Applications and Applied Mathematics: An International Journal (AAM)

Volume 8 | Issue 2 Article 5

12-2013

Analysis of Mixed Correlated Bivariate Negative Binomial and Continuous Responses

F. Razie
Shahid Beheshti University

E. B. Samani
Shahid Beheshti University

M. Ganjali
Shahid Beheshti University

Available at
http://pvamu.edu/aam
Appl. Appl. Math.
ISSN: 1932-9466

Applications and Applied Mathematics:
An International Journal
(AAM)

Vol. 8, Issue 2 (December 2013), pp. 404 – 415

Analysis of Mixed Correlated Bivariate Negative Binomial and Continuous Responses

F. Razie, E. Bahrami Samani and M. Ganjali
Department of Statistics
Shahid Beheshti University
Tehran, Iran
ehsan_bahrami_samani@yahoo.com

Received: April 31, 2013; Accepted: November 26, 2013

Abstract

A general model for the mixed correlated negative binomial and continuous responses is proposed. It is shown how to construct parameter of the models, using the maximization of the full likelihood. Influence of a small perturbation of correlation parameter of the model on the likelihood displacement is also studied. The model is applied to a medical data, obtained from an observational study on women, where the correlated responses are the negative binomial response of joint damage and continuous responses of body mass index. Simultaneous effects of some covariates on both responses are investigated.

Keywords: Latent variable models, Factorization Models, Mixed Correlated Responses, Likelihood Displacement, Body Mass Index, Joint Damage.

AMS-MSC (2010) No.: 62F03

1. Introduction

Some medical science data include correlated discrete and continuous outcomes. The example is in the study of the effect of type of accommodation on body mass index as continuous response and joint damage as negative binomial response (vide, our application in Section 4), where body mass index (BMI) and joint damage are correlated responses in an observational study on women. Furthermore, separate analyses give biased estimates for the parameters and misleading inference. Consequently, we need to consider a method in which these variables can be modeled jointly, for example one may use the factorization of the joint distribution of the outcomes or introduce an unobserved (latent) variable to model the correlation among the multiple outcomes. Many researchers have investigated the mixed correlated data, for example, Olkin and Tate

(1961), Heckman (1978), Poon and Lee (1987), Catalano and Ryan (1992), Fitzmaurice and Laird (1995), Sammel et al. (1997), Lin et al. (2000), Gueorguieva and Agresti (2001), Gueorguieva and Sancora (2006), McClluch (2007), Deleon and Carrier (2007) Yang et al. (2007) and Bahrami Samani et al. (2008).

The main idea of the factorization method is to write the likelihood as the product of the marginal distribution of one outcome and conditional distribution of the second outcome given the first outcomes. Cox and Wermuth (1992), Fitzmaurice and Laird (1995) and Catalano and Ryan (1992) discussed and extended two possible factorizations for modeling a continuous and binary outcome as functions of covariates.

Several models using latent variables have been proposed to analyze multiple non-commensurate outcomes as functions of covariates. Sammel et al. (1997) discussed a model where the outcomes are assumed to be a physical manifestation of a latent variable and conditional on this latent variable. Another approach based on latent variables was proposed by Dunson (2000). A major difference between this approach and Sammel's approach relates to the association between the responses and the covariates. In Dunson's approach, the covariates are not included in the model through the latent variable but rather introduced separately. Pinto and Normand (2009) used an idea similar to the scaled multivariate mixed model proposed by Lin et al. (2000). They introduced a new latent variable model by constraining the parameters of latent model proposed by Dunson (2000) for identifiability without restrictions on the correlation. Yang and Kang (2011) investigate the inferential method for mixed Poisson and continuous longitudinal data with non-ignorable missing values.

The aim of this paper is to use and extend an approach similar to that of Sammel et al. (1997) and Dunson (2000), for modeling of a negative binominal and a continuous variable, by factorization of the joint distribution and the use of latent modeling of bivariate negative binomial and continuous outcomes.

In Section 2, the models and likelihoods are given. In Section 3, simulation studies are used to compare consistency, efficiency and coverage of the multivariate approach with those of the univariate approach. In Section 4, the models are used on a medical data set where joint damage and body mass index (BMI) are correlated responses in an observational study on women. In these models joint damage is a negative binomial response and BMI is continuous response and age, the amount of total body calcium (Ca), job status (employee or housekeeper) and type of accommodation (house or apartment) are explanatory variables. We shall investigate the effects of these explanatory variables on responses simultaneously. The influence of a small perturbation of correlation parameter of the model on the likelihood displacement is also studied. Finally, in Section 5, the paper concludes with some remarks.

