INFERENCES IN LONGITUDINAL MULTINOMIAL FIXED AND MIXED MODELS

MD. RAFIQUL I CHOWDHURY

Inferences in Longitudinal Multinomial Fixed and Mixed Models

by

©Md. Rafiqul I Chowdhury

A thesis submitted to the School of Graduate Studies

in partial fulfillment of the requirement for the Degree of

Doctor of Philosophy in Statistics

Department of Mathematics and Statistics

Memorial University of Newfoundland

August, 2011

St. John's Newfoundland Canada

Abstract

Analyzing categorical data collected over time is an important research topic. Even though there exists numerous studies on analysis of categorical data in cross sectional setup, the analysis of this type of data in the longitudinal setup is, however, not adequately addressed. In this thesis, we develop two correlation models for multinomial (> 2 categories) longitudinal data, namely, a conditional linear probability based model and a non-linear logistic probability based model; and provide likelihood inferences for category effects, fixed covariate effects and correlations or dynamic dependence parameters. The inferences are done for both complete history and contingency tables based data. For the history based data, the thesis also models the influences of individual random effects in addition to the fixed covariate effects. Furthermore, as in many practical situations the number of individuals involved in the study may be small, in the thesis, we have examined the finite sample performance of the likelihood estimates both in fixed and mixed model setups.

Acknowledgements

Any word would be an understatement to acknowledge the immense role of my supervisor,Professor B.C. Sutradhar, during my Ph.D. program and towards the completion of my thesis. With this, I would like to express my heartiest gratitude to my supervisor, Prof. B.C. Sutradhar for supervision, advice, encouragement, enduring patience and constant support through out the program. He has always been there when I was stuck, constantly encouraging me do well. It was undoubtedly a great privilege for me to work under his supervision for which I shall forever be grateful.

I would like to thank Dr. Zhao Zhi Fan and Dr. Asokan Variyath of Department of Mathematics & Statistics, Memorial University of Newfoundland for taking the time to serve as the member of the supervisory committee of this thesis and also would like to thank all the members of the examination committee. Their constructive comments and suggestions served to improve the quality of the thesis.

I want to sincerely acknowledge the financial support provided by the School of Graduate Studies, Department of Mathematics and Statistics, and Prof. B.C. Sutradhar in the form of Graduate Assistantship & Teaching Assistantships. Further, I wish to thank the Department, the lovely and helpful staff at our General Office for providing a very friendly atmosphere, nice conversations and the necessary facilities, especially the stationary material for which they have been very generous to me.

It is my great pleasure to thank all my friends and well wishers who directly or indirectly encouraged and helped me in the whole period of my study and day to day life.

Last and most importantly, I am utterly grateful to my parents, my brother and sisters, and specially my wife & my lovely daughter, for their strong belief in my capabilities and for all of their eternal love, support and encouragement in my life.

I'm ever grateful to my parents for what I'm now,

also, my wife, Shuhana, and my lovely daughter, Sheza,

for the heavenly home

Contents

List of Tables

Chapter 1

Introduction

1.1 Motivation

In practice there are many situations where categorical responses with more than two categories along with information on multidimensional covariates are collected from a large number of independent individuals over a small period of time. In such a situation, it is likely that repeated categorical responses will be correlated. However, there does not appear to be adequate discussions on the analysis of such data mainly because of the difficulties of modeling longitudinal correlations for multinomial responses. For some studies either using time as a fixed covariate or modeling association through 'working' equi-correlation or independence and/or adhoc transition probabilities, we refer to Conaway (1989), Fienberg et al. (1985), Agresti (1990, 2002), Lipsitz et al. (1994), Stram et al. (1988), and Li and Chan (2006). The purpose of this thesis is to develop suitable correlation models and provide inferences for the regression parameters under categories by taking the longitudinal correlations into account.

As a motivation for the longitudinal multinomial data modeling, in this chapter, we first briefly demonstrate how multinomial regression analysis are usually done in the independent setup. Next, we provide some historical development on longitudinal binary modeling before we consider the multinomial generalization.

1.1.1 Multinomial Model in Independent Set-up

In the independent setup, there exist many analysis using multinomial models for univariate categorical responses at cross-sectional level. For example, in the 'Aspirin and Heart Attacks' problem discussed by Agresti (1990, Table 2.3, page 17) the status of heart attack, namely fatal attack, non-fatal attack and no attack, were recorded from 22071 independent individuals along with covariate information on whether the individual had aspirin or not during a clinical trial period. Here it is of interest to understand the effect of aspirin on heart attack status. Similarly, in socio-economic field, one may be interested in studying the effect of gender or say educational level on the categorical response variable, namely jobless spell in a given year.

Let

$$\mathbf{Y}_i = \left(y_{i1}, \cdots, y_{ij}, \cdots, y_{iK}\right)' \tag{1.1}$$

be a univariate K-dimensional nominal categorical response variable for the ith ($i = 1, \cdots, I$) individual. Here the K-dimensional variable implies that the response of the ith individual belongs to one of the K+1 categories. Suppose that a p-dimensional covariate vector $\mathbf{x}_i = \left(x_{i(1)}, \cdots, x_{i(p)}\right)'$ is recorded along with the response y_i from the ith individual, and all I individuals are independent. Since the response of the ith individual can belong to one of $K+1$ categories, we denote the jth ($j = 1, \cdots, K$) category response of the ith individual by

$$\begin{aligned} \mathbf{y}_i^{(j)} &= \left(y_{i1}^{(j)}, \cdots, y_{ij}^{(j)}, \cdots, y_{iK}^{(j)}\right)' \\ &= \left(0\mathbf{1}'_{j-1}, 1, 0\mathbf{1}'_{K-j}\right)', \end{aligned} \tag{1.2}$$

i.e $y_{ij}^{(j)} = 1$ and $y_{ij}^{(c)} = 0$ for $j \neq c$,

so that the response in the last category can be identified by

$$\mathbf{y}_i^{(K+1)} = (0, \cdots, 0, \cdots, 0)' = 0\,\mathbf{1}'_K.$$

Suppose that $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{jp})'$ denotes the effect of $\mathbf{x}_i$ on $\mathbf{y}_i^{(j)}$ for $j = 1, \cdots, K$ with $\boldsymbol{\beta}_{K+1} = (0, \cdots, 0)'$ by convention. Also suppose that

$$\begin{aligned} P\Big(\mathbf{Y_i} = \mathbf{y_i^{(j)}}\Big) &= \frac{\exp\left(\beta_{j0} + \sum_{u=1}^{p} \beta_{ju} x_{i(u)}\right)}{\sum_{c=1}^{K+1} \exp\left(\beta_{c0} + \sum_{u=1}^{p} \beta_{cu} x_{i(u)}\right)} \\ &= \frac{\exp\left(\beta_{j0} + \mathbf{x}_i'\boldsymbol{\beta}_j\right)}{1 + \sum_{c=1}^{K} \exp\left(\beta_{c0} + \mathbf{x}_i'\boldsymbol{\beta}_c\right)} \\ &= \pi_i^{(j)}, \text{ say, } \mathrm{i} = 1, \cdots, \mathrm{I};\ \mathrm{j} = 1, \cdots, \mathrm{K}; \end{aligned} \tag{1.3}$$

denote the probability that the response from the ith ($i = 1, \cdots, I$) individual belong to the jth ($j = 1, \cdots, K$) category. In (1.3), β_{j0} denotes an intercept parameter under the jth category. One may then obtain the likelihood estimate of

$$\boldsymbol{\mu} = \left(\beta_{10}, \boldsymbol{\beta}_1', \cdots, \beta_{j0}, \boldsymbol{\beta}_j', \cdots, \beta_{K0}, \boldsymbol{\beta}_K'\right)' \tag{1.4}$$

by maximizing the multinomial likelihood function given by

$$L\left(\boldsymbol{\mu} \,|\, X_1', \cdots, X_I'\right) = \prod_{i=1}^{I} \prod_{j=1}^{K+1} \frac{1! \times \left\{\pi_i^{(j)}\right\}^{y_{ij}}}{y_{ij}!} \tag{1.5}$$

where $y_{i,K+1} = \left(1 - \sum_{j=1}^{K} y_{ij}\right)$ and $\pi_i^{(K+1)} = \left(1 - \sum_{j=1}^{K} \pi_i^{(j)}\right)$; which is equivalent to solving the log likelihood estimating equation for $\boldsymbol{\beta}_j^* = \left(\beta_{j0}, \boldsymbol{\beta}_j'\right)'$

$$\frac{\partial \ln L(\boldsymbol{\mu})}{\partial \boldsymbol{\beta}_j^*} = \frac{\partial}{\partial \boldsymbol{\beta}_j^*}\left[C + \sum_{i=1}^{I}\sum_{l=1}^{K} y_{ij} \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix}' \boldsymbol{\beta}_l^* - \sum_{i=1}^{I} \ln\left\{1 + \sum_{l=1}^{K} \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix}' \boldsymbol{\beta}_l^*\right\}\right]$$

$$= \sum_{i=1}^{I} \left[\begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} y_{ij} - \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} \pi_{ij}^{(j)} \right]$$

$$= \sum_{i=1}^{I} \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} \left[y_{ij} - \pi_i^{(j)} \right], \tag{1.6}$$

leading to

$$\frac{\partial \ln L(\boldsymbol{\mu})}{\partial \boldsymbol{\mu}} = \sum_{i=1}^{I} \left[\mathbf{I}_K \otimes \begin{pmatrix} 1 \\ \mathbf{x}_i \end{pmatrix} \right] \left[\mathbf{y}_i - \boldsymbol{\pi}_i \right] \tag{1.7}$$

where $\boldsymbol{\pi}_i = \left(\pi_i^{(1)}, \cdots, \pi_i^{(K)}\right)'$ corresponding to $\mathbf{y}_i$; $\mathbf{x}_i$ is the $p \times 1$ design vector defined as $\mathbf{x}_i = \left(x_{i(1)}, \cdots, x_{i(p)}\right)'$ and I_K is the identity matrix of order K.

Aspirin and Heart Attacks Data Example: An Illustration

Recall that in the 'Aspirin and Heart Attacks' example, $I = 22071$ individuals were studied (Agresti, 1990) to understand the effect of aspirin on heart attack status. For convenience we display this data set in the Table 1.1.

In ordinal multinomial study, when cell observations are small under a given category, it is standard to combine tow such adjacent categories and deal with a inferences based on a lesser number of categories. However, in the nominal study, this type of merging does not make sense, where it is standard to assume that the

Table 1.1: *Cross-Classification of Aspirin Use and Myocardial Infarction.*

	Myocardial Infarction			
	Fatal Attack	Non-Fatal Attack	No Attack	Total
Placebo	18	171	10,845	11,034
Aspirin	5	99	10,933	11,037
Total	23	270	21,778	22,071

cell frequencies are reasonably large.

For $p = 1$, by using

$$x_{i(1)} = \begin{cases} 1 & \text{for aspirin taken by the } i\text{th individual} \\ 0 & \text{otherwise} \end{cases}$$

and multinomial response

$$\mathbf{y}_i = \begin{cases} (1,\ 0)' & \text{when ith individual had fatal attack} \\ (0,\ 1)' & \text{when ith individual had non fatal attack} \\ (0,\ 0)' & \text{when ith individual had no attack,} \end{cases}$$

it is of interest to estimate the multinomial probabilities $\pi_i^{(j)}$; for $i = 1, \cdots, I$; $j = 1, \cdots, K$. Note that the aforementioned values for $\mathbf{y}_i$ have been assigned by treating the Myocardia infarction status as nominal, whereas it is more appropriate to consider these status as ordinal. However, in the present thesis we develop the longitudinal models for nominal multinomial variable. Thus, a detailed discussion on ordinal multinomial case will be beyond the scope of the present thesis. As far as this example is concerned, we are using this data for the illustration of nominal multinomial model only.

Turning back to the estimation of the multinomial probabilities, we solve the likelihood equation (1.7) and obtain the estimates for the marginal category effect (intercept) and regression parameters (1.4), namely $\boldsymbol{\mu} = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})'$. Note that as far as the interpretation of these parameters are concerned, in absence of any covariates, two intercept parameters β_{10} and β_{20} will reflect the effect of 'Fatal attack' (category 1) and 'Non-fatal attack' (category 2) as compared to 'No attack' (category 3) on the corresponding multinomial probability, because of the fact that 'No attack' has been considered as the reference category. Similarly, β_{11} and β_{21} will reflect the effect of aspirin on the heart attack status to be in 'Fatal attack' and 'Non-fatal attack' category, respectively, as compared to the 'No attack' category. Thus, the likelihood estimates of the parameters are

$$\begin{aligned}\hat{\boldsymbol{\mu}} &= \left(\hat{\beta'_{10}}, \hat{\beta'_1}, \hat{\beta'_{20}}, \hat{\beta'_2}\right)' = \left(\hat{\beta_{10}}, \hat{\beta}_{11}, \hat{\beta_{20}}, \hat{\beta}_{21}\right)' \\ &= (-6.401, -1.289, -4.15, -0.555)' \qquad (1.8)\end{aligned}$$

with corresponding standard errors; $s.e\,(\hat{\beta}_{10}) = 0.2360$, $s.e\,(\hat{\beta}_{11}) = 0.5057$, $s.e\,(\hat{\beta}_{20}) = 0.0771$ and $s.e\,(\hat{\beta}_{21}) = 0.1270$. These estimates lead to the multinomial probabilities as in Table 1.2.

Table 1.2: *Observed and Estimated Multinomial Probabilities Corresponding to Table 1.1.*

		Myocardial Infarction			
	Proportion/ Probability	Fatal Attack	Non-Fatal Attack	No Attack	Total
Placebo	Observed	0.00163	0.01550	0.98287	1.00
	Estimated	0.00164	0.01562	0.98274	1.00
Aspirin	Observed	0.00045	0.00897	0.99058	1.00
	Estimated	0.00045	0.00897	0.99058	1.00

In Table 1.2, we have also displayed the observed probabilities. For example, the observed proportion of individuals whose heart attack was either fatal or non-fatal is shown to be (5+99)/11,037 = 0.00942 for the aspirin group and (18+171)/11,034 = 0.01713 for the placebo group. These are also available in Agresti (1990, Section 2.2.4, page 17) where the author has exploited these to compute the relative risk of heart attack as

$$\frac{(18+171)/11,034}{(5+99)/11,037} = \frac{0.01713}{0.00942} = 1.82.$$

Note that because of the availability of the estimated probabilities as in Table 1.2, we may now compute the estimated relative risk of heart attack which is given by

$$\frac{0.01726}{0.00942} = 1.83,$$

which agrees well with the observed relative risk. Thus based on the estimated relative risk, unlike using the observed relative risk used by Agresti (1990), we can infer that the proportion of individuals suffering heart attack was 1.83 times higher for patients taking placebo than for the patients taking aspirin.

Note that the aforementioned multinomial model (1.3) and the inferences (1.7) for this model are described for a cross-sectional study, where the multinomial response with corresponding covariates are collected from a large number of independent individuals at a single point of time. There are, however, situations in practice where this type of multinomial responses are collected over a small period of time. But, the modeling and inferences for such repeated multinomial data are not addressed adequately in the literature. As oppose to the multinomial case ($K > 1$) there, however, exists some studies in the longitudinal binary ($K = 1$) set up. For example, we refer to the binary logit models involving time as a fixed covariate considered by Agresti [1990 (Chapter 11, page 395) and 1997], conditional linear binary probability models discussed by Sutradhar (2010) and a conditional non-linear binary dynamic models suggested by Sutradhar and Farrell (2007). For convenience, we briefly discuss these longitudinal binary models in the following sections.

1.1.2 Fixed Binary Models in Longitudinal Set-up

1.1.2.1 Binary Longitudinal Models with Time as Fixed Covariate

(a) Individual History Based Model

Note that in a cross-sectional binary set up, when the history of every individual's covariates (common for continuous covariates) are available, it is standard to use a binary logistic model to fit such data. This model may be written as a special case of the multinomial probability model (1.3), and is given by

$$\pi_i^{(1)} = P(Y_i = 1) = \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{i(u)}\,\beta_u\right)}{1 + \exp\left(\beta_0 + \sum_{u=1}^{p} x_{i(u)}\,\beta_u\right)}, \tag{1.9}$$

and

$$\pi_i^{(2)} = 1 - \pi_i^{(1)}; \qquad \text{where i} = 1, \cdots, \text{I}.$$

In a longitudinal set up, the binary responses are collected from all I individuals over a small period of time T. By considering time as a fixed covariate with T different levels and assigning I individuals belong to $p+1$ groups, Agresti (1990, page 396) has used a model to accommodate the time effect on the response probability (see 1.15) in contingency table form. In the history based setup, one can write a general model as

$$\pi_{it}^{(1)} = P\left(Y_{it} = 1 \,|\, x_{it(1)}, \cdots, x_{it(p)}\right) = \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\,\beta_u \,+\, \lambda_t\right)}{1 \,+\, \exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\,\beta_u \,+\, \lambda_t\right)}, \quad (1.10)$$

with $\pi_{it}^{(2)} = 1 - \pi_{it}^{(1)}$, where $i = 1, \cdots, I$ and $t = 1, \cdots, T$. In (1.10), similar to (1.15), $\lambda_T = 0$ is considered. Alternatively, one can consider the restriction $\sum_{t=1}^{T} \lambda_t = 0$, so that $T-1$ effects are independent. Note that this model (1.10) can be considered as a generalization of model (1.9).

(b) Contingency Tables Based Model

Suppose that in model (1.9), there are I_u fixed individuals with a common covariate x_{iu} for $u = 1, \cdots, p+1$. Also suppose that $p+1$ levels are identified as follows,

$$\left(x_{i(1)}, \cdots, x_{i(p)}\right) \equiv \begin{cases} (\,1,\ 0,\ \cdots,\ 0\,) & \longrightarrow \text{ Level 1} \\ (\,0,\ 1,\ \cdots,\ 0\,) & \longrightarrow \text{ Level 2} \\ (\cdots\cdots\cdots\cdots) & \\ (\,0,\ 0,\ \cdots,\ 1\,) & \longrightarrow \text{ Level p} \\ (\,0,\ 0,\ \cdots,\ 0\,) & \longrightarrow \text{ Level p} + 1 \end{cases} \quad (1.11)$$

In the cross-sectional set up, one may then write a contingency table of the form:

Table 1.3: *Contingency Table in the Cross-sectional Set up (T = 1) for Binary Responses.*

Level	1	2	Total
1	$I_{(1)}^{(1)}$	$I_{(1)}^{(2)}$	$I_{(1)}$
2	$I_{(2)}^{(1)}$	$I_{(2)}^{(2)}$	$I_{(2)}$
.	.	.	.
.	.	.	.
u	$I_{(u)}^{(1)}$	$I_{(u)}^{(2)}$	$I_{(u)}$
.	.	.	.
.	.	.	.
p + 1	$I_{(p+1)}^{(1)}$	$I_{(p+1)}^{(1)}$	$I_{(p+1)}$
Total	$I_{(.)}^{(1)}$	$I_{(.)}^{(1)}$	I

Suppose that for the individuals with covariate level u (u = 1, $\cdots$, p +1), the probability that the response of an individual in this group belongs to the first category is denoted by $\pi_{(u)}^{(1)}$. Then, by using (1.11) and (1.9), the probability that the response of the ith individual with uth level of the covariate belongs to this first category can be expressed as

$$\begin{aligned} \pi_{(u)}^{(1)} &= P\Big(Y_i = 1 \,\big|\, x_{i(1)}, \cdots, x_{i(p)}\,;\, i \in I_{(u)}\Big) \\ &= \begin{cases} \dfrac{\exp(\beta_0 + \beta_u)}{1 + \exp(\beta_0 + \beta_u)}; & \text{for} \quad u = 1, \cdots, p \\ \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)}; & \text{for} \quad u = p+1. \end{cases} \end{aligned} \tag{1.12}$$

and $\pi_{(u)}^{(2)} = 1 - \pi_{(u)}^{(1)}$. Note that in writing the model (1.12), we however used $\beta_{p+1} = 0$ without any loss of generality. Now, the parameters in (1.12) may be estimated by maximizing the product binomial likelihood given by

$$L = \prod_{u=1}^{p+1} L_{(u)} \tag{1.13}$$

where

$$L_{(u)} = \binom{I_{(u)}}{I^{(1)}_{(u)}} \left\{\pi^{(1)}_{(u)}\right\}^{I^{(1)}_{(u)}} \left\{\pi^{(2)}_{(u)}\right\}^{I^{(2)}_{(u)}},$$

with $I^{(2)}_{(u)} = I_{(u)} - I^{(1)}_{(u)}$, where $I_{(u)}$ is fixed.

When the history of response is known, in the binary longitudinal set up with time period T, one can construct a contingency table of dimension $(p+1) \times 2^T$ as a generalization of Table 1.3. We display this form in Table 1.4 for convenience for $T = 3$. This table has similar structure as that of the contingency Table 11.2 in Agresti (2002, sec.11.2.1, p.459) which was constructed for a cross classification of responses on depression at three times by diagnosis and treatment.

