Bivariate Multinomial Models

by

© Bingrui Sun

A thesis submitted to the

School of Graduate Studies

in partial fulfilment of the

requirements for the degree of

Doctor of Philosophy in Statistics

Department of Mathematics and Statistics

Memorial University of Newfoundland

May 2014

St. John's Newfoundland Canada

Abstract

Analyzing multivariate categorical data is an important and practical research topic. Even though there exist many studies on the analysis of bivariate (possibly multivariate) categorical data, the modeling of correlations among the bivariate multinomial variables is, however, not adequately addressed. In this thesis, we develop three correlation models for bivariate multinomial data. The first model accommodates fully specified marginal probabilities and uses a bivariate normal type conditional probability relationship to model the correlations of the bivariate multinomial variables. Next, we propose a random effects based familial type model to accommodate the correlations, where conditional on the random effects the marginal probabilities are fully specified. The third model is developed by considering the marginal probabilities of one variable as fully specified, and using conditional multinomial logistic type probability model to accommodate correlations. The estimation of the parameters for all three models is discussed in details through both simulation studies and analysis of real data.

Acknowledgements

I would like to sincerely thank my thesis supervisor Dr. Brajendra Sutradhar for his advice, inspiration, and unconditional moral support throughout the four years of my PhD program and in preparation of my thesis. I would also like to thank my PhD program co-supervisor Dr. Alwell Oyet for his support throughout the program. I also want to thank my internal as well as external examiners for their valuable comments during my defense presentation, which have been helpful to prepare this final version.

I want to acknowledge the financial support provided by the School of Graduate Studies, Department of Mathematics and Statistics, and Professors Brajendra Sutradhar and Alwell Oyet in the form of Graduate Assistantship and Teaching Assistantship. Further I want to thank the Dean of Science and Graduate Students Union at Memorial University of Newfoundland for providing me financial support to attend national and international conferences.

It is my great pleasure to thank my family and all my friends for their support and encouragement.

Contents

Abstract **ii**

Acknowledgements **iii**

List of Tables **vii**

List of Figures **x**

1 Introduction **1**

1.1 Models for univariate multinomial data 2

1.1.1 Basic properties of the univariate multinomial model 3

1.1.2 Likelihood estimation for the univariate multinomial logit model 3

1.1.3 Contingency table based univariate multinomial logit model . 4

1.1.3.1 One categorical covariate with L levels 4

1.1.3.2 Remarks on an alternative loglinear model 11

1.1.3.3 More than one categorical covariate having interactions 13

1.2 Existing bivariate multinomial models 18

1.2.1 Existing bivariate multinomial models involving individual level covariates 23

1.2.2 Existing bivariate multinomial models with categorical covariates 26

1.3 Plan of the thesis 28

2 Linear Conditional Bivariate Multinomial (LCBM) Fixed Effects Model 31

2.1 Fully Specified Marginal Probabilities Based LCBM Fixed Effects Model and Properties 32

2.1.1 Basic properties of the LCBM fixed effects model 37

2.2 Estimation for the LCBM fixed model 42

2.2.1 JGQL approach 43

2.2.2 SSGQL approach 45

2.2.3 A simulation study 46

2.2.4 Diabetic retinopathy data analysis 63

2.2.4.1 An application of the linear conditional bivariate binary (LCBB) fixed model 66

2.2.4.2 An application of the linear conditional bivariate multinomial (LCBM) model 71

3 Individual Random Effects Based Bivariate Multinomial Mixed (BMM) Model 79

3.1 The model 80

3.1.1 Marginal probabilities conditional on individual specific random effect 80

3.1.2 Unconditional moment properties of the model 81

3.1.3 Remarks on similar random effects based models 82

3.2 Inferences for the BMM model 83
3.2.1 Joint GQL approach 84
3.2.2 MLE approach 85
3.2.3 Remarks on properties of JGQL and MLE estimates 87
3.2.4 A simulation study 89
3.2.5 Reanalysis of diabetic retinopathy data 100
3.2.5.1 An application of the bivariate binary mixed (BBM) model 100
3.2.5.2 An application of the bivariate multinomial mixed (BMM) model 104

4 Nonlinear Conditional Bivariate Multinomial (NLCBM) Fixed Model 108
4.1 The model 109
4.2 Likelihood estimation for the NLCBM fixed model 111
4.2.1 Estimation of the parameters 112

5 Concluding Remarks 115

Bibliography 117

List of Tables

1.1 Contingency table in the cross-sectional setup based on one covariate with L levels. 6

1.2 Contingency table of multinomial response with two categorical covariates. 16

1.3 Marginal and joint probabilities for bivariate multinomial responses Y and Z. 19

1.4 Bivariate multinomial observed data and underlying marginal and joint probabilities with no covariate. 20

2.1 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D1 with sample size n=200. 54

2.2 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D2 with sample size n=200. 55

2.3 The SM (simulated mean), SSE (simulated standard error), and ESE (estimated standard error) of the JGQL, SSGQL and QL estimates under D3 with sample size n=200. 56

2.4 The SM (simulated mean), SSE (simulated standard error), and ESE (estimated standard error) of the JGQL, SSGQL and QL estimates under D4 with sample size n=200. 57
2.5 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D5 with sample size n=200. 58
2.6 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D6 with sample size n=200. 59
2.7 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D7 with sample size n=200. 60
2.8 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D8 with sample size n=200. 61
2.9 Descriptive statistics of left and right eyes diabetic retinopathy status. 67
2.10 JGQL and SSGQL estimation results for the diabetic retinopathy data under the LCBB model. 69
2.11 Descriptive statistics of left and right eyes diabetic retinopathy status. 73
2.12 SSGQL estimation results for the diabetic retinopathy data under the LCBM model. 78

3.1 The SM (simulated mean), SSE (simulated standard error) and MSE (mean squared error) of the JGQL, MLE and QL estimates for selected regression parameter values and $\sigma_\gamma = 0.1$. 95

3.2 The SM (simulated mean), SSE (simulated standard error) and MSE (mean squared error) of the JGQL, MLE and QL estimates for selected regression parameter values and $\sigma_\gamma = 0.35$. 96

3.3 The SM (simulated mean), SSE (simulated standard error) and MSE (mean squared error) of the JGQL, MLE and QL estimates for selected regression parameter values and $\sigma_\gamma = 0.5$. 97

3.4 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and MLE estimates for selected regression parameter values and $\sigma_\gamma = 0.75$. 98

3.5 The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and MLE estimates for selected regression parameter values and $\sigma_\gamma = 1.0$. 99

3.6 JGQL and MLE estimation results for the diabetic retinopathy data under the BBM model. 102

3.7 JGQL estimation results for the diabetic retinopathy data under the BMM model. 107

List of Figures

2.1 MSE comparison of regression parameters between JGQL and SSGQL under D1 for n=200, 300 and 1000. 62

2.2 MSE comparison of correlation index parameters between JGQL and SSGQL under D1 for n=200, 300 and 1000. 63

2.3 MSE comparison of category regression parameters between JGQL and SSGQL under D7 for n=200, 300 and 1000. 64

2.4 MSE comparison of covariate regression parameters between JGQL and SSGQL under D7 for n=200, 300 and 1000. 65

2.5 MSE comparison of correlation index parameters between JGQL and SSGQL under D7 for n=200, 300 and 1000. 66

2.6 Histogram of correlations between left and right eye binary retinopathy status under the LCBB model. 72

3.1 Histogram of correlations between left and right eye binary retinopathy status under the BBM model. 104

Chapter 1

Introduction

There are many practical situations, for example, in many biomedical studies, where categorical responses are collected from a large number of independent individuals. In these situations, covariates are also collected. For example, in the Wisconsin Epidemiologic Study of Diabetic Retinopathy (WESDR) [Williamson, Kim and Lipsitz (1995)], diabetic retinopathy status on a ten point interval scale for left and right eyes, along with information on various associated covariates such as duration of diabetes, glycosylated hemoglobin level and so on, were collected from 996 independent patients. For convenience, the responses collected on ten point interval scale were grouped into four categories: none, mild, moderate and proliferative. These authors addressed the scientific question whether or not covariates have any effects on the categorical/multinomial retinopathy status of the left and right eyes. The modeling and analysis for this type of bivariate multinomial data will be discussed in Section 1.2 and subsequently in other chapters of the thesis. Note that when bivariate data are modeled, it requires the modeling of correlations on top of the marginal probabilities for each of the multinomial variables. Before we begin discussing such correlation models, we first review the existing modeling for univariate categorical data involving individual level or categorical covariates.

1.1 Models for univariate multinomial data

Let Y_i denote the multinomial response variable, for example, the right eye diabetic retinopathy status in the above WESDR study. If there are J categories or status for the response, then we denote this variable by $Y_i = (Y_{i1}, \ldots, Y_{ij}, \ldots, Y_{i,J-1})'$. Assuming that the response belongs to the jth ($j = 1, \ldots, J$) category, one represents this jth category response by $y_i^{(j)}$, which is naturally a realized value of y_i. Thus

$$y_i^{(j)} = \begin{cases} (y_{i1}^{(j)}, \ldots, y_{ij}^{(j)}, \ldots, y_{i,J-1}^{(j)})' = (\mathbf{0}'_{j-1}, 1, \mathbf{0}'_{J-1-j})', \; j = 1, \ldots, J-1, \\ (0, \ldots, 0, \ldots, 0)', \; j = J. \end{cases} \tag{1.1}$$

Because covariates play a role for an individual response to be in a particular category, it is standard to use the regression based multinomial logits to model the probability for such responses. Let $x_i = (x_{i1}, \ldots, x_{im}, \ldots, x_{ip})'$ be the p-dimensional covariate associated with y_i. One then expresses the marginal probabilities for $y_i = y_i^{(j)}$ as

$$\begin{aligned} P[Y_{ij} = 1] = Pr(Y_i = y_i^{(j)}) &= \begin{cases} \frac{\exp(\beta_{j0} + x_i'\beta_j)}{1 + \sum_{u=1}^{J-1} \exp(\beta_{u0} + x_i'\beta_u)}, \; j = 1, \ldots, J-1, \\ \frac{1}{1 + \sum_{u=1}^{J-1} \exp(\beta_{u0} + x_i'\beta_u)}, \; j = J. \end{cases} \\ &= \pi_{iy}^{(j)}, j = 1, \; \ldots, J. \end{aligned} \tag{1.2}$$

This type of marginal probability for a multinomial response to be in a particular category has been well discussed in the literature. For example, we refer to Agresti (2002, Section 7.1.3, Eqn. (7.2)). The basic properties of this model can be written as follows.

1.1.1 Basic properties of the univariate multinomial model

Because $E(Y_{ij}) = P[Y_{ij} = 1] = \pi_{iy}^{(j)}$, the mean vector of the multinomial response Y_i is given by

$$\begin{aligned} E(Y_i) &= [E(Y_{i1}), \ldots, E(Y_{ij}), \ldots, E(Y_{i,J-1})]' \\ &= (\pi_{iy}^{(1)}, \ldots, \pi_{iy}^{(j)}, \ldots, \pi_{iy}^{(J)})', \\ &= \pi_{iy}, \end{aligned}$$

with $\pi_{iy}^{(j)}$ as given in (1.2). Similarly, the covariance matrix of Y_i has the form

$$Cov(Y_i) = \begin{pmatrix} var(Y_{i1}) & \ldots & cov(Y_{i1}, Y_{ij}) & \ldots & cov(Y_{i1}, Y_{iJ}) \\ \vdots & & \vdots & & \vdots \\ & & var(Y_{ij}) & \ldots & cov(Y_{ij}, Y_{iJ}) \\ \vdots & & \vdots & & \vdots \\ cov(Y_{i1}, Y_{iJ}) & \ldots & \ldots & \ldots & var(Y_{iJ}) \end{pmatrix},$$

where $var(Y_{ij}) = \pi_{iy}^{(j)}(1 - \pi_{iy}^{(j)})$ and $cov(Y_{ij}, Y_{iu}) = -\pi_{iy}^{(j)}\pi_{iy}^{(u)}$ for $j \neq u$.

1.1.2 Likelihood estimation for the univariate multinomial logit model

In the multinomial probability model (1.2), β_{j0} denotes the intercept parameter under the jth category with $\beta_{J0} = 0$, and β_j is the effect of x_i on y_{ij} for $j = 1, \ldots, J-1$ with $\beta_J = (0, \ldots, 0)'$ by convention. One may use the likelihood method and estimate

$$\beta = (\beta_{10}, \beta_1', \ldots, \beta_{j0}, \beta_j', \ldots, \beta_{J-1,0}, \beta_{J-1}')' \tag{1.3}$$

by maximizing the multinomial likelihood function

$$L(\beta) = \prod_{i=1}^{n} \prod_{j=1}^{J} (\pi_{iy}^{(j)})^{y_{ij}}. \tag{1.4}$$

It is equivalent to solving the log likelihood estimating equation for $\beta_j^* = (\beta_{j0}, \beta_j')'$

$$\begin{aligned} \frac{\partial L(\beta)}{\partial \beta_j^*} &= \frac{\partial}{\partial \beta_j^*} \left[\sum_{i=1}^{n} \sum_{j=1}^{J-1} \binom{1}{x_i}' \beta_j^* - \sum_{i=1}^{n} ln \left\{ 1 + \sum_{j=1}^{J-1} \binom{1}{x_i}' \beta_j^* \right\} \right] y_{ij} \\ &= \sum_{i=1}^{n} \left[\binom{1}{x_i} y_{ij} - \binom{1}{x_i} \pi_{iy}^{(j)} \right] \\ &= \sum_{i=1}^{n} \left[\binom{1}{x_i} (y_{ij} - \pi_{iy}^{(j)}) \right] \end{aligned}$$

[Agresti (2002, Section 7.1.4, p. 273)], leading to the likelihood equation

$$\frac{\partial L(\beta)}{\partial \beta} = \sum_{i=1}^{n} \left[\mathbf{I}_{J-1} \bigotimes \binom{1}{x_i} \right] (y_i - \pi_{iy}) = 0, \tag{1.5}$$

for the estimation of β. One can use the well-known Newton-Raphson method to solve this equation.

1.1.3 Contingency table based univariate multinomial logit model

1.1.3.1 One categorical covariate with L levels

Note that the multinomial probability model (1.2) is written using individual level general covariate x_i. But in practice, the covariates may be of categorical nature with various levels. If all covariates involved in the study are categorical, one may then express the likelihood function in (1.4) in a simpler product multinomial likelihood function form. We demonstrate this below for single covariate ($p = 1$) with L levels. In this case we can write x_i only for x_{i1}. Suppose that we use $L-1$ dummy covariates

$x_{i(1)}, \ldots, x_{i(l)}, \ldots, x_{i(L-1)}$ to represent the L levels. These covariates can take the values as follows.

$$(x_{i(1)}, \ldots, x_{i(l)}, \ldots, x_{i(L-1)}) = \begin{cases} (1, 0, \ldots, 0), & \text{level } 1 \\ \vdots & \\ (\mathbf{0}'_{l-1}, 1, \mathbf{0}'_{L-l-1}), & \text{level } l \\ \vdots & \\ (0, \ldots, 0, 1), & \text{level } L-1 \\ (0, \ldots, 0), & \text{level } L. \end{cases} \tag{1.6}$$

Following (1.2) we may then write the marginal probabilities $\pi_{iy}^{(j)}$ for $y_i = y_i^{(j)}$ as

$$\pi_{iy}^{(j)} = Pr(y_i = y_i^{(j)}) = \begin{cases} \frac{\exp(\beta_{j0} + \sum_{l}^{L-1} x_{i(l)}\beta_{jl})}{1 + \sum_{u=1}^{J-1} \exp(\beta_{u0} + \sum_{l=1}^{L-1} x_{i(l)}\beta_{ul})}, & j = 1, \ldots, J-1, \\ \frac{1}{1 + \sum_{u=1}^{J-1} \exp(\beta_{u0} + \sum_{l=1}^{L-1} x_{i(l)}\beta_{ul})}, & j = J, \end{cases} \tag{1.7}$$

where β_{jl} is the effect of $x_{i(l)}$ on y_{ij} with $\beta_{Jl} = 0$ $(l = 1, \ldots, L)$ and $\beta_{jL} = 0$ $(j = 1, \ldots, J)$. Next, suppose that for the individuals with covariate level l $(l = 1, \ldots, L)$, the probability that the response of an individual in this group belongs to the jth category is denoted by $\pi_{y(l)}^{(j)}$. Note that in this problem n individuals can be grouped into L distinct (or non-overlapping) subgroups based on their covariate levels. For this reason, we use the notation $i \in l$ to represent that the ith individual has covariate level l and for this group i ranges from 1 to n_l, such that $\sum_{l=1}^{L} n_l = n$. We express

Table 1.1: Contingency table in the cross-sectional setup based on one covariate with L levels.

$X \setminus Y$	1	$\ldots$	j	$\ldots$	J	Total
level 1	$n_{[1]1}$	$\ldots$	$n_{[1]j}$	$\ldots$	$n_{[1]J}$	$\tilde{n}_1$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
level l	$n_{[l]1}$	$\ldots$	$n_{[l]j}$	$\ldots$	$n_{[l]J}$	$\tilde{n}_l$
$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
level L	$n_{[L]1}$	$\ldots$	$n_{[L]j}$	$\ldots$	$n_{[L]J}$	$\tilde{n}_L$
Total	n_1	$\ldots$	n_j	$\ldots$	n_J	n

$\pi^{(j)}_{y(l)}$ as

$$\begin{aligned} \pi^{(j)}_{y(l)} &= Pr(Y_i = y_i^{(j)} | i \in l) \\ &= \begin{cases} \frac{\exp(\beta_{j0}+\beta_{jl})}{1+\sum_{u=1}^{J-1}\exp(\beta_{u0}+\beta_{ul})}, \; j = 1, \ldots, J-1; \; l = 1, \ldots, L-1 \\ \frac{1}{1+\sum_{u=1}^{J-1}\exp(\beta_{u0}+\beta_{ul})}, \; j = J; \; l = 1, \ldots, L-1 \\ \frac{\exp(\beta_{j0})}{1+\sum_{u=1}^{J-1}\exp(\beta_{u0})}, \; j = 1, \ldots, J-1; \; l = L \\ \frac{1}{1+\sum_{u=1}^{J-1}\exp(\beta_{u0})}, \; j = J; l = L. \end{cases} \end{aligned} \tag{1.8}$$

Suppose that the observed counts under all levels ($l = 1, \ldots, L$) are given as in Table 1.1 above.

Now using (1.8), the product multinomial likelihood function for the observed data in Table 1.1 may be written as

$$L = \prod_{l=1}^{L} L_{(l)}, \tag{1.9}$$

where

$$L_{(l)} = \frac{\tilde{n}_l!}{\prod_{j=1}^{J} n_{[l]j}!} \prod_{j=1}^{J} (\pi_{y(l)}^{(j)})^{n_{[l]j}}. \tag{1.10}$$

Note that one may estimate the parameters involved in (1.8) by maximizing the product multinomial likelihood in (1.9). However, by expressing the exponents in the probabilities (1.8) in a linear regression form involving all parameters, one may obtain a simpler likelihood estimating equation. For this purpose, let $\theta^* = (\beta_1^{*'}, \ldots, \beta_j^{*'}, \ldots, \beta_{J-1}^{*'})'$ denote the vector of parameters involved in model (1.8), with $\beta_j^* = (\beta_{j0}, \beta_j')'$, where $\beta_j = (\beta_{j1}, \ldots, \beta_{jl}, \ldots, \beta_{jL-1})'$. Next let X_l denote the matrix of dummy covariates for the lth level, which is defined as follows:

$$\begin{aligned} X_l &= \begin{pmatrix} x'_{[l]1} \\ \vdots \\ x'_{[l]j} \\ \vdots \\ x'_{[l](J-1)} \\ x'_{[l]J} \end{pmatrix} \\ &= \begin{pmatrix} 1 & \mathbf{0}'_{l-1} & 1 & \mathbf{0}'_{L-l-1} & \ldots & 0 & & \mathbf{0}'_{L-1} & & \ldots & 0 & & \mathbf{0}'_{L-1} & \\ & & \vdots & & & & & \vdots & & & & & \vdots & \\ 0 & & \mathbf{0}'_{L-1} & & \ldots & 1 & \mathbf{0}'_{l-1} & 1 & \mathbf{0}'_{L-l-1} & \ldots & 0 & & \mathbf{0}'_{L-1} & \\ & & \vdots & & & & & \vdots & & & & & \vdots & \\ 0 & & \mathbf{0}'_{L-1} & & \ldots & 0 & & \mathbf{0}'_{L-1} & & \ldots & 1 & \mathbf{0}'_{l-1} & 1 & \mathbf{0}'_{L-l-1} \\ 0 & & \mathbf{0}'_{L-1} & & \ldots & 0 & & \mathbf{0}'_{L-1} & & \ldots & 0 & & \mathbf{0}'_{L-1} & \end{pmatrix} \end{aligned} \tag{1.11}$$

for $l = 1, \ldots, L-1$, and

$$X_L = \begin{pmatrix} x'_{[L]1} \\ \vdots \\ x'_{[L]j} \\ \vdots \\ x'_{[L](J-1)} \\ x'_{[L]J} \end{pmatrix} = \begin{pmatrix} 1 & \mathbf{0}'_{L-1} & \ldots & 0 & \mathbf{0}'_{L-1} & \ldots & 0 & \mathbf{0}'_{L-1} \\ & \vdots & & & \vdots & & & \vdots \\ 0 & \mathbf{0}'_{L-1} & \ldots & 1 & \mathbf{0}'_{L-1} & \ldots & 0 & \mathbf{0}'_{L-1} \\ & \vdots & & & \vdots & & & \vdots \\ 0 & \mathbf{0}'_{L-1} & \ldots & 0 & \mathbf{0}'_{L-1} & \ldots & 1 & \mathbf{0}'_{L-1} \\ 0 & \mathbf{0}'_{L-1} & \ldots & 0 & \mathbf{0}'_{L-1} & \ldots & 0 & \mathbf{0}'_{L-1} \end{pmatrix}. \tag{1.12}$$

By using the jth row $(x'_{[l]j})$ of the $J \times (J-1)L$ matrix X_l, we rewrite the probabilities in (1.8) with exponents in linear regression form as

$$\pi_{y(l)}^{(j)} = \frac{\exp(x'_{[l]j}\theta^*)}{\sum_{u=1}^{J} \exp(x'_{[l]u}\theta^*)}, \; j = 1, \ldots, J. \tag{1.13}$$

Now putting (1.13) in (1.10), one may obtain the likelihood estimating equation for

θ^* as

$$
\begin{aligned}
f(\theta^*) &= \frac{\partial l(\theta^*)}{\partial \theta^*} = \frac{\partial logL(\theta^*)}{\partial \theta^*} \\
&= \frac{\partial}{\partial \theta^*} \sum_{l=1}^{L} \sum_{j=1}^{J} n_{[l]j} log \pi_{y(l)}^{(j)} \\
&= \sum_{l=1}^{L} \sum_{j=1}^{J} n_{[l]j} \left(x_{[l]j} - \frac{\sum_{j=1}^{J} \exp(x'_{[l]j}\theta^*) x_{[l]j}}{\sum_{j=1}^{J} \exp(x'_{[l]j}\theta^*)} \right) \\
&= \sum_{l=1}^{L} \sum_{j=1}^{J} n_{[l]j} \left(x_{[l]j} - \sum_{j=1}^{J} \pi_{y(l)}^{(j)} x_{[l]j} \right) \\
&= \sum_{l=1}^{L} \left[\sum_{j=1}^{J} n_{[l]j} x_{[l]j} - \sum_{j=1}^{J} n_{[l]j} \sum_{j=1}^{J} \pi_{y(l)}^{(j)} x_{[l]j} \right] \\
&= \sum_{l=1}^{L} \left[\sum_{j=1}^{J} n_{[l]j} x_{[l]j} - \tilde{n}_l \sum_{j=1}^{J} \pi_{y(l)}^{(j)} x_{[l]j} \right] \\
&= \sum_{l=1}^{L} \left[\sum_{j=1}^{J} x_{[l]j} \left(n_{[l]j} - \tilde{n}_l \pi_{y(l)}^{(j)} \right) \right] \\
&= \sum_{l=1}^{L} X'_l \left[\left(n_{[l]1}, \ldots, n_{[l]j}, \ldots, n_{[l]J} \right)' - \tilde{n}_l \left(\pi_{y(l)}^{(1)}, \ldots, \pi_{y(l)}^{(j)}, \ldots, \pi_{y(l)}^{(J)} \right)' \right] \\
&= \sum_{l=1}^{L} X'_l \left[\mathbf{n}_l - \tilde{n}_l \pi_{y(l)} \right] = 0, \qquad (1.14)
\end{aligned}
$$

where

$$\mathbf{n}_l = \left[n_{[l]1}, \ldots, n_{[l]j}, \ldots, n_{[l](J-1)}, n_{[l]J} \right]'_{J \times 1},$$

and

$$\pi_{y(l)} = \left[\pi_{y(l)}^{(1)}, \ldots, \pi_{y(l)}^{(j)}, \ldots, \pi_{y(l)}^{(J-1)}, \pi_{y(l)}^{(J)} \right]'_{J \times 1}.$$

Notice that this likelihood equation in (1.14) has a simple form which is easy to solve for θ^*. Also note from (1.8) that the exponent in the probability functions does not use any linear addition of the regression parameters and hence there does not arise any question of confounding one parameter with another. Thus, all parameters unlike in (1.15) below do not encounter any identification problems. Furthermore even if β_{j0}

is added to many other parameters, this β_{j0} parameter is common at all probability levels making it different than other parameters.

Note that the equation (1.14) is similar to the likelihood equation (8.22) in Agresti (2002) developed for the log linear models. To be specific, θ^* from (1.14) may be obtained iteratively by using

$$\hat{\theta}^*_{k+1} = \hat{\theta}^*_k - \left[\frac{\partial^2 l(\theta^*)}{\partial\theta^* \partial\theta^{*\prime}}\right]^{-1}_k \left[\frac{\partial l(\theta^*)}{\partial\theta^*}\right]_k,$$

where

$$\frac{\partial^2 l(\theta^*)}{\partial\theta^* \partial\theta^{*\prime}} = \frac{\partial}{\partial\theta^*}\left[\sum_{l=1}^{L} X_l' \left(\mathbf{n}_l - \tilde{n}_l \pi_{y(l)}\right)\right],$$

which can be reexpressed as

$$\frac{\partial^2 l(\theta^*)}{\partial\theta^* \partial\theta^{*\prime}} = \sum_{l=1}^{L} \tilde{n}_l X_l' \left[Diag(\pi_{y(l)}) - \pi_{y(l)}\pi'_{y(l)}\right] X_l,$$

where $Diag(\pi_{y(l)}) = diag[\pi^{(1)}_{y(l)}, \ldots, \pi^{(j)}_{y(l)}, \ldots, \pi^{(J-1)}_{y(l)}, \pi^{(J)}_{y(l)}]_{J\times J}$. The variance of $\hat{\theta}^*$ is given by

$$Var(\hat{\theta}^*) = \left[\sum_{l=1}^{L} \tilde{n}_l X_l' \left[Diag(\pi_{y(l)}) - \pi_{y(l)}\pi'_{y(l)}\right] X_l\right]^{-1}.$$

Note that some of the existing studies model the relationship between y and x ignoring the fixed covariate nature of x, i.e., by treating x also as a response variable. See, for example, the modeling for the heart attack and aspirin use data discussed by Agresti (2002, Section 2.1.1, Table 2.1). In this approach, by considering the treatment (aspirin use/placebo) as a response variable, for example, the joint probability for the bivariate response is written as

$$\pi^*_{lj} = \frac{\exp(\alpha_l + \beta_j + \lambda_{lj})}{\sum_{l=1}^{L}\sum_{j=1}^{J}\exp(\alpha_l + \beta_j + \lambda_{lj})},\ l = 1, \ldots, L,\ j = 1, \ldots, J. \tag{1.15}$$

To be specific, π^*_{lj} represents the probability for y_i to be in the jth category with x_i at the lth level. Here β_j is the jth category effect as β_{j0} defined in (1.7); and α_l

determines the effect of the lth level of the x variable. Furthermore, λ_{lj} ($l = 1, \ldots, L$, $j = 1, \ldots, J$) denotes the interaction effect between y and x variables. Note that the parameters involved in (1.15) are supposed to satisfy the restrictions: $\sum_{l=1}^{L} \alpha_l = 0$, $\sum_{j=1}^{J} \beta_j = 0$, and $\sum_{l=1}^{L} \lambda_{lj} = \sum_{j=1}^{J} \lambda_{lj} = 0$. However, this type of modeling encounters several confusions. This is because λ_{lj} in (1.15) represents the effect of $x_{i(l)}$ for y to be in the jth category. This is well understood from the probability model for $\pi_{iy}^{(j)}$ given in (1.7). Notice that (1.7) contains β_{j0} which is the same as β_j in (1.15), and β_{jl}, which is the same as λ_{lj} in (1.15) when x is a fixed covariate. In such cases when x is fixed covariate, α_l is redundant. Thus, this approach of treating a fixed covariate as a categorical response variable is inappropriate. This may further be explained through interpreting λ_{lj}. For example, λ_{lj} in (1.15) is treated to be an odds ratio parameter which is a function of some correlations between two random variables. But, when x is a fixed covariate, the correlation interpretation is quite inappropriate. To make it even clearer, notice that the modeling by (1.7) cannot incorporate any correlation or odds ratio parameters, rather it is a regression way of modeling.