2. Models for Mixed Correlated Negative Binomial and Continuous Responses

Let Y_{c_i} denote a continuous response and Y_{nb_i} denote a negative binomial response for the ith of n individuals and X_{c_i} and X_{nb_i} denote $r_c \times 1$ and $r_{nb} \times 1$ vectors of covariates associated with

each response, respectively.

2.1. Univariate Models

One common approach to model multiple responses as functions of covariates is to ignorable the correlation between the responses and fit a separate model to each response variable. We are using a linear regression model for continuous response and a negative binomial regression ($NB(\mu_i, \sigma)$) model,

$$
\begin{aligned}
&Y_{c_i} = X_{c_i}{}'\beta_c + \varepsilon_i,\\
&P(Y_{nb_i} = y_{nb_i}) = \frac{\Gamma\left(y_{nb_i} + \sigma\right)}{\Gamma(\sigma)\Gamma\left(y_{nb_i} + 1\right)} \left(\frac{\mu_i}{\mu_i + \sigma}\right)^{y_{nb_i}} \left(1 - \frac{\mu_i}{\mu_i + \sigma}\right)^{\sigma},\\
&\log\ \mu_i = X_{nb_i}{}'\beta_{nb},\\
&\varepsilon_i \sim N(0, \sigma_c^2),
\end{aligned}
\tag{1}
$$

where $\beta_c = (\beta_{c1}, \ldots, \beta_{cr_c})'$, $\beta_{nb} = (\beta_{nb1}, \ldots, \beta_{nbr_{nb}})'$ and $\sigma > 0$ is a dispersion parameter.

2.2. A Factorization Models

We proposed a model for a correlated negative binomial and a continuous responses based on the factorization of the joint distribution of the responses, $f(y_{nb}, y_c) = f(y_{nb}|y_c)f(y_c)$. The model for the two responses is written as:

$$
\begin{aligned}
&Y_{c_i} = X_{c_i}{}'\beta_c + \varepsilon_i,\\
&Y_{nb_i}|Y_{c_i}, x_{c_i}, x_{nb_i} \sim NB(\mu_i, \sigma),\\
&log\ \mu_i = X_{nb_i}{}'\beta_{nb} + \eta(Y_{c_i} - X_{c_i}{}'\beta_c),\\
&\varepsilon_i \sim N(0, \sigma_c^2),
\end{aligned}
\tag{2}
$$

where η is the parameter for the regression coefficient of Y_{nb_i} on Y_{c_i}. Large absolute values of η indicate a strong correlation between the two responses. if $\eta = 0$, the two responses are independent given the covariates.

Maximum likelihood estimates for the parameters of the factorization method can be obtained with commonly used algorithms for maximizing the likelihood. The log-likelihood function under the factorization model (2) is

$$
\begin{aligned}
l\left(y_{nb}, y_c\right) &= \sum_{i=1}^{n} \{logf(y_{nb_i} / y_{c_i}, x_{c_i}, x_{nb_i}) + logf(y_{c_i} / x_{c_i})\},\\
&= \sum_{i=1}^{n} \{ln[\frac{\Gamma\left(y_{nb_i} + \sigma\right)}{\Gamma\left(\sigma\right)\Gamma\left(y_{nb_i} + 1\right)} - \left(y_{nb_i} + \sigma\right) ln\left(\sigma + \mu_i\right),\\
&\quad + \sigma ln\sigma + y_{nb_i} ln\mu_i + lnf(y_{c_i} / x_{c_i})\}.
\end{aligned}
$$

The vector of parameters β_c and β_{nb} , the parameters of σ_c^2 and η should be estimated.

