Winter 2011

Modeling and Analysis of Repeated Ordinal Data Using Copula Based Likelihoods and Estimating Equation Methods

Raghavendra Rao Kurada
Old Dominion University

MODELING AND ANALYSIS OF REPEATED ORDINAL DATA USING COPULA BASED LIKELIHOODS AND ESTIMATING EQUATION METHODS

by

Raghavendra Rao Kurada
M.Sc. Jul 2005, University of Hyderabad, India
M.S. May 2010, Old Dominion University, USA

A Dissertation Submitted to the Faculty of
Old Dominion University in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

MATHEMATICS AND STATISTICS

OLD DOMINION UNIVERSITY
December 2011

Approved by:

N. Rao Chaganty (Director)

Dayanand Naik (Member)

Nak-Kyeong Kim (Member)

Larry Filer (Member)

ABSTRACT

MODELING AND ANALYSIS OF REPEATED ORDINAL DATA USING COPULA BASED LIKELIHOODS AND ESTIMATING EQUATION METHODS

Raghavendra Rao Kurada
Old Dominion University, 2011
Director: Dr. N. Rao Chaganty

Repeated or longitudinal ordinal data occur in many fields such as biology, epidemiology, and finance. These data normally are analyzed using both likelihood and non-likelihood methods. The first part of this dissertation discusses the multivariate ordered probit model which is a likelihood method based on latent variables. We show that this latent variable model belong to a very general class of *Copula* models. We use the copula representation for the multivariate ordered probit model to obtain maximum likelihood estimates of the parameters. We apply the methodology in the analysis of real life data examples.

Though likelihood methods are preferable, there are computational challenges implementing them. Alternatives are the non-likelihood models. These are partially specified models, that is, in these models only the functional forms of the marginals are known but joint distributions are unknown. In addition, the dependence among the observations is modeled using an appropriate correlation structure. The second part of the dissertation outlines the estimating equations approach for the analysis of longitudinal ordinal data for these non-likelihood models. We study the asymptotic properties of the estimates for both likelihood and non-likelihood methods. Comparisons based on simulations show that the maximum likelihood estimates arising from copula models are more efficient than the estimates obtained from estimating equations.

The third part of this dissertation describes how ordinal data can be viewed as multinomial random vectors and points out the theoretical challenges in finding restrictions on the correlation parameters for dependent multinomial random vectors.

I dedicate this dissertation to my parents, Ratnamala and Narasimha Rao, my sister, Madhavi, and my first statistics teacher Dr. S. A. Jyothi Rani.

ACKNOWLEDGEMENTS

I wish to express my heartfelt and deepest gratitude to my mentor and dissertation advisor Professor N. Rao Chaganty for his suggestions and guidance during this arduous task. This dissertation would not be possible without his persistent help and constant support in understanding the subject from the initial to the final level, for which I am eternally grateful.

I would like to thank Professors Dayanand Naik, Nak-Kyeong Kim and Larry Filer for serving on my dissertation committee. Thanks are also to Dr. Dayanand Naik for his enormous encouragement during my graduate studies. I am indebted to Professor and Eminent Scholar Daniel M. Dauer in the department of Biological Sciences for arranging research support for me at the Benthic Ecology Lab. In addition, I personally thank Mike Lane at the Benthic lab for his camaraderie during the last four years. It is a pleasure to thank my friends Janardan Upadhyay and P. V. S. S. Vikas who gave me the moral support to complete this laborious work.

Finally, I would also like to thank my fellow students, friends, professors and staff members in the Department of Mathematics and Statistics at Old Dominion University, especially Professor Dorrepaal, for providing a pleasant and stimulating environment. I also wish to acknowledge the staff members Barbara Jeffrey and Sheila Hegwood, for their help during my graduate studies.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I

INTRODUCTION

I.1 ORDINAL LONGITUDINAL RESPONSES

One of the main objectives in statistics is to model the expectation of a response variable as a function of independent or predictor variables. When the response is observed on several occasions on the same subject, then the observed data is known as repeated or longitudinal data. Even though some differences between the definitions of "repeated" and "longitudinal" data exist, we use both terms broadly in the sense of observing multiple measurements on the same subject over time and therefore they are not statistically independent (Davis, 2002). Therefore, repeated measurements models should take into account the dependence of the responses for individual subjects. Although statistical tools used to model continuous longitudinal data are well developed (Laird and Ware, 1982; Ware, 1985), there is no unified methodology to model all types of non-continuous repeated measurements such as binary, count or categorical responses. Multi-category responses are a type of discrete responses which can be classified into two cases, nominal and ordinal. This dissertation addresses the challenging problems associated with modeling longitudinal or repeated *ordinal* categorical responses.

We use the following notation to represent longitudinal data in this dissertation. Let Y_{ij} be a response observed on subject i at time point j and $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2}, \cdots, x_{ijp})'$ be the $p \times 1$ vector of covariates associated with Y_{ij} for $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, t_i$. Assume that observations on different subjects are independent. A typical longitudinal (or clustered) data representation based on this notation is given in Table 1.

In this dissertation we deal with responses Y_{ij} which takes one of the K- ordered categories which can be modeled using a multinomial distribution with K categories. For example, pain status of a patient can be expressed as none (1), mild (2), moderate (3), and severe (4).

Table 1: Typical longitudinal data structure

Subject	Time	Response	Covariates			
	1	Y_{11}	x_{111}	x_{112}	$\cdots$	x_{11p}
	2	Y_{12}	x_{121}	x_{122}	$\cdots$	x_{12p}
1	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t_1	Y_{1t_1}	x_{1t_11}	x_{1t_12}	$\cdots$	x_{1t_1p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	1	Y_{i1}	x_{i11}	x_{i12}	$\cdots$	x_{i1p}
	2	Y_{i2}	x_{i21}	x_{i22}	$\cdots$	x_{i2p}
i	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t_i	Y_{it_i}	x_{it_i1}	x_{it_i2}	$\cdots$	x_{it_ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	1	Y_{n1}	x_{n11}	x_{n12}	$\cdots$	x_{n1p}
	2	Y_{n2}	x_{n21}	x_{n22}	$\cdots$	x_{n2p}
n	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t_n	Y_{nt_n}	x_{nt_n1}	x_{nt_n2}	$\cdots$	x_{nt_np}

I.2 EXAMPLES

We give four real life examples which serve as a motivation for the methods that we develop in this dissertation. Later we analyze these example data sets using those methods.

I.2.1 SKIN CONDITION CLINICAL TRIAL DATA

A clinical trial was conducted to test the efficacy and safety of a new drug for skin conditions in six clinics. Each patient was assigned to one of the two treatments, drug or placebo, and prior to treatment, a response was recorded to determine the initial severity of the skin condition. After the treatment, each patient has three follow-up visits and in each visit a response was observed on a 5-point ordinal response scale that defines the extent of improvement. The ordinal scale is, 1 = Rapidly Improving, 2 = Slowly Improving, 3 = Stable, 4 = Slowly Worsening, 5 = Rapidly Worsening. These data are provided in Table 1 of Stanish et al. (1978). A subset of data is given in Table 2 below.

Table 2: Skin condition clinical trial data

Inv	Trt	Baseline	R_1	R_2	R_3	Inv	Trt	Baseline	R_1	R_2	R_3
5	1	3	3	.	3	⋮	⋮	⋮	⋮	⋮	⋮
5	1	3	3	2	2	11	2	4	4	3	3
5	1	4	3	2	2	11	2	4	2	2	2
5	1	3	2	2	1	11	2	3	4	4	.
5	1	3	3	2	2	11	2	5	4	3	3
5	1	4	2	1	3	11	2	4	4	3	3
5	1	4	1	1	1	11	2	4	3	3	3
5	1	4	1	1	1	11	2	4	2	2	1
5	1	5	5	.	.	11	2	3	4	3	3
5	1	3	1	1	1	11	2	4	4	4	4
5	1	4	4	4	4	11	2	3	4	4	3
5	1	4	3	1	1	11	2	4	4	3	3
⋮	⋮	⋮	⋮	⋮	⋮	11	2	3	4	3	3

R_i = Response at Time i (1: Rapidly Improving; 2: Slowly Improving; 3: Stable; 4: Slowly Worsening; 5: Rapidly Worsening)
Inv = Investigator Identification Number (5, 6, 8, 9, 10, 11)
Trt = Treatment (1 = Test drug, 2 = Placebo)
Baseline = Initial Stage of Disease (3 = Fair, 4 = Poor, 5 = Exacerbation)

I.2.2 SIX CITIES LONGITUDINAL DATA

Ware et al. (1984) studied the respiratory health effects of white children living in six cities in the United States, examining the relationship of respiratory illness in the children exposed to various levels of indoor and outdoor air pollution and other factors, such as parental smoking habits, fuel used for cooking in the child's home, among other things. Lipsitz et al. (1994) analyzed a subset of the data set by modeling wheezing status as a function of age, smoking status of the mother at the particular age of the child, and city of residence for the child. The repeated ordered multinomial response is the wheezing status (no wheeze, wheeze with cold, wheeze apart from cold) of child at ages 9, 10, 11, and 12 years. The wheezing status is modeled as a function of three covariates, namely age (time-varying), smoking status of the mother at the particular age of children (time-varying) and city (time-stationary). These data originally given in Table II in Lipsitz et al. (1994) are reproduced below in Table 3.

Table 3: Data from six cities

		Maternal Smoking				Wheeze			
Case	City	9	10	11	12	9	10	11	12
1	Portage	.	0	0	0	.	1	1	1
2	Kingston	.	.	0	.	.	.	1	.
3	Kingston	0	0	0	0	1	1	1	1
4	Kingston	0	0	0	0	1	1	1	1
5	Portage	.	0	0	0	.	3	2	2
6	Portage	0	0	0	0	1	1	1	1
⋮			⋮					⋮	
327	Portage	0	0	0	0	1	1	1	1
328	Kingston	1	1	1	.	2	1	3	.

The four columns under the maternal smoking and wheeze represent the ages 9, 10, 11, 12.

I.2.3 RESPIRATORY DATA

A randomized controlled clinical trial tested a new treatment for a respiratory disorder is explained in Koch et al. (1989). Each of the 111 patients were randomly assigned to one of the two treatments, active and placebo. During the four follow-up visits, a response on 5-point ordinal scale (0 = terrible, 1 = poor, 2 = fair, 3 = good, 4 = excellent) was recorded for each patient. Miller et al. (1993) analyzed the data by collapsing the 5-point ordinal scale to 3-point ordinal scale (0-1 = poor, 2-3 = good, 4 = excellent). The data are summarized in Table 4.

I.2.4 INSOMNIA CLINICAL TRIAL DATA

Agresti and Natarajan (2001) describe a randomized double blind clinical trial involving two dependent multinomial variables. A pharmaceutical firm compares an active hypnotic drug with a placebo on patients with insomnia. A response to the question "How quickly did you fall asleep after going to bed?" for each patient is recorded at the beginning and at the conclusion of a two-week treatment period. The response to this question has four categories, $<$ 20 minutes, 20m - 30m, 30m -

Table 4: Responses of 111 patients at each of four time points

Visit				No. of patients		Visit				No. of patients	
1	2	3	4	Active	Placebo	1	2	3	4	Active	Placebo
p	p	p	p	1	6	g	g	e	e	1	2
p	p	g	p	1	0	g	e	g	g	0	1
p	p	g	g	0	2	g	e	g	e	2	1
p	g	p	p	1	0	g	e	e	g	3	0
p	g	g	g	0	2	g	e	e	e	7	1
p	g	e	e	1	0	e	p	p	p	0	1
g	p	p	p	0	4	e	p	e	g	1	0
g	p	p	g	0	1	e	g	p	g	0	1
g	p	g	g	1	2	e	g	g	p	1	1
g	g	p	p	1	2	e	g	g	e	1	1
g	g	p	g	2	2	e	g	e	g	0	2
g	g	g	p	4	1	e	g	e	e	0	2
g	g	g	g	8	12	e	e	g	g	2	0
g	g	g	e	2	2	e	e	g	e	2	0
g	g	e	g	1	0	e	e	e	g	3	1
						e	e	e	e	8	7

60m, and > 60m. The data in Table 5 is reproduced from Table 1 in Agresti and Natarajan (2001).

I.3 BACKGROUND

For modeling the repeated or longitudinal ordinal responses several authors have attempted to generalize the methods that are available for repeated binary responses. Some other authors constructed a full-likelihood through copulas for the analysis of repeated ordinal data. In this section we briefly give a survey of these methods. The generalized estimating equation (GEE) approach was applied to polytomous data in Miller et al. (1993) and Lipsitz et al. (1994), both of which generalized the GEE approach that was used for the binary case in Liang and Zeger (1986) and Prentice (1988). To make use of GEE approach for the polytomous data, for each Y_{ij} given in Table 1, we construct a binary choice vector $\boldsymbol{Y}_{ij}^* = (Y_{ij1}, Y_{ij2}, \cdots, Y_{ijK})'$ such that $Y_{ijr} = 1$ if Y_{ij} equals the rth category, and 0 otherwise. Since $\sum_{k=1}^{K} Y_{ijk} = 1$, it is not necessary to work with all the K indicator variables. Instead we drop the

Table 5: Distribution of time to fall asleep

Treatment	Initial occasion	Follow-up occasion < 20	20-30	30-60	> 60
Active	< 20	7	4	1	0
	20-30	11	5	2	2
	30-60	13	23	3	1
	> 60	9	17	13	8
Placebo	< 20	7	4	2	1
	20-30	14	5	1	0
	30-60	6	9	18	2
	> 60	4	11	14	22

last indicator variable, Y_{ijK}, and replace $\boldsymbol{Y}^*_{ij}$ with the binary choice vector $\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \cdots, Y_{ijK-1})'$ of dimension $(K-1)$. For simplicity of notation we assume $t_i = t$ for all i. The t binary vectors observed on subject i could be stacked as a column vector $\boldsymbol{Y}_i = (\boldsymbol{Y}'_{i1}, \boldsymbol{Y}'_{i2}, \cdots, \boldsymbol{Y}'_{it})'$, which is the complete data vector for the ith subject.

Let p_{ijr} be the probability that Y_{ij} equals the rth category. That is $p_{ijr} = P(Y_{ij} = r) = P(Y_{ijr} = 1)$. The expected value of $\boldsymbol{Y}_i$ is given by $\boldsymbol{p}_i = (\boldsymbol{p}'_{i1}, \boldsymbol{p}'_{i2}, \cdots, \boldsymbol{p}'_{it})'$ where $\boldsymbol{p}_{ij} = (p_{ij1}, p_{ij2}, \cdots, p_{ijK-1})'$. Normally this mean vector is modeled as a function of covariates along with an unknown parameter vector $\boldsymbol{\beta}$. In generalized linear models theory this relationship is specified through a link function g, that is, $g(\boldsymbol{p}_i) = \boldsymbol{X}\boldsymbol{\beta}$.

Suppose the covariance matrix of $\boldsymbol{Y}_i$ is denoted by $\boldsymbol{V}_i$ then the elements of $\boldsymbol{V}_i$ are given by the following expressions,

$$\text{Cov}(Y_{ijk}, Y_{ij'k'}) = \begin{cases} p_{ijk}(1-p_{ijk}) & \text{if } j = j', k = k' \\ \\ -p_{ijk}p_{ijk'} & \text{if } j = j', k \neq k' \\ \\ \text{Corr}(Y_{ijk}, Y_{ij'k'})\, \sigma_{ijk}\, \sigma_{ij'k'} & \text{if } j \neq j', \text{ for any } k, k'. \end{cases}$$

where $\sigma_{ijk} = (p_{ijk}(1-p_{ijk}))^{1/2}$ and $\sigma_{ij'k'} = (p_{ij'k'}(1-p_{ij'k'}))^{1/2}$. As it can be seen from the above expressions, except $\text{Corr}(Y_{ijk}, Y_{ij'k'})$ all the quantities depend only on $\boldsymbol{p}_{ij}$. These between-occasion category-specific correlations can be modeled in several ways in terms of an unknown parameter vector $\boldsymbol{\lambda}$. Notice that if we assume the measurements on same subject at different occasions are independent then $\boldsymbol{V}_i$ is a block diagonal matrix. On the other hand, the "saturated" model can be obtained by assuming a unique parameter for each between-occasion category-specific correlations. Several models can be constructed by assuming different structures for $\text{Corr}(Y_{ijk}, Y_{ij'k'})$ that have more number of parameters than the independence model but fewer number of parameters than the "saturated" model. The $\boldsymbol{V}_i$ obtained with this type of modeling is commonly known as "working" covariance matrix which depends on the parameter vectors $\boldsymbol{\lambda}$ and $\boldsymbol{\beta}$. In the literature several authors suggested methods for estimating $\boldsymbol{\lambda}$ using different approaches.

Based on the above setup for the polytomous data, if $\boldsymbol{\lambda}$ is known (so is $\boldsymbol{V}_i$), the GEE estimator of $\boldsymbol{\beta}$ is obtained solving the estimating equations

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{p}_i) \;=\; \boldsymbol{0}, \tag{1}$$

where $\boldsymbol{D}_i' = \partial \boldsymbol{p}_i / \partial \boldsymbol{\beta}'$. Suppose $\widehat{\boldsymbol{\beta}}$ is the solution of equation (1). Then we can show that $\widehat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$ and the asymptotic distribution of $\sqrt{n}\,(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is normal with mean $\boldsymbol{0}$ and covariance $\boldsymbol{V}_{\boldsymbol{\beta}}$ given by

$$\boldsymbol{V}_{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \right)^{-1} \left(\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \,\text{Cov}(\boldsymbol{Y}_i)\, \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \right) \left(\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \boldsymbol{D}_i \right)^{-1}$$

where $\text{Cov}(\boldsymbol{Y}_i)$ is true covariance matrix of $\boldsymbol{Y}_i$.

In general the vector $\boldsymbol{\lambda}$ is unknown and we need to estimate it to construct $\boldsymbol{V}_i$ in equation (1). Miller et al. (1993) suggested estimating $\boldsymbol{\lambda}$ using another set of equations based on sample correlations. Suppose $\boldsymbol{Z}_i = \boldsymbol{Z}_i(\boldsymbol{\beta})$ is a $((K-1)^2 t(t-1))/2 \times 1$ vector (referred as "working" correlation estimates) defined as

$$\boldsymbol{Z}_i(\boldsymbol{\beta}) = (Z_{i(11)(21)}(\boldsymbol{\beta}), Z_{i(11)(22)}(\boldsymbol{\beta}), \cdots, Z_{i(11)(2(K-1))}(\boldsymbol{\beta}), \cdots, Z_{i((t-1)(K-1))(t(K-1))}(\boldsymbol{\beta}))'$$

where

$$Z_{i(jk)(j'k')}(\boldsymbol{\beta}) = \frac{(y_{ijk} - \pi_{ijk}(\boldsymbol{\beta}))(y_{ij'k'} - \pi_{ij'k'}(\boldsymbol{\beta}))}{(\pi_{ijk}(\boldsymbol{\beta})(1 - \pi_{ijk}(\boldsymbol{\beta}))\pi_{ij'k'}(\boldsymbol{\beta})(1 - \pi_{ij'k'}(\boldsymbol{\beta})))^{1/2}},$$

for any $j \neq j'$ and k, k' between 1 and K. Note that $E(Z_{i(jk)(j'k')}(\boldsymbol{\beta})) = \mathrm{Corr}(Y_{ijk}, Y_{ij'k'}) = \eta_{i(jk)(j'k')}(\boldsymbol{\lambda})$. Miller et al. (1993) considered correlations $\boldsymbol{f}_{i(jk)(j'k')}$ induced by Fisher transformation given by the relation

$$\eta_{i(jk)(j'k')}(\boldsymbol{\lambda}) = \frac{\exp(\boldsymbol{\lambda}' \boldsymbol{f}_{i(jk)(j'k')} - 1)}{\exp(\boldsymbol{\lambda}' \boldsymbol{f}_{i(jk)(j'k')} + 1)}.$$

To estimate $\boldsymbol{\lambda}$ based on these transformed correlations, Miller et al. (1993) used the following second set of equations:

$$U(\boldsymbol{\lambda}) = \sum_{i=1}^{n} \boldsymbol{E}_i' \boldsymbol{W}_i^{-1}(\boldsymbol{Z}_i - \boldsymbol{\eta}_i(\boldsymbol{\lambda})) \ = \ \boldsymbol{0}, \tag{2}$$

where $\boldsymbol{E}_i = \partial \boldsymbol{\eta}_i(\boldsymbol{\lambda})/\partial \boldsymbol{\lambda}$, $\boldsymbol{W}_i = \mathrm{Cov}(\boldsymbol{Z}_i)$ is the covariance matrix that depends on a "working" covariance assumption for $\boldsymbol{Z}_i(\boldsymbol{\beta})$, and

$$\boldsymbol{\eta}_i(\boldsymbol{\lambda}) = (\eta_{i(11)(21)}(\boldsymbol{\lambda}), \eta_{i(11)(22)}(\boldsymbol{\lambda}), \cdots, \eta_{i(11)(2(K-1))}(\boldsymbol{\lambda}), \cdots, \eta_{i((t-1)(K-1))(t(K-1))}(\boldsymbol{\lambda}))'.$$

The estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are obtained recursively solving equations (1) and (2) until convergence.

Lipsitz et al. (1994) suggested an alternative estimate of $\boldsymbol{\lambda}$ using method of moments. In their paper they considered several structures for $\boldsymbol{\rho}_{ijj'}$, which is the correlation matrix between $\boldsymbol{Y}_{ij}$ and $\boldsymbol{Y}_{ij'}$. The structures include: (i) compound symmetry, which is defined as $\boldsymbol{\rho}_{ijj'} = \boldsymbol{\rho}$ for all $j \neq j'$, (ii) one-dependence, which is defined as $\boldsymbol{\rho}_{ij,j+1} = \boldsymbol{\rho}_j$ for all $j = 1, 2, \cdots, t-1$, and $\boldsymbol{\rho}_{ijj'} = \boldsymbol{0}$ otherwise, (iii) banded, which is defined as $\boldsymbol{\rho}_{ijj'} = \boldsymbol{\rho}_\tau$ when $|j' - j| = \tau$ for $\tau = 1, 2, \cdots, t-1$, and (iv) unstructured, which is defined as $\boldsymbol{\rho}_{ijj'} = \boldsymbol{\rho}_{jj'}$. Lipsitz et al. (1994) gave estimates of the correlations for each structure using method of moments. Define the "residual" for Y_{ijk} as

$$e_{ijk} = \frac{Y_{ijk} - p_{ijk}}{(p_{ijk}(1 - p_{ijk}))^{1/2}}.$$

Then clearly $E(e_{ijk}e_{ij'k'}) = \mathrm{Corr}(Y_{ijk}, Y_{ij'k'})$. In vector form the residuals can be written as $\boldsymbol{e}_{ij} = \boldsymbol{A}_{ij}^{-1/2}(\boldsymbol{Y}_{ij} - \boldsymbol{p}_{ij})$ where $\boldsymbol{A}_{ij}$ is a diagonal matrix with $\mathrm{Var}(Y_{ijk})$'s on the main diagonal. Using this notation, for each structure the moment estimators given in Lipsitz et al. (1994) are as follows.

1. Compound symmetry: For this structure $\boldsymbol{\rho}_{ijj'} = \boldsymbol{\rho}$ for all $j \neq j'$, that is $\boldsymbol{\rho} = E(\boldsymbol{e}_{ij}\boldsymbol{e}_{ij}')$ for all pairs ij and ij'. Therefore an estimate of the common correlation is given by

$$\widehat{\boldsymbol{\rho}} = \frac{\sum_{i=1}^{n}\sum_{j>j'}\widehat{\boldsymbol{e}}_{ij}\widehat{\boldsymbol{e}}_{ij'}'}{\left(\sum_{i=1}^{n}\frac{1}{2}t(t-1)\right)-p}$$

where $\widehat{\boldsymbol{e}}_{ij} = \widehat{\boldsymbol{A}}^{-1/2}(\boldsymbol{Y}_{ij} - \widehat{\boldsymbol{p}}_{ij})$. Recall that p is the number of covariates considered in the model.

2. One-dependence: Here $\boldsymbol{\rho}_{ij,j+1} = \boldsymbol{\rho}_j$ for all $j = 1, 2, \cdots, t-1$, and $\boldsymbol{\rho}_{ijj'} = \mathbf{0}$ otherwise. Since $\boldsymbol{\rho}_j = E(\boldsymbol{e}_{ij}\boldsymbol{e}_{i,j+1}')$ for $j = 1, 2, \cdots, t-1$, the moment estimate is given by

$$\widehat{\boldsymbol{\rho}}_j = \frac{\sum_{i=1}^{n}\widehat{\boldsymbol{e}}_{ij}\widehat{\boldsymbol{e}}_{i,j+1}'}{n-p}.$$

3. Banded: Here $\boldsymbol{\rho}_{ijj'} = \boldsymbol{\rho}_\tau$ when $|j'-j| = \tau$ for $\tau = 1, 2, \cdots, t-1$. Since $\boldsymbol{\rho}_\tau = E(\boldsymbol{e}_{ij}\boldsymbol{e}_{i,j+\tau}')$, a moment estimate of $\boldsymbol{\rho}_\tau$ is

$$\widehat{\boldsymbol{\rho}}_\tau = \frac{\sum_{i=1}^{n}\sum_{j=1}^{t-\tau}\widehat{\boldsymbol{e}}_{ij}\widehat{\boldsymbol{e}}_{i,j+\tau}'}{n(t-\tau)-p},$$

for $\tau = 1, 2, \cdots, t-1$.

4. Unstructured: Here $\boldsymbol{\rho}_{ijj'} = \boldsymbol{\rho}_{jj'} = E(\boldsymbol{e}_{ij}\boldsymbol{e}_{ij'}')$, and the moment estimate is

$$\widehat{\boldsymbol{\rho}}_{jj'} = \frac{\sum_{i=1}^{n}\widehat{\boldsymbol{e}}_{ij}\widehat{\boldsymbol{e}}_{ij'}'}{n-p}.$$

Lipsitz et al. (1994) used the above moment estimators to update $\boldsymbol{V}_i$ at each iteration when solving equation (1) using numerical routines. Lumley (1996) modeled the associations between the repeated ordinal measurements for polytomous data using cumulative odds ratios, rather than correlations as in the GEE framework.

Complementing the estimating equations approach, Meester and MacKay (1994) outlined a copula-based parametric approach for analyzing repeated-measure ordered categorical data featuring compound symmetry dependence. Copula based techniques are at the cutting edge for constructing a joint distribution for a given set of marginal distribution functions. A detailed discussion of copulas is given later in this dissertation in Chapter II. In a nutshell as stated in Meester and MacKay (1994), a copula, $C(.)$, is a multivariate cumulative distribution function on $[0,1]^t$ with uniform marginals. Meester and MacKay (1994) discussed the analysis of ordinal data using a bivariate copula that belongs to Frank's family of copulas given by,

$$C_\alpha(u_1, u_2) = \phi_\alpha^{-1}(\phi_\alpha(u_1) + \phi_\alpha(u_2)),\ 0 \leq u_1, u_2 \leq 1, \tag{3}$$

where

$$\phi_\alpha(t) = -\alpha^{-1} \log[(e^{-\alpha} - 1)/(e^{-\alpha t} - 1)] \tag{4}$$

and $-\infty < \alpha < \infty$ indexes the family.

Suppose Y_1 and Y_2 have marginal distribution functions F_1 and F_2 that depend on an unknown parameter vector $\boldsymbol{\theta}$. Then by substituting, $u_1 = F_1(y_1)$ and $u_2 = F_2(y_2)$ in equation (3) we get a joint cumulative distribution function $F_{\boldsymbol{Y}}(y_1, y_2; \boldsymbol{\eta})$ for $\boldsymbol{Y} = (Y_1,\ Y_2)$ with marginals F_1 and F_2. Here $\boldsymbol{\eta} = (\boldsymbol{\theta}', \alpha)'$ is the parameter vector. The bivariate probability mass function of $\boldsymbol{Y}$ is given by

$$P_{\boldsymbol{\eta}}(Y_1 = y_1,\ Y_2 = y_2) = F_{\boldsymbol{Y}}(y_1, y_2; \boldsymbol{\eta}) - F_{\boldsymbol{Y}}(y_1 - 1, y_2; \boldsymbol{\eta}) - F_{\boldsymbol{Y}}(y_1, y_2 - 1; \boldsymbol{\eta}) + F_{\boldsymbol{Y}}(y_1 - 1, y_2 - 1; \boldsymbol{\eta}).$$

The above probability mass function could be used to construct a likelihood which can be maximized to get an estimate of $\boldsymbol{\eta} = (\boldsymbol{\theta}', \alpha)'$. The maximization could be done via the method of scoring using either the expected or the observed information. The initial estimate of $\boldsymbol{\theta}$ for the iterative scoring method can be obtained from an independence model ($\alpha = 0$). Also, an initial estimate of α is obtained from the approximate relation,

$$\rho_s \approx (1 - \alpha e^{-\alpha/2} - e^{-\alpha})(e^{-\alpha/2} - 1)^{-2}$$

where ρ_s is the sample Spearman's correlation.

Meester and MacKay (1994) have extended their results to the general case where there are $t > 2$ repeated measurements. They used a generalized Frank's copula given by

$$C(\boldsymbol{u}) = C(u_1, u_2, \cdots, u_t) = \phi_\alpha^{-1}\left(\sum_{i=1}^{t} \phi_\alpha(u_i)\right), \ \boldsymbol{u} \in [0,1]^t$$

where $\phi_\alpha(t)$ is defined in (4). In the next section we give an overview of this dissertation.

I.4 OVERVIEW OF THE DISSERTATION

In Chapter II, we introduce the multivariate ordered probit model which is a likelihood approach for modeling repeated ordered responses. We show that the ordered probit model belongs to a very general class of Multivariate Copula Discrete (MCD) models. We provide a brief summary of the theory of copulas and how they are used to construct joint distributions with specified marginals, with special emphasis on the MCD models. Next we discuss the likelihood estimation for the MCD models, and derive the score equations for both the regression and the latent correlation parameters. These score equations are solved to get the maximum likelihood estimates using Quasi-Newton numerical method given as Algorithm 21 in Nash (1979). The R code that we developed is used to analyze the four examples discussed earlier in this chapter.

In Chapter III, we introduce Generalized Estimating Equations (GEE) (see Lipsitz et al., 1994), a non-likelihood approach for analyzing the repeated ordered responses. This approach requires only the specification of the link function which relates expectation of the responses with predictors and the dependence nature of the repeated responses. The method estimates the correlation between the responses on the same subject by moment estimators. Despite its simplicity the GEE method has several drawbacks, see Sabo and Chaganty (2010). In this chapter we also study large sample efficiencies between the multivariate ordered probit and the GEE estimates. The efficiency calculations show that the ML estimates are uniformly more efficient than GEE estimates for any choice of dependence parameters when the true model is the multivariate ordered probit model.