In Table 1.4, $I_{111(u)}$, for example, indicates the number of individuals out of $I_{(u)}$ (total number at covariate level u) who responded under category 1 at all three time points. Let

$$I_{111(u)} + I_{112(u)} + I_{121(u)} + I_{122(u)} = I^{(1)}_{(u,1)}$$

Table 1.4: *Contingency Table in the Longitudinal Setup Over $T = 3$ Periods for Binary Responses.*

	Response at three times								
Level	111	112	121	122	211	112	221	222	Total
1	$I_{111(1)}$	$I_{112(1)}$	$I_{121(1)}$	$I_{122(1)}$	$I_{211(1)}$	$I_{212(1)}$	$I_{221(1)}$	$I_{222(1)}$	$I_{(1)}$
2	$I_{111(2)}$	$I_{112(2)}$	$I_{121(2)}$	$I_{122(2)}$	$I_{211(2)}$	$I_{212(2)}$	$I_{221(2)}$	$I_{222(2)}$	$I_{(2)}$
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
u	$I_{111(u)}$	$I_{112(u)}$	$I_{121(u)}$	$I_{122(u)}$	$I_{211(u)}$	$I_{212(u)}$	$I_{221(u)}$	$I_{222(u)}$	$I_{(u)}$
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
p+1	$I_{111(p+1)}$	$I_{112(p+1)}$	$I_{121(p+1)}$	$I_{122(p+1)}$	$I_{211(p+1)}$	$I_{212(p+1)}$	$I_{221(p+1)}$	$I_{222(p+1)}$	$I_{(p+1)}$

denote the total number of individuals with uth level of covariate who responded in category (1) at time point 1. Similarly

$$I_{211(u)} + I_{212(u)} + I_{221(u)} + I_{222(u)} = I^{(2)}_{(u,1)}$$

represents the total number of individuals with uth level of covariate who responded in category (2) at time point 1. Thus,

$$I^{(1)}_{(u,1)} + I^{(2)}_{(u,1)} = I_{(u,1)}$$

is the total number of individuals at uth level of covariate responded at time $t = 1$. In general, for any t we can write

$$I^{(1)}_{(u,t)} + I^{(2)}_{(u,t)} = I_{(u,t)} \tag{1.14}$$

as the total number of individuals at uth level of covariate responded at time t. In fact, if there is no missing, then

$$I_{(u,1)} = \cdots = I_{(u,t)} = \cdots = I_{(u,T)} = I_{(u)}, \text{ say.}$$

Now to analyze the data in Table 1.4 and other similar data, the existing studies such as Agresti (1990) used the marginal probability model at a given time point t. For an individual i belonging to the group of $I_{(u,t)} = I_{(u)}$ individuals, i.e. $i \in I_{(u,t)}$ $(= I_{(u)})$, by using (1.11), these marginal probabilities at time point t given by (1.10) may be written as

$$\begin{aligned} \pi^{(1)}_{(u,t)} &= P\left(Y_{it} = 1 \,|\, x_{it1}, \cdots, x_{itp}; i \in I_{(u)}\right) \\ &= \begin{cases} \dfrac{\exp(\beta_0 + \beta_u + \lambda_t)}{1 + \exp(\beta_0 + \beta_u + \lambda_t)}; & \text{for } u = 1, \cdots, p; \\ & t = 1, \cdots, T-1 \\ \dfrac{\exp(\beta_0 + \lambda_t)}{1 + \exp(\beta_0 + \lambda_t)}; & \text{for } u = p+1; \\ & t = 1, \cdots, T-1 \\ \dfrac{\exp(\beta_0 + \beta_u)}{1 + \exp(\beta_0 + \beta_u)}; & \text{for } u = 1, \cdots, p; \; t = T \end{cases} \end{aligned} \qquad (1.15)$$

and $\pi^{(2)}_{(u,t)} = 1 - \pi^{(1)}_{(u,t)}$.

Note that by treating the time variable as a fixed covariate, the binary likelihood in the longitudinal set up can be written by exploiting the marginal probability (1.15) and the tabular counts from Table 1.4. To be specific, the product binomial likelihood function has the form

$$L = \prod_{u=1}^{p+1} \prod_{t=1}^{T} L_{(u,t)} \tag{1.16}$$

with

$$L_{(u,t)} = \binom{I_{(u,t)}}{I_{(u,t)}^{(1)}} \left\{\pi_{(u,t)}^{(1)}\right\}^{I_{(u,t)}^{(1)}} \left\{\pi_{(u,t)}^{(2)}\right\}^{I_{(u,t)}^{(2)}}$$

where $\pi_{(u,t)}^{(1)}$ and $\pi_{(u,t)}^{(2)}$ are as in (1.15), $I_{(u,t)}^{(1)}$, $I_{(u,t)}^{(2)}$ and $I_{(u,t)}$ for all u and t are as in Table 1.4. This likelihood can be maximized with respect to the desired parameters β_0, β_1, ..., β_p, λ_1, ..., λ_{T-1}. This likelihood analysis by treating time as a fixed covariate is similar to that of Agresti (2002, Sec. 11.2.1, p. 459-461).

In practice, it is, however, not sensible to treat the time factor as a fixed covariate. Suppose that time is an index variable and λ_t in (1.15) indicates the marginal effect of time t on the binary response y_{it}. Under this set up, logit from (1.15) has the linear form

$$\text{logit}\,(\beta_{\text{u}}, \lambda_{\text{t}}) = \beta_0 + \beta_{\text{u}} + \lambda_{\text{t}} \tag{1.17}$$

which is the same as Equation (11.8) in Agresti (1990, p.396). Note however that the binary responses $y_{i1}, \cdots, y_{it}, \cdots, y_{iT}$ are supposed to be correlated as they are collected from the same ith individual over T time points. For example, in Table 1.4, $I_{111(u)}$ binary responses at time points $t = 1, 2 \text{ and } 3$, at category 1 with uth level of covariate are correlated. It is necessary to write a joint probability function in such a set up or use the conditional (correlated) probability models for proper likelihood analysis. This observation that the marginal likelihood analysis cannot be used in such a longitudinal set up, was also pointed out by Agresti (1990, Sec.11.3.1, p. 395-396). In this thesis, as opposed to the marginal analysis, we propose two correlation models for longitudinal multinomial data, namely, a conditional linear probability based model in Chapter 2, and a non-linear logistic probability based model in Chapter 3, and use conditional cell probabilities to develop a likelihood estimation approach. This correlation modeling approach for the longitudinal binary data has been recently discussed in the literature [Sutradhar and Farrell (2007), Sutradhar (2010)]. We review these models in the next sections for convenience of their generalization to the multinomial cases.

1.1.2.2 Conditional Linear Binary Longitudinal Models with Time as an Index variable

(a) Individual History Based Model

Sutradhar (2010, Section 2.2) has proposed a general non-stationary auto correlations structure based longitudinal binary probability model. For the stationary AR(1) case, the conditional linear probability model from Sutradhar (2010) [see also Qaqish (2003)] may be written as

$$P(Y_{i1} = 1) = \pi_i^{(1)} \quad (1.18)$$

$$P(Y_{it} = 1 \mid Y_{i,t-1} = y_{i,t-1}) = \eta_{it|t-1}^{(1)} = \pi_i^{(1)} + \rho\left(y_{i,t-1} - \pi_i^{(1)}\right); \; t = 2, \cdots, T; \quad (1.19)$$

with

$$\pi_i^{(1)} = \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{i(u)}\beta_u\right)}{1 + \exp\left(\beta_0 + \sum_{u=1}^{p} x_{i(u)}\beta_u\right)}; \quad \text{for } t = 1, ..., T.$$

and $\pi_i^{(2)} = 1 - \pi_i^{(1)}$.

It can be shown that the means and variances of this model are

$$E(Y_{it}) = \pi_i^{(1)} \text{ and } \mathrm{Var}(Y_{it}) = \pi_i^{(1)}[1 - \pi_i^{(1)}] = \pi_i^{(1)}\pi_i^{(2)}; \quad \text{for } t = 1, ..., T.$$

Also, the stationary correlation between Y_{im} and Y_{it} for all $m, t = 1, \cdots, T$, is given by

$$\text{corr}(Y_{im}, Y_{it}) = \rho^{|t-m|} \quad \text{for all } m \text{ and } t. \tag{1.20}$$

where ρ parameter in (1.20) must satisfy the range restriction

$$\max\left[-\frac{\pi_i^{(1)}}{1-\pi_i^{(1)}}, -\frac{1-\pi_i^{(1)}}{\pi_i^{(1)}}\right] \leq \rho \leq 1$$

For the purpose of estimation purpose, one can write the following likelihood function by exploiting the marginal and conditional binary probabilities from model (1.18) and (1.19),

$$\mathrm{L} = \prod_{i=1}^{I}\left[f(y_{i1})\prod_{t=2}^{T} f(y_{it} \mid y_{i,t-1})\right] \tag{1.21}$$

where

$$f(y_{i1}) = \left\{\pi_i^{(1)}\right\}^{y_{i1}}\left\{1 - \pi_i^{(1)}\right\}^{1-y_{i1}}$$

and

$$f(y_{it} \mid y_{i,t-1}) = \left\{\eta_{it|t-1}^{(1)}\right\}^{y_{it}}\left\{1 - \eta_{it|t-1}^{(1)}\right\}^{1-y_{it}}; \qquad \text{for t} = 2, \cdots, \text{T}.$$

Then, for $T = 3$, by using (1.13), the log likelihood function for the conditional counts in Table 4 can be written as

$$\ln L(\beta_0, \beta_1, \cdots, \beta_p, \rho) = \sum_{u=1}^{p+1} \left[I_{(u,1)}^{(1)} \ln \left\{ \pi_{(u,1)}^{(1)} \right\} + I_{(u,1)}^{(2)} \ln \left\{ \pi_{(u,1)}^{(2)} \right\} \right] + \sum_{u=1}^{p+1} \sum_{t=2}^{T} \left[I_{(u,t)}^{(1)} \ln \left\{ \eta_{(u,t)}^{(1)} \right\} + I_{(u,t)}^{(2)} \ln \left\{ \eta_{(u,t)}^{(2)} \right\} \right] \quad (1.22)$$

where, by using the covariate levels as in (1.11),

$$\pi_{(u,t)}^{(1)} = \begin{cases} \dfrac{\exp(\beta_0 + \beta_u)}{1 + \exp(\beta_0 + \beta_u)}; & \text{for } u = 1, \cdots, p; \ t = 1, \cdots, T \\ \dfrac{\exp(\beta_0)}{1 + \exp(\beta_0)}; & \text{for } u = p+1; \ t = 1, \cdots, T \end{cases} \quad (1.23)$$

with $\pi_{(u,t)}^{(2)} = 1 - \pi_{(u,t)}^{(1)}$ and

$$\eta_{(u,t|t-1)}^{(1)} = \pi_{(u,t)}^{(1)} + \rho \left\{ y_{i,t-1} - \pi_{(u,t-1)}^{(1)} \right\}; \text{ for } u = 1, \cdots, p; +1 \ t = 2, \cdots, T \quad (1.24)$$

with $\eta_{(u,t|t-1)}^{(2)} = 1 - \eta_{(u,t|t-1)}^{(1)}$.

In this binary setup, for known ρ ; $\beta_0, \beta_1, \cdots, \beta_p$ can be estimated by exploiting the log likelihood function (1.22). As far as ρ is concerned, ρ is estimated by method of moment by checking its range restriction.

(b) Conditional Contingency Tables Based Model (Without any Covariate)

It should be clear from (a) that the likelihood function (1.21) was constructed by collapsing the history based counts from Table 1.4. However, in practice, the longitudinal binary data collection can be easier when only the counts are recorded at a given time t based on the outcomes from $t-1$ time only. Thus, the counts may not be available in the history based form of Table 1.4. To reflect this conditional data collection mechanism we display a format for the conditional counts as in Table 1.5.

In Table 1.5(a) and 1.5(b), $I^{(1)}_{(t-1)}$, for example, refers to the number of individuals who responded for category 1 at time point $t-1$ $(t = 2, \cdots, T)$. The cell counts in Table 1.5(b) at time t are conditional on the response category at time $t-1$. For example, $I_{11(t|t-1)}$ indicates the number of individuals who responded for category 1 at time t, given that these individuals also responded for category 1 at time $t-1$.

Note that the cell probabilities corresponding to the counts in Table 1.5 may be written from (1.23) and (1.24) by using $\beta_u = 0$ $(u = 1, \cdots, p)$ without any loss of generality. In this case, the probabilities in (1.23) and (1.24) were written for a model involving covariates, whereas Table 1.5 (a) and (b) are constructed for situation without any covariates. Thus, corresponding to the cell counts in Table 1.5, we write

Table 1.5: *Conditional Contingency Tables in the Longitudinal Setup for Binary Response Model.*

(a) For any time t

Category	1	2	Total
Counts	$I^{(1)}_{(t)}$	$I^{(2)}_{(t)}$	$I_{(t)} = I$

(b) Time, t − 1 (rows); Time, t (columns); for t = 2, ⋯ , T

Category	1	2	Total
1	$I_{11(t\|t-1)}$	$I_{12(t\|t-1)}$	$I^{(1)}_{(t-1)}$
2	$I_{21(t\|t-1)}$	$I_{22(t\|t-1)}$	$I^{(2)}_{(t-1)}$
Total	$I^{(1)}_{(t)}$	$I^{(2)}_{(t)}$	I

$$\pi^{(1)} = \frac{\exp(\beta_0)}{1 + \exp(\beta_0)}; \quad \text{for } t = 1, \cdots, T \tag{1.25}$$

with $\pi^{(2)} = 1 - \pi^{(1)}$ and

$$\eta^{(1)}_{(t|t-1)}(1) = \pi^{(1)} + \rho\{1 - \pi^{(1)}\}; \quad \text{for } t = 2, \cdots, T \tag{1.26}$$

$$\eta^{(1)}_{(t|t-1)}(0) = \pi^{(1)} - \rho\,\pi^{(1)} = \pi^{(1)}(1-\rho); \quad \text{for } t = 2, \cdots, T \tag{1.27}$$

with $\eta^{(2)}_{(t|t-1)}(1) = 1 - \eta^{(1)}_{(t|t-1)}(1)$ and $\eta^{(2)}_{(t|t-1)}(0) = 1 - \eta^{(1)}_{(t|t-1)}(0)$.

Now by using (1.26) and (1.27), one can write a product binomial likelihood for the cell counts in Table 1.5. For example, for $T = 3$ we can write the product binomial likelihood function as

$$L = f\left(I_{(1)}^{(1)}\right) f\left(I_{11(2|1)}, I_{21(2|1)} \mid I_{(1)}^{(1)}, I_{(1)}^{(2)}\right) f\left(I_{11(3|2)}, I_{21(3|2)} \mid I_{(2)}^{(1)}, I_{(2)}^{(2)}\right) \quad (1.28)$$

where

$$f\left(I_{(1)}^{(1)}\right) = \{\pi^{(1)}\}^{I_{(1)}^{(1)}} \{1 - \pi^{(1)}\}^{I_{(1)}^{(2)}} \quad (1.29)$$

$$\begin{aligned} f\left(I_{11(2|1)}, I_{21(2|1)} \mid I_{(1)}^{(1)}, I_{(1)}^{(2)}\right) &= \binom{I_{(1)}^{(1)}}{I_{11(2|1)}} \left\{\eta_{2|1}^{(1)}(1)\right\}^{I_{11(2|1)}} \left\{1 - \eta_{2|1}^{(1)}(1)\right\}^{I_{12(2|1)}} \\ &\times \binom{I_{(1)}^{(2)}}{I_{21(2|1)}} \left\{\eta_{2|1}^{(1)}(0)\right\}^{I_{21(2|1)}} \left\{1 - \eta_{2|1}^{(1)}(0)\right\}^{I_{22(2|1)}} \quad (1.30) \end{aligned}$$

and

$$\begin{aligned} f\left(I_{11(3|2)}, I_{21(3|2)} \mid I_{(2)}^{(1)}, I_{(2)}^{(2)}\right) &= \binom{I_{(2)}^{(1)}}{I_{11(3|2)}} \left\{\eta_{3|2}^{(1)}(1)\right\}^{I_{11(3|2)}} \left\{1 - \eta_{3|2}^{(1)}(1)\right\}^{I_{12(3|2)}} \\ &\times \binom{I_{(2)}^{(2)}}{I_{21(3|2)}} \left\{\eta_{3|2}^{(1)}(0)\right\}^{I_{21(3|2)}} \left\{1 - \eta_{3|2}^{(1)}(0)\right\}^{I_{22(3|2)}} \quad (1.31) \end{aligned}$$

Now, the estimation of the parameters β_0 and ρ can be done similar to the history based approach.

1.1.2.3 Conditional Non-Linear Binary Longitudinal Models with Time as an Index variable

(a) Individual History Based Model

As opposed to the linear models discussed in the previous sections, to analyze binary longitudinal data, Sutradhar and Farrell (2007) have used a non-linear binary dynamic model [see also, Amemiya, 1985, p. 422 and Manski, 1987];

$$\mu_{i1}^{*(1)} = P(Y_{i1} = 1) = \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{i1(u)}\beta_u\right)}{1 + \exp\left(\beta_0 + \sum_{u=1}^{p} x_{i1(u)}\beta_u\right)}, \tag{1.32}$$

$$\begin{aligned} \pi_{it|t-1}^{*(1)} &= P(Y_{it} = 1|\, y_{i,t-1}) \\ &= \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\beta_u + \gamma\, y_{i,t-1}\right)}{1 + \exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\beta_u + \gamma\, y_{i,t-1}\right)}, \end{aligned} \tag{1.33}$$

for $i = 1, \cdots, I$; $t = 2, \cdots, T$ where, β is the regression parameter of x on y, and γ is the dynamic dependence parameter.

For $t = 2, \cdots, T$, let

$$\begin{aligned} \mu^*_{it} &= \pi^{*(1)}_{it|t-1}|_{y_{i,t-1}=0} \\ &= \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\beta_u\right)}{1 + \exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\beta_u\right)} \end{aligned} \tag{1.34}$$

$$\begin{aligned} \tilde{\mu}_{it} &= \pi^{*(1)}_{it|t-1}|_{y_{i,t-1}=1} \\ &= \frac{\exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\beta_u + \gamma\right)}{1 + \exp\left(\beta_0 + \sum_{u=1}^{p} x_{it(u)}\beta_u + \gamma\right)} \end{aligned} \tag{1.35}$$

Further let μ_{it} denotes the unconditional expectation of y_{it} for all $t = 1, \cdots, T$. In general by using (1.34) and (1.35), it then follows that μ_{it} maintains a recursive relationship given by

$$\mu^{(1)}_{i1} = \mu^{*(1)}_{i1} \tag{1.36}$$

$$\begin{aligned} \mu^{(1)}_{it} &= E(Y_{it}) = P(Y_{it} = 1) \\ &= \mu^*_{it} + \mu_{i,t-1}\,(\tilde{\mu}_{it} - \mu^*_{it}); \quad t = 2, \cdots, T \end{aligned} \tag{1.37}$$

and

$$Var(Y_{it}) = \sigma_{itt} = \mu^{(1)}_{it}(1 - \mu^{(1)}_{it}); \quad t = 1, \cdots, T. \tag{1.38}$$

Next, following Sutradhar and Farrell [2007, equation (1.6)], we write the unconditional correlation between y_{im} and y_{it} as

$$Corr\,(y_{im},\, y_{it}) \;=\; \sqrt{\frac{\mu_{im}^{(1)}(1-\mu_{im}^{(1)})}{\mu_{it}^{(1)}(1-\mu_{it}^{(1)})}}\;\prod_{j=m+1}^{t}\,(\tilde{\mu}_{ij}-\mu_{ij}^{*})\,; \quad m<t. \qquad (1.39)$$

Note that this correlation in (1.39) ranges from -1 to 1, as $0 < \tilde{\mu}_{ij},\, \mu_{ij}^{*} < 1$.

As far as the inferences for β and γ are concerned, they may be estimated by maximizing the likelihood function given by

$$L = \prod_{i=1}^{I}\left[f(y_{i1})\prod_{t=2}^{T} f(y_{it}\mid y_{i,t-1})\right] \qquad (1.40)$$

where

$$f(y_{i1}) \;=\; \left\{\mu_{i1}^{(1)}\right\}^{y_{i1}}\left\{1\;-\;\mu_{i1}^{(1)}\right\}^{1-y_{i1}}$$

and

$$f(y_{it}\mid y_{i,t-1}) \;=\; \left\{\pi_{it|t-1}^{*(1)}\right\}^{y_{it}}\left\{1\;-\;\pi_{it|t-1}^{*(1)}\right\}^{1-y_{it}}; \qquad \text{for t} \;=\; 2, \cdots, \text{T}.$$

with $\pi_{it|t-1}^{*(1)}$ as given by (1.33).