The factorization of the joint distribution of y_{nb} and y_c can also be consider in reverse order: $f(y_{nb}, y_c) = f(y_c|y_{nb})f(y_{nb})$. The model for the two responses is written as:

$$
\begin{aligned}
&Y_{c_i}|Y_{nb_i}, x_{c_i}, x_{nb_i} = X_{c_i}{}'\beta_c + \xi(Y_{nb_i} - X_{nb_i}{}'\beta_{nb}) + \varepsilon_i \\
&Y_{NB_i} \sim NB(\mu_i, \sigma), \\
&log\ \mu_i = X_{nb_i}{}'\beta_{nb}, \\
&\varepsilon_i \sim N(0, \sigma_c^2),
\end{aligned} \tag{2'}
$$

where ξ is the parameter for the regression coefficient of Y_{c_i} on Y_{nb_i}.

2.3. A Latent Variable Model

We presented a latent variable model where it is assumed that the observed responses are physical manifestations of a latent variable. Conditional on this latent variable, the responses are assumed to be independent and are modeled as functions of fixed covariates and a subject-specific latent variable. Let b_i denote the latent variable. The responses are modeled as function of the latent variable

$$
\begin{aligned}
&Y_{c_i}|b_i, x_{c_i} = X_{c_i}{}'\beta_c + \lambda_c b_i + \varepsilon_i \\
&Y_{nb_i}|b_i, x_{nb_i} \sim NB(\mu_i, \sigma), \\
&log\ \mu_i = X_{nb_i}{}'\beta_{nb} + \lambda_b b_i, \\
&\varepsilon_i \sim N(0, \sigma_c^2), \\
&b_i \sim N(0, \sigma_b^2),
\end{aligned} \tag{3'}
$$

where b_i is a subject -specific latent vatiable. The latent variable shared by both responses induces the correlation and it is assumed that given the latent variable, the two responses are independent. Also b_i is independent of X_{c_i} and X_{nb_i}.

However, $\lambda_b, \lambda_c, \sigma_b^2$ and σ_c^2 are not identifiable. There are four parameters to be estimated but only information from the $Var(Y_c), Var(Y_{nb})$ and $Cov(Y_c, Y_{nb})$. We have to restrict at least two parameters to obtain an identifiable model. Here, we assume $\sigma_b^2 = 1$ and $\lambda_b = \lambda_c = \lambda$.

We can rewrite $(3')$ and obtain the final expression for a latent model for two responses:

$$
\begin{aligned}
&Y_{c_i}|b_i, x_{c_i} = X_{c_i}{}'\beta_c + \lambda b_i + \varepsilon_i, \\
&Y_{nb_i}|b_i, x_{nb_i} \sim NB(\mu_i, \sigma), \\
&log\ \mu_i = X_{nb_i}{}'\beta_{nb} + \lambda b_i, \\
&b_i \sim N(0,1), \\
&\varepsilon_i \sim N(0, \sigma_c^2).
\end{aligned} \tag{3}
$$

The log likelihood for the model is written as:

$$
\begin{aligned}
l(y_{nb}, y_c) &= \sum_{i=1}^{n} logf(y_{nb_i}, y_{c_i} / x_{nb_i}, x_{c_i}), \\
&= \sum_{i=1}^{n} log \int_{-\infty}^{\infty} f(y_{nb_i} / x_{nb_i}, b_i) f(y_{c_i} / x_{c_i}, b_i) f(b_i)\, db_i, \\
&= \sum_{i=1}^{n} log \int_{-\infty}^{\infty} \frac{\Gamma(y_i + \sigma)}{\Gamma(\sigma)\Gamma(y_i + 1)} \left(\frac{exp\left[X_{nb_i}'\beta_{nb} + \lambda b_i\right]}{exp\left[X_{nb_i}'\beta_{nb} + \lambda b_i\right] + \sigma}\right)^{y_i}, \\
&\quad \left(1 - \frac{exp\left[X_{nb_i}'\beta_{nb} + \lambda b_i\right]}{exp\left[X_{nb_i}'\beta_{nb} + \lambda b_i\right] + \sigma}\right)^{\sigma}, \\
&\quad \frac{exp\left(-\frac{(y_{c_i} - x_{c_i}'\beta_c - \lambda b_i)^2}{2\sigma_c^2}\right)}{\sqrt{2\pi\sigma_c^2}} \frac{exp\left(-\frac{b_i^2}{2}\right)}{\sqrt{2\pi}} db_i.
\end{aligned}
$$

The vector of parameters β_c and β_{nb} , the parameters of σ_c^2, λ and σ should be estimated.