In Chapter IV, we study the restrictions on the correlations for dependent ordered categorical random variables. First we describe possible correlations that can arise when we view the categorical response as a multinomial random vector. We derive the ranges of these correlations for two dependent multinomial random vectors with specified means. Some extensions are given for three correlated multinomial random vectors assuming a parsimonious structure. Our results can be viewed as a generalization of the results given in Chaganty and Joe (2006).

Finally we close this dissertation with an Appendix that contains proofs and derivatives for the multivariate ordered probit model and an R program that uses those derivatives to fit the model for real life data.

CHAPTER II

LIKELIHOOD INFERENCE

II.1 LATENT VARIABLE MODELING

Ashford and Sowden (1970) presented a convenient way of generating an ordinal response from a continuous *latent* random variable. More specifically, given a continuous random variable Z, we can generate a K-ordered categorical random variable Y by categorizing Z using $K-1$ thresholds. Thus a probability model for Y can be obtained by partitioning the range of the unobserved continuous random variable Z using these ordered thresholds. This approach is widely known as latent variable modeling.

II.2 MULTIVARIATE ORDERED PROBIT MODEL

When the ordinal responses are generated from latent random variables that are Gaussian then the resulting distribution for the ordered response Y is known as the *Ordered Probit Model.* Suppose that Z is a latent variable that is normally distributed. Let $\gamma(k)$, $1 \leq k \leq K-1$, be the ordered thresholds $-\infty = \gamma(0) < \gamma(1) < \gamma(2) < \cdots < \gamma(K-1) < \gamma(K) = \infty$. Then the ordinal variable Y has the stochastic representation,

$$Y = \begin{cases} 1 & \text{if } \gamma(0) \leq Z < \gamma(1) \\ 2 & \text{if } \gamma(1) \leq Z < \gamma(2) \\ \vdots & \qquad \vdots \\ K & \text{if } \gamma(K-1) \leq Z < \gamma(K). \end{cases}$$

The thresholds $\gamma(k) = \alpha_k + \boldsymbol{x}'\boldsymbol{\beta}$ are assumed to depend on the covariates $\boldsymbol{x}$. Here $\boldsymbol{\beta}$ is the regression parameter and α_k are level specific unknown parameters. Notice that the monotonic increasing nature of the thresholds accounts for the ordered nature of observed outcomes.

Suppose $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_t)'$ is a vector of t dependent ordinal response variables.

We assume that corresponding to each Y_j there is a underlying latent random variable Z_j and ordered thresholds $\gamma_j(k) = \alpha_k + \boldsymbol{x}_j'\boldsymbol{\beta}$, which are functions of the covariate vector $\boldsymbol{x}_j$ that is associated with Y_j. The *Multivariate Ordered Probit Model* is obtained by assuming that $\boldsymbol{Z} = (Z_1, Z_2, \cdots, Z_t)'$ is distributed as multivariate normal (MVN) with mean $\boldsymbol{0}$ and covariance matrix $\boldsymbol{R}$. For model identification we assume that Z_i's have unit variance, that is, $\boldsymbol{R}$ is a correlation matrix.

II.2.1 LIKELIHOOD CONSTRUCTION

Based on the assumption that $\boldsymbol{Z}$ follows t-variate normal distribution, the joint probability mass function of $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_t)'$ can be written as

$$\begin{aligned}
\pi_t(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{R}) &= P\big(Y_1 = y_1, Y_2 = y_2, ..., Y_t = y_t\big) \\
&= P\big(\gamma_1(y_1 - 1) \leq Z_1 < \gamma_1(y_1), \cdots, \gamma_t(y_t - 1) \leq Z_t < \gamma_t(y_t)\big) \\
&= \int_{\gamma_1(y_1-1)}^{\gamma_1(y_1)} \int_{\gamma_2(y_2-1)}^{\gamma_2(y_2)} \cdots \int_{\gamma_t(y_t-1)}^{\gamma_t(y_t)} \phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R})\, d\boldsymbol{z} \qquad (5)
\end{aligned}$$

where $\boldsymbol{y} = (y_1, y_2, \ldots, y_t)'$ and $\phi_t(\boldsymbol{z}; \boldsymbol{0}, \boldsymbol{R})$ is the density function of the multivariate normal distribution with mean $\boldsymbol{0}$ and correlation matrix $\boldsymbol{R}$. The correlation matrix $\boldsymbol{R}$ is commonly known as the latent correlation matrix, and it could be unstructured or a structured matrix such as AR(1), compound symmetry.

Several authors provided numerical approximations to the multiple integral given in equation (5). Two widely used approximations for the multidimensional integral (5) are due to Genz (1992, 1993) and Joe (1995). We used Joe (1995)'s approximation to numerically compute the probability mass function (5).

Given t-dimensional vector of observations $\boldsymbol{y}_i$ on n independent subjects, the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \pi_t(\boldsymbol{y}_i; \boldsymbol{0}, \boldsymbol{R}),$$

and the log-likelihood is

$$
\begin{aligned}
l(\boldsymbol{\theta}) &= \log L(\boldsymbol{\theta}; \boldsymbol{Y}) = \sum_{i=1}^{n} \log\big(\pi_t(\boldsymbol{y}_i; \mathbf{0}, \boldsymbol{R})\big) \\
&= \sum_{i=1}^{n} \log \int_{\gamma_{i1}(y_{i1}-1)}^{\gamma_{i1}(y_{i1})} \int_{\gamma_{i2}(y_{i2}-1)}^{\gamma_{i2}(y_{i2})} \cdots \int_{\gamma_{it}(y_{it}-1)}^{\gamma_{it}(y_{it})} \phi_t(\boldsymbol{z}_i; \mathbf{0}, \boldsymbol{R})\, d\boldsymbol{z}_i \qquad (6)
\end{aligned}
$$

where $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\rho})$. Here $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_{K-1})'$ is the vector of threshold intercepts, $\boldsymbol{\beta}$ is the regression parameter vector, and $\boldsymbol{\rho}$ is a parameter vector that characterizes the correlation matrix $\boldsymbol{R} = \boldsymbol{R}(\boldsymbol{\rho})$.

II.2.2 ESTIMATION

Since the likelihood (6) is non-linear, we need a numerical optimization routine to obtain the maximum likelihood estimator of $\boldsymbol{\theta}$. A good choice is the quasi-Newton (or variable metric) algorithm given in Nash (1979, p. 192). The algorithm can be described as follows:

Step 1. Start with an initial estimate $\widehat{\boldsymbol{\theta}}_{int}$ of $\boldsymbol{\theta}$.

Step 2. At the ith step compute $\widehat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i - c\,B(\widehat{\boldsymbol{\theta}}_i) g(\widehat{\boldsymbol{\theta}}_i)$ where $g(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and $B(\boldsymbol{\theta})$ is an approximation to the inverse of Hessian matrix, $[\partial^2 l(\boldsymbol{\theta})/\partial\theta_j \partial\theta_k]^{-1}$, and c is a constant. See Algorithm 21 in Nash (1979) for more details.

Step 3. Repeat Step 2 until $\widehat{\boldsymbol{\theta}}_{i+1} \cong \widehat{\boldsymbol{\theta}}_i$ and take $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{i+1}$ as the MLE of $\boldsymbol{\theta}$.

The Mprobit package in R software gives the MLE of $\boldsymbol{\theta}$ for multivariate ordered probit model. However, this package uses the numerical derivatives, and not the analytical derivatives, to calculate $g(\widehat{\boldsymbol{\theta}})$ where $g(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ in Step 2 of the above quasi-Newton algorithm.

II.2.3 ALTERNATIVE REPRESENTATION

The probability mass function $\pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R})$ of $\boldsymbol{Y}$, given in (5) is essentially a rectangle probability of the t-variate multivariate normal distribution. This rectangle probability can be expressed as a function of the multivariate normal cumulative distribution function. Thus an alternative expression for equation (5) is

$$\pi_t(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{R}) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2} \dots \sum_{i_t=1}^{2}(-1)^{i_1+i_2+\dots+i_t}\Phi_t(b_{1i_1}, b_{2i_2}, \dots, b_{ti_t}; \boldsymbol{0}, \boldsymbol{R}) \qquad (7)$$

where $b_{j1} = \gamma_j(y_j - 1)$; $b_{j2} = \gamma_j(y_j)$ and Φ_t is the cumulative distribution function of t-variate normal distribution. The equivalence of (5) and (7) can easily be verified, for example, when $t = 2$. The expression (7) is convenient for finding the analytical derivatives of the log-likelihood (6). Moreover, we can see that equation (7) is a special case of general class of *Multivariate Copula Discrete* (MCD) models. Before presenting a description of MCD models, we give a brief introduction to copula theory in the next section.

II.3 COPULAS

II.3.1 INTRODUCTION

One of the modern techniques of constructing joint distributions with specified marginal distributions is through *copulas*, see Joe (1997). Copula is a multivariate distribution with univariate margins that are uniform on the interval [0,1]. The basic idea behind the construction of a multivariate distribution using copulas is the following. It is well known that for any continuous random variable X with distribution function $F(\cdot)$, the transformation $F(X)$ follows a uniform distribution on [0, 1]. As a result, a joint distribution with specified marginals can be constructed using a multivariate distribution with uniform marginals.

Definition. A t-dimension copula is a function $C : [0,1]^t \to [0,1]$ with the following properties.

1. $C(1, \cdots, 1, a_i, 1, \cdots, 1) = a_i \; \forall \, i = 1, 2, \cdots, t$ and $a_i \in [0,1]$.

2. $C(a_1, a_2, \cdots, a_t) = 0$ if at least one $a_i = 0$ for $i = 1, 2, \cdots, t$.

3. For any $a_{i1}, a_{i2} \in [0,1]$ with $a_{i1} \leq a_{i2}$, for $i = 1, 2, \cdots, t$,

$$\sum_{j_1=1}^{2}\sum_{j_2=1}^{2} \cdots \sum_{j_t=1}^{2}(-1)^{j_1+j_2+\dots+j_t} C(a_{1j_1}, a_{2j_2}, \cdots, a_{tj_t}) \geq 0.$$

II.3.2 EXAMPLES

Below are some examples of some well known and widely used copulas.

Example 1. The *Independence Copula* is a function given by

$$C(a_1, a_2, \cdots, a_t) = \prod_{j=1}^{t} a_i \tag{8}$$

Example 2. The *Comonotonicity Copula* is a function given by

$$C(a_1, a_2, \cdots, a_t) = \min\{a_1, a_2, \cdots, a_t\} \tag{9}$$

Example 3. When $t = 2$, the *Countermonotonicity Copula* is a function given by

$$C(a_1, a_2) = \max\{a_1 + a_2 - 1, 0\} \tag{10}$$

Example 4. The *Multivariate Normal (Gaussian) Copula* with latent correlation matrix $\boldsymbol{R}$ is a function given by

$$C(a_1, a_2, \cdots, a_t; \boldsymbol{R}) = \Phi_t(\Phi^{-1}(a_1), \Phi^{-1}(a_2), \cdots, \Phi^{-1}(a_t); \mathbf{0}, \boldsymbol{R}) \tag{11}$$

where Φ is the cumulative distribution function of standard normal distribution and $\Phi_t(.; \boldsymbol{\mu}, \Sigma)$ is the distribution function of a t-variate normal with mean $\boldsymbol{\mu}$ and variance covariance matrix Σ given by

$$\Phi_t(z_1, z_2, \cdots, z_t) = \int_{-\infty}^{z_t} \cdots \int_{-\infty}^{z_1} \frac{1}{(2\pi)^{t/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{u}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{u}-\boldsymbol{\mu})} du_1 \ldots du_t \tag{12}$$

Note that the t-dimensional normal copula reduces to the Independence copula when $\Sigma = I$. We will be using this copula later in Section II.4.2.

Example 5. Let M be a univariate distribution function of a positive random variable. Note that $M(0) = 0$. Let

$$\phi(a) = \int_0^\infty e^{-au}\, dM(u),\ a \geq 0$$

be the Laplace transform of M. The t-dimensional *Archimedean Copula* is defined as

$$C(a_1, a_2, \cdots, a_t) = \phi\left(\sum_{j=1}^{t} \phi^{-1}(a_j)\right). \tag{13}$$

This copula is useful to model compound symmetry dependence.

A fundamental result for copulas is Sklar's theorem given below.

THEOREM 1. *(Sklar's Theorem). Let $Y_1, Y_2, \cdots, Y_t$ be random variables with marginal distribution functions $F_1, F_2, \cdots, F_t$ and joint cumulative distribution function F. Then the following hold.*

1. *There exists a t-dimensional copula C such that for all $y_1, y_2, \cdots, y_t \in (-\infty, \infty)$,*

$$F(y_1, y_2, \cdots, y_t) = C(F_1(y_1), F_2(y_2), \cdots, F_t(y_t)).$$

2. *If $Y_1, Y_2, \cdots, Y_t$ are continuous random variables defined on real line, then C is unique. Otherwise, C is uniquely determined on the t dimensional rectangle $Range(F_1) \times Range(F_2) \times \cdots \times Range(F_t)$.*

A more comprehensive discussion of the theory of copulas is in the classic books by Joe (1997), Nelson (2006) and Jaworski et al. (2010).

II.3.3 MULTIVARIATE PROBABILITY DENSITY FUNCTIONS

Suppose F_i is a marginal cumulative distribution function of Y_i, $i = 1, 2, \cdots, t$. For a copula model, the cumulative distribution function of a random vector $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_t)'$ is given by

$$F(\boldsymbol{y}) = C(F_1(y_1), F_2(y_2), \cdots, F_t(y_t)), \tag{14}$$

where C is a t-dimensional copula. If $\boldsymbol{Y}$ is continuous then its probability density function is

$$f(\boldsymbol{y}) = \prod_{i=1}^{t} f_i(y_i) c(F_1(y_1), F_2(y_2), \cdots, F_t(y_t)), \tag{15}$$

where $f_i(y) = \partial F_i(y)/\partial y$ is the marginal probability density function of Y_i and

$$c(a_1, a_2, \cdots, a_t) = \frac{\partial^t C(a_1, a_2, \cdots, a_t)}{\partial a_1 \partial a_2 \cdots \partial a_t}$$

is the density of copula C. For discrete random variables, the multivariate probability mass function of $\boldsymbol{Y}$ is given by

$$P(y_1, y_2, \cdots, y_t) = \sum_{j_1=1}^{2}\sum_{j_2=1}^{2}\cdots\sum_{j_t=1}^{2}(-1)^{j_1+j_2+\cdots+j_t} C(a_{1j_1}, a_{2j_2}, \cdots, a_{tj_t}), \tag{16}$$

where $a_{i1}(y_i) = F_i(y_i^-)$ and $a_{i2}(y_i) = F_i(y_i)$. Here $F_i(y_i^-)$ is the left hand limit of F_i at y_i. When the support of F_i is the set of integers then $F_i(y_i^-) = F_i(y_i - 1)$. Equation (16) is the probability mass function of a *Multivariate Copula Discrete* (MCD) model.

II.4 MULTIVARIATE COPULA DISCRETE MODELS FOR ORDINAL DATA

II.4.1 INTRODUCTION

Suppose $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_t)'$ is a t repeated ordinal response vector with each Y_j being an ordinal response random variate with K categories. Denote $p_{j,k}$ as the probability that Y_j takes the kth ordered category. Then define,

$$G_j(y_j) = \begin{cases} 0 & \text{if } y_j < 1 \\ \sum_{k=1}^{\lfloor y_j \rfloor} p_{j,k} & \text{if } 1 \le y_j < K \\ 1 & \text{if } y_j \ge K \end{cases}$$

where $\lfloor y_j \rfloor$ means the largest integer less than or equal to y_j. If we assume this is a distribution function of Y_j then for any given t dimensional copula, $C(a_1, a_2, \cdots, a_t)$, $C(G_1(a_1), G_2(a_2), \cdots, G_t(a_t); \boldsymbol{\theta})$ is a well defined joint cumulative distribution function for the ordinal random vector $\boldsymbol{Y}$. Using this copula based joint distribution, the joint probability mass function can be written as,

$$P(y_1, y_2, \cdots, y_t) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_t=1}^{2}(-1)^{i_1+i_2+\cdots+i_t} C(a_{1i_1}, a_{2i_2}, \cdots, a_{ti_t}; \boldsymbol{\theta}). \tag{17}$$

where $a_{j1}(y_j) = G_j(y_j - 1)$ and $a_{j2}(y_j) = G_j(y_j)$. This method of constructing joint probability mass functions for an ordinal response vector is known as *Multivariate Copula Discrete Model for Ordinal Data.*

Consider the following re-parametrization for $p_{j,k}$. Let Z_j be a continuous random variable with distribution function F_j and define $G_j(y_j) = F_j(\gamma_j(y_j))$. Then $p_{j,k} = F_j(\gamma_j(y_j^{(k)})) - F_j(\gamma_j(y_j^{(k-1)}))$ where $y_j^{(k)} = k$. This is equivalent to

$$Y_j = \begin{cases} 1 & \text{if } \gamma_j(0) \le Z_j < \gamma_j(1) \\ \\ 2 & \text{if } \gamma_j(1) \le Z_j < \gamma_j(2) \\ \vdots & \qquad \vdots \\ K & \text{if } \gamma_j(K-1) \le Z_j < \gamma_j(K) \end{cases}$$

where $-\infty = \gamma_j(0) < \gamma_j(1) < \cdots < \gamma_j(K-1) < \gamma_j(K) = \infty$ are constants for all $j = 1, 2, \cdots, t$. The above is simply a latent variable model for the ordinal response. Here $\gamma_j(y_j)$ is called the y_jth cut off point for the random variable Z_j. In general, $\gamma_j(y_j) = \alpha_{y_j} + \boldsymbol{x}_j'\boldsymbol{\beta}$ are functions of the covariates. This approach gives rise several models for the ordinal responses. For example, there are multiple choices for F_j such as logistic, normal, extreme value, gamma, lognormal, etc. Similarly, we have several choices for the copula such as multivariate normal copula, mixture of max-id copula and so on. In the following section we study the multivariate normal copula model.

II.4.2 MULTIVARIATE NORMAL COPULA MODELS

Recall that the multivariate normal (or) Gaussian copula is

$$C(a_1, a_2, \cdots, a_t; \boldsymbol{R}) = \Phi_t(\Phi^{-1}(a_1), \Phi^{-1}(a_2), \cdots, \Phi^{-1}(a_t); \boldsymbol{0}, \boldsymbol{R}) \tag{18}$$

where Φ^{-1} is inverse of a univariate standard normal distribution $\Phi(.; 0, 1)$ and $\Phi_t(.; \boldsymbol{\mu}, \Sigma)$ is a t-variate normal cumulative distribution function with mean $\boldsymbol{\mu}$ and variance covariance matrix Σ.

In MCD models for ordinal response, if we choose F_j as standard normal cumulative distribution function, then it is equivalent to choosing $G_j(y_j) = \Phi(\gamma_j(y_j))$. Furthermore, if we chose the multivariate normal copula then the joint probability mass function for t repeated ordinal response vector is

$$P(y_1, y_2, \cdots, y_t) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_t=1}^{2}(-1)^{i_1+i_2+\quad+i_t}C(a_{1i_1}, a_{2i_2}, \cdots, a_{ti_t}; \boldsymbol{\theta})$$

$$= \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_t=1}^{2}(-1)^{i_1+i_2+\quad+i_t}\Phi_t(\Phi^{-1}(a_{1i_1}), \Phi^{-1}(a_{2i_2}), \cdots, \Phi^{-1}(a_{ti_t}); \mathbf{0}, \boldsymbol{R}) \tag{19}$$

where $a_{j1}(y_j) = G_j(y_j - 1) = \Phi(\gamma_j(y_j - 1))$ and $a_{j2}(y_j) = G_j(y_j) = \Phi(\gamma_j(y_j))$. With this substitution, the joint probability mass function becomes

$$P(y_1, y_2, \cdots, y_t) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}...\sum_{i_t=1}^{2}(-1)^{i_1+i_2+\quad+i_t}\Phi_t(b_{1i_1}, b_{2i_2}, ..., b_{ti_t}; \mathbf{0}, \boldsymbol{R}) \tag{20}$$

with $b_{j1} = \gamma_j(y_j - 1)$ and $b_{j2} = \gamma_j(y_j)$. Notice that equation (20) is same as equation (7) which is an alternative representation of the multivariate ordered probit model. If we apply an MCD model using a multivariate normal copula and assume Z_j is standard normal then the MCD model is equivalent to the multivariate ordered probit model. We call this model as Gaussian copula based ordered probit model. On the other hand if Z_j is standard logistic then the resulting MCD model is known as Gaussian copula based ordered logit model.

II.4.3 LIKELIHOOD ESTIMATION

The joint distribution of a t repeated ordered response vector constructed using a copula model is useful to construct a likelihood when we have a sample of independent observations on n subjects. Recall that we need numerical routines for the maximum likelihood estimation. These numerical methods can be run efficiently if there are analytical expressions for the first order derivatives of the log-likelihood. One major advantage of representing the latent variable models through copulas is that we can derive analytical expressions for the derivatives. Here in this section we provide the first derivatives of the log-likelihood function for the MCD models based on Gaussian copula. Below we provide some notation to obtain the first derivatives of $l(\boldsymbol{\theta})$, with respect to $\boldsymbol{\theta}$, that are required by the optimization routines. We introduce some notation first.

For a vector $\boldsymbol{y} = (y_1, y_2, \cdots, y_t)'$, we denote by $\boldsymbol{y}_{-j}$ the vector obtained after leaving out y_j, the jth component of $\boldsymbol{y}$. Similarly $\boldsymbol{Z}_{-j}$ denotes the latent random vector $\boldsymbol{Z}$ after deleting the jth component Z_j. Suppose $\boldsymbol{R}$ is the correlation matrix of $\boldsymbol{Z}$ then $\boldsymbol{R}_{11}^{(j)}$ and $\boldsymbol{R}_{22}^{(j)}$ denote the correlation matrices correspond to $\boldsymbol{Z}_{-j}$ and Z_j respectively. Also $\boldsymbol{R}_{12}^{(j)}$ denotes the correlation matrix between $\boldsymbol{Z}_{-j}$ and Z_j. If $\boldsymbol{Z}_{-j/j}$ denotes the conditional random vector $\boldsymbol{Z}_{-j}$ given Z_j, then the conditional mean of $\boldsymbol{Z}_{-j/j}$ is $\boldsymbol{\mu}_{-j/j} = \boldsymbol{R}_{12}^{(j)}(\boldsymbol{R}_{22}^{(j)})^{-1}Z_j$ and the conditional covariance of $\boldsymbol{Z}_{-j/j}$ is $\boldsymbol{R}_{-j/j} = \boldsymbol{R}_{11}^{(j)} - \boldsymbol{R}_{12}^{(j)}(\boldsymbol{R}_{22}^{(j)})^{-1}\boldsymbol{R}_{21}^{(j)}$.

In a similar fashion $\boldsymbol{y}_{-(ls)}$ denotes the vector $\boldsymbol{y}$ obtained leaving out the lth and sth components y_l and y_s. Also denote $\boldsymbol{y}_{(ls)} = (y_l, y_s)'$. Now $\boldsymbol{R}_{11}^{(ls)}$ and $\boldsymbol{R}_{22}^{(ls)}$ denote the correlation matrices correspond to $\boldsymbol{Z}_{-(ls)}$ and $\boldsymbol{Z}_{(ls)}$ respectively. Also $\boldsymbol{R}_{12}^{(ls)}$ is the correlation matrix between $\boldsymbol{Z}_{-(ls)}$ and $\boldsymbol{Z}_{(ls)}$. If we denote $\boldsymbol{Z}_{-(ls)/(ls)}$ the conditional random vector $\boldsymbol{Z}_{-(ls)}$ given $\boldsymbol{Z}_{(ls)}$, then the conditional mean and covariance matrix of $\boldsymbol{Z}_{-(ls)/(ls)}$ are $\boldsymbol{\mu}_{-(ls)/(ls)} = \boldsymbol{R}_{12}^{(ls)}\left(\boldsymbol{R}_{22}^{(ls)}\right)^{-1}\boldsymbol{Z}_{(ls)}$ and $\boldsymbol{R}_{-(ls)/(ls)} = \boldsymbol{R}_{11}^{(ls)} - \boldsymbol{R}_{12}^{(ls)}\left(\boldsymbol{R}_{22}^{(ls)}\right)^{-1}\boldsymbol{R}_{21}^{(ls)}$, respectively.

Based on the above notations, we have the following derivatives for the probability mass function of the Gaussian copula MCD model, with respect to the regression and correlation parameters.

1. $$\frac{\partial}{\partial\beta_l}\pi_t(\boldsymbol{y};0,\boldsymbol{R}) = \sum_{j=1}^{t} x_{jl}\left\{\sum_{i_j=1}^{2}(-1)^{i_j}\phi(b_{ji_j};0,1)\left[\pi_{t-1}(\boldsymbol{y}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})\right]\right\}$$

where $b_{j1} = \gamma_j(y_j - 1)$ and $b_{j2} = \gamma_j(y_j)$.

2. $$\frac{\partial}{\partial\alpha_l}\pi_t(\boldsymbol{y};\boldsymbol{0},\boldsymbol{R}) = \sum_{j=1}^{t}\sum_{i_j=1}^{2}\frac{\partial b_{ji_j}}{\partial\alpha_l}(-1)^{i_j}\phi(b_{ji_j};0,1)\left[\pi_{t-1}(\boldsymbol{y}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})\right]$$

where

$$\frac{\partial b_{ri_r}}{\partial\alpha_l} = \begin{cases} 1 & \text{if } (l = y_r - 1 \text{ and } i_r = 1) \text{ or } (l = y_r \text{ and } i_r = 2) \\ 0 & \text{otherwise.} \end{cases}$$

3. $$\frac{\partial}{\partial r_{ls}}\pi_t(\boldsymbol{y};\boldsymbol{0},\boldsymbol{R}) = \sum_{i_l=1}^{2}\sum_{i_s=1}^{2}\left[(-1)^{i_l+i_s}\phi_2((b_{li_l},b_{si_s});\boldsymbol{0},\boldsymbol{R}_{22}^{(ls)}).\right.$$
$$\left.\pi_{t-2}\big(\boldsymbol{y}_{-(ls)};\boldsymbol{\mu}_{-(ls)/(ls)},\boldsymbol{R}_{-(ls)/(ls)}\big)\right]$$

where $b_{j1} = \gamma_j(y_j - 1)$ and $b_{j2} = \gamma_j(y_j)$ and

$$\frac{\partial}{\partial \boldsymbol{\rho}}\pi_t(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{R}) = \sum_{i=1}^{t-1} \sum_{j=i+1}^{t} \frac{\partial}{\partial r_{ij}}\pi_t(\boldsymbol{y}; \boldsymbol{0}, \boldsymbol{R}). \frac{\partial r_{ij}}{\partial \boldsymbol{\rho}}.$$

The proofs of the above derivatives are given in Appendix 1. Plackett's identities given in Kotz et al. (2000) were used to find the derivatives with respect to correlation parameters r_{ls}. The score function for the multivariate ordered probit model is

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{1}{\pi_t(\boldsymbol{y}_i; \boldsymbol{0}, \boldsymbol{R})} \frac{\partial}{\partial \boldsymbol{\theta}}\pi_t(\boldsymbol{y}_i; \boldsymbol{0}, \boldsymbol{R})$$

where $\frac{\partial}{\partial \boldsymbol{\theta}}\pi_t(\boldsymbol{y}_i; \boldsymbol{0}, \boldsymbol{R}) = \left(\frac{\partial \pi_t}{\partial \boldsymbol{\alpha}} \;\; \frac{\partial \pi_t}{\partial \boldsymbol{\beta}} \;\; \frac{\partial \pi_t}{\partial \boldsymbol{\rho}}\right)'$. We developed an R code to solve the score equation and obtain the maximum likelihood estimates for the Gaussian copula based ordered probit-logit models using the following algorithm.

Step 1. Start with an initial estimate $\widehat{\boldsymbol{\theta}}_{int}$ of $\boldsymbol{\theta}$.

Step 2. At the ith step compute $\widehat{\boldsymbol{\theta}}_{i+1} = \widehat{\boldsymbol{\theta}}_i - c\, B(\widehat{\boldsymbol{\theta}}_i) g(\widehat{\boldsymbol{\theta}}_i)$ where $g(\boldsymbol{\theta}) = \partial l(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and $B(\boldsymbol{\theta})$ is an approximation to the inverse of Hessian matrix, $[\partial^2 l(\boldsymbol{\theta})/\partial \theta_j \partial \theta_k]^{-1}$, and c is a constant. See algorithm 21 in Nash (1979) for more details.

Step 3. Repeat Step 2 until $\widehat{\boldsymbol{\theta}}_{i+1} \cong \widehat{\boldsymbol{\theta}}_i$ and take $\widehat{\boldsymbol{\theta}} = \widehat{\boldsymbol{\theta}}_{i+1}$ as the MLE of $\boldsymbol{\theta}$.

Note that we use the analytical derivative of log-likelihood function, $g(\boldsymbol{\theta})$, in the iterative steps of the quasi-Newton algorithm. Even though the theory presented here is for balanced data $t_i = t$, the R code that we developed can handle unbalanced data, that is, t_i's could vary with i. In the following section we provide the analysis of real life data using Gaussian copula based ordered probit-logit models.

II.5 APPLICATIONS

II.5.1 MODEL INTERPRETATION

In this section we discuss some interpretations useful for data analysis for the Gaussian copula based ordered probit-logit models. In order to facilitate the discussion let $p_{j,r} = P(Y_j = r)$ as before, the probability that the response takes the rth category, and $\pi_{j,r} = P(Y_j \leq r)$ as the cumulative probability. From latent variable modeling we have seen that the marginal distribution of Y_j depends on the underlying latent random variable Z_j's distribution. If we assume that Z_j is distributed as standard normal then

$$p_{j,r} = P(\gamma_j(r-1) < Z_j \leq \gamma_j(r)) = \Phi(\gamma_j(r)) - \Phi(\gamma_j(r-1))$$

and

$$\pi_{j,r} = P(Y_j \leq r) = \sum_{l=1}^{r} p_{j,l} = \sum_{l=1}^{r} (\Phi(\gamma_j(l)) - \Phi(\gamma_j(l-1)) = \Phi(\gamma_j(r)))$$

because $\Phi(\gamma_j(0)) = \Phi(-\infty) = 0$. More precisely the relationship is

$$\pi_{j,r} = \Phi(\gamma_j(r)) \qquad \text{which implies} \qquad \Phi^{-1}(\pi_{j,r}) = \alpha_r + \boldsymbol{x}_j'\boldsymbol{\beta} \tag{21}$$

This relationship suggests that if we apply an ordered probit model then we link the marginal cumulative probabilities with the independent variables using a *probit* link function. Instead of standard normal if the distribution of Z_j is standard logistic distribution, that is,

$$F_{Z_j}(x) = \frac{1}{1+e^{-x}}; \qquad x \in (-\infty, \infty) \tag{22}$$

then

$$\pi_{j,r} = F_{Z_j}(\gamma_j(r)) \qquad \text{which implies} \qquad \log\left(\frac{\pi_{j,r}}{1-\pi_{j,r}}\right) = \alpha_r + \boldsymbol{x}_j'\boldsymbol{\beta} \tag{23}$$

This relationship suggests that if we assume that the underlying latent variable is distributed as standard logistic distribution then we link the marginal cumulative probabilities to the covariates using a *logit* link function. It is important to note that even in this case, the joint distribution of $\boldsymbol{Y} = (Y_1, Y_2, \cdots, Y_t)'$ can be constructed using multivariate normal copula.