1.1.3.2 Remarks on an alternative loglinear model

Note that to analyze the data shown in the contingency table 1.1, it is standard to use multinomial distribution as indicated in (1.10). However, there exists a basic alternative loglinear model (see, for example, Agresti (2009), Chapter 7 and Christensen (1997)) where poisson distributions are fitted. To be specific, in such a setup, it is assumed that

$$n_{[l]j} \sim Poisson(\mu_{[l]j} = \exp(\beta^* + \beta_j^Y + \beta_l^X)) \tag{1.16}$$

with category dependent restrictions $\sum_{j=1}^{J} \beta_j^Y = \sum_{l=1}^{L} \beta_l^X = 0$. It is further assumed that these cell counts are independent. It then follows that $n = \sum_{l=1}^{L} \sum_{j=1}^{J} n_{[l]j}$ has the poisson distribution with parameter $\mu = \sum_{l=1}^{L} \sum_{j=1}^{J} \mu_{[l]j} = \sum_{l=1}^{L} \sum_{j=1}^{J} \exp(\beta^* + \beta_j^Y + \beta_l^X)$. Thus, realizing that n is a random variable, an independent poisson likelihood, i.e.,

$$\prod_{l=1}^{L} \prod_{j=1}^{J} \frac{e^{-\mu_{[l]j}} \mu_{[l]j}^{n_{[l]j}}}{n_{[l]j}!} \tag{1.17}$$

is fitted to estimate the parameters β^*, β_j^Y $(j = 1, \ldots, J-1)$ and β_l^X $(l = 1, \ldots, L-1)$.

However, there are at least two reasons why multinomial distribution is preferred to the poisson distribution for analyzing such contingency table based data. First, in many studies in practice, n is prespecified, and then these n individuals are distributed in JK cells according to their individual responses. For this reason, conditional inference, where n is a specified value from a poisson distribution becomes more appropriate. Consequently, the cell counts with given n follow the multinomial distribution. Notice however that, because we have written Table 1.1 for X covariate with L levels, we have given a product multinomial instead of a full multinomial likelihood function.

Second, when multinomial likelihood (1.9) is used instead of the poisson likelihood (1.17), the multinomial model does not contain β^* any more, which is sensible in categorical data analysis. This is because β^* does not contribute any additional information when categories are compared, but the poisson approach requires this parameter to be estimated.

Moreover, when n is given, the category dependence, that is, the structural correlations of the responses are understood well from the multinomial setup, as opposed to the poisson setup.

For the above reasons, we will deal with the multinomial model only in this thesis.

1.1.3.3 More than one categorical covariate having interactions

Suppose that there are more than one categorical covariate associated with the multinomial response y_i. To be specific, for simplicity, we consider two categorical covariates x_{i1} and x_{i2} with L and Q levels, respectively. Note that this can be generalized to accommodate any number of categorical covariates. Because L levels of a covariate can be represented by $L-1$ dummy covariates, we denote the dummy covariates for x_{i1} as $x_{i1(1)}, \ldots, x_{i1(l)}, \ldots, x_{i1(L-1)}$, and the $Q-1$ dummy covariates for x_{i2} as $x_{i2(1)}, \ldots, x_{i2(q)}, \ldots, x_{i2(Q-1)}$. Remark that in some situations, the two such covariates may have interactions. For generality we use the interaction factor in modeling the probabilities for the multinomial response y_i. Similar to the one categorical covariate case (1.6), we write the values for the two covariates as follows.

$$(x_{i1(1)}, \ldots, x_{i1(l)}, \ldots, x_{i1(L-1)}) = \begin{cases} (1, 0, \ldots, 0), & \text{level } 1 \\ \vdots & \\ (\mathbf{0}'_{l-1}, 1, \mathbf{0}'_{L-l-1}), & \text{level } l \\ \vdots & \\ (0, \ldots, 0, 1), & \text{level } L-1 \\ (0, \ldots, 0), & \text{level } L, \end{cases} \tag{1.18}$$

and

$$
(x_{i2(1)}, \ldots, x_{i2(q)}, \ldots, x_{i2(Q-1)}) = \begin{cases} (1, 0, \ldots, 0), & \text{level } 1 \\ \vdots & \\ (\mathbf{0'}_{q-1}, 1, \mathbf{0'}_{Q-q-1}), & \text{level } q \\ \vdots & \\ (0, \ldots, 0, 1), & \text{level } Q-1 \\ (0, \ldots, 0), & \text{level } Q. \end{cases} \tag{1.19}
$$

Suppose that β_{jl} denotes the effect of $x_{i1(l)}$ on y_{ij}, $\tilde{\beta}_{jq}$ as the effect of $x_{i2(q)}$ on y_{ij}, and we further denote λ_{jlq} as the effect of the interaction between $x_{i1(l)}$ and $x_{i2(q)}$ on y_{ij}. By treating the Lth level of x_{i1} and the Qth level of x_{i2} as the reference level, one requires the following restrictions on the values of β_{jl}, $\tilde{\beta}_{jq}$ and λ_{jlq}:

$$
\begin{aligned}
&\beta_{Jl} = 0,\ l = 1, \ldots, L,\ \beta_{jL} = 0,\ j = 1, \ldots, J, \\
&\tilde{\beta}_{Jq} = 0,\ q = 1, \ldots, Q,\ \tilde{\beta}_{jQ} = 0,\ j = 1, \ldots, J, \\
&\lambda_{Jlq} = 0,\ l = 1, \ldots, L,\ q = 1, \ldots, Q, \\
&\lambda_{jLq} = 0,\ j = 1, \ldots, J,\ q = 1, \ldots, Q, \\
&\lambda_{jlQ} = 0,\ j = 1, \ldots, J,\ l = 1, \ldots, L.
\end{aligned}
$$

Now as an extension of the one categorical covariate case (1.7), we accommodate two covariates along with their interactions and write the marginal probabilities as

$$
\begin{aligned}
\pi_{iy}^{(j)} &= Pr(y_i = y_i^{(j)}) \\
&= \begin{cases} \frac{\exp(\beta_{j0} + \sum_{l=1}^{L-1} x_{i1(l)}\beta_{jl} + \sum_{q=1}^{Q-1} x_{i2(q)}\tilde{\beta}_{jq} + \sum_{l=1}^{L-1}\sum_{q=1}^{Q-1} x_{i1(l)}x_{i2(q)}\lambda_{jlq})}{1 + \sum_{u=1}^{J-1}\exp(\beta_{u0} + \sum_{l=1}^{L-1} x_{i1(l)}\beta_{ul} + \sum_{q=1}^{Q-1} x_{i2(q)}\tilde{\beta}_{uq} + \sum_{l=1}^{L-1}\sum_{q=1}^{Q-1} x_{i1(l)}x_{i2(q)}\lambda_{ulq})}, & j = 1, \ldots, J-1, \\ \frac{1}{1 + \sum_{u=1}^{J-1}\exp(\beta_{u0} + \sum_{l=1}^{L-1} x_{i1(l)}\beta_{ul} + \sum_{q=1}^{Q-1} x_{i2(q)}\tilde{\beta}_{uq} + \sum_{l=1}^{L-1}\sum_{q=1}^{Q-1} x_{i1(l)}x_{i2(q)}\lambda_{ulq})}, & j = J. \end{cases}
\end{aligned} \tag{1.20}
$$

Note that these probability models are available in many standard textbooks. See, Agresti (2002, Section 8.2.1, eqn. (8.8)), for example, for similar modeling for Poisson count data.

Next, based on the covariates levels, we group the individuals into LQ levels, the observed data is given in Table 2 below.

Let

$$\tilde{\theta}^* = (\tilde{\beta}_1^{*\prime}, \ldots, \tilde{\beta}_j^{*\prime}, \ldots, \tilde{\beta}_{J-1}^{*\prime})'_{(J-1)LQ\times 1}$$

denote the vector of parameters involved in model (1.20), with

$$\tilde{\beta}_j^* = (\beta_{j0}, \beta_j', \tilde{\beta}_j', \lambda_j')'_{LQ\times 1},$$

where

$$\begin{aligned} \beta_j &= (\beta_{j1}, \ldots, \beta_{jl}, \ldots, \beta_{jL-1})', \\ \tilde{\beta}_j &= (\tilde{\beta}_{j1}, \ldots, \tilde{\beta}_{jq}, \ldots, \tilde{\beta}_{jQ-1})', \\ \text{and } \lambda_j &= (\lambda_{j11}, \ldots, \lambda_{jlq}, \ldots, \lambda_{jL-1,Q-1})'. \end{aligned}$$

Similar to the single categorical covariate case discussed in Section 1.1.3.1, by expressing the exponents in a regression form involving all parameters in $\tilde{\theta}^*$, we write the marginal probabilities in (1.20) as

$$\pi_{y(\nu)}^{(j)} = \frac{\exp(x'_{[\nu]j}\tilde{\theta}^*)}{\sum_{u=1}^{J}\exp(x'_{[\nu]u}\tilde{\theta}^*)},\ j = 1, \ldots, J, \tag{1.21}$$

where for $\nu = 1, \ldots, LQ$, $x'_{[\nu]j}$ is the jth row of the $J \times (J-1)LQ$ matrix X_ν, which

Table 1.2: Contingency table of multinomial response with two categorical covariates.

x_{i1}	x_{i2}	Level ν	1	...	j	...	J	Total
1	1	1	$n_{[1]1}$	...	$n_{[1]j}$	...	$n_{[1]J}$	$\tilde{n}_1$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	q	q	$n_{[q]1}$	...	$n_{[q]j}$	...	$n_{[q]J}$	$\tilde{n}_q$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	Q	Q	$n_{[Q]1}$	...	$n_{[Q]j}$	...	$n_{[Q]J}$	$\tilde{n}_Q$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
l	1	$(l-1)Q+1$	$n_{[(l-1)Q+1]1}$	...	$n_{[(l-1)Q+1]j}$	...	$n_{[(l-1)Q+1]J}$	$\tilde{n}_{(l-1)Q+1}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	q	$(l-1)Q+q$	$n_{[(l-1)Q+q]1}$	...	$n_{[(l-1)Q+q]j}$	...	$n_{[(l-1)Q+q]J}$	$\tilde{n}_{(l-1)Q+q}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	Q	lQ	$n_{[lQ]1}$	...	$n_{[lQ]j}$	...	$n_{[lQ]J}$	$\tilde{n}_{lQ}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
L	1	(L-1)Q+1	$n_{[(L-1)Q+1]1}$	...	$n_{[(L-1)Q+1]j}$	...	$n_{[(L-1)Q+1]J}$	$\tilde{n}_{(L-1)Q+1}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	q	(L-1)Q+q	$n_{[(L-1)Q+q]1}$	...	$n_{[(L-1)Q+q]j}$	...	$n_{[(L-1)Q+q]J}$	$\tilde{n}_{(L-1)Q+q}$
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$	$\vdots$
	Q	LQ	$n_{[LQ]1}$	...	$n_{[LQ]j}$	...	$n_{[LQ]J}$	$\tilde{n}_{LQ}$
Total			n_1	...	n_j	...	n_J	n

has the form

$$X_\nu = \begin{pmatrix} x'_{[\nu]1} \\ \vdots \\ x'_{[\nu]j} \\ \vdots \\ x'_{[\nu](J-1)} \\ x'_{[\nu]J} \end{pmatrix}.$$

Following the one categorical covariate case (1.14), we may write the likelihood equation for $\tilde{\theta}^*$ as

$$f(\tilde{\theta}^*) = \sum_{\nu=1}^{LQ} X'_\nu \left[\mathbf{n}_\nu - \tilde{n}_\nu \pi_{y(\nu)}\right] = 0, \tag{1.22}$$

which can be solved iteratively for the estimation of $\tilde{\theta}^*$.

1.2 Existing bivariate multinomial models

In this section, we provide a brief review of the existing modeling and inference for bivariate multinomial data. This type of bivariate multinomial data exhibit two types of structural correlations. First, the marginal multinomial responses for one response variable are correlated. Second, the multinomial responses of one variable is correlated to the multinomial responses for the other variable. This correlation is referred to as the familial correlation which is caused by a common individual random effect shared by both response variables. Thus, for two multinomial responses with dimensions J and K, respectively, there is a $(J-1)\times(K-1)$ structural correlation matrix for a given individual. It is important to take these correlations into account to obtain consistent and as efficient as possible estimates for the effects of the covariates. For the purpose, in this section we indicate how some of the existing inference approaches are developed and also indicate their drawbacks.

Let y_i and z_i denote the two multinomial response variables with $J \geq 2$ and $K \geq 2$. We denote them as $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{i,J-1})'$ and $z_i = (z_{i1}, \ldots, z_{ik}, \ldots, z_{i,K-1})'$. Recall from Section 1.1 that we represented the jth category response of y_i by $y_i^{(j)}$, similarly, we represent the kth category response of z_i by $z_i^{(k)}$. Thus

$$z_i^{(k)} = \begin{cases} (z_{i1}^{(k)}, \ldots, z_{ik}^{(k)}, \ldots, z_{i,K-1}^{(k)})' = (\mathbf{0}'_{k-1}, 1, \mathbf{0}'_{K-1-k})', \; k = 1, \ldots, K-1, \\ (0, \ldots, 0, \ldots, 0)', \; k = K. \end{cases} \tag{1.23}$$

In this setup, one is interested in understanding the association between the two multinomial variables. The marginal effect of each variable is also of interest. This requires one to model the joint probabilities for understanding the associations. Note that the modeling for the joint probabilities is, however, not so straightforward. Many

Table 1.3: Marginal and joint probabilities for bivariate multinomial responses Y and Z.

$Z_i \setminus Y_i$	1	$\ldots$	j	$\ldots$	J
1	π_{i11}	$\ldots$	π_{i1j}	$\ldots$	π_{i1J}
$\vdots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
k	π_{ik1}	$\ldots$	π_{ikj}	$\ldots$	π_{ikJ}
$\vdots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
K	π_{iK1}	$\ldots$	π_{iKj}	$\ldots$	π_{iKJ}

existing studies have modeled these joint probabilities directly. To be specific, when the data are available in a contingency table form, the joint probabilities are modeled using functions similar to (1.15). To make it more clear how these joint probabilities are modeled for an individual, we, for convenience, display the joint probabilities for the response of an individual to be in a particular cell out of all KJ cells in Table 1.3.

Note that when all n individuals in a study are categorized based on the two responses only, one may write $\pi_{ikj} \equiv \pi_{kj}$ for all i. This gives the contingency (or cross-classified) Table 1.4 containing observed cell counts along with their joint probabilities.

As pointed out earlier, the probabilities shown in Table 4 can be modeled in the fashion similar to (1.15) without considering any covariates. Thus, in the existing modeling approach (see for example, Agresti (2002), Eqn. (8.4), Fienberg (2007))

Table 1.4: Bivariate multinomial observed data and underlying marginal and joint probabilities with no covariate.

$Z \setminus Y$	1	...	j	...	J
1	n_{11}, π_{11}	...	n_{1j}, π_{1j}	...	n_{1J}, π_{1J}
$\vdots$	...	...	...	...	...
k	n_{k1}, π_{k1}	...	n_{kj}, π_{kj}	...	n_{kJ}, π_{kJ}
$\vdots$	...	...	...	...	...
K	n_{K1}, π_{K1}	...	n_{Kj}, π_{Kj}	...	n_{KJ}, π_{KJ}

one writes

$$\begin{aligned} \pi_{kj} &= Pr(z_i = z_i^{(k)}, y_i = y_i^{(j)}) \\ &= \frac{\exp(\alpha_k + \beta_j + \lambda_{kj})}{\sum_{k=1}^{K}\sum_{j=1}^{J}\exp(\alpha_k + \beta_j + \lambda_{kj})},\ k = 1, \ldots, K,\ j = 1, \ldots, J, \end{aligned} \quad (1.24)$$

along with the restrictions: $\sum_{k=1}^{K}\alpha_k = 0$, $\sum_{j=1}^{J}\beta_j = 0$, and $\sum_{k=1}^{K}\lambda_{kj} = \sum_{j=1}^{J}\lambda_{kj} = 0$, or equivalently $\alpha_K = -\sum_{k=1}^{K-1}\alpha_k$, $\beta_J = -\sum_{j=1}^{J-1}\beta_j$ and $\lambda_{KJ} = -\sum_{j=1}^{J-1}\lambda_{Kj} = -\sum_{k=1}^{K-1}\lambda_{kJ} = \sum_{k=1}^{K-1}\sum_{j=1}^{J-1}\lambda_{kj}$.

Now let $\theta = (\alpha_1, \ldots, \alpha_{K-1}, \beta_1, \ldots, \beta_{J-1}, \lambda_{11}, \lambda_{12}, \ldots, \lambda_{K-1,J-1})'_{(KJ-1)\times 1}$ denote the vector of parameters involved in the joint probability (1.24), one may then estimate θ by solving appropriate likelihood equations derived from the likelihood function given by

$$L(\theta) = \frac{n!}{\prod_{k=1}^{K}\prod_{j=1}^{J} n_{kj}!}\prod_{k=1}^{K}\prod_{j=1}^{J}\pi_{kj}^{n_{kj}}. \quad (1.25)$$

Next, for simplicity of writing the likelihood estimating equation, we express the exponents in (1.24) in a regression form involving all parameters in θ as

$$\pi_{kj} = \frac{\exp(\omega'_{kj}\theta)}{\sum_{k=1}^{K}\sum_{j=1}^{J}\exp(\omega'_{kj}\theta)}, \tag{1.26}$$

where ω_{kj} is the $(KJ-1)$-dimensional dummy covariate vector corresponding to the (k,j)th cell of Table 4, yielding the $KJ \times (KJ-1)$ dummy covariate matrix W, which is defined as

$$W = \begin{pmatrix} \omega'_{11} \\ \omega'_{12} \\ \vdots \\ \omega'_{1J} \\ \omega'_{21} \\ \omega'_{kj} \\ \vdots \\ \omega'_{KJ} \end{pmatrix},$$

with

$$\begin{aligned}
\omega'_{kj} &= \left[\mathbf{0}'_{(k-1)\times 1}, 1, \mathbf{0}'_{(K-k-1)\times 1}, \mathbf{0}'_{(j-1)\times 1}, 1, \mathbf{0}'_{(J-j-1)\times 1}, \mathbf{0}'_{[(k-1)(J-1)+j-1]\times 1}, 1, \mathbf{0}'_{[(K-k)(J-1)-j]\times 1}\right], \\
& \quad k = 1, \ldots, K-1, \; j = 1, \ldots, J-1, \\
\omega'_{kJ} &= \left[\mathbf{0}'_{(k-1)\times 1}, 1, \mathbf{0}'_{(K-k-1)\times 1}, -\mathbf{1}'_{(J-1)\times 1}, \mathbf{0}'_{(k-1)(J-1)\times 1}, -\mathbf{1}'_{(J-1)\times 1}, \mathbf{0}'_{[(K-k-1)(J-1)]\times 1}\right], \\
& \quad k = 1, \ldots, K-1, \\
\omega'_{Kj} &= \left[-\mathbf{1}'_{(K-1)\times 1}, \mathbf{0}'_{(j-1)\times 1}, 1, \mathbf{0}'_{(J-j-1)\times 1}, \mathbf{0}'_{(j-1)\times 1}, -1, \mathbf{0}'_{(J-j-1)\times 1}, \ldots, \mathbf{0}'_{(j-1)\times 1}, -1, \mathbf{0}'_{(J-j-1)\times 1}\right], \\
& \quad j = 1, \ldots, J-1, \\
\omega'_{KJ} &= \left[-\mathbf{1}'_{(K-1)\times 1}, -\mathbf{1}'_{(J-1)\times 1}, \mathbf{1}'_{(K-1)(J-1)\times 1}\right].
\end{aligned}$$

Let $Y = (n_{11}, \ldots, n_{1J}, n_{21}, \ldots, n_{kj}, \ldots, n_{KJ})'$ denote the $KJ \times 1$ vector of counts for all KJ cells and $\Pi = (\pi_{11}, \ldots, \pi_{1J}, \pi_{21}, \ldots, \pi_{kj}, \ldots, \pi_{KJ})'$ be the vector of corresponding cell probabilities. Then, by using (1.25) and (1.26), the log likelihood estimating equation for θ can be written as

$$f(\theta) = W'(Y - n\Pi) = 0, \tag{1.27}$$

where $n = \sum_{k=1}^{K}\sum_{j=1}^{J} n_{kj}$.

We point out the following advantages and drawbacks of this modeling approach below.

Advantages:

The advantage of modeling joint probabilities by (1.24) is that the estimation of the parameters by solving (1.25) is relatively straightforward. To be specific, $\hat{\theta}$ for θ can be obtained by using the simple iterative equation

$$\hat{\theta}_{k+1} = \hat{\theta}_k - \left[f'(\theta)\right]^{-1}_{\theta_k=\hat{\theta}_k}\left[f(\theta)\right]_{\theta_k=\hat{\theta}_k},$$

where $f'(\theta) = nW'\left[Diag(\Pi) - \Pi\Pi'\right]W$ with $Diag(\Pi) = diag[\pi_{11}, \ldots, \pi_{kj}, \ldots, \pi_{KJ}]_{KJ\times KJ}$.

Drawbacks:

(1) The joint probabilities (1.24) yield complicated marginal probabilities given by

$$Pr(z_i = z_i^{(k)}) = \sum_{j=1}^{J}\pi_{kj} = \frac{\sum_{j=1}^{J}\exp(\alpha_k + \beta_j + \lambda_{kj})}{\sum_{k=1}^{K}\sum_{j=1}^{J}\exp(\alpha_k + \beta_j + \lambda_{kj})}, \tag{1.28}$$

and similarly for $Pr(y_i = y_i^{(j)})$. For simplicity, consider the bivariate binary case with $J = 2$ and $K = 2$. Using the restrictions $\alpha_1 + \alpha_2 = 0$, $\beta_1 + \beta_2 = 0$, and $\lambda_{11} + \lambda_{12} = 0$, $\lambda_{21} + \lambda_{22} = 0$, $\lambda_{11} + \lambda_{21} = 0$, and $\lambda_{12} + \lambda_{22} = 0$, the marginal probability for $z_i = z_i^{(1)}$, has the formula

$$\begin{aligned} & Pr(z_i = z_i^{(1)}) = \pi_{11} + \pi_{12} \\ = & \frac{\exp(\alpha_1 + \beta_1 + \lambda_{11}) + \exp(\alpha_1 - \beta_1 - \lambda_{11})}{\exp(\alpha_1 + \beta_1 + \lambda_{11}) + \exp(\alpha_1 - \beta_1 - \lambda_{11}) + \exp(-\alpha_1 + \beta_1 - \lambda_{11}) + \exp(-\alpha_1 - \beta_1 + \lambda_{11})}. \end{aligned}$$

It is clear that this marginal probability is a complicated function of all marginal and association parameters, namely, α_1, β_1 and λ_{11}. Similarly, $Pr(y_i = y_i^{(j)})$ is also a complicated function of all marginal and association parameters. Thus, there is no clear cut marginal parameters to define the marginal probabilities. Furthermore, the use of association parameters to explain marginal probabilities appears to be counter intuitive. This difficulty to explain marginal probabilities through all parameters arise because of modeling the joint probabilities first.

(2) The association parameter λ_{kj} in (1.24) are also referred to as odds ratio parameters. For example, for the above bivariate binary case, λ_{11} satisfies the formula

$$\lambda_{11} = \frac{1}{4} log \frac{\pi_{11}\pi_{22}}{\pi_{12}\pi_{21}},$$

which is proportional to the log of odds ratio. However, when this type of odds ratio change from individual to individual because of individual level covariate effects, the analysis becomes difficult and many existing studies (see for example, Williamson, Kim and Lipsitz (1995)) attempted to model such variable odds ratios through some 'working' linear models, which is however arbitrary. This 'working' modeling approach is described further in the following section.

1.2.1 Existing bivariate multinomial models involving individual level covariates

Suppose now that there are individual level covariates associated with the two multinomial responses. Let x_{i1} and x_{i2} denote the covariate vector associated with z_i and y_i, respectively. Note that x_{i1} and x_{i2} may contain certain common and fixed covariates. It is of scientific interest to understand the effect of x_{i1} on z_i and the effect of

x_{i2} on y_i as well. Let α_k $(k = 1, \ldots, K-1)$ and β_j $(j = 1, \ldots, J-1)$ represent the intercept parameters reflecting the categories, and let θ_1 and θ_2 denote the effects of x_{i1} on z_i and x_{i2} on y_i, respectively. Some authors, for example, Williamson, Kim and Lipsitz (1995), write the marginal probabilities for the multinomial variable z_i as

$$\begin{aligned} \pi_{iz}^{(k)} &= Pr(z_i = z_i^{(k)}) = \frac{\exp(\alpha_k + x_{i1}'\theta_{k1})}{1 + \sum_{q=1}^{K-1} \exp(\alpha_q + x_{i1}'\theta_{q1})}, \quad \text{for } k = 1, \ldots, K-1, \\ \text{and } \pi_{iz}^{(K)} &= Pr(z_i = z_i^{(K)}) = 1 - \sum_{k=1}^{K-1} \pi_{iz}^{(k)} = \frac{1}{1 + \sum_{q=1}^{K-1} \exp(\alpha_q + x_{i1}'\theta_{q1})}. \end{aligned} \quad (1.29)$$

Similarly, the marginal probabilities for the multinomial variable y_i are given by

$$\begin{aligned} \pi_{iy}^{(j)} &= Pr(y_i = y_i^{(j)}) = \frac{\exp(\beta_j + x_{i2}'\theta_{j2})}{1 + \sum_{l=1}^{J-1} \exp(\beta_l + x_{i2}'\theta_{l2})}, \quad \text{for } j = 1, \ldots, J-1, \\ \text{and } \pi_{iy}^{(J)} &= Pr(y_i = y_i^{(J)}) = 1 - \sum_{j=1}^{J-1} \pi_{iy}^{(j)} = \frac{1}{1 + \sum_{l=1}^{J-1} \exp(\beta_l + x_{i2}'\theta_{l2})}. \end{aligned} \quad (1.30)$$

It then follows that the marginal mean, variance and structural covariance of these two multinomial variables are given by:

$$E(Z_{ik}) = \pi_{iz}^{(k)}, \; Var(Z_{ik}) = \pi_{iz}^{(k)}(1 - \pi_{iz}^{(k)}), \; Cov(Z_{ik}, Z_{iq}) = -\pi_{iz}^{(k)}\pi_{iz}^{(q)}, k \neq q; \quad (1.31)$$

$$\text{and } E(Y_{ij}) = \pi_{iy}^{(j)}, \; Var(Y_{ij}) = \pi_{iy}^{(j)}(1 - \pi_{iy}^{(j)}), \; Cov(Y_{ij}, Y_{il}) = -\pi_{iy}^{(j)}\pi_{iy}^{(l)}, j \neq l.$$

On top of the marginal properties (1.29), it is necessary for the bivariate multinomial data analysis to model the joint probabilities for y_i and z_i, so that correlations between y_i and z_i can be accommodated for any inferences mainly for the parameters involved in the marginal probabilities. To address the above correlation issues, that is, (1) estimation of α_k and θ_{k1} using marginal information $z_i = z_i^{(k)}$ and similarly estimating β_j and β_{j2} by exploiting $y_i = y_i^{(j)}$ can not be the same thing as estimating these parameters by accommodating correlations between y_i and z_i. It is well known that in such cases the marginal estimates loose efficiency. (2) Further because in the bivariate binary or multinomial setup, it is also important to know the joint probabilities $\pi_{ikj} = Pr(y_i = y_i^{(j)}, z_i = z_i^{(k)})$. However, it should be clear that if correlations are ignored and joint probabilities π_{ikj} are computed using $\pi_{ikj} = \pi_{iz}^{(k)}\pi_{iy}^{(j)}$, then they will be biased estimates for actual probabilities. For these two reasons, it is important to model π_{ikj} as a function of suitable dependence between y_i and z_i.

There exist many studies, see for example, Williamson, Kim and Lipsitz (1995) for bivariate multinomial data analysis at a cross-sectional setup; and Lipsitz, Laird and Harrington (1991), Yi and Cook (2002), Ten Have and Morabia (1999) for correlated binary data in longitudinal setup. In these studies, the marginal probabilities are modeled in a fashion similar to the models (1.29) and (1.30), but, as indicated in the last section, the joint probabilities are defined through certain odds ratios approach. To be specific, the odds ratio in terms of the joint probability π_{ikj} corresponding to response $(z_i = z_i^{(k)}, y_i = y_i^{(j)})$ is defined as:

$$\tau_{ikj} = \frac{Pr(z_i = z_i^{(k)}, y_i = y_i^{(j)})Pr(z_i \neq z_i^{(k)}, y_i \neq y_i^{(j)})}{Pr(z_i = z_i^{(k)}, y_i \neq y_i^{(j)})Pr(z_i \neq z_i^{(k)}, y_i = y_i^{(j)})} = \frac{\pi_{ikj}(1-\pi_{ikj})}{(\pi_{iz}^{(k)} - \pi_{ikj})(\pi_{iy}^{(j)} - \pi_{ikj})},$$

for $k = 1, \ldots, K-1$ and $j = 1, \ldots, J-1$. Notice that if these odds ratio parameters are the same for all i (which is however in general not the case in practice), then $\tau_{ikj} \equiv \tau_{kj}$, which is related to λ_{kj} used in the last section where $\lambda_{kj} = \frac{1}{4} log\tau_{kj}$. Further notice that, the computation of the joint probabilities using the above odds ratio parameters naturally become complex. This complexity is clear from the relationship:

$$\pi_{ikj} = \begin{cases} \frac{f_{ikj} - [f_{ikj}^2 - 4\tau_{ikj}(\tau_{ikj}+1)\pi_{iy}^{(j)}\pi_{iz}^{(k)}]^{\frac{1}{2}}}{2(\tau_{ikj}+1)} & (\tau_{ikj} \neq 1), \\ \pi_{iz}^{(k)}\pi_{iy}^{(j)} & (\tau_{ikj} = 1), \end{cases} \quad (1.32)$$

where $f_{ikj} = 1 + \tau_{ikj}(\pi_{iz}^{(k)} + \pi_{iy}^{(j)})$ (see Lipsitz et al. (1991), Yi and Cook (2002), for example). Remark that for the purpose of computing the joint probabilities by (1.32), one needs to estimate the individual specific odds ratios τ_{ikj} for all individuals, which is however not possible without further modeling or assumptions. Thus, Williamson et al. (1995) (see also Williamson and Kim (1996)), for example, have used the linear regression model

$$log\tau_{ikj} = \Delta + \Delta_k + \tilde{\Delta}_j + \Delta_{kj} + \zeta x'_{ic}, \tag{1.33}$$

where Δ is an intercept parameter, Δ_k and $\tilde{\Delta}_j$ are the effects of z and y, respectively, Δ_{kj} is the interaction parameter and x'_{ic} is a suitable vector of covariates responsible to correlate y and z, and ζ is the effect of x'_{ic}. This type of regression model to explain association parameters lacks theoretical justification and hence appears to be arbitrary. More specifically, because odds ratios are equivalent to correlations between the multinomial responses, and because correlations are usually functions of the main covariate through the marginal probabilities, this extra model (1.33), however, does not address this issue at all.