3. Simulation Study

We used a simulation study to investigate estimates obtained by the univariate model, factorization model and latent variable model. In this section, simulation study is used to illustrate the application of our proposed models. In this simulation, ε, Y_{nb} and b were generated from a normal distribution, negative binomial distribution and a normal distribution. We thus generated data sets with different cases (1, 2 and 3). For each case, we generated 1000 samples. The data generated from the following cases:

Case 1:

$$
\begin{aligned}
Y_c &= \beta_{0c} + \beta_{1c}X_{1c} + \beta_{2c}X_{2c} + \beta_{3c}X_{3c} + \beta_{4c}X_{4c} + \varepsilon, \\
Y_{nb} / Y_c &\sim NB(\mu, 0.66), \\
log(\mu) &= \beta_{b0} + \beta_{b1}X_{1nb} + \beta_{b2}X_{2nb} + \beta_{b3}X_{3nb} + \beta_{b4}X_{4nb} + \eta(Y_c - E(Y_c)), \\
E(Y_c) &= \beta_{0c} + \beta_{1c}X_{1c} + \beta_{2c}X_{2c} + \beta_{3c}X_{3c} + \beta_{4c}X_{4c}, \\
\varepsilon &\sim N(0, 5).
\end{aligned}
$$

Case 2:

$$
\begin{aligned}
Y_c / Y_{nb} &= \beta_{0c} + \beta_{1c}X_{1c} + \beta_{2c}X_{2c} + \beta_{3c}X_{3c} + \beta_{4c}X_{4c} + \xi\left(Y_{nb} - E\left(Y_{nb}\right)\right) + \varepsilon, \\
E\left(Y_{nb}\right) &= \mu, \\
Y_{nb} &\sim NB(\mu, 0.66), \\
log(\mu) &= \beta_{b0} + \beta_{b1}X_{1nb} + \beta_{b2}X_{2nb} + \beta_{b3}X_{3nb} + \beta_{b4}X_{4nb}, \\
\varepsilon &\sim N(0, 5).
\end{aligned}
$$

Case 3:

$$Y_c = \beta_{0c} + \beta_{1c}X_{1c} + \beta_{2c}X_{2c} + \beta_{3c}X_{3c} + \beta_{4c}X_{4c} + \lambda b + \varepsilon,$$
$$Y_{nb} \sim NB(\mu, 0.66),$$
$$log(\mu) = \beta_{b0} + \beta_{b1}X_{1nb} + \beta_{b2}X_{2nb} + \beta_{b3}X_{3nb} + \beta_{b4}X_{4nb} + \lambda b,$$
$$\varepsilon \sim N(0,5),$$
$$b \sim N(0,1).$$

Also X_{1c} and X_{1nb} are generated from $gamma(100,2)$, X_{2c} and X_{2nb} are generated from gamma$(2,2)$ and X_{3c}, X_{4c}, X_{3nb} and X_{4nb} are generated from Bernilli(0.5).

The vector of coefficients associated with the covariate was chosen $(\beta_{0c}, \beta_{1c}, \beta_{2c}, \beta_{3c}, \beta_{4c}) = (30, 0.1, -0.130, 1.715, 0.981)$, $(\beta_{0b}, \beta_{1b}, \beta_{2b}, \beta_{3b}, \beta_{4b}) = (2, 0.003, -0.223, 0.39, -0.437)$ and $(\sigma_c^2, \sigma, \lambda, \xi, \eta) = (5, 0.66, 0.435, 0.305, 0.25)$.

The data generated from each case is modeled using that case and univariate approach (ignoring the correlation between the outcomes)

$$Y_c = \beta_{0c} + \beta_{1c}X_{1c} + \beta_{2c}X_{2c} + \beta_{3c}X_{3c} + \beta_{4c}X_{4c} + \varepsilon,$$
$$Y_{nb} \sim NB(\mu, 0.66),$$
$$log\ (\mu) = \beta_{b0} + \beta_{b1}X_{1nb} + \beta_{b2}X_{2nb} + \beta_{b3}X_{3nb} + \beta_{b4}X_{4nb},$$
$$\varepsilon \sim N(0,5).$$

The models were fitted using nlminb from R to assure that the same numerical algorithms were used to maximize the likelihoods.