(b) Conditional Contingency Tables Based Model (Without any Covariate)

In this covariates free conditional table set up, by following (1.32) and (1.33), we now write

$$\pi^{*(1)}_{t|t-1}(1) = \frac{e^{\beta_0+\gamma}}{1+e^{\beta_0+\gamma}} = \eta^{(1)}_{t|t-1}(1) \tag{1.41}$$

$$\pi^{*(1)}_{t|t-1}(0) = \frac{e^{\beta_0}}{1+e^{\beta_0}} = \eta^{(1)}_{t|t-1}(0) = \mu^{*(1)}_{(1)} \tag{1.42}$$

It then follows that the likelihood for the counts in Table 1.5 has the same form (1.28) as in the conditional linear binary set up. The difference lies in the fact that conditional probabilities $\eta^{(1)}_{t|t-1}(1)$ and $\eta^{(1)}_{t|t-1}(0)$ have the formulas (1.26) and (1.27) under the linear binary set up, whereas in the present covariates free non-linear binary set up, these conditional probabilities are given by (1.41) and (1.42).

Now, one can use these marginal and conditional probabilities to construct the likelihood function similar to (1.28) and obtain the maximum likelihood estimates of the parameters β_0 and γ.

1.2 Objective of the Thesis

In section 1.1.2.1, we have reviewed the existing (e.g. Agresti, 1990) binary longitudinal data analysis where time is considered as a fixed covariate. See, for example, the binary logistic probability model (1.10) where λ_t was used to represent the t-th time effect on the probability of the response. As mentioned earlier, this approach does not accommodate any correlations among longitudinal binary responses. The purpose of the thesis is to take such longitudinal correlations into account in multinomial setup. For this reason, we will not follow the fixed time covariate approach in the thesis any more.

Note that by taking the longitudinal correlations into account, a conditional linear binary probability model is provided in Section 1.1.2.2 for the analysis of history and conditional contingency tables based data. The objective of Chapter 2 is to generalize the history and conditional tables based longitudinal binary models to the longitudinal multinomial setup. The basic properties and the likelihood inferences for these models are given in details.

In Chapter 3, we consider a non-linear multinomial fixed effects model in longitudinal setup as a generalization of the longitudinal binary model discussed in Section 1.1.2.3. Note that as opposed to the conditional linear probability model discussed in

Chapter 2, this type of non-linear models produces longitudinal correlations satisfying full range. The basic properties and the likelihood inferences for these models are given in both history and contingency tables setup. In this chapter, we also carry out a simulation study to examine the small sample performance of the estimates in the non-linear setup. The Three Mile Island Stress-Level data set is analyzed by applying this non-linear model.

Note however that, in practice, it may happen that the longitudinal responses of an individual may also be affected by an unobserved random effect of the individual. In such cases, a longitudinal mixed model is used to accommodate both fixed regression as well as individual's random effects. For example, for a non-linear longitudinal binary mixed model, we refer to Sutradhar, Rao and Pandit (2008). The purpose of Chapter 4 is to generalize this binary mixed model to the multinomial setup. The properties and likelihood inferences for this multinomial mixed model are given in details. Furthermore, an extensive simulation study is conducted in this chapter to examine the small sample performance of the multinomial likelihood estimation approach.

In Chapter 5, we provide some concluding remarks and also indicate some future research in the longitudinal multinomial setup.

Chapter 2

Multinomial Linear Dynamic Fixed Probability Model

Inferences in conditional linear dynamic binary fixed models (1.18)-(1.19) have been studied recently in details by some authors such as Sutradhar (2010) [see also Sutradhar (2011)]. There are, however, many situations where one needs to deal with longitudinal responses with multiple (more than two) categories. For example, Conaway (1989) has studied such repeated categorial data with an application to the Three Mile Island Stress-Level data set with three categories, namely: low, medium and high stress levels. Similarly, Agresti (2002, Section 11.2.3, p. 462) has studied for an insomnia problem to examine the relationship between using a hypnotic drug (sleeping pill) and time to falling asleep with four categories.

These authors have, however, considered the time as a fixed covariate in order to understand the time effect on the multinomial responses. For this type of fixed covariates based analysis, we also refer to Agresti (1989, 1999). As far as the inference technique is concerned, they have used likelihood approach for the estimation of the category and time effects. But, in this thesis, we consider time as a nominal or an index variable and develop the longitudinal correlation models for repeated multinomial responses. Agresti (1993) used a correlation model where correlations are generated through random effects which does not appear to address longitudinal correlations as random effects remain the same over time. Note however that there is no unique way to model the longitudinal correlations, whether the responses are binary or multinomial. As mentioned earlier there exists certain conditional linear and non-linear models in longitudinal setup for binary data. In this chapter, we generalize the inferences in conditional linear binary dynamic models to the multinomial setup. This new model is refered to as the Multinomial Linear Dynamic Fixed Probability (MLDFP) model. This we do for two situations. First, for the history based data, i.e., when responses at every time point for all individuals are known. Second, when data are available in contingency table forms. This history based generalized model and its basic properties are discussed in Section 2.1. In the same section, we provide the likelihood inferences for the parameters of the proposed history based longitudinal multinomial model. In Section 2.2, we deal with longitudinal multinomial data in contingency table form for the cases when the covariates are time independent. In the same section, we provide the product multinomial likelihood approach for the estimation of the parameters of the contingency tables based stationary models.

Note that this extended multinomial model, similar to the conditional linear binary model, has nice features with respect to interpretation of the dynamic dependence parameter. However, the linearity may require some heavy restriction for the range of such parameter. Thus, for the sake of descriptive advantage we provide this model as an extension of the binary model. Instead of showing further application of this model, two non-linear multinomial dynamic models will be discussed in Chapters 3 and 4.

2.1 History Based Non-stationary MLDFP Model

Recall from Section 1.1.1 that in a cross-sectional setup (i.e. $T = 1$), the K-dimensional multinomial response under $K + 1$ categories of an individual i ($i = 1, \cdots, I$) was denoted by $\mathbf{y}_i = (y_{i1}, ..., y_{ij}, ..., y_{iK})'$. We now consider $T > 1$, and define

$$\mathbf{y}_{it} = (y_{it1}, \cdots, y_{itj}, \cdots, y_{itK})'; \quad t = 1, \cdots, T, \tag{2.1}$$

as the K-dimensional multinomial response for the ith individual at time point t. Suppose that $\mathbf{x}_{it} = (x_{it1}, \cdots, x_{itp})'$ is the p-dimensional time dependent covariate vector associated with $\mathbf{y}_{it}$ from the ith individual at time t. Note that these time dependent covariates will cause non-stationary correlations among the repeated multinomial responses. In the present history based longitudinal set up, it is assumed that all $\mathbf{y}_{it}$ and $\mathbf{x}_{it}$ are available for all $i = 1, \cdots, I$ and $t = 1, \cdots, T$.

Suppose that in the longitudinal setup, the jth ($j = 1, \cdots, K$) category response of the ith individual at time $t\,(t = 1, \cdots, T)$ is denoted by

$$\begin{aligned} \mathbf{y}_{it}^{(j)} &= (y_{it1}^{(j)}, \cdots, y_{itj}^{(j)}, \cdots, y_{itK}^{(j)})' \\ &= (0\mathbf{1}'_{j-1}, 1, 0\mathbf{1}'_{K-j})'; \quad j = 1, \cdots, K, \end{aligned} \tag{2.2}$$

i.e $y_{itj}^{(j)} = 1$ and $y_{itj}^{(c)} = 0$ for $j \neq c,\ j,\ c = 1, \cdots, K$. Note that the response in the last $[(K+1)th]$ category can be identified using

$$\begin{aligned} \mathbf{y}_{it}^{(K+1)} &= (0, \cdots, 0, \cdots, 0)' \\ &= 0\,\mathbf{1}'_K. \end{aligned}$$

Further note that because time is an index variable in our approach, by that we mean it should be latent and there is no meaning of attaching any quantitative value to any time points, even though as they can change the responses based on the length. It should be understood that time intervals may be meaningful and there is no reason to consider different regression effects due to change in time. Thus, we use the same regression parameter $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{ju}, \cdots, \beta_{jp})'$ to explain the effect of all time dependent covariates $\mathbf{x}_{it}$ on $\mathbf{y}_{it}^{(j)}$ for $j = 1, \cdots, K$ with $\boldsymbol{\beta}_{K+1} = (0, \cdots, 0)'$ by convention. As far as the time effect is concerned, it will be assumed that the multinomial responses collected over time will follow a suitable correlation structure.

In order to write a probability model for longitudinal multinomial responses denoted by (2.2), we, first, refer to the longitudinal correlated binary model (1.18)-(1.19) and then extend it to the multinomial setup as follows. Similar to the binary case, the marginal probability for the multinomial response to be in jth $(j = 1, \cdots, K)$ category at time $t = 1$ may be written as

$$P\left(\mathbf{Y}_{i1} = \mathbf{y}_{i1}^{(j)}\right) = \frac{\exp\left(\beta_{j0} + \sum_{u=1}^{p} \beta_{ju} x_{i1(u)}\right)}{\sum_{c=1}^{K+1} \exp\left(\beta_{c0} + \sum_{u=1}^{p} \beta_{cu} x_{i1(u)}\right)} = \frac{\exp\left(\beta_{j0} + \mathbf{x}_{i1}'\boldsymbol{\beta}_j\right)}{1 + \sum_{c=1}^{K} \exp\left(\beta_{c0} + \mathbf{x}_{i1}'\boldsymbol{\beta}_c\right)} = \pi_{i1}^{(j)}, \text{ say,} \tag{2.3}$$

the probability for the response to be in the last category is being given by

$$\pi_{i1}^{(K+1)} = 1 - \sum_{j=1}^{K} \pi_{i1}^{(j)} = \frac{1}{1 + \sum_{c=1}^{K} \exp\left(\beta_{c0} + \mathbf{x}_{i1}'\boldsymbol{\beta}_c\right)}. \tag{2.4}$$

Now, the conditional probabilities at time t ($t = 2, \cdots, T$) given the response at time $t-1$, may be written as

$$\begin{aligned} P\Big(\mathbf{Y}_{it} = \mathbf{y}_{it}^{(j)} \,|\, \mathbf{Y}_{i,t-1} = \mathbf{y}_{i,t-1}^{(l)}\Big) &= \pi_{it}^{(j)} + \sum_{q=1}^{K} \rho_{jq} \left(y_{i,t-1,q}^{(l)} - \pi_{i,t-1}^{(q)} \right) \\ &= \pi_{it}^{(j)} + \boldsymbol{\rho}_j' \left(\mathbf{y}_{i,t-1}^{(l)} - \mathbf{\Pi}_{i,t-1} \right) \\ &= \eta_{it|t-1}^{(j)}(l), \text{ say}, \quad j = 1, ..., K, \end{aligned} \tag{2.5}$$

where for any $t = 1, \cdots, T$, $\pi_{it}^{(j)} = \dfrac{\exp(\beta_{j0} + \mathbf{x}_{it}'\boldsymbol{\beta}_j)}{1 + \sum_{c=1}^{K} \exp(\beta_{c0} + \mathbf{x}_{it}'\boldsymbol{\beta}_c)}$.

For $j = K+1$, i.e., for the $(K+1)$th category, the conditional probability is given by

$$\eta_{it|t-1}^{(K+1)}(l) = \left(1 - \sum_{j=1}^{K} \eta_{it|t-1}^{(j)}(l) \right) \tag{2.6}$$

Note that in (2.3), β_{j0} represents an intercept parameter under jth category for $j = 1, \cdots, K$, and β_{ju} in $\boldsymbol{\beta}_j = (\beta_{j1}, \cdots, \beta_{ju}, \cdots, \beta_{jp})'$ represents the regression effect of the uth covariate for the individuals belonging to the jth category. Further note that as opposed to the binary variable $\mathbf{y}_i$ in (1.9), $\mathbf{y}_{i1}$ in (2.3) is a multinomial variable defined as in (2.1). Consequently, in the present multinomial setup, the parameters are denoted by β_{j0} and β_{ju} for $j = 1, \cdots, K$ categories, where as in the binary model (1.18) these parameters were denoted by β_0 and β_u, respectively, for $K = 1$, i.e., for 2 categories.

As far as the correlations among multinomial responses are concerned, they are generated by the conditional model (2.5), where ρ_{jl} refers to the dynamic dependence parameter relating the multinomial response belonging to the jth category at time t with the previous response being in lth category at time $t-1$. Note that it is enough to consider ρ_{jl} for $l = 1, \cdots, K$, because of the fact that $\rho_{j,K+1} = 0$ by convention, for all $j = 1, \cdots, K+1$.

Note that as it is not easy to model the longitudinal correlations, some authors such as Miller et al. (1993) and Lipsitz et al. (1994), used 'working' correlations (equi-correlation or independence) approach which however suffers from definition problem as discussed by Sutradhar (2003) [see also Crowder (1995), Sutradhar (2011)]. However, because it is most likely that the correlations for longitudinal data decay as lag increases, in the thesis, we have considered an autoregressive type dynamic model that accommodates this decaying property for the correlations. For some alternative modeling for longitudinal correlations for multinomial responses, we refer to Sutradhar and Kovacevic (2000), and Molenberghs and Lesaffre (1994).

2.1.1 Basic Properties of the History Based Model

Note that it is standard to interpret the data through their mean, variance and correlations. For the purpose, in this section, we provide these basic properties for the longitudinal multinomial responses following the model (2.3)-(2.6). Further note that as these basic statistics will be functions of the parameters β_{j0}, β_{ju} and ρ_{jl} $(j = 1, \cdots, K\,;\, l = 1, \cdots, K\,;\, u = 1, \cdots, p)$ involved in the model, it will be necessary to estimate them as efficiently as possible. We will use the well-known likelihood method for such inferences in Section 2.2.

We provide the means and the variances of the multinomial responses in Lemma 2.1 and the covariances between multinomial responses at any two time points in Lemma 2.2.

Lemma 2.1: For $i = 1, \cdots, I$ and $t = 1, \cdots, T$, the unconditional mean vector and the covariance matrix of the multinomial response vector $\mathbf{Y}_{it} = (Y_{it1}, \cdots, Y_{itj}, \cdots, Y_{itK})$ have the forms

$$E\left(\mathbf{Y}_{it}\right) = \left(\pi_{it}^{(1)}, \cdots, \pi_{it}^{(j)}, \cdots, \pi_{it}^{(K)}\right)' = \mathbf{\Pi}_{it}, \tag{2.7}$$

and

$$Var(\mathbf{Y}_{it}) = diag\left[\pi_{it}^{(1)}, \cdots, \pi_{it}^{(j)}, \cdots, \pi_{it}^{(K)}\right] - \mathbf{\Pi}_{it}\mathbf{\Pi}_{it}', \tag{2.8}$$

where $\pi_{it}^{(j)}$ is defined in (2.5) for all $t = 1, \cdots, T$; and $j = 1, \cdots, K$.

Proof: At initial time point $t = 1$, for a given individual i $(i = 1, \cdots, I)$ the categorical response vector $\mathbf{Y}_{i1}$ marginally follows a multinomial distribution with density function

$$\begin{aligned} P(Y_{i11} = y_{i11}, \cdots, Y_{i1K} = y_{i1K}) &= \frac{1}{y_{i11}! \cdots y_{i1K}!, \, (1 - \sum_{q=1}^{K} y_{i1q})!} \prod_{q=1}^{K} \left(\pi_{i1}^{(q)}\right)^{y_{i1q}} \\ &\times \left(1 - \sum_{q=1}^{K} \pi_{i1}^{(q)}\right)^{(1-\sum_{q=1}^{K} y_{i1q})}, \end{aligned} \tag{2.9}$$

yielding the marginal mean and marginal variance of $\mathbf{Y}_{i1}$

$$E(\mathbf{Y}_{i1}) = \mathbf{\Pi}_{i1} = \left(\pi_{i1}^{(1)}, \cdots, \pi_{i1}^{(j)}, \cdots, \pi_{i1}^{(K)}\right)' \tag{2.10}$$

$$Var(\mathbf{Y}_{i1}) = diag\left[\pi_{i1}^{(1)}, \cdots, \pi_{i1}^{(j)}, \cdots, \pi_{i1}^{(K)}\right] - \mathbf{\Pi}_{i1}\mathbf{\Pi}_{i1}'. \tag{2.11}$$

Now, in general, at time point $t = 2, \cdots, T$, we can write the conditional distribution of the multinomial response vector $\mathbf{Y}_{it}$ given $\mathbf{y}_{i,t-1}$ as

$$\begin{aligned} P(Y_{it1} = y_{it1}, \cdots, Y_{itk} = y_{itk} | \, y_{i,t-1}^{(l)}) &= \frac{1}{y_{it1}! \cdots y_{itk}!, \, (1 - \sum_{q=1}^{K} y_{itq})!} \prod_{q=1}^{K} \left(\eta_{it|t-1}^{(q)}(l)\right)^{y_{itq}} \\ &\times \left(1 - \sum_{q=1}^{K} \eta_{it|t-1}^{(q)}(l)\right)^{(1-\sum_{q=1}^{K} y_{itq})}, \end{aligned} \tag{2.12}$$

where the conditional probability $\eta^{(q)}_{it|t-1}(l)$ are defined in (2.5). It then follows from (2.5) that the conditional mean and conditional variance have the formulas

$$\begin{aligned} E(\mathbf{Y}_{it}|\mathbf{y}^{(l)}_{i,t-1}) &= \tilde{\mathbf{\Pi}}_{it} = \left(\eta^{(1)}_{it|t-1}(l),\cdots,\eta^{(j)}_{it|t-1}(l),\cdots,\eta^{(K)}_{it|t-1}(l)\right)' \quad (2.13)\\ Var(\mathbf{Y}_{it}|\mathbf{y}^{(l)}_{i,t-1}) &= diag\left[\eta^{(1)}_{it|t-1}(l),\cdots,\eta^{(j)}_{it|t-1}(l),\cdots,\eta^{(K)}_{it|t-1}(l)\right] - \tilde{\mathbf{\Pi}}_{it}\tilde{\mathbf{\Pi}}'_{it}. \quad (2.14)\end{aligned}$$

We now derive the formulas for $E(\mathbf{Y}_{it})$ and $Cov(\mathbf{Y}_{it})$ for $t=2$ using conditioning and unconditioning properties of the expectations. Note that for $t=2$, it follows from (2.12) that

$$\begin{aligned} E(\mathbf{Y}_{i2}|\mathbf{y}^{(l)}_{i1}) &= \tilde{\mathbf{\Pi}}_{i2} = \left(\eta^{(1)}_{i2|1}(l),\cdots,\eta^{(j)}_{i2|1}(l),\cdots,\eta^{(K)}_{i2|1}(l)\right)' \\ &= \begin{bmatrix}\pi^{(1)}_{i2}\\ \vdots \\ \pi^{(j)}_{i2} \\ \vdots \\ \pi^{(K)}_{i2}\end{bmatrix} + \begin{bmatrix}\rho_{11} & \cdots & \rho_{1l} & \cdots & \rho_{2K}\\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho_{j1} & \cdots & \rho_{jl} & \cdots & \rho_{jK} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \rho_{K1} & \cdots & \rho_{Kl} & \cdots & \rho_{KK}\end{bmatrix} \times \begin{bmatrix} y^{(l)}_{i11} - \pi^{(1)}_{i1} \\ \vdots \\ y^{(l)}_{i1j} - \pi^{(j)}_{i1} \\ \vdots \\ y^{(l)}_{i1K} - \pi^{(K)}_{i1}\end{bmatrix} \\ &= \mathbf{\Pi}_{i2} + \boldsymbol{\rho}_M\left(\mathbf{Y}^{(l)}_{i1} - \mathbf{\Pi}_{i1}\right), \text{ say.} \quad (2.15)\end{aligned}$$

Consequently, we can write the unconditional mean at time $t=2$ as

$$E(\mathbf{Y}_{i2}) = E_{\mathbf{Y}_{i1}}E\left(\mathbf{Y}_{i2}|\mathbf{y}_{i1}\right) = E_{\mathbf{Y}_{i1}}\left[\mathbf{\Pi}_{i2} + \boldsymbol{\rho}_M\left(\mathbf{Y}_{i1} - \mathbf{\Pi}_{i1}\right)\right] = \mathbf{\Pi}_{i2}$$

because $E_{\mathbf{Y}_{i1}}(\mathbf{Y}_{i1}) = \mathbf{\Pi}_{i1}$ by (2.10).