1.2.2 Existing bivariate multinomial models with categorical covariates

In this section, we briefly review the modeling of bivariate multinomial responses with categorical covariates. For simplicity, suppose that we deal with a situation where the models (1.29)-(1.30) contain one covariate x_i instead of x_{i1} and x_{i2}. Also suppose that x_i is a categorical covariate with L levels. To represent these L levels, we use $L-1$ dummy covariates $x_{i1}, \ldots, x_{il}, \ldots, x_{i,L-1}$. As pointed out in Section 1.1.3.1 (Eqn. (1.15)), some authors treated the categorical covariate x also as a multinomial response, see, for example, Agresti (2002, Section 8.4.2, Table 8.8). Thus, treating

X as the third response variable, the joint probability for a response to be in the lth level of x, jth and kth categories of y and z, respectively, has been written as

$$\pi^*_{lkj} = \frac{\exp(\psi_l + \alpha_k + \beta_j + \lambda^{XZ}_{lk} + \lambda^{XY}_{lj} + \lambda^{ZY}_{kj} + \lambda^{XZY}_{lkj})}{\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{j=1}^{J} \exp(\psi_l + \alpha_k + \beta_j + \lambda^{XZ}_{lk} + \lambda^{XY}_{lj} + \lambda^{ZY}_{kj} + \lambda^{XZY}_{lkj})}, \tag{1.34}$$

[Agresti (2002), Eqn. (8.12), Fienberg (2007), Eqn. (3.11)], where ψ_l, α_k and β_j are the level/category effect of x, z and y to influence the response to be in the (l, k, j)th cell. In (1.32), λ^{XZ}_{lk}, λ^{XY}_{lj}, and λ^{ZY}_{kj} are second order interaction effects between x, z; x, y; and z, y, respectively. Also λ^{XZY}_{lkj} is the third order interaction effect among x, y and z. These parameters in (1.34) are supposed to satisfy the following restrictions:

$$\begin{aligned}
&\sum_{l=1}^{L}\psi_l = 0, \sum_{k=1}^{K}\alpha_k = 0, \sum_{j=1}^{J}\beta_j = 0,\\
&\sum_{l=1}^{L}\lambda^{XZ}_{lk} = \sum_{k=0}^{K}\lambda^{XZ}_{lk} = 0,\ \sum_{l=1}^{L}\lambda^{XY}_{lj} = \sum_{j=0}^{J}\lambda^{XY}_{lj} = 0,\\
&\sum_{k=1}^{K}\lambda^{ZY}_{kj} = \sum_{j=1}^{J}\lambda^{ZY}_{kj} = 0,\\
&\sum_{l=1}^{L}\lambda^{XZY}_{lkj} = \sum_{k=1}^{K}\lambda^{XZY}_{lkj} = \sum_{j=1}^{J}\lambda^{XZY}_{lkj} = 0.
\end{aligned}$$

This type of joint probability models constructed by treating a categorical covariate as a response variable suffers from several drawbacks. For example, as pointed out for (1.15), ψ_l parameters would be redundant when x is a fixed covariate. This means one has to use product multinomial modeling instead of full multinomial models. By the same token, because y and z are two response variables, the interaction effect λ^{ZY}_{kj} is quite meaningful to interpret the association between y and z, whereas λ^{XZ}_{lk}, for example, can not be used as an association parameter when x is a fixed covariate. Furthermore, even if x is a true categorical response variable, the joint probability in

(1.34) produces extremely complicated marginal probabilities. For example,

$$Pr(y_i = y_i^{(j)}) = \frac{\sum_{l=1}^{L}\sum_{k=1}^{K}\exp(\psi_l + \alpha_k + \beta_j + \lambda_{lk}^{XZ} + \lambda_{lj}^{XY} + \lambda_{kj}^{ZY} + \lambda_{lkj}^{XZY})}{\sum_{l=1}^{L}\sum_{k=1}^{K}\sum_{j=1}^{J}\exp(\psi_l + \alpha_k + \beta_j + \lambda_{lk}^{XZ} + \lambda_{lj}^{XY} + \lambda_{kj}^{ZY} + \lambda_{lkj}^{XZY})},$$

which involves all parameters to explain the marginal effect of y. Note that we will return to the proper modeling for this type of bivariate multinomial data in the presence of one or more categorical covariates in Chapter 2.

Remark that there are some studies on univariate longitudinal multinomial models. See, for example, Fienberg, Bromet, Follmann, Lambert and May (1985), Conaway (1989), Frees (2004), Fitzmaurice, Laird, and Ware (2004), Lipsitz et. al. (1991), Williamson et. al. (1995) and Chen, Yi, and Cook (2009). However, because univariate longitudinal data generates a multivariate distribution for the clustered data, some authors have used certain correlation structures from such a setup to model the bivariate binary and/or multinomial data. For example, the odds ratios used in longitudinal modeling have been exploited to model bivariate or multivariate categorical data including binary cases. For example, one may refer to Chen, Yi and Cook (2009), Williamson et. al. (1995) and Lipsitz et. al. (1991). But, as we discussed in Section 1.2, this approach encounters difficulties with the estimation of odds ratios based correlations. This is because odds ratios are not model parameters, rather, they are "working" parameters.

1.3 Plan of the thesis

In Section 1.1 we have reviewed the univariate multinomial model along with existing approaches for the estimation of the model parameters. Next, the existing extension of the univariate multinomial model to the bivariate case has been reviewed in Section 1.2. It has been demonstrated that the existing models fall short to address the correlations between two multinomial variables. For example, it was demonstrated that when interaction effects based joint probabilities are used to model the correlations, the resulting marginal probabilities remain complicated, more specifically, they involve the correlation or interaction parameters in a complex way. As discussed in Section 1.2, some authors use odds ratios to model the correlations and hence joint bivariate probabilities, but the estimation of the odds ratios is done arbitrarily using extra 'working' linear models.

For the above reason, i.e., because the existing models to deal with bivariate multinomial data are not adequately developed, in this thesis, we address this important modeling issue and develop bivariate multinmial correlation models using fully or partly specified marginal probabilities. Both fixed and mixed effect approaches are considered to model the correlations where marginal probabilities are kept fully specified. Conditional model is also considered where marginal probabilities are specified for one of the two response variables. The estimation of parameters for all these models is discussed in details both analytically and numerically. Several real life data analysis are also conducted. The specific plan of the thesis is as follows.

In Chapter 2, we first specify the marginal probabilities and develop a correlation model between two variables following the bivariate normality model. This model is referred to as the linear conditional bivariate multinomial (LCBM) fixed model. Both joint generalized quasi-likelihood (JGQL) and single stage GQL (SSGQL) estimation methods for this model are given in details. An extensive finite sample simulation study is conducted to examine the performance of these estimation approaches. A real life bivariate diabetic retinopathy data set is reanalyzed, first by using a simpler bivariate binary model, and then by using a trinomial categories based model for both response variables.

In Chapter 3, as opposed to the bivariate normal correlation structure used in Chpater 2, we propose a random effects based bivariate familial model. In this approach, the marginal probabilities of both multinomial variables are fully specified conditional on the random effects. Unconditionally the two multinomial variables become correlated. The joint GQL and likelihood inferences for associated regression and random effect variance parameters are given in details. Both simulation study and real life data analysis are given to illustrate the model and estimation empirically.

In Chapter 4, we first specify the marginal probabilities of one response variable and then develop a conditional multinomial logistic type probability model to accommodate correlations. As opposed to the linear conditional probability model discussed in Chapter 2, this type of partly specified non-linear models produce bivariate correlations satisfying full range. An outline of inferences is given using the likelihood approach.

The thesis concludes in Chapter 5.

Chapter 2

Linear Conditional Bivariate Multinomial (LCBM) Fixed Effects Model

In Chapter 1, we briefly reviewed the existing modeling for bivariate multinomial data, namely, the direct modeling approach of joint probabilities available in Agresti (2002) and Fienberg (2007), for example. However, as discussed there, this type of joint probabilities produce complex marginal probabilities. To be specific, in this existing approach, the joint probabilities are modeled without any prior or specified forms for the marginal probabilities. Thus this approach may be referred to as the fully unspecified marginal probability approach. Also, in a situation where joint probabilities should be individual covariates dependent, the use of constant (i.e. equal) interaction effects or odds ratio parameters in the joint probability formula, will not be appropriate. Note that as we discussed in Section 1.2.1 (see equations

(1.30)-(1.31)), many authors such as Williamson et al. (1995) attempted to tackle the latter problem by using individual covariates dependent odds ratios to define the joint probabilities. But, as indicated in Section 1.2, this odds ratio approach use an extra modeling for the estimation purpose, which is arbitrary.

In this chapter, as opposed to the fully unspecified marginal models, we discuss a fully specified marginal probability based linear conditional bivariate model to compute the joint probabilities and hence covariances and correlations. We explain the model along with its properties in Section 2.1. The likelihood and quasi-likelihood approaches are discussed for inferences in Section 2.2. In the same section, the performances of these inference techniques are examined through an intensive simulation study. Also, we illustrate the application of the model and inference methodologies by reanalyzing the so-called WESDR (Wisconsin Epidemiologic Study of Diabetic Retinopathy).

2.1 Fully Specified Marginal Probabilities Based LCBM Fixed Effects Model and Properties

Recall that the joint bivariate probabilities in (1.22) were modeled without prior forms for the marginal probabilities, whereas some authors (Williamson et al. (1995)) have used pre-specified marginal probabilities first to define an odds ratio in terms of the joint and the marginal probabilities, so that the joint probabilities can be computed as a function of odds ratios (see eqn. (1.30)). But as mentioned earlier this approach encounters problems in estimating the odds ratios. However, because the marginal probabilities are important to understand the variables separately, similar to Williamson et al. (1995), we prefer to use a pre-specified marginal probabilities based model, but unlike these authors, in this chapter, we use a standard normal regression type approach to model the conditional probabilities.

For the purpose, using the notations from Section 1.2, we assume that the categorical variables $Z_i : (K-1) \times 1$ and $Y_i : (J-1) \times 1$, marginally, have the multinomial distribution given by

$$\begin{aligned} \pi_{iz}^{(k)} &= Pr(z_i = z_i^{(k)}) = \frac{\exp(\alpha_k + x_{i1}'\theta_{k1})}{1 + \sum_{u=1}^{K-1} \exp(\alpha_u + x_{i1}'\theta_{u1})}, \quad \text{for } k = 1, \ldots, K-1, \\ \pi_{iz}^{(K)} &= Pr(z_i = z_i^{(K)}) = 1 - \sum_{k=1}^{K-1} \pi_{iz}^{(k)} = \frac{1}{1 + \sum_{u=1}^{K-1} \exp(\alpha_u + x_{i1}'\theta_{u1})}; \end{aligned} \tag{2.1}$$

and

$$\begin{aligned} \pi_{iy}^{(j)} &= Pr(y_i = y_i^{(j)}) = \frac{\exp(\beta_j + x_{i2}'\theta_{j2})}{1 + \sum_{l=1}^{J-1} \exp(\beta_l + x_{i2}'\theta_{l2})}, \quad \text{for } j = 1, \ldots, J-1, \\ \pi_{iy}^{(J)} &= Pr(y_i = y_i^{(J)}) = 1 - \sum_{j=1}^{J-1} \pi_{iy}^{(j)} = \frac{1}{1 + \sum_{l=1}^{J-1} \exp(\beta_l + x_{i2}'\theta_{l2})}, \end{aligned} \tag{2.2}$$

respectively. In (2.1) and (2.2), α_k $(k = 1, \ldots, K-1)$ and β_j $(j = 1, \ldots, J-1)$ are category oriented parameters that influence the response of the ith individual to be in the kth and jth categories of the respective response; θ_{k1} and θ_{j2} are the effects of the covariate vector of dimensions p and q, say, on the response variables z_i and y_i, respectively. Note that x_{i1} and x_{i2} may contain certain common and fixed covariates. For example, one may consider

$$x_{i1} = (x_{iz}' : 1 \times p_1, x_{ic}' : 1 \times p_2)' : p \times 1, \; x_{i2} = (x_{iy}' : 1 \times q_1, x_{ic}' : 1 \times q_2)' : q \times 1,$$

where x_{iz} and x_{iy} are individual response specific covariates and x_{ic} is a common covariate vector influencing both responses of the ith individual.

Because y_i and z_i are recorded from the same ith individual, they are likely to be correlated. Recall from (1.22) that some of the existing approaches accommodate this type of correlations or associations by introducing certain joint categorical based association parameters. But as explained previously, this approach produces complicated marginal probabilities. Also this approach encounters a major problem when associations are likely to be individual covariates dependent. To avoid this type of modeling problem, we now discuss a linear probability model by pretending as though the variables were normal. Thus, we write

$$\begin{aligned}
\eta^{(y)}_{ij|k} &= Pr(y_i = y_i^{(j)} | z_i = z_i^{(k)}) \\
&= \pi^{(j)}_{iy} + \sum_{u=1}^{K-1} \rho_{uj}(z^{(k)}_{iu} - \pi^{(u)}_{iz}),\ j = 1, \ldots, J-1,\ k = 1, \ldots, K; \\
\text{and } \eta^{(y)}_{iJ|k} &= Pr(y_i = y_i^{(J)} | z_i = z_i^{(k)}) \\
&= 1 - \sum_{j=1}^{J-1} \eta^{(y)}_{ij|k},\ k = 1, \ldots, K, \qquad (2.3)
\end{aligned}$$

where $z^{(k)}_{iu}$ is the uth $(u = 1, \ldots, K-1)$ component of $z^{(k)}_i$, with $z^{(k)}_{iu} = 1$ if $u = k$, and 0 otherwise; ρ_{uj} is referred to as the dependence parameter relating y_{ij} with z_{iu}.

Note that in writing (2.3), we have used the conditioning on z_i, i.e., we assume that z_i acts as a fixed covariate which is the realized value of the random variable Z_i. One may also use alternatively the conditional probability for z_i given y_i. To be specific, by changing the dependence parameters, this can be written as

$$\begin{aligned}
\tilde{\eta}^{(z)}_{ik|j} &= Pr(z_i = z_i^{(k)} | y_i = y_i^{(j)}) \\
&= \pi^{(k)}_{iz} + \sum_{l=1}^{J-1} \tilde{\rho}_{lk}(y^{(j)}_{il} - \pi^{(l)}_{iy}),\ k = 1, \ldots, K-1,\ j = 1, \ldots, J; \qquad (2.4) \\
\text{and } \tilde{\eta}^{(z)}_{iK|j} &= Pr(z_i = z_i^{(K)} | y_i = y_i^{(j)}) \\
&= 1 - \sum_{k=1}^{K-1} \tilde{\eta}^{(z)}_{ik|j},\ j = 1, \ldots, J.
\end{aligned}$$

However, in this chapter, we follow the model in (2.3) only. Remark that if necessary one can derive the relationship between $\{\rho_{uj}\}$ and $\{\tilde{\rho}_{lk}\}$. For example, in the simple bivariate binary case, suppose the response variable z_i follows a binary distribution with marginal probability $\pi_{iz} = Pr(z_i = 1)$, and the other binary response variable y_i has marginal probability $\pi_{iy} = Pr(y_i = 1)$. Following model (2.3) We then write

$$\begin{aligned} \rho_{iyz} = corr(y_i, z_i) &= \frac{\pi_{i11} - \pi_{iz}\pi_{iy}}{\sqrt{\pi_{iz}(1-\pi_{iz})\pi_{iy}(1-\pi_{iy})}} \\ &= \rho_{11}\sqrt{\frac{\pi_{iz}(1-\pi_{iz})}{\pi_{iy}(1-\pi_{iy})}}, \end{aligned} \tag{2.5}$$

where $\pi_{i11} = Pr(y_i = 1, z_i = 1) = Pr(z_i = 1)P(y_i = 1|z_i = 1) = \pi_{iz}\eta^{(y)}_{i1|1} = \pi_{iz}[\pi_{iy} + \rho_{11}(1-\pi_{iz})]$. Similarly, we can write $\rho_{iyz} = \tilde{\rho}_{11}\sqrt{\frac{\pi_{iy}(1-\pi_{iy})}{\pi_{iz}(1-\pi_{iz})}}$ by using the alternative modeling in (2.4), yielding that $\rho_{iyz} = \sqrt{\tilde{\rho}_{11}\rho_{11}}$.

Note that some authors have studied correlated multinomial data in the univariate longitudinal setup. See, for example, the unpublished PhD thesis by Chowdhury (2011). In that thesis, for example, when a univariate categorical response y_i for the ith individual is collected at $T = 2$ time points, i.e., $y_i = (y^{(j)}_{i1}, y^{(k)}_{i2})$, it is of scientific interest to understand the correlations between y_{i1} and y_{i2}. Similar to but different than this univariate longitudinal setup, in this thesis we deal with correlated bivariate multinomial responses collected from the same person at a single point of time. Because the two responses are collected from the same person, it is also of scientific interest to understand the correlations between them.

Remark that conditional linear models similar to (2.3) were also used in the literature (Zeger, Liang and Self (1985), Qaqish (2003)) to explain the dependence among repeated binary responses. The difference lies in the dimensions as in the present model one usually deals with $K \geq 2$, and $J \geq 2$. Further remark that the dependence parameter ρ_{kj} $(k = 1, \ldots, K-1,\ j = 1, \ldots, J-1)$ in the linear conditional relationship (2.3) has to satisfy certain range restrictions. For example, when $J = 2$ and $K = 2$, the dependence parameter ρ_{11} in the conditional probability of y_i given z_i has the restriction given by

$$max\{-\frac{\pi_{iy}}{1-\pi_{iz}}, -\frac{1-\pi_{iy}}{\pi_{iz}}\} \leq \rho_{11} \leq min\{\frac{1-\pi_{iy}}{1-\pi_{iz}}, \frac{\pi_{iy}}{\pi_{iz}}\},$$

where $\pi_{iy} = Pr(Y_i = 1)$ and $\pi_{iz} = Pr(Z_i = 1)$, respectively. Note that this range indicates that correlations can be negative. See, for example, Sutradhar (2011, Table 7.1, P255) for correlation ranges under different models. Further note that these correlation parameters, whether they take positive or negative values, ultimately play roles to influence the joint probabilities. In the thesis, we are interested in studying models for joint probabilities in the bivariate multinomial setup. Further note that these range restrictions are usually taken care of during estimation of the parameters by checking the range for conditional probabilities at every stage. In general, if a proper efficient method is used for estimation, one can obtain the estimates for these parameters whatever narrow ranges they might have to satisfy.

As we will discuss below, this linear conditional model is very simple for inferences. However, we will deal with alternative bivariate modeling in Chapters 3 and 4 which do not have any range restrictions for the dependence parameters of the model. As far as the estimation of the regression parameters involved in marginal (2.1)-(2.2) and conditional probabilities (2.3) (also functions of the marginal probabilities) in this chapter is concerned, the regression parameters in the marginal probabilities do not arise through any addition among them and the covariates involved in x_{i1} and x_{i2} are independent (mutually exclusive). Consequently there does not arise any identification problems among these parameters.

We now provide the basic properties of the model (2.1)-(2.3) in Section 2.1.1. The inferences will be discussed in Section 2.2.

2.1.1 Basic properties of the LCBM fixed effects model

The marginal means and variances of the bivariate responses are given in Lemma 2.1, and the joint moment between two multinomial responses are given in Lemma 2.2.

Lemma 2.1: For $i = 1, \ldots, n$, the unconditional mean vector and the covariance matrix of the multinomial response vector $Z_i = (Z_{i1}, \ldots, Z_{ik}, \ldots, Z_{i,K-1})'$ have the forms

$$E(Z_i) = (\pi_{iz}^{(1)}, \ldots, \pi_{iz}^{(k)}, \ldots, \pi_{iz}^{(K-1)})' = \Pi_{iz}, \tag{2.6}$$

and

$$Var(Z_i) = diag[\pi_{iz}^{(1)}, \ldots, \pi_{iz}^{(k)}, \ldots, \pi_{iz}^{(K-1)}] - \Pi_{iz}\Pi_{iz}'; \tag{2.7}$$

similarly, the unconditional mean vector and the covariance matrix of the multinomial response vector $Y_i = (Y_{i1}, \ldots, Y_{ij}, \ldots, Y_{i,J-1})'$ have the forms

$$E(Y_i) = (\pi_{iy}^{(1)}, \ldots, \pi_{iy}^{(j)}, \ldots, \pi_{iy}^{(J-1)})' = \Pi_{iy}, \tag{2.8}$$

and

$$Var(Y_i) = diag[\pi_{iy}^{(1)}, \ldots, \pi_{iy}^{(j)}, \ldots, \pi_{iy}^{(J-1)}] - \Pi_{iy}\Pi_{iy}'. \tag{2.9}$$

Proof: These properties follow from the assumed marginal distributions of Z_i and Y_i given by:

$$f(Z_{i1} = z_{i1}, \ldots, Z_{i,K-1} = z_{i,K-1}) = \frac{1!}{z_{i1}! \ldots z_{iu}! \ldots z_{iK}!} \prod_{u=1}^{K} \left(\pi_{iz}^{(u)}\right)^{z_{iu}},$$

and

$$f(Y_{i1} = y_{i1}, \ldots, Y_{i,J-1} = y_{i,J-1}) = \frac{1!}{y_{i1}! \ldots y_{il}! \ldots y_{iJ}!} \prod_{l=1}^{J} \left(\pi_{iy}^{(l)}\right)^{y_{il}}.$$

respectively. This is because it can be shown that Z_{ik}, for example, follows the binary distribution $Bin(\pi_{iz}^{(k)})$, yielding

$$\begin{aligned} E(Z_{ik}) &= \pi_{iz}^{(k)}, \\ \text{and } Var(Z_{ik}) &= \pi_{iz}^{(k)}(1 - \pi_{iz}^{(k)}). \end{aligned}$$

Furthermore, $cov(Z_{ik}, Z_{iu})$ for $k \neq u$ is given by

$$cov(Z_{ik}, Z_{iu}) = E(Z_{ik} Z_{iu}) - E(Z_{ik})E(Z_{iu}) = -\pi_{iz}^{(k)} \pi_{iz}^{(u)},$$

as the quantity $Z_{ik} Z_{iu}$ represents an impossible event.

Lemma 2.2: For $i = 1, \ldots, n$, the covariance matrix $Cov(Z_i, Y_i')$ of the bivariate multinomial responses Z_i and Y_i is given by:

$$Cov(Z_i, Y_i') = [Var(Z_i)]\, \Phi : (K-1) \times (J-1), \tag{2.10}$$

where $Var(Z_i)$ is given in (2.5), and Φ is the $(K-1) \times (J-1)$ matrix involving dependence parameters $\rho_{kj} : k = 1, \ldots, K-1, j = 1, \ldots, J-1$, and is given by

$$\begin{aligned} \Phi = \left(\rho_{kj} \right)_{(K-1)\times(J-1)} &= [\rho_1, \ldots, \rho_j, \ldots, \rho_{J-1}] \\ &= \begin{pmatrix} \rho_{11} & \cdots & \rho_{1j} & \cdots & \rho_{1,J-1} \\ \rho_{21} & \cdots & \rho_{2j} & \cdots & \rho_{2,J-1} \\ & & \vdots & & \\ \rho_{k1} & \cdots & \rho_{kj} & \cdots & \rho_{k,J-1} \\ & & \vdots & & \\ \rho_{K-1,1} & \cdots & \rho_{K-1,j} & \cdots & \rho_{K-1,J-1} \end{pmatrix}. \end{aligned} \tag{2.11}$$

Proof: To prove (2.8), we first derive the covariance between two general elements Z_{ik} and Y_{ij} for $k = 1, \ldots, K-1$ and $j = 1, \ldots, J-1$. That is, we write

$$\begin{aligned} cov(Z_{ik}, Y_{ij}) &= E(Y_{ij}Z_{ik}) - E(Z_{ik})E(Y_{ij}) \\ &= \pi_{ikj} - \pi_{iz}^{(k)}\pi_{iy}^{(j)}, \end{aligned}$$

where by (2.1) and (2.3), one computes

$$\pi_{ikj} = \pi_{iz}^{(k)}\eta_{ij|k}^{(y)} = \pi_{iz}^{(k)}\left[\pi_{iy}^{(j)} + \sum_{u=1}^{K-1}\rho_{uj}(z_{iu}^{(k)} - \pi_{iz}^{(u)})\right], \tag{2.12}$$

yielding the covariance as:

$$\begin{aligned} cov(Z_{ik}, Y_{ij}) &= \pi_{iz}^{(k)}\left[\pi_{iy}^{(j)} + \sum_{u=1}^{K-1}\rho_{uj}(z_{iu}^{(k)} - \pi_{iz}^{(u)})\right] - \pi_{iz}^{(k)}\pi_{iy}^{(j)} \\ &= \pi_{iz}^{(k)}\sum_{u=1}^{K-1}\rho_{uj}(z_{iu}^{(k)} - \pi_{iz}^{(u)}) \\ &= \rho_{kj}\pi_{iz}^{(k)}(1 - \pi_{iz}^{(k)}) - \sum_{u=1, u\neq k}^{K-1}\rho_{uj}\pi_{iz}^{(k)}\pi_{iz}^{(u)}. \end{aligned} \tag{2.13}$$

Now, following (2.11) we write the $K-1$ covariance quantities in $cov(Z_i, Y_{ij}) = [cov(Z_{i1}, Y_{ij}), \ldots, cov(Z_{ik}, Y_{ij}), \ldots, cov(Z_{i,K-1}, Y_{ij})]'$ as follows:

$$\begin{aligned} cov(Z_{i1}, Y_{ij}) &= \rho_{1j}\pi_{iz}^{(1)}(1 - \pi_{iz}^{(1)}) - \sum_{u=2}^{K-1}\rho_{uj}\pi_{iz}^{(1)}\pi_{iz}^{(u)}, \\ &\vdots \\ cov(Z_{ik}, Y_{ij}) &= \rho_{kj}\pi_{iz}^{(k)}(1 - \pi_{iz}^{(k)}) - \sum_{u=1, u\neq k}^{K-1}\rho_{uj}\pi_{iz}^{(k)}\pi_{iz}^{(u)}, \\ &\vdots \\ cov(Z_{i,K-1}, Y_{ij}) &= \rho_{K-1,j}\pi_{iz}^{(K-1)}(1 - \pi_{iz}^{(K-1)}) - \sum_{u=1}^{K-2}\rho_{uj}\pi_{iz}^{(K-1)}\pi_{iz}^{(u)}; \end{aligned}$$

which, in a matrix form, is given by:

$$
\begin{aligned}
\begin{pmatrix} cov(Z_{i1}, Y_{ij}) \\ \vdots \\ cov(Z_{ik}, Y_{ij}) \\ \vdots \\ cov(Z_{i,K-1}, Y_{ij}) \end{pmatrix}
&= \begin{pmatrix}
\pi_{iz}^{(1)}(1-\pi_{iz}^{(1)}) & \dots & \dots & \dots & -\pi_{iz}^{(1)}\pi_{iz}^{(K-1)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
-\pi_{iz}^{(k)}\pi_{iz}^{(1)} & \dots & \pi_{iz}^{(k)}(1-\pi_{iz}^{(k)}) & \dots & -\pi_{iz}^{(k)}\pi_{iz}^{(K-1)} \\
\vdots & \vdots & \vdots & \vdots & \vdots \\
-\pi_{iz}^{(K-1)}\pi_{iz}^{(1)} & \dots & \dots & \dots & \pi_{iz}^{(K-1)}(1-\pi_{iz}^{(K-1)})
\end{pmatrix}
\begin{pmatrix} \rho_{1j} \\ \vdots \\ \rho_{kj} \\ \vdots \\ \rho_{K-1,j} \end{pmatrix} \\
&= [diag(\Pi_{iz}) - \Pi_{iz}\Pi_{iz}']\rho_j \\
&= Var(Z_i)\rho_j. \qquad (2.14)
\end{aligned}
$$

We now combine the results from (2.12) for every $j = 1, \dots, J-1$, and write

$$
\begin{aligned}
Cov(Z_i, Y_i') &= [Var(Z_i)\rho_1, \dots, Var(Z_i)\rho_j, \dots, Var(Z_i)\rho_{J-1}] \\
&= Var(Z_i)\,[\rho_1, \dots, \rho_j, \dots, \rho_{J-1}] \\
&= Var(Z_i)\Phi.
\end{aligned}
$$

Hence Lemma 2.2 follows.

Lemma 2.3: The joint probabilities based on (2.1)-(2.3) retain the specified marginal probabilities.

Proof: Note that $\pi_{ikj} = Pr(z_i = z_i^{(k)}, y_i = y_i^{(j)}) = \pi_{iz}^{(k)}\eta_{ij|k}^{(y)}$ holds for any specified marginal probability for z_i, i.e., for $\pi_{iz}^{(k)}$. Now because

$$\begin{aligned}
\sum_{k=1}^{K} \pi_{ikj} &= \sum_{k=1}^{K} \pi_{iz}^{(k)} \eta_{ij|k}^{(y)} \\
&= \sum_{k=1}^{K} \pi_{iz}^{(k)} \left[\pi_{iy}^{(j)} + \sum_{u=1}^{K-1} \rho_{uj} (z_{iu}^{(k)} - \pi_{iz}^{(u)}) \right] \\
&= \sum_{k=1}^{K} \pi_{iz}^{(k)} \pi_{iy}^{(j)} + \sum_{k=1}^{K} \pi_{iz}^{(k)} \sum_{u=1}^{K-1} \rho_{uj} (z_{iu}^{(k)} - \pi_{iz}^{(u)}) \\
&= \pi_{iy}^{(j)} + \sum_{u=1}^{K-1} \rho_{uj} \left[\sum_{k=1}^{K} \pi_{iz}^{(k)} z_{iu}^{(k)} - \sum_{k=1}^{K} \pi_{iz}^{(k)} \pi_{iz}^{(u)} \right] \\
&= \pi_{iy}^{(j)} + \sum_{u=1}^{K-1} \rho_{uj} (\pi_{iz}^{(u)} - \pi_{iz}^{(u)}) \\
&= \pi_{iy}^{(j)},
\end{aligned}$$

one may use any desired formula for $\pi_{iy}^{(j)}$.