Table 1. Results of the simulations study for case 1

Case 1		Model (2)		Uni. model	
Parameter	Real value	Est.	S.E	Est.	S.E.
NJ					
Constant	2	2.513	0.470	-0.957	0.291
Age	0.003	-0.008	0.009	0.006	0.006
Ca	-0.223	- 0.323	0.071	0.015	0.046
Job	0.039	0.112	0.089	0	0.054
TA	-0.437	-0.517	0.086	0.051	0.065
σ	0.66	0.647	0.029	0.078	0.010
BMI					
Constant	30	32.074	1.669	36.737	1.642
Age	0.100	0.057	0.033	-0.012	0.032
Ca	-0.130	-0.45	0.261	-0.082	0.219
Job	1.715	1.930	0.33	-0.508	0.331
TA	0.981	0.744	0.318	0.165	0.319
σ_c^2	5	5.054	0.110	5.173	0.108
η	0.250	0.249	0.003	-	-

Table 2. Results of the simulations study for case 2

Case 2		Model (2')		Uni. model	
Parameter	Real value	Estimate	S.E.	Este	S.E.
NJ					
Constant	2	2.001	0.289	1.458	0.189
Age	0.003	0.003	0.004	-0.001	0.004
Ca	-0.223	- 0.239	0.034	0.037	0.027
Job	0.039	0.049	0.044	0.001	0.044
TA	-0.437	-0.487	0.045	-0.099	0.037
σ	0.66	0.666	0.045	0.578	0.073
BMI					
Constant	30	30.41	1.727	33.645	1.612
Age	0.100	0.103	0.032	0.032	0.32
Ca	-0.130	-0.289	0.251	0.148	0.227
Job	1.715	2.186	0.327	-0.294	0.328
TA	0.981	0.659	0.367	0.145	0.319
σ_c^2	5	5.009	0.112	5.271	0.105
ξ	0.305	0.469	0.077	-	-

Table 3. Results of the simulations study for case 3

Case 3		Model (3)		Uni. model	
Parameter	Real value	Est.	S.E.	Est.	S.E.
NJ					
Constant	2	1.649	0.386	1.141	0.272
Age	0.003	0.006	0.005	-0.007	0.005
Ca	-0.223	- 0.204	0.039	0.018	0.038
Job	0.039	0.083	0.052	-0.088	0.046
TA	-0.437	-0.468	0.052	0.019	0.049
σ	0.66	0.612	0.064	0.392	0.023
BMI					
Constant	30	27.186	1.672	34.356	1.649
Age	0.100	0.158	0.032	0.039	0.32
Ca	-0.130	-0.219	0.242	-0.026	0.231
Job	1.715	1.998	0.318	0.288	0.323
TA	0.981	0.833	0.319	-0.339	0.322
σ_c^2	5	5.115	0.109	5.295	0.104
λ	0.435	0.407	0.058	-	-

Tables 1-3 contains the average estimated values of

$$(\beta_{0c}, \beta_{1c}, \beta_{2c}, \beta_{3c}, \beta_{4c}), (\beta_{0b}, \beta_{1b}, \beta_{2b}, \beta_{3b}, \beta_{4b}), \sigma_c^2, \sigma$$

and λ (for model (3)), ξ (for model (2')) and η (for model (2)) for n = 1000. The results are summarized as follows. The parameter estimates by the model (2), model (2') and model (3) are close to the true values of the parameters.

4. Application and Sensitivity Analysis

4.1. Application

In this section, we use the Mixed correlated models in (2) and (3) for the medical data set describe in the following subsection. The medical data set is obtained from an observational study on women in the Taleghani hospital of Tehran, Iran. These data record the number of joint

damage (*NJ*) as negative binomial responses and body mass index (*BMI*) as continuous responses for 163 patients. These patients are heavy body.

Joint damage is a disease of bone in which the bone mineral density (BMD) is reduced, bone micro architecture is disrupted and the amount and variety of non-collage nous proteins in bone is altered. *BMI* is a statistical measure of the weight of body mass index. A person body mass index may be accurately calculated using any of the formulas such as $BMI = \frac{W}{H^2}$ where W is weight and *H* is height. Also, The heavy body can result in damages to joints of knee and ankle, etc. These two variables, joint damage and *BMI* correlated variables, and they have to be modeled. Explanatory variables which affect these variables are: (1) amount of total body calcium (*Ca*), (2) job status (*Job*, employee or housekeeper), (3) type of the accommodation (*Ta*, house or apartment) and (4) *age*.