Since $E(\mathbf{Y}_{i2}) = \mathbf{\Pi}_{i2}$ has the same form as $E(\mathbf{Y}_{i1}) = \mathbf{\Pi}_{i1}$ and because by (2.13) the conditional mean i.e., $E(\mathbf{Y}_{it}|\mathbf{y}_{i,t-1}) = \tilde{\mathbf{\Pi}}_{it}$ has the same structure as in (2.15) for all $t = 2, \cdots, T$, it then follows that

$$E(\mathbf{Y}_{it}) = \mathbf{\Pi}_{it} \qquad \text{for all} \quad t = 1, \cdots, T. \tag{2.16}$$

Next to derive the unconditional covariance matrix of $\mathbf{Y}_{it}$, we will use the same conditioning and unconditioning properties of expectations as we have used for the derivation of the unconditional mean vector. To be specific, we first write

$$Var(\mathbf{Y}_{it}) = E_{\mathbf{Y}_{i,t-1}} Var\left[\mathbf{Y}_{it}\middle|\mathbf{Y}_{i,t-1}\right] + Var_{\mathbf{Y}_{i,t-1}} E\left[\mathbf{Y}_{it}\middle|\mathbf{Y}_{i,t-1}\right]. \tag{2.17}$$

Now by using (2.14) and (2.16), the first term in the right hand side of (2.17) may be expressed as

$$\begin{aligned} E_{\mathbf{Y}_{i,t-1}} Var\left[\mathbf{Y}_{it}\middle|\mathbf{Y}_{i,t-1}\right] &= E_{\mathbf{Y}_{i,t-1}}\left\{diag\left[\eta^{(1)}_{it|t-1}(l), \cdots, \eta^{(j)}_{it|t-1}(l), \cdots, \eta^{(K)}_{it|t-1}(l)\right] \right. \\ &\qquad \left. - \tilde{\mathbf{\Pi}}_{it}\tilde{\mathbf{\Pi}}'_{it}\right\} \\ &= diag\left[\pi^{(1)}_{it}, \cdots, \pi^{(j)}_{it}, \cdots, \pi^{(K)}_{it}\right] - E_{Y_{i,t-1}}\left[\tilde{\mathbf{\Pi}}_{it}\tilde{\mathbf{\Pi}}'_{it}\right], \end{aligned} \tag{2.18}$$

where

$$\begin{aligned} E_{Y_{i,t-1}}\left[\tilde{\mathbf{\Pi}}_{it}\tilde{\mathbf{\Pi}}'_{it}\right] &= E_{Y_{i,t-1}}\left[\mathbf{\Pi}_{it}+\boldsymbol{\rho}_M\Big(\mathbf{Y}_{i,t-1}-\mathbf{\Pi}_{i,t-1}\Big)\right]\left[\mathbf{\Pi}_{it}+\boldsymbol{\rho}_M\Big(\mathbf{Y}_{i,t-1}-\mathbf{\Pi}_{i,t-1}\Big)\right]' \\ &= \mathbf{\Pi}_{it}\mathbf{\Pi}'_{it} + \boldsymbol{\rho}_M\, E_{Y_{i,t-1}}\left[\Big(\mathbf{Y}_{i,t-1}-\mathbf{\Pi}_{i,t-1}\Big)\Big(\mathbf{Y}_{i,t-1}-\mathbf{\Pi}_{i,t-1}\Big)'\right]\boldsymbol{\rho}'_M \\ &= \mathbf{\Pi}_{it}\mathbf{\Pi}'_{it} + \boldsymbol{\rho}_M\, Var\left[\mathbf{Y}_{i,t-1}\right]\boldsymbol{\rho}'_M, \end{aligned} \tag{2.19}$$

because $E\big(\mathbf{Y}_{i,t-1}-\mathbf{\Pi}_{i,t-1}\big)=0$.

Similarly, the second term in the right hand side of (2.17) may be expressed as

$$\begin{aligned} Var_{Y_{i,t-1}}E\left[\mathbf{Y}_{it}\,|\,\mathbf{Y}_{i,t-1}\right] &= Var_{Y_{i,t-1}}\left[\tilde{\mathbf{\Pi}}_{it}\right] \\ &= Var_{Y_{i,t-1}}\left[\mathbf{\Pi}_{it}+\boldsymbol{\rho}_M\Big(\mathbf{Y}_{i,t-1}-\mathbf{\Pi}_{i,t-1}\Big)\right] \\ &= \boldsymbol{\rho}_M\, Var\left[\mathbf{Y}_{i,t-1}\right]\boldsymbol{\rho}'_M \end{aligned} \tag{2.20}$$

Now by using (2.18) and (2.20) in (2.17), we obtain

$$Var\Big(\mathbf{Y}_{it}\Big) = diag\left[\pi_{it}^{(1)},\cdots,\pi_{it}^{(j)},\cdots,\pi_{it}^{(K)}\right] - \mathbf{\Pi}_{it}\mathbf{\Pi}'_{it}; \quad \text{for } t=1,\cdots,T. \ \Diamond$$

Note that the multinomial mean vector and the covariance matrix given in Lemma 2.1 are simply the generalization of the binary case ($K=1$). That is, for $K=1$, $E(Y_{it})=\pi_{it}^{(1)}$ and $Var(Y_{it})=\pi_{it}^{(1)}(1-\pi_{it}^{(1)})$. Also, it follows from the formula for $Var(Y_{it})$ that at time point t,

$$Cov(Y_{it}^{(j)},Y_{it}^{(l)}) = -\pi_{it}^{(j)}\pi_{it}^{(l)}.$$

This shows that even though multinomial responses at two consecutive times maintain a relationship through the dependence parameters ρ_{jl} $(j = 1, \cdots, K; l = 1, \cdots, K)$, the mean, variance and covariances (between categories) at any given time are same as those of a marginal multinomial distribution at that time point.

We now proceed to understand the covariances between any two multinomial responses recorded at two distinct time points m and t such that $m < t$. For the purpose, we, for clarity, do this when $t - m$ is small which is practically meaningful. To be specific, we compute the covariance matrices successively for lags 1,2 and 3; and then provide a general formula. Thus, for lag 1, we compute $Cov(\mathbf{Y}_{i2}, \mathbf{Y}_{i1})$, $Cov(\mathbf{Y}_{i3}, \mathbf{Y}_{i2})$ and write the general form for $Cov(\mathbf{Y}_{it}, \mathbf{Y}_{i,t-1})$.

Computation of Lag 1 Covariances:

We first compute

$$\begin{aligned}
E[Y_{i2}Y'_{i1}] &= E_{Y_{i1}}\, E\,[Y_{i2}\, Y'_{i1}\,|\, y_{i1}] \\
&= E_{Y_{i1}}\left\{\tilde{\Pi}_{i2}\, Y'_{i1}\right\} \qquad \text{by (2.13)} \\
&= E_{Y_{i1}}\,[\{\Pi_{i2} + \rho_M\,(Y_{i1} - \Pi_{i1})\}\, Y'_{i1}] \\
&= \Pi_{i2}\,\Pi'_{i1} + \rho_M\, Var(Y_{i1}) \qquad (2.21)
\end{aligned}$$

Thus,

$$
\begin{aligned}
Cov(Y_{i2}, Y_{i1}) &= E\left[Y_{i2}\, Y_{i1}'\right] - E(Y_{i2})\left\{E(Y_{i1})\right\}' \\
&= \Pi_{i2}\, \Pi_{i1}' + \rho_M\, Var(Y_{i1}) - \Pi_{i2}\, \Pi_{i1}' \\
&= \rho_M\, Var(Y_{i1})
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
E[Y_{i3} Y_{i2}'] &= E_{Y_{i2}}\, E\left[Y_{i3}\, Y_{i2}' \,|\, y_{i2}\right] \\
&= E_{Y_{i2}}\left\{\tilde{\Pi}_{i3}\, Y_{i2}'\right\} \quad \text{by (2.13)} \\
&= E_{Y_{i2}}\left[\left\{\Pi_{i3} + \rho_M\, (Y_{i2} - \Pi_{i2})\right\} Y_{i2}'\right] \\
&= \Pi_{i3}\, \Pi_{i2}' + \rho_M\, Var(Y_{i2}), \qquad (2.22)
\end{aligned}
$$

by Lemma 2.1. Thus,

$$
\begin{aligned}
Cov(Y_{i3}, Y_{i2}) &= E\left[Y_{i3}\, Y_{i2}'\right] - E(Y_{i3})\left\{E(Y_{i2})\right\}' \\
&= \Pi_{i3}\, \Pi_{i2}' + \rho_M\, Var(Y_{i2}) - \Pi_{i3}\, \Pi_{i2}' \\
&= \rho_M\, Var(Y_{i2}).
\end{aligned}
$$

Note that because $E(Y_{it}\, Y_{i,t-1}')$ for any $t = 4, \cdots, T$, have the same structures as those of $E(Y_{i2}\, Y_{i1}')$ in (2.22) and $E(Y_{i3}\, Y_{i2}')$ in (2.23), it then follows that

$$
Cov(Y_{it}, Y_{i,t-1}) = \rho_M\, Var(Y_{i,t-1}); \qquad for\ t = 2, \cdots, T. \qquad (2.23)
$$

Computation of Lag 2 Covariances:

We compute

$$\begin{aligned}
E[Y_{i3}Y_{i1}'] &= E_{Y_{i1}}\, E_{Y_{i2}}\, E\,[Y_{i3}\, Y_{i1}' \,|\, y_{i2}\, y_{i1}] \\
&= E_{Y_{i1}}\, E_{Y_{i2}}\, [\tilde{\Pi}_{i3}\, Y_{i1}' \,|\, y_{i1}] \quad \text{by (2.13)} \\
&= E_{Y_{i1}}\, [E_{Y_{i2}}\, \{\, \Pi_{i3} \,+\, \rho_M\, (Y_{i2} \,-\, \Pi_{i2})\}\, Y_{i1}' \,|\, y_{i1}] \\
&= E_{Y_{i1}}\, \left[\left\{ \Pi_{i3} \,+\, \rho_M \left(\tilde{\Pi}_{i2} \,-\, \Pi_{i2} \right) \right\} Y_{i1}' \right] \quad \text{by (2.13)} \\
&= E_{Y_{i1}}\, [\, \Pi_{i3}\, Y_{i1}' + \rho_M\, \{\, \Pi_{i2} \,+\, \rho_M\, (Y_{i1} \,-\, \Pi_{i1})\}\, Y_{i1}' \,-\, \rho_M\, \Pi_{i2}\, Y_{i1}'] \\
&= \Pi_{i3}\, \Pi_{i1}' + \rho_M\, \Pi_{i2}\, \Pi_{i1}' \,+\, \rho_M^2\, Var(Y_{i1}) \,-\, \rho_M\, \Pi_{i2}\, \Pi_{i1}' \\
&= \Pi_{i3}\, \Pi_{i1}' \,+\, \rho_M^2\, Var(Y_{i1}). \qquad (2.24)
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
Cov(Y_{i3}, Y_{i1}) &= E\,[Y_{i3}\, Y_{i1}'] \,-\, E(Y_{i3})\, \{E(Y_{i1})\}' \\
&= \Pi_{i3}\, \Pi_{i1}' \,+\, \rho_M^2\, Var(Y_{i1}) \,-\, \Pi_{i3}\, \Pi_{i1}' \\
&= \rho_M^2\, Var(Y_{i1}). \qquad (2.25)
\end{aligned}$$

Similarly,

$$\begin{aligned}
E[Y_{i4}Y_{i2}'] &= E_{Y_{i2}}\, E_{Y_{i3}}\, E\,[Y_{i4}\, Y_{i2}' \,|\, y_{i3}\, y_{i2}] \\
&= E_{Y_{i2}}\, E_{Y_{i3}}\, [\tilde{\Pi}_{i4}\, Y_{i2}' \,|\, y_{i2}] \\
&= E_{Y_{i2}}\, [E_{Y_{i3}}\, \{\, \Pi_{i4} \,+\, \rho_M\, (Y_{i3} \,-\, \Pi_{i3})\}\, Y_{i2}' \,|\, y_{i2}]
\end{aligned}$$

$$= E_{Y_{i2}}\left[\left\{\Pi_{i4} + \rho_M\left(\tilde{\Pi}_{i3} - \Pi_{i3}\right)\right\} Y'_{i2}\right]$$

$$= E_{Y_{i2}}\left[\Pi_{i4} Y'_{i2} + \rho_M\left\{\Pi_{i3} + \rho_M\left(Y_{i2} - \Pi_{i2}\right)\right\} Y'_{i2} - \rho_M \Pi_{i3} Y'_{i2}\right]$$

$$= \Pi_{i4}\Pi'_{i2} + \rho_M \Pi_{i3}\Pi'_{i2} + \rho^2_M Var(Y_{i2}) - \rho_M \Pi_{i3}\Pi'_{i2}$$

$$= \Pi_{i4}\Pi'_{i2} + \rho^2_M Var(Y_{i2}). \quad (2.26)$$

Thus,

$$\begin{aligned} Cov(Y_{i4}, Y_{i2}) &= E\left[Y_{i4} Y'_{i2}\right] - E(Y_{i4})\{E(Y_{i2})\}' \\ &= \Pi_{i4}\Pi'_{i2} + \rho^2_M Var(Y_{i2}) - \Pi_{i4}\Pi'_{i2} \\ &= \rho^2_M Var(Y_{i2}). \end{aligned} \quad (2.27)$$

Note that as the formulas for $E(Y_{it} Y'_{i,t-2})$ for $t = 5, \cdots, T$ will have the same structures as in (2.25) and (2.27), we can write

$$Cov(Y_{it}, Y_{i,t-2}) = \rho^2_M Var(Y_{i,t-2}); \quad for\ t = 3, \cdots, T. \quad (2.28)$$

Computation of Lag 3 Covariances:

Here, we compute

$$\begin{aligned} E[Y_{i4} Y'_{i1}] &= E_{Y_{i1}} E_{Y_{i2}} E_{Y_{i3}} E\left[Y_{i4} Y'_{i1} \mid y_{i3}\, y_{i2}\, y_{i1}\right] \\ &= E_{Y_{i1}} E_{Y_{i2}} E_{Y_{i3}}\left[\tilde{\Pi}_{i4} Y'_{i1} \mid y_{i2}\, y_{i1}\right] \\ &= E_{Y_{i1}} E_{Y_{i2}} E_{Y_{i3}}\left[\left\{\Pi_{i4} + \rho_M\left(Y_{i3} - \Pi_{i3}\right)\right\} Y'_{i1} \mid y_{i2}\, y_{i1}\right] \\ &= E_{Y_{i1}} E_{Y_{i2}}\left[\left\{\Pi_{i4} + \rho_M\left(\tilde{\Pi}_{i3} - \Pi_{i3}\right)\right\} Y'_{i1} \mid y_{i1}\right] \end{aligned}$$

$$
\begin{aligned}
&= E_{Y_{i1}} E_{Y_{i2}} \left[\Pi_{i4} Y'_{i1} + \rho_M \left\{ \Pi_{i3} + \rho_M \left(Y_{i2} - \Pi_{i2} \right) \right\} Y'_{i1} - \rho_M \Pi_{i3} Y'_{i1} | y_{i1} \right] \\
&= \Pi_{i4} E_{Y_{i1}} [Y'_{i1}] + E_{Y_{i1}} \left[\rho_M^2 \left(\tilde{\Pi}_{i2} - \Pi_{i2} \right) Y'_{i1} \right] \\
&= \Pi_{i4} \Pi'_{i1} + E_{Y_{i1}} \left[\rho_M^2 \left\{ \Pi_{i2} + \rho_M \left(Y_{i1} - \Pi_{i1} \right) \right\} Y'_{i1} \right] - \rho_m^2 \Pi_{i2} E_{Y_{i1}} [Y'_{i1}] \\
&= \Pi_{i4} \Pi'_{i1} + \rho_M^2 \Pi_{i2} \Pi'_{i1} + \rho_M^3 Var(Y_{i1}) - \rho_m^2 \Pi_{i2} \Pi'_{i1} \\
&= \Pi_{i4} \Pi'_{i1} + \rho_M^3 Var(Y_{i1}) \qquad (2.29)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
Cov(Y_{i4}, Y_{i1}) &= E[Y_{i4} Y'_{i1}] - E(Y_{i4}) \{E(Y_{i1})\}' \\
&= \Pi_{i4} \Pi'_{i1} + \rho_M^3 Var(Y_{i1}) - \Pi_{i4} \Pi'_{i1} \\
&= \rho_M^3 Var(Y_{i1}) \qquad (2.30)
\end{aligned}
$$

Note that because the formulas for the lag covariances in (2.23), (2.28) and (2.30) reveal a clear dynamic pattern, one may exploit this pattern and write the lag $(t-m)$ $(m < t)$ covariance between $\mathbf{Y}_{it}$ and $\mathbf{Y}_{im}$ as

$$
Cov\left(\mathbf{Y}_{it}, \mathbf{Y}_{im}\right) = \boldsymbol{\rho}_M^{t-m} Var\left(\mathbf{Y}_{im}\right), \qquad (2.31)
$$

where

$$
\boldsymbol{\rho}_M = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1K} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{K1} & \rho_{K2} & \cdots & \rho_{KK} \end{bmatrix}_{K \times K}
$$

and

$$Var\left(\mathbf{Y}_{im}\right) = diag\left[\pi_{im}^{(1)}, \cdots, \pi_{im}^{(j)}, \cdots, \pi_{im}^{(K)}\right] - \mathbf{\Pi}_{im}\mathbf{\Pi}_{im}',$$

as in (2.8). $\diamond$

2.1.2 Likelihood Estimation for the History Based Model

In the previous section we have derived the basic properties such as mean, variance and covariances under the conditional linear dynamic multinomial model (2.3)-(2.6). Since these basic properties are interpreted in terms of the parameters β_{j0}, β_{ju} and ρ_{jl}, we now estimate them efficiently by exploiting the well known likelihood approach. Note that the regression parameters β_{j0} and β_{jl} are also of primary interest.

Now by writing $\boldsymbol{\beta} \equiv (\beta_{j0}, \beta_{ju})$ and $\boldsymbol{\rho} \equiv (\rho_{jl})$ for $j = 1, \cdots, K$, $l = 1, \cdots, K$, $u = 1, \cdots, p$, it follows from the model (2.3)-(2.6) that the likelihood function has the form

$$L(\boldsymbol{\beta}, \boldsymbol{\rho}) = \prod_{i=1}^{I}\left[f(\mathbf{y}_{i1})\prod_{t=2}^{T} f(\mathbf{y}_{it} \mid \mathbf{y}_{i,t-1})\right] \tag{2.32}$$

where $f(\mathbf{y}_{i1})$, the multinomial density of $\mathbf{y}_{i1}$ has the form as in (2.9), and $f(\mathbf{y}_{it} \mid \mathbf{y}_{i,t-1})$, the conditional multinomial density of $\mathbf{y}_{it}$ given $\mathbf{y}_{i,t-1}$ has the form given by (2.12).