Note from (2.11) that on top of the marginal probabilities $\pi_{iz}^{(k)}$, the correlation between Y_{ij} and Z_{ik}, say $corr(Z_{ik}, Y_{ij})$, is a function of the components of ρ_j vector, where $\rho_j = (\rho_{1j}, \ldots, \rho_{kj}, \ldots, \rho_{K-1,j})'$. In this sense, ρ_{kj} can also be viewed as a correlation index parameter. Further note that the LCBM fixed model defined by (2.1)-(2.3) may also be referred to as the bivariate multinomial fixed effects (BMFE) model. This is because the covariates used in marginal probabilities (2.1) and (2.2) are considered to be fixed.

In the next section, we demonstrate how the marginal and joint moment properties given in Lemmas 2.1-2.2 of the LCBM fixed model (2.1)-(2.3) can be exploited for developing suitable estimating equations for all regression parameters $\psi = (\alpha_1, \ldots, \alpha_{K-1}, \theta_1',$

$\beta_1, \ldots, \beta_{J-1}, \theta_2')'$, with $\theta_1 = (\theta_{11}', \ldots, \theta_{k1}', \ldots, \theta_{K-1,1}')'$ and $\theta_2 = (\theta_{12}', \ldots, \theta_{j2}', \ldots, \theta_{J-1,2}')'$; and correlation index parameters $\rho^* = (\rho_1', \ldots, \rho_j', \ldots, \rho_{J-1}')' = (\rho_{11}, \rho_{21}, \ldots, \rho_{K-1,1},$

$\rho_{12}, \ldots, \rho_{kj}, \ldots, \rho_{K-1,J-1})'$.

2.2 Estimation for the LCBM fixed model

Let $\phi = (\psi', \rho^{*'})'$ denote the vector of all regression and correlation index parameters. One may estimate these parameters by solving likelihood estimating equations where the likelihood construction requires knowledge of joint probabilities. To be specific, the likelihood function for ϕ may be written as:

$$L(\phi) = \prod_{i=1}^{n} \pi_{i11}^{z_{i1}y_{i1}} \ldots \pi_{ikj}^{z_{ik}y_{ij}} \ldots \pi_{iKJ}^{z_{iK}y_{iJ}}.$$

Note, however, that because

$$\pi_{ikj} = \pi_{iz}^{(k)} \left[\pi_{iy}^{(j)} + \sum_{u=1}^{K-1} \rho_{uj} (z_{iu}^{(k)} - \pi_{iz}^{(u)}) \right],$$

solving the exact likelihood equations, i.e.,

$$\frac{\partial lnL(\phi)}{\partial \phi} = \sum_{i=1}^{n} \sum_{k=1}^{K} \sum_{j=1}^{J} z_{ik} y_{ij} \frac{\partial log \pi_{ikj}}{\partial \phi},$$

by exploiting the complicated second order derivatives is algebraically cumbersome. Thus we will exploit an alternatively simpler user-friendly GQL (generalized quasi-likelihood) [Sutradhar (2004)] approach to estimate the parameters in ϕ.

Remark that the GQL approach was suggested by Sutradhar (2003) [see also Sutradhar (2004), Sutradhar (2010) (a), and Sutrahdar (2010) (b)] as a generalization of the quasi-likelihood (QL) approach for independent data suggested by Wedderburn (1974). To be specific, this approach minimizes a generalized quadratic distance function, where the distance function is constructed based on true mean, variance, and correlation structures of the data, whereas the QL approach was constructed based on true mean and variance structures only. To be precise and clear, consider Z_i variable which has mean $\Pi_{iz}(\psi_z)$ and covariance structure $Var(Z_i) = \Sigma_{iz}$ involving variances and correlations. One then write the GQL estimating equation for ψ_z by

$$f(\psi_z) = \sum_{i=1}^{n} \frac{\partial \Pi'_{iz}}{\partial \psi_z} \Sigma_{iz}^{-1}(z_i - \Pi_{iz}(\psi_z)) = 0,$$

[Sutradhar (2003), Section 3].

Turning back to the estimation of ϕ, we discuss two versions of the GQL approach, namely, (i) a joint GQL (JGQL) approach for $\phi = (\psi', \rho^{*'})'$, and (ii) a single stage GQL (SSGQL) for ψ.

2.2.1 JGQL approach

In this approach, we exploit the marginal and product moments directly to construct the desired JGQL estimating equation for ϕ. Note that this type of equation, as shown below, requires only first derivative of the moments with respect to the parameters.

Recall from Lemma 2.1 that the first order moments

$$\Pi_{iz} = (\pi_{iz}^{(1)}, \ldots, \pi_{iz}^{(k)}, \ldots, \pi_{iz}^{(K-1)})', \text{ and } \Pi_{iy} = (\pi_{iy}^{(1)}, \ldots, \pi_{iy}^{(j)}, \ldots, \pi_{iy}^{(J-1)})'$$

are functions of ψ only, whereas ρ^* is involved in the second order moments in (2.10). Consequently, to estimate ψ and ρ^* jointly, we develop a GQL estimating equation based on first and second order responses. To be specific, we exploit z_i, y_i and the elements of the $z_i y_i'$ matrix to construct the estimating equations. Let

$$g_i = (z_{i1}y_{i1}, \ldots, z_{i1}y_{i,J-1}, z_{i2}y_{i1}, \ldots, z_{ik}y_{ij}, \ldots, z_{i,K-1}y_{i,J-1})'$$

be a stacked vector of second order responses from the $z_i y_i'$ matrix and we write its expectation as

$$\Pi_{izy} = E(g_i) = (\pi_{i11}, \ldots, \pi_{i1,J-1}, \pi_{i21}, \ldots, \pi_{ikj}, \ldots, \pi_{iK-1,J-1})'.$$

Note that the joint probabilities in Π_{izy} are functions of both ψ and ρ^* parameters. We may now write the joint GQL estimating equations for ϕ as

$$f(\phi) = \sum_{i=1}^{n} \frac{\partial(\Pi'_{iz}, \Pi'_{iy}, \Pi'_{izy})}{\partial\phi} \Sigma_i^{-1} \begin{pmatrix} z_i - \Pi_{iz} \\ y_i - \Pi_{iy} \\ g_i - \Pi_{izy} \end{pmatrix} = 0, \qquad (2.15)$$

[Sutradhar (2004)]. In (2.13) Σ_i is the covariance matrix of $(z'_i, y'_i, g'_i)'$, which has the form

$$\Sigma_i = Var\begin{pmatrix} \begin{pmatrix} z_i \\ y_i \end{pmatrix} \\ g_i \end{pmatrix} = \begin{pmatrix} \Sigma_{i11} & \Sigma_{i12} \\ & \Sigma_{i22} \end{pmatrix},$$

where

$$\begin{aligned} \Sigma_{i11} &= Var\begin{pmatrix} z_i \\ y_i \end{pmatrix} = \begin{pmatrix} Var(z_i) & Cov(z_i, y'_i) \\ & Var(y_i) \end{pmatrix} \\ &= \begin{pmatrix} diag(\Pi_{iz}) - \Pi_{iz}\Pi'_{iz} & [diag(\Pi_{iz}) - \Pi_{iz}\Pi'_{iz}]\Phi \\ & diag(\Pi_{iy}) - \Pi_{iy}\Pi'_{iy} \end{pmatrix}, \end{aligned}$$

with Φ as given in (2.9). Next,

$$\Sigma_{i12} = \begin{pmatrix} Cov(z_i, g'_i) \\ Cov(y_i, g'_i) \end{pmatrix},$$

and

$$\Sigma_{i22} = Var(g_i) = diag(\Pi_{izy}) - \Pi_{izy}\Pi'_{izy}.$$

2.2.2 SSGQL approach

The SSGQL (single-stage GQL) is slightly different than the JGQL approach. In this approach, for known ρ^*, we exploit the first order moments to estimate ψ parameter at the first stage. Once an estimate of ψ is available, we use it as a known value of ψ in the estimating equation for ρ^*, which is developed exploiting both first and second order moments. Thus we first write the GQL estimating equation for ψ as

$$f(\psi) = \sum_{i=1}^{n} \frac{\partial(\Pi'_{iz}, \Pi'_{iy})}{\partial\psi} \Sigma_{i11}^{-1} \begin{pmatrix} z_i - \Pi_{iz} \\ y_i - \Pi_{iy} \end{pmatrix} = 0, \tag{2.16}$$

where Σ_{i11} is given in (2.13). At the second stage we estimate ρ^* or equivalently Φ by using the well-known method of moments (MM). Recall from (2.8) that

$$Cov(Z_i, Y_i') = [Var(Z_i)]\, \Phi.$$

Because this relationship holds for all $i = 1, \ldots, n$, by taking averages on both sides, we obtain the moment estimator of Φ as

$$\hat{\Phi}_{MM} = \left[\frac{1}{n}\sum_{i=1}^{n} \hat{Var}(Z_i)\right]^{-1} \left[\frac{1}{n}\sum_{i=1}^{n} \hat{Cov}(Z_i, Y_i')\right], \tag{2.17}$$

where $\hat{cov}(Z_{ik}, Y_{ij})$ in $\hat{Cov}(Z_i, Y_i')$, for example, has the formula, $\hat{cov}(Z_{ik}, Y_{ij}) = (z_{ik} - \hat{\pi}_{iz}^{(k)})(y_{ij} - \hat{\pi}_{iy}^{(j)})$.

Note that as mentioned earlier the dependence parameters ρ_{kj} in Φ have certain range restrictions. This is because the conditional probabilities in (2.3) may not satisfy the range $0 < \eta_{ij|k}^{(y)} < 1$ for all values of ρ_{kj}. However, the aforementioned inference procedures, specifically the method of moments for the estimation of Φ by (2.15) yields consistent estimates for the true parameter values whatever their range may be. In any cases, because the LCBM fixed model discussed in this chapter, in spite of its simplicity, encounters the range restriction problem discussed above, we will develop a random effects based general non-linear bivariate multinomial model in Chapter 3, which is not adequately discussed in the literature.

2.2.3 A simulation study

In this section, we conduct a small sample study to examine the relative performance of the JGQL and SSGQL methods discussed in Sections 2.2.1 and 2.2.2, respectively. For sample size we consider $n = 200$.

Note that we consider these two GQL estimation approaches in this simulation study, as they are founded on the same principle of the GQL approach which is known to produce consistent estimates. However, their empirical performances may be slightly different from each other because of the different ways the correlations are estimated. Further note that even though we have included (1) ”working” odds ratio and (2) joint probability modeling approaches in our discussion in Section 1.2.1 and 1.2.2, these approaches however do not appear to be feasible for any comparison with the above GQL approaches. To be specific, even though we are considering some common covariates, there is no guidence to chose x_{ic} covariate vector in the extra model (1.31) for odds ratios. Also, as we indicated in Section 1.2.1, even though it is quite reasonable to model individual based odds ratio (τ_{ikj}) in terms of few parameters, there is no justification why a linear model is appropriate. In fact, there is no unique way of choosing the ”working” model. Thus we do not include such arbitrary model based estimation approach in our comparison.

As far as the joint probability modeling in Section 1.2.2 is concerned, this type of joint modeling does not produce our marginal probabilities. Consequently, this approach is not feasible for comparison.

We now turn back to consider the GQL approaches. To generate the data following models (2.1)-(2.3) in a given simulation, we first consider the simulation design as follows:

Covariate selection and marginal specification:

- We set $K = J = 3$ for number of categories for z and y.
- To represent the category effects we specify the marginal probabilities as in (2.1)-(2.2) using $\alpha_1 = 0.4$, $\alpha_2 = 0.3$; and $\beta_1 = 0.35$, $\beta_2 = 0.25$.
- Next, we consider that x_{i1} and x_{i2} are of dimension 2×1, with $x_{i1} = (x_{iz}, x_{ic})'$ and $x_{i2} = (x_{iy}, x_{ic})'$, where x_{iz} is the covariate specific to response variable z and x_{iy} is the covariate specific to response variable y, and x_{ic} is the common covariate shared by both response variables. We choose the covariates design as follows:

$$x_{iz} \sim Binary(0.4),\ x_{iy} \sim Binary(0.7), \text{ and } x_{ic} \sim Standardized\, U(0,1).$$

Design selection:

With regard to selection of regression parameters, namely, $\theta_1 \equiv (\theta_{z1}, \theta_{cz1}, \theta_{z2}, \theta_{cz2})'$, $\theta_2 \equiv (\theta_{y1}, \theta_{cy1}, \theta_{y2}, \theta_{cy2})'$ and $\rho^* \equiv (\rho_{11}, \rho_{21}, \rho_{12}, \rho_{22})'$, we consider the following eight designs with various magnitude (small and big) for the parameters:

- Design 1 (D1):

$\theta_{z1} = \theta_{z2} = \theta_z = 0.25$, $\theta_{y1} = \theta_{y2} = \theta_y = 0.4$, $\theta_{cz1} = \theta_{cz2} = \theta_{cy1} = \theta_{cy2} = \theta_c = 0.2$;

$\rho_{11} = 0.4$, $\rho_{21} = 0.2$, $\rho_{12} = 0.15$ and $\rho_{22} = 0.35$.

- Design 2 (D2):

  $\theta_{z1} = \theta_{z2} = \theta_z = 0.25$, $\theta_{y1} = \theta_{y2} = \theta_y = 0.4$, $\theta_{cz1} = \theta_{cy1} = \theta_{c1} = 0.2$, $\theta_{cz2} = \theta_{cy2} = \theta_{c2} = 0.3$;

  $\rho_{11} = 0.4$, $\rho_{21} = 0.2$, $\rho_{12} = 0.15$ and $\rho_{22} = 0.35$.

- Design 3 (D3):

  $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = \theta_{cy1} = \theta_{c1} = 0.2$, $\theta_{cz2} = \theta_{cy2} = \theta_{c2} = 0.3$;

  $\rho_{11} = 0.4$, $\rho_{21} = 0.2$, $\rho_{12} = 0.15$ and $\rho_{22} = 0.35$.

- Design 4 (D4):

  $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = \theta_{cy1} = \theta_{c1} = 0.2$, $\theta_{cz2} = \theta_{cy2} = \theta_{c2} = 0.3$;

  $\rho_{11} = 0.2$, $\rho_{21} = 0.0$, $\rho_{12} = 0.0$ and $\rho_{22} = 0.1$.

- Design 5 (D5):

  $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = \theta_{cy1} = \theta_{c1} = 0.2$, $\theta_{cz2} = \theta_{cy2} = \theta_{c2} = 0.3$;

  $\rho_{11} = 0.4$, $\rho_{21} = -0.2$, $\rho_{12} = -0.2$ and $\rho_{22} = 0.4$.

- Design 6 (D6):

  $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = 0.2$, $\theta_{cz2} = 0.3$, $\theta_{cy1} = 0.35$, $\theta_{cy2} = 0.45$;

  $\rho_{11} = 0.4$, $\rho_{21} = 0.2$, $\rho_{12} = 0.15$ and $\rho_{22} = 0.35$.

- Design 7 (D7):

  $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = 0.2$, $\theta_{cz2} = 0.3$, $\theta_{cy1} = 0.35$, $\theta_{cy2} = 0.45$;

  $\rho_{11} = 0.2$, $\rho_{21} = 0.0$, $\rho_{12} = 0.0$ and $\rho_{22} = 0.1$.

- Design 8 (D8):

  $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = 0.2$, $\theta_{cz2} = 0.3$, $\theta_{cy1} = 0.35$, $\theta_{cy2} = 0.45$;

  $\rho_{11} = 0.3$, $\rho_{21} = -0.2$, $\rho_{12} = -0.2$ and $\rho_{22} = 0.3$.

Data generation:

As far as the generation of response variables z_i and y_i are concerned, we follow the steps given below:

- Step 1: We first generate the trinomial response z_i following model (2.1) by using suitable regression parameters selected above.

- Step 2: Given the value of z_i, we follow models (2.2) and (2.3) to generate the response y_i by taking suitable regression and correlation index parameters into consideration.

Simulation results:

In a given simulation, we then use the data along with design covariates to estimate the parameters by the JGQL (Section 2.2.1) and SSGQL (Section 2.2.2) approaches. These approaches require initial estimates for all regression and correlation parameters, we have chosen these initial values close to zero. The data generation and estimation is repeated for 500 times. The convergence of estimates was quick in all cases and initial values were not any issue. The simulated mean (SM), simulated standard error (SSE), mean squared error (MSE), and estimated standard error (ESE) of the JGQL and SSGQL estimates under the above eight designs are reported in Tables 2.1-2.8, respectively.

The results from Table 2.1-2.8 show that both methods are performing well in estimating parameters involved in the LCBM fixed model. However, for some designs the JGQL approach appears to produce estimates with larger standard errors than the SSGQL approach, but for some other designs the SSGQL approach produces estimates with larger standard errors. This can happen because of the design matrix used in the study, not because of any convergence problems for the method themselves. Thus, the two approaches appear to be quite competitive because the overall MSEs are not too different under these two approaches. For example, under D1, the JGQL estimated $\theta_y = 0.4$ as $\hat{\theta}_{y,JGQL} = 0.382$ with standard error 0.316, and the SSGQL estimated this parameter as $\hat{\theta}_{y,SSGQL} = 0.388$ with standard error 0.373, which is larger than the SSE under the JGQL approach. On the other hand, under D2, the JGQL approach estimated $\beta_1 = 0.35$ as $\hat{\beta}_{1,JGQL} = 0.344$ with standard error 0.328, whereas $\hat{\beta}_{1,SSGQL} = 0.345$ with standard error 0.301, which is smaller than the SSE under JGQL. Now because we have considered a wide range of designs to examine the performances of JGQL and SSGQL approaches, based on the simulation results in Tables 2.1-2.8, we can see that both approaches are competitive. Note, however, that the SSGQL approach is simpler than the JGQL approach computationally, even though it is slightly less efficient.

Next, the aforementioned tables show that the estimated standard errors (ESE) of the JGQL and SSGQL estimation approaches are close to their corresponding SSE's.

For example, under D1, the JGQL approach estimates $\alpha_2 = 0.3$ as $\hat{\alpha}_{2,JGQL} = 0.314$ with SSE equal to 0.213 and ESE equal to 0.211; similarly the SSGQL approach estimates the same parameter as $\hat{\alpha}_{2,SSGQL} = 0.317$ with SSE equal to 0.234 and ESE equal to 0.230. Note that in both cases the ESE's are very close to the corresponding values of SSE. Furthermore, under D5, the JGQL approach estimates $\theta_{z2} = 0.35$ as $\hat{\theta}_{z2,JGQL} = 0.357$ with SSE equal to 0.359 and ESE equal to 0.360, and the SSGQL approach estimates the same parameter as $\hat{\theta}_{z2,SSGQL} = 0.379$ with SSE equal to 0.379 and ESE equal to 0.389, showing that the ESE under the two estimation approaches are again close to their corresponding SSE's. Thus, in general, the estimated standard errors formulas obtained from the estimating equations (2.13) and (2.14), for the JGQL and SSGQL approaches, respectively, work well, as expected.

With regard to the role of the correlation index parameters on the regression estimates, we observe from the Tables 2.1-2.8 that when correlations change, the standard errors of the regression estimates appear to change sometimes substantially. For example, when Tables 2.3, 2.4, 2.5 are compared, the same regression parameters $\theta_{z1} = 0.25$ and $\theta_{z2} = 0.35$ are estimated with standard errors 0.191 and 0.173 under Table 2.4 with $\rho_{11} = 0.2$, $\rho_{21} = 0.0$, $\rho_{12} = 0.0$ and $\rho_{22} = 0.1$; and were estimated with standard errors 0.177 and 0.152 under Table 2.3 with $\rho_{11} = 0.4$, $\rho_{21} = 0.2$, $\rho_{12} = 0.15$ and $\rho_{22} = 0.35$; and they are estimated with standard errors 0.122 and 0.129 under Table 2.5 with $\rho_{11} = 0.4$, $\rho_{21} = -0.2$, $\rho_{12} = -0.2$ and $\rho_{22} = 0.4$.

Note that the JGQL and SSGQL approaches were developed by accommodating the correlations between bivariate responses. This was done to obtain improved regression estimates in the sense of MSE efficiency as compared to the independence assumption based such as the quasi-likelihood (QL) and other possible approaches.

Now to understand the efficiency gain due to the use of correlations involved in JGQL and SSGQL approaches, we include the QL approach (does not involve correlations) as well in two of our simulation experiments, namely, the studies using designs D3 and D4. For convenience, we provide below the QL estimating equations for the regression parameters, where no correlation parameters are needed:

$$f(\psi) = \sum_{i=1}^{n} \frac{\partial(\Pi'_{iz}, \Pi'_{iy})}{\partial\psi} [\Sigma^*_{i11}]^{-1} \begin{pmatrix} z_i - \Pi_{iz} \\ y_i - \Pi_{iy} \end{pmatrix} = 0, \tag{2.18}$$

with

$$\Sigma^*_{i11} = \begin{pmatrix} Var(z_i) & 0 \\ & Var(y_i) \end{pmatrix} = \begin{pmatrix} diag(\Pi_{iz}) - \Pi_{iz}\Pi'_{iz} & 0 \\ & diag(\Pi_{iy}) - \Pi_{iy}\Pi'_{iy} \end{pmatrix}.$$

The simulation results under the QL approach are added to Table 2.3 for D3 and Table 2.4 for D4. Due to space limitation, the mean squared errors are not reported but they can be easily computed by using the SM and SSE values along with the true parameter values. The results of these tables show that the QL estimates are also almost unbiased, indicating that the correlation index parameters do not play any roles in producing unbiased and hence consistent estimates. However, when SSEs and/or MSEs under the QL approach are compared with those of the JGQL and SSGQL approach, the QL approach, as expected, appears to be relatively inefficient. For example, when $\rho_{11} = 0.4$, $\rho_{21} = 0.2$, $\rho_{12} = 0.15$ and $\rho_{22} = 0.35$, the SSGQL approach estimated θ_{z1} and θ_{z2} with MSEs 0.154 and 0.159, respectively, whereas the QL approach produced the estimates with larger MSEs, namely, 0.177 and 0.180, respectively. Similar comparative results follow for Table 2.4 where the SSGQL approach appeared to be slightly competitive in the sense of MSE efficiency. This is because the correlations are not so large, namely they are $\rho_{11} = 0.2$, $\rho_{21} = 0.0$, $\rho_{12} = 0.0$ and

$\rho_{22} = 0.1$. Note that these behavior of the simulation results are not unexpected, because in finite sample cases one method may produce slightly different estimates and/or standard errors than the other. But overall both methods produce consistent estimates.

Table 2.1: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D1 with sample size n=200.

	JGQL				SSGQL			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.403	0.218	0.047	0.208	0.405	0.225	0.051	0.227
$\alpha_2 = 0.3$	0.314	0.213	0.046	0.211	0.317	0.234	0.055	0.230
$\theta_z = 0.25$	0.262	0.291	0.085	0.260	0.275	0.335	0.113	0.347
$\beta_1 = 0.35$	0.379	0.316	0.101	0.275	0.371	0.323	0.105	0.314
$\beta_2 = 0.25$	0.289	0.316	0.101	0.277	0.278	0.329	0.109	0.316
$\theta_y = 0.4$	0.382	0.316	0.100	0.280	0.388	0.373	0.140	0.371
$\theta_c = 0.2$	0.197	0.195	0.038	0.143	0.208	0.153	0.024	0.125
$\rho_{11} = 0.4$	0.403	0.083	0.007	0.077	0.403	0.074	0.006	-
$\rho_{21} = 0.2$	0.200	0.084	0.007	0.078	0.197	0.078	0.006	-
$\rho_{12} = 0.15$	0.145	0.081	0.007	0.077	0.142	0.076	0.006	-
$\rho_{22} = 0.35$	0.351	0.082	0.007	0.080	0.348	0.082	0.007	-

Table 2.2: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D2 with sample size n=200.

	JGQL				SSGQL			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.413	0.241	0.058	0.217	0.411	0.239	0.057	0.231
$\alpha_2 = 0.3$	0.322	0.244	0.060	0.220	0.319	0.240	0.058	0.234
$\theta_z = 0.25$	0.248	0.310	0.096	0.252	0.256	0.342	0.117	0.344
$\beta_1 = 0.35$	0.344	0.328	0.108	0.274	0.345	0.301	0.090	0.290
$\beta_2 = 0.25$	0.253	0.348	0.121	0.276	0.250	0.296	0.087	0.292
$\theta_y = 0.4$	0.411	0.364	0.132	0.285	0.413	0.363	0.132	0.355
$\theta_{c1} = 0.2$	0.193	0.195	0.038	0.160	0.201	0.167	0.028	0.137
$\theta_{c2} = 0.3$	0.292	0.196	0.039	0.162	0.307	0.168	0.028	0.141
$\rho_{11} = 0.4$	0.408	0.098	0.010	0.076	0.405	0.076	0.006	-
$\rho_{21} = 0.2$	0.204	0.099	0.010	0.077	0.196	0.081	0.007	-
$\rho_{12} = 0.15$	0.145	0.085	0.007	0.076	0.140	0.074	0.006	-
$\rho_{22} = 0.35$	0.348	0.089	0.008	0.080	0.347	0.081	0.007	-

Table 2.3: The SM (simulated mean), SSE (simulated standard error), and ESE (estimated standard error) of the JGQL, SSGQL and QL estimates under D3 with sample size n=200.

	JGQL			SSGQL			QL		
Parameter	SM	SSE	ESE	SM	SSE	ESE	SM	SSE	ESE
$\alpha_1 = 0.4$	0.414	0.243	0.219	0.403	0.225	0.230	0.411	0.238	-
$\alpha_2 = 0.3$	0.315	0.240	0.224	0.314	0.228	0.235	0.314	0.238	-
$\theta_{z1} = 0.25$	0.268	0.421	0.315	0.272	0.392	0.394	0.268	0.421	-
$\theta_{z2} = 0.35$	0.376	0.389	0.319	0.366	0.398	0.397	0.374	0.423	-
$\beta_1 = 0.35$	0.371	0.328	0.274	0.373	0.304	0.316	0.351	0.319	-
$\beta_2 = 0.25$	0.276	0.333	0.280	0.266	0.322	0.325	0.263	0.332	-
$\theta_{y1} = 0.4$	0.390	0.400	0.308	0.381	0.402	0.398	0.420	0.406	-
$\theta_{y2} = 0.5$	0.495	0.419	0.315	0.506	0.418	0.405	0.516	0.413	-
$\theta_{c1} = 0.2$	0.201	0.243	0.178	0.202	0.167	0.138	0.220	0.159	-
$\theta_{c2} = 0.3$	0.303	0.286	0.191	0.301	0.162	0.140	0.315	0.162	-
$\rho_{11} = 0.4$	0.397	0.080	0.077	0.401	0.074	-	-	-	-
$\rho_{21} = 0.2$	0.192	0.075	0.077	0.199	0.073	-	-	-	-
$\rho_{12} = 0.15$	0.148	0.099	0.076	0.143	0.075	-	-	-	-
$\rho_{22} = 0.35$	0.351	0.096	0.080	0.343	0.085	-	-	-	-

Table 2.4: The SM (simulated mean), SSE (simulated standard error), and ESE (estimated standard error) of the JGQL, SSGQL and QL estimates under D4 with sample size n=200.

	JGQL			SSGQL			QL		
Parameter	SM	SSE	ESE	SM	SSE	ESE	SM	SSE	ESE
$\alpha_1 = 0.4$	0.402	0.259	0.236	0.402	0.237	0.238	0.407	0.243	-
$\alpha_2 = 0.3$	0.311	0.261	0.243	0.301	0.241	0.243	0.329	0.240	-
$\theta_{z1} = 0.25$	0.275	0.437	0.376	0.282	0.401	0.387	0.278	0.434	-
$\theta_{z2} = 0.35$	0.370	0.415	0.386	0.398	0.382	0.390	0.359	0.407	-
$\beta_1 = 0.35$	0.344	0.348	0.320	0.370	0.330	0.345	0.345	0.324	-
$\beta_2 = 0.25$	0.246	0.354	0.329	0.279	0.345	0.352	0.261	0.343	-
$\theta_{y1} = 0.4$	0.419	0.446	0.382	0.405	0.413	0.417	0.421	0.399	-
$\theta_{y2} = 0.5$	0.524	0.465	0.391	0.503	0.411	0.423	0.514	0.421	-
$\theta_{c1} = 0.2$	0.213	0.163	0.147	0.215	0.152	0.140	0.210	0.152	-
$\theta_{c2} = 0.3$	0.315	0.169	0.141	0.315	0.148	0.142	0.309	0.146	-
$\rho_{11} = 0.2$	0.200	0.096	0.087	0.199	0.086	0.007	-	-	-
$\rho_{21} = 0.0$	-0.005	0.086	0.086	-0.003	0.086	0.007	-	-	-
$\rho_{12} = 0.0$	0.001	0.094	0.086	-0.003	0.084	0.007	-	-	-
$\rho_{22} = 0.1$	0.110	0.094	0.089	0.101	0.090	0.008	-	-	-

Table 2.5: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D5 with sample size n=200.

	JGQL				SSGQL			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.366	0.223	0.051	0.231	0.392	0.236	0.056	0.237
$\alpha_2 = 0.3$	0.284	0.222	0.050	0.235	0.303	0.239	0.057	0.242
$\theta_{z1} = 0.25$	0.266	0.349	0.122	0.355	0.275	0.397	0.159	0.386
$\theta_{z2} = 0.35$	0.357	0.359	0.129	0.360	0.369	0.379	0.144	0.389
$\beta_1 = 0.35$	0.397	0.377	0.144	0.338	0.355	0.345	0.119	0.335
$\beta_2 = 0.25$	0.306	0.347	0.124	0.313	0.267	0.340	0.116	0.342
$\theta_{y1} = 0.4$	0.339	0.426	0.185	0.375	0.409	0.428	0.183	0.410
$\theta_{y2} = 0.5$	0.445	0.410	0.171	0.352	0.499	0.414	0.171	0.416
$\theta_{c1} = 0.2$	0.183	0.141	0.020	0.141	0.203	0.157	0.025	0.139
$\theta_{c2} = 0.3$	0.296	0.140	0.020	0.150	0.309	0.154	0.024	0.141
$\rho_{11} = 0.4$	0.388	0.086	0.008	0.081	0.401	0.087	0.008	-
$\rho_{21} = -0.2$	-0.195	0.079	0.006	0.071	-0.198	0.077	0.006	-
$\rho_{12} = -0.2$	-0.183	0.076	0.006	0.074	-0.198	0.072	0.005	-
$\rho_{22} = 0.4$	0.400	0.090	0.008	0.081	0.398	0.082	0.007	-

Table 2.6: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D6 with sample size n=200.