II.5.2 ANALYSIS OF SKIN CONDITION CLINICAL TRIAL DATA

We return to the skin condition clinical trial experiment described in Chapter I. The main objective of this experiment is to test the efficacy of a new drug for skin conditions. To this data, we apply the Gaussian copula based ordered probit-logit models with AR(1) and compound symmetry (CS) structured latent correlation matrices. As explained in the model interpretation section the links for the cumulative probabilities are

(i) Gaussian copula-logit model:

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \gamma_{ij}(k) = \alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2};$$

(ii) Gaussian copula-probit model:

$$\pi_{ijk} = \Phi(\gamma_j(k)) = \Phi(\alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2})$$

for $i = 1, 2, \cdots, 171$, $j = 1, 2, 3$, and $k = 1, 2, 3, 4, 5$. The independent variables are $x_{ij1} = 1$ if the ith subject was given the new drug and $x_{ij1} = 2$ if ith subject was given the placebo. The covariate x_{ij2} is time related, that is, $x_{ij2} = j$.

Table 6: Parameter estimates and standard errors for the skin condition clinical trial data obtained fitting Gaussian copula models with AR(1) latent correlation structure.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	0.9590	0.2164	< 0.0001	0.2354	0.2174	0.2788
α_2	1.7343	0.2050	< 0.0001	1.4639	0.2097	< 0.0001
α_3	2.6054	0.2394	< 0.0001	2.8882	0.2196	< 0.0001
α_4	3.8988	0.2639	< 0.0001	4.9852	0.2063	< 0.0001
Treat	-1.3946	0.1329	< 0.0001	-1.4571	0.1389	< 0.0001
Time	0.1413	0.0310	< 0.0001	0.2454	0.0428	< 0.0001
ρ	0.8724	0.0099	< 0.0001	0.8590	0.0129	< 0.0001
-loglik	463.6943			471.0650		

Table 6 contains the parameter estimates along with standard errors and p-values for both Gaussian copula models when the latent correlation structure is assumed to be AR(1). Similarly Table 7 has the parameter estimates along with standard errors and p-values for both Gaussian copula models when the latent correlation structure is assumed to be CS. Treatment and time are significant in both the models and for both latent correlation structures. Interpretation of parameters for the model which assumes AR(1) structure for latent correlation matrix is given below. Similar explanations can be given for the other models.

Table 7: Parameter estimates and standard errors for the skin condition clinical trial data obtained fitting Gaussian copula models with CS latent correlation structure.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	0.7227	0.2461	0.0033	0.3250	0.2616	0.2141
α_2	1.3957	0.2510	< 0.0001	1.1250	0.2554	< 0.0001
α_3	2.2415	0.2800	< 0.0001	1.9829	0.2628	< 0.0001
α_4	3.6675	0.3467	< 0.0001	3.4545	0.2652	< 0.0001
Treat	-1.1470	0.1728	< 0.0001	-0.8868	0.1445	< 0.0001
Time	0.0973	0.0238	< 0.0001	0.1734	0.0249	< 0.0001
ρ	0.8485	0.0116	< 0.0001	0.9362	0.0028	< 0.0001
-loglik	474.1160			491.9667		

For the Gaussian copula-ordered logit model, we have

$$\frac{P(Y \le r/x_{ij1}=1, x_{ij2}=c)}{P(Y > r/x_{ij1}=1, x_{ij2}=c)} \Bigg/ \frac{P(Y \le r/x_{ij1}=2, x_{ij2}=c)}{P(Y > r/x_{ij1}=2, x_{ij2}=c)}$$

$$= \frac{P(Y \le r/ActiveDrug)}{P(Y > r/ActiveDrug)} \Bigg/ \frac{P(Y \le r/Placebo)}{P(Y > r/Placebo)} = \exp(\beta_1 - 2\beta_1) = \exp(-\beta_1)$$

Based on the above equation we can interpret the parameter β_1 in terms of the odds ratios. Odds of response level r or lesser in the treatment group is $\exp(-\beta_1) = \exp(0.8859) = 2.4252$ times odds of response level r or lesser in the placebo group.

Since the ordering nature of the response variable is improving towards the lower direction (1: Rapidly improving, ..., 5: Rapidly worsening) we can draw the following conclusions for the treatment group. For any level r, the odds that a treatment group patient response is in the improving direction rather than in the worsening direction is approximately 2.5 times the odds of placebo group patient.

II.5.3 ANALYSIS OF SIX CITIES LONGITUDINAL DATA

Next we analyze the six cities data. The goal of the data is to study the effect of the parental smoking on the child's wheezing status. We analyze the data using the Gaussian copula both the probit and logit marginals. The covariates used in the model are the city of residence for the child, age of the child and the smoking status of the mother at that particular age of the child. The marginal models are

(i) Gaussian copula-logit model:

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \gamma_{ij}(k) = \alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2} + \beta_3 \times x_{ij3};$$

(ii) Gaussian copula-probit model:

$$\pi_{ijk} = \Phi(\gamma_j(k)) = \Phi(\alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2} + \beta_3 \times x_{ij3})$$

for $i = 1, 2, \cdots, 297$, $j = 1, 2, 3, 4$, and $k = 1, 2, 3$. The independent variable x_{ij1} indicates the city of residence, x_{ij2} indicates mothers smoking status, and x_{ij3} is the normalized age.

Table 8: Parameter estimates and standard errors for the six cities data obtained fitting Gaussian copula models with AR(1) latent correlation structure.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	0.6590	0.3353	0.0494	1.3761	0.2551	< 0.0001
α_2	1.1299	0.3337	0.0007	2.1924	0.2628	< 0.0001
City	-0.3305	0.1378	0.0165	-0.5743	0.1638	0.0005
Smoke	-0.2576	0.0985	0.0089	-0.4550	0.1311	0.0005
Age	0.0317	0.0334	0.3420	0.0243	0.0259	0.3476
ρ	0.7361	0.0307	0.0000	0.7517	0.0344	< 0.0001
-loglik	482.0618			485.4908		

Parameter estimates for the six cities data are given in Table 8 and Table 9. The results for both the models for both the correlation structures are similar. Age is not a significant factor for both the models and since $\widehat{\beta_2}$ corresponding to mothers smoking status (Smoke) is negative and significant, we can conclude that the probability of improvement in the wheezing status for smoking mothers' children is less than that for the non-smoking mothers' children.

II.5.4 ANALYSIS OF RESPIRATORY DATA

Next we analyze the data set taken from a clinical trial to test a new treatment for a respiratory disorder. The marginal mean models for this data are as follows.

Table 9: Parameter estimates and standard errors for the six cities data obtained fitting Gaussian copula models with CS latent correlation structure.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	0.8633	0.3018	0.0042	0.6512	0.7080	0.3577
α_2	1.3528	0.3045	< 0.0001	1.7945	0.7017	0.0105
City	-0.2923	0.1292	0.0236	-0.3900	0.5886	0.5077
Smoke	-0.2069	0.1175	0.0782	-0.3158	0.9054	0.7273
Age	0.0102	0.0296	0.7299	0.1229	0.0856	0.1510
ρ	0.6750	0.0406	< 0.0001	0.7302	0.1512	< 0.0001
-loglik	477.0067			495.2465		

(i) Gaussian copula-logit model:

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \gamma_{ij}(k) = \alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2};$$

(ii) Gaussian copula-probit model:

$$\pi_{ijk} = \Phi(\gamma_j(k)) = \Phi(\alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2})$$

for $i = 1, 2, \cdots, 111$, $j = 1, 2, 3, 4$, and $k = 1, 2, 3$. The independent variable x_{ij1} takes values 1,2,3,4, corresponding to the visit, and x_{ij2} is an indicator for the treatment (1 = active treatment, 0 = placebo).

Table 10: Parameter estimates and standard errors for the respiratory data obtained fitting Gaussian copula models with AR(1) latent correlation structure.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	-1.2094	0.1860	< 0.0001	-1.2369	0.7843	0.1148
α_2	0.0852	0.1581	0.5899	0.7078	0.8604	0.4107
Visit	0.0567	0.0221	0.0103	0.0135	1.0968	0.9902
Treat	0.5506	0.1961	0.0050	0.0522	1.0067	0.9586
ρ	0.7445	0.0315	< 0.0001	0.7239	0.1916	0.0002
-loglik	357.0737			363.4830		

Table 11: Parameter estimates and standard errors for the respiratory data obtained fitting Gaussian copula models with CS latent correlation structure.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	-1.2972	0.1756	< 0.0001	-1.2896	1.0806	0.2327
α_2	0.0829	0.1396	0.5528	0.7567	0.9531	0.4272
Visit	0.0587	0.0272	0.0306	0.0351	0.2054	0.8642
Treat	0.5245	0.1750	0.0027	0.0399	0.9935	0.9680
ρ	0.5582	0.0569	< 0.0001	0.6621	0.6919	0.3386
-loglik	364.3997			367.6742		

Unlike the previous examples, here the results from the two Gaussian copula models are not the same. For example, we can see from the results in Table 10 and Table 11, both visit and treatment are significant in the probit marginal model whereas they both are insignificant in the logit marginal model.

II.5.5 ANALYSIS OF INSOMNIA CLINICAL TRIAL DATA

Finally we analyze the data collected on insomnia patients to compare a hypnotic drug with placebo. The marginal mean models for this data are as follows.

(i) Gaussian copula-logit model:

$$\log\left(\frac{\pi_{ijk}}{1-\pi_{ijk}}\right) = \gamma_{ij}(k) = \alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2} + \beta_3 \times x_{ij1} * x_{ij2};$$

(ii) Gaussian copula-probit model:

$$\pi_{ijk} = \Phi(\gamma_j(k)) = \Phi(\alpha_k + \beta_1 \times x_{ij1} + \beta_2 \times x_{ij2} + \beta_3 \times x_{ij1} * x_{ij2})$$

for $i = 1, 2, \cdots, 239$, $j = 1, 2$, and $k = 1, 2, 3, 4$. The independent variables are x_{ij1} which is an indicator for the treatment (1=hypnotic drug, 0=placebo), and x_{ij2} indicates the occasion (1 = follow up occasion, 0 = initial occasion).

Since each subject is observed only at two time points, there is no difference between the AR(1) and CS correlations structures. Table 12 provides the parameter estimates along with standard errors and the p-values. The interaction term is significant in both the marginal probit and the logit model. Note that for the logistic

Table 12: Parameter estimates and standard errors for the insomnia data obtained fitting Gaussian copula model.

	(i) Ordered Probit			(ii) Ordered Logit		
Parameter	Estimate	SE	p - value	Estimate	SE	p - value
α_1	-1.3582	0.1192	< 0.0001	-2.1303	0.1438	< 0.0001
α_2	-0.5812	0.1088	< 0.0001	-0.8603	0.1304	< 0.0001
α_3	0.2452	0.1064	0.0212	0.4565	0.1314	0.0005
Treat	0.0108	0.1434	0.9398	-0.0731	0.1766	0.6789
Occasion	0.6272	0.1106	< 0.0001	0.9662	0.1195	< 0.0001
Treat*Occasion	0.4430	0.1539	0.0040	0.7685	0.1666	< 0.0001
ρ	0.5263	0.0555	< 0.0001	0.5518	0.0523	< 0.0001
-loglik	594.8495			596.9802		

marginal,

$$\frac{\pi_{ijk}}{1-\pi_{ijk}} = \exp(\alpha_k + \beta_1\, Treat + \beta_2\, Occasion + \beta_3\, Treat * Occasion) \tag{24}$$

From the above (24) relation we can see that the odds of response level k or lesser at the follow-up occasion is $\exp(\beta_2 + \beta_3) = \exp(0.9662 + 0.7685) = 5.67$ times odds of response level k or lesser at the initial occasion in the treatment group. On the other hand, odds of response level k or lesser at the follow-up occasion is $\exp(\beta_2) = \exp(0.9662) = 2.62$ times odds of response level k or lesser at the initial occasion in the placebo group.

Similarly, odds of response level k or lesser in the treatment group is $\exp(\beta_1+\beta_3) = \exp(-0.0731 + 0.7685) = 2.00$ times odds of response level k or lesser in the placebo group at the follow up occasion. On the other hand, odds of response level k or lesser in the treatment group is $\exp(\beta_1) = \exp(-0.0731) = 0.93$ times odds of response level k or lesser in the placebo group at the initial occasion.

CHAPTER III

ESTIMATING EQUATIONS

III.1 INTRODUCTION

In Chapter II we have discussed modeling repeated or longitudinal ordered categorical data based on latent variables. However in the literature there are other models and these can be classified broadly into three types:

1. Marginal Models.
2. Random Effect Models.
3. Transitional Models.

The above three models account for the dependence among the observations on the same subject (or in a cluster) in different ways. In marginal models, marginal expectation and the dependence among the repeated observations are modeled separately. Whereas in random effect models, dependence is accounted using subject-specific random effects. Finally, in transitional models the dependency is measured by including the subject's past history into the model. Marginal models are non-likelihood models, whereas the other two models, random effects and transition are likelihood models. The construction of the likelihood is different for each of the two likelihood models is different but the parameter estimation is simply the maximum likelihood. Marginal models can be viewed as a generalization of generalized linear models for univariate responses. In this chapter we study marginal models and estimating equation techniques for categorical repeated measurements data.

III.2 UNIVARIATE GENERALIZED LINEAR MODELS

III.2.1 INTRODUCTION

McCullagh and Nelder (1989) gave a unified regression approach for the response random variables that belong to the exponential family of distributions and is known

Table 13: Univariate data structure

Subject	Response	Covariates			
1	Y_1	x_{11}	x_{12}	$\cdots$	x_{1p}
2	Y_2	x_{21}	x_{22}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
i	Y_i	x_{i1}	x_{i2}	$\cdots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	Y_n	x_{n1}	x_{n2}	$\cdots$	x_{np}

as Generalized Linear Models (GLM). These GLM regression models have three parts: (i) a response distribution that belongs to the exponential family, (ii) a linear combination of independent variables known as linear predictors, and (iii) a link function connecting the mean response variable to the linear predictors. Suppose data consists of Y_i, a response on the ith subject and $\boldsymbol{x}_i$ a $p \times 1$ vector of covariates associated with Y_i for $i = 1, 2, ..., n$ as given in Table 13. If $E(Y_i) = \mu_i$, then the fundamental assumption in GLM regression is given by

$$g(\mu_i) = g(E(Y_i)) = \eta_i = \boldsymbol{x}_i'\boldsymbol{\beta} \tag{25}$$

for some link function $g(\cdot)$. If Y_i is a categorical random variable that takes one of the K categories, then to apply the GLM regression we need to create for each Y_i a binary choice vector $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \cdots, Y_{iK-1})'$ where $Y_{ir} = 1$ if $Y_i = r$, and 0 otherwise. The resulting data structure is given in Table 14.

The binary choice vector $\boldsymbol{Y}_i$ is simply a multinomial random vector with one trial and K - categories ; $i = 1, 2, ..., n$, that is, $\boldsymbol{Y}_i \sim$ Multinomial$(1, \mu_{i1}, \mu_{i2}, ..., \mu_{iK-1})$ where μ_{ir} is the probability that Y_i chooses the rth category. Using this representation, we can describe the three components of GLM for categorical variables as follows.

Table 14: Binary choice vector representation

Subject	Response	Binary choice vector	Covariates			
1	Y_1	$(Y_{11}, Y_{12}, \cdots, Y_{1K-1})'$	x_{11}	x_{12}	$\cdots$	x_{1p}
2	Y_2	$(Y_{21}, Y_{22}, \cdots, Y_{2K-1})'$	x_{21}	x_{22}	$\cdots$	x_{2p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
i	Y_i	$(Y_{i1}, Y_{i2}, \cdots, Y_{iK-1})'$	x_{i1}	x_{i2}	$\cdots$	x_{ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
n	Y_n	$(Y_{n1}, Y_{n2}, \cdots, Y_{nK-1})'$	x_{n1}	x_{n2}	$\cdots$	x_{np}

III.2.2 RESPONSE DISTRIBUTION

Since $\boldsymbol{Y}_i$ is a multinomial random vector, we have

$$\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{iK-1}) \sim \text{Mult}(1, \mu_{i1}, \mu_{i2}, ..., \mu_{iK-1}) \text{ and}$$
$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \prod_{k=1}^{K} \mu_{ik}^{y_{ik}} \quad \text{with} \sum_{k=1}^{K} \mu_{ik} = 1 \text{ and } \sum_{k=1}^{K} y_{ik} = 1$$

where

$$Y_{ik} = \begin{cases} 1 & \text{if } Y_i \text{ takes the category } k \\ 0 & \text{otherwise} \end{cases} \quad ; \ k = 1, 2, ..., K-1$$

Note that $E(Y_{ik}) = \mu_{ik}$. Because of the restrictions on the Y_{ik}'s and μ_{ik}'s we have

$$Y_{iK} = 1 - \sum_{k=1}^{K-1} Y_{ik} \quad \text{and} \quad \mu_{iK} = 1 - \sum_{k=1}^{K-1} \mu_{ik}.$$

and only $(K-1)$ probabilities are independent. Therefore we can rewrite the probability distribution more explicitly as

$$P(\boldsymbol{Y}_i = \boldsymbol{y}_i) = \prod_{k=1}^{K} \mu_{ik}^{y_{ik}} = \prod_{k=1}^{K-1} \mu_{ik}^{y_{ik}} \ \mu_{iK}^{y_{iK}}.$$

This probability mass function belongs to the exponential family because

$$
\begin{aligned}
P(\boldsymbol{Y}_i = \boldsymbol{y}_i) &= \prod_{k=1}^{K} \mu_{ik}^{y_{ik}} \\
&= \exp\left\{\sum_{j=1}^{K-1} y_{ij}\log\mu_{ij} + \left(1 - \sum_{r=1}^{K-1} y_{ir}\right)\log\left(1 - \sum_{r=1}^{K-1}\mu_{ir}\right)\right\} \\
&= \left(1 - \sum_{r=1}^{K-1}\mu_{ir}\right)\exp\left\{\sum_{j=1}^{K-1} y_{ij}\log\left(\frac{\mu_{ij}}{1-\mu_{i1}-\mu_{i2}-\cdots-\mu_{iK-1}}\right)\right\} \\
&= c(\boldsymbol{\theta}_i)h(\boldsymbol{y})\exp\left\{\sum_{j=1}^{K-1} w_j(\boldsymbol{\theta}_i)t_j(\boldsymbol{y})\right\}
\end{aligned}
$$

where $c(\boldsymbol{\theta_i}) = \left(1 - \sum_{r=1}^{K-1}\mu_{ir}\right)$, $h(\boldsymbol{y}) = 1$, $w_j(\boldsymbol{\theta_i}) = \log\left(\frac{\mu_{ij}}{1-\mu_{i1}-\mu_{i2}- \quad -\mu_{iK-1}}\right)$, and $t_j(\boldsymbol{y}) = y_j$. Here $\boldsymbol{\theta}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{i(K-1)})$.

III.2.3 LINEAR PREDICTOR

Since the response variable has $K-1$ independent levels, we need to consider $K-1$ linear combinations of the predictors in the regression model. Though not necessary but usually the same $\boldsymbol{x}_i$ vector of covariates is associated with each level of the multinomial responses. The kth level linear predictor η_{ik} is given by

$$\eta_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta}_k = \sum_{l=1}^{p} x_{il}\beta_{lk}$$

In matrix form the regression parameter for all the levels can be written as

$$\begin{pmatrix} \boldsymbol{\beta}_{1_{p\times 1}} & \boldsymbol{\beta}_{2_{p\times 1}} & \cdots & \boldsymbol{\beta}_{K-1_{p\times 1}} \end{pmatrix} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1K-1} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2K-1} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{p1} & \beta_{p2} & \cdots & \beta_{pK-1} \end{pmatrix}_{(p\,\times\,K-1)} .$$

Depending on the model assumptions these parameters can have different interpretations. For example, $\beta_{11}, \beta_{12}, \cdots, \beta_{1K-1}$ can be treated as intercepts for each of the $(K-1)$ levels. In addition we can decrease the number of parameters with additional assumptions. For example we can assume $\beta_{ij} = \beta_i$ for all $j = 1, 2, ..., K-1$; $i = 1, 2, 3, ..., p$. In this case, the number of parameters depend only on p and the resulting model is known as " Proportional odds model" or "Parallel slopes model".

III.2.4 LINK FUNCTIONS

In univariate generalized linear models for binary or binomial random variables the mean is related to the covariates through a link function. Similarly, for multinomial random vectors, the mean μ_{ik} is related to the linear predictor η_{ik} through a monotone differential link function as

$$g(\mu_{ik}) = \eta_{ik} \text{ or } \mu_{ik} = g^{-1}(\eta_{ik}). \tag{26}$$

Popular choice for $g(\cdot)$ in the case of binary or binomial responses are the logit and the probit link functions. These link functions are natural candidates for the multinomial responses as well. In the multinomial case we have other possibilities, which include linking the cumulative probabilities instead of marginal means. Thus the possible link functions are

1. Multinomial Logits Model (MNL):

$$g(\mu_{ik}) = logit(\mu_{ik}) = \log\left(\frac{\mu_{ik}}{1-\mu_{ik}}\right) = \eta_{ik} = \alpha_k + \boldsymbol{x}_i'\boldsymbol{\beta}_k$$

2. Multinomial Probit (MNP):

$$g(\mu_{ik}) = \Phi^{-1}(\mu_{ik}) = \eta_{ik} = \alpha_k + \boldsymbol{x}_i'\boldsymbol{\beta}_k$$

3. Cumulative Logit (Odds-Proportional Model):

$$g(\pi_{ik}) = logit(\pi_{ik}) = \log\left(\frac{\pi_{ik}}{1-\pi_{ik}}\right) = \eta_{ik} = \alpha_k + \boldsymbol{x}_i'\boldsymbol{\beta}, \text{ where } \pi_{il} = \sum_{k=1}^{l} \mu_{ik}.$$

4. Cumulative Probit (Ordered Probit Model):

$$g(\pi_{ik}) = \Phi^{-1}(\pi_{ik}) = \eta_{ik} = \alpha_k + \boldsymbol{x}_i'\boldsymbol{\beta}, \text{ where } \pi_{il} = \sum_{k=1}^{l} \mu_{ik}.$$

In all cases α_k are the model intercepts. The MNL and the MNP link functions could also be used for nominal categorical variables as well. Other link functions, which are not studied further in this dissertation, are the complementary log-log and cumulative complementary log-log link functions:

1. Complementary log-log:

$$g(\mu_{ik}) = \log(-\log(1-\mu_{ik})) = \eta_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta}_k.$$

2. Cumulative Complementary log-log:

$$g(\pi_{ik}) = \log(-\log(1-\pi_{ik})) = \eta_{ik} = \boldsymbol{x}_i'\boldsymbol{\beta}_k, \text{ where } \pi_{il} = \sum_{k=1}^{l} \mu_{ik}.$$

A detailed discussion of these link functions can be found in Agresti (2002).

III.2.5 MAXIMUM LIKELIHOOD ESTIMATION FOR SINGLE OBSERVATIONS

In this section we discuss maximum likelihood estimation for ordinal data when we have only a single ordinal measurement for each subject ($t_i = 1$). The likelihood equations in this case form the basis for developing estimating equations for parameter estimation for $t_i > 1$ and when the within subject observations are dependent. Let $\boldsymbol{Y}_i = (y_{i1}, y_{i2}, \ldots, y_{i(K-1)})'$, $1 \le i \le n$, be independent multinomial random vectors. Assume that $E(\boldsymbol{Y}_i) = \boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{i(K-1)})'$, where μ_{ik} is a function of the covariates and an unknown regression parameter $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ as given in previous section. Then the likelihood function can be written as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \left[\left(\prod_{k=1}^{K-1} \mu_{ik}^{y_{ik}} \right) \mu_{iK}^{y_{iK}} \right],$$

and the log-likelihood is

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \left[\left(\sum_{k=1}^{K-1} y_{ik} \log \mu_{ik} \right) + y_{iK} \log \mu_{iK} \right], \tag{27}$$

where $y_{iK} = 1 - \sum_{k=1}^{K-1} y_{ik}$ and $\mu_{iK} = 1 - \sum_{k=1}^{K-1} \mu_{ik}$. Taking the derivative of (27) with respect to $\boldsymbol{\theta}$ we get the score equation as

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\theta}} &= \sum_{i=1}^{n} \left[\sum_{k=1}^{K-1} \frac{y_{ik}}{\mu_{ik}} \frac{\partial \mu_{ik}}{\partial \boldsymbol{\theta}} + \frac{y_{iK}}{\mu_{iK}} \frac{\partial \mu_{iK}}{\partial \boldsymbol{\theta}} \right] = \mathbf{0} \\ &= \sum_{i=1}^{n} \left\{ \sum_{k=1}^{K-1} \frac{\partial \mu_{ik}}{\partial \boldsymbol{\theta}} \left[\frac{y_{ik}}{\mu_{ik}} - \frac{y_{iK}}{\mu_{iK}} \right] \right\} = 0. \end{aligned} \tag{28}$$

In the above simplification we used the identity

$$\frac{\partial \mu_{iK}}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} \left[1 - \sum_{k=1}^{K-1} \mu_{ik} \right] = -\sum_{k=1}^{K-1} \frac{\partial \mu_{ik}}{\partial \boldsymbol{\theta}}.$$

Equation (28), in matrix notation, can be written as

$$S(\boldsymbol{\theta}) = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\theta}} \Sigma_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}_{p+(K-1)\times 1} \tag{29}$$

where $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{iK-1})'$, $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, ..., \mu_{iK-1})'$ and

$$\mathrm{Cov}(\boldsymbol{Y}_i) = \Sigma_i = \begin{pmatrix} \mu_{i1}(1-\mu_{i1}) & -\mu_{i1}\mu_{i2} & \cdots & -\mu_{i1}\mu_{iK-1} \\ -\mu_{i2}\mu_{i1} & \mu_{i2}(1-\mu_{i2}) & \cdots & -\mu_{i2}\mu_{iK-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\mu_{iK-1}\mu_{i1} & -\mu_{iK-1}\mu_{i2} & \cdots & \mu_{iK-1}(1-\mu_{iK-1}) \end{pmatrix}_{K-1 \times K-1}$$

and the inverse is

$$\Sigma_i^{-1} = \begin{pmatrix} \frac{1}{\mu_{i1}} + \frac{1}{\mu_{iK}} & \frac{1}{\mu_{iK}} & \cdots & \frac{1}{\mu_{iK}} \\ \frac{1}{\mu_{iK}} & \frac{1}{\mu_{i2}} + \frac{1}{\mu_{iK}} & \cdots & \frac{1}{\mu_{iK}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{\mu_{iK}} & \frac{1}{\mu_{iK}} & \cdots & \frac{1}{\mu_{iK-1}} + \frac{1}{\mu_{iK}} \end{pmatrix}, \text{ where } \mu_{iK} = 1 - \sum_{k=1}^{K-1} \mu_{ik}.$$

Note that

$$\Sigma_i^{-1}(\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \begin{pmatrix} \frac{Y_{i1}}{\mu_{i1}} - \frac{Y_{iK}}{\mu_{iK}} \\ \frac{Y_{i2}}{\mu_{i2}} - \frac{Y_{iK}}{\mu_{iK}} \\ \vdots \\ \frac{Y_{iK-1}}{\mu_{iK-1}} - \frac{Y_{iK}}{\mu_{iK}} \end{pmatrix}.$$

The maximum likelihood estimate of $\boldsymbol{\theta}$ is the solution of the score equation $S(\boldsymbol{\theta}) = \mathbf{0}$. We need to use numerical routines, such as Newton-Raphson method, to solve this score equation iteratively as follows:

Step 1. Start with a trial value for $\boldsymbol{\theta}$, say $\widehat{\boldsymbol{\theta}}_0$.

Step 2. Calculate $\widehat{\boldsymbol{\theta}}_1 = \widehat{\boldsymbol{\theta}}_0 - H(\widehat{\boldsymbol{\theta}}_0)^{-1}S(\widehat{\boldsymbol{\theta}}_0)$.

Step 3. Replace $\widehat{\boldsymbol{\theta}}_0$ by $\widehat{\boldsymbol{\theta}}_1$ and repeat Step 2.

Step 4. Stop when $\widehat{\boldsymbol{\theta}}_{r+1} \approx \widehat{\boldsymbol{\theta}}_r$.

In the above algorithm $H(\widehat{\boldsymbol{\theta}})$, which is the derivative of $S(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$, is known as the Hessian matrix. If we replace $H(\boldsymbol{\theta})$ with $I(\boldsymbol{\theta}) = -E(H(\boldsymbol{\theta}))$, the Fisher information, then the above iterative algorithm is known as Fisher-scoring method. In the literature, the Fisher-scoring method is also known as iterated reweighted least squares (IRLS). See McCullagh and Nelder (1989).

III.3 MARGINAL MODELS

III.3.1 INTRODUCTION

Suppose on each subject we have repeated measurements that are categorical in nature. A simple alternative to the challenging likelihood approach is marginal modeling. These models can be regarded as generalization of the univariate GLM methodology to the multivariate situation. As we mentioned before, in marginal models the within subject correlation is modeled separately from the marginal mean. Suppose Y_{ij} is a response observed on the ith subject at the jth time point and $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2}, \cdots, x_{ijp})'$ is the $p \times 1$ vector of covariates associated with Y_{ij} ; $i = 1, 2, \cdots, n$ and $j = 1, 2, \cdots, t$. The assumptions for the marginal models are as follows:

1. The marginal expectation μ_{ij} of Y_{ij} is related to the covariates through a known monotone differentiable link function

$$g(\mu_{ij}) = \boldsymbol{x}_{ij}'\boldsymbol{\beta}$$

   where $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression parameters.

2. The marginal variance and marginal mean of the response variable are related through a known function.

$$\text{Var}(Y_{ij}) = \phi\ \nu(\mu_{ij})$$

where $\nu(.)$ is a known function, and $\phi > 0$ is a scale parameter.

3. The correlation between y_{ij} and $y_{ij'}$ is a function of $\mu_{ij}, \mu_{ij'}$ and a vector of unknown parameters $\boldsymbol{\lambda}$

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \rho(\mu_{ij}, \mu_{ij'}; \boldsymbol{\lambda})$$

where $\rho(.)$ is a known function.

In marginal models, as we discussed before, the categorical variable Y_{ij} is replaced by a binary choice vector $\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \cdots, Y_{ijK-1})'$ where $Y_{ijr} = 1$ if $Y_{ij} = r$, and 0 otherwise. The data layout for this binary choice vector representation is given in Table 15.

