y - E(Y|R_y = 1)}{\sqrt{E(Y|R_y = 1) \times (1 - E(Y|R_y = 1))}} \tag{2b}$$

where

$$E(Z|R_z = 1) = \beta'_2 X_2 + \sigma \rho_{24} M,$$

$$\text{var}(Z|R_z = 1) = \sigma^2 \left[(1 - \rho_{24}^2) + \rho_{24}^2 (1 - M \alpha'_2 X_4 - M^2) \right] \tag{3}$$

where $M = \frac{\Phi(-\alpha'_2 X_4)}{\Phi(\alpha'_2 X_4)}$ is the Mill's ratio [7,12] and

$$E(Y|R_y = 1) = P_r(Y = 1|R_y = 1) = \frac{1 - \Phi(-\beta'_1 X_1) - \Phi(-\alpha'_3 X_3) + \Phi_{12}(-\beta'_1 X_1, -\alpha'_3 X_3, ; \rho_{13})}{1 - \Phi(-\alpha'_3 X_3)} \tag{4}$$

where Φ and Φ_{12} are respectively the cumulative univariate and bivariate normal distributions.

The estimated Pearson residuals (\hat{r}_z, \hat{r}_y) can be found by using the maximum likelihood estimates of the parameters in System (1). As residuals in (2a) and (2b) are defined by conditioning on observing the responses, they differ from those of Ten Have *et al.* [15]. Ten Have *et al.* [15] found the expectation and variances of responses, and consequently the residuals, unconditional on the fact that responses should be observed. Then, to calculate the residuals they assumed no link between

response and nonresponse. This gives biased estimates of the means and variances of responses if missingness is not at random [10].

3. Application: The Effect of Psychological Disorders on the Verbal Comprehension Score

3.1. Psychological Disorders Data

The data set extracted from Little and Schluchter [11] has been used also by Fitzmaurice and Laird [3]. These authors assumed a mechanism by which observations were MAR and then used the general model of Olkin and Tate [14] to decompose the joint distribution of the continuous and discrete variables. In these data, school-age children are classified according to the risk of their parents to psychological disorders (normal, moderate risk and high risk). Little and Schluchter [11] and Fitzmaurice and Laird [3] used two continuous responses, a standardized reading score and a standardized verbal comprehension score, and a single binary response, indicating the presence of adverse symptoms, which was obtained for each child. Here, only the standardized verbal comprehension score and the binary response for the first child in family are used. The data include 69 families with a general pattern of missing responses. Table 1 gives the different patterns of missing response for each child.

For example, 15 first-born children responded for both variables and the observed value of their binary response is 0.

Since a preliminary analysis indicated that there was no significant difference between moderate and high risk groups of psychological disorders [3], in this paper, the moderate and high risk groups are combined into a single group.

Table 1. Different patterns of missing responses for psychological disorders data

R_y	R_z	y	No.
1	1	0	15
1	1	1	12
0	1	-	12
1	0	0	6
1	0	1	8
0	0	-	16

Table 2. Descriptive statistics for psychological disorders observed data

		Score		Ad. Sym		
Low Risk	No	mean	S.D.	0	1	Total

0	22	105.455	30.764	10** (24.4)	11(26.8)	21(51.2)
1	17	146.118	19.647	16(39)	4(9.8)	20(48.8)
Total	39	123.176*	33.210*	26(63.4)	15(36.6)	41

* Calculated ignoring low risk, ** Count(%)

Table 2 presents summary statistics for presence of the adverse symptoms (Ad. Sym. in Table 2) and the standardized verbal comprehension score using only observed data. In this table the value of 0 for 'Low Risk' means that we have an individual from moderate or high risk group and the value of 1 means that individual is from low risk group.

Table 2 shows that the mean (standard deviation) score of children from a family with a low risk of psychological disorders (Low Risk=1 in Table 2) is more (lower) than families with moderate or high risk of psychological disorders. It also shows that the occurrence of adverse symptoms is less in family with a low risk of psychological disorders.

As the variance of scores is less for a family with a low risk of psychological disorders, the problem of heteroscedasticity should be taken into account.

A Kolmogorov Smirnov test of the assumption of normality for scores of the children of a low family of psychological disorders is not rejected (p-value=0.200). The same test does not reject the assumption of the normality for scores of the children from a moderate or high risk of psychological disorders (p-value=0.200). The Pearson correlation between two responses is $r=-0.449$. This emphasizes that two responses should be modeled simultaneously.