	JGQL				SSGQL			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.410	0.241	0.058	0.231	0.407	0.240	0.057	0.234
$\alpha_2 = 0.3$	0.329	0.210	0.045	0.235	0.321	0.237	0.057	0.238
$\theta_{z1} = 0.25$	0.263	0.347	0.121	0.311	0.261	0.392	0.154	0.392
$\theta_{z2} = 0.35$	0.342	0.347	0.121	0.313	0.344	0.390	0.152	0.396
$\beta_1 = 0.35$	0.371	0.267	0.072	0.279	0.361	0.326	0.106	0.340
$\beta_2 = 0.25$	0.287	0.288	0.084	0.284	0.262	0.334	0.112	0.347
$\theta_{y1} = 0.4$	0.380	0.344	0.119	0.318	0.412	0.409	0.167	0.417
$\theta_{y2} = 0.5$	0.484	0.363	0.132	0.324	0.531	0.404	0.164	0.423
$\theta_{cz1} = 0.2$	0.222	0.223	0.050	0.190	0.209	0.192	0.037	0.190
$\theta_{cz2} = 0.3$	0.328	0.208	0.044	0.192	0.322	0.204	0.042	0.193
$\theta_{cy1} = 0.35$	0.306	0.266	0.072	0.159	0.368	0.209	0.044	0.202
$\theta_{cy2} = 0.45$	0.412	0.208	0.045	0.167	0.473	0.213	0.046	0.203
$\rho_{11} = 0.4$	0.381	0.080	0.007	0.077	0.400	0.080	0.006	-
$\rho_{21} = 0.2$	0.175	0.079	0.007	0.077	0.198	0.078	0.006	-
$\rho_{12} = 0.15$	0.153	0.081	0.007	0.076	0.145	0.081	0.007	-
$\rho_{22} = 0.35$	0.361	0.085	0.007	0.079	0.347	0.083	0.007	-

Table 2.7: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D7 with sample size n=200.

	JGQL				SSGQL			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.413	0.220	0.049	0.226	0.413	0.237	0.056	0.238
$\alpha_2 = 0.3$	0.317	0.237	0.056	0.233	0.319	0.250	0.063	0.242
$\theta_{z1} = 0.25$	0.260	0.421	0.178	0.389	0.259	0.409	0.167	0.397
$\theta_{z2} = 0.35$	0.387	0.414	0.173	0.398	0.356	0.394	0.156	0.399
$\beta_1 = 0.35$	0.348	0.310	0.096	0.295	0.339	0.327	0.107	0.333
$\beta_2 = 0.25$	0.263	0.313	0.098	0.303	0.252	0.324	0.105	0.340
$\theta_{y1} = 0.4$	0.410	0.371	0.137	0.367	0.441	0.418	0.177	0.411
$\theta_{y2} = 0.5$	0.489	0.369	0.137	0.376	0.536	0.413	0.172	0.417
$\theta_{cz1} = 0.2$	0.198	0.196	0.038	0.189	0.202	0.202	0.041	0.192
$\theta_{cz2} = 0.3$	0.298	0.188	0.035	0.192	0.313	0.195	0.038	0.195
$\theta_{cy1} = 0.35$	0.360	0.217	0.047	0.181	0.388	0.199	0.041	0.201
$\theta_{cy2} = 0.45$	0.448	0.210	0.044	0.185	0.487	0.207	0.044	0.203
$\rho_{11} = 0.2$	0.203	0.091	0.008	0.086	0.201	0.089	0.008	-
$\rho_{21} = 0.0$	-0.004	0.084	0.007	0.085	-0.006	0.083	0.007	-
$\rho_{12} = 0.0$	-0.002	0.084	0.007	0.085	0.000	0.087	0.007	-
$\rho_{22} = 0.1$	0.113	0.087	0.008	0.088	0.108	0.094	0.009	-

Table 2.8: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and SSGQL estimates under D8 with sample size n=200.

	JGQL				SSGQL			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.354	0.228	0.054	0.234	0.393	0.222	0.049	0.236
$\alpha_2 = 0.3$	0.292	0.220	0.048	0.237	0.294	0.224	0.050	0.241
$\theta_{z1} = 0.25$	0.271	0.350	0.123	0.368	0.255	0.355	0.126	0.390
$\theta_{z2} = 0.35$	0.327	0.368	0.136	0.371	0.355	0.368	0.136	0.394
$\beta_1 = 0.35$	0.385	0.317	0.102	0.300	0.327	0.332	0.111	0.316
$\beta_2 = 0.25$	0.309	0.317	0.104	0.298	0.257	0.342	0.117	0.321
$\theta_{y1} = 0.4$	0.337	0.382	0.149	0.356	0.437	0.409	0.169	0.397
$\theta_{y2} = 0.5$	0.446	0.379	0.147	0.355	0.523	0.424	0.180	0.402
$\theta_{cz1} = 0.2$	0.211	0.189	0.036	0.187	0.202	0.178	0.032	0.189
$\theta_{cz2} = 0.3$	0.318	0.182	0.033	0.189	0.307	0.192	0.037	0.193
$\theta_{cy1} = 0.35$	0.346	0.194	0.038	0.176	0.351	0.198	0.039	0.197
$\theta_{cy2} = 0.45$	0.438	0.198	0.040	0.177	0.443	0.197	0.039	0.198
$\rho_{11} = 0.3$	0.290	0.095	0.009	0.084	0.305	0.085	0.007	-
$\rho_{21} = -0.2$	-0.199	0.084	0.007	0.077	-0.195	0.074	0.006	-
$\rho_{12} = -0.2$	-0.189	0.080	0.007	0.076	-0.202	0.081	0.007	-
$\rho_{22} = 0.3$	0.301	0.088	0.008	0.085	0.301	0.086	0.007	-

Figure 2.1: MSE comparison of regression parameters between JGQL and SSGQL under D1 for n=200, 300 and 1000.

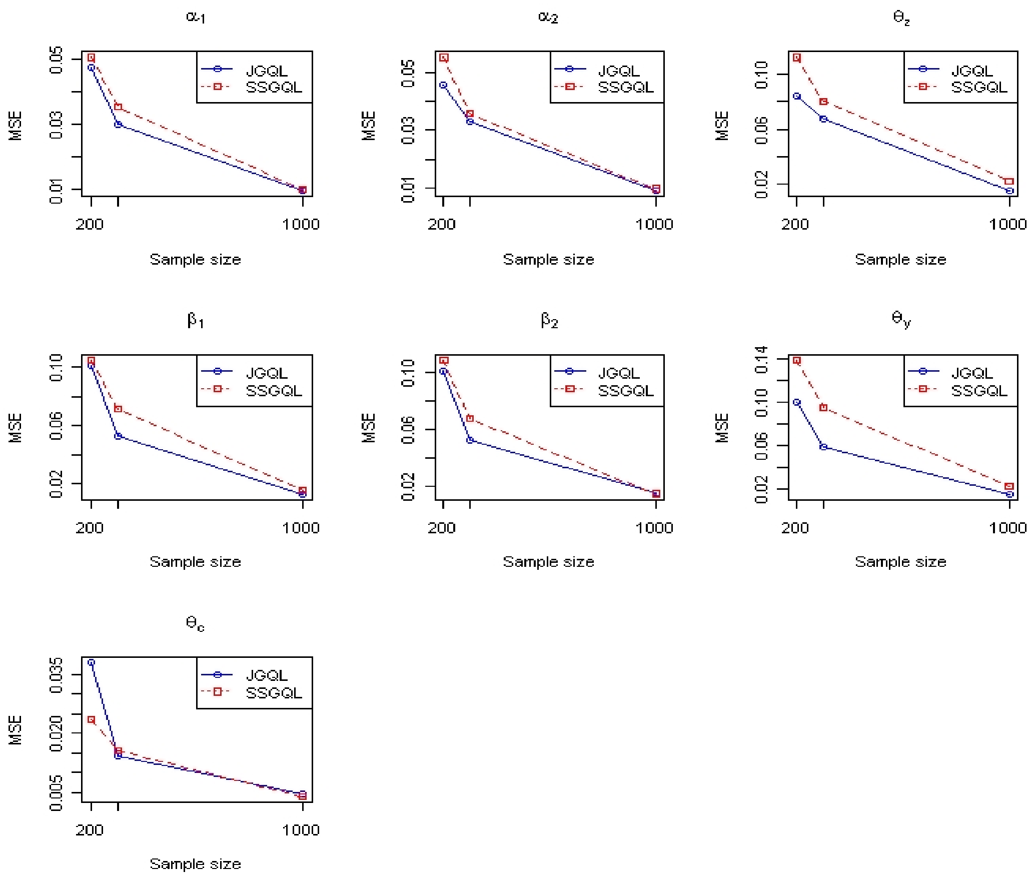
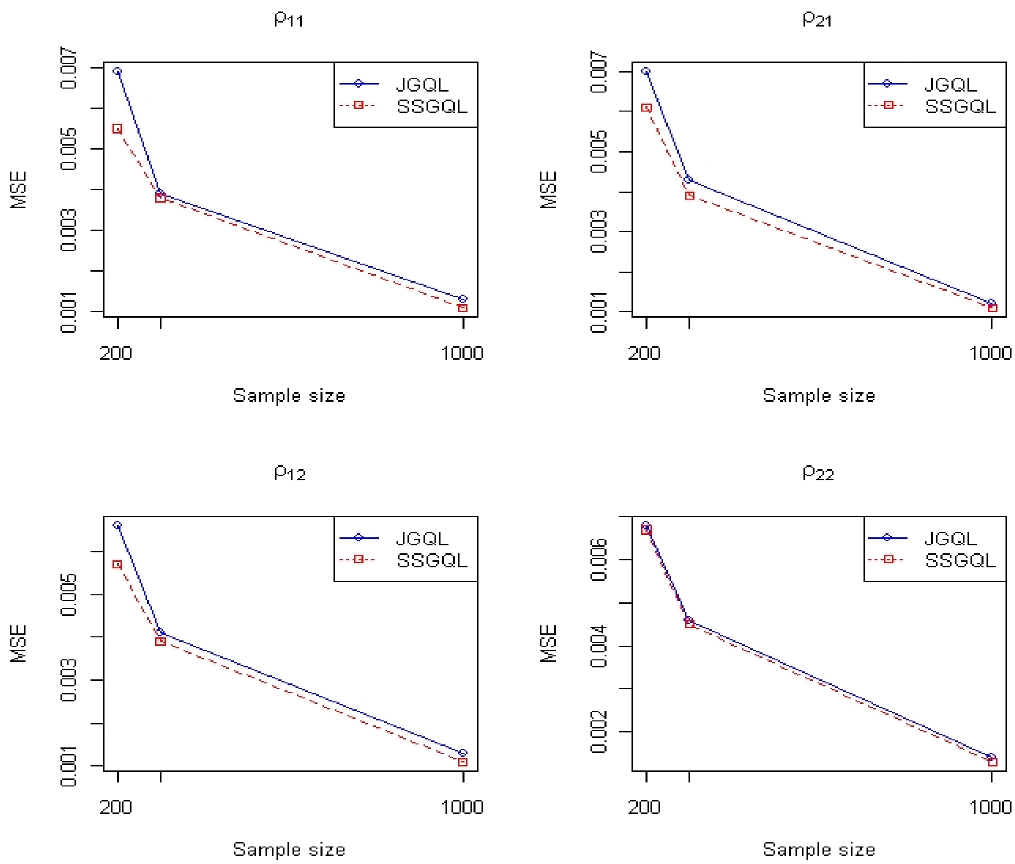


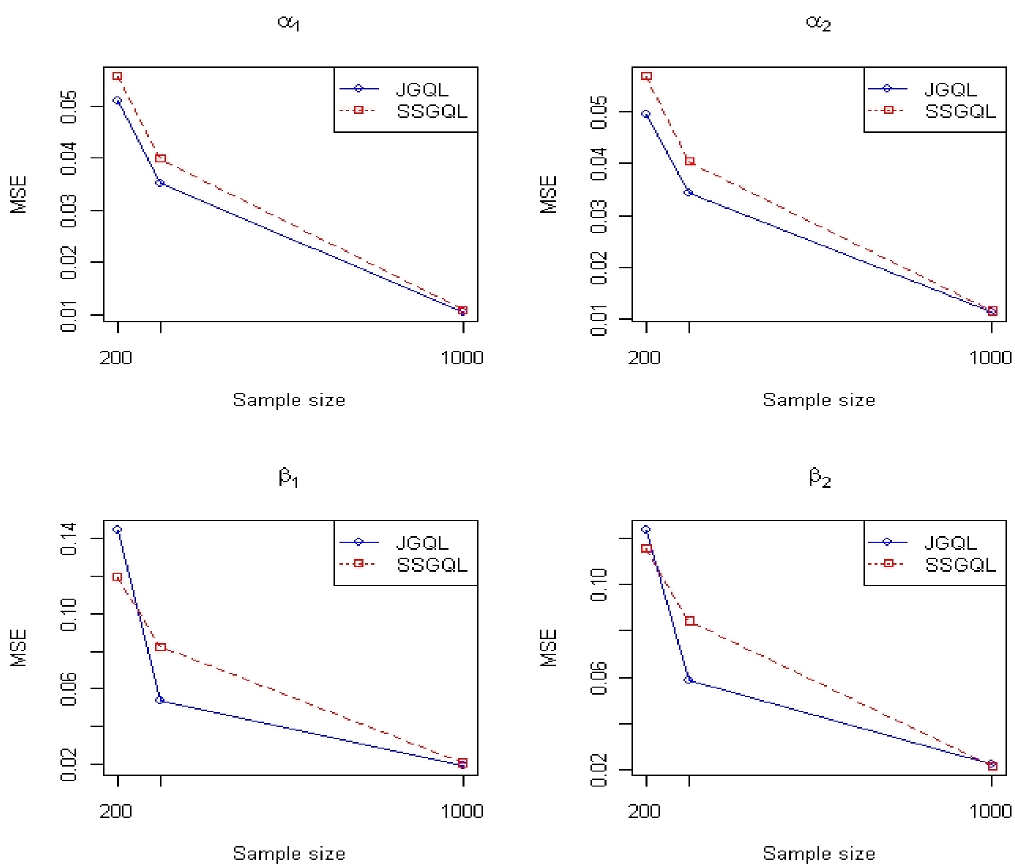
Figure 2.2: MSE comparison of correlation index parameters between JGQL and SSGQL under D1 for n=200, 300 and 1000.



2.2.4 Diabetic retinopathy data analysis

In the last section, we demonstrated through an intensive simulation study that the JGQL and SSGQL approaches perform very well in estimating the effects of the associated covariates on the bivariate multinomial responses. In this section, we illustrate an application of these inference techniques by fitting the proposed LCBM model to the so-call WESDR (Wisconsin Epidemiologic Study of Diabetic Retinopathy), which was analyzed earlier by some authors such as Williamson et al. (1995).

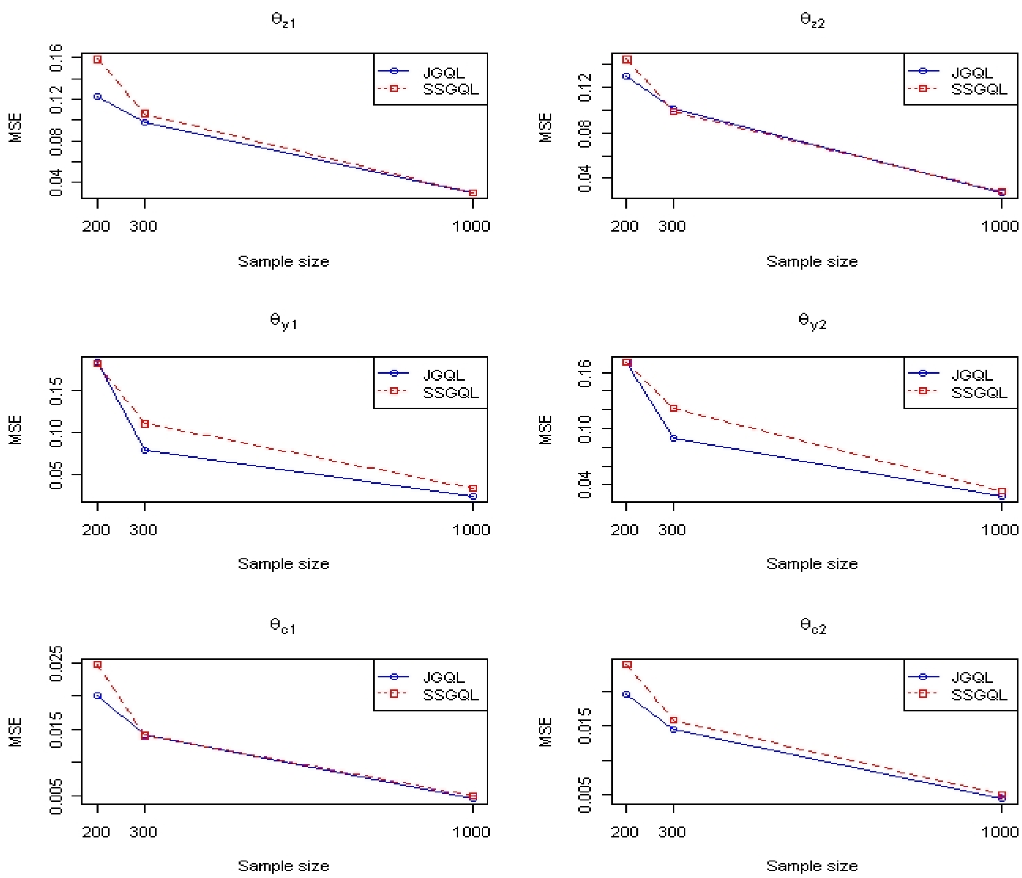
Figure 2.3: MSE comparison of category regression parameters between JGQL and SSGQL under D7 for n=200, 300 and 1000.



We explain the WESDR data set in brief as follows. This data set contains diabetic retinopathy status on a ten point interval scale for left and right eyes of 996 independent patients, along with information on various associated covariates. Some of the important covariates are: (1) duration of diabetes (DD), (2) glycosylated hemoglobin level (GHL), (3) diastolic blood pressure (DBP), (4) gender, (5) proteinuria (Pr), (6) dose of insulin per day (DI), and (7) macular edema (ME). There are 743 subjects with complete response and covariate data.

Because the bivariate responses from an individual are supposed to be correlated,

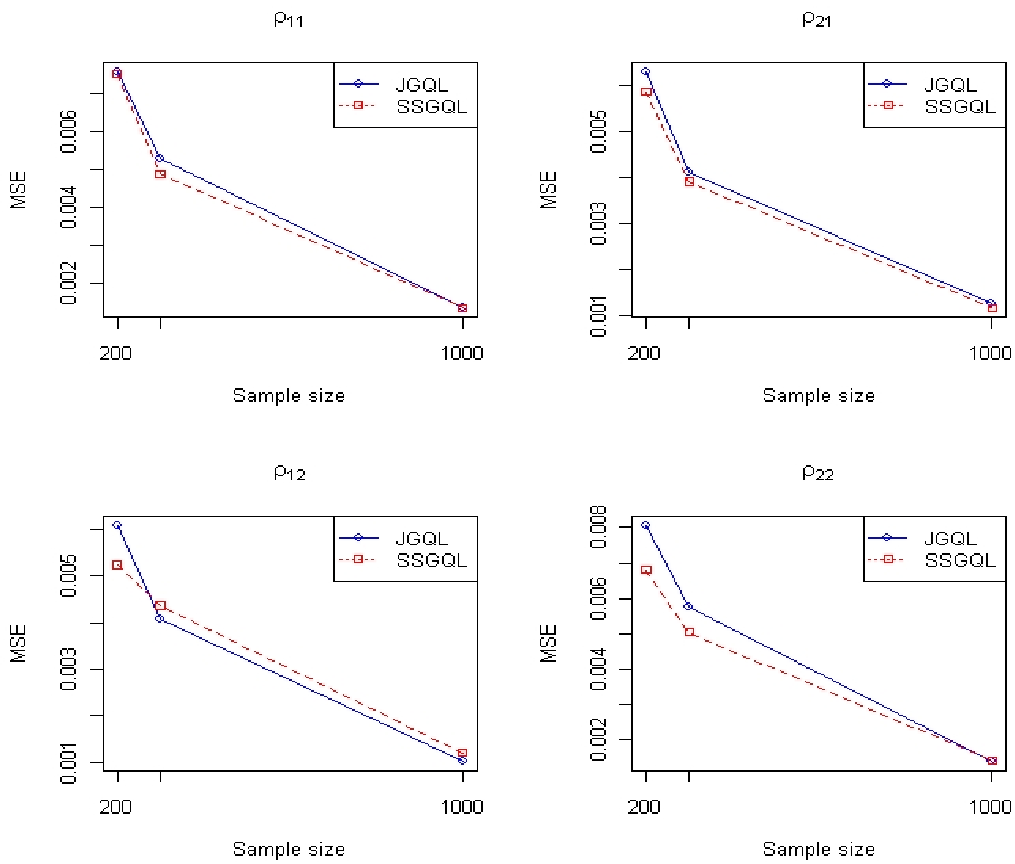
Figure 2.4: MSE comparison of covariate regression parameters between JGQL and SSGQL under D7 for n=200, 300 and 1000.



it is of interest to accommodate bivariate correlations and examine the effects of these covariates on the bivariate responses. Williamson et. al. (1995) used four categories, namely, none, mild, moderate, and proliferative, and exploited an odds ratio approach to accommodate the correlations and they used estimating equations approach to compute the effects of the covariates.

Note however that as mentioned in Section 1.2, this odds ratio approach use an extra arbitrary or 'working' regression relationship to model the association through odds ratios. To avoid this arbitrariness in modeling correlations, we have used the

Figure 2.5: MSE comparison of correlation index parameters between JGQL and SSGQL under D7 for n=200, 300 and 1000.



LCBM model in this chapter. We now fit the model to the DR data by estimating the parameters through proposed JGQL and SSGQL approaches.

2.2.4.1 An application of the linear conditional bivariate binary (LCBB) fixed model

In this section, for simplicity, we collapsed the four categories of left and eye diabetic retinopathy status in Williamson et. al. (1995) into 2 categories for each of the bivariate responses. The DR responses in the bivariate binary format is shown in

Table 2.9.

Table 2.9: Descriptive statistics of left and right eyes diabetic retinopathy status.

right eye \ left eye	Y=1 (presence of DR)	Y=0 (absence of DR)	Total
Z=1 (presence of DR)	424	31	455
Z=0 (absence of DR)	39	249	288
Total	463	280	743

As far as the covariates are concerned, we denote the 7 covariates as follows. First, we categorize duration of diabetes (DD) into three categories, to do so we use two dummy covariates x_{i11} and x_{i12} defined as follows:

$$(x_{i11}, x_{i12}) = \begin{cases} (1,0), & DD \ < 5 \ years \\ (0,0), & DD \ between \ 5 \ and \ 10 \ years \\ (0,1), & DD \ > 10 \ years. \end{cases}$$

The other six covariates are denoted as:

$$x_{i2} = \frac{GHL_i - \overline{GHL}}{se(GHL)}, \ x_{i3} = \begin{cases} 0, & DBP \ < 80 \\ 1, & DBP \ \geq 80, \end{cases} \quad x_{i4} = \begin{cases} 0, & male \\ 1, & female, \end{cases}$$

$$x_{i5} = \begin{cases} 0, & Pr \ absence \\ 1, & Pr \ presence, \end{cases} \quad x_{i6} = \begin{cases} 0, & DI \leq 1 \\ 1, & DI \equiv 2, \end{cases} \quad x_{i7} = \begin{cases} 0, & ME \ absence \\ 1, & ME \ presence. \end{cases}$$

For convenience we now use $x_i = (x_{i11}, x_{i12}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}, x_{i7})'$ to represent all 7 covariates, and $\theta = (\theta_{11}, \theta_{12}, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6, \theta_7)'$ to represent the effects of x_i on the binary response variables y_i and z_i. Note that in addition to x_i, the probabilities for the response variables z_i and y_i are functions of marginal parameters α_1 and β_1, respectively. Thus, following Section 2.1, we now spell out the linear conditional bivariate binary (LCBB) model relating y and z with x as follows,

LCBB Model:

$$\begin{aligned} \pi_{iz} &= Pr(z_i = 1) = \frac{\exp(\alpha_1 + x_i'\theta)}{1 + \exp(\alpha_1 + x_i'\theta)}, \\ \pi_{iy} &= Pr(y_i = 1) = \frac{\exp(\beta_1 + x_i'\theta)}{1 + \exp(\beta_1 + x_i'\theta)}, \\ \text{and } \eta_{i1|k}^{(y)} &= Pr(y_i = 1|z_i, x_i) = \pi_{iy} + \rho_{11}(z_i - \pi_{iz}), k = 0, 1. \end{aligned} \tag{2.19}$$

The JGQL and SSGQL estimates of all model parameters and their estimated standard errors (ESE) are reported in Table 2.10. Note that following (2.13) the estimated standard errors for $\hat{\phi} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\theta}, \hat{\rho}_{11})'$ under the JGQL approach were computed from the covariance matrix of $\hat{\phi}$ given by

$$Cov(\hat{\phi}) = \left\{ \sum_{i=1}^{n} \left[\frac{\partial(\Pi_{iz}', \Pi_{iy}', \Pi_{izy}')}{\partial\phi} \right] \Sigma_i^{-1} \left[\frac{\partial(\Pi_{iz}', \Pi_{iy}', \Pi_{izy}')}{\partial\phi} \right]' \right\}^{-1}. \tag{2.20}$$

Similarly, following (2.14), the estimated standard error of $\hat{\psi} = (\hat{\alpha}_1, \hat{\beta}_1, \hat{\theta})'$ under the SSGQL approach were computed from the covariance matrix of $\hat{\psi}$ given by

$$Cov(\hat{\psi}) = \left\{ \sum_{i=1}^{n} \left[\frac{\partial(\Pi_{iz}', \Pi_{iy}')}{\partial\psi} \right] \Sigma_{i11}^{-1} \left[\frac{\partial(\Pi_{iz}', \Pi_{iy}')}{\partial\psi} \right]' \right\}^{-1}. \tag{2.21}$$

From Table 2.10 we can see that the JGQL estimates are very close to the SSGQL estimates. The simulation results reported in Tables 2.1-2.8 appear to support this closeness. Next, when ESEs are compared, it is clear that the ESEs of the JGQL estimates are smaller than the corresponding SSGQL estimates, which is also supported by simulation results shown in Tables 2.1-2.8 and Figures 2.1-2.5.

The results in Table 2.10 show that the propensity of diabetic retinopathy (probability of having diabetic retinopathy problem) tends to increase with longer DD,

Table 2.10: JGQL and SSGQL estimation results for the diabetic retinopathy data under the LCBB model.

Approach	JGQL		SSGQL	
Parameter (Effect of)	Estimate	ESE	Estimate	ESE
α_1	-0.3166	0.1974	-0.3203	0.2005
β_1	-0.2146	0.1968	-0.2379	0.2003
θ_{11} (DD low)	-2.0402	0.2741	-2.1187	0.2867
θ_{12} (DD high)	2.2349	0.2064	2.2376	0.2096
θ_2 (GHL)	0.3871	0.0925	0.4168	0.0951
θ_3 (DBP)	0.5729	0.1889	0.5538	0.1926
θ_4 (Gender)	-0.2491	0.1829	-0.2297	0.1867
θ_5 (Pr)	0.5271	0.3208	0.5099	0.3274
θ_6 (DI)	0.0026	0.1835	0.0177	0.1874
θ_7 (ME)	2.0638	1.0428	2.6025	1.3779
ρ_{11}	0.6372	0.0393	0.6361	-

higher GHL, higher DBP, male gender, presence of Pr, more DI per day and presence of ME. Note that the estimates of effects of DD and ME are found to deviate from zero clearly, indicating that these two covariates are important risk factors of diabetic retinopathy problem. To be specific, (1) the marginal parameter estimates $\hat{\alpha}_{1,JGQL} = -0.3166$ and $\hat{\beta}_{1,JGQL} = -0.2146$ indicate that when other covariates are fixed, an individual has small probabilities to develop left and right eye retinopathy problem. Next, because DD was coded as $(0, 0)$ for duration between 5 and 10 years, the large positive value of $\hat{\theta}_{12,JGQL} = 2.2349$ and negative value of $\hat{\theta}_{11,JGQL} = -2.0402$ show that as DD increases, the probability of an individual to have retinopathy problem increases. (3) The positive values of $\hat{\theta}_{2,JGQL} = 0.3871$ and $\hat{\theta}_{3,JGQL} = 0.5729$, indicate that an individual with high GHL and DBP has greater probability to have retinopathy problem given the other covariates fixed, respectively. (4) The negative value of $\hat{\theta}_{4,JGQL} = -0.2491$ indicate that males are more likely to develop retinopathy problem compared with females. Next, $\hat{\theta}_{5,JGQL} = 0.5271$ show that presence of Pr (proteinuria) increases one's probability to develop retinopathy compared with those who don't have Pr problem. (6) The small values of $\hat{\theta}_6$ under both approaches, to be specific, $\hat{\theta}_{6,SSGQL} = 0.0177$, indicate that dose of insulin per day (DI) does not have much influence on one's propensity to have retinopathy problem. (7) The regression effect of ME (macular edema) on the probability of having diabetic retinopathy in left or right eye was found to be $\hat{\theta}_{7,SSGQL} = 2.60$. Because ME was coded as $x_7 = 1$ in the presence of ME, this high positive value $\hat{\theta}_{7,SSGQL} = 2.60$ indicates that ME has great effects on the retinopathy status.