Table 15: Longitudinal data layout with binary choice vector representation

Subject	Time	Response	Binary choice vector	Covariates			
	1	Y_{11}	$(Y_{111}, Y_{112}, \cdots, Y_{11K-1})$	x_{111}	x_{112}	$\cdots$	x_{11p}
	2	Y_{12}	$(Y_{121}, Y_{122}, \cdots, Y_{12K-1})$	x_{121}	x_{122}	$\cdots$	x_{12p}
1	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t_1	Y_{1t_1}	$(Y_{1t_11}, Y_{1t_12}, \cdots, Y_{1t_1K-1})$	x_{1t_11}	x_{1t_12}	$\cdots$	x_{1t_1p}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	
	1	Y_{i1}	$(Y_{i11}, Y_{i12}, \cdots, Y_{i1K-1})$	x_{i11}	x_{i12}	$\cdots$	x_{i1p}
	2	Y_{i2}	$(Y_{i21}, Y_{i22}, \cdots, Y_{i2K-1})$	x_{i21}	x_{i22}	$\cdots$	x_{i2p}
i	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t_i	Y_{it_i}	$(Y_{it_i1}, Y_{it_i2}, \cdots, Y_{it_iK-1})$	x_{it_i1}	x_{it_i2}	$\cdots$	x_{it_ip}
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	1	Y_{n1}	$(Y_{n11}, Y_{n12}, \cdots, Y_{n1K-1})$	x_{n11}	x_{n12}	$\cdots$	x_{n1p}
	2	Y_{n2}	$(Y_{n21}, Y_{n22}, \cdots, Y_{n2K-1})$	x_{n21}	x_{n22}	$\cdots$	x_{n2p}
n	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
	t_n	Y_{nt_n}	$(Y_{nt_n1}, Y_{nt_n2}, \cdots, Y_{nt_nK-1})$	x_{nt_n1}	x_{nt_n2}	$\cdots$	x_{nt_np}

III.3.2 CORRELATION MODELS FOR REPEATED MULTINOMIAL RESPONSES

Since $\boldsymbol{Y}_{ij}$ is a multinomial random vector with one trial with mean $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \mu_{ij2}, \cdots, \mu_{ijK-1})'$, we can write the probability mass function as,

$$P(\boldsymbol{Y}_{ij} = \boldsymbol{y}_{ij}) = \prod_{k=1}^{K} \mu_{ijk}^{y_{ijk}} \quad \text{with } \sum_{k=1}^{K} \mu_{ijk} = 1 \text{ and } \sum_{k=1}^{K} y_{ijk} = 1$$

Note that $E(Y_{ijk}) = \mu_{ijk}$ and in the marginal model we assume

$$g(\mu_{ijk}) = g(E(Y_{ijk})) = \eta_{ijk} = \alpha_k + \boldsymbol{x}'_{ij}\boldsymbol{\beta}. \tag{30}$$

Common choices for $g(\cdot)$ are the generalized logit and multinomial probit link functions. Although some authors have used cumulative logit and cumulative probit link functions for the cumulative probabilities of the ordinal responses. Next the covariance matrix of the multinomial random vector, $\boldsymbol{Y}_{ij}$ is

$$\begin{aligned} \text{Cov}(\boldsymbol{Y}_{ij}) = \Sigma_{ij} &= \begin{pmatrix} \mu_{ij1}(1-\mu_{ij1}) & -\mu_{ij1}\mu_{ij2} & \cdots & -\mu_{ij1}\mu_{ijK-1} \\ -\mu_{ij2}\mu_{ij1} & \mu_{ij2}(1-\mu_{ij2}) & \cdots & -\mu_{ij2}\mu_{ijK-1} \\ \vdots & \vdots & \ddots & \vdots \\ -\mu_{ijK-1}\mu_{ij1} & -\mu_{ijK-1}\mu_{ij2} & \cdots & \mu_{ijK-1}(1-\mu_{ijK-1}) \end{pmatrix} \\ &= \text{diag}(\boldsymbol{\mu}_{ij}) - \boldsymbol{\mu}_{ij}\boldsymbol{\mu}'_{ij} \end{aligned} \tag{31}$$

where $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \mu_{ij2}, \cdots, \mu_{ijK-1})'$. If $\boldsymbol{A}_{ij} = \text{diag}(\Sigma_{ij})$, the $\text{Corr}(\boldsymbol{Y}_{ij}) = \boldsymbol{R}_{ij} = \boldsymbol{A}_{ij}^{-1/2}\,\Sigma_{ij}\,\boldsymbol{A}_{ij}^{-1/2}$. Finally, we need to model $\Sigma_{ijj'}$ the covariance between $\boldsymbol{Y}_{ij}$ and $\boldsymbol{Y}_{ij'}$ for $j \neq j'$. More generally we need to model the covariance Σ_i of $\boldsymbol{Y}_i = (\boldsymbol{Y}'_{i1}, \boldsymbol{Y}'_{i2}, \cdots, \boldsymbol{Y}'_{it})'$. In marginal models this is obtained via a model for the correlation. Let $\boldsymbol{A}_i = \text{diag}(\boldsymbol{A}_{i1}, \boldsymbol{A}_{i2}, \ldots, \boldsymbol{A}_{it})$. Then $\text{Cov}(\boldsymbol{Y}_i) = \Sigma_i = \boldsymbol{A}_i^{1/2}\,\boldsymbol{R}(\boldsymbol{\lambda})\,\boldsymbol{A}_i^{1/2}$, where $\boldsymbol{R}(\boldsymbol{\lambda})$ is a structured correlation matrix determined by the parameter vector $\boldsymbol{\lambda}$. The preceding notation is best understood in a simple case. Suppose $K = 3$ and $t_i = 4$ for all i. Then

$$\text{Cov}(\boldsymbol{Y}_i) \;=\; \Sigma_i = \begin{pmatrix} \Sigma_{i1} & \Sigma_{i,12} & \Sigma_{i,13} & \Sigma_{i,14} \\ \Sigma_{i,21} & \Sigma_{i2} & \Sigma_{i,23} & \Sigma_{i,24} \\ \Sigma_{i,31} & \Sigma_{i,32} & \Sigma_{i3} & \Sigma_{i,34} \\ \Sigma_{i,41} & \Sigma_{i,42} & \Sigma_{i,43} & \Sigma_{i4} \end{pmatrix} \tag{32}$$

where the matrices on the diagonal are

$$\Sigma_{ij} = \begin{pmatrix} \mu_{ij1}(1-\mu_{ij1}) & -\mu_{ij1}\mu_{ij2} \\ -\mu_{ij2}\mu_{ij1} & \mu_{ij2}(1-\mu_{ij2}) \end{pmatrix} \quad \text{for } j = 1,2,3,4.$$

The off-diagonal matrices $\Sigma_{ijj's}$ for $j \neq j'$ are determined by the diagonal matrix $\boldsymbol{A}_i$ given by

$$\boldsymbol{A}_i \;=\; \begin{pmatrix} \text{diag}(\Sigma_{i1}) & 0 & 0 & 0 \\ 0 & \text{diag}(\Sigma_{i2}) & 0 & 0 \\ 0 & 0 & \text{diag}(\Sigma_{i3}) & 0 \\ 0 & 0 & 0 & \text{diag}(\Sigma_{i4}) \end{pmatrix}$$

and the correlation matrix $\boldsymbol{R}(\boldsymbol{\lambda})$ of $\boldsymbol{Y}_i$ given by

$$\boldsymbol{R}(\boldsymbol{\lambda}) = \begin{pmatrix} \boldsymbol{R}_{i11} & \boldsymbol{R}_{12} & \boldsymbol{R}_{13} & \boldsymbol{R}_{14} \\ \boldsymbol{R}_{21} & \boldsymbol{R}_{i22} & \boldsymbol{R}_{23} & \boldsymbol{R}_{24} \\ \boldsymbol{R}_{31} & \boldsymbol{R}_{32} & \boldsymbol{R}_{i33} & \boldsymbol{R}_{34} \\ \boldsymbol{R}_{41} & \boldsymbol{R}_{42} & \boldsymbol{R}_{43} & \boldsymbol{R}_{i44} \end{pmatrix}. \tag{33}$$

The matrices on the diagonal $\boldsymbol{R}_{ijj} = (\text{diag}(\Sigma_{ij}))^{-1/2}\,\Sigma_{ij}\,(\text{diag}(\Sigma_{ij}))^{-1/2}$ are the correlation matrices corresponding to Σ_{ij}. Note that $\boldsymbol{R}_{ijj}$ is independent of $\boldsymbol{\lambda}$ and depends only on $\boldsymbol{\mu}_i$. But the off-diagonal matrices $\boldsymbol{R}_{jj'}$ for $1 \leq j \neq j' \leq 4$ are functions of the parameter $\boldsymbol{\lambda}$. Commonly used structures are

$$\text{Compound symmetry (CS):} \qquad \boldsymbol{R}_{jj'}(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_1 & \lambda_2 \\ \lambda_3 & \lambda_4 \end{pmatrix} \text{ for all } j, j'.$$

$$\text{Autoregressive order 1 (AR(1)):} \qquad \boldsymbol{R}_{jj'}(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_1^{|j-j'|} & \lambda_2^{|j-j'|} \\ \lambda_3^{|j-j'|} & \lambda_4^{|j-j'|} \end{pmatrix} \text{ for all } j, j'.$$

$$\text{Unstructured (UN):} \qquad \boldsymbol{R}_{jj'}(\boldsymbol{\lambda}) = \begin{pmatrix} \lambda_{jj',1} & \lambda_{jj',2} \\ \lambda_{jj',3} & \lambda_{jj',4} \end{pmatrix} \text{ for all } j, j'.$$

III.3.3 GENERALIZED ESTIMATING EQUATIONS

A popular methodology for estimating parameters in marginal models is the Generalized Estimating Equations (GEE) proposed by Liang and Zeger (1986). To understand the construction of the GEE, we first look at the case where the $\boldsymbol{Y}_{ij}$'s (distributed as Multinomial$(1, \boldsymbol{\mu}_{ij})$) are independent for all $i = 1, 2, ..., n$ and $j = 1, 2, ..., t$. In this case the likelihood is

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} \prod_{j=1}^{t} \left[\prod_{k=1}^{K-1} \mu_{ijk}^{y_{ijk}} \mu_{ijK}^{y_{ijK}} \right]$$

where $\mu_{ijk} = g^{-1}(\alpha_k + \boldsymbol{x}'_{ij}\boldsymbol{\beta})$ for some link function $g(\cdot)$ and $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$. Then the log-likelihood is

$$l(\boldsymbol{\theta}) = \log L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{t} \left[\left(\sum_{k=1}^{K-1} y_{ijk} \log \mu_{ijk} \right) + y_{ijK} \log \mu_{ijK} \right].$$

Differentiating the above with respect to $\boldsymbol{\theta}$, we get the score equation as

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \sum_{j=1}^{t} \left[\sum_{k=1}^{K-1} \frac{y_{ijk}}{\mu_{ijk}} \frac{\partial \mu_{ijk}}{\partial \boldsymbol{\theta}} + \frac{y_{ijK}}{\mu_{ijK}} \frac{\partial \mu_{ijK}}{\partial \boldsymbol{\theta}} \right] = \mathbf{0}.$$

Since

$$\frac{\partial \mu_{ijK}}{\partial \boldsymbol{\theta}} = \frac{\partial}{\partial \boldsymbol{\theta}} \left[1 - \sum_{k=1}^{K-1} \mu_{ijk} \right] = - \sum_{k=1}^{K-1} \frac{\partial \mu_{ijk}}{\partial \boldsymbol{\theta}},$$

we have

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \; = \; \sum_{i=1}^{n} \sum_{j=1}^{t} \left[\sum_{k=1}^{K-1} \frac{\partial \mu_{ijk}}{\partial \boldsymbol{\theta}} \left[\frac{y_{ijk}}{\mu_{ijk}} - \frac{y_{ijK}}{\mu_{ijK}} \right] \right] = \mathbf{0}. \tag{34}$$

In matrix notation equation (34) can be written as

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \sum_{i=1}^{n} \frac{\partial \mu'_i}{\partial \boldsymbol{\theta}} \boldsymbol{B}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0}, \tag{35}$$

where $\boldsymbol{\mu}_i = (\boldsymbol{\mu}'_{i1}, \boldsymbol{\mu}'_{i2}, \cdots, \boldsymbol{\mu}'_{it})'$ and $\boldsymbol{B}_i = \text{diag}(\Sigma_{i1}, \Sigma_{i2}, \cdots, \Sigma_{it})$.

The unbiased estimating equation (35) is known as Independent Estimating Equation (IEE). It is also the score equation under the assumption of independence of the repeated measurements on each subject. The GEE is an extension of IEE for correlated repeated measurements. It is obtained by replacing $\boldsymbol{B}_i$ in (35) with a symmetric weight matrix $\boldsymbol{W}_i$ that has the same diagonal block matrices as $\boldsymbol{B}_i$. Thus the GEE is given by

$$\psi = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}'_i}{\partial \boldsymbol{\theta}} \boldsymbol{W}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = \mathbf{0} \tag{36}$$

In the above equation the weight matrices are constructed using a *working correlation matrix*, $\boldsymbol{R}(\boldsymbol{\lambda})$, as $= \boldsymbol{W}_i = \boldsymbol{W}_i(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \boldsymbol{A}_i^{1/2}\boldsymbol{R}(\boldsymbol{\lambda})\boldsymbol{A}_i^{1/2}$. The correlation matrix $\boldsymbol{R}(\boldsymbol{\lambda})$ can be any structured matrix such as CS, AR(1) or UN determined by an unknown parameter $\boldsymbol{\lambda}$ as defined in (33). Since the equations (36) are not linear, the solution for $\boldsymbol{\theta}$ can be obtained by an iterative procedure described below.

Step 1. Initially a solution $\widehat{\boldsymbol{\theta}}_I$ is obtain by solving IEE's.

Step 2. Consider $\widehat{\boldsymbol{\theta}}_I$ as current estimate of $\boldsymbol{\theta}$ and update the weight matrices and $\widehat{\boldsymbol{\theta}}$ as,

$$\widehat{\boldsymbol{\theta}}_{m+1} = \widehat{\boldsymbol{\theta}}_m + \left[\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\theta}} \widehat{\boldsymbol{W}}_i^{-1}(\widehat{\boldsymbol{\lambda}}_m, \widehat{\boldsymbol{\theta}}_m) \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}}\right]^{-1} \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i'}{\partial \boldsymbol{\theta}} \widehat{\boldsymbol{W}}_i^{-1}(\widehat{\boldsymbol{\lambda}}_m, \widehat{\boldsymbol{\theta}}_m)(\boldsymbol{Y}_i - \boldsymbol{\mu}_i)$$

Step 3. Iterate until $\widehat{\boldsymbol{\theta}}_{m+1} \cong \widehat{\boldsymbol{\theta}}_m$.

In the above algorithm $\widehat{\boldsymbol{\lambda}}_m$ is an estimate of $\boldsymbol{\lambda}$. For binary data $\boldsymbol{\lambda}$ is a real valued parameter and Liang and Zeger (1986) gave methods of estimation for various correlation structures by the method of moments. These estimates were extended to the multinomial case by Lipsitz et al. (1994). Also as stated in Lipsitz et al. (1994), $\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ is asymptotically normal. Statistical software packages SAS, S − Plus and R have procedures for fitting the GEE models for binary outcomes but not for correlated repeated multinomial outcomes.

III.3.4 DRAWBACKS OF GEE

One of the advantages of the GEE methodology is that it requires minimal assumptions such as specification of the first two moments. Unlike the computations for fully specified models such as latent variable models, GEE computations are relatively easy. Regardless of these advantages, the GEE method has a number of theoretical flaws, see Sabo and Chaganty (2010). One major flaw is that the working correlation matrix may not correspond to any joint distribution for $\boldsymbol{Y}_i$, that is, there may not exist a joint distribution for $\boldsymbol{Y}_i$ that has the specified working correlation matrix $\boldsymbol{R}(\boldsymbol{\lambda})$. Secondly, for multinomial random variables there are severe restrictions on the parameter $\boldsymbol{\lambda}$ imposed by the marginal means, and the GEE method ignores these restrictions and thus the method could lead to incorrect inferences. We study these restrictions on the parameter $\boldsymbol{\lambda}$ in Chapter IV. In the following section, we compare the large sample efficiencies of GEE estimates for several choices of $\boldsymbol{R}(\boldsymbol{\lambda})$ with respect to the maximum likelihood estimates of multivariate ordered probit models.

III.4 EFFICIENCY COMPARISONS

III.4.1 ASYMPTOTIC VARIANCES

In Section II.4.3, we have discussed maximum likelihood estimation of the threshold intercepts $\boldsymbol{\alpha}$, regression parameter $\boldsymbol{\beta}$ and latent correlation parameter $\boldsymbol{\rho}$, for the MCD model constructed with Gaussian copula. From the general theory of maximum likelihood estimation, it follows that the asymptotic covariance matrix of the maximum likelihood estimates is given by the inverse of the Fisher information matrix. For a known structured latent correlation matrix, $\boldsymbol{R}(\boldsymbol{\rho})$, the Fisher information matrix for $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ can be calculated as

$$\mathcal{I} = \sum_{i=1}^{n} \sum_{\boldsymbol{y}} \frac{\partial \pi_t(\boldsymbol{y}_i; \mathbf{0}, \boldsymbol{R}(\boldsymbol{\rho}))}{\partial \boldsymbol{\theta}} \frac{\partial \pi_t(\boldsymbol{y}_i; \mathbf{0}, \boldsymbol{R}(\boldsymbol{\rho}))}{\partial \boldsymbol{\theta}'} / \pi_t(\boldsymbol{y}_i; \mathbf{0}, \boldsymbol{R}(\boldsymbol{\rho})) \tag{37}$$

where the inner sum is taken over K^t possible vectors of $\boldsymbol{y}$. On the other hand, if $\widehat{\boldsymbol{\theta}}_{gee} = (\widehat{\boldsymbol{\alpha}}_{gee}, \widehat{\boldsymbol{\beta}}_{gee})$ is the solution for the weighted estimation equation (36) then the asymptotic covariance matrix of $\widehat{\boldsymbol{\theta}}_{gee}$ is given by $\boldsymbol{V}_\psi = (-D_\psi^{-1}) M_\psi (-D_\psi^{-1})$, where $M_\psi = \mathrm{Cov}(\boldsymbol{\psi})$ and $D_\psi = E(\partial \boldsymbol{\psi} / \partial \boldsymbol{\theta}')$. See Chaganty and Joe (2004) for details. Suppose the true covariance matrix of $\boldsymbol{Y}_i$, is Σ_i then

$$M_\psi = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_i^{-1} \Sigma_i \boldsymbol{W}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}'}, \tag{38}$$

$$-D_\psi = \sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\theta}} \boldsymbol{W}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}'} \tag{39}$$

As shown in Chaganty and Joe (2004), the optimal choice for the weight matrix $\boldsymbol{W}_i$ is Σ_i and for this choice the covariance matrix $\boldsymbol{V}_\psi$ reduces

$$\boldsymbol{V}_{opt} = \left(\sum_{i=1}^{n} \frac{\partial \boldsymbol{\mu}_i{}'}{\partial \boldsymbol{\theta}} \Sigma_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\theta}'} \right)^{-1}. \tag{40}$$

For efficiency comparisons between the maximum likelihood and the weighted estimating equation we compute the Fisher information (37), the matrix $\boldsymbol{V}_{\psi}$ and $\boldsymbol{V}_{opt}$. The later two matrices require calculation of the true covariance matrix Σ_i of $\boldsymbol{Y}_i$ based on multivariate ordered probit model. These calculations are described below.

Note that the binary choice vector representation of ordinal response Y_{ij} is $\boldsymbol{Y}_{ij} = (Y_{ij1}, Y_{ij2}, \cdots, Y_{ijK-1})'$ where $Y_{ijr} = 1$ if $Y_{ij} = r$ and 0 otherwise. According to this representation $\boldsymbol{Y}_{ij}$ is distributed as multinomial with 1-trial and mean $\boldsymbol{\mu}_{ij} = (\mu_{ij1}, \mu_{ij2}, \cdots, \mu_{ijK-1})'$. The covariance matrix of $\boldsymbol{Y}_{ij}$ is $\Sigma_{ij} = \text{diag}(\boldsymbol{\mu}_{ij}) - \boldsymbol{\mu}_{ij}\boldsymbol{\mu}_{ij}'$.

For any $j \neq k$ and $r, s \in \{1, 2, \cdots, K-1\}$ we have

$$\begin{aligned} \text{Corr}(y_{ijr}, y_{iks}) &= \frac{E(y_{ijr}y_{iks}) - E(y_{ijr})E(y_{iks})}{(V(y_{ijr})V(y_{iks}))^{1/2}} \\ &= \frac{E(y_{ijr}y_{iks}) - \mu_{ijr}\mu_{iks}}{(\mu_{ijr}(1-\mu_{ijr})\mu_{iks}(1-\mu_{iks}))^{1/2}}. \end{aligned} \tag{41}$$

Using the multivariate ordered probit model we have

$$\mu_{ijr} = \Phi(\alpha_r + \boldsymbol{x}_{ij}'\boldsymbol{\beta}) \quad - \quad \Phi(\alpha_{r-1} + \boldsymbol{x}_{ij}'\boldsymbol{\beta}) \tag{42}$$

and

$$\begin{aligned} E(Y_{ijr}Y_{iks}) &= P(y_{ij} = r, y_{ik} = s) \\ &= \int_{\alpha_{r-1}+\boldsymbol{x}_{ij}'\boldsymbol{\beta}}^{\alpha_r+\boldsymbol{x}_{ij}'\boldsymbol{\beta}} \int_{\alpha_{s-1}+\boldsymbol{x}_{ik}'\boldsymbol{\beta}}^{\alpha_s+\boldsymbol{x}_{ik}'\boldsymbol{\beta}} \phi_2(z_j, z_k; (0,0), (1,1), \rho_{jk})\, dz_j\, dz_k, \end{aligned} \tag{43}$$

where $\rho_{jk} = \text{Corr}(Z_j, Z_k)$ is the latent correlation. The $\text{Cov}(\boldsymbol{Y}_{ij}, \boldsymbol{Y}_{ik}) = \Sigma_{i,jk}$ can be calculated using (41), (42) and (43). These matrices can be put together to obtain

$$\text{Cov}(\boldsymbol{Y}_i) = \Sigma_i = \begin{pmatrix} \Sigma_{i1} & \Sigma_{i,12} & \Sigma_{i,13} & \cdots & \Sigma_{i,1t} \\ \Sigma_{i,21} & \Sigma_{i2} & \Sigma_{i,23} & \cdots & \Sigma_{i,2t} \\ \Sigma_{i,31} & \Sigma_{i,32} & \Sigma_{i3} & \cdots & \Sigma_{i,3t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Sigma_{i,t1} & \Sigma_{i,t2} & \Sigma_{i,t3} & \cdots & \Sigma_{it} \end{pmatrix}$$

III.4.2 COMPARISONS

We computed scaled variances of $\widehat{\boldsymbol{\theta}}_{mle}$, $\widehat{\boldsymbol{\theta}}_{opt}$, and $\widehat{\boldsymbol{\theta}}_{gee}$ taking the diagonal elements of $\mathcal{I}^{-1}$, $\boldsymbol{V}_{\psi}$ and $\boldsymbol{V}_{opt}$ respectively for different values of $\boldsymbol{\theta} = (\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $\boldsymbol{\rho}$, assuming the true model is multivariate ordered probit model. For each set of $\boldsymbol{\theta}$ values, we consider several choices for $\boldsymbol{\lambda}$ to compute $\boldsymbol{V}_{\psi}$ and $\boldsymbol{V}_{opt}$. Also note that the $\boldsymbol{\lambda}$ values are chosen such that the correlation matrix $\boldsymbol{R}(\boldsymbol{\lambda})$ given in equation (33) is positive definite for all subjects. In the model we took two covariates, $\boldsymbol{x}_{ij} = (x_{ij1}, x_{ij2})'$, where x_{ij1} and x_{ij2} are taken from uniform random variables in the interval [-1,1] and [0,1] respectively. The scaled variances and efficiencies are given in Tables 16, 17, 18, and 19. In Table 16 we choose $t = 3$ whereas in Table 17 we consider $t = 4$ with CS latent correlation structure. Similarly in Table 18 we took $t = 3$ whereas in Table 19 we choose $t = 4$ with AR(1) latent correlation structure. The results in the four tables show that the GEE method is inefficient when compared to maximum likelihood estimates arising from the multivariate ordered probit model.

Table 16: Scaled diagonal elements of $\mathcal{I}^{-1}$, V_{ψ}, and V_{opt} with CS correlation structures. Efficiencies are given in parenthesis.

Method				$nV(\alpha_1)$	$nV(\alpha_2)$	$nV(\beta_1)$	$nV(\beta_2)$
MLE				1.764(1.000)	1.806(1.000)	0.715(1.000)	2.803(1.000)
Optimal				1.948(0.906)	1.979(0.913)	0.939(0.761)	3.650(0.768)
	GEE						
λ_1	λ_2	λ_3	λ_4				
0.016	0.044	0.016	0.008	2.357(0.749)	2.398(0.753)	1.365(0.524)	5.294(0.529)
0.093	0.084	0.044	0.064	2.164(0.815)	2.206(0.819)	1.156(0.618)	4.489(0.624)
0.004	0.084	0.060	0.014	2.343(0.753)	2.381(0.759)	1.348(0.530)	5.228(0.536)
0.092	0.027	0.055	0.024	2.170(0.813)	2.213(0.816)	1.168(0.612)	4.532(0.618)
0.072	0.040	0.092	0.011	2.192(0.805)	2.233(0.809)	1.189(0.601)	4.616(0.607)
0.063	0.071	0.059	0.028	2.212(0.798)	2.252(0.802)	1.210(0.591)	4.694(0.597)
0.023	0.064	0.067	0.025	2.298(0.768)	2.337(0.773)	1.301(0.549)	5.047(0.555)
0.054	0.081	0.035	0.022	2.231(0.791)	2.271(0.795)	1.231(0.581)	4.775(0.587)
0.078	0.001	0.016	0.088	2.217(0.796)	2.260(0.799)	1.217(0.587)	4.722(0.594)
0.072	0.076	0.008	0.045	2.202(0.801)	2.244(0.805)	1.200(0.596)	4.657(0.602)

Parameter values are $t = 3$, $\rho = 0.7$, $\alpha_1 = 0$, $\alpha_2 = 0.42$, $\beta_1 = 0.25$, $\beta_2 = 0.45$ and $n = 5000$.

Table 17: Scaled diagonal elements of $\mathcal{I}^{-1}$, V_ψ, and V_{opt} with CS correlation structures. Efficiencies are given in parenthesis.

Method				$nV(\alpha_1)$	$nV(\alpha_2)$	$nV(\beta_1)$	$nV(\beta_2)$
MLE				2.007(1.000)	2.109(1.000)	1.177(1.000)	4.736(1.000)
Optimal				2.060(0.974)	2.190(0.963)	1.226(0.961)	4.951(0.956)
GEE							
λ_1	λ_2	λ_3	λ_4				
0.034	0.061	0.067	0.016	2.223(0.902)	2.354(0.896)	1.393(0.845)	5.608(0.844)
0.088	0.019	0.053	0.098	2.157(0.930)	2.296(0.918)	1.320(0.892)	5.320(0.890)
0.026	0.020	0.086	0.066	2.228(0.901)	2.361(0.893)	1.397(0.843)	5.624(0.842)
0.025	0.014	0.080	0.026	2.225(0.902)	2.356(0.895)	1.396(0.844)	5.618(0.843)
0.078	0.062	0.031	0.046	2.164(0.927)	2.298(0.917)	1.329(0.886)	5.355(0.884)
0.012	0.072	0.037	0.096	2.262(0.887)	2.396(0.880)	1.432(0.822)	5.762(0.822)
0.060	0.006	0.035	0.090	2.178(0.921)	2.313(0.912)	1.345(0.875)	5.419(0.874)
0.059	0.056	0.029	0.073	2.184(0.919)	2.319(0.909)	1.350(0.872)	5.440(0.870)
0.026	0.009	0.038	0.036	2.232(0.899)	2.363(0.892)	1.403(0.839)	5.647(0.839)
0.004	0.090	0.050	0.065	2.280(0.880)	2.412(0.874)	1.451(0.811)	5.839(0.811)

Parameter values are $t = 4$, $\rho = 0.5$, $\alpha_1 = 0.64$, $\alpha_2 = 0.92$, $\beta_1 = 0.1$, $\beta_2 = 0.4$ and $n = 5000$.

Table 18: Scaled diagonal elements of $\mathcal{I}^{-1}$, V_ψ, and V_{opt} with AR(1) correlation structures. Efficiencies are given in parenthesis.

Method				$nV(\alpha_1)$	$nV(\alpha_2)$	$nV(\beta_1)$	$nV(\beta_2)$
MLE				1.837(1.000)	1.868(1.000)	0.868(1.000)	3.464(1.000)
Optimal				1.926(0.953)	1.959(0.954)	0.975(0.891)	3.869(0.895)
GEE							
λ_1	λ_2	λ_3	λ_4				
0.092	0.054	0.007	0.077	2.197(0.840)	2.241(0.838)	1.222(0.705)	4.941(0.708)
0.084	0.089	0.024	0.073	2.191(0.838)	2.234(0.836)	1.230(0.706)	4.900(0.707)
0.098	0.059	0.094	0.011	2.162(0.849)	2.203(0.848)	1.201(0.723)	4.788(0.723)
0.005	0.026	0.046	0.011	2.367(0.776)	2.407(0.776)	1.410(0.616)	5.621(0.616)
0.049	0.043	0.055	0.046	2.256(0.814)	2.296(0.813)	1.297(0.669)	5.171(0.670)
0.018	0.055	0.026	0.083	2.323(0.791)	2.363(0.790)	1.364(0.636)	5.438(0.637)
0.006	0.048	0.079	0.024	2.331(0.788)	2.370(0.788)	1.373(0.632)	5.474(0.633)
0.000	0.047	0.094	0.080	2.334(0.787)	2.374(0.787)	1.374(0.632)	5.476(0.632)
0.046	0.004	0.041	0.078	2.281(0.805)	2.323(0.804)	1.324(0.656)	5.276(0.657)
0.019	0.062	0.038	0.053	2.324(0.794)	2.363(0.794)	1.355(0.641)	5.460(0.642)

Parameter values are $t = 3$, $\rho = 0.7$, $\alpha_1 = 0$, $\alpha_2 = 0.42$, $\beta_1 = 0.25$, $\beta_2 = 0.45$ and $n = 5000$.

Table 19: Scaled diagonal elements of $\mathcal{I}^{-1}$, V_ψ, and V_{opt} with AR(1) correlation structures. Efficiencies are given in parenthesis.