We used a test to investigate over dispersion for count response. Deviance and Pearson Chi-Square divided by the degrees of freedom are used to detect over dispersion or under dispersion in the Poisson regression. Values greater than 1 indicate over dispersion, that is, the true variance is bigger than the mean, values smaller than 1 indicate under dispersion, the true variance is smaller than the mean. Evidence of under dispersion or over dispersion indicates inadequate fit of the Poisson model. We can test for over dispersion with a likelihood ratio test based on Poisson and negative binomial distributions. This test tests equality of the mean and the variance imposed by the Poisson distribution against the alternative that the variance exceeds the mean.

For the negative binomial distribution, the variance of count response (Y_{count}) is

$$Var\left(Y_{count}\right) = E\left(Y_{count}\right) + kE\left(Y_{count}^2\right),$$

where $k > 0$, the negative binomial distribution reduces to Poisson when $k = 0$. The null hypothesis is $H_0: k = 0$ and the alternative hypothesis is $H_1: k > 0$). Use the *LR*(likelihood ratio) test, that is, compute *LR* statistic, *-2(L L (Poisson) - LL (negative binomial)*), where *LL* is *log(likelihood).* The asymptotic distribution of the *LR* statistic has probability mass of one half at zero and one half Chi-sq distribution with 1 *df* (see Cameron and Trivedi, 1998). To test the null hypothesis at the significance level α, use the critical value of Chi-sq distribution corresponding to significance level α , that is reject H_0 if *LR* statistic $> \chi^2_{1-2\alpha,1}$.

In this data, we calculated *LL* (Poisson for *NJ*)=533.513, *LL* (negative binomial for *NJ*) = 541.701 and *-2(LL (Poisson) - LL (negative binomial))* = -16.376 (with 1 *d.f.* and *P-value*=0.003). So, *NJ* has over dispersion and *NJ* is negative binomial distribution.

Results of using three models (model 1, 2 and 3) are given in Table 4. We used the univariate model (model 1), the factorization model (model 2) and the latent variable model (model 3) as described Section 2.4) to estimate the parameters of the models.

Univariate model (model 1) shows significant no effect of covariates on BMI and the number of joint damage. For the factorization models (model 2) shows significant effect of amount of total body calcium and job status on the frequency of joint damage. From these effects we can infer

that the amount of total body calcium have a negative impact on the frequency of number of joint damage. Job status has a positive impact on the frequency of number of joint damage. $\hat{\sigma}$ indicates that the increase of dispersion has a positive impact on the frequency of number of joint damage. In these models, correlation parameter $\hat{\eta}$ is strongly significant. It shows a positive correlation between BMI and the number of joint damage. The estimated variance of BMI ($\hat{\sigma}_c^2$ obtained by the factorization model is less than those of univariate model. The factorization model (model 2) gives the same results as the latent variable model (model 3). In the latent variable model, correlation parameter $\hat{\lambda}$ is strongly significant. The better performance of the latent variable model over the factorization model.

Table 4. Estimation results of the four models (NJ: Negative Binomial Response and BMI: Continuous Response) of real data, parameter estimates highlighted in **bold** are significant at 5 % level.)

Model	Model (1)		Model (2)		Model (3)	
Parameter	Est.	S.E.	Est.	S.E.	Est.	S.E.
NJ						
Constant	**.432**	**.233**	**.805**	**.130**	**.731**	**.116**
Age	0.003	0.016	0.002	0.011	0.002	0.010
Ca	-0.223	0.517	**-0.977**	**.401**	**-0.952**	**.391**
Job	0.039	0.031	**.057**	**.020**	**.055**	**.017**
TA	-0.437	0.441	-0.405	0.333	-0.404	0.333
σ	**.250**	**.029**	**.243**	**.015**	**.240**	**.090**
BMI						
Constant	**.526**	**.231**	**.517**	**.665**	**.437**	**.570**
Age	0.100	0.131	0.101	0.140	0.101	0.130
Ca	-0.130	0.085	-0.104	0.074	-0.089	0.069
Job	1.715	2.168	1.721	2.421	1.712	2.420
TA	0.981	0.775	0.980	0.773	0.978	0.765
σ_c^2	**.524**	**.588**	**.919**	**.588**	**.341**	**.522**
η	-	-	**.251**	**.070**	-	-
λ	-	-	-	-	**.435**	**.055**
-loglike	1156.013		1047.013		1037.311	

4.2. Sensitivity Analysis

Likelihood displacement is a very important concept as it provides a general approach to study the problem of influence. The method of local influence was introduced by Cook (1986) and modified by Billor and Loynes (1993) as a general tool for assessing the influence of local departures from the assumptions underlying the statistical models.