Next, the correlation index parameter $\hat{\rho}_{11,JGQL} = 0.6372$ ($\hat{\rho}_{11,SSGQL} = 0.6361$) implies that right eye retinopathy status is highly correlated with the retinopathy status of left eye. This high correlation appears to reflect well the correlation indicated by the observations in Table 2.9. Note that this correlation value was accommodated in obtaining the above efficient regression estimates under both approaches. Note that the aforementioned regression estimates (effects of risk factors) also agree with the corresponding estimates obtained by Williamson et al. (1995, Table 2), for example. However, in our approach, unlike Williamson et al. (1995), we do not require any extra modeling for any association such as their odds ratio parameters. By the same token, we have avoided the odds ratio based models used by Agresti (2002) which complicates the marginal probabilities involving marginal parameters of both variables. In contrary, in fitting the LCBB model to the data we easily computed the correlation parameter $\hat{\rho}_{11}$ which appears to be quite high under both estimation approaches. Thus, the present approach explains both marginal effects and correlations relatively easily.

Now because regression effects and correlation index parameter are estimated, we can examine the bivariate correlation pattern by computing the individual correlations which will be functions of ψ and ρ_{11}. To be specific, Recall from (2.5) that $\rho_{iyz} = corr(y_i, z_i) = \rho_{11}\sqrt{\frac{\pi_{iz}(1-\pi_{iz})}{\pi_{iy}(1-\pi_{iy})}}$. Now, by using the model parameter estimates given in Table 2.10, we can calculate the correlation ρ_{iyz} for each $i = 1, \ldots, n$. This we do by using the SSGQL estimates. We give the histogram of correlations in Figure 2.6 below. From Figure 2.6 we can see that a large number of individuals have big correlations such as 0.66 or even higher. To be precise, the minimum of ρ_{iyz} is found to be 0.6120, and the maximum is 0.6628, with average of ρ_{iyz} given by $\bar{\rho}_{yz} = 0.6426$. Thus, the present model helps to understand the correlation between bivariate multinomial data, the binary data being the special case.

2.2.4.2 An application of the linear conditional bivariate multinomial (LCBM) model

In the last section, the LCBB model was fitted to the bivariate binary diabetic retinopathy (DR) data set and it was found that duration of diabetes (DD) and macular edema (ME), among other covariates, have played an important role on dichotomous diabetic retinopathy status. Note that when DR is present, it may be

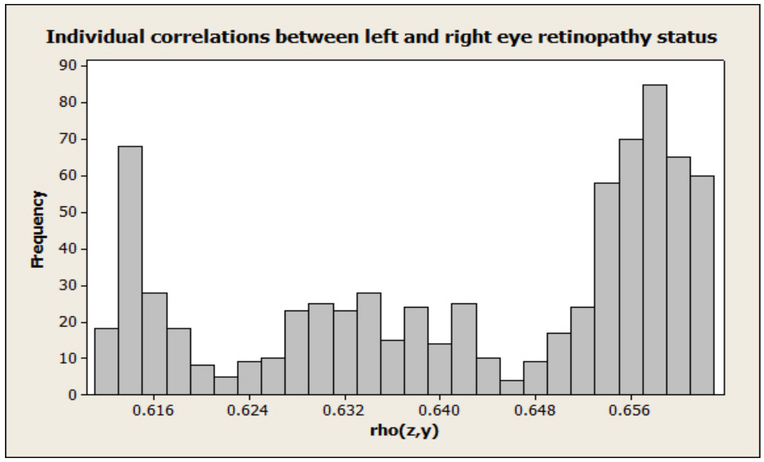


Figure 2.6: Histogram of correlations between left and right eye binary retinopathy status under the LCBB model.

useful to understand the effects of associated covariates on various levels of DR. To be specific, when DR is present, it may be however severe or non-severe. For this reason, in this section, we categorize the diabetic retinopathy status into three categories, namely, absence of DR, non-severe DR, and severe DR; and fit the LCBM model to this data set by applying the SSGQL method to examine the effects of selected covariates on DR. To represent three categories of right eye diabetic retinopathy status, we use two dummy variables z_{i1} and z_{i2} defined as follows:

$$(z_{i1}, z_{i2}) = \begin{cases} (1,0), \ non-severe\ DR\ (category\ 1) \\ (0,1), \ severe\ DR\ (category\ 2) \\ (0,0), \ absence\ of\ DR\ (category\ 3). \end{cases}$$

Similarly, we use two dummy variables y_{i1} and y_{i2} to represent the three categories of left eye diabetic retinopathy status as follows:

$$(y_{i1}, y_{i2}) = \begin{cases} (1,0), \ non - severe \ DR \ (category \ 1) \\ (0,1), \ severe \ of \ DR \ (category \ 2) \\ (0,0), \ absence \ of \ DR \ (category \ 3). \end{cases}$$

The distribution of the 743 individuals under 3 categories of each of y and z are shown in Table 2.11.

Table 2.11: Descriptive statistics of left and right eyes diabetic retinopathy status.

right eye \ left eye	non-severe DR	severe DR	absence of DR	Total
non-severe DR	354	15	31	400
severe DR	12	43	0	55
absence of DR	39	0	249	288
Total	405	58	280	743

As far as the covariates are concerned, in the bivariate binary analysis in the last section, we consider 7 covariates. However, one of the covariates, namely, dose of insulin per day (DI) was found to have no obvious effect on DR evident from the JGQL and SSGQL estimates for this effect, which were found to be $\hat{\theta}_{6,JGQL} = 0.0026$ and $\hat{\theta}_{6,SSGQL} = 0.0177$. Thus, we do not include DI in the present multinomial analysis. The rest of the covariates are: (1) duration of diabetes (DD), (2) glycosylated hemoglobin level (GHL), (3) diastolic blood pressure (DBP), (4) gender, (5) proteinuria (Pr), and (6) macular edema (ME); and it is of interest to find the effects of the 6 covariates on the trinomial status of DR. Furthermore, unlike in the previous section, in this section, we use standardized DD to estimate the effect of DD on DR. There are two obvious advantages of doing so, first the total number of model parameters can be reduced by two, yielding simpler calculations; second it is easier to interpret effects of DD on different categories of DR. We give the formula for standardizing DD as follows:

$$x_{i1} = \frac{DD_i - \overline{DD}}{se(DD)}.$$

Next, to specify the bivariate multinomial probabilities following (2.1)-(2.2), we use the notation $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6})'$ to represent aforementioned 6 covariates, and use $\theta_1 = (\theta_{11}, \theta_{21}, \theta_{31}, \theta_{41}, \theta_{51}, \theta_{61})'$ to represent the effects of x_i on the response variables y_{i1} and z_{i1}, and $\theta_2 = (\theta_{12}, \theta_{22}, \theta_{32}, \theta_{42}, \theta_{52}, \theta_{62})'$ to represent the effects of x_i on the response variables y_{i2} and z_{i2}. For example, θ_{11} is the effect of DD on non-severe DR, and θ_{12} represent the effect of DD on severe retinopathy problem. Note that in addition to x_i, the probabilities for the response variables z_{i1} and z_{i2} are functions of marginal parameters α_1 and α_2, respectively; similarly, the probabilities for the response variables y_{i1} and y_{i2} are functions of marginal parameters β_1 and β_2, respectively. Following (2.1)-(2.2), we now spell out the linear conditional bivariate binary (LCBM) fixed model relating y and z with x as follows,

LCBM Model:

$$\begin{aligned}
\pi_{iz}^{(1)} &= Pr(z_i = z_i^{(1)}) = \frac{\exp(\alpha_1 + x_i'\theta_1)}{1 + \exp(\alpha_1 + x_i'\theta_1) + \exp(\alpha_2 + x_i'\theta_2)}, \\
\pi_{iz}^{(2)} &= Pr(z_i = z_i^{(2)})) = \frac{\exp(\alpha_2 + x_i'\theta_2)}{1 + \exp(\alpha_1 + x_i'\theta_1) + \exp(\alpha_2 + x_i'\theta_2)}, \qquad (2.22) \\
\pi_{iz}^{(3)} &= Pr(z_i = z_i^{(3)}) = 1 - \pi_{iz}^{(1)} - \pi_{iz}^{(2)},
\end{aligned}$$

$$
\begin{aligned}
\pi_{iy}^{(1)} &= Pr(y_i = y_i^{(1)}) = \frac{\exp(\beta_1 + x_i'\theta_1)}{1 + \exp(\beta_1 + x_i'\theta_1) + \exp(\beta_2 + x_i'\theta_2)}, \\
\pi_{iy}^{(2)} &= Pr(y_i = y_i^{(2)}) = \frac{\exp(\beta_2 + x_i'\theta_2)}{1 + \exp(\beta_1 + x_i'\theta_1) + \exp(\beta_2 + x_i'\theta_2)}, \\
\pi_{iy}^{(3)} &= Pr(y_i = y_i^{(3)}) = 1 - \pi_{iy}^{(1)} - \pi_{iy}^{(2)};
\end{aligned} \tag{2.23}
$$

and

$$
\begin{aligned}
\eta_{i1|1}^{(y)} &= Pr(y_i = y_i^{(1)} | z_i = z_i^{(1)}) = \pi_{iy}^{(1)} + \rho_{11}(1 - \pi_{iz}^{(1)}) - \rho_{21}\pi_{iz}^{(2)}, \\
\eta_{i2|1}^{(y)} &= Pr(y_i = y_i^{(2)} | z_i = z_i^{(1)}) = \pi_{iy}^{(2)} + \rho_{12}(1 - \pi_{iz}^{(1)}) - \rho_{22}\pi_{iz}^{(2)}, \\
\eta_{i3|1}^{(y)} &= Pr(y_i = y_i^{(3)} | z_i = z_i^{(1)}) = 1 - \eta_{i1|1}^{(y)} - \eta_{i2|1}^{(y)}, \\
\eta_{i1|2}^{(y)} &= Pr(y_i = y_i^{(1)} | z_i = z_i^{(2)}) = \pi_{iy}^{(1)} - \rho_{11}\pi_{iz}^{(1)} + \rho_{21}(1 - \pi_{iz}^{(2)}), \\
\eta_{i2|2}^{(y)} &= Pr(y_i = y_i^{(2)} | z_i = z_i^{(2)}) = \pi_{iy}^{(2)} - \rho_{12}\pi_{iz}^{(1)} + \rho_{22}(1 - \pi_{iz}^{(2)}), \\
\eta_{i3|2}^{(y)} &= Pr(y_i = y_i^{(3)} | z_i = z_i^{(2)}) = 1 - \eta_{i1|1}^{(y)} - \eta_{i2|1}^{(y)}, \\
\eta_{i1|3}^{(y)} &= Pr(y_i = y_i^{(1)} | z_i = z_i^{(3)}) = \pi_{iy}^{(1)} - \rho_{11}\pi_{iz}^{(1)} - \rho_{21}\pi_{iz}^{(2)}, \\
\eta_{i2|3}^{(y)} &= Pr(y_i = y_i^{(2)} | z_i = z_i^{(3)}) = \pi_{iy}^{(2)} - \rho_{12}\pi_{iz}^{(1)} - \rho_{22}\pi_{iz}^{(2)}, \\
\eta_{i3|3}^{(y)} &= Pr(y_i = y_i^{(3)} | z_i = z_i^{(3)}) = 1 - \eta_{i1|1}^{(y)} - \eta_{i2|1}^{(y)}.
\end{aligned} \tag{2.24}
$$

The SSGQL estimates of all model parameters and the estimated standard errors (ESE) of all regression parameters (α_1, α_2, β_1, β_2, θ_1 and θ_2) are reported in Table 2.12.

The results in Table 2.12 show that the propensity of diabetic retinopathy (probability of having diabetic retinopathy problem) tends to increase with longer DD, higher GHL, higher DBP, male gender, presence of proteinuria, and presence of ME. This observation agrees with the results in Table 2.10 under the bivariate binary analysis. To be specific, (1) the marginal parameter estimates $\hat{\alpha}_{1,SSGQL} = 0.6817$ and

$\hat{\alpha}_{2,SSGQL} = -2.5275$, along with the marginal parameter estimates $\hat{\beta}_{1,SSGQL} = 0.7531$ and $\hat{\beta}_{1,SSGQL} = -2.3879$, indicate that when other covariates are fixed, an diabetic patient tends to develop retinopathy problem. However, the probability to have moderate (non-severe) retinopathy problem is larger as compared to the probability of having severe retinopathy problem. This observation agrees with the descriptive statistics in Table 2.11. (2) The large positive value of $\hat{\theta}_{11,SSGQL} = 2.1768$ and $\hat{\theta}_{12,SSGQL} = 2.5909$ show that as DD increases, the probability of an individual to have retinopathy problem increases, the longer DD, the severer retinopathy status will be. (3) The positive values of $\hat{\theta}_{31,SSGQL} = 0.6728$ and $\hat{\theta}_{32,SSGQL} = 1.1458$ indicate that an individual with higher DBP has greater probability to have retinopathy problem given the other covariates fixed. The positive values of $\hat{\theta}_{21}$ and $\hat{\theta}_{22}$ give similar interpretation of the effects of GHL on one's retinopathy status. (4) The negative value of $\hat{\theta}_{41,SSGQL} = -0.1899$ and $\hat{\theta}_{42,SSGQL} = -0.3735$ indicate that males are more likely to develop retinopathy problem as compared to females, and males are more likely to develop severe retinopathy problem than females. (5) The large positive values of $\hat{\theta}_{61} = 2.0768$ and $\hat{\theta}_{62} = 4.1538$ indicate that ME has a strong influence on one's propensity of diabetic retinopathy, and that presence of ME leads to severe DR more likely than moderate retinopathy problems.

Next, the large correlation index parameter values $\hat{\rho}_{11,SSGQL} = 0.6405$ and $\hat{\rho}_{22,SSGQL} = 0.6740$, and the small values of $\hat{\rho}_{21,SSGQL} = 0.0173$ and $\hat{\rho}_{12,SSGQL} = 0.0086$ imply that right eye retinopathy severity is highly correlated with the retinopathy severity of left eye. For example, for individuals whose left eye retinopathy status is non-severe, it is highly possible for them to have non-severe right eye retinopathy problem. Similarly, for those who have severe left eye retinopathy problem, it is greatly possible for them to have severe right eye retinopathy problem as well. This high correlation appears to reflect well the correlation indicated by the observations in Table 2.11.

Table 2.12: SSGQL estimation results for the diabetic retinopathy data under the LCBM model.

Parameter (Effect of)	Estimate	ESE
α_1	0.6817	0.1473
α_2	-2.5275	0.3120
β_1	0.7531	0.1479
β_2	-2.3879	0.3083
θ_{11} (DD on non-severe DR)	2.1768	0.1412
θ_{12} (DD on severe DR)	2.5909	0.1772
θ_{21} (GHL on non-severe DR)	0.3667	0.0695
θ_{22} (GHL on severe DR)	0.3911	0.1321
θ_{31} (DBP on non-severe DR)	0.6728	0.1415
θ_{32} (DBP on severe DR)	1.1458	0.2868
θ_{41} (Gender on non-severe DR)	-0.1899	0.1383
θ_{42} (Gender on severe DR)	-0.3735	0.2609
θ_{51} (Pr on non-severe DR)	0.5446	0.2454
θ_{52} (Pr on severe DR)	1.7405	0.3348
θ_{61} (ME on non-severe DR)	2.0768	1.0346
θ_{62} (ME on severe DR)	4.1538	1.0504
ρ_{11}	0.6405	-
ρ_{21}	0.0173	-
ρ_{12}	0.0086	-
ρ_{22}	0.6740	-

Chapter 3

Individual Random Effects Based Bivariate Multinomial Mixed (BMM) Model

In the last chapter, we have discussed the LCBM (linear conditional bivariate multinomial) fixed model where the correlations between two multinomial variables were introduced through a conditional linear relationship between two multinomial response variables. There are however situations in practice where the correlations between two responses from the same individual may arise because of the influence of common individual random effect shared by both responses. This type of random effects model produces in general equal-correlations and they are referred to as structural or familial correlations. Familial correlations, specially in the GLMM (generalized linear mixed models) set up [Breslow and Clayton (1993), Lee and Nelder (1996), Sutradhar (2011, Chapter 5)], are usually constructed among the responses of the members of the same family, whereas in the present setup, we develop correlations between bivariate responses as they may share the same individual specific invisible random effect. Thus, in this chapter, we generalize the GLMM for binary data to the multinomial setup for an individual with bivariate multinomial responses. Further note that this type of model, unlike the LCBM fixed model discussed in Chapter 2, does not encounter any restrictions on correlations. This is mainly because in this approach a common random effect is added to the linear predictions involved in the marginal multinomial probabilities for both multinomial variables causing the responses to be correlated. The correlations mainly arise through the variance of the random effects whatever large or small this variance may be. We describe the proposed model in the following section.

3.1 The model

Suppose that the marginal multinomial probabilities for $Z_i = (Z_{i1}, \ldots, Z_{iK-1})'$ and $Y_i = (Y_{i1}, \ldots, Y_{iJ-1})'$ defined in (2.1) and (2.2), respectively, are now influenced by a common random effect $\gamma_i^* \overset{iid}{\sim} N(0, \sigma_\gamma^2)$ associated with the ith individual for all $i = 1, \ldots, n$.

3.1.1 Marginal probabilities conditional on individual specific random effect

As a generalization of the existing GLMM for binary data, in this section we consider similar mixed model but for the multinomial data. To do this, we recall the fully specified marginal multinomial probabilities from (2.1)-(2.2) and introduce random effects to the linear predictors involved in these probabilities. To be specific, let $\gamma_i = \frac{\gamma_i^*}{\sigma_\gamma} \stackrel{iid}{\sim} N(0,1)$, where σ_γ is a common scaler parameter irrespective of the categories. Next, suppose that the marginal probabilities conditional on the individual random effects are given by:

$$
\begin{aligned}
\tilde{\pi}_{iz}^{(k)}(\gamma_i) &= Pr(z_i = z_i^{(k)}|\gamma_i) = \frac{\exp(\alpha_k + x'_{i1}\theta_{k1} + \sigma_\gamma\gamma_i)}{1 + \sum_{q=1}^{K-1}\exp(\alpha_q + x'_{i1}\theta_{q1} + \sigma_\gamma\gamma_i)},\ k = 1, \ldots, K-1, \\
\tilde{\pi}_{iz}^{(K)}(\gamma_i) &= Pr(z_i = z_i^{(K)}|\gamma_i) = \frac{1}{1 + \sum_{q=1}^{K-1}\exp(\alpha_q + x'_{i1}\theta_{q1} + \sigma_\gamma\gamma_i)}; \qquad (3.1)
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\pi}_{iy}^{(j)}(\gamma_i) &= Pr(y_i = y_i^{(j)}|\gamma_i) = \frac{\exp(\beta_j + x'_{i2}\theta_{j2} + \sigma_\gamma\gamma_i)}{1 + \sum_{l=1}^{J-1}\exp(\beta_l + x'_{i2}\theta_{l2} + \sigma_\gamma\gamma_i)},\ j = 1, \ldots, J-1, \\
\tilde{\pi}_{iy}^{(J)}(\gamma_i) &= Pr(y_i = y_i^{(J)}|\gamma_i) = \frac{1}{1 + \sum_{l=1}^{J-1}\exp(\beta_l + x'_{i2}\theta_{l2} + \sigma_\gamma\gamma_i)}. \qquad (3.2)
\end{aligned}
$$

Note that $\sigma_\gamma\gamma_i$ is added to the linear predictor $\alpha_k + x'_{i1}\theta_{k1}$ involved in the probability for $Z_i = Z_i^{(k)}$, and it is added to the linear predictor $\beta_j + x'_{i2}\theta_{j2}$ involved in the probability for $Y_i = Y_i^{(j)}$.

3.1.2 Unconditional moment properties of the model

(a) Unconditional means:

It follows from (3.1)-(3.2) that the unconditional marginal probabilities have the forms

$$
\begin{aligned}
\pi_{iz}^{(k)} = Pr(z_i = z_i^{(k)}) = E_{\gamma_i}E[Z_{ik}|\gamma_i] = E_{\gamma_i}[\tilde{\pi}_{iz}^{(k)}(\gamma_i)|\gamma_i], \qquad (3.3) \\
\text{and } \pi_{iy}^{(j)} = Pr(y_i = y_i^{(j)}) = E_{\gamma_i}E[Y_{ij}|\gamma_i] = E_{\gamma_i}[\tilde{\pi}_{iy}^{(j)}(\gamma_i)|\gamma_i].
\end{aligned}
$$

Note that there is no closed form expressions for these expectations. However, they can be computed empirically.

(b) Unconditional covariance and joint moment:

The unconditional covariance between z_i and y_i may be computed using

$$Cov(Y_i, Z_i') = E(Y_i Z_i') - E(Y_i)E(Z_i'),$$

where $E(Y_i Z_i')$ is computed by

$$E(Z_{ik}Y_{ij}) = \pi_{ikj} = E_{\gamma_i}E[Z_{ik}Y_{ij}|\gamma_i] = E_{\gamma_i}[E(Z_{ik}|\gamma_i)E(Y_{ij}|\gamma_i)]. \tag{3.4}$$

Note that the computation of the above marginal probabilities in (3.3) and the joint probability in (3.4) requires the distribution of γ_i to be known. Under normality assumption (see Breslow and Clayton (1993) for binary case), these moments for example, $\pi_{iy}^{(j)}$ and $E(Y_{ij}Z_{ik})$ can be calculated as

$$\begin{aligned} \pi_{iy}^{(j)} &= Pr(y_i = y_i^{(j)}) = E_{\gamma_i}(\tilde{\pi}_{iy}^{(j)}|\gamma_i) = \int_{-\infty}^{\infty} \tilde{\pi}_{iy}^{(j)} f_N(\gamma_i) d\gamma_i, \\ \text{and } E(Y_{ij}Z_{ik}) &= \int_{-\infty}^{\infty} \tilde{\pi}_{iz}^{(k)} \tilde{\pi}_{iy}^{(j)} f_N(\gamma_i) d\gamma_i, \end{aligned}$$

where $f_N(\gamma_i) = \frac{\exp(\frac{-\gamma_i^2}{2})}{\sqrt{2\pi}}$.

Further note that because σ_γ is involved in all conditional probabilities $(\tilde{\pi}_{iz}^{(k)}(\gamma_i), \tilde{\pi}_{iy}^{(j)}(\gamma_i))$ and unconditional probabilities $(\pi_{iz}^{(k)}, \pi_{iy}^{(j)})$, this parameter (σ_γ) plays a complex role in the correlations between y_i and z_i. However, as this parameter is essential to explain the joint probability for y_i and z_i, it is important that we estimate this parameter. In some situations, the correlations themselves may be of interest for the purpose of data interpretation.

3.1.3 Remarks on similar random effects based models

Note that some authors such as MacDonald (1994) used individual random effects to construct correlation models for repeated binary data. Various scenarios for the distribution of the random effects are considered. This approach appears to be more suitable in the present bivariate multinomial setup as opposed to the univariate longitudinal setup. We however, use normal random effects similar to Breslow and Clayton (1993), for example, and develop the familial correlation model through such random effects. Further note that a binary mixed model similar to (3.1)-(3.2) was used by Ten Have and Morabia (1999) in a familial longitudinal setup. They have used two different random effects for two binary responses to represent possible overdispersion, which however do not cause any familial or structural correlations between the bivariate binary responses at a given time. Thus, it remains as a short fall of these approaches as the familial or structural correlations have to be accommodated. The bivariate association between the two binary responses was modeled through certain additional random effects based odds ratios, but the estimation of the odds ratios requires extra regression modeling as pointed out in Chapter 1, which is a limitation to this approach.

3.2 Inferences for the BMM model

Recall that the LCBM model (2.1)-(2.3) contains regression parameters ψ and the linear dependence parameter ρ^*, whereas the present BMM model (3.1)-(3.2) involves the regression parameter ψ (which has different interpretation than ψ parameter in (2.1)-(2.2)) and the variance of the random effects σ_γ^2. Note that the roles played by the correlation index parameters are, however, different in these models. This is because in the LCBM fixed model, the marginal probabilities are not influenced by the correlation index parameters, whereas in the present BMM model the marginal probabilities are functions of σ_γ, the correlation index parameter. In fact, for this complex role of σ_γ, we find it reasonable to use the JGQL approach as opposed to the SSGQL approach under the LCBM model. Thus, in this section, we use the JGQL approach to estimate all parameters, we will also use the likelihood approach. The estimation performance of these two approaches will be compared through a simulation study. Furthermore, the bivariate binary diabetic retinopathy data will be reanalyzed by fitting the BMM model.

3.2.1 Joint GQL approach

Note that the computation for the marginal mean, variance and covariances to construct the GQL estimating equations under the present BMM model are relatively cumbersome. This is because the moments computation under the BMM model requires an integration (3.3)-(3.4) over the distribution of the random effects γ_i. To be clear and precise, we therefore, write the GQL estimating equations with slightly different notations than in (2.13) for the LCBM fixed model. The GQL estimating equations now have the form

$$f(\phi^*) = \sum_{i=1}^{n} \frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \phi^*} [\bar{\Sigma}_i]^{-1} \begin{pmatrix} z_i - \bar{\Pi}_{iz} \\ y_i - \bar{\Pi}_{iy} \\ g_i - \bar{\Pi}_{izy} \end{pmatrix} = 0, \quad (3.5)$$

where $\phi^* = (\psi', \sigma_\gamma)'$, and by using binomial approximation, for example, for the integration (Ten Have and Morabia (1999)), the probabilities involved in $\bar{\Pi}_{iz}$, $\bar{\Pi}_{iy}$, and $\bar{\Pi}_{izy}$ have the formulas

$$\begin{aligned}
\bar{\pi}_{iz}^{(k)} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iz}^{(k)}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu},\\
\bar{\pi}_{iy}^{(j)} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iy}^{(j)}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu},\\
\text{and } \bar{\pi}_{ikj} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iz}^{(k)}(\gamma_{i\nu})\tilde{\pi}_{iy}^{(j)}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu},
\end{aligned}$$

respectively, for $k = 1, \ldots, K-1$, and $j = 1, \ldots, J-1$, where $\gamma_{i\nu} = \frac{\nu - N(0.5)}{\sqrt{N(0.5)(0.5)}}$, and we use $N = 40$ for the simulation study in Section 3.2.3.

3.2.2 MLE approach

In this section, we discuss the maximum likelihood estimation approach for the BMM model. Given the individual specific random effect γ_i, the two multinomial response variables z_i and y_i are known to be independent, and the conditional likelihood function can be written as follow:

$$L(\phi^*) = \prod_{i=1}^{n} \int_{\gamma_i} \prod_{k=1}^{K} \left[\tilde{\pi}_{iz}^{(k)}(\gamma_i)\right]^{z_{ik}} \prod_{j=1}^{J} \left[\tilde{\pi}_{iy}^{(j)}(\gamma_i)\right]^{y_{ij}} f_N(\gamma_i)d\gamma_i, \tag{3.6}$$

and after some algebras, it reduces to

$$\begin{aligned}
L(\phi^*) &= \exp\left[\sum_{i=1}^{n}\sum_{k=1}^{K-1} z_{ik}(\alpha_k + x'_{i1}\theta_{k1})\right] \times \exp\left[\sum_{i=1}^{n}\sum_{j=1}^{J-1} y_{ij}(\beta_j + x'_{i2}\theta_{j2})\right]\\
&\quad \times \prod_{i=1}^{n}\int_{\gamma_i} \frac{\exp[\sigma_\gamma\gamma_i(\sum_{k=1}^{K-1} z_{ik} + \sum_{j=1}^{J-1} y_{ij})]}{[1+\sum_{q=1}^{K-1}\exp(\alpha_q + x'_{i1}\theta_{q1} + \sigma_\gamma\gamma_i)][1+\sum_{l=1}^{J-1}\exp(\beta_l + x'_{i2}\theta_{l2} + \sigma_\gamma\gamma_i)]} f_N(\gamma_i)d\gamma_i.
\end{aligned}$$

Next, for notational simplicity, by using

$$V_i = \int_{\gamma_i} \exp(\delta_i\gamma_i)u_i(\gamma_i)v_i(\gamma_i)f_N(\gamma_i)d\gamma_i,$$

with $\delta_i = \sigma_\gamma(\sum_{k=1}^{K-1} z_{ik} + \sum_{j=1}^{J-1} y_{ij})$, and $u_i(\gamma_i) = [1 + \sum_{q=1}^{K-1}\exp(\alpha_q + x'_{i1}\theta_{q1} + \sigma_\gamma\gamma_i)]^{-1}$ and $v_i(\gamma_i) = [1+\sum_{l=1}^{J-1}\exp(\beta_l + x'_{i2}\theta_{l2} + \sigma_\gamma\gamma_i)]^{-1}$, the log-likelihood function from (3.6) has the form

$$l(\phi^*) = logL(\phi^*) = \sum_{i=1}^{n}\sum_{k=1}^{K-1} z_{ik}(\alpha_k + x'_{i1}\theta_{k1}) + \sum_{i=1}^{n}\sum_{j=1}^{J-1} y_{ij}(\beta_j + x'_{i2}\theta_{j2}) + \sum_{i=1}^{n} lnV_i,$$

yielding the desired likelihood estimating equations for α_k and θ_{k1} $(k = 1, \ldots, K-1)$ as

$$\frac{\partial l(\phi^*)}{\partial \alpha_k} = \sum_{i=1}^{n} z_{ik} + \sum_{i=1}^{n} \frac{M_{i\alpha_k}}{V_i} = 0, \tag{3.7}$$

$$\frac{\partial l(\phi^*)}{\partial \theta_{k1}} = \sum_{i=1}^{n} z_{ik}x_{i1} + \sum_{i=1}^{n} \frac{M_{i\theta_{k1}}}{V_i} = 0, \tag{3.8}$$

and for β_j and θ_{j2} $(j = 1, \ldots, J-1)$ as

$$\frac{\partial l(\phi^*)}{\partial \beta_j} = \sum_{i=1}^{n} y_{ij} + \sum_{i=1}^{n} \frac{M_{i\beta_j}}{V_i} = 0, \tag{3.9}$$

$$\frac{\partial l(\phi^*)}{\partial \theta_{j2}} = \sum_{i=1}^{n} y_{ij}x_{i2} + \sum_{i=1}^{n} \frac{M_{i\theta_{j2}}}{V_i} = 0, \tag{3.10}$$

and for σ_γ as

$$\frac{\partial l(\phi^*)}{\partial \sigma_\gamma} = \sum_{i=1}^{n} \frac{M_{i\gamma}}{V_i} = 0, \tag{3.11}$$

where, for example, for $k = 1, \ldots, K-1$,

$$M_{i\alpha_k} = \frac{\partial V_i}{\partial \alpha_k} = -\int_{\gamma_i} \exp(\delta_i\gamma_i)u_i(\gamma_i)v_i(\gamma_i)\tilde{\pi}_{iz}^{(k)} f_N(\gamma_i)d\gamma_i,$$

$$M_{i\theta_{k1}} = \frac{\partial V_i}{\partial \theta_{k1}} = -\int_{\gamma_i} \exp(\delta_i\gamma_i)u_i(\gamma_i)v_i(\gamma_i)\tilde{\pi}_{iz}^{(k)} x_{i1} f_N(\gamma_i)d\gamma_i,$$

and for $j = 1, \ldots, J-1$, $M_{i\beta_j}$ and $M_{i\theta_{j2}}$ can be computed similarly. Furthermore, in (3.11,)

$$M_{i\gamma} = \frac{\partial V_i}{\partial \sigma_\gamma} = \int_{\gamma_i} \exp(\delta_i\gamma_i)u_i(\gamma_i)v_i(\gamma_i)\gamma_i[\sum_{k=1}^{K-1} z_{ik} + \sum_{j=1}^{J-1} y_{ij} - (2 - \tilde{\pi}_{iz}^{(K)} - \tilde{\pi}_{iy}^{(J)})] f_N(\gamma_i)d\gamma_i.$$

Let $\theta_1 = (\theta'_{11}, \ldots, \theta'_{k1}, \ldots, \theta'_{K-1,1})'$ and $\theta_2 = (\theta'_{12}, \ldots, \theta'_{j2}, \ldots, \theta'_{J-1,2})'$. The likelihood estimating equation for ϕ^* is given by

$$f(\phi^*) = \Big(\frac{\partial l(\phi^*)}{\partial \alpha_1}, \ldots, \frac{\partial l(\phi^*)}{\partial \alpha_{K-1}}, \frac{\partial l(\phi^*)}{\partial \theta'_1}, \frac{\partial l(\phi^*)}{\partial \beta_1}, \ldots, \frac{\partial l(\phi^*)}{\partial \beta_{J-1}}, \frac{\partial l(\phi^*)}{\partial \theta'_2}, \frac{\partial l(\phi^*)}{\partial \sigma_\gamma}\Big)' = 0. \tag{3.12}$$

Note that the aforementioned likelihood estimating equations involve V_i which requires an integral over the distribution of γ_i. Similar to Section 3.2.1, we approximate this integral by using the binomial approximation technique and use $\bar{V}_i$ for V_i, where

$$\bar{V}_i = \sum_{\nu=0}^{N} \exp(\delta_i \gamma_{i\nu}) u_i(\gamma_{i\nu}) v_i(\gamma_{i\nu}) \binom{N}{\nu} (\frac{1}{2})^{\nu} (1 - \frac{1}{2})^{N-\nu},$$

with $\gamma_{i\nu} = \frac{\nu - N(0.5)}{\sqrt{N(0.5)(0.5)}}$, where $N = 40$ is used for the simulation study in Section 3.2.4 and the diabetic retinopathy data analysis in Section 3.2.5.