Thus, by using (2.9) and (2.12) in (2.32), one writes the log likelihood as

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\rho}) \;=\; C + \sum_{i=1}^{I} \Big[g_i(\boldsymbol{\beta}) + h_i(\boldsymbol{\beta}, \boldsymbol{\rho}) \Big] \tag{2.33}$$

where

$$g_i(\boldsymbol{\beta}) \;=\; \sum_{q=1}^{K} y_{i1q}\Big(\beta_{q0} + x_{i1}'\boldsymbol{\beta}_q\Big) - \ln\left\{1 + \sum_{l=1}^{K} \exp\Big(\beta_{l0} + \mathbf{x}_{i1}'\boldsymbol{\beta}_l\Big)\right\} \tag{2.34}$$

$$\begin{aligned} h_i(\boldsymbol{\beta}, \boldsymbol{\rho}) \;=\; & \sum_{t=2}^{T}\left[\sum_{q=1}^{K} y_{itq} \ln\left\{\pi_{it}^{(q)} + \rho_q'\left(y_{i,t-1}^{(l)} - \Pi_{i,t-1}\right)\right\} + \left(1 - \sum_{q=1}^{K} y_{itq}\right)\right. \\ & \left. \times \ln\left\{1 - \sum_{q=1}^{K} \pi_{it}^{(q)} - \sum_{q=1}^{K} \rho_q'\left(y_{i,t-1}^{(l)} - \Pi_{i,t-1}\right)\right\}\right] \end{aligned} \tag{2.35}$$

For convenience of estimation, we denote the elements of $\boldsymbol{\beta}$ as

$$\boldsymbol{\beta} \equiv \left(\beta_1^{*\prime}, \cdots, \beta_j^{*\prime}, \cdots, \beta_K^{*\prime}\right)'$$

where

$$\begin{aligned} \boldsymbol{\beta}_j^* \;&=\; (\beta_{j0}, \beta_{j1}, \cdots, \beta_{ju}, \cdots, \beta_{jp})' \\ &=\; \left(\beta_{j0}, \boldsymbol{\beta}_j'\right)'. \end{aligned} \tag{2.36}$$

And the corresponding (p+1)-dimensional covariate vector is denoted by

$$\boldsymbol{x}_{it}^* \;=\; \left(1, x_{it(1)}, \cdots, x_{it(u)}, \cdots, x_{it(p)}\right)' \;=\; \left(1, \boldsymbol{x}_{it}'\right)'. \tag{2.37}$$

Similarly the elements of $\boldsymbol{\rho}$ are expressed as

$$\boldsymbol{\rho} \equiv \left(\boldsymbol{\rho}'_1, \cdots, \boldsymbol{\rho}'_j, \cdots, \boldsymbol{\rho}'_K\right)'$$

where

$$\boldsymbol{\rho}'_j = (\rho_{j1}, \cdots, \rho_{jl}, \cdots, \rho_{jK}). \tag{2.38}$$

We now write the likelihood estimating equations for $\boldsymbol{\beta}^*_j$ and $\boldsymbol{\rho}_j$ for all $j = 1, \cdots, K$. These equations are given by

$$\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta^*_j} = \sum_{i=1}^{I} \left[\frac{\partial g_i(\beta)}{\partial \beta^*_j} + \frac{\partial h_i(\beta, \rho)}{\partial \beta^*_j}\right] = 0, \tag{2.39}$$

and

$$\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_j} = \sum_{i=1}^{I} \left[\frac{\partial g_i(\beta)}{\partial \rho_j} + \frac{\partial h_i(\beta, \rho)}{\partial \rho_j}\right] = 0, \tag{2.40}$$

respectively. The derivatives in (2.39) have the formulas as

$$\frac{\partial g_i(\beta)}{\partial \beta^*_j} = \left[y_{i1j} - \pi^{(j)}_{i1}\right] x^*_{i1}, \tag{2.41}$$

$$\frac{\partial h_i(\beta, \rho)}{\partial \beta^*_j} = \sum_{t=2}^{T} \sum_{q=1}^{K} \left[\left\{\left(\frac{y_{itq}}{\eta^{(q)}_{it|t-1}(l)}\right) - \frac{\left(1 - \sum_{q=1}^{K} y_{itq}\right)}{\left(1 - \sum_{q=1}^{K} \eta^{(q)}_{it|t-1}(l)\right)}\right\} \times \left(\frac{\partial \eta^{(q)}_{it|t-1}(l)}{\partial \beta^*_j}\right)\right], \tag{2.42}$$

where

$$\frac{\partial\, \eta^{(q)}_{it|t-1}(l)}{\partial \beta^*_j} = \begin{cases} \pi^{(j)}_{it}\left(1-\pi^{(j)}_{it}\right) x^*_{it} - \rho_{jj}\left\{\pi^{(j)}_{i,t-1}\left(1-\pi^{(j)}_{i,t-1}\right)\right\} x^*_{i,t-1} \\ \qquad + \sum_{l\neq j}^{K} \rho_{j\,l}\left\{\pi^{(j)}_{i,t-1}\,\pi^{(l)}_{i,t-1}\right\} x^*_{i,t-1}, \quad \text{for } q = j \\ -\,\pi^{(q)}_{it}\,\pi^{(j)}_{it}\,x^*_{it} - \rho_{qj}\,\pi^{(j)}_{i,t-1}\left(1-\pi^{(j)}_{i,t-1}\right) x^*_{i,t-1} \\ \qquad + \sum_{l\neq j}^{K} \rho_{ql}\left\{\pi^{(l)}_{i,t-1}\pi^{(j)}_{i,t-1}\right\} x^*_{i,t-1}, \quad \text{for } q \neq j \end{cases} \tag{2.43}$$

Similarly the derivatives in (2.40) have the formulas

$$\frac{\partial\, g_i(\beta)}{\partial\, \rho_j} = 0 \; : \; K \times 1, \tag{2.44}$$

$$\frac{\partial\, h_i(\beta,\, \rho)}{\partial\, \rho_j} = \sum_{t=2}^{T}\left[\left\{\left(\frac{y_{itj}}{\eta^{(j)}_{it|t-1}(l)}\right) - \frac{\left(1-\sum_{q=1}^{K} y_{itq}\right)}{\left(1-\sum_{q=1}^{K}\eta^{(q)}_{it|t-1}(l)\right)}\right\} \times \left(\frac{\partial\, \eta^{(j)}_{it|t-1}(l)}{\partial \rho_j}\right)\right] \; : \; K \times 1, \tag{2.45}$$

where

$$\frac{\partial\, \eta^{(j)}_{it|t-1}(l)}{\partial \rho_j} = \left(y^{(l)}_{i,t-1} - \Pi_{i,t-1}\right) \quad \text{and} \quad \frac{\partial\, \eta^{(q)}_{it|t-1}(l)}{\partial \rho_j} = 0\;; \;\; q \neq j, \tag{2.46}$$

with $y_{i,t-1} = (y_{i,t-1,1}, \cdots, y_{i,t-1,j}, \cdots, y_{i,t-1,K})'$ as in (2.1)

and $\Pi_{i,t-1} = \left(\pi^{(1)}_{i,t-1}, \cdots, \pi^{(j)}_{i,t-1}, \cdots, \pi^{(K)}_{i,t-1}\right)'$ as in (2.7).

Next by using the well known Newton-Raphson iterative procedure, we compute the likelihood estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ by

$$\hat{\boldsymbol{\beta}}_{(r+1)} = \hat{\boldsymbol{\beta}}_{(r)} + \left[\frac{\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \boldsymbol{\beta}\, \partial \boldsymbol{\beta}'}\right]^{-1} \frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \boldsymbol{\beta}} \tag{2.47}$$

and

$$\hat{\boldsymbol{\rho}}_{(r+1)} = \hat{\boldsymbol{\rho}}_{(r)} + \left[\frac{\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}\, \partial \boldsymbol{\rho}'}\right]^{-1} \frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \tag{2.48}$$

where

$$\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta} = \left[\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta_1^{*\prime}}, \ldots, \frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta_j^{*\prime}}, \ldots, \frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta_K^{*\prime}}\right]'$$

$$\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho} = \left[\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_1'}, \ldots, \frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_j'}, \ldots, \frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_K'}\right]',$$

with $\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta_j^*}$ and $\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_j}$ as in (2.39) and (2.40) respectively.

2.1.2.1 Formulas for the Second Derivatives With Respect to β

Note that to compute the second derivative matrix in (2.47), it is sufficient to compute the following two derivatives

$$\frac{\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta_j^* \, \partial \beta_m^{*\prime}} = \begin{cases} -\sum_{i=1}^{I} \left[\pi_{i1}^{(j)} \left\{ 1 - \pi_{i1}^{(j)} \right\} x_{i1}^* \, x_{i1}^{*\prime} - h_{ijj}''(\beta, \rho) \right] \\ \\ \sum_{i=1}^{I} \left[\pi_{i1}^{(j)} \, \pi_{i1}^{(m)} \, x_{i1}^* \, x_{i1}^{*\prime} + h_{ijm}''(\beta, \rho) \right] \end{cases} \tag{2.49}$$

where

$$\begin{aligned} h_{ijm}''(\beta, \rho) = & \sum_{i=1}^{I} \sum_{t=2}^{T} \sum_{q=1}^{K} \left[\left(\frac{y_{itq}}{\eta_{it|t-1}^{(q)}(l)} \right) \left\{ \left(\frac{\partial^2 \eta_{it|t-1}^{(q)}(l)}{\partial \beta_j^* \, \partial \beta_m^{*\prime}} \right) - \left(\frac{1}{\eta_{it|t-1}^{(q)}(l)} \right) \times \right. \right. \\ & \left. \left(\frac{\partial \eta_{it|t-1}^{(q)}(l)}{\partial \beta_j^*} \right) \left(\frac{\partial \eta_{it|t-1}^{(q)}(l)}{\partial \beta_m^*} \right) \right\} - \left(\frac{1 - \sum_{r=1}^{K} y_{itr}}{1 - \sum_{r=1}^{K} \eta_{it|t-1}^{(r)}(l)} \right) \times \\ & \left\{ \frac{\partial^2 \eta_{it|t-1}^{(q)}(l)}{\partial \beta_j^* \, \partial \beta_m^{*\prime}} - \left(\frac{1}{1 - \sum_{r=1}^{K} \eta_{it|t-1}^{(r)}(l)} \right) \left(\frac{\partial \eta_{it|t-1}^{(q)}(l)}{\partial \beta_j^*} \right) \times \right. \\ & \left. \left. \sum_{r=1}^{K} \left(\frac{\partial \eta_{it|t-1}^{(r)}(l)}{\partial \beta_m^*} \right) \right\} \right]. \end{aligned}$$

$$\frac{\partial^2 \eta^{(q)}_{it|t-1}(l)}{\partial \beta^*_j \, \partial \beta^{*\prime}_m} = \begin{cases} A_{ijt}\, x^*_{it} x^{*\prime}_{it} - \left[\rho_{jj}\, A_{ij,t-1} - \sum_{l \neq j}^{K} \rho_{jl}\, B_{ilj,t-1} \right] x^*_{i,t-1}\, x^{*\prime}_{i,t-1}\,; & \text{for } q = j = m \\ -B_{iqjt}\, x^*_{it}\, x^{*\prime}_{it} - \left[\rho_{qj}\, A_{ij,t-1} - \sum_{l \neq j}^{K} \rho_{ql}\, B_{ilj,t-1} \right] x^*_{i,t-1}\, x^{*\prime}_{i,t-1}\,; & \text{for } q \neq j = m \\ -B_{imjt}\, x^*_{it}\, x^{*\prime}_{it} + \Bigg[\rho_{jj}\, B_{imj,t-1} + \rho_{jm}\, B_{ijm,t-1} & \\ \qquad\qquad -2 \sum_{l \neq j,m}^{K} \rho_{jl}\, D_{ijlm,t-1} \Bigg] x^*_{i,t-1}\, x^{*\prime}_{i,t-1}\,; & \text{for } q = j \neq m \\ -B_{ijmt}\, x^*_{it}\, x^{*\prime}_{it} + \Bigg[\rho_{mj}\, B_{imj,t-1} + \rho_{mm}\, B_{ijm,t-1} & \\ \qquad\qquad -2 \sum_{l \neq j,m}^{K} \rho_{ml}\, D_{ijlm,t-1} \Bigg] x^*_{i,t-1}\, x^{*\prime}_{i,t-1}\,; & \text{for } q = m \neq j \\ 2\, D_{ijqmt}\, x^*_{it}\, x^{*\prime}_{it} + \Bigg[\rho_{qj}\, B_{imj,t-1} + \rho_{qm}\, B_{ijm,t-1} & \\ \qquad\qquad -2 \sum_{l \neq j,m}^{K} \rho_{ql}\, D_{ijlm,t-1} \Bigg] x^*_{i,t-1}\, x^{*\prime}_{i,t-1}\,; & \text{for } q \neq j \neq m \end{cases}$$

where

$$A_{ijt} = \pi^{(j)}_{it} \left(1 - \pi^{(j)}_{it}\right) \left(1 - 2\pi^{(j)}_{it}\right)$$

$$B_{iljt} = \pi^{(l)}_{it}\, \pi^{(j)}_{it} \left(1 - 2\, \pi^{(j)}_{it}\right)$$

$$D_{ijlmt} = \pi^{(j)}_{it}\, \pi^{(l)}_{it}\, \pi^{(m)}_{it}$$

2.1.2.2 Formulas for the Second Derivatives With Respect to ρ

Similarly to compute the second derivative matrix in (2.48), we note from (2.46) that

$\frac{\partial \eta^{(j)}_{it|t-1}(l)}{\partial \rho_j}$ is free from ρ_j for any j. Thus, $\frac{\partial^2 \eta^{(j)}_{it|t-1}(l)}{\partial \rho_j \, \partial \rho_j'} = 0$. Consequently, we obtain

$$\frac{\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_j \, \partial \rho_j'} = \sum_{i=1}^{I} \sum_{t=2}^{T} \sum_{j=1}^{K} \left[\left\{ \frac{y_{itj}}{\left(\eta^{(j)}_{it|t-1}(l)\right)^2} \right\} \left\{ - \left(\frac{\partial \eta^{(j)}_{it|t-1}(l)}{\partial \rho_j} \right)^2 \right\} - \left\{ \frac{1 - \sum_{q=1}^{K} y_{itq}}{\left(1 - \sum_{q=1}^{K} \eta^{(q)}_{it|t-1}(l)\right)^2} \right\} \left\{ \left(\frac{\partial \eta^{(j)}_{it|t-1}(l)}{\partial \rho_j} \right)^2 \right\} \right], \quad (2.50)$$

for $j = 1, \cdots, K$, and

$$\frac{\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \rho_j \, \partial \rho_m'} = 0 \quad (2.51)$$

for $j \neq m$; $j, m = 1, \cdots, K$. This completes the computation for the second order derivatives needed for (2.48).

Note that we have shown in (2.48) how to obtain the MLE for $\boldsymbol{\rho}$ parameters in a standard fashion. However, it is understandable that solution by (2.48) requires the knowledge of the range restrictions for $\boldsymbol{\rho}$ which are functions of the marginal probabilities containing the given design covariates and $\boldsymbol{\beta}$ parameters. But, finding these restrictions, unlike in the binary case, may be cumbersome, which we do not emphasize much any way as our main intension is to deal with a more robust model that we present in Chapter 3 and 4.

2.1.2.3 Asymptotic Properties of the Likelihood Regression Estimator

Let $\hat{\beta}_{ML} = (\hat{\beta}^{*\prime}_{1,ML}, \cdots, \hat{\beta}^{*\prime}_{j,ML}, \cdots, \hat{\beta}^{*\prime}_{K,ML})'$ and $\hat{\rho}_{ML} = (\hat{\rho}'_{1,ML}, \cdots, \hat{\rho}'_{j,ML}, \cdots, \hat{\rho}'_{K,ML})'$ be the likelihood estimates for β and ρ, respectively.

For known ρ, it follows from likelihood iterative equation (2.47) for β that

$$\hat{\beta}_{ML} \xrightarrow{p} \beta \tag{2.52}$$

This is because $E\left(\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta}\right) = 0$ in (2.47) and when covariates are bounded, $E\left(\frac{\partial^2 \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta \, \partial \beta'}\right)$ in (2.49) is finite. To show $E\left(\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta}\right) = 0$, we note that

$$E(Y_{i1j}) = \pi_{i1}^{(j)}$$

and

$$E(Y_{itq} | Y_{i,t-1,q}) = \eta_{it|t-1}^{(j)}$$

leading to $E\left(\frac{\partial \ln g_i(\boldsymbol{\beta})}{\partial \beta_j}\right) = 0$ and $E\left(\frac{\partial \ln h_i(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta_j}\right) = 0$ by (2.41)-(2.42). Thus by (2.39), one obtains

$$E\left(\frac{\partial \ln L(\boldsymbol{\beta}, \boldsymbol{\rho})}{\partial \beta}\right) = 0.$$

It also follows that conditional on the history, the variance of the estimator $\hat{\beta}_{ML}$ is given by

$$\begin{aligned} Var(\hat{\beta}_{ML}) &= \left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)^{-1} Var\left(\frac{\partial \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta}\right)\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)^{-1} \\ &= \left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)^{-1} E\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)^{-1}, \qquad (2.53) \end{aligned}$$

which converges to

$$\begin{aligned} \lim_{I\to\infty} Var(\hat{\beta}_{ML}) &= \left[E\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)\right]^{-1} E\left[Var\left(\frac{\partial \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta}\right)\right]\left[E\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)\right]^{-1} \\ &= \left[E\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)\right]^{-1}. \qquad (2.54) \end{aligned}$$

Thus, asymptotically as $I \to \infty$, $\hat{\beta}_{ML}$ follows a Gaussion distribution with mean β [by (2.52)] and covariance matrix $\left[E\left(\frac{\partial^2 \ln L(\boldsymbol{\beta},\boldsymbol{\rho})}{\partial\beta\,\partial\beta'}\right)\right]^{-1}$.

Note that the derivation of the above asymptotic properties is quite standard [see for example, Rao (1973)]. Also we would be able to derive the asymptotic properties of similar likelihood estimators in the future chapters by using such standard approach. However, throughout the thesis, we will concentrate on the finite sample performances of the likelihood estimators, which is more practical to examine.

2.2 Conditional Contingency Table Based Inferences

Note that in the last section we have dealt with non-stationary multinomial models where covariates were considered to be time dependent for a given individual. There may be a simpler situation where this covariates are time independent and consequently for a given combination level of multiple covariates, the same stationary multinomial response probability will be obtained for a group of individuals. For simplicity, we now consider only one covariate with $p+1$ levels, and use $x_{i(1)}, \cdots, x_{i(p)}$ as in (1.11) to represent the $p+1$ levels of the covariate for the ith individual. Suppose that under the uth level of the covariate there are $I_{(u,1)}^{(j)}$ individuals belonging to the jth category at time $t=1$ with the same probability. These marginal counts at time $t=1$ are shown in Table 2.2(a) for convenience. Under this circumstance, for an individual i belonging to this group of $I_{(u,1)}^{(j)}$ individuals, the marginal probability in (2.3) may be written as

$$\begin{aligned} P\left(Y_{i1}=y_{i1}^{(j)} \,|\, x_{i(1)}, \cdots, x_{i(p)};\, i \in I_{(u,1)}^{(j)}\right) &= \frac{\exp\left(\beta_{j0}+\beta_{ju}\right)}{1+\sum_{l=1}^{K}\exp\left(\beta_{l0}+\beta_{lu}\right)}; \quad u=1,\cdots,p+1 \\ &= \pi_{(u)}^{(j)} \text{ say}, \qquad j=1,\cdots,K \end{aligned} \tag{2.55}$$

and $\pi_{(u)}^{(K+1)} = \left(1-\sum_{j=1}^{K}\pi_{(u)}^{(j)}\right)$. Note that in (2.55), $\beta_{j,p+1}=0$. Further note that this probability in (2.55) is a direct generalization of the binary marginal probability at $T = 1$ (cross-sectional) shown in (1.11). By the same token, under the uth level of the covariate, let $I_{lj(u,t|t-1)}$ be the number of individuals transmitted from lth category at time $t-1$ to the jth category at time t. Consequently under the uth level of the covariate, for an individual i; $i \in I_{lj(u,t|t-1)}$, the conditional multinomial probability in (2.5) may be written as

$$
\begin{aligned}
P\left(Y_{i1} = y_{i1}^{(j)} \,|\, Y_{i,t-1} = y_{i,t-1}\,;\, i \in I_{lj(u,t|t-1)}\right) &= \pi_{(u)}^{(j)} + \rho_{jl}\left(1 - \pi_{(u)}^{(l)}\right) - \sum_{k \neq l}^{K} \rho_{jk}\,\pi_{(u)}^{(k)} \\
&= \pi_{(u)}^{(j)} + \rho_{jl} - \sum_{k=1}^{K} \rho_{jk}\,\pi_{(u)}^{(k)} \\
&= \eta_{(u,t|t-1)}^{(j)}(l), \text{ say;} \qquad (2.56)
\end{aligned}
$$

where $j = 1, \cdots, K$; $u = 1, \cdots, p+1$ and

$$
\eta_{(u,t|t-1)}^{(K+1)}(l) = \left(1 - \sum_{j=1}^{K} \eta_{(u,t|t-1)}^{(j)}(l)\right). \qquad (2.57)
$$

For convenience, for the uth level of the covariate, we now summarize the multinomial marginal probabilities (2.55) and corresponding counts in Tables 2.1(a) and 2.2(a), respectively. Similarly, the conditional probabilities (2.56)-(2.57) and corresponding counts are summarized in Tables 2.1(b) and 2.2(b), respectively.

The likelihood estimation for the parameters involved in the model (2.55)-(2.57) is discussed in the next section.

Table 2.1: *Marginal and Conditional Probability Tables in the Longitudinal Setup for the Multinomial Response Model at the uth Level of the Covariate for* $u = 1, \cdots, p+1$.

Table 2.1(a): Marginal

For any time t

Category	1	$\cdots$	j	$\cdots$	K+1	Total
Counts	$\pi_{(u)}^{(1)}$	$\cdots$	$\pi_{(u)}^{(j)}$	$\cdots$	$\pi_{(u)}^{(K+1)}$	1.0

Table 2.1(b): Conditional

Time $t = 2, \cdots, T$

	Category	1	$\cdots$	j	$\cdots$	K+1	Total
	1	$\eta_{(u,t\|t-1)}^{(1)}(1)$	$\cdots$	$\eta_{(u,t\|t-1)}^{(j)}(1)$	$\cdots$	$\eta_{(u,t\|t-1)}^{(K+1)}(1)$	1.0
Time	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
t-1	l	$\eta_{(u,t\|t-1)}^{(1)}(l)$	$\cdots$	$\eta_{(u,t\|t-1)}^{(j)}(l)$	$\cdots$	$\eta_{(u,t\|t-1)}^{(K+1)}(l)$	1.0
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	K+1	$\eta_{(u,t\|t-1)}^{(1)}(K+1)$	$\cdots$	$\eta_{(u,t\|t-1)}^{(j)}(K+1)$	$\cdots$	$\eta_{(u,t\|t-1)}^{(K+1)}(K+1)$	1.0

Table 2.2: *Conditional Contingency Tables in the Longitudinal Setup for Multinomial Response Model at the uth Level of the Covariate for* $u = 1, \cdots, p+1$.