Method				$nV(\alpha_1)$	$nV(\alpha_2)$	$nV(\beta_1)$	$nV(\beta_2)$
MLE				1.947(1.000)	2.048(1.000)	1.251(1.000)	5.000(1.000)
Optimal				1.968(0.989)	2.084(0.983)	1.273(0.983)	5.093(0.982)
GEE							
λ_1	λ_2	λ_3	λ_4				
0.028	0.088	0.055	0.031	2.116(0.920)	2.232(0.918)	1.415(0.885)	5.653(0.885)
0.027	0.061	0.024	0.024	2.113(0.921)	2.229(0.919)	1.414(0.885)	5.647(0.885)
0.094	0.015	0.062	0.047	2.056(0.947)	2.174(0.942)	1.355(0.923)	5.418(0.923)
0.027	0.094	0.043	0.003	2.111(0.922)	2.225(0.920)	1.411(0.887)	5.636(0.887)
0.093	0.088	0.021	0.039	2.062(0.944)	2.182(0.939)	1.361(0.919)	5.442(0.919)
0.080	0.092	0.047	0.048	2.088(0.933)	2.210(0.927)	1.385(0.904)	5.537(0.903)
0.032	0.075	0.081	0.020	2.114(0.921)	2.230(0.919)	1.412(0.886)	5.644(0.886)
0.009	0.055	0.030	0.029	2.137(0.911)	2.253(0.909)	1.437(0.871)	5.742(0.871)
0.056	0.085	0.069	0.016	2.095(0.929)	2.212(0.926)	1.393(0.898)	5.569(0.898)
0.073	0.095	0.021	0.013	2.067(0.942)	2.184(0.938)	1.367(0.915)	5.465(0.915)

Parameter values are $t = 4$, $\rho = 0.5$, $\alpha_1 = 0.64$, $\alpha_2 = 0.92$, $\beta_1 = 0.1$, $\beta_2 = 0.4$ and $n = 5000$.

CHAPTER IV

CORRELATED MULTINOMIAL RANDOM VECTORS

IV.1 INTRODUCTION

The GEE methodology that we discussed in Chapter III is a non-likelihood method. Indeed a joint distribution between two dependent multinomial random vectors with specified mean and correlations as in the GEE model assumptions, may or may not exist. In this chapter we investigate the conditions on the correlations which guarantee the existence of a joint distribution for two or more dependent multinomial random vectors. Unlike Gaussian random variables, for binary random variables the correlations are restricted by some functions of the marginal means. As a simple example consider two binary random variables Y_1 and Y_2 with means p_1 and p_2 and correlation ρ. It is well known that the joint distribution for Y_1 and Y_2 exists if and only if $L(p_1, p_2) \leq \rho \leq U(p_1, p_2)$, where $L(p_1, p_2) = \max\left\{-\sqrt{\frac{p_1 p_2}{q_1 q_2}}, -\sqrt{\frac{q_1 q_2}{p_1 p_2}}\right\}$ and $U(p_1, p_2) = \min\left\{\sqrt{\frac{p_1 q_2}{p_2 q_1}}, \sqrt{\frac{p_2 q_1}{p_1 q_2}}\right\}$. These lower and upper limits are known as Fréchet bounds, see Chaganty and Joe (2006). In this chapter, we study the relationship of marginal means and correlation matrix elements between multinomial random vectors. In the next section we focus on the correlation bounds for two dependent multinomial random vectors, and later we extend these results to three dependent multinomial random vectors.

IV.2 BIVARIATE MULTINOMIAL RANDOM VECTORS

IV.2.1 INTRODUCTION

Suppose Y_1 and Y_2 are two categorical random variables that take values $1, 2, \ldots, K$. Then as discussed in Chapter III, for each Y_i we can associate a binary choice vector $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, \cdots, Y_{iK-1})'$ as

$$Y_{ij} = \begin{cases} 1 & \text{if } Y_i = j \\ 0 & \text{otherwise} \end{cases} \quad \text{for } j = 1, 2, ..., K-1.$$

Normally we denote $Y_{iK} = 1 - \sum_{j=1}^{K-1} Y_{ij}$. Note that $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are two multinomial random vectors with K categories and one trial, that is, $\boldsymbol{Y}_i \sim Mult(1, p_{i1}, p_{i2}, ..., p_{iK-1})$, $i = 1, 2$. By definition, the multinomial random vector $\boldsymbol{Y}_i = (Y_{i1}, Y_{i2}, ..., Y_{iK-1})'$ is a restricted binary vector. For example, when $K = 3$, it can take only three possible values $(1,0), (0,1)$ and $(0,0)$. We need some new notation to facilitate further discussion. Let $p_{i,c_i} = P(Y_i = c_i)$ and $p_{12,c_1c_2} = P(Y_1 = c_1, Y_2 = c_2)$ where $c_1, c_2 \in \{1, 2, ..., K-1\}$. Note that $p_{i,c_i} = P(Y_{ic_i} = 1)$ and $p_{12,c_1c_2} = P(Y_{1c_1} = 1, Y_{2c_2} = 1)$. With this notation the bivariate distribution of (Y_1, Y_2) can be expressed as in Table 20 below.

Table 20: Joint distribution of Y_1 and Y_2

$Y_1 \backslash Y_2$	1	2	...	K	
1	$p_{12,11}$	$p_{12,12}$	...	$p_{12,1K}$	$p_{1,1}$
2	$p_{12,21}$	$p_{12,22}$	...	$p_{12,2K}$	$p_{1,2}$
⋮	⋮	⋮	⋱	⋮	⋮
K	$p_{12,K1}$	$p_{12,K2}$	...	$p_{12,KK}$	$p_{1,K}$
	$p_{2,1}$	$p_{2,2}$	...	$p_{2,K}$	1

IV.2.2 BETWEEN AND WITHIN CORRELATIONS FOR MULTINOMIAL VECTORS

The correlation concept is well defined for two binary random variables. However for two dependent multinomial random vectors there are several correlations to consider. We know that the covariance matrix of a multinomial random vector, $\boldsymbol{Y}_i$, is

$$\text{Cov}\begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{iK-1} \end{pmatrix} = \begin{pmatrix} p_{i,1}(1-p_{i,1}) & -p_{i,1}p_{i,2} & \cdots & -p_{i,1}p_{i,K-1} \\ -p_{i,2}p_{i,1} & p_{i,2}(1-p_{i,2}) & \cdots & -p_{i,2}p_{i,K-1} \\ \vdots & \vdots & \ddots & \vdots \\ -p_{i,K-1}p_{i,1} & -p_{i,K-1}p_{i,2} & \cdots & p_{i,K-1}(1-p_{i,K-1}) \end{pmatrix}$$

$$= \begin{pmatrix} \sigma^2_{i,1} & \sigma_{ii,12} & \cdots & \sigma_{ii,1K-1} \\ \sigma_{ii,21} & \sigma^2_{i,2} & \cdots & \sigma_{ii,2K-1} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{ii,K-11} & \sigma_{ii,K-12} & \cdots & \sigma^2_{i,K-1} \end{pmatrix}$$

The above covariance matrix measures the association within levels of the multinomial random vector $\boldsymbol{Y}_i$. Next, the correlation *between* levels of two multinomial random vectors is defined as

$$\rho_{12,c_1c_2} = \frac{\text{Cov}(Y_{1c_1}, Y_{2c_2})}{\sqrt{(V(Y_{1c_1})V(Y_{2c_2}))}} = \frac{E(Y_{1c_1}Y_{2c_2}) - p_{1,c_1}p_{2,c_2}}{\sigma_{1,c_1}\sigma_{2,c_2}} = \frac{P(Y_{1c_1}=1, Y_{2c_2}=1) - p_{1,c_1}p_{2,c_2}}{\sigma_{1,c_1}\sigma_{2,c_2}} = \frac{p_{12,c_1c_2} - p_{1,c_1}p_{2,c_2}}{\sigma_{1,c_1}\sigma_{2,c_2}}. \quad (44)$$

Using these definitions, we can write the correlation matrix between two multinomial random vectors as

$$\text{Corr}\left(\begin{array}{c} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1K-1} \\ \hline Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2K-1} \end{array}\right) = \left(\begin{array}{cccc|cccc} 1 & \rho_{11,12} & \cdots & \rho_{11,1K-1} & \rho_{12,11} & \rho_{12,12} & \cdots & \rho_{12,1K-1} \\ \rho_{11,21} & 1 & \cdots & \rho_{11,2K-1} & \rho_{12,21} & \rho_{12,22} & \cdots & \rho_{12,2K-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{11,K-11} & \rho_{11,K-12} & \cdots & 1 & \rho_{12,K-11} & \rho_{12,K-12} & \cdots & \rho_{12,K-1K-1} \\ \hline \rho_{12,11} & \rho_{12,21} & \cdots & \rho_{12,K-11} & 1 & \rho_{22,12} & \cdots & \rho_{22,1K-1} \\ \rho_{12,12} & \rho_{12,22} & \cdots & \rho_{12,K-12} & \rho_{22,21} & 1 & \cdots & \rho_{22,2K-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{12,1K-1} & \rho_{12,2K-1} & \cdots & \rho_{12,K-1K-1} & \rho_{22,K-11} & \rho_{22,K-12} & \cdots & 1 \end{array}\right)$$

$$= \left(\begin{array}{c|c} \boldsymbol{R}_{11} & \boldsymbol{R}_{12} \\ \hline \boldsymbol{R}_{21} & \boldsymbol{R}_{22} \end{array}\right) = \boldsymbol{R} \quad (45)$$

Note that $\boldsymbol{R}_{11}$ and $\boldsymbol{R}_{22}$ are symmetric and $\boldsymbol{R}'_{12} = \boldsymbol{R}_{21}$. Suppose we assume for parsimony $\rho_{12,c_1c_2} = \rho_{12} = \rho$ for all c_1, c_2, then the correlation matrix (45) becomes

$$\text{Corr}\left(\begin{array}{c} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1K-1} \\ \hline Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2K-1} \end{array}\right) = \left(\begin{array}{cccc|cccc} 1 & \rho_{11,12} & \cdots & \rho_{11,1K-1} & \rho & \rho & \cdots & \rho \\ \rho_{11,21} & 1 & \cdots & \rho_{11,2K-1} & \rho & \rho & \cdots & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{11,K-11} & \rho_{11,K-12} & \cdots & 1 & \rho & \rho & \cdots & \rho \\ \hline \rho & \rho & \cdots & \rho & 1 & \rho_{22,12} & \cdots & \rho_{22,1K-1} \\ \rho & \rho & \cdots & \rho & \rho_{22,21} & 1 & \cdots & \rho_{22,2K-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \cdots & \rho & \rho_{22,K-11} & \rho_{22,K-12} & \cdots & 1 \end{array}\right)$$

$$= \left(\begin{array}{cc} \boldsymbol{R}_{11} & \rho\text{J} \\ \rho\text{J} & \boldsymbol{R}_{22} \end{array}\right) = \boldsymbol{R} \quad (46)$$

where $\mathbf{J}$ is a matrix of ones. It can be easily checked that the determinant of $\boldsymbol{R}_{ii}$ matrix is $p_{i,K}/(\prod_{c_i=1}^{K-1}(1-p_{i,c_i}))$ and it is always positive. The following theorem provides the range of ρ for which the determinant of the matrix $\boldsymbol{R}$ is nonnegative.

THEOREM 1. *Suppose $\boldsymbol{R}$ is a correlation matrix between two multinomial random vectors with the assumption that $\rho_{12,c_1c_2} = \rho_{12} = \rho$ for all c_1, c_2. Then the range of ρ for which the determinant of the matrix $\boldsymbol{R}$ is non negative is*

$$\rho \leq \left| \frac{1}{\sqrt{(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})}} \right|.$$

Proof. First note that

$$\begin{aligned} |\boldsymbol{R}| &= |\boldsymbol{R}_{11}|\,|\boldsymbol{R}_{22} - \rho^2 \mathrm{J}\boldsymbol{R}_{11}^{-1}\mathrm{J}| \\ &= |\boldsymbol{R}_{11}|\,|\boldsymbol{R}_{22}|\,|\mathrm{I} - \rho^2 \mathrm{J}\boldsymbol{R}_{11}^{-1}\mathrm{J}\boldsymbol{R}_{22}^{-1}| \\ &= |\boldsymbol{R}_{11}|\,|\boldsymbol{R}_{22}|\,|\mathrm{I} - \rho^2 \mathbf{1}\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1}\mathbf{1}'\boldsymbol{R}_{22}^{-1}| \\ &= |\boldsymbol{R}_{11}|\,|\boldsymbol{R}_{22}|\,|1 - \rho^2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})| \end{aligned}$$

because $|\mathrm{I} - \mathrm{AB}| = |\mathrm{I} - \mathrm{BA}|$. Since $|\boldsymbol{R}_{ii}| > 0$ we have,

$$\begin{aligned} |\boldsymbol{R}| \geq 0 &\Leftrightarrow 1 - \rho^2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1}) \geq 0 \\ &\Leftrightarrow \rho \leq \left| \frac{1}{\sqrt{(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})}} \right|. \end{aligned}$$

This completes the proof.

Remark: The range of ρ for which the determinant of the matrix $\boldsymbol{R}$ is nonnegative is same as the positive semi-definite range for $\boldsymbol{R}$. We verified this result numerically but we do not have a formal proof.

IV.2.3 BOUNDS ON CORRELATION

In this section we obtain necessary and sufficient bounds on the parameter ρ such that a joint distribution with correlation structure (46) for the two dependent multinomial random vectors $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ exists.

THEOREM 2. *Suppose $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are two multinomial random vectors with one trial and probabilities $\boldsymbol{p}_1 = (p_{1,1}, p_{1,2}, \cdots, p_{1,K})'$ and $\boldsymbol{p}_2 = (p_{2,1}, p_{2,2} \cdots, p_{2,K})'$ respectively. Suppose that the correlation matrix of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ is*

$$\boldsymbol{R} = \begin{pmatrix} \boldsymbol{R}_{11} & \rho J \\ \rho J & \boldsymbol{R}_{22} \end{pmatrix}.$$

Then the joint distribution for $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ exists if and only if

$$\max\{L_{21}(\boldsymbol{p}_1,\boldsymbol{p}_2)\ ,\ L_{22}(\boldsymbol{p}_1,\boldsymbol{p}_2)\} \leq \rho \leq \min\{U_{21}(\boldsymbol{p}_1,\boldsymbol{p}_2)\ ,\ U_{22}(\boldsymbol{p}_1,\boldsymbol{p}_2)\}, \tag{47}$$

where

$$\begin{aligned} L_{21}(\boldsymbol{p}_1,\boldsymbol{p}_2) &= \max\left\{\frac{-(\sum_{i=1}^{K-1}p_{1,i})(\sum_{i=1}^{K-1}p_{2,i})}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})}, \frac{-p_{1,K}p_{2,K}}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})}\right\}, \\ L_{22}(\boldsymbol{p}_1,\boldsymbol{p}_2) &= \max\left\{\frac{-p_{i,c_i}(\sum_{c_j=1}^{K-1}p_{j,c_j})}{\sigma_{i,c_i}(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}\ ; 1\leq i\neq j\leq 2, c_i, c_j\in\{1,2,...,K-1\}\right\}, \end{aligned}$$

and

$$\begin{aligned} U_{21}(\boldsymbol{p}_1,\boldsymbol{p}_2) &= \min\left\{\frac{(\sum_{i=1}^{K-1}p_{1,i})p_{2,K}}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})}, \frac{(\sum_{i=1}^{K-1}p_{2,i})p_{1,K}}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})}\right\}, \\ U_{22}(\boldsymbol{p}_1,\boldsymbol{p}_2) &= \min\left\{\frac{p_{i,c_i}p_{j,K}}{\sigma_{i,c_i}(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}\ 1\leq i\neq j\leq 2, c_i, c_j\in\{1,2,...,K-1\}\right\}. \end{aligned}$$

Proof. We prove the theorem for $K = 3$. The joint distribution of Y_1 and Y_2 given in Table 20 when $K = 3$ reduces to the form given in Table 21. The joint distribution in this table is completely determined by the probabilities $p_{12,11}, p_{12,12}, p_{12,21}, p_{12,22}$ and $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}$, as shown in Table 22. The probability distribution given in Table 22 is legitimate if and only if all the nine probabilities listed are greater than zero. And these nine restrictions lead to the following inequalities

Table 21: Joint distribution of Y_1 and Y_2 when $K = 3$

$Y_1\backslash Y_2$	1	2	3	
1	$p_{12,11}$	$p_{12,12}$	$p_{12,13}$	$p_{1,1}$
2	$p_{12,21}$	$p_{12,22}$	$p_{12,23}$	$p_{1,2}$
3	$p_{12,31}$	$p_{12,32}$	$p_{12,33}$	$p_{1,3}$
	$p_{2,1}$	$p_{2,2}$	$p_{2,3}$	1

Table 22: Bivariate probabilities and dependencies

Y_1	Y_2	$P(Y_1 = y_1, Y_2 = y_2)$
1	1	$p_{12,11}$
1	2	$p_{12,12}$
1	3	$p_{1,1} - p_{12,11} - p_{12,12}$
2	1	$p_{12,21}$
2	2	$p_{12,22}$
2	3	$p_{1,2} - p_{12,21} - p_{12,22}$
3	1	$p_{2,1} - p_{12,11} - p_{12,21}$
3	2	$p_{2,2} - p_{12,12} - p_{12,22}$
3	3	$1 - p_{1,1} - p_{1,2} - p_{2,1} - p_{2,2} + p_{12,11} + p_{12,12} + p_{12,21} + p_{12,22}$

$$\begin{aligned} \max\{0, p_{1,1} + p_{1,2} + p_{2,1} + p_{2,2} - 1\} &\le p_{12,11} + p_{12,12} + p_{12,21} + p_{12,22} \le \\ &\quad \min\{p_{1,1} + p_{1,2}, p_{2,1} + p_{2,2}\} \qquad (48) \\ 0 \le p_{12,11} + p_{12,12} \le p_{1,1};\ 0 &\le p_{12,21} + p_{12,22} \le p_{1,2} \\ 0 \le p_{12,11} + p_{12,21} \le p_{2,1};\ 0 &\le p_{12,12} + p_{12,22} \le p_{2,2} \qquad (49) \end{aligned}$$

Since $\rho_{12,c_1c_2} = \rho_{12}$, we have $p_{12,c_1c_2} = p_{1,c_1}p_{2,c_2} + \rho_{12}\left(\sigma_{1,c_1}\sigma_{2,c_2}\right)$ for all c_1, c_2. Hence

$$p_{12,11} + p_{12,12} + p_{12,21} + p_{12,22} = \sum_{i=1}^{2}\sum_{j=1}^{2} p_{1,i}p_{2,j} + \rho \sum_{i=1}^{2}\sum_{j=1}^{2} \sigma_{1,i}\sigma_{2,j}.$$

It is easy to check that

$$\begin{aligned} (p_{1,1} + p_{1,2} + p_{2,1} + p_{2,2} - 1) - \sum_{i=1}^{2}\sum_{j=1}^{2} p_{1,i}p_{2,j} &= -p_{1,3}p_{2,3} \\ (p_{1,1} + p_{1,2}) - \sum_{i=1}^{2}\sum_{j=1}^{2} p_{1,i}p_{2,j} &= (p_{1,1} + p_{1,2})p_{2,3} \\ (p_{2,1} + p_{2,2}) - \sum_{i=1}^{2}\sum_{j=1}^{2} p_{1,i}p_{2,j} &= (p_{2,1} + p_{2,2})p_{1,3} \\ \sum_{i=1}^{2}\sum_{j=1}^{2} p_{1,i}p_{2,j} &= (p_{1,1} + p_{1,2})(p_{2,1} + p_{2,2}) \end{aligned}$$

Using the above identities we can rewrite the inequalities (48) as

$$\frac{\max(0, p_{1,1}+p_{1,2}+p_{2,1}+p_{2,2}-1)-\sum_{i=1}^{2}\sum_{j=1}^{2}p_{1,i}p_{2,j}}{\sum_{i=1}^{2}\sum_{j=1}^{2}\sigma_{1,i}\sigma_{2,j}} \le \rho \le \frac{\min(p_{1,1}+p_{1,2}, p_{2,1}+p_{2,2})-\sum_{i=1}^{2}\sum_{j=1}^{2}p_{1,i}p_{2,j}}{\sum_{i=1}^{2}\sum_{j=1}^{2}\sigma_{1,i}\sigma_{2,j}},$$

or simply

$$\max\left\{\frac{-(p_{1,1}+p_{1,2})(p_{2,1}+p_{2,2})}{(\sigma_{1,1}+\sigma_{1,2})(\sigma_{2,1}+\sigma_{2,2})}, \frac{-p_{1,3}p_{2,3}}{(\sigma_{1,1}+\sigma_{1,2})(\sigma_{2,1}+\sigma_{2,2})}\right\} \le \rho \le \min\left\{\frac{(p_{1,1}+p_{1,2})p_{2,3}}{(\sigma_{1,1}+\sigma_{1,2})(\sigma_{2,1}+\sigma_{2,2})}, \frac{(p_{2,1}+p_{2,2})p_{1,3}}{(\sigma_{1,1}+\sigma_{1,2})(\sigma_{2,1}+\sigma_{2,2})}\right\}. \quad (50)$$

Similar simplification of the inequalities (49) gives us

$$\begin{aligned}
\frac{-p_{1,1}(p_{2,1}+p_{2,2})}{\sigma_{1,1}(\sigma_{2,1}+\sigma_{2,2})} &\le \rho \le \frac{p_{1,1}p_{2,3}}{\sigma_{1,1}(\sigma_{2,1}+\sigma_{2,2})},\\
\frac{-p_{1,2}(p_{2,1}+p_{2,2})}{\sigma_{1,2}(\sigma_{2,1}+\sigma_{2,2})} &\le \rho \le \frac{p_{1,2}p_{2,3}}{\sigma_{1,2}(\sigma_{2,1}+\sigma_{2,2})},\\
\frac{-p_{2,1}(p_{1,1}+p_{1,2})}{\sigma_{2,1}(\sigma_{1,1}+\sigma_{1,2})} &\le \rho \le \frac{p_{2,1}p_{1,3}}{\sigma_{2,1}(\sigma_{1,1}+\sigma_{1,2})},\\
\frac{-p_{2,2}(p_{1,1}+p_{1,2})}{\sigma_{2,2}(\sigma_{1,1}+\sigma_{1,2})} &\le \rho \le \frac{p_{2,2}p_{1,3}}{\sigma_{2,2}(\sigma_{1,1}+\sigma_{1,2})},
\end{aligned}$$

which can be written compactly as

$$\max\left\{\frac{-p_{1,1}(p_{2,1}+p_{2,2})}{\sigma_{1,1}(\sigma_{2,1}+\sigma_{2,2})}, \frac{-p_{1,2}(p_{2,1}+p_{2,2})}{\sigma_{1,2}(\sigma_{2,1}+\sigma_{2,2})}, \frac{-p_{2,1}(p_{1,1}+p_{1,2})}{\sigma_{2,1}(\sigma_{1,1}+\sigma_{1,2})}, \frac{-p_{2,2}(p_{1,1}+p_{1,2})}{\sigma_{2,2}(\sigma_{1,1}+\sigma_{1,2})}\right\} \le \rho \le \min\left\{\frac{p_{1,1}p_{2,3}}{\sigma_{1,1}(\sigma_{2,1}+\sigma_{2,2})}, \frac{p_{1,2}p_{2,3}}{\sigma_{1,2}(\sigma_{2,1}+\sigma_{2,2})}, \frac{p_{2,1}p_{1,3}}{\sigma_{2,1}(\sigma_{1,1}+\sigma_{1,2})}, \frac{p_{2,2}p_{1,3}}{\sigma_{2,2}(\sigma_{1,1}+\sigma_{1,2})}\right\}. \quad (51)$$

Equations (50) and (51) together give the necessary and sufficient range for ρ such that the distribution in Table 22 is a proper probability distribution. In general for any K, define,

$$\begin{aligned}
L_{21}(\boldsymbol{p}_1, \boldsymbol{p}_2) &= \max\left\{\frac{-(\sum_{i=1}^{K-1}p_{1,i})(\sum_{i=1}^{K-1}p_{2,i})}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})}, \frac{-p_{1,K}p_{2,K}}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})}\right\},\\
L_{22}(\boldsymbol{p}_1, \boldsymbol{p}_2) &= \max\left\{\frac{-p_{i,c_i}(\sum_{c_j=1}^{K-1}p_{j,c_j})}{\sigma_{i,c_i}(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}; i \ne j \in \{1,2\}, c_i, c_j \in \{1,2,...,K-1\}\right\}.
\end{aligned}$$

and

$$U_{21}(\boldsymbol{p}_1, \boldsymbol{p}_2) = \min \left\{ \frac{(\sum_{i=1}^{K-1} p_{1,i}) p_{2,K}}{(\sum_{i=1}^{K-1} \sigma_{1,i})(\sum_{i=1}^{K-1} \sigma_{2,i})}, \frac{(\sum_{i=1}^{K-1} p_{2,i}) p_{1,K}}{(\sum_{i=1}^{K-1} \sigma_{1,i})(\sum_{i=1}^{K-1} \sigma_{2,i})} \right\},$$

$$U_{22}(\boldsymbol{p}_1, \boldsymbol{p}_2) = \min \left\{ \frac{p_{i,c_i} p_{j,K}}{\sigma_{i,c_i}(\sum_{c_j=1}^{K-1} \sigma_{j,c_j})} \; ; i, j \in \{1, 2\},\ i \neq j, c_i \in \{1, 2, ..., K-1\} \right\}.$$

Using the above notation the bounds (50) and (51) can be written as

$$\max \{L_{21}(\boldsymbol{p}_1, \boldsymbol{p}_2)\ ,\ L_{22}(\boldsymbol{p}_1, \boldsymbol{p}_2)\} \leq \rho \leq \min \{U_{21}(\boldsymbol{p}_1, \boldsymbol{p}_2)\ ,\ U_{22}(\boldsymbol{p}_1, \boldsymbol{p}_2)\} \qquad (52)$$

This completes the proof of the theorem when $K = 3$.

IV.2.4 CONSTRUCTION OF BIVARIATE MULTINOMIAL DISTRIBUTION

We have seen the feasible range for ρ such that a joint distribution for $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ exists. Now given a value for ρ within this feasible range the joint probabilities can be calculated using equation (44). For example, let $\boldsymbol{p}_1 = (0.2,\ 0.3,\ 0.4,\ 0.1)$ and $\boldsymbol{p}_2 = (0.4,\ 0.1,\ 0.2,\ 0.3)$ be fixed. The feasible range of the correlation ρ is $(-0.0187,\ 0.0247)$. A joint distribution with fixed marginals $\boldsymbol{p}_1$, $\boldsymbol{p}_2$ and correlation $\rho = -0.0181$ is given by

$Y_1 \backslash Y_2$	1	2	3	4	
1	0.0765	0.0178	0.0371	0.0686	0.2
2	0.1159	0.0275	0.0567	0.0999	0.3
3	0.1557	0.0373	0.0765	0.1305	0.4
4	0.0519	0.0173	0.0297	0.0010	0.1
	0.4	0.1	0.2	0.3	1

For different marginal multinomial distributions, the range of ρ such that the correlation matrix (46) is positive definite and the bounds (47) given in Theorem 2 are tabulated in Table 23.

Table 23 clearly shows that the positive definiteness range of ρ is wider than the range of ρ for which the joint distribution exists. Furthermore, the range of ρ for which the joint distribution exists is a proper sub-interval for which the correlation matrix is positive definite.

Table 23: Positive definite ranges and bounds on correlation ρ of the matrix $\boldsymbol{R}$ for different marginal probabilities

Marginal Probabilities*		Bounds on ρ			Positive Definite Range		
$\boldsymbol{p}_1$	$\boldsymbol{p}_2$	Lower	Upper	Range	Lower	Upper	Range
(0.1,0.1,0.8)	(0.1,0.1,0.8)	-0.1111	0.4444	0.5556	-0.4444	0.4444	0.8889
(0.1,0.1,0.8)	(0.1,0.4,0.5)	-0.1111	0.2110	0.3221	-0.4022	0.4022	0.8043
(0.1,0.1,0.8)	(0.1,0.7,0.2)	-0.1111	0.0879	0.1990	-0.3303	0.3303	0.6605
(0.1,0.1,0.8)	(0.4,0.1,0.5)	-0.1111	0.2110	0.3221	-0.4022	0.4022	0.8043
(0.1,0.1,0.8)	(0.4,0.4,0.2)	-0.2722	0.0680	0.3402	-0.2722	0.2722	0.5443
(0.1,0.1,0.8)	(0.7,0.1,0.2)	-0.1111	0.0879	0.1990	-0.3303	0.3303	0.6605
(0.1,0.4,0.5)	(0.1,0.4,0.5)	-0.2110	0.2110	0.4220	-0.3639	0.3639	0.7278
(0.1,0.4,0.5)	(0.1,0.7,0.2)	-0.1670	0.0879	0.2549	-0.2988	0.2988	0.5977
(0.1,0.4,0.5)	(0.4,0.1,0.5)	-0.2110	0.2110	0.4220	-0.3639	0.3639	0.7278
(0.1,0.4,0.5)	(0.4,0.4,0.2)	-0.1292	0.0680	0.1973	-0.2463	0.2463	0.4926
(0.1,0.4,0.5)	(0.7,0.1,0.2)	-0.1670	0.0879	0.2549	-0.2988	0.2988	0.5977
(0.1,0.7,0.2)	(0.1,0.7,0.2)	-0.0696	0.0879	0.1575	-0.2454	0.2454	0.4908
(0.1,0.7,0.2)	(0.4,0.1,0.5)	-0.1670	0.0879	0.2549	-0.2988	0.2988	0.5977
(0.1,0.7,0.2)	(0.4,0.4,0.2)	-0.0538	0.0680	0.1219	-0.2022	0.2022	0.4045
(0.1,0.7,0.2)	(0.7,0.1,0.2)	-0.0696	0.0879	0.1575	-0.2454	0.2454	0.4908
(0.4,0.1,0.5)	(0.4,0.1,0.5)	-0.2110	0.2110	0.4220	-0.3639	0.3639	0.7278
(0.4,0.1,0.5)	(0.4,0.4,0.2)	-0.1292	0.0680	0.1973	-0.2463	0.2463	0.4926
(0.4,0.1,0.5)	(0.7,0.1,0.2)	-0.1670	0.0879	0.2549	-0.2988	0.2988	0.5977
(0.4,0.4,0.2)	(0.4,0.4,0.2)	-0.0417	0.1667	0.2083	-0.1667	0.1667	0.3333
(0.4,0.4,0.2)	(0.7,0.1,0.2)	-0.0538	0.0680	0.1219	-0.2022	0.2022	0.4045
(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.0696	0.0879	0.1575	-0.2454	0.2454	0.4908

* $\boldsymbol{p}_i = (p_{i,1}, p_{i,2}, p_{i,3})$ for $i = 1, 2, 3$.