3.2.3 Remarks on properties of JGQL and MLE estimates

Once the JGQL and MLE estimates for ϕ^* are found by solving (3.5) and (3.12), respectively, it is important to compute the estimated variances of these estimates.

As far as the asymptotic property of $\hat{\phi}^*_{MLE}$ is concerned, we note that as $n \to \infty$, $\hat{\phi}^*_{MLE}$ converge to ϕ^* in probability (Newey and McFadden (1993)), with the covariance matrix computed from the Fisher information matrix $Cov(\hat{\phi}^*) = -[E(\frac{\partial^2 l(\phi^*)}{\partial \phi^* \partial \phi^{*\prime}})]^{-1}$. However, the computation of this covariance matrix for the likelihood estimators and its estimation is relatively complex because of the involvement of integration over γ_i to compute the second derivatives. In the next section, we rather concentrate on the finite sample performance of the MLE of ϕ^* through simulations. To be more specific, we will examine the relative performance of the MLE approach to the JGQL approach discussed in Section 3.2.1 by comparing the estimates and their standard errors.

Note that unlike the computation for the ML estimators, one may however obtain the asymptotic properties of the JGQL estimators relatively easily. To be specific, to find the asymptotic variance of the JGQL estimates, we first write the Gauss-Newton iterative equation to solve the JGQL estimating equation (3.5). The iterative equation has the form:

$$\begin{aligned}
\hat{\phi}^*_{JGQL,(r+1)} &= \hat{\phi}^*_{JGQL,(r)} + \left[\sum_{i=1}^{n} \frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \phi^*} \bar{\Sigma}_i^{-1} \left(\frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \phi^*}\right)'\right]^{-1}_{(r)} \\
&\quad \times \left[\sum_{i=1}^{n} \frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \phi^*} \bar{\Sigma}_i^{-1} \begin{pmatrix} z_i - \bar{\Pi}_{iz} \\ y_i - \bar{\Pi}_{iy} \\ g_i - \bar{\Pi}_{izy} \end{pmatrix}\right]_{(r)} \\
&= \hat{\phi}^*_{JGQL,(r)} + \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} D_i\right]^{-1}_{(r)} \times \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} (f_i - \xi_i)\right]_{(r)}, (say)
\end{aligned}$$

where $(\ )_r$ denotes that the expression within the square bracket is evaluated at $\phi^* = \hat{\phi}^*_{JGQL,(r)}$, the estimate obtained for the r-th iteration. Note that the iterative convergence to obtain the final JGQL estimates, i.e., $\hat{\phi}^*_{JGQL}$, requires

$$\left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} D_i\right]^{-1} \times \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} (f_i - \xi_i)\right] \to 0$$

in probability. This probability convergence is achieved because $E(f_i) = \xi_i$. This implies that $E(\hat{\phi}^*_{JGQL}) = \phi^*$. The convergence also requires that

$$Var\left(\left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} D_i\right]^{-1} \times \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} (f_i - \xi_i)\right]\right)$$

to be finite, where the variance is given by

$$\begin{aligned}
&\left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} D_i\right]^{-1} \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} Cov(f_i) \bar{\Sigma}_i^{-1} D_i\right] \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} D_i\right]^{-1} \\
&= \left[\sum_{i=1}^{n} D'_i \bar{\Sigma}_i^{-1} D_i\right]^{-1},
\end{aligned}$$

which is also the variance of $\hat{\phi}^*_{JGQL}$. In fact, for $n \to \infty$, by applying Lindeberg-Feller central limit theory (Amemiya (1985), Theorem 3.3.6, p. 92), developed based on non-identical distributions for independent random variables, one may show that $\hat{\phi}^*_{JGQL}$ follows the $p-$dimensional multivariate normal distribution, that is

$$\sqrt{n}(\hat{\phi}^*_{JGQL} - \phi^*) \sim N\left(0, n\left[\sum_{i=1}^{n} \frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \phi^*} \bar{\Sigma}_i^{-1} \left(\frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \phi^*}\right)'\right]^{-1}\right), \tag{3.13}$$

where ϕ^* has p dimensions.

3.2.4 A simulation study

In this section, we fit the proposed BMM model discussed in Section 3.1 to examine the role of common random effects that cause the correlation between two multinomial response variables. Because the random effects variance σ^2_γ is involved in all marginal and joint probabilities, obtaining a reasonable estimate would require large sample size. For this reason, in the simulation study we use $n = 1000$, whereas for the LCBM model in Chapter 2 we use sample size as small as 200, where marginal probabilities were fully specified and free from correlation index parameters.

Simulation design:

Similar to the simulation study for the LCBM model, we consider $K = J = 3$ for the response variables z and y. Also we consider the same marginal parameter values, namely, $\alpha_1 = 0.4$, $\alpha_2 = 0.3$; and $\beta_1 = 0.35$, $\beta_2 = 0.25$. As far as selection of covariates x_{iz}, x_{iy} and x_{ic} is concerned, we use same covariates structure as in the LCBM model based simulation study. That is, we consider x_{i1} and x_{i2} as $x_{i1} = (x_{iz}, x_{ic})'$ and $x_{i2} = (x_{iy}, x_{ic})'$. We choose the covariates design as follows:

$$x_{iz} \sim Binary(0.4),\ x_{iy} \sim Binary(0.7), \text{ and } x_{ic} \sim Standardized\, U(0, 1).$$

Next, we consider a set of regression parameters, namely, $\theta_{z1} = 0.25$, $\theta_{z2} = 0.35$, $\theta_{y1} = 0.4$, $\theta_{y2} = 0.5$, $\theta_{cz1} = \theta_{cy1} = \theta_{c1} = 0.2$, $\theta_{cz2} = \theta_{cy2} = \theta_{c2} = 0.3$. We choose both small and large values for σ_γ, specifically, we use $\sigma_\gamma = 0.1$, 0.35 and 0.5 to reflect small correlations between z_i and y_i; and we use $\sigma_\gamma = 0.75$ and 1.0 to reflect large correlations.

Data generation:

To generate

$$\begin{aligned} z_i &= (z_{i1}, z_{i2}) \\ &= \begin{cases} (1,0), & category\ 1 \\ (0,1), & category\ 2 \\ (0,0), & category\ 3 \end{cases} \end{aligned}$$

and

$$\begin{aligned} y_i &= (y_{i1}, y_{i2}) \\ &= \begin{cases} (1,0), & category\ 1 \\ (0,1), & category\ 2 \\ (0,0), & category\ 3 \end{cases} \end{aligned}$$

for $i = 1, \ldots, 1000$, we first generate γ_i for $i = 1, \ldots, 1000$ from the standard normal distribution, namely, $N(0, 1)$. Note that we have chosen the size $n = 1000$ because in socioeconomic studies the sample size are in general large. We then use $\sigma_\gamma \gamma_i = \gamma_i^*$ in (3.1) and (3.2) to compute the multinomial probabilities for z_i and y_i, respectively. We then use these probabilities and use IMSL subroutine to generate the multinomial observations $z_i = (z_{i1}, z_{i2})$ and $y_i = (y_{i1}, y_{i2})$.

Estimation:

We estimate the regression and variance component (σ_γ) parameters by using the JGQL and MLE estimation approaches. Specifically, we solve the JGQL estimating equation (3.5) and MLE estimating equation (3.12) to obtain the estimates at a given simulation. We repeat the simulations for 500 times. The simulated mean (SM), simulated standard error (SSE) and mean squared error (MSE) for estimates of all model parameters are reported in Tables 3.1-3.5. In addition, we report the estimated standard errors (ESE) of the JGQL estimates in Tables 3.4 and 3.5. Note however that due to the computational complexity involved in the covariance matrix of the MLE estimates, we choose not to compute the ESE's under the MLE approach and thus there is no reporting on the ESE's of MLE estimates in Tables 3.4 and 3.5.

Simulation results:

From simulation results reported in Tables 3.1-3.5, we can see that both JGQL and MLE approaches produced almost unbiased estimates for all regression parameters in general. For example, when $\sigma_\gamma = 0.5$, the JGQL approach yielded the estimate of $\beta_2 = 0.25$ as $\hat{\beta}_{2,JGQL} = 0.244$ with mean squared error 0.024, and $\hat{\beta}_{2,MLE} = 0.235$ with MSE 0.045. Thus both estimates are close to the true value of the parameter. However, in general the MLE approach appear to produce regression estimates with same or larger SSE as compared to the JGQL approach. Consequently, the MLE approach produced regression estimates with same or larger MSE as compared to the JGQL approach. For example, when $\sigma_\gamma = 0.75$, the results in Table 3.4 showed that the MLE approach estimated $\theta_{z1} = 0.25$ as $\hat{\theta}_{z1,MLE} = 0.246$ with SSE 0.169 and MSE 0.029, whereas $\hat{\theta}_{z1,JGQL} = 0.245$ and the SSE and MSE for this estimate are 0.268 and 0.072, respectively. Remark that MLE produces consistent estimates similar to the GQL estimates. However, finite sample behavior can vary. In fact, MLE is known to produce optimal (highly efficient) estimates.

Next, it is clear from Tables 3.4 and 3.5 that the JGQL approach produces close ESE to the corresponding SSE values when $\sigma_\gamma = 0.75$ and $\sigma_\gamma = 1.0$, respectively. For example, when $\sigma_\gamma = 0.75$, the JGQL approach estimates σ_γ as $\hat{\sigma_\gamma}_{JGQL} = 0.736$ with ESE equal to 0.169, which is very close to the SSE value of 0.162. When $\sigma_\gamma = 1.0$, the JGQL approach estimates $\alpha_2 = 0.3$ as $\hat{\alpha}_{2,JGQL} = 0.274$ with SSE 0.125 and ESE 0.127, showing that the SSE and ESE are very close to each other. Thus, in general, the estimated standard error formula derived from the JGQL estimating equation works well as expected.

As far as the estimation of σ_γ is concerned, the MLE approach produced more biased estimates with larger SSE than the JGQL approach. This makes the MLE approach worse than the JGQL approach in the sense of MSE efficiency. For example, for small σ_γ, such as $\sigma_\gamma = 0.5$, $\hat{\sigma}_{\gamma,MLE} = 0.346$ with SSE 0.281 as compared to $\hat{\sigma}_{\gamma,JGQL} = 0.508$ with SSE 0.182. Thus, in this case, MLE produced an estimate with MSE 0.102 and JGQL estimated this parameter with MSE 0.033. The performance of MLE becomes worse when σ_γ increases. For example, for $\sigma_\gamma = 1.0$, MLE estimated this parameter as $\hat{\sigma}_{\gamma,MLE} = 0.866$ with MSE 0.642, whereas $\hat{\sigma}_{\gamma,JGQL} = 0.983$ with MSE 0.021. Thus MLE performed much worse when $\sigma_\gamma = 1.0$ as compared to the case for $\sigma_\gamma = 0.5$.

Now to understand the effect of correlation index parameter σ_γ on regression estimates, we have used the quasi-likelihood (QL) estimation technique for the regression parameters by ignoring the correlations for all simulation designs with different values of σ_γ, namely, $\sigma_\gamma = 0.1$, 0.35, 0.5, 0.75 and 1.0. For convenience, we first write the

QL estimating equation for the regression parameter vector ψ as follows

$$f(\psi) = \sum_{i=1}^{n} \frac{\partial(\bar{\Pi}'_{iz}, \bar{\Pi}'_{iy}, \bar{\Pi}'_{izy})}{\partial \psi} [\bar{\Sigma}_i]^{-1} \begin{pmatrix} z_i - \bar{\Pi}_{iz} \\ y_i - \bar{\Pi}_{iy} \\ g_i - \bar{\Pi}_{izy} \end{pmatrix} = 0, \qquad (3.14)$$

where $\bar{\pi}_{iz}^{(k)}$ in $\bar{\Pi}_{iz}$, for example, has the formula

$$\begin{aligned} \bar{\pi}_{iz}^{(k)} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iz}^{(k)}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu}, \\ &= \sum_{\nu=0}^{N} \frac{\exp(\alpha_k + x'_{i1}\theta_{k1} + 0 \times \gamma_{i\nu})}{1 + \sum_{q=1}^{K-1} \exp(\alpha_q + x'_{i1}\theta_{q1} + 0 \times \gamma_{i\nu})} \binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu}, \end{aligned}$$

with $\sigma_\gamma = 0$ and $\gamma_{i\nu} = \frac{\nu - N(0.5)}{\sqrt{N(0.5)(0.5)}}$, and $N = 40$ is used for the simulation study.

The simulation results under the QL approach were reported in Tables 3.1-3.3 for correlation index parameter values $\sigma_\gamma = 0.1$, 0.35, and 0.5, respectively. With regard to the QL estimation results under $\sigma_\gamma = 0.75$ and 1.0, we chose not to report them due to two fold problems encountered by the QL approach. One problem is that the QL approach produces biased estimates when correlation index parameter σ_γ gets larger. This is evident from the pattern exhibited in Tables 3.1-3.3. The other problem is that the QL approach encountered serious convergence problems especially in large correlation scenarios, namely, the QL estimating equation (3.14) failed to produce appropriate inverse matrix $[\bar{\Sigma}_i]^{-1}$ in a large number of simulations where σ_γ was large.

From the results in Tables 3.1-3.3 we can see that with small correlation, namely, when $\sigma_\gamma = 0.1$, the QL approach, as expected, produced competitive regression parameter estimates as compared with the JGQL and MLE estimates. However, as σ_γ increased, the QL approach was found to yield significantly biased estimates for the regression parameters along with larger SSE and MSE. For example, when $\sigma_\gamma = 0.35$, the QL approach estimated $\alpha_2 = 0.3$ as $\hat{\alpha}_{2,QL} = 0.253$ with SSE 0.105 and MSE 0.075, whereas the JGQL approach yielded $\hat{\alpha}_{2,JGQL} = 0.312$ with SSE 0.108 and MSE 0.012, indicating larger bias in the QL approach. By the same token, when $\sigma_\gamma = 0.5$, the QL approach estimated θ_{z1} as $\hat{\theta}_{z1,QL} = 0.188$ with SSE 0.178 and MSE 0.036, whereas the JGQL approach yielded $\hat{\theta}_{z1,JGQL} = 0.255$ with SSE 0.176 and MSE 0.031. Thus, by applying the QL method to the BMM model, unlike for the LCBM model discussed in Chapter 2, we see that the QL method would produce biased regression estimate when σ_γ is ignored. This indicates that one should estimate regression and correlation index parameters jointly as far as consistent estimation for the regression parameters is desired under the BMM model. This also means that unlike the LCBM model, the proposed BMM model is more general, but, the joint estimation of the correlation index parameter is needed.

Table 3.1: The SM (simulated mean), SSE (simulated standard error) and MSE (mean squared error) of the JGQL, MLE and QL estimates for selected regression parameter values and $\sigma_\gamma = 0.1$.

	JGQL			MLE			QL		
Parameter	SM	SSE	MSE	SM	SSE	MSE	SM	SSE	MSE
$\alpha_1 = 0.4$	0.425	0.108	0.012	0.427	0.137	0.019	0.390	0.104	0.011
$\alpha_2 = 0.3$	0.333	0.116	0.014	0.313	0.147	0.022	0.292	0.109	0.012
$\theta_{z1} = 0.25$	0.263	0.180	0.033	0.279	0.192	0.038	0.259	0.178	0.032
$\theta_{z2} = 0.35$	0.353	0.175	0.031	0.397	0.195	0.040	0.353	0.173	0.030
$\beta_1 = 0.35$	0.380	0.146	0.022	0.374	0.245	0.061	0.345	0.150	0.023
$\beta_2 = 0.25$	0.280	0.139	0.020	0.285	0.266	0.072	0.237	0.156	0.024
$\theta_{y1} = 0.4$	0.431	0.186	0.036	0.434	0.217	0.048	0.402	0.180	0.032
$\theta_{y2} = 0.5$	0.529	0.184	0.035	0.507	0.214	0.046	0.511	0.184	0.034
$\theta_{c1} = 0.2$	0.209	0.058	0.003	0.206	0.058	0.003	0.202	0.059	0.003
$\theta_{c2} = 0.3$	0.312	0.064	0.004	0.306	0.061	0.004	0.305	0.061	0.004
$\sigma_\gamma = 0.1$	0.355	0.161	0.091	0.292	0.237	0.093	-	-	-

Table 3.2: The SM (simulated mean), SSE (simulated standard error) and MSE (mean squared error) of the JGQL, MLE and QL estimates for selected regression parameter values and $\sigma_\gamma = 0.35$.

	JGQL			MLE			QL		
Parameter	SM	SSE	MSE	SM	SSE	MSE	SM	SSE	MSE
$\alpha_1 = 0.4$	0.416	0.109	0.012	0.403	0.122	0.015	0.352	0.100	0.134
$\alpha_2 = 0.3$	0.312	0.108	0.012	0.317	0.131	0.018	0.253	0.105	0.075
$\theta_{z1} = 0.25$	0.267	0.191	0.037	0.259	0.201	0.040	0.287	0.174	0.113
$\theta_{z2} = 0.35$	0.366	0.179	0.032	0.345	0.226	0.051	0.397	0.176	0.188
$\beta_1 = 0.35$	0.374	0.151	0.023	0.320	0.214	0.047	0.326	0.142	0.127
$\beta_2 = 0.25$	0.265	0.145	0.021	0.222	0.220	0.049	0.220	0.157	0.073
$\theta_{y1} = 0.4$	0.393	0.184	0.034	0.449	0.188	0.038	0.384	0.178	0.179
$\theta_{y2} = 0.5$	0.506	0.176	0.031	0.557	0.178	0.035	0.496	0.192	0.282
$\theta_{c1} = 0.2$	0.198	0.062	0.004	0.201	0.086	0.007	0.192	0.058	0.040
$\theta_{c2} = 0.3$	0.304	0.064	0.004	0.298	0.091	0.008	0.288	0.055	0.086
$\sigma_\gamma = 0.35$	0.428	0.166	0.034	0.321	0.204	0.043	-	-	-

Table 3.3: The SM (simulated mean), SSE (simulated standard error) and MSE (mean squared error) of the JGQL, MLE and QL estimates for selected regression parameter values and $\sigma_\gamma = 0.5$.

	JGQL			MLE			QL		
Parameter	SM	SSE	MSE	SM	SSE	MSE	SM	SSE	MSE
$\alpha_1 = 0.4$	0.417	0.111	0.013	0.363	0.194	0.039	0.360	0.107	0.013
$\alpha_2 = 0.3$	0.321	0.106	0.012	0.272	0.198	0.040	0.260	0.108	0.013
$\theta_{z1} = 0.25$	0.255	0.176	0.031	0.305	0.189	0.039	0.188	0.178	0.036
$\theta_{z2} = 0.35$	0.354	0.171	0.029	0.385	0.183	0.035	0.294	0.169	0.032
$\beta_1 = 0.35$	0.356	0.156	0.024	0.333	0.215	0.047	0.333	0.149	0.022
$\beta_2 = 0.25$	0.244	0.155	0.024	0.235	0.211	0.045	0.228	0.150	0.023
$\theta_{y1} = 0.4$	0.427	0.188	0.036	0.412	0.183	0.034	0.325	0.174	0.036
$\theta_{y2} = 0.5$	0.538	0.181	0.034	0.508	0.154	0.024	0.431	0.176	0.036
$\theta_{c1} = 0.2$	0.228	0.061	0.005	0.189	0.065	0.004	0.214	0.056	0.003
$\theta_{c2} = 0.3$	0.331	0.062	0.005	0.288	0.060	0.004	0.319	0.055	0.003
$\sigma_\gamma = 0.5$	0.508	0.182	0.033	0.346	0.281	0.102	-	-	-

Table 3.4: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and MLE estimates for selected regression parameter values and $\sigma_\gamma = 0.75$.

	JGQL				MLE			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.368	0.114	0.014	0.120	0.546	0.757	0.594	-
$\alpha_2 = 0.3$	0.268	0.116	0.014	0.122	0.459	0.752	0.591	-
$\theta_{z1} = 0.25$	0.246	0.169	0.029	0.178	0.245	0.268	0.072	-
$\theta_{z2} = 0.35$	0.351	0.172	0.030	0.180	0.333	0.258	0.067	-
$\beta_1 = 0.35$	0.378	0.153	0.024	0.164	0.531	0.739	0.578	-
$\beta_2 = 0.25$	0.279	0.162	0.027	0.168	0.421	0.754	0.598	-
$\theta_{y1} = 0.4$	0.320	0.181	0.039	0.192	0.358	0.248	0.063	-
$\theta_{y2} = 0.5$	0.416	0.195	0.045	0.195	0.468	0.240	0.059	-
$\theta_{c1} = 0.2$	0.171	0.063	0.005	0.065	0.226	0.285	0.082	-
$\theta_{c2} = 0.3$	0.271	0.059	0.004	0.066	0.322	0.264	0.070	-
$\sigma_\gamma = 0.75$	0.736	0.162	0.026	0.169	0.809	1.019	1.041	-

Table 3.5: The SM (simulated mean), SSE (simulated standard error), MSE (mean squared error), and ESE (estimated standard error) of the JGQL and MLE estimates for selected regression parameter values and $\sigma_\gamma = 1.0$.

	JGQL				MLE			
Parameter	SM	SSE	MSE	ESE	SM	SSE	MSE	ESE
$\alpha_1 = 0.4$	0.372	0.116	0.014	0.124	0.429	0.580	0.338	-
$\alpha_2 = 0.3$	0.274	0.125	0.016	0.127	0.309	0.540	0.292	-
$\theta_{z1} = 0.25$	0.220	0.174	0.031	0.185	0.242	0.279	0.078	-
$\theta_{z2} = 0.35$	0.323	0.180	0.033	0.186	0.340	0.277	0.077	-
$\beta_1 = 0.35$	0.310	0.157	0.026	0.164	0.296	0.198	0.042	-
$\beta_2 = 0.25$	0.215	0.163	0.028	0.167	0.191	0.160	0.029	-
$\theta_{y1} = 0.4$	0.402	0.188	0.035	0.196	0.505	0.615	0.389	-
$\theta_{y2} = 0.5$	0.494	0.192	0.037	0.198	0.623	0.576	0.347	-
$\theta_{c1} = 0.2$	0.211	0.063	0.004	0.070	0.221	0.279	0.078	-
$\theta_{c2} = 0.3$	0.310	0.067	0.005	0.071	0.323	0.281	0.080	-
$\sigma_\gamma = 1.0$	0.983	0.145	0.021	0.145	0.866	0.790	0.642	-

3.2.5 Reanalysis of diabetic retinopathy data

In this section, we reanalyze the diabetic retinopathy data by using the BMM model discussed in this chapter, whereas the same data were analyzed earlier in Section 2.2.4 by fitting the LCBM model.

3.2.5.1 An application of the bivariate binary mixed (BBM) model

Similar to Section 2.2.4.1, we treat the diabetic retinopathy status as a binary (absence or presence) variable. But the correlation between two such binary variables are thought to be generated through individual random effects common to both response variables. Considering z_i and y_i as the binary retinopathy status of left and right eyes, respectively, we now precisely write the bivariate binary mixed (BBM) model as follows as a special case of the BMM model described in (3.1)-(3.2). The BBM model is given by

$$\tilde{\pi}_{iz}(\gamma_i) = Pr(z_i = 1|\gamma_i) = \frac{\exp(\alpha_1 + x_i'\theta + \sigma_\gamma\gamma_i)}{1 + \exp(\alpha_1 + x_i'\theta + \sigma_\gamma\gamma_i)}, \tag{3.15}$$

$$and\ \tilde{\pi}_{iy}(\gamma_i) = Pr(y_i = 1|\gamma_i) = \frac{\exp(\beta_1 + x_i'\theta + \sigma_\gamma\gamma_i)}{1 + \exp(\beta_1 + x_i'\theta + \sigma_\gamma\gamma_i)}. \tag{3.16}$$

In (3.15) and (3.16), $x_i = (x_{i11}, x_{i12}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6}, x_{i7})'$ is the 8-dimensional covariate vector as in (2.17). Further, as in (3.1)-(3.2), γ_i in (3.15) and (3.16) is the common random effect of the ith individual causing z_i and y_i to be correlated unconditionally. That is, as discussed in (3.4), π_{i11}, to be precise, $\pi_{i11}(\sigma_\gamma) = Pr(z_i = 1, y_i = 1) = \int_{-\infty}^{\infty} \tilde{\pi}_{iz}\tilde{\pi}_{iy} f_N(\gamma_i)d\gamma_i$ involves the correlation through (σ_γ) between z_i and y_i. It is important to accommodate these correlations in order to obtain θ, the effect of x_i on y_i and z_i. We need to compute α_1, β_1 and σ_γ as well. Thus we estimate

$\phi^* = (\alpha_1, \beta_1, \theta', \sigma_\gamma)'$.

Next, to estimate ϕ^*, we turn back to the JGQL and ML estimation equations (3.5) and (3.12), respectively. By solving them iteratively, as discussed in Sections 3.2.1 and 3.2.2, we obtain the estimates of ϕ^*. The JGQL estimates were obtained in 10 iterations and ML estimates in 35 iterations. As far as the standard errors of these estimates are concerned, as discussed in Section 3.2.3, the computation of the standard errors by MLE is very complicated due to integration over the distribution of γ_i. However, the standard errors of JGQL estimates were obtained easily by using (3.13). We, therefore, provide the JGQL and ML estimates but the standard errors for JGQL estimates only. These estimates and standard errors are reported in Table 3.6.

The results in Table 3.6 show that the propensity of diabetic retinopathy (probability of having diabetic retinopathy problem) tends to increase with longer DD, higher GHL, higher DBP, male gender, presence of Pr, more DI per day and presence of ME. This observation agrees with the diabetic retinopathy data analysis results reported in Table 2.10 under the LCBB model perfectly, even though the magnitude of the covariate estimates along with the estimates of the intercepts were found to be different under the BBM and LCBB models. For example, because DD was coded as $(0, 0)$ for duration between 5 and 10 years, the large negative value of $\hat{\theta}_{11,JGQL} = -5.780$ and positive value of $\hat{\theta}_{12,JGQL} = 6.423$ under the present BBM model show that as DD increases, the probability of an individual to have retinopathy problem increases, whereas under the LCBB model the estimates of θ_{11} and θ_{12} were found to be $\hat{\theta}_{11,SSGQL} = -2.1187$ and $\hat{\theta}_{12,SSGQL} = 2.2376$, respectively. But the differences in magnitude are reasonable, because under the LCBB model correlation

Table 3.6: JGQL and MLE estimation results for the diabetic retinopathy data under the BBM model.