3.2. Models for Psychological Disorders Data

For comparative purposes, three models are considered. The first model (model I) uses only complete cases and does not consider the correlation between two responses. This model is

$$y^* = \beta_{01} + \beta_{11}L + \varepsilon_1, \tag{5a}$$

$$z = \beta_{02} + \beta_{12}L + \sigma\varepsilon_2, \tag{5b}$$

$$\sigma = \exp(\gamma_0 + \gamma_1L) \tag{5c}$$

where there is no correlation between ε_1 and ε_2 . The second model (model II) uses model I and takes into account the correlation between two responses. The third model (model III) considers the full model, model with missing mechanism, with the following form

$$y^* = \beta_{01} + \beta_{11}L + \varepsilon_1, \tag{6a}$$

$$z = \beta_{02} + \beta_{12}L + \sigma\varepsilon_2, \tag{6b}$$

$$R_y^* = \alpha_{01} + \alpha_{11}L + \alpha_{21}M + \varepsilon_3, \tag{6c}$$

$$R_z^* = \alpha_{02} + \alpha_{12}L + \alpha_{22}M + \varepsilon_4, \tag{6d}$$

$$\sigma = \exp(\gamma_0 + \gamma_1L). \tag{6e}$$

In these models

$$L = \begin{cases} 1 & \text{if child is from a low risk group} \\ 0 & \text{Otherwise} \end{cases}$$

and

$$M = \begin{cases} 1 & \text{if child is from a Moderate risk group} \\ 0 & \text{Otherwise} \end{cases}$$

In model (III) non-response models (R_y^* and R_z^*) let have one more explanatory variable [6].

3.3. Results for Psychological Disorders Data

For the standardized verbal comprehension score and the binary response for the first born child in family, model (III) finds no connection between missing mechanism and responses with a change of deviance of 2.4116 on 4 d.f. (p-value=0.661). Consequently, using only the complete data, without considering the missing mechanism, can give unbiased estimates of the parameters. Results of using three models are given in table 3. For model (III) as the correlation parameters $\rho_{jj'}$ for $j=1,2$ and $j'=3,4$ are not significant (as mentioned above) results with removing these parameters are given.

These results show that there is a negative correlation (a change in deviance of 4.742 with 1 df) between two responses (the more the score the less likely the presence of adverse symptoms) and separate analysis of the responses gives biased estimates of the standard errors of the parameters. For example, under model (I), the standard errors of the estimates of β_{02} and β_{12} , in comparison with the results of the model (II), are overestimated and the standard errors of the estimates of β_{01} and β_{11} are underestimated. Results also show that children from a low risk family can obtain a better standardized verbal comprehension score and the variance of their scores is less than that of the family with a high or moderate risk ($\gamma_1=-0.786$ for model II).

The estimate of the parameter by model (III) shows a positive correlation between two underlying variables for missing mechanisms. This means that children who do not respond their score are more likely to not to respond their discrete response.

Table 4 shows values of residuals calculated using

(2b) and model (II) for the discrete response.

Residuals for continuous response calculated using (2a) do not include any value larger than 3 (in absolute value) and with the residual values in Table 4, model (II) can be considered as a good fit for the psychological disorders data.

Table 3. Results using three models for the psychological disorders data

Parameter	Model(III)		Model(II)		Model(I)	
	Estimate	S.E.	Estimate	S.E.	Estimate	S.E.
β_{01}	0.288	0.251	0.173	0.335	0.180	0.337
β_{11}	-0.786	0.434	-0.671	0.495	-0.682	0.496
β_{02}	104.783	6.338	105.214	9.300	105.214	9.297
β_{12}	41.870	7.782	45.786	10.341	45.786	10.315
α_{01}	0.141	0.300	-	-	-	-
α_{11}	0.002	0.387	-	-	-	-
α_{21}	0.271	0.394	-	-	-	-
α_{02}	0.141	0.304	-	-	-	-
α_{12}	0.201	0.393	-	-	-	-
α_{22}	-0.137	0.398	-	-	-	-
γ_0	3.410	0.151	3.556	0.187	3.549	0.189
γ_1	-0.475	0.224	-0.786	0.263	-0.770	0.278
ρ_{12}	-0.474	0.205	-0.512	0.194	-	-
ρ_{34}	0.373	0.173	-	-	-	-
-loglik	295.218		139.322		141.693	

Table 4. Residuals for discrete response using model (II) for psychological disorders data

	Y=0	Y=1
L=0	-0.871	1.149
L=1	-1.494	0.670

4. Discussion

In this paper a latent variable model is presented for simultaneously modelling of binary and continuous correlated responses. The model is formulated in such a way that can handle the possibility of missing responses. The latent variable model in system (1) can also be used for a mixture of ordinal and continuous data (with usual changes for ordinality of the discrete response). It can also be used for more than two variables with missing responses. In this case the more variables with missing values we have the more

response indicators should be defined and the more parameters coming to the estimation procedure.