IV.3 TRIVARIATE MULTINOMIAL RANDOM VECTORS

IV.3.1 INTRODUCTION

In the previous subsection we studied bivariate joint distributions for correlated multinomial random vectors. Here we extend the results further to trivariate multinomial random vectors. Suppose Y_1, Y_2 and Y_3 are three dependent categorical random variables that can take one of the K categories. Let $\boldsymbol{Y}_1, \boldsymbol{Y}_2$, and $\boldsymbol{Y}_3$ be the binary choice vectors corresponding to Y_1, Y_2 and Y_3 respectively. We need some additional notation. Let $p_{i,c_i} = P(Y_i = c_i)$ and $p_{ij,c_ic_j} = P(Y_i = c_i,\, Y_j = c_j)$ where $1 \leq i, j \leq 3$ and $c_i, c_j \in \{1, 2, ..., K-1\}$. Further, let $p_{123,c_1c_2c_3} = P(Y_1 = c_1,\, Y_2 = c_2,\, Y_3 = c_3)$.

Similar to the bivariate situation, in the trivariate case, among the 3^K joint probabilities only few probabilities are flexible to vary when the marginal and bivariate joint probabilities are fixed. The dependent constraints among the probabilities can be summarized as follows.

$$\begin{aligned} p_{i,K} = 1 - \sum_{t=1}^{K-1} p_{i,t},\ p_{ij,c_iK} &= p_{i,c_i} - \sum_{t=1}^{K-1} p_{ij,c_it}, \\ p_{123,c_1c_2K} &= p_{12,c_1c_2} - \sum_{t=1}^{K-1} p_{123,c_1c_2t}, \\ p_{123,c_1Kc_3} &= p_{13,c_1c_3} - \sum_{t=1}^{K-1} p_{123,c_1tc_3}, \\ p_{123,Kc_2c_3} &= p_{23,c_2c_3} - \sum_{t=1}^{K-1} p_{123,tc_2c_3}, \end{aligned}$$

for $i, j \in \{1, 2, 3\}$ and $c_i \in \{1, 2, ..., K-1\}$. We focus on the special case $K = 3$, the general case can be handled similarly but the notation is cumbersome. The $3^3 = 27$ joint probabilities are summarized explicitly in Table 24.

IV.3.2 POSITIVE DEFINITE RANGES

When there are three categories ($K = 3$), the covariance matrix of the multinomial random vector is

$$\text{Cov}\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} = \begin{pmatrix} p_{i,1}(1-p_{i,1}) & -p_{i,1}p_{i,2} \\ -p_{i,2}p_{i,1} & p_{i,2}(1-p_{i,2}) \end{pmatrix} = \begin{pmatrix} \sigma_{i,1}^2 & \sigma_{ii,12} \\ \sigma_{ii,21} & \sigma_{i,2}^2 \end{pmatrix}, i = 1, 2, 3.$$

Similar to the bivariate case, denote the correlations between the levels of any two multinomial random variables as,

Table 24: Trivariate probabilities and dependencies

Y_1	Y_2	Y_3	$P(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3)$
1	1	1	$p_{123,111}$
1	1	2	$p_{123,112}$
1	1	3	$p_{12,11} - p_{123,111} - p_{123,112}$
1	2	1	$p_{123,121}$
1	2	2	$p_{123,122}$
1	2	3	$p_{12,12} - p_{123,121} - p_{123,122}$
1	3	1	$p_{13,11} - p_{123,111} - p_{123,121}$
1	3	2	$p_{13,12} - p_{123,112} - p_{123,122}$
1	3	3	$p_{1,1} - p_{12,11} - p_{12,12} - p_{13,11} - p_{13,12} + p_{123,111} + p_{123,121} + p_{123,112} + p_{123,122}$
2	1	1	$p_{123,211}$
2	1	2	$p_{123,212}$
2	1	3	$p_{12,21} - p_{123,211} - p_{123,212}$
2	2	1	$p_{123,221}$
2	2	2	$p_{123,222}$
2	2	3	$p_{12,22} - p_{123,221} - p_{123,222}$
2	3	1	$p_{13,21} - p_{123,211} - p_{123,221}$
2	3	2	$p_{13,22} - p_{123,212} - p_{123,222}$
2	3	3	$p_{1,2} - p_{12,21} - p_{12,22} - p_{13,21} - p_{13,22} + p_{123,211} + p_{123,221} + p_{123,212} + p_{123,222}$
3	1	1	$p_{23,11} - p_{123,111} - p_{123,211}$
3	1	2	$p_{23,12} - p_{123,112} - p_{123,212}$
3	1	3	$p_{2,1} - p_{12,11} - p_{12,21} - p_{23,11} - p_{23,12} + p_{123,111} + p_{123,211} + p_{123,112} + p_{123,212}$
3	2	1	$p_{23,21} - p_{123,121} - p_{123,221}$
3	2	2	$p_{23,22} - p_{123,122} - p_{123,222}$
3	2	3	$p_{2,2} - p_{12,12} - p_{12,22} - p_{23,21} - p_{23,22} + p_{123,121} + p_{123,221} + p_{123,122} + p_{123,222}$
3	3	1	$p_{3,1} - p_{13,11} - p_{13,21} - p_{23,11} - p_{23,21} + p_{123,111} + p_{123,211} + p_{123,121} + p_{123,221}$
3	3	2	$p_{3,2} - p_{13,12} - p_{13,22} - p_{23,12} - p_{23,22} + p_{123,112} + p_{123,212} + p_{123,122} + p_{123,222}$
3	3	3	$1 - p_{1,1} - p_{1,2} - p_{2,1} - p_{2,2} - p_{3,1} - p_{3,2} + p_{12,11} + p_{12,12} + p_{12,21} + p_{12,22} + p_{13,11} + p_{13,12} + p_{13,21} + p_{13,22} + p_{23,11} + p_{23,12} + p_{23,21} + p_{23,22} - p_{123,111} - p_{123,211} - p_{123,121} - p_{123,221} - p_{123,112} - p_{123,212} - p_{123,122} - p_{123,222}$

$$
\begin{aligned}
\rho_{ij,c_ic_j} = \frac{\text{Cov}(Y_{ic_i}, Y_{jc_j})}{\sqrt{[V(Y_{ic_i})V(Y_{jc_j})]}} &= \frac{E(Y_{ic_i}Y_{jc_j}) - p_{i,c_i}p_{j,c_j}}{\sigma_{i,c_i}\sigma_{j,c_j}} \\
&= \frac{P(Y_{ic_i} = 1, Y_{jc_j} = 1) - p_{i,c_i}p_{j,c_j}}{\sigma_{i,c_i}\sigma_{j,c_j}} \\
&= \frac{p_{ij,c_ic_j} - p_{i,c_i}p_{j,c_j}}{\sigma_{i,c_i}\sigma_{j,c_j}} \text{ for } i \neq j. \qquad (53)
\end{aligned}
$$

Using the above notation the correlation matrix for three multinomial random vectors when there are three categories can be written as

$$
\text{Corr}\begin{pmatrix} Y_{11} \\ Y_{12} \\ \hline Y_{21} \\ Y_{22} \\ \hline Y_{31} \\ Y_{32} \end{pmatrix} = \left(\begin{array}{cc|cc|cc}
1 & \rho_{11,12} & \rho_{12,11} & \rho_{12,12} & \rho_{13,11} & \rho_{13,12} \\
\rho_{11,21} & 1 & \rho_{12,21} & \rho_{12,22} & \rho_{13,21} & \rho_{13,22} \\ \hline
\rho_{12,11} & \rho_{12,21} & 1 & \rho_{22,12} & \rho_{23,11} & \rho_{23,12} \\
\rho_{12,12} & \rho_{12,22} & \rho_{22,21} & 1 & \rho_{23,21} & \rho_{23,22} \\ \hline
\rho_{13,11} & \rho_{13,21} & \rho_{23,11} & \rho_{23,21} & 1 & \rho_{33,12} \\
\rho_{13,12} & \rho_{13,22} & \rho_{23,12} & \rho_{23,22} & \rho_{33,21} & 1
\end{array}\right)
$$

As in the bivariate case, for parsimonious modeling we assume $\rho_{ij,c_ic_j} = \rho_{ij}$ for all $i \neq j$. This is also known as the unstructured (UN) correlation matrix. We can further reduce the number of correlation parameters by considering structured correlation matrices such as (i) compound symmetry (CS), $\rho_{ij} = \rho$ for all $i \neq j$, (ii) AR(1), $\rho^{|i-j|}$ for all $i \neq j$. When $K = 3$, these structured matrices take the form

$$
\boldsymbol{R}_{cs} = \left(\begin{array}{cc|cc|cc}
1 & \rho_{11,12} & \rho & \rho & \rho & \rho \\
\rho_{11,21} & 1 & \rho & \rho & \rho & \rho \\ \hline
\rho & \rho & 1 & \rho_{22,12} & \rho & \rho \\
\rho & \rho & \rho_{22,21} & 1 & \rho & \rho \\ \hline
\rho & \rho & \rho & \rho & 1 & \rho_{33,12} \\
\rho & \rho & \rho & \rho & \rho_{33,21} & 1
\end{array}\right) \qquad (54)
$$

$$
\boldsymbol{R}_{ar1} = \left(\begin{array}{cc|cc|cc}
1 & \rho_{11,12} & \rho & \rho & \rho^2 & \rho^2 \\
\rho_{11,21} & 1 & \rho & \rho & \rho^2 & \rho^2 \\ \hline
\rho & \rho & 1 & \rho_{22,12} & \rho & \rho \\
\rho & \rho & \rho_{22,21} & 1 & \rho & \rho \\ \hline
\rho^2 & \rho^2 & \rho & \rho & 1 & \rho_{33,12} \\
\rho^2 & \rho^2 & \rho & \rho & \rho_{33,21} & 1
\end{array}\right) \qquad (55)
$$

$$\boldsymbol{R}_{un} = \left(\begin{array}{cc|cc|cc} 1 & \rho_{11,12} & \rho_{12} & \rho_{12} & \rho_{13} & \rho_{13} \\ \rho_{11,21} & 1 & \rho_{12} & \rho_{12} & \rho_{13} & \rho_{13} \\ \hline \rho_{12} & \rho_{12} & 1 & \rho_{22,12} & \rho_{23} & \rho_{23} \\ \rho_{12} & \rho_{12} & \rho_{22,21} & 1 & \rho_{23} & \rho_{23} \\ \hline \rho_{13} & \rho_{13} & \rho_{23} & \rho_{23} & 1 & \rho_{33,12} \\ \rho_{13} & \rho_{13} & \rho_{23} & \rho_{23} & \rho_{33,21} & 1 \end{array}\right) \tag{56}$$

We will study properties of these correlation matrices in the next few sections.

Properties of the Correlation matrix $\boldsymbol{R}_{cs}$

The correlation matrix (54) can be written in partitioned form compactly as

$$\boldsymbol{R}_{cs} = \left(\begin{array}{cc|c} \boldsymbol{R}_{11} & \rho\,\mathrm{J} & \rho\,\mathrm{J} \\ \rho\,\mathrm{J} & \boldsymbol{R}_{22} & \rho\,\mathrm{J} \\ \hline \rho\,\mathrm{J} & \rho\,\mathrm{J} & \boldsymbol{R}_{33} \end{array}\right) = \left(\begin{array}{cc} \boldsymbol{T}_1 & \boldsymbol{T}_2 \\ \boldsymbol{T}_2' & \boldsymbol{T}_3 \end{array}\right), \tag{57}$$

where J is a matrix of ones. The following theorem provides the range of ρ for which the determinant of the matrix $\boldsymbol{R}_{cs}$ is positive.

THEOREM 1. *Consider the correlation matrix $\boldsymbol{R}_{cs}$ defined in (57). The determinant of $\boldsymbol{R}_{cs}$ is positive if and only if the cubic polynomial*

$$1 - A\rho^2 + B\rho^3 > 0,$$

where

$$\begin{aligned} A &= (\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1}) + (\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1}) + (\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1}), \\ B &= 2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1}). \end{aligned}$$

Proof. From the formula of determinant for partitioned matrix we have

$$\begin{aligned}
|\boldsymbol{R}_{cs}| &= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3 - \boldsymbol{T}_2'\boldsymbol{T}_1^{-1}\boldsymbol{T}_2| \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,|\mathrm{I} - \boldsymbol{T}_2'\boldsymbol{T}_1^{-1}\boldsymbol{T}_2\boldsymbol{T}_3^{-1}| \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,\left|\mathrm{I} - \rho^2[\mathbf{J}:\mathbf{J}]\boldsymbol{T}_1^{-1}\begin{bmatrix}\mathbf{J}\\ \mathbf{J}\end{bmatrix}\boldsymbol{T}_3^{-1}\right| \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,\left|\mathrm{I} - \rho^2\mathbf{1}[\mathbf{1}':\mathbf{1}']\boldsymbol{T}_1^{-1}\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}\mathbf{1}'\boldsymbol{T}_3^{-1}\right| \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,\left[1 - \rho^2[\mathbf{1}':\mathbf{1}']\boldsymbol{T}_1^{-1}\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}\mathbf{1}'\boldsymbol{T}_3^{-1}\mathbf{1}\right]
\end{aligned}$$

because $|\mathrm{I} - \mathrm{AB}| = |\mathrm{I} - \mathrm{BA}|$ for any two matrices A and B. Since

$$\begin{aligned}
\boldsymbol{T}_1^{-1} &= \begin{pmatrix}\boldsymbol{T}^{11} & \boldsymbol{T}^{12}\\ \boldsymbol{T}^{21} & \boldsymbol{T}^{22}\end{pmatrix} \\
&= \begin{pmatrix}(\boldsymbol{R}_{11} - \rho^2\mathbf{J}\boldsymbol{R}_{22}^{-1}\mathbf{J})^{-1} & -\rho\boldsymbol{T}^{11}\mathbf{J}\boldsymbol{R}_{22}^{-1}\\ -\rho\boldsymbol{R}_{22}^{-1}\mathbf{J}\boldsymbol{T}^{11} & \boldsymbol{R}_{22}^{-1} + \rho^2\boldsymbol{R}_{22}^{-1}\mathbf{J}\boldsymbol{T}^{11}\mathbf{J}\boldsymbol{R}_{22}^{-1}\end{pmatrix},
\end{aligned}$$

we have

$$\begin{aligned}
|\boldsymbol{R}_{cs}| &= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,\left[1 - \rho^2[\mathbf{1}':\mathbf{1}']\begin{bmatrix}\boldsymbol{T}^{11} & \boldsymbol{T}^{12}\\ \boldsymbol{T}^{21} & \boldsymbol{T}^{22}\end{bmatrix}\begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}(\mathbf{1}'\boldsymbol{T}_3^{-1}\mathbf{1})\right] \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,\left[1 - \rho^2\{\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1} + \mathbf{1}'\boldsymbol{T}^{12}\mathbf{1} + \mathbf{1}'\boldsymbol{T}^{21}\mathbf{1} + \mathbf{1}'\boldsymbol{T}^{22}\mathbf{1}\}(\mathbf{1}'\boldsymbol{T}_3^{-1}\mathbf{1})\right] \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,[1 - \rho^2\{\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1} - (\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\rho - (\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1})\rho \\
&\quad + (\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1}) + \rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\}(\mathbf{1}'\boldsymbol{T}_3^{-1}\mathbf{1})] \\
&= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,\left[1 - \rho^2(\mathbf{1}'\boldsymbol{T}_3^{-1}\mathbf{1})\{\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1}[1 - \rho(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})]^2 + \mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1}\}\right]
\end{aligned} \tag{58}$$

where $\mathbf{1}$ is a column vector of ones. Now

$$\begin{aligned}\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1} &= \mathbf{1}'(\boldsymbol{R}_{11}-\rho^2\mathbf{J}\boldsymbol{R}_{22}^{-1}\mathbf{J})^{-1}\mathbf{1}\\ &= \mathbf{1}'(\boldsymbol{R}_{11}-\rho^2\mathbf{1}(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\mathbf{1}')^{-1}\mathbf{1}\\ &= \mathbf{1}'(\boldsymbol{R}_{11}-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\mathbf{J})^{-1}\mathbf{1}.\end{aligned}$$

To simplify, $\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1}$, we use the lemma given in Kenneth (1981) which states that if $\mathbf{H}$ has rank one, then

$$(\mathbf{G}+\mathbf{H})^{-1} \;=\; \mathbf{G}^{-1}-\frac{1}{1+g}\mathbf{G}^{-1}\mathbf{H}\mathbf{G}^{-1} \text{ where } g=\text{trace}(\mathbf{H}\mathbf{G}^{-1}).$$

In the present proof, we let $\mathbf{G}=\boldsymbol{R}_{11}$ and $\mathbf{H}=-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\mathbf{J}$. We can easily check that $\text{rank}(\mathbf{H})=1$ and

$$\begin{aligned}g=\text{trace}(\mathbf{H}\mathbf{G}^{-1}) &= \text{trace}(-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\mathbf{J}\boldsymbol{R}_{11}^{-1})\\ &= \text{trace}(-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})\\ &= -\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1}).\end{aligned}$$

We also have

$$\begin{aligned}\mathbf{1}'\mathbf{G}^{-1}\mathbf{H}\mathbf{G}^{-1}\mathbf{1} &= \mathbf{1}'\boldsymbol{R}_{11}^{-1}[-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})\mathbf{J}]\boldsymbol{R}_{11}^{-1}\mathbf{1}\\ &= -\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})^2.\end{aligned}$$

Using the above results we get

$$\begin{aligned}\mathbf{1}'(\mathbf{G}+\mathbf{H})^{-1}\mathbf{1} &= \mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1}-\frac{1}{1-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})}\left[-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})^2\right]\\ &= \frac{\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1}}{1-\rho^2(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})}.\end{aligned}$$

Therefore from equation (58) we have

$$\begin{aligned}|\boldsymbol{R}_{cs}| &= |\boldsymbol{T}_1|\,|\boldsymbol{T}_3|\,[1-\rho^2(\mathbf{1}'\boldsymbol{T}_3^{-1}\mathbf{1})\{\mathbf{1}'\boldsymbol{T}^{11}\mathbf{1}[1-\rho(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})]^2+\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1}\}]\\ &= |\boldsymbol{T}_1|\,|\boldsymbol{R}_{33}|\\ &\qquad\left[1-\rho^2(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})\left\{\frac{\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1}+\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1}-2\rho(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})}{1-\rho^2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})}\right\}\right]\end{aligned}$$

$$= |\boldsymbol{T}_1|\,|\boldsymbol{R}_{33}| \left[\frac{1-\rho^2[(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})+(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})+(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})]+2\rho^3(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})}{1-\rho^2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})}\right]$$

Using the fact that $|\boldsymbol{T}_1| = |\boldsymbol{R}_{11}||\boldsymbol{R}_{22}|[1-\rho^2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})]$ we get

$$\begin{aligned} |\boldsymbol{R}_{cs}| &= |\boldsymbol{R}_{11}|\,|\boldsymbol{R}_{22}|\,|\boldsymbol{R}_{33}| \\ &\quad \big(1-\rho^2[(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})+(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})+(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})] \\ &\quad +2\rho^3(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})\big). \end{aligned}$$

Thus $|\boldsymbol{R}_{cs}| > 0$ if and only if $1 - A\rho^2 + B\rho^3 > 0$ where

$$\begin{aligned} A &= (\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})+(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1})+(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1}) \text{ and} \\ B &= 2(\mathbf{1}'\boldsymbol{R}_{11}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{22}^{-1}\mathbf{1})(\mathbf{1}'\boldsymbol{R}_{33}^{-1}\mathbf{1}). \end{aligned}$$

This completes the proof of the theorem.

Remark: We observed numerically the cubic equation in ρ in Theorem 1. has one root ($\rho_1 < 0$) that is negative always and two positive roots ($0 < \rho_2 < \rho_3$). Further the cubic equation is nonnegative if and only if $\rho_1 < \rho < \rho_2$. This range (ρ_1, ρ_2) for ρ is also the range where the correlation matrix $\boldsymbol{R}_{cs}$ is positive semi-definite. We verified this result numerically but were unable to prove this analytically.

IV.3.3 EXISTENCE OF A JOINT DISTRIBUTION WITH CS STRUCTURE

In this section, we derive the range for the ρ in the correlation structure (54) such that a joint distribution exists for three multinomial random vectors $\boldsymbol{Y}_1$, $\boldsymbol{Y}_2$ and $\boldsymbol{Y}_3$ exists with correlation structure $\boldsymbol{R}_{cs}$. The main theorem in this section is the following.

THEOREM 2. *Suppose $\boldsymbol{Y}_1, \boldsymbol{Y}_2$ and $\boldsymbol{Y}_3$ are three multinomial random vectors with one trial and probabilities $\boldsymbol{p}_1 = (p_{1,1}, p_{1,2}, \cdots, p_{1,K-1})'$, $\boldsymbol{p}_2 = (p_{2,1}, p_{2,2}, \cdots, p_{2,K-1})'$, and $\boldsymbol{p}_3 = (p_{3,1}, p_{3,2}, \cdots, p_{3,K-1})'$ respectively. Assume that the correlation matrix of the three multinomial vectors is given by $\boldsymbol{R}_{cs}$ defined in (54). Then the joint distribution for $\boldsymbol{Y}_1, \boldsymbol{Y}_2$, and $\boldsymbol{Y}_3$ exists if and only if the parameter ρ satisfies the* *inequalities*

$$\max\{L_{31}(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3), L_{32}(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3), L_{33}(\boldsymbol{p}_i,\boldsymbol{p}_j),\ 1\le i<j\le\ 3\}\ \le\ \rho\ \le \\ \min\{\min\{U_{31}(\boldsymbol{p}_i,\boldsymbol{p}_j),\ 1\le i<j\le\ 3\}, U_{32}(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3)\}$$

where

$$L_{31}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \max\left\{\frac{-(\sum_{c_i=1}^{K-1}p_{i,c_i})(\sum_{c_j=1}^{K-1}p_{j,c_j})}{(\sum_{c_i=1}^{K-1}\sigma_{i,c_i})(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}, \frac{-p_{i,K}p_{j,K}}{(\sum_{c_i=1}^{K-1}\sigma_{i,c_i})(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}\right\},$$

$$L_{32}(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3) = \max\left\{\frac{-p_{i,c_i}(\sum_{c_j=1}^{K-1}p_{j,c_j})}{\sigma_{i,c_i}(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}; i\neq j\in\{1,2,3\}, c_i\in\{1,...,K-1\}\right\},$$

$$L_{33}(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3) = \frac{-\{(\sum_{i=1}^{K-1}p_{1,i})(\sum_{i=1}^{K-1}p_{2,i})(\sum_{i=1}^{K-1}p_{3,i})+(p_{1,K}p_{2,K}p_{3,K})\}}{(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{2,i})+(\sum_{i=1}^{K-1}\sigma_{1,i})(\sum_{i=1}^{K-1}\sigma_{3,i})+(\sum_{i=1}^{K-1}\sigma_{2,i})(\sum_{i=1}^{K-1}\sigma_{3,i})},$$

and

$$U_{31}(\boldsymbol{p}_i,\boldsymbol{p}_j) = \min\left\{\frac{(\sum_{c_i=1}^{K-1}p_{i,c_i})p_{j,K}}{(\sum_{c_i=1}^{K-1}\sigma_{i,c_i})(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}, \frac{(\sum_{c_j=1}^{K-1}p_{j,c_j})p_{i,K}}{(\sum_{c_i=1}^{K-1}\sigma_{i,c_i})(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}\right\},$$

$$U_{32}(\boldsymbol{p}_1,\boldsymbol{p}_2,\boldsymbol{p}_3) = \min\left\{\frac{p_{i,c_i}p_{j,K}}{\sigma_{i,c_i}(\sum_{c_j=1}^{K-1}\sigma_{j,c_j})}; i\neq j\in\{1,2,3\}, c_i\in\{1,...,K-1\}\right\}.$$

Proof. The theorem can be proved imitating the proof of Theorem 2.. We omit the details since the notation is cumbersome.

IV.3.4 CONSTRUCTION OF TRIVARIATE MULTINOMIAL DISTRIBUTIONS

We have seen in bivariate case the joint probabilities are determined completely by the marginal probabilities and the correlations. However, in dimensions more than two, there can be many joint probability with specified marginals and correlations. For example, the Table 25 we give three joint probability mass functions (PMF) all having the same marginal means $\boldsymbol{p}_1$ = (0.2, 0.3, 0.5), $\boldsymbol{p}_2$= (0.4, 0.1, 0.5), and $\boldsymbol{p}_3$ = (0.3, 0.1, 0.6) and correlation value $\rho = 0.02$, which is within the feasible range (−0.1297, 0.1942).

Table 25: Trivariate joint probability mass functions

Y_1	Y_2	Y_3	PMF - 1	PMF - 2	PMF - 3
1	1	1	0.0062	0.0042	0.0056
1	1	2	0.0028	0.0059	0.0051
1	1	3	0.0750	0.0739	0.0732
1	2	1	0.0061	0.0066	0.0043
1	2	2	0.0054	0.0024	0.0038
1	2	3	0.0109	0.0134	0.0143
1	3	1	0.0514	0.0529	0.0537
1	3	2	0.0143	0.0141	0.0135
1	3	3	0.0280	0.0267	0.0264
2	1	1	0.0031	0.0021	0.0026
2	1	2	0.0022	0.0042	0.0052
2	1	3	0.1191	0.1183	0.1168
2	2	1	0.0048	0.0028	0.0065
2	2	2	0.0015	0.0020	0.0064
2	2	3	0.0265	0.0280	0.0198
2	3	1	0.0863	0.0894	0.0851
2	3	2	0.0290	0.0266	0.0212
2	3	3	0.0274	0.0268	0.0365
3	1	1	0.1152	0.1182	0.1163
3	1	2	0.0379	0.0329	0.0327
3	1	3	0.0385	0.0404	0.0426
3	2	1	0.0219	0.0234	0.0219
3	2	2	0.0049	0.0074	0.0016
3	2	3	0.0181	0.0141	0.0213
3	3	1	0.0051	0.0005	0.0039
3	3	2	0.0020	0.0046	0.0105
3	3	3	0.2565	0.2585	0.2491
			1.0000	1.0000	1.0000

Table 26 and Table 27 contain positive definite ranges for ρ given in Theorem 1. and feasible ranges given in Theorem 2. for numerous marginal distributions. Note that there are many joint distributions even for a specified ρ within the feasible range.

Table 26: Positive definite ranges and bounds on correlation ρ of the matrix $\boldsymbol{R}_{cs}$ for different marginal probabilities

Marginal Probabilities*			Bounds on ρ		Positive Definite Range Roots†		
$\boldsymbol{p}_1$	$\boldsymbol{p}_2$	$\boldsymbol{p}_3$	Lower	Upper	ρ_1	ρ_2	ρ_3
(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.1,0.1,0.8)	-0.111	0.444	-0.222	0.444	0.444
(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.1,0.4,0.5)	-0.111	0.211	-0.208	0.390	0.444
(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.1,0.7,0.2)	-0.111	0.088	-0.180	0.303	0.444
(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.4,0.1,0.5)	-0.111	0.211	-0.208	0.390	0.444
(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.4,0.4,0.2)	-0.104	0.068	-0.155	0.239	0.444
(0.1,0.1,0.8)	(0.1,0.1,0.8)	(0.7,0.1,0.2)	-0.111	0.088	-0.180	0.303	0.444
(0.1,0.1,0.8)	(0.1,0.4,0.5)	(0.1,0.4,0.5)	-0.111	0.211	-0.194	0.364	0.416
(0.1,0.1,0.8)	(0.1,0.4,0.5)	(0.1,0.7,0.2)	-0.105	0.088	-0.169	0.289	0.407
(0.1,0.1,0.8)	(0.1,0.4,0.5)	(0.4,0.1,0.5)	-0.111	0.211	-0.194	0.364	0.416
(0.1,0.1,0.8)	(0.1,0.4,0.5)	(0.4,0.4,0.2)	-0.087	0.068	-0.146	0.228	0.406
(0.1,0.1,0.8)	(0.1,0.4,0.5)	(0.7,0.1,0.2)	-0.105	0.088	-0.169	0.289	0.407
(0.1,0.1,0.8)	(0.1,0.7,0.2)	(0.1,0.7,0.2)	-0.070	0.088	-0.148	0.245	0.370
(0.1,0.1,0.8)	(0.1,0.7,0.2)	(0.4,0.1,0.5)	-0.105	0.088	-0.169	0.289	0.407
(0.1,0.1,0.8)	(0.1,0.7,0.2)	(0.4,0.4,0.2)	-0.054	0.068	-0.128	0.199	0.357
(0.1,0.1,0.8)	(0.1,0.7,0.2)	(0.7,0.1,0.2)	-0.070	0.088	-0.148	0.245	0.370
(0.1,0.1,0.8)	(0.4,0.1,0.5)	(0.4,0.1,0.5)	-0.111	0.211	-0.194	0.364	0.416
(0.1,0.1,0.8)	(0.4,0.1,0.5)	(0.4,0.4,0.2)	-0.087	0.068	-0.146	0.228	0.406
(0.1,0.1,0.8)	(0.4,0.1,0.5)	(0.7,0.1,0.2)	-0.105	0.088	-0.169	0.289	0.407
(0.1,0.1,0.8)	(0.4,0.4,0.2)	(0.4,0.4,0.2)	-0.042	0.068	-0.111	0.167	0.333
(0.1,0.1,0.8)	(0.4,0.4,0.2)	(0.7,0.1,0.2)	-0.054	0.068	-0.128	0.199	0.357
(0.1,0.1,0.8)	(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.070	0.088	-0.148	0.245	0.370
(0.1,0.4,0.5)	(0.1,0.4,0.5)	(0.1,0.4,0.5)	-0.134	0.345	-0.182	0.364	0.364
(0.1,0.4,0.5)	(0.1,0.4,0.5)	(0.1,0.7,0.2)	-0.137	0.225	-0.159	0.281	0.364
(0.1,0.4,0.5)	(0.1,0.4,0.5)	(0.4,0.1,0.5)	-0.134	0.345	-0.182	0.364	0.364
(0.1,0.4,0.5)	(0.1,0.4,0.5)	(0.4,0.4,0.2)	-0.115	0.183	-0.137	0.221	0.364
(0.1,0.4,0.5)	(0.1,0.4,0.5)	(0.7,0.1,0.2)	-0.137	0.225	-0.159	0.281	0.364
(0.1,0.4,0.5)	(0.1,0.7,0.2)	(0.1,0.7,0.2)	-0.070	0.157	-0.139	0.245	0.321
(0.1,0.4,0.5)	(0.1,0.7,0.2)	(0.4,0.1,0.5)	-0.137	0.225	-0.159	0.281	0.364
(0.1,0.4,0.5)	(0.1,0.7,0.2)	(0.4,0.4,0.2)	-0.054	0.122	-0.121	0.198	0.311
(0.1,0.4,0.5)	(0.1,0.7,0.2)	(0.7,0.1,0.2)	-0.070	0.157	-0.139	0.245	0.321
(0.1,0.4,0.5)	(0.4,0.1,0.5)	(0.4,0.1,0.5)	-0.134	0.345	-0.182	0.364	0.364
(0.1,0.4,0.5)	(0.4,0.1,0.5)	(0.4,0.4,0.2)	-0.115	0.183	-0.137	0.221	0.364
(0.1,0.4,0.5)	(0.4,0.1,0.5)	(0.7,0.1,0.2)	-0.137	0.225	-0.159	0.281	0.364
(0.1,0.4,0.5)	(0.4,0.4,0.2)	(0.4,0.4,0.2)	-0.042	0.110	-0.105	0.167	0.287
(0.1,0.4,0.5)	(0.4,0.4,0.2)	(0.7,0.1,0.2)	-0.054	0.122	-0.121	0.198	0.311
(0.1,0.4,0.5)	(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.070	0.157	-0.139	0.245	0.321
(0.1,0.7,0.2)	(0.1,0.7,0.2)	(0.1,0.7,0.2)	-0.070	0.157	-0.123	0.245	0.245
(0.1,0.7,0.2)	(0.1,0.7,0.2)	(0.4,0.1,0.5)	-0.070	0.157	-0.139	0.245	0.321

* $\boldsymbol{p}_i = (p_{i,1}, p_{i,2}, p_{i,3})$ for $i = 1, 2, 3$.