Perturbations of the model influence key results of the analysis are to compare the results derived from the original and perturbed models. The influence graphs introduced in this Section are simply devices to facilitate such comparisons when the behavior of the parameter estimates is of interest. This article shows that local-influence analysis of perturbations of the correlation parameters of models. The log-likelihood for the unperturbed and perturbed models are denoted by $L(\theta)$ and $L(\theta|\omega)$, respectively. The perturbed likelihood $L(\theta|\omega)$obtained after the likelihood have been perturbed by an amount ω where ω is a $q \times 1$ vector which is restricted to some open subset Ω of R^q.

Then the likelihood displacement LD(ω) is defined by

$$LD(\omega) = 2[L(\hat{\theta}) - L(\hat{\theta}_\omega)]. \quad (4.1)$$

Generally, one introduce perturbations into the model through the $q \times 1$ vector ω which is restricted to some open subset Ω of R^q and θ is $p \times 1$ vector of unknown parameters. Cook (1986) proposed the maximum normal curvature C_{max}. The C_{max} is defined by

$$C_{max} = max_l C_l ,$$

where C_l is the lifted line in the direction l can be easily calculated by

$$C_l = 2\left|l^T \Delta^T (\ddot{L})^{-1} \Delta l\right|, \qquad (4.2)$$

where $\Delta_i = \left.\frac{\partial^2 l_i(\theta / \omega_i)}{\partial \omega_i \partial \eta}\right|_{\theta=\hat{\theta},\omega_i=0}$ and define Δ as the $p \times n$ matrix with Δ_i as its *ith* column and $\ddot{L}$ denote the $p \times p$ matrix of second-order derivatives of $l(\theta / \omega_0)$, where there is an ω_0 in Ω, with respect to θ, also evaluated at $\theta = \hat{\theta}$. Obviously, C_l can be calculated for any direction *l*. One evident choice is the vector l_i containing one in the *ith* position and zero elsewhere, corresponding to the perturbation of the *ith* weight only. The corresponding local influence measure, denoted by C_l, then becomes $C_l = 2\left|\Delta_i^T (\ddot{L})^{-1} \Delta_i\right|$. Another important direction is the direction l_{max} of maximal normal curvature. C_{max} Also, C_{max} is the largest Eigen value of $\Delta_i^T (\ddot{L})^{-1} \Delta_i$ and l_{max} is the corresponding eigenvector.

Let see how we can use this approach for our purposes. Condition for independent responses $(\eta = 0)$ and condition for Poisson distribution for NJ is $(DI = \frac{Var(NJ)}{E(NJ)} = 1)$ which gives the following condition for not having over dispersion $h = \frac{Var(NJ)}{E(NJ)} - 1 = 0$ in the model (3). We can use maximal normal curvature for the effect of perturbation from independent responses to correlated responses and the perturbation from Poisson distribution to negative binomial distribution (or over dispersion).

Let $= (\eta, h)$. Here, $\omega_0 = (0,0)$ for each model and $q = 2$. Denote the log-likelihood function by

$$L(\theta|\omega) = \sum_{i=1}^{n} L_i\,(\theta|\omega),$$

where $L_i(\theta|\omega)$ is the contribution of the ith individual to the log-likelihood and θ is the parameter vector. Here, $L(\eta|\omega = \omega_0)$ is the log-likelihood function which corresponds to independent responses. Suppose ω can be perturbed around 0. Let $\hat{\theta}$ be MLE estimator for θ obtained by maximizing $L(\theta) = L(\theta|\omega = \omega_0)$ and let $\hat{\theta}_\omega$ denote the MLE estimator for θ under $L(\theta|\omega)$. Now one compare $\hat{\theta}_\omega$ and $\hat{\theta}$ as local influence. Strongly different estimates show that the estimation procedure is highly sensitive to such modification. We can quantify the differences using maximal normal curvature defined as (4.2).