The selection models in systems (1) can be sensitive to the assumption of normality for error distributions [10]. However, the residuals introduced in this paper can be used to practically examine this assumption.

Acknowledgments

The author is grateful to Shahid Beheshti University for the awarded research grant.

References

1. Catalano P. and Ryan L.M. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **87**(419): 651-658 (1992).
2. Cox D.R. and Wermuth N. Response models for mixed binary and quantitative variables. *Biometrika*, **79**(3): 441-461 (1992).
3. Fitzmaurice G.M. and Laird N.M. Regression models for mixed discrete and continuous responses with potentially missing values. *Biometrics*, **53**: 110-122 (1997).
4. Fletcher R. *Practical Methods of Optimization*. John Wiley (2000).
5. Harvey A. Estimating regression models with multiplicative heteroscedasticity. *Econometrica*, **44**: 461-465 (1976).
6. Heckman J.J. Dummy endogenous variables in a simultaneous equation system. *Econometrica*, **46**(6): 931-959 (1978).
7. Heckman J.J. Sample selection bias as a specification error. *Ibid.*, **47**: 153-161 (1979).
8. Hirano K. Imbens, G. Ridder, G. and Rubin, D. Combining panel data sets with attrition and refreshment samples. Technical working paper, National Bureau of Economic Research. Cambridge, Massachusetts (1998).
9. Leon A.R. and Carriere K.C. On the one sample location hypothesis for mixed bivariate data. *Commun. statist. Theory and Meth.*, **29**(11): 2573-2581 (2000).
10. Little R.J and Rubin D. *Statistical Analysis with Missing Data*. New York; Wiley (1987).
11. Little R.J. and Schluchter M. Maximum likelihood estimation for mixed continuous and categorical data with missing values. *Biometrika*, **72**(3): 497-512 (1985).
12. Maddala G.S. Limited dependent and qualitative variables in econometrics. Cambridge, (1983).
13. NAG. Numerical Algorithms Group Manual. Mark 16. Oxford, U.K. (1996).
14. Olkin I. and Tate R.F. Multivariate correlation models with mixed discrete and continuous variables. *Annals of Mathematical Statistics*, **32**: 448-465 (1997).
15. Ten Have T.R.T., Kunselman A.R., Pulkstenis E.P., and Landis J.R. Mixed effects logistic regression models for longitudinal binary response data with informative dropout. *Biometrics*, **54**: 367-383 (1998).
16. Schafer J.L. *Analysis of Incomplete Multivariate Data*, Chapman & Hall (1997).

Appendix: Likelihood for Mixed Data with Missing Mechanism

For individuals who observe neither y nor z the likelihood is

$$L = \Pr(R_y = 0, R_z = 0) = \Phi_{12}(-\alpha'_1 X_3, -\alpha'_2 X_4; \rho_{34}), \tag{7}$$

where Φ_{12} is the cumulative bivariate normal distribution.

For individuals who observe only y and the value of y is 0 the likelihood is

$$L = \Pr(y = 0, R_y = 1, R_z = 0) = \Pr(y = 0, R_z = 0) - \Pr(y = 0, R_y = 0, R_z = 0) = \Phi_{12}(-\beta'_1 X_1, -\alpha'_2 X_4; \rho_{14}) - \Phi_{123}(-\beta'_1 X_1, -\alpha'_1 X_3, -\alpha'_2 X_4; \Sigma_{134}), \tag{8}$$

where

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

and Φ_{123} is cumulative three-variate normal distribution.