† The range of ρ for the correlation matrix $\boldsymbol{R}_{cs}$ is positive semi-definite is $\rho_1 < \rho < \rho_2$.

Table 27: Positive definite ranges and bounds on correlation ρ of the matrix $\boldsymbol{R}_{cs}$ for different marginal probabilities (Continued.)

Marginal Probabilities*			Bounds on ρ		Positive Definite Range Roots†		
$\boldsymbol{p}_1$	$\boldsymbol{p}_2$	$\boldsymbol{p}_3$	Lower	Upper	ρ_1	ρ_2	ρ_3
(0.1,0.7,0.2)	(0.1,0.7,0.2)	(0.4,0.4,0.2)	-0.054	0.122	-0.107	0.191	0.245
(0.1,0.7,0.2)	(0.1,0.7,0.2)	(0.7,0.1,0.2)	-0.070	0.157	-0.123	0.245	0.245
(0.1,0.7,0.2)	(0.4,0.1,0.5)	(0.4,0.1,0.5)	-0.137	0.225	-0.159	0.281	0.364
(0.1,0.7,0.2)	(0.4,0.1,0.5)	(0.4,0.4,0.2)	-0.054	0.122	-0.121	0.198	0.311
(0.1,0.7,0.2)	(0.4,0.1,0.5)	(0.7,0.1,0.2)	-0.070	0.157	-0.139	0.245	0.321
(0.1,0.7,0.2)	(0.4,0.4,0.2)	(0.4,0.4,0.2)	-0.042	0.110	-0.094	0.167	0.217
(0.1,0.7,0.2)	(0.4,0.4,0.2)	(0.7,0.1,0.2)	-0.054	0.122	-0.107	0.191	0.245
(0.1,0.7,0.2)	(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.070	0.157	-0.123	0.245	0.245
(0.4,0.1,0.5)	(0.4,0.1,0.5)	(0.4,0.1,0.5)	-0.134	0.345	-0.182	0.364	0.364
(0.4,0.1,0.5)	(0.4,0.1,0.5)	(0.4,0.4,0.2)	-0.115	0.183	-0.137	0.221	0.364
(0.4,0.1,0.5)	(0.4,0.1,0.5)	(0.7,0.1,0.2)	-0.137	0.225	-0.159	0.281	0.364
(0.4,0.1,0.5)	(0.4,0.4,0.2)	(0.4,0.4,0.2)	-0.042	0.110	-0.105	0.167	0.287
(0.4,0.1,0.5)	(0.4,0.4,0.2)	(0.7,0.1,0.2)	-0.054	0.122	-0.121	0.198	0.311
(0.4,0.1,0.5)	(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.070	0.157	-0.139	0.245	0.321
(0.4,0.4,0.2)	(0.4,0.4,0.2)	(0.4,0.4,0.2)	-0.042	0.208	-0.083	0.167	0.167
(0.4,0.4,0.2)	(0.4,0.4,0.2)	(0.7,0.1,0.2)	-0.042	0.110	-0.094	0.167	0.217
(0.4,0.4,0.2)	(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.054	0.122	-0.107	0.191	0.245
(0.7,0.1,0.2)	(0.7,0.1,0.2)	(0.7,0.1,0.2)	-0.070	0.157	-0.123	0.245	0.245

* $\boldsymbol{p}_i = (p_{i,1}, p_{i,2}, p_{i,3})$ for $i = 1, 2, 3$.

† The range of ρ for the correlation matrix $\boldsymbol{R}_{cs}$ is positive semi-definite is $\rho_1 < \rho < \rho_2$.

CHAPTER V

SUMMARY

For dependent Gaussian random variables with correlation matrix $\boldsymbol{R}$, it suffices that $\boldsymbol{R}$ be positive definite. This is not the case, however, for discrete random variables. There are additional restrictions on the correlation matrix $\boldsymbol{R}$ to guarantee a joint distribution for dependent discrete random variables. In Chapter IV we discussed these additional restrictions for dependent multinomial vectors. These complexity of these restrictions increases as the dimension increases. However, understanding these restrictions is necessary when constructing the likelihood for dependent multinomial random vectors, even though specification of these restrictions is nearly impossible for dimensions greater than 3. The GEE methodology that we discussed in Chapter III estimates the correlation parameters ignoring these additional restrictions, and thus the methodology provides estimates of the regression parameter which may lack a probabilistic basis. An alternative and promising solution which bypasses these difficulties is the use of latent variables or more generally copula models. These models make it possible to construct proper likelihoods for dependent multinomial random vectors. We have discussed these likelihoods and maximum likelihood estimation in Chapter II of this dissertation.

REFERENCES

Agresti, A. (2002). *Categorical Data Analysis, Second Edition.* John Wiley & Sons Inc., New Jersy.

Agresti, A. and Natarajan, R. (2001). Modeling clustered ordered categorical data: A Survey. *International Statistical Review* **69**, 345-371.

Ashford, J. R. and Sowden, R. R. (1970). Multi-variate probit analysis. *Biometrics* **26**, 535-546.

Chaganty, N. R. and Joe, H. (2004). Efficiency of generalized estimating equations for binary responses. *Journal of Royal Statistics Society, B* **66**, 851-860.

Chaganty, N. R. and Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika* **93**, 197-206.

Davis, C. S. (2002). *Statistical methods for the analysis of repeated measurements.* Springer series in statistics. Springer, New York.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics* **1**, 141-150.

Genz, A. (1993). Comparison of methods for the computation of multivariate normal probabilities. *Computing Science and Statistics* **25**, 400-405.

Jaworski, P., Durante, F., Härdle, W., and Rychlik, T. (2010). *Copula Theory and its Applications.* Springer-Verlag Berlin Heidelberg.

Joe, H. (1995). Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of American Statistical Association* **90**, 957–964.

Joe, H. (1997). *Multivariate Models and Dependence Concepts.* Monographs on Statistics and Applied Probability, vol.73. Chapman & Hall, London.

Kenneth, S. M. (1981). On the inverse of the sum of matrices. *Mathematics Magazine* **54**, 67-72.

Koch, G. G., Carr, G. J., Amara, I. A., Stokes, M. E., and Uryniak, T. J. (1989). Categorical data analysis. *Statistical Methodology in the Pharmaceutical Sciences*, D. A. Berry(ed.), 391-475.

Kotz, S., Balakrishnan, N., and Johnson, N. L. (2000). *Continuous Multivariate Distributions. Volume 1: Models and Applications.* Second Edition. John Wiley & Sons, Inc.

Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics* **38**, 963-974.

Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.

Lipsitz, S. R., Kim, K., and Zhao, L. (1994). Analysis of repeated categorical data using generalized estimating equations. *Statistics in Medicine* **13**, 1149-1163.

Lumley, T. (1996). Generalized Estimating Equations for Ordinal Data: A Note on Working Correlation Structures. *Biometrics* **52**, 354-361.

Meester, G. S. and MacKay, J. (1994). A Parametric Model for Cluster Correlated Categorical Data. *Biometrics* **50**, 954-963.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models, Second Edition.* Chapman and Hall, London.

Miller, M. E., Davis, C. S., and Richard Landis, J. (1993). The Analysis of Longitudinal Polytomous Data: Generalized Estimating Equations and Connections with Weighted Least Squares. *Biometrics* **49**, 1033-1044.

Nash, J. C. (1979). *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation, Second Edition.* Hilger, New York.

Nelson, R. B. (2006). *An Introduction to Copulas, Second Edition.* Springer series in Statistics. Springer, New York.

Prentice, R. L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics* **44**, 1033-1048.

Sabo, R. T. and Chaganty, N. R. (2010). What can go wrong when ignoring correlation bounds in the use of generalized estimating equations. *Statistics in Medicine* **29**, 2501–2507.

Stanish, W. M., Gillings, D. B., and Koch, G. G. (1978). An Application of Multivariate Ratio Methods for the Analysis of a Longitudinal Clinical Trial with Missing Data. *Biometrics* **34**, 305-317.

Ware, J. H. (1985). Linear Models for the Analysis of Longitudinal Studies. *The American Statistician* **39**, 95-101.

Ware, J. H., Dockery, D. W., Spiro, A. III, Speizer, F. E., and Ferris, B. G. Jr. (1984). Passive smoking, gas cooking and respiratory health of children living in six cities. *American Review of Respiratory Diseases* **129**, 366–374.

APPENDIX A

ORDERED PROBIT MODEL DERIVATIVES

Here we obtain the derivatives for the multivariate ordered probit model that we discussed in Section II.2 and Section II.4.

A.1 DERIVATIVES WITH RESPECT TO REGRESSION PARAMETERS

The following notation is needed for the derivatives of the multivariate ordered probit model (5) with respect to the regression parameter. For a vector $\boldsymbol{a} = (a_1, a_2, ..., a_t)'$, we denote the vector deleting the lth component by $\boldsymbol{a}_{-l} = (a_1, a_2, ..., a_{l-1}, a_{l+1}, ..., a_t)'$. For a correlation matrix

$$\boldsymbol{R} = \begin{pmatrix} 1 & r_{12} & r_{13} & \cdots & r_{1t} \\ r_{21} & 1 & r_{23} & \cdots & r_{2t} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{t1} & r_{t2} & r_{t3} & \cdots & 1 \end{pmatrix},$$

we denote the matrix obtained by permutating the lth column (and row) with the tth column (and row) by

$$\boldsymbol{R}^{(l)} = \left(\begin{array}{ccccccc|c} 1 & r_{12} & \cdots & r_{1(l-1)} & r_{1(l+1)} & \cdots & r_{1t} & r_{1l} \\ r_{21} & 1 & \cdots & r_{2(l-1)} & r_{2(l+1)} & \cdots & r_{2t} & r_{2l} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{(l-1)1} & r_{(l-1)2} & \cdots & 1 & r_{(l-1)(l+1)} & \cdots & r_{(l-1)t} & r_{(l-1)l} \\ r_{(l+1)1} & r_{(l+1)2} & \cdots & r_{(l+1)(l-1)} & 1 & \cdots & r_{(l+1)t} & r_{(l+1)l} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots \\ r_{t1} & r_{t2} & \cdots & r_{t(l-1)} & r_{t(l+1)} & \cdots & 1 & r_{tl} \\ \hline r_{l1} & r_{l2} & \cdots & r_{l(l-1)} & r_{l(l+1)} & \cdots & r_{lt} & 1 \end{array}\right)$$

$$= \begin{pmatrix} \boldsymbol{R}_{11}^{(l)} & \boldsymbol{R}_{12}^{(l)} \\ \boldsymbol{R}_{21}^{(l)} & \boldsymbol{R}_{22}^{(l)} \end{pmatrix}.$$

Define

$$\begin{aligned}
\boldsymbol{R}_{-l/l} &= \boldsymbol{R}_{11}^{(l)} - \boldsymbol{R}_{12}^{(l)}(\boldsymbol{R}_{22}^{(l)})^{-1}\boldsymbol{R}_{21}^{(l)} = \boldsymbol{R}_{11}^{(l)} - \boldsymbol{R}_{12}^{(l)}\boldsymbol{R}_{21}^{(l)}, \\
\boldsymbol{\mu}_{-l/l} &= \boldsymbol{R}_{12}^{(l)}(\boldsymbol{R}_{22}^{(l)})^{-1}a_l = \boldsymbol{R}_{12}^{(l)}a_l.
\end{aligned}$$

The following lemma provides the derivative of a multivariate normal cumulative distribution function.

LEMMA 1. *Suppose* $\Phi_t(a_1, a_2, ..., a_t; \mathbf{0}, \boldsymbol{R})$ *denotes a t-variate normal distribution function with mean* $\mathbf{0}$ *and correlation matrix* $\boldsymbol{R}$. *Then the derivative of* Φ_t *with respect to* a_l *is*

$$\frac{\partial}{\partial a_l}\Phi_t(a_1, a_2, ..., a_t; \mathbf{0}, \boldsymbol{R}) = \Phi_{t-1}(\boldsymbol{a}_{-l}; \boldsymbol{\mu}_{-l/l}, \boldsymbol{R}_{-l/l})\ \phi(a_l; 0, 1).$$

Proof. Let $P = \Phi_t(a_1, a_2, ..., a_t; \mathbf{0}, \boldsymbol{R})$. Then

$$\begin{aligned}
\frac{\partial P}{\partial a_l} &= \frac{\partial}{\partial a_l}\int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_t}\phi_t(\boldsymbol{z}; \mathbf{0}, \boldsymbol{R})\ d\boldsymbol{z} \\
&= \frac{\partial}{\partial a_l}\int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_t}\phi_{t-1}(\boldsymbol{z}_{-l}; \boldsymbol{R}_{12}^{(l)}z_l, \boldsymbol{R}_{-l/l})\ \phi(z_l; 0, 1)\ d\boldsymbol{z} \\
&= \frac{\partial}{\partial a_l}\left(\int_{-\infty}^{a_l}\Phi_{t-1}(\boldsymbol{a}_{-l}; \boldsymbol{R}_{12}^{(l)}z_l, \boldsymbol{R}_{-l/l})\ \phi(z_l; 0, 1)\ dz_l\right) \\
&= \Phi_{t-1}(\boldsymbol{a}_{-l}; \boldsymbol{R}_{12}^{(l)}a_l, \boldsymbol{R}_{-l/l})\ \phi(a_l; 0, 1) \\
&= \Phi_{t-1}(\boldsymbol{a}_{-l}; \boldsymbol{\mu}_{-l/l}, \boldsymbol{R}_{-l/l})\ \phi(a_l; 0, 1) \qquad \square
\end{aligned}$$

The probability mass function $\pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R})$ of the multivariate ordered probit model defined in equation (5) can be written as 2^t differences of t-variate normal cumulative distribution function as

$$\pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R}) = \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_t=1}^{2}(-1)^{i_1+i_2+..+i_t}\Phi_t(b_{1i_1}, b_{2i_2}, ..., b_{ti_t}; \mathbf{0}, \boldsymbol{R}) \qquad (59)$$

with $b_{j1} = \gamma_j(y_j - 1)$; $b_{j2} = \gamma_j(y_j)$. Recall that the ordered thresholds $\gamma_j(k) = \alpha_k + \boldsymbol{x}_j'\boldsymbol{\beta}$, where $\boldsymbol{\beta}$ is the regression parameter.

Theorem 1. *The derivative of $\pi_t(\boldsymbol{y};\mathbf{0},\boldsymbol{R})$ with respect β_l, the lth component of $\boldsymbol{\beta}$ is given by*

$$\frac{\partial}{\partial\beta_l}\pi_t(\boldsymbol{y};\mathbf{0},\boldsymbol{R}) = \sum_{j=1}^{t} x_{jl}\left\{\sum_{i_j=1}^{2}(-1)^{i_j}\phi(b_{ji_j};0,1)\left[\pi_{t-1}(\boldsymbol{y}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})\right]\right\}$$

where $b_{j1} = \gamma_j(y_j - 1)$ and $b_{j2} = \gamma_j(y_j)$.

Proof. Note that from equation (59) we have

$$\pi_t(\boldsymbol{y};\mathbf{0},\boldsymbol{R}) \;=\; \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}...\sum_{i_t=1}^{2}(-1)^{i_1+i_2+\quad+i_t}\Phi_t(b_{1i_1},b_{2i_2},...,b_{ti_t};\mathbf{0},\boldsymbol{R}),$$

where b_{ji_j}'s are functions of the regression parameter $\boldsymbol{\beta}$. For notational convenience we write π_t for $\pi_t(\boldsymbol{y};\mathbf{0},\boldsymbol{R})$. Then the derivative of π_t with respect to β_l is

$$\begin{aligned}\frac{\partial\pi_t}{\partial\beta_l} &= \frac{\partial\pi_t}{\partial b_{1i_1}}.\frac{\partial b_{1i_1}}{\partial\beta_l} + \frac{\partial\pi_t}{\partial b_{2i_2}}.\frac{\partial b_{2i_2}}{\partial\beta_l} + \;...\; + \frac{\partial\pi_t}{\partial b_{ti_t}}.\frac{\partial b_{ti_t}}{\partial\beta_l}\\ &= \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}...\sum_{i_t=1}^{2}\left(\frac{\partial}{\partial b_{1i_1}}\Phi_t(\boldsymbol{b};\mathbf{0},\boldsymbol{R}).\frac{\partial b_{1i_1}}{\partial\beta_l} + \frac{\partial}{\partial b_{2i_2}}\Phi_t(\boldsymbol{b};\mathbf{0},\boldsymbol{R}).\frac{\partial b_{2i_2}}{\partial\beta_l}\right.\\ &\qquad\left. + \cdots + \frac{\partial}{\partial b_{ti_t}}\Phi_t(\boldsymbol{b};\mathbf{0},\boldsymbol{R}).\frac{\partial b_{ti_t}}{\partial\beta_l}\right)(-1)^{i_1+i_2+\quad+i_t},\end{aligned}$$

where $\boldsymbol{b} = (b_{1i_1}, b_{2i_2}, ..., b_{ti_t})'$. By Lemma 1 we have

$$\begin{aligned}\frac{\partial\pi_t}{\partial\beta_l} &= \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}...\sum_{i_t=1}^{2}\left(\Phi_{t-1}(\boldsymbol{b}_{-1};\boldsymbol{R}_{12}^{(1)}b_{1i_1},\boldsymbol{R}_{-1/1})\;\phi(b_{1i_1};0,1)\frac{\partial b_{1i_1}}{\partial\beta_l}\right.\\ &\qquad\left. + \cdots + \Phi_{t-1}(\boldsymbol{b}_{-t};\boldsymbol{R}_{12}^{(t)}b_{ti_t},\boldsymbol{R}_{-t/t})\;\phi(b_{ti_t};0,1)\frac{\partial b_{ti_t}}{\partial\beta_l}\right)(-1)^{i_1+i_2+\quad+i_t}\end{aligned}$$

which implies

$$\begin{aligned}\frac{\partial\pi_t}{\partial\beta_l} &= \sum_{j=1}^{t}\sum_{i_j=1}^{2}\frac{\partial b_{ji_j}}{\partial\beta_l}(-1)^{i_j}\phi(b_{ji_j};0,1)\times\\ &\quad\sum_{i_1=1}^{2}...\sum_{i_{j-1}=1}^{2}\sum_{i_{j+1}=1}^{2}...\sum_{i_t=1}^{2}\Phi_{t-1}(\boldsymbol{b}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})(-1)^{i_1+\quad+i_{j-1}+i_{j+1}+\quad+i_t}\\ &= \sum_{j=1}^{t}\sum_{i_j=1}^{2}\frac{\partial b_{ji_j}}{\partial\beta_l}(-1)^{i_j}\phi(b_{ji_j};0,1)\times\left[\pi_{t-1}(\boldsymbol{y}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})\right]. \qquad (60)\end{aligned}$$

Since $\partial b_{ri_r}/\partial\beta_l = x_{rl}$ for all $i_r = 1, 2$, we get

$$\frac{\partial}{\partial\beta_l}\pi_t(\boldsymbol{y};\mathbf{0},\boldsymbol{R}) = \sum_{j=1}^{t} x_{jl}\left\{\sum_{i_j=1}^{2}(-1)^{i_j}\phi(b_{ji_j};0,1)\left[\pi_{t-1}(\boldsymbol{y}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})\right]\right\}.$$

This completes the proof. □

Corollary. The derivative of $\pi_t(\boldsymbol{y};\mathbf{0},\boldsymbol{R})$ with respect to α_l is

$$\frac{\partial\pi_t}{\partial\alpha_l} = \sum_{j=1}^{t}\sum_{i_j=1}^{2}\frac{\partial b_{ji_j}}{\partial\alpha_l}(-1)^{i_j}\phi(b_{ji_j};0,1)\left[\pi_{t-1}(\boldsymbol{y}_{-j};\boldsymbol{R}_{12}^{(j)}b_{ji_j},\boldsymbol{R}_{-j/j})\right],$$

where

$$\frac{\partial b_{ri_r}}{\partial\alpha_l} = \begin{cases} 1 & \text{if } (l = y_r - 1 \;\&\; i_r = 1) \text{ or } (l = y_r \;\&\; i_r = 2) \\ 0 & \text{otherwise.} \end{cases}$$

This follows replacing β_l by α_l in equation (60).

A.2 DERIVATIVES WITH RESPECT TO LATENT CORRELATIONS

We need some additional notation to express the derivatives of the multivariate ordered probit model (5) with respect to the latent correlations. For a vector $\boldsymbol{a} = (a_1, a_2, ..., a_t)'$, we denote the vector obtained deleting the l and sth components as $\boldsymbol{a}_{-(ls)} = (a_1, a_2, \cdots, a_{l-1}, a_{l+1}, \cdots, a_{s-1}, a_{s+1}, \cdots, a_t)'$ and $\boldsymbol{a}_{(ls)} = (a_l, a_s)'$. Assuming $l < s$, the matrix obtained permuting the lth and sth columns (and rows) with the $(t-1)$th and tth columns (and rows) of the correlation matrix $\boldsymbol{R}$, is denoted by

$$\boldsymbol{R}^{(ls)} = \left(\begin{array}{cccc|cc} 1 & r_{12} & \cdots & r_{1t} & r_{1l} & r_{1s} \\ r_{21} & 1 & \cdots & r_{2t} & r_{2l} & r_{2s} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ r_{t1} & r_{t2} & \cdots & 1 & r_{tl} & r_{ts} \\ \hline r_{l1} & r_{l2} & \cdots & r_{lt} & 1 & r_{ls} \\ r_{s1} & r_{s2} & \cdots & r_{st} & r_{sl} & 1 \end{array}\right) = \begin{pmatrix} \boldsymbol{R}_{11}^{(ls)} & \boldsymbol{R}_{12}^{(ls)} \\ \boldsymbol{R}_{21}^{(ls)} & \boldsymbol{R}_{22}^{(ls)} \end{pmatrix}.$$

Let

$$\boldsymbol{R}_{-(ls)/(ls)} = \boldsymbol{R}_{11}^{(ls)} - \boldsymbol{R}_{12}^{(ls)}\left(\boldsymbol{R}_{22}^{(ls)}\right)^{-1}\boldsymbol{R}_{21}^{(ls)}; \qquad \boldsymbol{\mu}_{-(ls)/(ls)} = \boldsymbol{R}_{12}^{(ls)}\left(\boldsymbol{R}_{22}^{(ls)}\right)^{-1}\boldsymbol{a}_{(ls)}$$

$$\boldsymbol{R}_{(ls)/-(ls)} = \boldsymbol{R}_{22}^{(ls)} - \boldsymbol{R}_{21}^{(ls)}\left(\boldsymbol{R}_{11}^{(ls)}\right)^{-1}\boldsymbol{R}_{12}^{(ls)}; \qquad \boldsymbol{\mu}_{(ls)/-(ls)} = \boldsymbol{R}_{21}^{(ls)}\left(\boldsymbol{R}_{11}^{(ls)}\right)^{-1}\boldsymbol{a}_{-(ls)}$$

For a multivariate normal random vector $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_t)$ with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{R}$, the above expressions are precisely the variance and mean of the conditional random vector $\boldsymbol{Z}_{-(ls)/(ls)}$, which is $\boldsymbol{Z}_{-(ls)}$ given $\boldsymbol{Z}_{(ls)}$ is equal to $\boldsymbol{a}_{(ls)}$ and the conditional random vector $\boldsymbol{Z}_{(ls)/-(ls)}$, which is $\boldsymbol{Z}_{(ls)}$ given $\boldsymbol{Z}_{-(ls)}$ is equal to $\boldsymbol{a}_{-(ls)}$ respectively. The next lemma provides the derivative of a multivariate normal cumulative distribution function with respect to the latent correlations.

LEMMA 1. *Suppose $\Phi_t(a_1, a_2, ..., a_t; \mathbf{0}, \boldsymbol{R})$ denotes a t-variate normal distribution function with mean $\mathbf{0}$ and correlation matrix $\boldsymbol{R}$. Let r_{ls} be the (ls)th element of the correlation matrix $\boldsymbol{R}$. Then the derivative of Φ_t with respect to r_{ls} is*

$$\frac{\partial}{\partial r_{ls}}\Phi_t(a_1, a_2, ..., a_t; \mathbf{0}, \boldsymbol{R}) = \Phi_{t-2}(\boldsymbol{a}_{-(ls)}; \boldsymbol{\mu}_{-(ls)/(ls)}, \boldsymbol{R}_{-(ls)/(ls)})\ \phi_2(\boldsymbol{a}_{(ls)}; \mathbf{0}, \boldsymbol{R}_{22}^{(ls)}).$$

Proof. Let $P = \Phi_t(a_1, a_2, ..., a_t; \mathbf{0}, \boldsymbol{R})$. Then

$$\begin{aligned} P &= \int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_t} \phi_t(\boldsymbol{z}; \mathbf{0}, \boldsymbol{R})\, d\boldsymbol{z} \\ &= \int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_t} \phi_{t-2}(\boldsymbol{z}_{-(ls)}; \mathbf{0}, \boldsymbol{R}_{11}^{(ls)})\ \phi_2(\boldsymbol{z}_{(ls)}; M_{(ls)/-(ls)}, \boldsymbol{R}_{(ls)/-(ls)})\, d\boldsymbol{z} \\ &= \int_{-\infty}^{\boldsymbol{a}_{-(ls)}} \phi_{t-2}(\boldsymbol{z}_{-(ls)}; \mathbf{0}, \boldsymbol{R}_{11}^{(ls)})\ \Phi_2(\boldsymbol{a}_{(ls)}; M_{(ls)/-(ls)}, \boldsymbol{R}_{(ls)/-(ls)})\, d\boldsymbol{z}_{-(ls)}, \end{aligned} \tag{61}$$

where

$$\int_{-\infty}^{\boldsymbol{a}_{-(ls)}} = \int_{-\infty}^{a_1}\int_{-\infty}^{a_2}\cdots\int_{-\infty}^{a_{l-1}}\int_{-\infty}^{a_{l+1}}\cdots\int_{-\infty}^{a_{s-1}}\int_{-\infty}^{a_{s+1}}\cdots\int_{-\infty}^{a_t},$$

$$M_{(ls)/-(ls)} = \boldsymbol{R}_{21}^{(ls)}\left(\boldsymbol{R}_{11}^{(ls)}\right)^{-1}\boldsymbol{z}_{(ls)} = \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} \quad \text{and}$$

$$\boldsymbol{R}_{(ls)/-(ls)} = \begin{pmatrix} 1 & r_{ls} \\ r_{ls} & 1 \end{pmatrix} - \begin{pmatrix} \delta_{11} & \delta_{12} \\ \delta_{12} & \delta_{22} \end{pmatrix} = \begin{pmatrix} 1-\delta_{11} & r_{ls}-\delta_{12} \\ r_{ls}-\delta_{12} & 1-\delta_{22} \end{pmatrix}.$$

Note that η_i and δ_{ij} do not depend on r_{ls}. Consider

$$\begin{aligned} \Phi_2\left[(a_l, a_s); M_{(ls)/-(ls)}, \boldsymbol{R}_{(ls)/-(ls)}\right] &= P\left(W_1 \le a_l, W_2 \le a_s\right) \\ \text{where} \begin{pmatrix} W_1 \\ W_2 \end{pmatrix} &\sim N\left(\begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix}, \begin{pmatrix} 1-\delta_{11} & r_{ls}-\delta_{12} \\ r_{ls}-\delta_{12} & 1-\delta_{22} \end{pmatrix}\right). \end{aligned}$$

After standardizing we get

$$\Phi_2\left((a_l,a_s);M_{(ls)/-(ls)},\boldsymbol{R}_{(ls)/-(ls)}\right)$$

$$\begin{aligned} &= P\left(\frac{W_1-\eta_1}{\sqrt{1-\delta_{11}}}\le\frac{a_l-\eta_1}{\sqrt{1-\delta_{11}}},\frac{W_2-\eta_2}{\sqrt{1-\delta_{22}}}\le\frac{a_s-\eta_2}{\sqrt{1-\delta_{22}}}\right)\\ &= P\left(W_1^*\le a_l^*,W_2^*\le a_s^*\right),\end{aligned}$$

$$\text{where}\begin{pmatrix}W_1^*\\W_2^*\end{pmatrix}\sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1 & r_{ls}^*\\ r_{ls}^* & 1\end{pmatrix}\right)\sim N\left(\mathbf{0},\Lambda_{ls}\right).$$

$$\text{and } r_{ls}^* = \frac{r_{ls}-\delta_{12}}{\sqrt{1-\delta_{11}}\sqrt{1-\delta_{22}}}.$$

Therefore,

$$\begin{aligned}\frac{\partial}{\partial r_{ls}}\Phi_2\left((a_l,a_s);M_{(ls)/-(ls)},\boldsymbol{R}_{(ls)/-(ls)}\right) &= \frac{\partial}{\partial r_{ls}}\Phi_2\left((a_l^*,a_s^*);\mathbf{0},\Lambda_{ls}\right)\\ &= \frac{\partial}{\partial r_{ls}^*}\Phi_2\left((a_l^*,a_s^*);\mathbf{0},\Lambda_{ls}\right)\ \frac{\partial r_{ls}^*}{\partial r_{ls}}\end{aligned}$$

$$=\int_{-\infty}^{a_l^*}\int_{-\infty}^{a_s^*}\frac{\partial}{\partial r_{ls}^*}\phi_2(w_1^*,w_2^*;\mathbf{0},\Lambda_{ls})dw_1^*dw_2^* \quad\times\quad \frac{1}{\sqrt{1-\delta_{11}}\sqrt{1-\delta_{22}}}.$$

Using Plackett's identity (Kotz et al. (2000), page 259) we have

$$\begin{aligned} &= \int_{-\infty}^{a_l^*}\int_{-\infty}^{a_s^*}\frac{\partial^2}{\partial w_1^*\partial w_2^*}\phi_2(w_1^*,w_2^*;\mathbf{0},\Lambda_{ls})dw_1^*dw_2^*.\ \frac{1}{\sqrt{1-\delta_{11}}\sqrt{1-\delta_{22}}}\\ &= \phi_2((a_l^*,a_s^*);\mathbf{0},\Lambda_{ls})\ \frac{1}{\sqrt{1-\delta_{11}}\sqrt{1-\delta_{22}}}\\ &= \phi_2\left((a_l,a_s);M_{(ls)/-(ls)},\boldsymbol{R}_{(ls)/-(ls)}\right).\end{aligned}$$

Using equation (61) we get

$$\begin{aligned}\frac{\partial P}{\partial r_{ls}} &= \frac{\partial}{\partial r_{ls}}\int_{-\infty}^{\boldsymbol{a}_{-(ls)}}\phi_{t-2}(\boldsymbol{z}_{-(ls)};\mathbf{0},\boldsymbol{R}_{11}^{(ls)})\ \Phi_2(\boldsymbol{a}_{(ls)};M_{(ls)/-(ls)},\boldsymbol{R}_{(ls)/-(ls)})\ d\boldsymbol{z}_{-(ls)}\\ &= \int_{-\infty}^{\boldsymbol{a}_{-(ls)}}\phi_{t-2}(\boldsymbol{z}_{-(ls)};\mathbf{0},\boldsymbol{R}_{11}^{(ls)})\ \frac{\partial}{\partial r_{ls}}\Phi_2(\boldsymbol{a}_{(ls)};M_{(ls)/-(ls)},\boldsymbol{R}_{(ls)/-(ls)})\ d\boldsymbol{z}_{-(ls)}\\ &= \int_{-\infty}^{\boldsymbol{a}_{-(ls)}}\phi_{t-2}(\boldsymbol{z}_{-(ls)};\mathbf{0},\boldsymbol{R}_{11}^{(ls)})\ \phi_2(\boldsymbol{a}_{(ls)};M_{(ls)/-(ls)},\boldsymbol{R}_{(ls)/-(ls)})\ d\boldsymbol{z}_{-(ls)}\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\boldsymbol{a}_{-(ls)}} \phi_t \left[(z_1, \cdots, z_{l-1}, a_l, z_{l+1}, \cdots, z_{s-1}, a_s, z_{s+1}, \cdots, z_t); \mathbf{0}, \boldsymbol{R}\right] \, d\boldsymbol{z}_{-(ls)} \\
&= \int_{-\infty}^{\boldsymbol{a}_{-(ls)}} \phi_2(\boldsymbol{a}_{(ls)}; \mathbf{0}, \boldsymbol{R}_{22}^{(ls)}) \, \phi_{t-2}(\boldsymbol{z}_{-(ls)}; M_{-(ls)/(ls)}, \boldsymbol{R}_{-(ls)/(ls)}) \, d\boldsymbol{z}_{-(ls)} \\
&= \phi_2(\boldsymbol{a}_{(ls)}; \mathbf{0}, \boldsymbol{R}_{22}^{(ls)}) \int_{-\infty}^{\boldsymbol{a}_{-(ls)}} \phi_{t-2}(\boldsymbol{z}_{-(ls)}; M_{-(ls)/(ls)}, \boldsymbol{R}_{-(ls)/(ls)}) \, d\boldsymbol{z}_{-(ls)} \\
&= \Phi_{t-2}(\boldsymbol{a}_{-(ls)}; \boldsymbol{\mu}_{-(ls)/(ls)}, \boldsymbol{R}_{-(ls)/(ls)}) \, \phi_2(\boldsymbol{a}_{(ls)}; \mathbf{0}, \boldsymbol{R}_{22}^{(ls)}). \qquad \square
\end{aligned}
$$

Now we are in a position to derive the derivatives of the multivariate probability mass function (59) with respect to the latent correlation parameters.