Table 2.2(a): Marginal

For any time t

Category	1	$\cdots$	j	$\cdots$	K+1	Total
Counts	$I_{(u,t)}^{(1)}$	$\cdots$	$I_{(u,t)}^{(j)}$	$\cdots$	$I_{(u,t)}^{(K)}$	$I_{(u,t)} = I_{(u)}$

Table 2.2(b): Conditional

Time $t = 2, \cdots, T$

	Category	1	$\cdots$	j	$\cdots$	K+1	Total
	1	$I_{11(u,t\|t-1)}$	$\cdots$	$I_{1j(u,t\|t-1)}$	$\cdots$	$I_{1,K+1(u,t\|t-1)}$	$I_{(u,t-1)}^{(1)}$
Time	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
t-1	l	$I_{l1(u,t\|t-1)}$	$\cdots$	$I_{lj(u,t\|t-1)}$	$\cdots$	$I_{l,K+1(u,t\|t-1)}$	$I_{(u,t-1)}^{(l)}$
	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
	K+1	$I_{K+1,1(u,t\|t-1)}$	$\cdots$	$I_{K+1,j(u,t\|t-1)}$	$\cdots$	$I_{K+1,K+1(u,t\|t-1)}$	$I_{(u,t-1)}^{(K+1)}$
	Total	$I_{(u,t)}^{(1)}$	$\cdots$	$I_{(u,t)}^{(j)}$	$\cdots$	$I_{(u,t)}^{(K+1)}$	$I_{(u)}$

2.2.1 Product Multinomial Likelihood Inferences

It is clear from the last section that at uth ($u = 1, \cdots, p+1$) level of the covariate there are $I_{(u)}$ individuals with their distribution as in Table 2.2 with corresponding probabilities as in Table 2.1. Let $L_{(u)}$ denote the likelihood for the $I_{(u)}$ individuals which by using the notations from Tables 2.1 and 2.2, has the form

$$L_{(u)} = f_{(u,1)} \prod_{t=2}^{T} \prod_{l=1}^{K+1} \left[f_{(u,t\,|\,t-1)}(l) \right] \tag{2.58}$$

where $f_{(u,1)}$ is the marginal multinomial probability at time $t = 1$ given by

$$f_{(u,1)} = \prod_{j=1}^{K+1} \frac{I_{(u)}!}{I_{(u,1)}^{(j)}!} \left\{ \pi_{(u)}^{(j)} \right\}^{I_{(u,1)}^{(j)}}, \tag{2.59}$$

and $f_{(u,t|t-1)}(l)$ is the conditional multinomial probability at time t given that the response was in the lth category at time $t-1$. This conditional distribution has the formula

$$f_{(u,t|t-1)}(l) = \prod_{j=1}^{K+1} \frac{I_{(u,t-1)}^{(l)}!}{I_{lj(u,t|t-1)}!} \left\{ \eta_{(u,t|t-1)}^{(j)}(l) \right\}^{I_{lj(u,t|t-1)}} \tag{2.60}$$

Next, because the $p+1$ levels of a single covariate are mutually exclusive, we may now write the overall likelihood function as

$$L(\beta, \rho) = \prod_{u=1}^{p+1} L_{(u)}, \tag{2.61}$$

yielding the log likelihood function given by

$$\ln L(\beta,\rho) \;=\; C + \sum_{u=1}^{p+1}\Big[\, g_u^*(\beta) \,+\, h_u^*(\beta,\,\rho)\Big] \tag{2.62}$$

where for $u = 1, \cdots, p+1$,

$$g_u^*(\beta) = \sum_{j=1}^{K} I_{(u,1)}^{(j)}\Big(\beta_{j0}+\beta_{ju}\Big) - I_{(u)}\,\ln\left\{1+\sum_{l=1}^{K}\exp\Big(\beta_{l0}+\beta_{lu}\Big)\right\} \tag{2.63}$$

and

$$h_u^*(\beta,\,\rho) = \sum_{t=2}^{T}\sum_{l=1}^{K+1}\sum_{j=1}^{K+1} I_{lj(u,t|t-1)}\,\ln\left\{\pi_{(u)}^{(j)} + \rho_{j\,l} - \sum_{k=1}^{K}\rho_{j\,k}\,\pi_{(u)}^{(k)}\right\} \tag{2.64}$$

and C is a normalizing constant.

For convenience of writing a single equation for the derivatives of all β_{j0} and β_{ju} ($u = 1, \cdots, p$), we use $u = 0, \cdots, p$ to represent them. Thus, we now write the likelihood estimating equations for β_{ju} for all $j = 1, \cdots, K$ and $u = 0, \cdots, p$ as

$$\frac{\partial \ln L(\beta,\rho)}{\partial \beta_{ju}} \;=\; \begin{cases} \displaystyle\sum_{u=1}^{p+1}\left[\frac{\partial\, g_u^*(\beta)}{\partial \beta_{j0}} + \frac{\partial\, h_u^*(\beta,\,\rho)}{\partial \beta_{j0}}\right] = 0\,, \quad \text{for} \quad u=0 \\[2ex] \left[\dfrac{\partial\, g_u^*(\beta)}{\partial \beta_{ju}} + \dfrac{\partial\, h_u^*(\beta,\,\rho)}{\partial \beta_{ju}}\right] = 0\,, \quad \text{for} \quad u=1,\cdots,p. \end{cases} \tag{2.65}$$

Next for $j = 1, \cdots, K$ and $l = 0, \cdots, K$, we write the likelihood estimating equations for ρ_{jl} as

$$\frac{\partial \ln L(\beta, \rho)}{\partial \rho_{jl}} = \sum_{u=1}^{p+1} \left[\frac{\partial g_u^*(\beta)}{\partial \rho_{jl}} + \frac{\partial h_u^*(\beta, \rho)}{\partial \rho_{jl}} \right] = 0. \tag{2.66}$$

The derivatives in (2.65) have the formulas as

$$\frac{\partial g_u^*(\beta)}{\partial \beta_{ju}} = \left[I_{(u,1)}^{(j)} - \pi_{(u)}^{(j)} \right]; \; j = 1, \cdots, K; \; u = 0, \cdots, p \tag{2.67}$$

and

$$\frac{\partial h_u^*(\beta, \rho)}{\partial \beta_{ju}} = \sum_{t=2}^{T} \sum_{l=1}^{K+1} \sum_{m=1}^{K+1} \left[\frac{I_{lm(u,t|t-1)}}{\eta_{(u,t|t-1)}^{(m)}(l)} \frac{\partial \eta_{(u,t|t-1)}^{(m)}(l)}{\partial \beta_{ju}} \right], \tag{2.68}$$

where

$$\frac{\partial \eta_{(u,t|t-1)}^{(m)}(l)}{\partial \beta_{ju}} = \begin{cases} \pi_{(u)}^{(j)} \left(1 - \pi_{(u)}^{(j)}\right) (1 - \rho_{jj}) + \sum_{k \neq j}^{K} \rho_{jk} \pi_{(u)}^{(k)} \pi_{(u)}^{(j)}, & \text{for } m = j \\ -\pi_{(u)}^{(m)} \pi_{(u)}^{(j)} - \rho_{mj} \pi_{(u)}^{(j)} \left(1 - \pi_{(u)}^{(j)}\right) + \sum_{k \neq j}^{K} \rho_{mk} \pi_{(u)}^{(k)} \pi_{(u)}^{(j)}, & \text{for } m \neq j \end{cases} \tag{2.69}$$

Similarly, the derivatives in (2.66) have the formulas

$$\frac{\partial g_u^*(\beta)}{\partial \rho_{jl}} = 0 \tag{2.70}$$

$$\frac{\partial h_u^*(\beta, \rho)}{\partial \rho_{jl}} = \sum_{t=2}^{T}\sum_{l=1}^{K+1}\sum_{m=1}^{K+1}\left[\frac{I_{lm(u,t|t-1)}}{\eta^{(m)}_{(u,t|t-1)}(l)}\frac{\partial \eta^{(m)}_{(u,t|t-1)}(l)}{\partial \rho_{jl}}\right], \tag{2.71}$$

where

$$\frac{\partial \eta^{(m)}_{(u,t|t-1)}(l)}{\partial \rho_{jl}} = \begin{cases} \left(1-\pi^{(l)}_{(u)}\right); & \text{for} \quad m = j \\ -\pi^{(l)}_{(u)}; & \text{for} \quad m \neq j \end{cases} \tag{2.72}$$

For convenience of writing the iterative equation, we denote the regression parameters, $\boldsymbol{\beta}$ as follows

$$\boldsymbol{\beta} \equiv \left(\tilde{\boldsymbol{\beta}}_0', \tilde{\boldsymbol{\beta}}_1', \cdots, \tilde{\boldsymbol{\beta}}_u', \cdots, \tilde{\boldsymbol{\beta}}_p'\right)' \; : \; K(p+1)\times 1$$

where

$$\tilde{\boldsymbol{\beta}}_u = (\beta_{1u}, \cdots, \beta_{ju}, \cdots, \beta_{Ku})', \;\text{ for } u = 0, \cdots, p. \tag{2.73}$$

Similarly the dynamic dependence parameters $\boldsymbol{\rho}$'s are expressed as

$$\boldsymbol{\rho} \equiv \left(\boldsymbol{\rho}_1', \cdots, \boldsymbol{\rho}_j', \cdots, \boldsymbol{\rho}_K'\right)' \; : \; K^2 \times 1$$

where

$$\boldsymbol{\rho}_j' = (\rho_{j1}, \cdots, \rho_{jl}, \cdots, \rho_{jK}). \tag{2.74}$$

Hence we write the the Newton-Raphson iterative equation for the likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$ as

$$\hat{\boldsymbol{\beta}}_{(r+1)} = \hat{\boldsymbol{\beta}}_{(r)} + \left\{ \left[\frac{\partial^2 \ln L(\beta, \rho)}{\partial \beta \, \partial \beta'} \right]^{-1} \frac{\partial \ln L(\beta, \rho)}{\partial \beta} \right\}_{(r)} \tag{2.75}$$

and

$$\hat{\boldsymbol{\rho}}_{(r+1)} = \hat{\boldsymbol{\rho}}_{(r)} + \left\{ \left[\frac{\partial^2 \ln L(\beta, \rho)}{\partial \rho \, \partial \rho'} \right]^{-1} \frac{\partial \ln L(\beta, \rho)}{\partial \rho} \right\}_{(r)} \tag{2.76}$$

where the elements of $\frac{\partial \ln L(\beta, \rho)}{\partial \beta}$ and $\frac{\partial \ln L(\beta, \rho)}{\partial \rho}$ can be obtained from (2.65) and (2.66) respectively. Furthermore, $\hat{\boldsymbol{\beta}}_{(r)}$ in (2.75) and $\hat{\boldsymbol{\rho}}_{(r)}$ in (2.76) the rth iterative value for $\boldsymbol{\beta}$ and $\boldsymbol{\rho}$, respectively. Also, in both (2.75) and (2.76), $\{ \}_{(r)}$ represents that the quantity in $\{ \}$ is evaluated at $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_{(r)}$ and $\boldsymbol{\rho} = \hat{\boldsymbol{\rho}}_{(r)}$, respectively.

The formulas for the second order derivatives in (2.75) and (2.76) are given in Section 2.2.1.1 and 2.2.1.2, respectively.

2.2.1.1 Formulas for the Second Order Derivatives With Respect to β

Note that to compute the second derivative matrix in (2.75), it is sufficient to compute the following four derivatives

$$\frac{\partial^2 \ln L(\beta,\rho)}{\partial\beta_{j0}\,\partial\beta_{j'0}} = \begin{cases} \sum_{u=1}^{p+1}\left[-\pi_{(u)}^{(j)}\left\{1-\pi_{(u)}^{(j)}\right\} + h^{**}_{ujj00}(\beta,\,\rho)\right] \\ \\ \sum_{u=1}^{p+1}\left[\pi_{(u)}^{(j)}\pi_{(u)}^{(j')} + h^{**}_{ujj'00}(\beta,\,\rho)\right] \end{cases} \tag{2.77}$$

$$\frac{\partial^2 \ln L(\beta,\rho)}{\partial\beta_{jv}\,\partial\beta_{j'w}} = \begin{cases} -\pi_{(u)}^{(j)}\left[1-\pi_{(u)}^{(j)}\right] + h^{**}_{ujj0u}(\beta,\,\rho) \\ -\pi_{(u)}^{(j)}\left[1-\pi_{(u)}^{(j)}\right] + h^{**}_{ujjuu}(\beta,\,\rho) \\ \pi_{(u)}^{(j)}\pi_{(u)}^{(j')} + h^{**}_{ujj'uu}(\beta,\,\rho) \end{cases} \tag{2.78}$$

with

$$h^{**}_{ujj'vw}(\beta,\,\rho) = \sum_{t=2}^{T}\sum_{l=1}^{K+1}\sum_{m=1}^{K+1}\left[\frac{I_{lm(u,\,t|t-1)}}{\eta^{(m)}_{(u,\,t|t-1)}(l)}\left\{\frac{\partial^2\eta^{(m)}_{(u,\,t|t-1)}(l)}{\partial\beta_{jv}\,\partial\beta_{j'w}} - \frac{1}{\eta^{(m)}_{(u,\,t|t-1)}(l)} \times \left(\frac{\partial\eta^{(m)}_{(u,\,t|t-1)}(l)}{\partial\beta_{jv}}\right)\left(\frac{\partial\eta^{(m)}_{(u,\,t|t-1)}(l)}{\partial\beta_{j'w}}\right)\right\}\right]$$

where for any $v = 0$ or u; and $w = o$ or u, we may write

$$\frac{\partial^2 \eta^{(m)}_{(u,t|t-1)}(l)}{\partial \beta_{jv}\, \partial \beta_{j'w}} = \begin{cases} A^*_{uj}\,(1-\rho_{jj}) - \sum_{k \neq j}^{K} \rho_{j\,k}\, B^*_{ukj}\,; & \text{for } j = j' = m \\ -B^*_{umj} - \rho_{mj}\, A^*_{uj} + \sum_{k \neq j}^{K} \rho_{m\,k}\, B^*_{ukj}\,; & \text{for } j = j' \neq m \\ -B^*_{uj'j}\,(1-\rho_{jj}) + \rho_{jj'}\, B^*_{ujj'} - 2\sum_{k \neq j}^{K} \rho_{j\,k}\, D^*_{ukjj'}\,; & \text{for } j = m \neq j' \\ -B^*_{ujj'} - (\rho_{j'j} - \rho_{j'j'})\, B^*_{uj'j} + 2\sum_{k \neq j, j'}^{K} \rho_{j'\,k}\, D^*_{ukjj'}\,; & \text{for } j \neq j' = m \\ 2D^*_{umjj'} + (\rho_{mj} - \rho_{mj'})\, B^*_{ujj'} + 2\sum_{k \neq j, j'}^{K} \rho_{m\,k}\, D^*_{ukjj'}\,; & \text{for } m \neq j,\, j' \end{cases}$$

where

$$\begin{aligned} A^*_{uj} &= \pi^{(j)}_{(u)} \left(1 - \pi^{(j)}_{(u)}\right) \left(1 - 2\pi^{(j)}_{(u)}\right) \\ B^*_{ukj} &= \pi^{(k)}_{(u)} \pi^{(j)}_{(u)} \left(1 - 2\pi^{(j)}_{(u)}\right) \\ D^*_{ukjj'} &= \pi^{(k)}_{(u)}\, \pi^{(j)}_{(u)}\, \pi^{(j')}_{(u)} \end{aligned}$$

This completes the calculation for the second order derivatives needed for (2.75).

2.2.1.2 Formulas for the Second Order Derivatives With Respect to ρ

The computation for the second derivative matrix in (2.76) with respect to ρ is much simpler as compare to β. Note that by using (2.70) and (2.71) into (2.66), we write

$$\frac{\partial^2 \ln L(\beta,\rho)}{\partial \rho_{jl}\,\partial \rho_{j'l'}} = \sum_{u=1}^{p+1}\sum_{t=2}^{T}\sum_{q=1}^{K+1}\sum_{m=1}^{K+1}\Bigg[I_{qm(u,t|t-1)}\frac{\partial\{\eta^{(m)}_{(u,t|t-1)}(q)\}^{-1}}{\partial \rho_{j'l'}}\left(\frac{\partial \eta^{(m)}_{(u,t|t-1)}(q)}{\partial \rho_{jl}}\right) + \frac{I_{qm(u,t|t-1)}}{\eta^{(m)}_{(u,t|t-1)}(q)}\left(\frac{\partial^2 \eta^{(m)}_{(u,t|t-1)}(q)}{\partial \rho_{jl}\,\partial \rho_{j'l'}}\right)\Bigg]. \tag{2.79}$$

Because the first derivative in (2.72) is free from ρ, it follows that the second quantity within the square bracket [ ] in (2.79) is zero. Next, by using the first derivative from (2.72), the first term within [ ] in (2.79) can be calculated easily, yielding the final form as

$$\frac{\partial^2 \ln L(\beta,\rho)}{\partial \rho_{jl}\,\partial \rho_{j'l'}} = -\sum_{u=1}^{p+1}\sum_{t=2}^{T}\sum_{q=1}^{K+1}\sum_{m=1}^{K+1}\left[\frac{I_{qm(u,t|t-1)}}{\left\{\eta^{(m)}_{(u,t|t-1)}(q)\right\}^2}\left(\frac{\partial \eta^{(m)}_{(u,t|t-1)}(q)}{\partial \rho_{jl}}\right) \times \left(\frac{\partial \eta^{(m)}_{(u,t|t-1)}(q)}{\partial \rho_{j'l'}}\right)\right]. \tag{2.80}$$

This completes the calculations for the second order derivatives, namely, $\dfrac{\partial^2 \ln L(\beta,\rho)}{\partial \rho\,\partial \rho'}$ in (2.76).

Chapter 3

Multinomial Dynamic Fixed Logit Models

In Chapter 2, we introduced a conditional linear dynamic probability model for the analysis of multinomial longitudinal data under two situations, (1) when the categorical data are available from an individual over the whole period of study; (2) when individual identity is not recorded, rather, his/her categorical responses are recorded at time t conditioning on time $t-1$. Also, in the second situation, the covariates were stationary, i.e., time independent. This type of models however can not accommodate the longitudinal correlations with full range. That is, the range for the correlation index parameters, namely ρ_{jl} $(j = 1, \cdots, K;\ l = 1, \cdots, K)$ can be narrower than from -1 to 1. As a remedy to this range issue, there exists situations both in Economics

(Amemiya, 1985) and Statistics (Sutradhar, 2011, Chapter 7) literature using a conditional non- linear probability model for longitudinal binary data. In this chapter, we follow these studies and generalize the binary dynamic fixed logit (BDFL) model to the multinomial longitudinal setup. We refer to such a model as the multinomial dynamic fixed logit (MDFL) model.

In Section 3.1, we use this MDFL model for history based data. This means that multinomial responses are available from all individuals over the whole period of time, as in Section 2.1 of Chapter 2. The basic properties of the proposed MDFL model are also discussed in this section. In the same section, we provide the likelihood inferences for the parameters under the proposed history based longitudinal multinomial logit models non-stationary data. A simulation study and a real life example are also given in the same section. In Section 3.2, we consider longitudinal multinomial data in the conditional contingency table form and use the proposed MDFL models to fit such data. In the same section, we describe the product multinomial likelihood approach for the estimation of parameters of such MDFL models.

Note that under the longitudinal setup, a non-linear multinomial mixed model will be discussed in Chapter 4.

3.1 MDFL Model For the History Based Data

Recall from Section 2.1 in Chapter 2 that the marginal multinomial multinomial probability $\pi_{it}^{(j)}$ for $t = 1, \cdots, T$, has the formula

$$\pi_{it}^{(j)} = P(Y_{it} = y_{it}^{(j)}) = \frac{\exp\left(\beta_{j0} + x_{it}'\boldsymbol{\beta}_j\right)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + x_{it}'\boldsymbol{\beta}_k\right)} \tag{3.1}$$

[see (2.5) and (2.3)], whereas the conditional probability for $y_{it} = (y_{it1}, \cdots, y_{itj}, \cdots, y_{itK})'$ given $y_{i,t-1}$ [say $y_{i,t-1} = y_{i,t-1}^{(l)}$] was modeled by a linear dynamic relationship (2.5) denoted by $\eta_{it|t-1}^{(j)}(l)$, where $\rho_j = (\rho_{j1}, \cdots, \rho_{jl}, \cdots, \rho_{jK})'$ involved in $\eta_{it|t-1}^{(j)}(l)$ is a vector of linear dynamic dependence index parameters.