For individuals who observe only y and the value of y is 1 the likelihood is

$$L = \Pr(y = 1, R_y = 1, R_z = 0) = \Pr(R_z = 0) - \Pr(y = 0, R_z = 0) - \Pr(R_y = 0, R_z = 0) + \Pr(y = 0, R_y = 0, R_z = 0) = \Phi(-\alpha'_2 X_4) - \Phi_{12}(-\beta'_1 X_1, -\alpha'_2 X_4; \rho_{14}) - \Phi_{12}(-\alpha'_1 X_3, -\alpha'_2 X_4; \rho_{34}) + \Phi_{123}(-\beta'_1 X_1, -\alpha'_1 X_3, -\alpha'_2 X_4; \Sigma_{134}). \tag{9}$$

For individuals who observe only z the likelihood is

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

$$= f(z) \left[\Phi \left(\frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}} \right) - \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}}; \Sigma_{134|2} \right), \quad (11)$$

$$\Phi \left(\frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}}, \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}}; \frac{\rho_{34} - \rho_{24} \rho_{23}}{\sqrt{1 - \rho^2_{23}} \sqrt{1 - \rho^2_{24}}} \right), \quad (10)$$

For individuals with both z and y observed and the value of y equal to 0 the likelihood is

$$\begin{aligned} L &= f(z) \Pr(Y = 0, R_y = 1, R_z = 1|z) \\ &= f(z) [\Pr(y = 0|z) - \Pr(Y = 0, R_y = 0|z) \\ &\quad - \Pr(Y = 0, R_z = 0|z) \\ &\quad + \Pr(Y = 0, R_y = 0, R_z = 0|z)] \\ &= f(z) \left[\Phi \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}} \right) - \Phi_{12} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}}, \frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}}; \frac{\rho_{13} - \rho_{12} \rho_{23}}{\sqrt{1 - \rho^2_{12}} \sqrt{1 - \rho^2_{23}}} \right) - \Phi_{12} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}}, \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}}; \frac{\rho_{14} - \rho_{12} \rho_{24}}{\sqrt{1 - \rho^2_{12}} \sqrt{1 - \rho^2_{24}}} \right) + \Phi_{123} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}}, \frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}}, \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}} \right) \right] \end{aligned}$$

where

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

For individuals with both z and y observed and the value of y equal to 1 the likelihood is

$$\begin{aligned} L &= f(z) \Pr(Y = 1, R_y = 1, R_z = 1|z) \\ &= f(z) [1 - \Pr(Y = 0|z) - \Pr(R_y = 0|z) - \Pr(R_z = 0|z) + \Pr(Y = 0, R_y = 0|z) + \Pr(Y = 0, R_z = 0|z) - \Pr(Y = 0, R_y = 0, R_z = 0|z)] \\ &= f(z) \left[1 - \Phi \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}} \right) - \Phi \left(\frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}} \right) - \Phi \left(\frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}} \right) + \Phi_{12} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}}, \frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}} \right), \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}}; \frac{\rho_{13} - \rho_{12} \rho_{23}}{\sqrt{1 - \rho^2_{12}} \sqrt{1 - \rho^2_{23}}} \right) + \Phi_{12} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}}, \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}}; \frac{\rho_{14} - \rho_{12} \rho_{24}}{\sqrt{1 - \rho^2_{12}} \sqrt{1 - \rho^2_{24}}} \right) + \Phi_{123} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{12}}}, \frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{23}}}, \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1 - \rho^2_{24}}} \right) \right] \end{aligned}$$

$$\begin{aligned}
 & \frac{\rho_{13} - \rho_{12}\rho_{23}}{\sqrt{1-\rho^2_{12}} \sqrt{1-\rho^2_{23}}}) + \\
 & \Phi_{12} \left(\frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{23}}}, \right. \\
 & \left. \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{24}}}; \right. \\
 & \left. \frac{\rho_{34} - \rho_{24}\rho_{23}}{\sqrt{1-\rho^2_{24}} \sqrt{1-\rho^2_{23}}}) + \right. \\
 & \left. \Phi_{12} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{12}}}, \right. \right. \\
 & \left. \left. \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{24}}}; \right. \right. \\
 & \left. \left. \frac{\rho_{14} - \rho_{12}\rho_{24}}{\sqrt{1-\rho^2_{12}} \sqrt{1-\rho^2_{24}}}) - \right. \right. \\
 & \left. \left. \Phi_{123} \left(\frac{-\beta'_1 X_1 - \frac{\rho_{12}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{12}}}, \right. \right. \right. \\
 & \left. \left. \frac{-\alpha'_1 X_3 - \frac{\rho_{23}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{23}}}, \right. \right. \\
 & \left. \left. \frac{-\alpha'_2 X_4 - \frac{\rho_{24}}{\sigma} (z - \beta'_2 X_2)}{\sqrt{1-\rho^2_{24}}}; \Sigma_{134|2} \right) \right]. \quad (12)
 \end{aligned}$$