THEOREM 1. *Let $\pi_t = \pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R})$ be probability mass function of the multivariate ordered probit model as defined in (59). Then the derivative of π_t with respect to the (ls)th element of matrix $\boldsymbol{R}$ is*

$$
\begin{aligned}
\frac{\partial}{\partial r_{ls}} \pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R}) \;=\; \sum_{i_l=1}^{2} \sum_{i_s=1}^{2} \Big((-1)^{i_l+i_s} \phi_2((b_{li_l}, b_{si_s}); \mathbf{0}, \boldsymbol{R}_{22}^{(ls)}). \\
\pi_{t-2}(\boldsymbol{y}_{-(ls)}; \boldsymbol{\mu}_{-(ls)/(ls)}, \boldsymbol{R}_{-(ls)/(ls)}) \Big)
\end{aligned}
$$

where $b_{j1} = \gamma_j(y_j - 1)$ and $b_{j2} = \gamma_j(y_j)$.

Proof. Note that from equation (59) we have

$$
\pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R}) = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \cdots \sum_{i_t=1}^{2} (-1)^{i_1+i_2+\cdots+i_t} . \Phi_t(b_{1i_1}, b_{2i_2}, \ldots, b_{ti_t}; \mathbf{0}, \boldsymbol{R}).
$$

So the derivative of $\pi_t(\boldsymbol{y}; \mathbf{0}, \boldsymbol{R})$ with respect to r_{ls} is

$$
\begin{aligned}
\frac{\partial \pi_t}{\partial r_{ls}} &= \frac{\partial}{\partial r_{ls}} \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \cdots \sum_{i_t=1}^{2} (-1)^{i_1+i_2+\cdots+i_t} \Phi_t(b_{1i_1}, b_{2i_2}, \ldots, b_{ti_t}; \mathbf{0}, \boldsymbol{R}) \\
&= \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \cdots \sum_{i_t=1}^{2} (-1)^{i_1+i_2+\cdots+i_t} \frac{\partial}{\partial r_{ls}} \Phi_t(b_{1i_1}, b_{2i_2}, \ldots, b_{ti_t}; \mathbf{0}, \boldsymbol{R}).
\end{aligned}
$$

By Lemma 2 we have

$$
\begin{aligned}
\frac{\partial \pi_t}{\partial r_{ls}} &= \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_t=1}^{2}(-1)^{i_1+i_2+\quad+i_t} \\
&\qquad\qquad \Big(\Phi_{t-2}\big(\boldsymbol{b}_{-(ls)};\boldsymbol{\mu}_{-(ls)/(ls)},\boldsymbol{R}_{-(ls)/(ls)}\big)\;\phi_2\big((b_{li_l},b_{si_s});\mathbf{0},\boldsymbol{R}_{22}^{(ls)}\big)\Big) \\
&= \sum_{i_1=1}^{2}\sum_{i_2=1}^{2}\cdots\sum_{i_t=1}^{2}\Big((-1)^{i_l+i_s}.\phi_2\big((b_{li_l},b_{si_s});\mathbf{0},\boldsymbol{R}_{22}^{(ls)}\big) \\
&\qquad (-1)^{i_1+\quad+i_{l-1}+i_{l+1}+\quad+i_{s-1}+i_{s+1}+\quad+i_t}\Phi_{t-2}\big(\boldsymbol{b}_{-(ls)};\boldsymbol{\mu}_{-(ls)/(ls)},\boldsymbol{R}_{-(ls)/(ls)}\big)\Big) \\
&= \sum_{i_l=1}^{2}\sum_{i_s=1}^{2}\Big((-1)^{i_l+i_s}\phi_2\big((b_{li_l},b_{si_s});\mathbf{0},\boldsymbol{R}_{22}^{(ls)}\big).\;\pi_{t-2}\big(\boldsymbol{y}_{-(ls)};\boldsymbol{\mu}_{-(ls)/(ls)},\boldsymbol{R}_{-(ls)/(ls)}\big)\Big)
\end{aligned}
$$

This completes the proof of the theorem. $\square$

APPENDIX B

R PROGRAM

Here we provide details of the R program that we developed for fitting the multivariate ordered probit-logit models. The program consists of several functions. Details of these functions are given below.

1. sigma: Constructs the correlation matrix for a given rho (ρ), size, and structure (only AR(1) and CS structures are allowed).
2. negloglik: Calculates the negative log-likelihood function for a given theta ($\boldsymbol{\theta}$) and data values based on given link function. This program requires MProbit package in R.
3. Grad_Anal_beta: Calculates the analytical derivatives of negative log-likelihood with respect to regression parameter ($\boldsymbol{\beta}$) for a given theta ($\boldsymbol{\theta}$) and data values.
4. Grad_Anal_R: Calculates the analytical derivatives of negative log-likelihood with respect to latent correlation parameters ($\boldsymbol{\rho}$) for a given theta ($\boldsymbol{\theta}$), data values, and correlation structure.
5. Grad_Anal: Combines the analytical derivatives of negative log-likelihood with respect to regression parameters ($\boldsymbol{\beta}$) and latent correlation parameters ($\boldsymbol{\rho}$).
6. QNMin: Minimizes the negative log-likelihood function for a given initial $\boldsymbol{\theta}$ values using a quasi-Newton algorithm. (See Chapter II, Section II.4.3 for details.)
7. OrdGuasCopula: Main function which uses all of the above functions, inputs the data values and other instructions such as correlation structure and link function, and outputs the final results.

Remarks:

1. The main function OrdGuasCopula takes the data and the initial parameter values and makes them global. Thus the data and the initial parameter values can be accessed by all the other functions readily.

2. The functions Grad_Anal_beta and Grad_Anal_R use negloglik function. The QNMin function uses both the negloglik and the Grad_Anal functions.

3. The data structure for the main program should have Id variable which is a subject indicator. Also for the categorical independent variables, dummy variables must be created manually.

```
library(mprobit)
# Construct the Correlation matrix for a given rho and size
sigma <- function(rho)
{
  if (struct == "ar")
  {
    m <- diag(nt)
    sigma <- rho^(abs(row(m)-col(m)))
  }
  if (struct == "cs")
  {
    sigma <- matrix(rep(rho,nt*nt),nt,nt)
    diag(sigma) <- 1.0
  }
  return(sigma)
}


# for a given theta and data values calculates the log-likelihood
# based on link function
negloglik <- function(theta)
{
  alpha <- theta[1:nc-1]
  alpha <- c(-6,alpha,6)
  beta <- theta[nc:(nc+np-1)]
  rho <- theta[length(theta)]
  if (struct == "ar" || struct == "cs")
    if (rho < 0.0 || rho >= 1.0) return(1.e10)

  value <- 0
  Sig <- sigma(rho)

  en <- cumsum(unlist(lapply(split(id,f=id),length)))#cluster ending points
  st <- c(1,en[-length(en)]+1)#cluster starting points
```

```
  for (i in 1:nsub)
  {
    xi <- XCov[st[i]:en[i],]
    if (np == 1)
      Mean <- as.vector(xi*beta)
    else
      Mean <- as.vector(xi%*%beta)

    yi <- YRes[st[i]:en[i]]
    Lower <- alpha[yi] + Mean
    Upper <- alpha[yi+1] + Mean

    # probit link is default, Lower = Lower and Upper = Upper.
    if (link == "logit")
    {
      Lower <- qnorm(exp(Lower)/(1+exp(Lower)))
      Upper <- qnorm(exp(Upper)/(1+exp(Upper)))
    }

    Sigi <- Sig[1:(en[i]-st[i]+1),1:(en[i]-st[i]+1)]
    if(length(yi) > 1)
    {
      TempObj <- mvnapp(Lower,Upper,rep(0,length(Lower)),Sigi,eps=1.e-03)
      if (TempObj$ifail > 0)
        return(1.e10)
      JointProbi <- TempObj$pr
    }
    else
      JointProbi <- (pnorm(Upper[1]) - pnorm(Lower[1]))

    if (is.nan(JointProbi)) return(1.e9)
    if (JointProbi <= -1) return(1.e10)
    if (JointProbi <= 0) JointProbi = 1.0e-15

    value <- value + log(JointProbi)
  }
  return(-value)
}
```

```
# for a given theta and data values calculates analytical derivatives
# with respect to regression parameters and cutoff points
Grad_Anal_beta <- function(theta)
{
  alpha <- theta[1:nc-1]
  alpha <- c(-6,alpha,6)
  beta <- theta[nc:(nc+np-1)]
  rho <- theta[length(theta)]
  if (struct == "ar" || struct == "cs")
    if (rho < 0.0 || rho >= 1.0) return(0)

  Deriv <- 0
  Sig <- sigma(rho)

  en <- cumsum(unlist(lapply(split(id,f=id),length))) #cluster ending points
  st <- c(1,en[-length(en)]+1) # cluster starting points

  for (i in 1:nsub)
  {
    if((en[i]-st[i]+1) > 1)
      xi <- XCov[st[i]:en[i],]
    else
      xi <- matrix(XCov[st[i]:en[i],],1,length(XCov[st[i]:en[i],]))

    if (np == 1)
      Mean <- as.vector(xi*beta)
    else
      Mean <- as.vector(xi%*%beta)

    yi <- YRes[st[i]:en[i]]
    Lower <- alpha[yi] + Mean
    Upper <- alpha[yi+1] + Mean

    # probit link is default, Lower = Lower and Upper = Upper.
    if (link == "logit")
    {
      Lower <- qnorm(exp(Lower)/(1+exp(Lower)))
      Upper <- qnorm(exp(Upper)/(1+exp(Upper)))
    }

    Sigi <- Sig[1:(en[i]-st[i]+1),1:(en[i]-st[i]+1)]
```

```
if(length(yi) > 1)
{
  TempObj <- mvnapp(Lower,Upper,rep(0,length(Lower)),Sigi,eps=1.e-03)

  if (is.nan(TempObj$pr)){next}
  if (TempObj$ifail > 0){next}
  if(TempObj$pr > 1){next}
  if(TempObj$pr == 0){next}

  JointProbi <- TempObj$pr
}
else
  JointProbi <- (pnorm(Upper[1]) - pnorm(Lower[1]))

alpha_deriv_i <- rep(0,nc-1)
beta_deriv_i <- rep(0,np)

for (j in 1:(length(Lower)))
{
  if(length(yi) > 1)
  {
    CondRj <- Sigi[-j,-j] - (Sigi[-j,j]%*%t(Sigi[j,-j]))
    CondMuj1 <- Sigi[-j,j]*Upper[j]
    CondMuj2 <- Sigi[-j,j]*Lower[j]
  }
  if(length(yi) > 2)
  {
    TempObj <- mvnapp(Lower[-j],Upper[-j],CondMuj1,CondRj,eps=1.e-03)
    if (TempObj$ifail > 0){TempObj$pr <- 0}
    if (is.nan(TempObj$pr)) {TempObj$pr <- 0}
    if(TempObj$pr > 1){TempObj$pr <- 0}
    term1 <- dnorm(Upper[j],0,1)*TempObj$pr
    if (is.nan(term1)) {term1 <- 0}

    TempObj <- mvnapp(Lower[-j],Upper[-j],CondMuj2,CondRj,eps=1.e-03)
    if (TempObj$ifail > 0){TempObj$pr <- 0}
    if (is.nan(TempObj$pr)) {TempObj$pr <- 0}
    if(TempObj$pr > 1){TempObj$pr <- 0}
    term2 <- dnorm(Lower[j],0,1)*TempObj$pr
    if (is.nan(term2)) {term2 <- 0}
  }
```

```
      else if (length(yi) == 2)
      {
        term1 <- dnorm(Upper[j],0,1)*(pnorm(Upper[-j],CondMuj1,CondRj)
                 - pnorm(Lower[-j],CondMuj1,CondRj))
        term2 <- dnorm(Lower[j],0,1)*(pnorm(Upper[-j],CondMuj2,CondRj)
                 - pnorm(Lower[-j],CondMuj2,CondRj))
      }
      else if (length(yi) == 1)
      {
        term1 <- dnorm(Upper[j],0,1)
        term2 <- dnorm(Lower[j],0,1)
      }

      if (np == 1)
        beta_deriv_i <- beta_deriv_i + (xi[j]*(term1-term2))
      else
        beta_deriv_i <- beta_deriv_i + (xi[j,]*(term1-term2))

      for (k in 1:(nc-1))
      {
        if (yi[j] == k) {alpha_deriv_i[k] = alpha_deriv_i[k] + term1}
        else if ((yi[j] - 1) == k) {alpha_deriv_i[k] = alpha_deriv_i[k] - term2}
      }
    }
    deriv_i <- c(alpha_deriv_i,beta_deriv_i)
    Deriv <- Deriv + (deriv_i/JointProbi)
  }
  return(-Deriv)
}


# for a given theta and data values calculates analytical derivatives
#with respect to R
Grad_Anal_R <- function(theta)
{
  alpha <- theta[1:nc-1]
  alpha <- c(-6,alpha,6)
  beta <- theta[nc:(nc+np-1)]
  rho <- theta[length(theta)]
  if (struct == "ar" || struct == "cs")
    if (rho < 0.0 || rho >= 1.0) return(0)
  DerivR <- 0
```

```
Sig <- sigma(rho)

en <- cumsum(unlist(lapply(split(id,f=id),length))) #cluster ending points
st <- c(1,en[-length(en)]+1) # cluster starting points

for (i in 1:nsub)
{
  if((en[i]-st[i]+1) > 1)
    xi <- XCov[st[i]:en[i],]
  else
    xi <- matrix(XCov[st[i]:en[i],],1,length(XCov[st[i]:en[i],]))

  if (np == 1)
    Mean <- as.vector(xi*beta)
  else
    Mean <- as.vector(xi%*%beta)

  yi <- YRes[st[i]:en[i]]
  Lower <- alpha[yi] + Mean
  Upper <- alpha[yi+1] + Mean

  # probit link is default, Lower = Lower and Upper = Upper.
  if (link == "logit")
  {
    Lower <- qnorm(exp(Lower)/(1+exp(Lower)))
    Upper <- qnorm(exp(Upper)/(1+exp(Upper)))
  }

  Sigi <- Sig[1:(en[i]-st[i]+1),1:(en[i]-st[i]+1)]
  if(length(yi) > 1)
  {
    TempObj <- mvnapp(Lower,Upper,rep(0,length(Lower)),Sigi,eps=1.e-03)
    if (is.nan(TempObj$pr)){next}
    if (TempObj$ifail > 0){next}
    if(TempObj$pr > 1){next}
    if(TempObj$pr == 0){next}
    JointProbi <- TempObj$pr
  }
  else
    JointProbi <- (pnorm(Upper[1]) - pnorm(Lower[1]))
```

```
r_jk_deriv_i <- c(0)

for (j in 1:(length(Lower)-1))
{
  for(k in (j+1):(length(Lower)))
  {
    if(length(yi) > 1)
      R22 <- Sigi[c(j,k),c(j,k)]

    if(length(yi) > 2)
    {
      CondRj <- Sigi[-c(j,k),-c(j,k)] - (Sigi[-c(j,k),c(j,k)]
                         %*%ginv(Sigi[c(j,k),c(j,k)])%*%Sigi[c(j,k),-c(j,k)])
      CondMuj1 <- Sigi[-c(j,k),c(j,k)]%*%ginv(Sigi[c(j,k),c(j,k)])
                         %*%matrix(c(Lower[j],Lower[k]),2,1)
      CondMuj2 <- Sigi[-c(j,k),c(j,k)]%*%ginv(Sigi[c(j,k),c(j,k)])
                         %*%matrix(c(Lower[j],Upper[k]),2,1)
      CondMuj3 <- Sigi[-c(j,k),c(j,k)]%*%ginv(Sigi[c(j,k),c(j,k)])
                         %*%matrix(c(Upper[j],Lower[k]),2,1)
      CondMuj4 <- Sigi[-c(j,k),c(j,k)]%*%ginv(Sigi[c(j,k),c(j,k)])
                         %*%matrix(c(Upper[j],Upper[k]),2,1)

      if (length(yi) == 3)
      {
        TempObj <- (pnorm(Upper[-c(j,k)],CondMuj1,CondRj)
                         - pnorm(Lower[-c(j,k)],CondMuj1,CondRj))
        term1 <- (dmvnorm(c(Lower[j],Lower[k]), rep(0,2), R22,
                                             log=FALSE))*TempObj
        TempObj <- (pnorm(Upper[-c(j,k)],CondMuj2,CondRj)
                         - pnorm(Lower[-c(j,k)],CondMuj2,CondRj))
        term2 <- (dmvnorm(c(Lower[j],Upper[k]), rep(0,2), R22,
                                             log=FALSE))*TempObj
        TempObj <- (pnorm(Upper[-c(j,k)],CondMuj3,CondRj)
                         - pnorm(Lower[-c(j,k)],CondMuj3,CondRj))
        term3 <- (dmvnorm(c(Upper[j],Lower[k]), rep(0,2), R22,
                                             log=FALSE))*TempObj
        TempObj <- (pnorm(Upper[-c(j,k)],CondMuj4,CondRj)
                         - pnorm(Lower[-c(j,k)],CondMuj4,CondRj))
        term4 <- (dmvnorm(c(Upper[j],Upper[k]), rep(0,2), R22,
                                             log=FALSE))*TempObj
      }
```

```
  else
  {
    TempObj <- mvnapp(Lower[-c(j,k)],Upper[-c(j,k)],CondMuj1,
                                                  CondRj,eps=1.e-03)
    if (is.nan(TempObj$pr)) {TempObj$pr <- 0}
    if (TempObj$ifail > 0){TempObj$pr <- 0}
    if(TempObj$pr > 1){TempObj$pr <- 0}
    term1 <- (dmvnorm(c(Lower[j],Lower[k]), rep(0,2), R22,
                                              log=FALSE))*TempObj$pr
    TempObj <- mvnapp(Lower[-c(j,k)],Upper[-c(j,k)],CondMuj2,
                                                  CondRj,eps=1.e-03)
    if (is.nan(TempObj$pr)) {TempObj$pr <- 0}
    if (TempObj$ifail > 0){TempObj$pr <- 0}
    if(TempObj$pr > 1){TempObj$pr <- 0}
    term2 <- (dmvnorm(c(Lower[j],Upper[k]), rep(0,2), R22,
                                              log=FALSE))*TempObj$pr
    TempObj <- mvnapp(Lower[-c(j,k)],Upper[-c(j,k)],CondMuj3,
                                                  CondRj,eps=1.e-03)
    if (is.nan(TempObj$pr)) {TempObj$pr <- 0}
    if (TempObj$ifail > 0){TempObj$pr <- 0}
    if(TempObj$pr > 1){TempObj$pr <- 0}
    term3 <- (dmvnorm(c(Upper[j],Lower[k]), rep(0,2), R22,
                                              log=FALSE))*TempObj$pr
    TempObj <- mvnapp(Lower[-c(j,k)],Upper[-c(j,k)],CondMuj4,
                                                  CondRj,eps=1.e-03)
    if (is.nan(TempObj$pr)) {TempObj$pr <- 0}
    if (TempObj$ifail > 0){TempObj$pr <- 0}
    if(TempObj$pr > 1){TempObj$pr <- 0}
    term4 <- (dmvnorm(c(Upper[j],Upper[k]), rep(0,2), R22,
                                              log=FALSE))*TempObj$pr
  }
}

if(length(yi) == 2)
{
  term1 <- dmvnorm(c(Lower[j],Lower[k]), rep(0,2), R22, log=FALSE)
  term2 <- dmvnorm(c(Lower[j],Upper[k]), rep(0,2), R22, log=FALSE)
  term3 <- dmvnorm(c(Upper[j],Lower[k]), rep(0,2), R22, log=FALSE)
  term4 <- dmvnorm(c(Upper[j],Upper[k]), rep(0,2), R22, log=FALSE)
}
```

```
        if(length(yi) < 2)
        {
          term1 <- term2 <- term3 <- term4 <- 0
        }
        if (struct == "ar")
          r_jk_deriv_i <- r_jk_deriv_i + ((k-j)*(rho^(k-j-1)))
                                              *(term1-term2-term3+term4)
        else if (struct == "cs")
          r_jk_deriv_i <- r_jk_deriv_i + (term1-term2-term3+term4)
        if (is.nan(r_jk_deriv_i)) return(0)
      }
    }
    DerivR <- DerivR + (r_jk_deriv_i/JointProbi)
  }
  return(-DerivR)
}


# Combines the derivative (w.r.t beta and R) vectors
Grad_Anal <- function(theta)
{
  Deriv_beta <- Grad_Anal_beta(theta)
  Deriv_R <- Grad_Anal_R(theta)
  return(c(Deriv_beta,Deriv_R))
}


# Quasi-Newton Minimization algorithm
QNMin <- function(Theta)
{
  np <- length(Theta); b <- Theta
  ig <- 1; ifn <- np+1; w <- 0.2; tol <- 0.0001; eps <- 0.00001
  bool1 <- bool2 <- bool3 <- "T"

  P0 = negloglik(b)
  if (P0 > 0.1e309)
  {
    Err <- 1
    return(rep(0,np))
  }

  g <-  Grad_Anal(b)
  l = 0
```

```
while (bool1 == "T")
{
  l <- l + 1
  if (bool2 == "T")
  {
    H <- diag(np)
    bool2 <- "F"
  }

  x <- b; c <- g
  t1 <- (-1)*H%*%g
  D1 <- (-1)%*%t(g)%*%t1
  sn <- t(t1)%*%t1
  if (D1 < 0)
  {
    bool2 <- "T"
    next
  }

  sn <- 0.5/sqrt(sn)
  k <- c(1)
  if (sn < k)
    k <- sn

  bool3 <- "T"
  while (bool3 == "T")
  {
    cnt <- 0
    for(i in 1:np)
    {
      b[i] <- x[i]+k*t1[i]
      if (abs(b[i] - x[i]) < eps)
        cnt <- cnt + 1
    }
    if (cnt == np)
      return(list(b = b,H = H,P0 = P0))
    else
    {
      P1 <- negloglik(b)
      ifn <- ifn + 1
    }
```

```
      if (P1 >= P0 - D1*k*tol)
      {
        k <- k*w
        next
      }
      else
        bool3 <- "F"
    }

    P0 <- P1
    g <- Grad_Anal(b);
    ig <- ig + 1; ifn <- ifn + np
    t1 <- k[1]*t1; c <- g - c; D1 <- c%*%t1
    if (D1 <= 0)
    {
      bool2 <- "T"
      next
    }
    else
    {
      x <- H%*%c
      D2 <- t(x)%*%c
    }

    D2 <- 1+(D2/D1)

    for(r in 1:np)
      for(s in 1:np)
        H[r,s] = H[r,s] - ((t1[r]*x[s] + x[r]*t1[s] - D2*t1[r]*t1[s])/D1)
  } # main while until loop;
}

# Main function starts here
OrdGuasCopula<- function(y,x,id,struct,link)
{
  nc <- max(y) # no.of categories of y
  np <- ncol(x) #no.of parameters
  nt <- max(unlist(lapply(split(id,f=id),length)))
              # no.of time points (maximum cluster size)
  nrec <- length(y) # Total no.of records
  nsub <- length(unlist(lapply(split(id,f=id),length))) # no.of subjects
```

```
#Global variables decleration
assign("XCov", x, envir = .GlobalEnv)
assign("YRes", y, envir = .GlobalEnv)
assign("Id", id, envir = .GlobalEnv)
assign("nc", nc, envir = .GlobalEnv)
assign("np", np, envir = .GlobalEnv)
assign("nt", nt, envir = .GlobalEnv)
assign("nsub", nsub, envir = .GlobalEnv)
assign("struct", struct, envir = .GlobalEnv)
assign("link", link, envir = .GlobalEnv)

if(nc == 2)   #for Initial values when the response has only 2 - categories
{
  ybin <- 2-y
  xx <- cbind(rep(1,nrec),x)
  names(xx)[1] <- "intcpt"
  th <- glm.fit(xx,ybin,family=binomial(link="probit"))$coef
  Initial <- c(th,.4)
}
else #for Initial values when the response more than 2 - categories
{
  cum <- (1:(nc-1))
  cut <- rep(0,nc-1)
  for(k in cum)
  {
    pr=sum(y<=k)
    if (pr==0) pr=1
    cut[k]=qnorm(pr/nrec)
  }
    Initial <- c(cut,rep(0,np),.4)
}

source(paste(CodeDir,"QNMin_Deriv.r",sep=""))
result <- QNMin(Initial)

if(link == "logit")
  Marginal <- "Logistic marginal distribution used with Gaussian Copula"
else
  Marginal <- "Normal marginal distribution used with Gaussian Copula"
```

```
  if(struct == "ar")
    CorrStr <- "Autoregression(1) correlation structure"
  else if (struct == "cs")
    CorrStr <- "Compound Symmetry correlation structure"

  Est <- result$b
  Hess <- result$H
  NegLogLik <- result$P0
  SE <- sqrt(diag(Hess))
  PVal <- 2*(1-pnorm(abs(Est),mean=0,sd=SE))

  ParEstTable <- round(cbind(Est, SE, PVal),8)
  return(list(Marginal = Marginal, CorrStr = CorrStr, Initial = Initial,
                       NegLogLik = NegLogLik, ParEstTable = ParEstTable))
}

# An illustrative example how to use the above program to fit the gaussian
# copula based ordered probit-logit models.

DataDir <- "H:/Research/Programs/DataSets/"
CodeDir <- "H:/Research/Programs/Ordered Probit/"
source(paste(CodeDir,"Deriv.r",sep=""))
library(mprobit)

Ex <- read.table(paste(DataDir,"Sixcities_Mult.csv",sep=""),header=T, sep=",")
attach(Ex)
y <- Y
SMOKE[which(SMOKE>0)]=1
x <- cbind(CITY,SMOKE,TIME+8)
id <- ID
detach(Ex)

ProbitOutputEx1 <- OrdGuasCopula(y,x,id,"ar","probit")
print(ProbitOutputEx1)
LogitOutputEx1 <- OrdGuasCopula(y,x,id,"ar","logit")
print(LogitOutputEx1)

ProbitOutputEx2 <- OrdGuasCopula(y,x,id,"cs","probit")
print(ProbitOutputEx2)
LogitOutputEx2 <- OrdGuasCopula(y,x,id,"cs","logit")
print(LogitOutputEx2)
```

VITA

Raghavendra Rao Kurada
Department of Mathematics and Statistics
Old Dominion University
Norfolk, VA 23529

Education

Ph.D	Old Dominion University, Norfolk, VA. (December 2011) Major: Computational and Applied Mathematics. (Statistics)
MS	Old Dominion University, Norfolk, VA. (May 2010) Major: Computational and Applied Mathematics. (Statistics)
M.Sc	University of Hyderabad, Hyderabad, AP, India. (June 2005) Major: Statistics.
B.Sc	Osmania University, Hyderabad, AP, India. (June 2003) Majors: Mathematics, Statistics, and Computer Science.

Experience

Graduate Research Assistant, Benthic Ecology Lab (08/2007 - 12/2011)
Old Dominion University, Norfolk, VA.

Statistical Analyst, Cranes Software International Limited (07/2005 - 06/2007)
Bangalore, KA, India.

Publications

Kurada, R. R., Sabo, R. T., and Chaganty, N. R. (2011). Copula Based Modeling and Estimation of Repeated Ordinal Data. In preparation.

Saxena, A. and **Kurada, R. R.** (2006). Regression Diagnostics. In *Statistical Analysis with Software*,**6**, Ed. Bhogle, S. Cranes Software International Ltd., Bangalore, India.

Typeset using LaTeX.