In contrast to that linear model (2.5), we now write a multinomial logit (non-linear) model consisting of the marginal probability at $t = 1$ given by

$$\begin{aligned} \tilde{\pi}_{i1}^{(j)} &= P(Y_{i1} = y_{i1}^{(j)}) \\ &= \frac{\exp\left(\beta_{j0} + x_{i1}'\boldsymbol{\beta}_j\right)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + x_{i1}'\boldsymbol{\beta}_k\right)}, \end{aligned} \tag{3.2}$$

as in (3.1), and for $t = 2, \cdots, T$, unlike in (2.5), a non-linear conditional probability given by

$$P(\boldsymbol{Y}_{it} = \boldsymbol{y}_{it}^{(j)} \,|\, \boldsymbol{Y}_{i,t-1} = \boldsymbol{y}_{i,t-1}^{(l)}) \;=\; \frac{\exp\left(\beta_{j0} + \sum_{u=1}^{p} \beta_{ju} x_{it(u)} + \sum_{c=1}^{K} \theta_{jc}\, y_{i,t-1,c}^{(l)}\right)}{\sum_{k=1}^{K+1} \exp\left(\beta_{k0} + \sum_{u=1}^{p} \beta_{ku} x_{it(u)} + \sum_{c=1}^{K} \theta_{kc}\, y_{i,t-1,c}^{(l)}\right)}$$

$$
\begin{aligned}
&= \frac{\exp\left(\beta_{j0} + x'_{it}\boldsymbol{\beta}_j + \boldsymbol{\theta}'_j \boldsymbol{y}^{(l)}_{i,t-1}\right)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + x'_{it}\boldsymbol{\beta}_k + \boldsymbol{\theta}'_k \boldsymbol{y}^{(l)}_{i,t-1}\right)} \\
&= \tilde{\eta}^{(j)}_{it|t-1}(l), \text{ say,} \quad \text{for } j, l = 1, \cdots, K. \quad (3.3)
\end{aligned}
$$

For the remaining cases, i.e., when $l = K + 1$, the conditional probability in (3.3) reduces to

$$
\begin{aligned}
P(Y_{it} = y^{(j)}_{it} \,|\, Y_{i,t-1} = y^{(K+1)}_{i,t-1}) &= \frac{\exp\left(\beta_{j0} + x'_{it}\boldsymbol{\beta}_j\right)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + x'_{it}\boldsymbol{\beta}_k\right)} \\
&= \tilde{\eta}^{(j)}_{it|t-1}(K+1), \text{ say,} \quad \text{for } j = 1, \cdots, K, \quad (3.4)
\end{aligned}
$$

and for any $l = 1, \cdots, K$, the probability for the response to be in the last category $K + 1$ (i.e., for $j = K + 1$) at time t is given by

$$
\tilde{\eta}^{(K+1)}_{it|t-1}(l) = 1 - \sum_{j=1}^{K} \tilde{\eta}^{(j)}_{it|t-1}(l) = \frac{1}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + x'_{it}\boldsymbol{\beta}_k + \boldsymbol{\theta}'_k \boldsymbol{y}^{(l)}_{i,t-1}\right)}. \quad (3.5)
$$

Note that $\tilde{\pi}^{(j)}_{i1} = \pi^{(j)}_{i1}$ as in (2.3). However, the marginal probabilities, say $\tilde{\pi}^{(j)}_{it}$, for $t = 2, \cdots, T$, under the proposed conditional model (3.2)-(3.5) will have a recursive relationship relating $\tilde{\pi}^{(j)}_{it}$ and $\tilde{\pi}^{(j)}_{i,t-1}$. As far as the parameters are concerned, $\boldsymbol{\beta} \equiv (\beta_{j0}, \beta_{ju})$ for $j = 1, \cdots, K$ and $u = 1, \cdots, p$, are the same regression parameters as in (2.3), but $\boldsymbol{\theta} \equiv (\theta_{jl})$ for $j, l = 1, \cdots, K$ in (3.2)-(3.5) are referred to as the dynamic dependence parameters, whereas $\boldsymbol{\rho} \equiv (\rho_{jl})$ in (2.5) are correlation index parameters.

We use different notation for the dynamic dependence parameters in this non-linear multinomial model (3.3) as compared to (2.5). This is because, $\boldsymbol{\theta}$ in (3.2)-(3.5) is not restricted for its range unlike $\boldsymbol{\rho}$ under the model (2.5). More clearly θ_{jl} ranges from $-\infty$ to ∞, yielding the correlation between $\boldsymbol{y}_{itj}$ and $\boldsymbol{y}_{i,t-1,l}$ from -1 to 1. This would be clear from the basic properties of the model which is given in Section 3.1.1. Some authors such as De Rooij (2011) has used a non-linear dynamic model to analyze repeated multinomial data, which is, however, similar but different than our model (3.2)-(3.5). The difference lies in the fact that De Rooij (2011), unlike (3.2)-(3.5), uses an exponent of squared distance function in dynamic variables to define the conditional multinomial probability. This model is extremely complicated leading to very complicated computation for the correlations.

Further note that in time series setup, i.e., when $T \rightarrow \infty$ and $I = 1$, this multinomial logit model (3.2)-(3.5) has been recently studied by Loredo-Osti and Sutradhar (2011). See also Fahrmeir and Kaufmann (1987), and Fokianos and Kedem (2003). Thus, basic properties of this model for $i = 1$ to be discussed below will be the same as in Loredo-Osti and Sutradhar (2011). Nevertheless, as in longitudinal set up T is small such as T=3 or 4, we provide these properties for T up to 4 in Section 3.1.1 by using directly the conditioning and un-conditioning principles.

3.1.1 Basic Properties of the History Based MDFL Model

Note that the MDFL model given by (3.2)-(3.5) is referred to as a history based model. This is because, this model accommodates all responses and covariates information for all individuals ($i = 1, \cdots, I$) over the whole duration of the study for $t = 1, \cdots, T$. We now provide the means and the variances under this model in Lemma 3.1 and the covariances between multinomial responses at any two time points in Lemma 3.2.

Lemma 3.1: For $i = 1, \cdots, I$ and $t = 1, \cdots, T$, the unconditional mean vector and the covariance matrix of the multinomial response vector $\mathbf{Y}_{it} = (Y_{it1}, \cdots, Y_{itj}, \cdots, Y_{itK})$ have the forms

$$\begin{aligned} E(\mathbf{Y}_{it}) &= \boldsymbol{\eta}_{it|t-1} + \left[W_{it} - \boldsymbol{\eta}_{it|t-1} 1'\right] \tilde{\mathbf{\Pi}}_{i,t-1} \\ &= \left(\tilde{\pi}_{it}^{(1)}, \cdots, \tilde{\pi}_{it}^{(j)}, \cdots, \tilde{\pi}_{it}^{(K)}\right)' \\ &= \tilde{\mathbf{\Pi}}_{it} \end{aligned} \tag{3.6}$$

and

$$Var(\mathbf{Y}_{it}) = diag\left[\tilde{\pi}_{it}^{(1)}, \cdots, \tilde{\pi}_{it}^{(j)}, \cdots, \tilde{\pi}_{it}^{(K)}\right] - \tilde{\mathbf{\Pi}}_{it} \tilde{\mathbf{\Pi}}_{it}', \tag{3.7}$$

for all $j = 1, \cdots, K$ and $t = 1, \cdots, T$.

Proof: In the time series setup, a direct proof of this lemma is available from Loredo-Osti and Sutradhar (2011). Nevertheless, we verify this result for some of the smaller lags which is practically useful. These are shown in the Appendix A (page 135). $\diamondsuit$

Lemma 3.2: For all $i = 1, \cdots, I$, the covariance between the multinomial response vector $\boldsymbol{Y}_{it}$ and $\boldsymbol{Y}_{it'}$ for $t = 1, \cdots, t'-1$; $t' = 2, \cdots, T$ is given by

$$Cov\left(\boldsymbol{Y}_{it}, \boldsymbol{Y}_{it'}\right) = Var\left(\boldsymbol{Y}_{it}\right) \prod_{s=t+1}^{t'} \left[\boldsymbol{W}_{is} - \boldsymbol{\eta}_{is|s-1} \mathbf{1}'\right], \quad t < t' \tag{3.8}$$

where

$$\boldsymbol{W}_{is} = \begin{bmatrix} \tilde{\eta}^{(1)}_{is|s-1}(1) & \cdots & \tilde{\eta}^{(1)}_{is|s-1}(l) & \cdots & \tilde{\eta}^{(1)}_{is|s-1}(K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\eta}^{(j)}_{is|s-1}(1) & \cdots & \tilde{\eta}^{(j)}_{is|s-1}(l) & \cdots & \tilde{\eta}^{(j)}_{is|s-1}(K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\eta}^{(K)}_{is|s-1}(1) & \cdots & \tilde{\eta}^{(K)}_{is|s-1}(l) & \cdots & \tilde{\eta}^{(K)}_{is|s-1}(K) \end{bmatrix}, \quad \boldsymbol{\eta}_{is|s-1} = \begin{bmatrix} \tilde{\eta}^{(1)}_{is|s-1}(K+1) \\ \vdots \\ \tilde{\eta}^{(j)}_{is|s-1}(K+1) \\ \vdots \\ \tilde{\eta}^{(K)}_{is|s-1}(K+1) \end{bmatrix}$$

Proof: Similar to the proof for Lemma 3.1, a direct proof for this lemma under time series setup is given in Loredo-Osti and Sutradhar (2011). We, however, provide a detailed induction based proof of the lemma in the Appendix A (page 142). ◇

3.1.1.1 Understanding the Correlations Through Special Cases

Trinomial (K = 2) and Binary (K = 1) Cases:

For the special trinomial case when $K = 2$, we use (3.8) and write

$$
\begin{aligned}
Cov\left(\mathbf{Y}_{i1}, \mathbf{Y}_{i2}\right) &= Var\left(\mathbf{Y}_{i1}\right)\left[\boldsymbol{W}_{i2} - \boldsymbol{\eta}_{i2|1}\mathbf{1}'\right] \\
&= \begin{bmatrix} \tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)}) & -\tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)} \\ -\tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)} & \tilde{\pi}_{i1}^{(2)}(1-\tilde{\pi}_{i1}^{(2)}) \end{bmatrix} \\
&\qquad \left\{\begin{bmatrix} \tilde{\eta}_{i2|1}^{(1)}(1) & \tilde{\eta}_{i2|1}^{(1)}(2) \\ \tilde{\eta}_{i2|1}^{(2)}(1) & \tilde{\eta}_{i2|1}^{(2)}(2) \end{bmatrix} - \begin{bmatrix} \tilde{\eta}_{i2|1}^{(1)}(3) & \tilde{\eta}_{i2|1}^{(1)}(3) \\ \tilde{\eta}_{i2|1}^{(2)}(3) & \tilde{\eta}_{i2|1}^{(2)}(3) \end{bmatrix}\right\} \\
&= \begin{bmatrix} \tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)}) & -\tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)} \\ -\tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)} & \tilde{\pi}_{i1}^{(2)}(1-\tilde{\pi}_{i1}^{(2)}) \end{bmatrix} \begin{bmatrix} \tilde{\eta}_{i2|1}^{(1)}(1)-\tilde{\eta}_{i2|1}^{(1)}(3) & \tilde{\eta}_{i2|1}^{(1)}(2)-\tilde{\eta}_{i2|1}^{(1)}(3) \\ \tilde{\eta}_{i2|1}^{(2)}(1)-\tilde{\eta}_{i2|1}^{(2)}(3) & \tilde{\eta}_{i2|1}^{(2)}(2)-\tilde{\eta}_{i2|1}^{(2)}(3) \end{bmatrix}
\end{aligned}
$$

Thus we can write the covariance and correlation between y_{i11} and y_{i21} as

$$
Cov(y_{i11}, y_{i21}) = \tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)})\left\{\tilde{\eta}_{i2|1}^{(1)}(1) - \tilde{\eta}_{i2|1}^{(1)}(3)\right\} - \tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)}\left\{\tilde{\eta}_{i2|1}^{(2)}(1) - \tilde{\eta}_{i2|1}^{(2)}(3)\right\}
$$

$$
\begin{aligned}
Corr(y_{i11}, y_{i21}) &= \frac{Cov(y_{i11}, y_{i21})}{\sqrt{Var(y_{i11})Var(y_{i21})}} \\
&= \frac{\tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)})\left\{\tilde{\eta}_{i2|1}^{(1)}(1) - \tilde{\eta}_{i2|1}^{(1)}(3)\right\} - \tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)}\left\{\tilde{\eta}_{i2|1}^{(2)}(1) - \tilde{\eta}_{i2|1}^{(2)}(3)\right\}}{\sqrt{\tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)})\,\tilde{\pi}_{i2}^{(1)}(1-\tilde{\pi}_{i2}^{(1)})}}
\end{aligned} \tag{3.9}
$$

Similarly when $t = 1$ and $t' = 2$, the correlation among y_{i11} and y_{i22}, for example, has the formula

$$\begin{aligned} Corr(y_{i11}, y_{i22}) &= \frac{Cov(y_{i11}, y_{i22})}{\sqrt{Var(y_{i11})Var(y_{i22})}} \\ &= \frac{\tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)})\left\{\tilde{\eta}_{i2|1}^{(1)}(2)-\tilde{\eta}_{i2|1}^{(1)}(3)\right\} - \tilde{\pi}_{i1}^{(1)}\tilde{\pi}_{i1}^{(2)}\left\{\tilde{\eta}_{i2|1}^{(2)}(2)-\tilde{\eta}_{i2|1}^{(2)}(3)\right\}}{\sqrt{\tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)})\,\tilde{\pi}_{i2}^{(2)}(1-\tilde{\pi}_{i2}^{(2)})}} \end{aligned} \tag{3.10}$$

Now, for the binary case when $K = 1$, the correlation between y_{i11} and y_{i22} in (3.9) will reduce to the following formula

$$Corr(y_{i11}, y_{i21}) = \sqrt{\left[\frac{\tilde{\pi}_{i1}^{(1)}(1-\tilde{\pi}_{i1}^{(1)})}{\tilde{\pi}_{i2}^{(1)}(1-\tilde{\pi}_{i2}^{(1)})}\right]} \times \left\{\tilde{\eta}_{i2|1}^{(1)}(1)-\tilde{\eta}_{i2|1}^{(1)}(2)\right\}, \tag{3.11}$$

which matches with the correlations discussed by Sutradhar and Farrell (2007, Eqn. (1.6), p.450). Note that because

$$\tilde{\eta}_{i2|1}^{(1)}(1) = \frac{\exp\left(\beta_{10} + x_{i2}'\beta_1 + \theta_{11}\right)}{1 + \exp\left(\beta_{10} + x_{i2}'\beta_1 + \theta_{11}\right)} \tag{3.12}$$

and

$$\tilde{\eta}_{i2|1}^{(1)}(2) = \frac{\exp\left(\beta_{10} + x_{i2}'\beta_1\right)}{1 + \exp\left(\beta_{10} + x_{i2}'\beta_1\right)} \tag{3.13}$$

by (3.3), for $-\infty < \theta_{11} < \infty$, it then follows that $0 < \tilde{\eta}^{(1)}_{i2|1}(1), \tilde{\eta}^{(1)}_{i2|1}(2) < 1$. Furthermore, when $\theta_{11} \rightarrow \infty$, $\tilde{\eta}^{(1)}_{i2|1}(1) \rightarrow 1$, yielding $\left\{\tilde{\eta}^{(1)}_{i2|1}(1) - \tilde{\eta}^{(1)}_{i2|1}(2)\right\}$ as a positive fraction. Similarly, for $\theta_{11} \rightarrow -\infty$, $\tilde{\eta}^{(1)}_{i2|1}(1) \rightarrow 0$, yielding $\left\{\tilde{\eta}^{(1)}_{i2|1}(1) - \tilde{\eta}^{(1)}_{i2|1}(2)\right\}$ as a negative fraction. Consequently it follows from (3.11) that

$$-1 < Corr(y_{i11}, y_{i21}) < 1$$

This full range property for the correlations appear to hold for any two components of a multinomial response vector over two different time points. When compared to the linear dynamic fixed probability model discussed in Chapter 2, it is natural that the present non-linear model has advantages over the linear dynamic model with regard to the ranges for the correlations. The inferences for this non-linear dynamic model is, however, not discussed in the literature. In the following section, we exploit the well-known likelihood approach for such inferences. In Section 3.1.3, we provide a simulation study to examine the finite sample performance of the likelihood approach. Also, a real-life data on 'Three Mile Island Stress-Level' is re-analyzed by using this likelihood approach, which was earlier analyzed by Fienberg et al. (1985) and Conaway (1989), for example.

3.1.2 Likelihood Estimation for the History Based MDFL Model

To understand the mean, variance and covariances under the MDFL model, we need to study the parameters of the model such as β_{j0}, β_{ju} and θ_{jc}. We use the maximum likelihood approach to estimate all parameters involved, even though the regression parameters β_{j0} and β_{ju} may be of primary interest.

3.1.2.1 Log-Likelihood Function

Now by writing $\boldsymbol{\beta} \equiv (\beta_{j0}, \beta_{ju})$ and $\boldsymbol{\theta} \equiv (\theta_{jc})$ for $j = 1, \cdots, K$, $c = 1, \cdots, K$, $u = 1, \cdots, p$, it follows from the model (3.2)-(3.5) that the likelihood function has the form

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{i=1}^{I} \left[f(y_{i1}) \prod_{t=2}^{T} f(y_{it} \mid y_{i,t-1}) \right] \tag{3.14}$$

where

$$\begin{aligned} f(y_{i1}) = & \frac{1}{y_{i11}! \cdots y_{i1K}!, (1 - \sum_{q=1}^{K} y_{i1q})!} \prod_{q=1}^{K} \left\{ \tilde{\pi}_{i1}^{(q)} \right\}^{y_{i1q}} \\ & \times \left\{ 1 - \sum_{q=1}^{K} \tilde{\pi}_{i1}^{(q)} \right\}^{(1 - \sum_{q=1}^{K} y_{i1q})}, \end{aligned} \tag{3.15}$$

and

$$f(y_{it} \mid y_{i,t-1}) = \frac{1}{y_{it1}! \cdots y_{itk}!, (1 - \sum_{q=1}^{K} y_{itq})!} \prod_{q=1}^{K} \left\{ \tilde{\eta}_{it|t-1}^{(q)} \right\}^{y_{itq}} \times \left\{ 1 - \sum_{q=1}^{K} \tilde{\eta}_{it|t-1}^{(q)} \right\}^{\left(1 - \sum_{q=1}^{K} y_{itq}\right)}, \quad (3.16)$$

where $\tilde{\eta}_{it|t-1}^{(q)}$ is written from (3.3) for $\tilde{\eta}_{it|t-1}^{(q)}(l)$ simply by writing $y_{i,t-1}$ for known $y_{i,t-1}^{(l)}$ without any loss of generality. Thus, by using (3.15) and (3.16) in (3.14), one writes the log likelihood as

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\theta}) = C + \sum_{i=1}^{I} \left[\tilde{g}_i(\boldsymbol{\beta}) + \tilde{h}_i(\boldsymbol{\beta}, \boldsymbol{\theta}) \right] \quad (3.17)$$

where

$$\tilde{g}_i(\boldsymbol{\beta}) = \sum_{q=1}^{K} y_{i1q} \left(\beta_{q0} + \boldsymbol{x}'_{i1} \boldsymbol{\beta}_q \right) - \ln \left\{ 1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + \boldsymbol{x}'_{i1} \boldsymbol{\beta} \right) \right\}, \quad (3.18)$$

and

$$\tilde{h}_i(\boldsymbol{\beta}, \boldsymbol{\theta}) = \sum_{t=2}^{T} \left[\sum_{q=1}^{K} y_{i1q} \left(\beta_{q0} + \boldsymbol{x}'_{it} \boldsymbol{\beta}_q + \boldsymbol{\theta}'_q y_{i,t-1} \right) - \ln \left\{ 1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + \boldsymbol{x}'_{it} \boldsymbol{\beta}_k + \boldsymbol{\theta}'_k y_{i,t-1} \right) \right\} \right], \quad (3.19)$$

with C is the normalizing constant.

3.1.2.2 Likelihood Estimating Equations

By maintaining the same notation for the vector of regression parameters as in Chapter 2, we write

$$\boldsymbol{\beta} \equiv \left(\boldsymbol{\beta}_1^{*\prime}, \cdots, \boldsymbol{\beta}_j^{*\prime}, \cdots, \boldsymbol{\beta}_K^{*\prime}\right)'$$

where

$$\boldsymbol{\beta}_j^* = (\beta_{j0}, \beta_{j1}, \cdots, \beta_{ju}, \cdots, \beta_{jp})'. \tag{3.20}$$

Similarly, we denote the (p+1)-dimensional vector of covariate as

$$\boldsymbol{x}_{it}^* = \left(1, x_{it(1)}, \cdots, x_{it(u)}, \cdots, x_{it(p)}\right)' = \left(1, \boldsymbol{x}_{it}'\right)'. \tag{3.21}$$

For the dynamic dependence parameter we use

$$\boldsymbol{\theta} \equiv \left(\boldsymbol{\theta}_1', \cdots, \boldsymbol{\theta}_j', \cdots, \boldsymbol{\theta}_K'\right)'$$

where

$$\boldsymbol{\theta}_j' = (\theta_{j1}, \cdots, \theta_{jl}, \cdots, \theta_{jK}). \tag{3.22}$$

We now write the likelihood estimating equations for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as in the following lemma.