	JGQL		MLE	
Parameter (Effect of)	Estimate	ESE	Estimate	ESE
α_1	-0.853	0.582	-0.848	-
β_1	-0.624	0.579	-0.620	-
θ_{11} (DD low)	-5.780	0.942	-5.728	-
θ_{12} (DD high)	6.423	0.870	6.365	-
θ_2 (GHL)	1.120	0.285	1.110	-
θ_3 (DBP)	1.515	0.551	1.502	-
θ_4 (Gender)	-0.668	0.515	-0.660	-
θ_5 (Pr)	1.512	0.885	1.497	-
θ_6 (DI)	0.016	0.514	0.017	-
θ_7 (ME)	4.596	2.154	4.567	-
σ_γ	4.528	0.563	4.484	-

index parameters do not enter into the marginal probabilities, whereas in the present BBM model the marginal probabilities are defined as functions of correlation index parameter σ_γ. Next, the regression effect of ME (macular edema) on the probability of having diabetic retinopathy in left or right eye was found to be $\hat{\theta}_{7,JGQL} = 4.596$, since ME was coded as $x_7 = 1$ in the presence of ME, this high positive value indicates that ME has great effects on retinopathy status.

Note that the random effect parameter estimate, i.e., $\hat{\sigma}_{\gamma,JGQL} = 4.528$ implies that retinopathy status of left and right eyes are highly correlated, this also agrees with the large correlation index parameter value for ρ_{11}, i.e., $\hat{\rho}_{11,SSGQL} = 0.6361$, found based on the LCBB model. Note however that it is not possible to find any theoretical relationship between σ_γ and ρ_{11} as the models are completely different.

Now similar to Section 2.2.4 where we have used the estimated regression effects (ψ) and correlation index parameter ρ_{11} to examine the bivariate correlation pattern between left and right eye retinopathy status by computing the individual correlations, under the present BBM model, we can also use the estimated regression parameter (ψ) and correlation index parameter (σ_γ) to calculate the individual correlations. To be specific, we compute

$$\rho_{iyz} = corr(y_i, z_i) = \frac{\bar{\pi}_{i11} - \bar{\pi}_{iz}\bar{\pi}_{iy}}{\sqrt{\bar{\pi}_{iz}(1-\bar{\pi}_{iz})\bar{\pi}_{iy}(1-\bar{\pi}_{iy})}}, \tag{3.17}$$

where

$$\begin{aligned}
\bar{\pi}_{iz} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iz}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu}, \\
\bar{\pi}_{iy} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iy}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu}, \\
\text{and } \bar{\pi}_{i11} &= \sum_{\nu=0}^{N} \tilde{\pi}_{iz}(\gamma_{i\nu})\tilde{\pi}_{iy}(\gamma_{i\nu})\binom{N}{\nu}(\frac{1}{2})^{\nu}(1-\frac{1}{2})^{N-\nu},
\end{aligned}$$

respectively, with $\gamma_{i\nu} = \frac{\nu - N(0.5)}{\sqrt{N(0.5)(0.5)}}$ and $N = 40$ as before.

Next, by using the JGQL estimates given in Table 3.6, we can calculate the correlation ρ_{iyz} for each $i = 1, \ldots, n$. We give the histogram of correlations in Figure 3.1 below. From Figure 3.1 we can see that most correlations lie between 0.58 and 0.68, the minimum of ρ_{iyz} is found to be 0.2376, and the maximum is 0.6727. To be specific, the average of ρ_{iyz} under the BBM model is given by $\bar{\rho}_{yz} = 0.6050$. When these correlations under the present BBM model are compared with those in Figure 2.6

under the LCBB model, the later model produced average correlation $\bar{\rho}_{yz} = 0.6426$. These average values are close to each other.

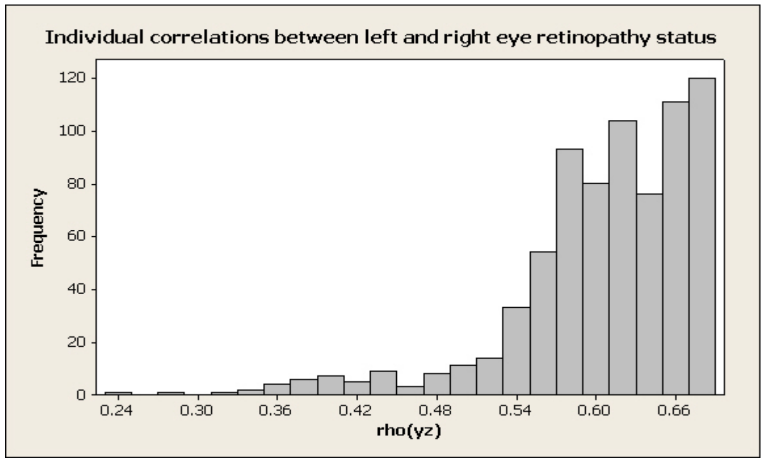


Figure 3.1: Histogram of correlations between left and right eye binary retinopathy status under the BBM model.

3.2.5.2 An application of the bivariate multinomial mixed (BMM) model

In this section, similar to Section 2.2.4.2, we treat the diabetic retinopathy status as a multinomial variable with three categories: absence, non-severity, and severity status of DR. However, as opposed to the LCBM model, here we consider the 3 category based BMM model, which is an extension of the BBM model discussed in the last section. Thus, on top of the notations used in the LCBM model, we now incorporate random effects which cause correlations among the two trinomial responses. More specifically, following (3.1)-(3.2), we display the trinomial mixed model as follows:

$$\begin{aligned}
\tilde{\pi}_{iz}^{(1)}(\gamma_i) = Pr(z_i = z_i^{(1)}|\gamma_i) &= \frac{\exp(\alpha_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i)}{1 + \exp(\alpha_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i) + \exp(\alpha_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}, \\
\tilde{\pi}_{iz}^{(2)}(\gamma_i) = Pr(z_i = z_i^{(2)}|\gamma_i) &= \frac{\exp(\alpha_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}{1 + \exp(\alpha_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i) + \exp(\alpha_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}, \quad (3.18) \\
\tilde{\pi}_{iz}^{(3)}(\gamma_i) = Pr(z_i = z_i^{(3)}|\gamma_i) &= \frac{1}{1 + \exp(\alpha_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i) + \exp(\alpha_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)};
\end{aligned}$$

$$\begin{aligned}
and\ \tilde{\pi}_{iy}^{(1)}(\gamma_i) = Pr(y_i = y_i^{(1)}|\gamma_i) &= \frac{\exp(\beta_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i)}{1 + \exp(\beta_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i) + \exp(\beta_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}, \\
\tilde{\pi}_{iy}^{(2)}(\gamma_i) = Pr(y_i = y_i^{(2)}|\gamma_i) &= \frac{\exp(\beta_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}{1 + \exp(\alpha_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i) + \exp(\beta_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}, \quad (3.19) \\
\tilde{\pi}_{iy}^{(3)}(\gamma_i) = Pr(y_i = y_i^{(3)}|\gamma_i) &= \frac{1}{1 + \exp(\beta_1 + x_i'\theta_1 + \sigma_\gamma\gamma_i) + \exp(\beta_2 + x_i'\theta_2 + \sigma_\gamma\gamma_i)}.
\end{aligned}$$

In (3.18) and (3.19), $x_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4}, x_{i5}, x_{i6})'$ is the 6-dimensional covariate vector as in (2.21)-(2.22).

To estimate the parameters in the mixed model, we have used the JGQL approach and these JGQL estimates and their standard errors are reported in Table 3.7.

When results in Table 3.7 are compared to those of Table 2.11, the effects of covariates appear to have similar interpretation, except the magnitude of the effects are different. For example, the large positive value of $\hat{\theta}_{11,JGQL} = 5.0229$ and $\hat{\theta}_{12,JGQL} = 5.4474$ under the present BMM model show that as DD increases, the probability of an individual to have retinopathy problem increases, and that longer DD leads to more severe retinopathy problem. Also Table 3.7 shows the estimate of σ_γ as $\hat{\sigma}_{\gamma,JGQL} = 4.9945$ as the standard deviation of the random effect or correlation index parameter. This large positive estimate of σ_γ implies that retinopathy status of left and right eyes are highly correlated. This large positive estimate $\hat{\sigma}_\gamma$ also agrees with the large correlation index parameter values for ρ_{11} and ρ_{22}, i.e., $\hat{\rho}_{11,SSGQL} = 0.6405$ and $\hat{\rho}_{22,SSGQL} = 0.6740$, found based on the LCBM model. Note that the results of Table 3.7 also agree with the binary analysis based results in Table 3.6, except that Table 3.7 provides more detailed information on effects of covaraites on various levels of DR status. For example, the regression effect of ME (macular edema) on the probability of having non-severe diabetic retinopathy in left or right eye was found to be $\hat{\theta}_{61,JGQL} = 5.3394$, and the regression effect of ME on the probability of having severe diabetic retinopathy was found to be $\hat{\theta}_{62} = 7.4150$ under the present trinomial analysis, whereas in the binary case, $\hat{\theta}_{7,SSGQL} = 2.6025$ shows the effect of ME on presence of DR.

Table 3.7: JGQL estimation results for the diabetic retinopathy data under the BMM model.

Parameter (Effect of)	Estimate	ESE
α_1	1.1527	0.4917
α_2	-2.0704	0.5644
β_1	1.3746	0.4970
β_2	-1.7808	0.5668
θ_{11} (DD on non-severe DR)	5.0229	0.6490
θ_{12} (DD on severe DR)	5.4474	0.6575
θ_{21} (GHL on non-severe DR)	1.1222	0.2915
θ_{22} (GHL on severe DR)	1.1315	0.3125
θ_{31} (DBP on non-severe DR)	2.5025	0.6028
θ_{32} (DBP on severe DR)	2.9889	0.6526
θ_{41} (Gender on non-severe DR)	-0.6301	0.5353
θ_{42} (Gender on severe DR)	-0.8000	0.5793
θ_{51} (Pr on non-severe DR)	2.5618	0.9609
θ_{52} (Pr on severe DR)	3.7453	0.9870
θ_{61} (ME on non-severe DR)	5.3394	2.4511
θ_{62} (ME on severe DR)	7.4150	2.4577
σ_γ	4.9945	0.5914

Chapter 4

Nonlinear Conditional Bivariate Multinomial (NLCBM) Fixed Model

Recall that in Chapter 2, we exploited the linear conditional bivariate multinomial (LCBM) fixed model for bivariate multinomial data analysis, which consists of fully specified marginal probabilities for both multinomial response variables z_i and y_i, as well as linear conditional probability that describes the correlation between z_i and y_i. In this LCBM model, we used the so-called dependence parameters ρ_{kj} to indicate the conditional relationship between z_{ik} and y_{ij}, which, however, as expected, suffers from certain range restriction problems. In the present chapter, similar to but different than the LCBM fixed model, we use a fully specified marginal probability model for one multinomial response variable, for example, z_i, and use a multinomial logistic approach to model the conditional probabilities of y_i given z_i through dependence parameters δ_{kj}, $k = 1, \ldots, K-1$ and $j = 1, \ldots, J-1$, whereas in the LCBM model the conditional probabilities of y_i given z_i were treated to be linear. We refer to this proposed model as the nonlinear conditional bivariate multinomial (NLCBM) fixed model. Note that in this model, the dependence parameters δ_{kj}, unlike their counterparts ρ_{kj} in the LCBM fixed model, do not have any range restriction problems, as δ_{kj} can range from $-\infty$ to ∞. However the present NLCBM fixed model is a partly specified model, this is because in this model, the marginal probabilities for one response variable, say, z_i are prespecified only, as opposed to the LCBM fixed model in Chapter 2. To be more clear, the marginal probabilities of the other multinomial response variable, say, y_i are not prespecified, instead, they can be obtained through summation of suitable joint probabilities computed by using certain marginal probabilities of z_i and conditional probabilities of y_i given z_i.

4.1 The model

To develop the desired NLCBM fixed model, as indicated above, we first consider that the multinomial response variable z_i has the specified marginal probability as in (2.1) under the LCBM fixed model. That is, we write

$$\begin{aligned} \pi_{iz}^{(k)} &= Pr(z_i = z_i^{(k)}) = \frac{\exp(\alpha_k + x'_{i1}\theta_{k1})}{1 + \sum_{u=1}^{K-1} \exp(\alpha_u + x'_{i1}\theta_{u1})}, \ k = 1, \ldots, K-1, \\ \pi_{iz}^{(K)} &= Pr(z_i = z_i^{(K)}) = 1 - \sum_{k=1}^{K-1} \pi_{iz}^{(k)} = \frac{1}{1 + \sum_{u=1}^{K-1} \exp(\alpha_u + x'_{i1}\theta_{u1})}, \end{aligned} \qquad (4.1)$$

where $x_{i1} = (x'_{iz} : 1 \times p_1, x'_{ic} : 1 \times p_2)' : p \times 1$ as in the model (2.1), and $\theta_{k1} = (\theta'_{kz}, \theta'_{kc})'$ is the $p-$dimensional vector of regression effects of x_{i1} on z_{ik}. However, unlike the linear conditional probability considered in (2.3), we now model the conditional probability of y_i given $z_i = z_i^{(k)}$ as follows:

$$\begin{aligned}
\eta_{ij|k}^{(y)} &= Pr(y_i = y_i^{(j)} | z_i = z_i^{(k)}) \\
&= \frac{\exp[\beta_j + x'_{iy}\theta_{jy} + \sum_{u=1}^{K-1} \delta_{uj}(z_{iu}^{(k)} - \pi_{iz}^{(u)})]}{1 + \sum_{l=1}^{J-1} \exp[\beta_l + x'_{iy}\theta_{ly} + \sum_{u=1}^{K-1} \delta_{ul}(z_{iu}^{(k)} - \pi_{iz}^{(u)})]}, \\
& \quad j = 1, \ldots, J-1,\ k = 1, \ldots, K; \\
\text{and } \eta_{iJ|k}^{(y)} &= Pr(y_i = y_i^{(J)} | z_i = z_i^{(k)}) \\
&= 1 - \sum_{j=1}^{J-1} \eta_{ij|k}^{(y)},\ k = 1, \ldots, K, \qquad (4.2)
\end{aligned}$$

where $z_{iu}^{(k)}$ is the uth $(u = 1, \ldots, K-1)$ component of $z_i^{(k)}$, with $z_{iu}^{(k)} = 1$ if $u = k$, and 0 otherwise; δ_{uj} is referred to as the dependence parameter relating y_{ij} with z_{iu}. Notice that in writing the conditional probability in (4.2), we have used the individual response specific covariate x_{iy}, whereas in (2.3), the marginal probability for y_i contains the covariates $x_{i2} = (x'_{iy}, x'_{ic})'$. This is quite reasonable, as the common covariates are already used in (4.1) to construct the probability model for z_i, because y_i depends on z_i through (4.2), it implies that the probability model for y_i uses the common covariates through z_i. Further note that one may compute the joint probability $\pi_{ikj} = Pr(z_i = z_i^{(k)}, y_i = y_i^{(j)})$ as $\pi_{ikj} = \pi_{iz}^{(k)} \eta_{ij|k}^{(y)}$ for $k = 1, \ldots, K$ and $j = 1, \ldots, J$.

Note that in writing (4.2), we have used the conditioning on z_i. One may also use alternatively the conditional probability for z_i given y_i. To be specific, one can write the marginal probabilities for y_i as

$$\begin{aligned}
\pi_{iy}^{(j)} &= Pr(y_i = y_i^{(j)}) = \frac{\exp(\beta_j + x'_{i2}\theta_{j2})}{1 + \sum_{l=1}^{J-1} \exp(\beta_l + x'_{i2}\theta_{l2})},\ j = 1, \ldots, J-1, \\
\pi_{iy}^{(J)} &= Pr(y_i = y_i^{(J)}) = 1 - \sum_{j=1}^{J-1} \pi_{iy}^{(j)} = \frac{1}{1 + \sum_{l=1}^{J-1} \exp(\beta_l + x'_{i2}\theta_{l2})}, \qquad (4.3)
\end{aligned}$$

where $x_{i2} = (x'_{iy}, x'_{ic})'$. Next, by changing the dependence parameters, the conditional probabilities of z_i given y_i can be written as

$$\begin{aligned}
\tilde{\eta}^{(z)}_{ik|j} &= Pr(z_i = z_i^{(k)} | y_i = y_i^{(j)}) \\
&= \frac{\exp[\alpha_k + x'_{iz}\theta_{kz} + \sum_{l=1}^{J-1} \tilde{\delta}_{lk}(y_{il}^{(j)} - \pi_{iy}^{(l)})]}{1 + \sum_{u=1}^{K-1} \exp[\alpha_u + x'_{iz}\theta_{uz} + \sum_{l=1}^{J-1} \tilde{\delta}_{lu}(y_{il}^{(j)} - \pi_{iy}^{(l)})]}, \qquad (4.4) \\
& \quad k = 1, \ldots, K-1,\ j = 1, \ldots, J; \\
\text{and } \tilde{\eta}^{(z)}_{iK|j} &= Pr(z_i = z_i^{(K)} | y_i = y_i^{(j)}) \\
&= 1 - \sum_{k=1}^{K-1} \tilde{\eta}^{(z)}_{ik|j},\ j = 1, \ldots, J.
\end{aligned}$$

However, in this chapter, we follow the models in (4.1)-(4.2) only.

4.2 Likelihood estimation for the NLCBM fixed model

Notice that the marginal model (4.1) involves the regression parameter $\psi^* = (\alpha_1, \ldots, \alpha_k, \ldots, \alpha_{K-1}, \theta'_{1z}, \ldots, \theta'_{kz}, \ldots, \theta'_{K-1,z}, \theta'_{1c}, \ldots, \theta'_{kc}, \ldots, \theta'_{K-1,c})'$, which are also involved in the conditional probability (4.2) through $\pi_{iz}^{(k)}$. Furthermore, the conditional probability involves the additional new parameter vector $\zeta = (\beta_1, \ldots, \beta_j, \ldots, \beta_{J-1}, \ldots, \theta'_{1y}, \ldots, \theta'_{jy}, \ldots, \theta'_{J-1,y}, \delta_{11}, \ldots, \delta_{k1}, \delta_{K-1,1}, \delta_{21}, \ldots, \delta_{kj}, \ldots, \delta_{K-1,J-1})'$. In order to derive the desired likelihood estimating equations for ψ^* and ζ, we first write the full likelihood function for ψ^* and ζ as

$$\begin{aligned} L(\psi^*,\zeta) &= \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{j=1}^{J}(\pi_{ikj})^{z_{ik}y_{ij}} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\prod_{j=1}^{J}(\pi_{iz}^{(k)}\eta_{ij|k}^{(y)})^{z_{ik}y_{ij}} \\ &= \prod_{i=1}^{n}\prod_{k=1}^{K}\left[(\pi_{iz}^{(k)})^{z_{ik}}\prod_{j=1}^{J}(\eta_{ij|k}^{(y)})^{z_{ik}y_{ij}}\right], \qquad (4.5) \end{aligned}$$

where $\pi_{iz}^{(k)}$ is defined in (4.1) and the conditional probability $\eta_{ij|k}^{(y)}$ is defined in (4.2). Next, we take the logarithm of $L(\psi^*,\zeta)$ given above and obtain the log likelihood function for (ψ^*,ζ) as follows

$$l(\psi^*,\zeta) = logL(\psi^*,\zeta) = \sum_{i=1}^{n}\sum_{k=1}^{K} z_{ik}log\pi_{iz}^{(k)} + \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{j=1}^{J} z_{ik}y_{ij}log\eta_{ij|k}^{(y)}. \qquad (4.6)$$

4.2.1 Estimation of the parameters

One may then construct the likelihood estimating equations for the parameters involved in the model (4.1)-(4.2).

Likelihood equation for ψ^*

By taking the derivatives of $l(\psi^*,\zeta)$ in (4.6) with respect to the components of ψ^*, we obtain the likelihood estimating equations for α_k and θ_{k1} $(k = 1,\ldots,K-1)$ as

$$\begin{aligned} \frac{\partial l(\psi^*,\zeta)}{\partial \alpha_k} &= \sum_{i=1}^{n}(z_{ik} - \pi_{iz}^{(k)}) - \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{j=1}^{J-1} z_{ik}\delta_{kj}(y_{ij} - \eta_{ij|k}^{(y)})\pi_{iz}^{(k)}(1-\pi_{iz}^{(k)}) = 0, \qquad (4.7) \\ \frac{\partial l(\psi^*,\zeta)}{\partial \theta_{k1}} &= \sum_{i=1}^{n}(z_{ik} - \pi_{iz}^{(k)})x_{i1} - \sum_{i=1}^{n}\sum_{k=1}^{K}\sum_{j=1}^{J-1} z_{ik}\delta_{kj}(y_{ij} - \eta_{ij|k}^{(y)})\pi_{iz}^{(k)}(1-\pi_{iz}^{(k)})x_{i1} \\ &= 0. \qquad (4.8) \end{aligned}$$

Likelihood equation for ζ

Next, By taking the derivatives of $l(\psi^*,\zeta)$ in (4.6) with respect to ζ, we obtain the likelihood estimating equation for β_j, θ_{jy} ($j = 1, \ldots, J-1$) as

$$\frac{\partial l(\psi^*, \zeta)}{\partial \beta_j} = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik}(y_{ij} - \eta_{ij|k}^{(y)}) = 0, \quad (4.9)$$

$$\frac{\partial l(\psi^*, \zeta)}{\partial \theta_{jy}} = \sum_{i=1}^{n} \sum_{k=1}^{K} z_{ik}(y_{ij} - \eta_{ij|k}^{(y)}) x_{iy} = 0, \quad (4.10)$$

and for δ_{kj} ($k = 1, \ldots, K-1$ and $j = 1, \ldots, J-1$) as

$$\frac{\partial l(\psi^*, \zeta)}{\partial \delta_{kj}} = \sum_{i=1}^{n} \sum_{u=1}^{K} z_{iu}(y_{ij} - \eta_{ij|u}^{(y)})(z_{iu}^{(k)} - \pi_{iz}^{(k)}) = 0. \quad (4.11)$$

Note that these likelihood equations in (4.7)-(4.11) can be solved jointly which will however requires extensive second derivatives computation. This computational burden can be reduced by solving these equations in two stages. To be specific, in the first stage, we solve (4.7)-(4.8) for ψ^* assuming that ζ is known, i.e., using some initial values for the parameters involved in ζ. In the second stage, the estimate of ψ^* obtained at the first stage is used in solving the estimating equation for ζ. This will constitute a cycle of iterations, and the iterations will continue until convergence is reached. For simplicity, we write the likelihood iterative equations for these two stages as follows.

Stage 1: Iterative equation for ψ^*

The iterative equation for ψ^* is given by:

$$\hat{\psi}^*_{r+1} = \hat{\psi}^*_r - \left[\frac{\partial^2 l(\psi^*, \zeta)}{\partial \psi^* \partial \psi^{*\prime}}\right]_r^{-1} \left(\frac{\partial l(\psi^*, \zeta)}{\partial \psi^*}\right)_r,$$

where the second derivatives can be computed from (4.7)-(4.8), which is straightforward but would require lengthy calculations. These are not given here as our purpose is to demonstrate how the likelihood method can be exploited.

Stage 2: Iterative equation for ζ

Similarly, the iterative equation for ζ is given by:

$$\hat{\zeta}_{r+1} = \hat{\zeta}_r - \left[\frac{\partial^2 l(\psi^*, \zeta)}{\partial \zeta \partial \zeta'}\right]_r^{-1} \left(\frac{\partial l(\psi^*, \zeta)}{\partial \zeta}\right)_r,$$

where the second derivatives can be computed from (4.9)-(4.11).

Note that one may also develop JGQL estimating equation approach to estimate the parameters ψ^* and ζ, which will naturally be more complicated as compared to the JGQL approach developed in Chapter 2.

Further note that in view of the computational results discussed in Chapters 2 and 3, it is reasonable to expect that both likelihood and the JGQL approaches will perform well in estimating the parameters of the model (4.1)-(4.2). We however do not undertake any further numerical computations at this stage.

Chapter 5

Concluding Remarks

Even though in many practical situations bivariate categorical responses are collected in a cross sectional setup, the existing inferences have drawbacks in analyzing this type of data due to improper modeling and/or difficult model parameter interpretation [Agresti (2002)] or arbitrary extra modeling posed on the correlations between two responses [Williamson et. al. (1995)]. In the thesis, we have developed three types of conditional probability models. One such model is constructed by linear probability function conditioning one response on the other, where marginal probabilities are fully specified. The second model is constructed by conditioning on suitable random effects so that unconditionally two multinomial variables become correlated. Also we have considered a conditional model similar to the first model but using a logistic (non-linear) probability function conditioning one response on the other.

As far as the inferences are concerned, because the likelihood approach for the first model considered in Chapter 2 is relatively complicated, we have used the JGQL (joint generalized quasi-likelihood) and SSGQL (single stage GQL) approaches for the estimation of the parameters of this model. The simulation study conducted in this chapter shows that these two estimation approaches are competitive and they estimated the parameters well. However, for simplicity, we have recommended the use of SSGQL as compared to the JGQL approach.

Note that the second conditional model discussed in Chapter 3 accommodates the correlations between two multinomial variables through common individual random effects. In developing this model, it was assumed that conditional on the common random effects, the marginal probabilities have specified multinomial logistic forms. Consequently this model allows full range for the correlations. The estimation of the parameters including the random effects variance (correlation index parameter) was done by using the JGQL and likelihood approaches. It was found that the JGQL approach estimates better or as well as the likelihood approach. However, because the likelihood estimation method for mixed model is computationally more involved, we prefer the JGQL approach over the likelihood approach.

In this thesis we have mainly dealt with bivariate multinomial responses collected from the same individual at a given point of time. However, there may be situations where more than two multinomial responses are collected from the same individual at a given point of time. The analysis of this type of data will require generalization of the bivariate multinomial data analysis discussed in this thesis. Furthermore, there may be situations that bivariate (possibly multivariate) multinomial data are collected from the same person over a short period of time. This type of longitudinal data analysis will require the generalization of the existing univariate longitudinal models (e.g., Chowdhury (2011)) to the multivariate setup. This generalization will be naturally more complex and is beyond the scope of the present thesis.

Bibliography

Agresti, A. (1996). *An Introduction to Categorical Data Analysis.* John Wiley and Sons.

Agresti, A. (2002). *Categorical data analysis*, 2nd edition. John Wiley and Sons.

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd edition. John Wiley and Sons.

Agresti, A. (2013). *Categorical data analysis*, 3rd edition. Wiley.

Amemiya, T. (1985). *Advanced Econometrics.* Harvard University Press.

Breslow, N.E. and Clayton, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association*, **88**, 9-25.

Chen B., Yi, G.Y., and Cook, R. (2009). Likelihood analysis of joint marginal and conditional models for longitudinal categorical data. *The Canadian Journal of Statistics*, **37**, 182-205.

Chowdhury, R. I. (2011). *Inference in Longitudinal Multinomial Fixed and Mixed Model.* An published PhD thesis, Department of Mathematics and Statistics, Memorial University of Newfoundland.

Christensen, R. (1997). *Log-linear Models and Logistic Regression.* New York: Springer.

Conaway, M.R. (1989). Analysis of repeated categorical measurements with conditional likelihood methods. *Journal of the American Statistical Association*, **84**, 53-62.

Fahrmeir, L. and Tutz, G. (1994). *Multivariate statistical modeling based on generalized linear models.* New York: Springer.

Fienberg, S.E., Bromet, E.J., Follmann, D., Lambert, D. and May, S.M. (1985). Longitudinal analysis of categorical epidemiological data: A study of three mile island. *Enviromental Health Perspectives*, **63**, 241-248.

Fienberg, S.E. (2007). *The Analysis of Cross-classified Categorical Data*, 2nd Edition. Springer-Verlag, New York.

Fitzmaurice, G., Laird, N.M., and Ware J. (2004). *Applied Longitudinal Analysis.* New York: Wiley.

Frees, E. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences.* Cambridge University Press.

Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *Journal of the American Statistical Association*, **93**, 720-729.

Lee, Y. and Nelder, J. A. (1996). Hierarchical generalized linear models (with discussion). *Journal of the Royal Statistical Society, Series B*, **58**, 619–678.

Lipsitz, S. R., Laird, N.M., and Harrington, D. P. (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, **78**, 153-60.

MacDonald, B. W. (1994). Two random effects models for multivariate binary data. *Biometrics*, **50**, 164-172.

Newey, W.K. and McFadden, D. (1993). Estimation in large samples. *Handbook of Econometrics*, **4**, eds. D. McFadden and R. Engler, Amesterdam: North-Holland.

Qaqish, B. F. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, **90**, 455-463.

Sutradhar, B. C. (2003). An review on regression models for discrete longitudinal responses. *Statistical Science*, **18**, 377-393.

Sutradhar, B. C. (2004). On exact Quasilikelihood inference in generalized linear mixed models. *Sankhya*, **66**, 263-291.

Sutradhar, B. C. (2010 a). Inferences in generalized linear longitudinal mixed models. *The Canadian Journal of Statistics*, **38**, 174-196.

Sutradhar, B. C. (2010 b). "Generalized Quasi-likelihood (GQL) Inference" (version 8). *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies*. Freely available at http://statprob.com/encyclopedia/GeneralizedQuasiLikelihoodGQLInferences.html.

Sutradhar, B. C. (2011). *Dynamic Mixed Models for Familial Longitudinal Data.* New York: Springer.

Ten Have, T. R. and Morabia, A. (1999). Mixed effects models with bivariate and univariate association parameters for longitudinal bivariate binary rsponse data. *Biometrics*, **55**, 85-93.

Wedderburn, R. (1974). Quasi-likelihood functions, generalized linear models, and the GaussNewton method. *Biometrika*, **61**, 439-447.

Williamson, J. M., Kim, K. M., and Lipsitz, S. R. (1995). Analyzing bivariate ordinal data using a global odds ratio. *Journal of the American Statistical Association*, **90**, 1432-1437.

Williamson JM, Kim K. (1996). A global odds ratio regression model for bivariate ordered categorical data from ophthalmological studies. *Statistics in Medicine*, **15** 1507-1518.

Yi, G. Y. and Cook, R. J. (2002). Marginal methods for incomplete longitudinal data arising in clusters. *Journal of the American Statistical Association*, **97**, 1071-1080.

Zeger, S. L., Liang, K.Y., and Self, S.G. (1985). The analysis of binary longitudinal data with time-independent covariates. *Biometrika*, **72**, 31-38.