To search for Sensitivity analysis we find C_{max}. This is confirmed by the curvature $C_{max} =$

12.013. This curvature indicates extreme local sensivity. These curves show a high curvature aroundω_0, so that the differing values of ω affects the model (3) results, hence final results of the model (3), is highly sensitive to correlated responses and negative distribution for NJ.

5. Conclusion

We presented different approaches to model correlated negative binomial and continuous outcomes. We proposed new multivariate variable models. We also implemented likelihood approach based. Simulation results suggest that the four approaches lead to consistent estimates of the regression parameters. This suggests that the correlation between the outcomes will not be worse than the assumption of independence. In contrast to the factorization approach, the latent variable model presented is easily extended to several continuous and/or several negative binomial outcomes by including additional latent variables as long as the outcomes are positively correlated. However, some of the assumptions of the model, such as the distribution of the latent variables, are not easily assessed. In the presence of missing observations in one of the outcomes, the factorization approach only uses the complete cases or it requires the *EM*-algorithm to include all the cases in the analysis (Fitzmaurice and Laird, 1997). This is not the case with the latent model. If the missing data is missing at random or missing completely at random (Little and Schluchte, 1987), this situation can be easily accommodated due to the conditional independence of the outcomes given the latent variable. Furthermore, the latent variable model is easily fitted using standard software.

REFERENCES

Bahrami Samani, E., Ganjali, M. and Khodaddadi, A. (2008). A Latent Variable Model for Mixed Continuous and Ordinal Responses. *Journal of Statistical Theory and Applications*, **(3)**, 337-349.

Billor, N. and Loynes, R.M. (1993). Local Influence: A New Approach, *Comm. Statist.-Theory Meth.,* 1595-1611.

Catalano, P. and Ryan, L. M. (1992). Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association*, **(3)**, 1078-1095.

Cameron, A.C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.

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

Cook, R. D. (1986). Assessment of Local Influence (with discussion). *Journal. Royal Statist. Soc*., Ser. B., , 133-169.

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

Dunson D. B. (2000). Bayesian latent variable models for clustered mixed outcomes, *Journal of the Royal Statistical Society, Series B: Statistical Methodology,* **(2)**, 355-366.

Fitzmaurice, G. M. and Laird, N. M. (1995). Regression models for Bivariate discrete and continuous outcome with clustering, *Journal of the American Statistical Association,* 845-852.

Fitzmaurice, G. M. and Laird, N. M. (1997). Regression models for mixed discrete and continuous

responses with potentially missing value. *Biometrics*, 110-122.

Gueorguieva, R.V. and Agresti, A. (2001). Correlated Probit Model for Joint Modeling of Clustered Binary and Continuous Response, *Journal of the American Statistical Association*, 1102-1112.

Gueorguieva, R.V. and Sanacora, G. (2006). Joint Analysis of Repeatedly Observed Continuous and Ordinal Measures of Disease Severity, *Statistics in Medicine* **(8)**, 1307-1322.

Heckman, J. J. D. (1978). Endogenous variable in a simultaneous Equation system, *Econometrical,* (6), 931-959.

Lin X., Ryan L., Sammel M., Zhang D., Padungtod C. and Xu X. (2000). A scaled linear mixed model for multiple outcomes. *Biometrics*, **(2)**, 593-601.

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

McCulloch, C. (2007). Joint modeling of mixed outcome type using latent variables*, statistical methods in Medical Research*, **(1)**, 53-73.

Olkin L. and Tate R. F. (1961). Multivariate correlation models with mixed discrete and continuous variables, *Annals of Mathematical Statistics*, 448-456.

Sammel M. D., Ryan L. M and Legler J. M. (1997). Latent variable models for mixed discrete and continuous outcomes, *Journal of the Royal Statistical Society, Series B: Methodological*, 667-678.

Pinto, A. T. and Normand, S. L. T. (2009). Correlated bivariate continuous and bi- nary outcomes: Issues and applications. *Statistics in Medicine*, 1753-1773.

Poon, W. Y. and Lee, S. Y. (1987). Maximum likelihood estimation of multivariate polyserial and polychromic correlation coefficients. *Psychometrika* **(3):** 409-430.

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

Yang, Y. and Kang, J. (2011). Joint analysis of mixed Poisson and continuous longitudinal data with non-ignorable missing values, *Computational Statistics Data Analysis,* 193-207.