Lemma 3.3: For the MDFL model (3.2)-(3.5), by using the log-likelihood function (3.17), we can write the likelihood estimating equation for $\boldsymbol{\alpha} = \left(\boldsymbol{\beta}', \boldsymbol{\theta}'\right)'$ as follows

$$\frac{\partial \ln L(\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \sum_{i=1}^{I}\sum_{t=1}^{T}\left[\boldsymbol{y}_{it} - \boldsymbol{\eta}_{it|t-1}\right] \otimes \begin{pmatrix} \boldsymbol{x}^*_{it} \\ \boldsymbol{y}_{i,t-1} \end{pmatrix} = 0 \qquad (3.23)$$

where for convenience we use $\boldsymbol{y}_{i0} = (0, \cdots, 0)'$ implying that $\tilde{\pi}^{(j)}_{i1} = \tilde{\eta}^{(j)}_{i1|0}$.

Proof: Using the notation from (3.20)-(3.22), we can write the log-likelihood function (3.17) as

$$\begin{aligned} \ln L(\boldsymbol{\alpha}) &= C + \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{q=1}^{K} y_{itq}\left(x^{*\prime}_{it}\beta^*_q + \theta'_q y_{i,t-1}\right) \\ &\quad - \sum_{i=1}^{I}\sum_{t=1}^{T}\ln\left[1 + \sum_{q=1}^{K}\exp\left(x^{*\prime}_{it}\beta^*_q + \theta'_q y_{i,t-1}\right)\right] \end{aligned} \qquad (3.24)$$

By taking derivatives of the log-likelihood function with respect to β^*_j and θ_j we obtain

$$\begin{aligned} \frac{\partial \ln L(\boldsymbol{\alpha})}{\partial \beta^*_j} &= \sum_{i,t} y_{itj}x_{it} - \sum_{i,t}\tilde{\eta}^{(j)}_{it|t-1}x^*_{it} \\ &= \sum_{i,t}\left[y_{itj} - \tilde{\eta}^{(j)}_{it|t-1}\right]x^*_{it} \end{aligned} \qquad (3.25)$$

$$\begin{aligned}\frac{\partial \ln L(\boldsymbol{\alpha})}{\partial \theta_j} &= \sum_{i,t} y_{itj} y_{i,t-1} - \sum_{i,t} \tilde{\eta}^{(j)}_{it|t-1} y_{i,t-1} \\ &= \sum_{i,t} \left[y_{itj} - \tilde{\eta}^{(j)}_{it|t-1} \right] y_{i,t-1} \end{aligned} \quad (3.26)$$

Thus, we can write the derivatives of the log-likelihood function with respect to β and θ in vector form as follows

$$\frac{\partial \ln L(\boldsymbol{\alpha})}{\partial \beta} = \sum_{i,t} \left[y_{it} - \boldsymbol{\eta}_{it|t-1} \right] \otimes x^*_{it} \quad (3.27)$$

$$\frac{\partial \ln L(\boldsymbol{\alpha})}{\partial \theta} = \sum_{i,t} \left[y_{it} - \boldsymbol{\eta}_{it|t-1} \right] \otimes y_{i,t-1} \quad (3.28)$$

Now, estimating functions in (3.27) and (3.28) together yield the estimating equation (3.23) as in the lemma. $\diamondsuit$

Lemma 3.4: For the MDFL model (3.2)-(3.5), we can write the Hessian matrix of the log-likelihood function (3.17) with respect to $\boldsymbol{\alpha} = \left(\boldsymbol{\beta}', \boldsymbol{\theta}'\right)'$ as

$$\begin{aligned} H[ln\, L(\boldsymbol{\alpha})] &= \left[\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \alpha\, \partial \alpha'} \right] \\ &= \sum_{i,t} \left[diag\left(\tilde{\eta}^{(1)}_{it|t-1}, \cdots, \tilde{\eta}^{(j)}_{it|t-1}, \cdots, \tilde{\eta}^{(K)}_{it|t-1} \right) \right. \\ &\qquad \left. - \boldsymbol{\eta}_{it|t-1}\, \boldsymbol{\eta}'_{it|t-1} \right] \otimes \begin{pmatrix} x^*_{it} \\ y_{i,t-1} \end{pmatrix} \begin{pmatrix} x^*_{it} \\ y_{i,t-1} \end{pmatrix}' \end{aligned} \quad (3.29)$$

Proof: Note that

$$\begin{aligned} \boldsymbol{\beta}_j^* &= (\beta_{j0}, \beta_{j1}, \cdots, \beta_{ju}, \cdots, \beta_{jp})'; && \text{by (3.20), and} \\ \boldsymbol{\theta}_j &= (\theta_{j1}, \cdots, \theta_{jl}, \cdots, \theta_{jK})'; && \text{by (3.22).} \end{aligned}$$

Now, taking the second derivatives of the log-likelihood function (3.17), that is, taking another derivative of the equations (3.25) and (3.26) with respect to β_j^* and θ_j, we obtain:

$$\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial V\, \partial W'} = \begin{cases} -\sum_{i,t} \lambda_{it}^{(j)}\, x_{it}^* x_{it}^{*\prime}; & \text{for } V = W = \beta_j^* \\ -\sum_{i,t} \lambda_{it}^{(j)}\, x_{it}^* y_{i,t-1}'; & \text{for } V = \beta_j^*,\, W = \theta_j \\ -\sum_{i,t} \lambda_{it}^{(j)}\, y_{i,t-1} y_{i,t-1}'; & \text{for } V = W = \theta_j \\ \sum_{i,t} \delta_{it}^{(jm)}\, x_{it}^* x_{it}^{*\prime}; & \text{for } V = \beta_j^*,\, W = \beta_m^* \\ \sum_{i,t} \delta_{it}^{(jm)}\, x_{it}^* y_{i,t-1}'; & \text{for } V = \beta_j^*,\, W = \theta_m \\ \sum_{i,t} \delta_{it}^{(jm)}\, y_{i,t-1} y_{i,t-1}'; & \text{for } V = \theta_j,\, W = \theta_m, \end{cases} \tag{3.30}$$

where

$$\lambda_{it}^{(j)} = \tilde{\eta}_{it|t-1}^{(j)} \left[1 - \tilde{\eta}_{it|t-1}^{(j)}\right],$$

and

$$\delta_{it}^{(jm)} = \tilde{\eta}_{it|t-1}^{(j)}\, \tilde{\eta}_{it|t-1}^{(m)}.$$

Hence we can write the second derivatives of the log-likelihood function (3.17) with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ as follows:

$$\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \beta\, \partial \beta'} = \sum_{i,t} \Big[diag\left(\tilde{\eta}^{(1)}_{it|t-1}, \cdots, \tilde{\eta}^{(j)}_{it|t-1}, \cdots, \tilde{\eta}^{(K)}_{it|t-1}\right) - \boldsymbol{\eta}_{it|t-1}\, \boldsymbol{\eta}'_{it|t-1}\Big] \otimes (x^*_{it}\, x^{*\prime}_{it}), \tag{3.31}$$

$$\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \beta\, \partial \theta'} = \sum_{i,t} \Big[diag\left(\tilde{\eta}^{(1)}_{it|t-1}, \cdots, \tilde{\eta}^{(j)}_{it|t-1}, \cdots, \tilde{\eta}^{(K)}_{it|t-1}\right) - \boldsymbol{\eta}_{it|t-1}\, \boldsymbol{\eta}'_{it|t-1}\Big] \otimes (x^*_{it}\, y'_{i,t-1}). \tag{3.32}$$

and

$$\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \theta\, \partial \theta'} = \sum_{i,t} \Big[diag\left(\tilde{\eta}^{(1)}_{it|t-1}, \cdots, \tilde{\eta}^{(j)}_{it|t-1}, \cdots, \tilde{\eta}^{(K)}_{it|t-1}\right) - \boldsymbol{\eta}_{it|t-1}\, \boldsymbol{\eta}'_{it|t-1}\Big] \otimes (y_{i,t-1}\, y'_{i,t-1}), \tag{3.33}$$

respectively.

Now by combining (3.31), (3.32), and (3.33), we obtain the Hessian matrix as in the lemma. ◇

Lemma 3.5: For the MDFL model (3.2)-(3.5), we can write the Fisher information matrix which is the expected value of the Hessian matrix of the log-likelihood function (3.17) as

$$\begin{aligned} I(\boldsymbol{\alpha}) &= E\left[H\left\{ln\,L(\boldsymbol{\alpha})\right\}\right] = \left[E\left\{\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \alpha\,\partial \alpha'}\right\}\right] \\ &= \begin{bmatrix} E\left\{\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \beta\,\partial \beta'}\right\} & E\left\{\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \beta\,\partial \theta'}\right\} \\ E\left\{\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \beta\,\partial \theta'}\right\} & E\left\{\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial \theta\,\partial \theta'}\right\} \end{bmatrix}, \end{aligned} \tag{3.34}$$

where

$$E\left[\frac{\partial^2 \ln L(\boldsymbol{\alpha})}{\partial V\,\partial W'}\right] = \begin{cases} -\sum\limits_{i,t}\left[D_{it} - W_{it}\right]; & \text{for } \; V = W = \boldsymbol{\beta} \\ -\sum\limits_{i,t}\left[D^*_{it} - W^*_{it}\right]; & \text{for } \; V = W = \boldsymbol{\theta} \\ -\sum\limits_{i,t}\left[\tilde{D}_{it} - \tilde{W}_{it}\right]; & \text{for } \; V = \boldsymbol{\beta},\, W = \boldsymbol{\theta} \end{cases} \tag{3.35}$$

with

$$D_{it} = \left[\bigoplus_{j=1}^{K} D_{itj}\right]_{Kp\times Kp} = \begin{bmatrix} D_{it1} & 0 & \cdots & 0 \\ 0 & D_{it2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{itK} \end{bmatrix}_{Kp\times Kp};$$

$$W_{it} = (V_{jj'})_{Kp\times Kp} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1K} \\ V_{21} & V_{22} & \cdots & V_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ V_{K1} & V_{K2} & \cdots & V_{KK} \end{bmatrix}_{Kp\times Kp},$$

$$D^*_{it} = \left[\bigoplus_{j=1}^{K} D^*_{itj}\right]_{K^2\times K^2}; \qquad W^*_{it} = \left(V^*_{jj'}\right)_{K^2\times K^2}, \text{ and}$$

$$\tilde{D}_{it} = \left[\bigoplus_{j=1}^{K} \tilde{D}_{itj}\right]_{Kp\times K^2}; \qquad \tilde{W}_{it} = \left(\tilde{V}_{jj'}\right)_{Kp\times K^2}$$

where $D_{itj} = \left[\tilde{\pi}^{(j)}_{it}(x^*_{it}x^{*\prime}_{it})\right]_{p\times p}$, $V_{jj'} = \left[\sum_{l=1}^{K+1}\tilde{\pi}^{(l)}_{i,t-1}\left\{\tilde{\eta}^{(j)}_{it|t-1}(l)\,\tilde{\eta}^{(j')}_{it|t-1}(l)\right\}(x^*_{it}x^{*\prime}_{it})\right]_{p\times p}$,

$$D^*_{itj} = diag\left[\tilde{\eta}^{(j)}_{it|t-1}(1)\tilde{\pi}^{(1)}_{i,t-1},\cdots,\tilde{\eta}^{(j)}_{it|t-1}(K)\tilde{\pi}^{(K)}_{i,t-1}\right],$$

$$V^*_{jj'} = diag\left[\tilde{\eta}^{(j)}_{it|t-1}(1)\tilde{\eta}^{(j')}_{it|t-1}(1)\tilde{\pi}^{(1)}_{i,t-1},\cdots,\tilde{\eta}^{(j)}_{it|t-1}(K)\tilde{\eta}^{(j')}_{it|t-1}(K)\tilde{\pi}^{(K)}_{i,t-1}\right],$$

$$\tilde{D}_{itj} = x_{it}\left[\boldsymbol{\eta}^{(j)}_{it|t-1}() \oplus \tilde{\Pi}_{i,t-1}\right]'_{p\times K}, \; \tilde{V}_{jj'} = x_{it}\left[\boldsymbol{\eta}^{(j)}_{it|t-1}() \oplus \boldsymbol{\eta}^{(j')}_{it|t-1}() \oplus \tilde{\Pi}_{i,t-1}\right]'_{p\times K},$$

with $\boldsymbol{\eta}^{(j')}_{it|t-1}() = \left(\tilde{\eta}^{(j)}_{it|t-1}(1),\cdots,\tilde{\eta}^{(j)}_{it|t-1}(K)\right)'$, and '$\oplus$' denoting the Hadamard, i.e., elements wise product.

Proof: The proof of the lemma requires the expectations of various elements involved in the second derivatives. For convenience, the derivation for these expectations for some special cases are shown in the Appendix A (page 148). ◇

3.1.3 A Simulation Study for the History Based Data

Asymptotic properties of the likelihood estimators are well-known. In this section, we rather examine the finite sample properties of such estimators through a simulation study. Recall that the parameters involved in the MDFL model (3.2)-(3.5) are $\boldsymbol{\beta} \equiv (\beta_{j0}, \beta_{ju})$ and $\boldsymbol{\theta} \equiv (\theta_{jl})$ for $j = 1, \cdots, K$, $u = 1, \cdots, p$, and $l = 1, \cdots, K$. For the simulation purpose, by using a true set of values for the components of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, we generate the initial multinomial response $\boldsymbol{y}_{i1} = (y_{i11}, \cdots, y_{i1K})'$ $(i = 1, \cdots, I)$ following (3.2), and $y_{it} = (y_{it1}, \cdots, y_{itK})'$ for $t = 2, \cdots, T$, following (3.3)-(3.5). The true parameter values $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ will then be estimated by solving the likelihood estimating equation (3.23) as in Lemma 3.3. This data generation and estimation process will be repeated for 500 times. Finally, these 500 likelihood estimates will be summarized to examine the performance of the MDFL model in estimating the true parameter values.

3.1.3.1 Simulation Design

We consider $I = 100$ independent individuals with trichotomous responses over $T = 4$ time points. As far as the time dependent covariates for these individuals are concerned, we select two covariates as follows:

$$x_{it(1)} = \begin{cases} 1 & \text{for } t = 4; & i = 1, \ldots, 25 \\ 0 & \text{for } t = 1,\ 2,\ 3; & i = 1, \ldots, 25 \\ 1 & \text{for } t = 3,\ 4; & i = 26, \ldots, 50 \\ 0 & \text{for } t = 1,\ 2; & i = 26, \ldots, 50 \\ 1 & \text{for } t = 2,\ 3,\ 4; & i = 51, \ldots, 75 \\ 0 & \text{for } t = 1; & i = 51, \ldots, 75 \\ 1 & \text{for } t = 1,\ 2,\ 3,\ 4; & i = 76, \ldots, 100 \end{cases}$$

$$x_{it(2)} \sim bin(p = 0.6); \quad \text{for } t = 1,\ 2,\ 3,\ 4; \quad i = 1, \ldots, 100$$

The above time dependent covariates values are chosen hypothetically, where, we have, however, followed four different patterns for four groups of individuals in selection of the first covariate. The second covariate, in contrast to the first one, has been chosen a random covariate allowing random differences among the individuals.

For the selection of the elements of the regression parameter vector $\boldsymbol{\beta}$, i.e., the category and covariate effects, we consider the following two sets of parameters;

$$\beta' = (\beta_{10}, \beta_{11}, \beta_{12}, \beta_{20}, \beta_{21}, \beta_{22}) = \begin{cases} (0.1, 0.0, 0.0, 0.2, \quad 0.0, 0.0) \\ (0.1, 0.3, 0.1, 0.2, -0.2, 0.0) \end{cases},$$

and for the selection of the elements of the dynamic dependence parameter vector $\boldsymbol{\theta}$, i.e., the transition effects, we consider following four sets of parameters;

$$\theta' = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22}) = \begin{cases} (0.2, \quad 0.0, \quad 0.0, \ 0.1) \\ (0.8, \quad 0.3, \quad 0.3, \ 0.8) \\ (0.3, \quad 0.7, \quad 0.8, \ 0.5) \\ (0.8, -0.5, -0.5, \ 0.8) \end{cases}$$

Note that the values of the elements of $\boldsymbol{\theta}$ were chosen to reflect large and small correlations both for an individual remaining in the same category and transiting to the different categories.

3.1.3.2 Simulated Estimates

Note that we have selected eight parameter combinations and for each of these combinations we compute the likelihood estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ for each of the 500 simulations. The simulated mean (SM) and the simulated standard errors (SSE) calculated from the 500 simulation results are reported in Table 3.1.

Table 3.1: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* θ *and the True Value of* $\boldsymbol{\beta}' = (\beta_{10} = 0.1, \beta_{11} = 0.0, \beta_{12} = 0.0, \beta_{20} = 0.2, \beta_{21} = 0.0, \beta_{22} = 0.0)$, *Under the MDFL Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2$, $\theta_{12} = 0.0$; $\theta_{21} = 0.0$, $\theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11552	-0.00581	0.00490	0.20714	-0.01081	0.00088
SSE	0.26896	0.27167	0.26180	0.26389	0.27296	0.25171
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.19725	-0.00858		0.01456	0.089217
SSE		0.30111	0.30654		0.29626	0.28396
True	$\theta_{11} = 0.8$, $\theta_{12} = 0.3$; $\theta_{21} = 0.3$, $\theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.14938	0.02928	0.01089	0.24043	0.02373	0.01021
SSE	0.28772	0.29446	0.27812	0.28791	0.29591	0.28558
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.73028	0.21649		0.25696	0.72751
SSE		0.30808	0.31691		0.34542	0.30724
True	$\theta_{11} = 0.3$, $\theta_{12} = 0.7$; $\theta_{21} = 0.8$, $\theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12221	0.01980	0.03138	0.23099	0.02147	0.02714
SSE	0.28124	0.30340	0.27766	0.26336	0.28877	0.25840
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.23891	0.66745		0.76224	0.44223
SSE		0.30000	0.31130		0.32821	0.30810
True	$\theta_{11} = 0.8$, $\theta_{12} = -0.5$; $\theta_{21} = -0.5$, $\theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11550	0.02822	0.01142	0.20945	0.02680	0.01295
SSE	0.27413	0.28632	0.27192	0.27746	0.27686	0.26380
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.75467	-0.51821		-0.54053	0.77984
SSE		0.32029	0.38801		0.38426	0.29824

Table 3.2: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* θ *and the True Value of* $\boldsymbol{\beta}' = (\beta_{10} = 0.1, \beta_{11} = 0.3, \beta_{12} = 0.1, \beta_{20} = 0.2, \beta_{21} = -2.0, \beta_{22} = 0.0)$*, Under the MDFL Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2$, $\theta_{12} = 0.0$; $\theta_{21} = 0.0$, $\theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12068	0.32086	0.09633	0.22241	-0.20433	-0.01038
SSE	0.26271	0.25500	0.25078	0.27245	0.26928	0.25383
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.17008	-0.01423		-0.02041	0.07084
SSE		0.27584	0.30368		0.31529	0.30380
True	$\theta_{11} = 0.8$, $\theta_{12} = 0.3$; $\theta_{21} = 0.3$, $\theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.13009	0.35445	0.09255	0.23090	-0.15832	-0.01879
SSE	0.26569	0.27808	0.26330	0.28284	0.29780	0.26793
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.72712	0.25674		0.23634	0.75366
SSE		0.28862	0.34284		0.32990	0.35244
True	$\theta_{11} = 0.3$, $\theta_{12} = 0.7$; $\theta_{21} = 0.8$, $\theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12427	0.32509	0.10172	0.22767	-0.15598	0.00768
SSE	0.28296	0.28629	0.27321	0.28069	0.28881	0.27561
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.26451	0.65446		0.75837	0.40014
SSE		0.30139	0.32535		0.32132	0.36367
True	$\theta_{11} = 0.8$, $\theta_{12} = -0.5$; $\theta_{21} = -0.5$, $\theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.13230	0.32497	0.09711	0.23023	-0.17362	-0.01400
SSE	0.28145	0.27740	0.26058	0.28077	0.29269	0.29067
		$\hat{\theta}_{11}$	$\hat{\theta}_{12}$		$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM		0.75217	-0.54321		-0.53428	0.76220
SSE		0.28441	0.35641		0.37447	0.32001

It is clear that the likelihood estimation works very well. For example, when $\boldsymbol{\beta}' = (0.1, 0.0, 0.0, 0.2, 0.0, 0.0)$ and $\boldsymbol{\theta}' = (0.2, 0.0, 0.0, 0.1)$ the results in Table 3.1 provides the estimates of $\boldsymbol{\beta}$ as $\hat{\boldsymbol{\beta}}' = (0.11552,\ -0.00581,\ 0.00490,\ 0.20714,\ -0.01081,\ 0.00088)$ and of $\boldsymbol{\theta}$ as $\hat{\boldsymbol{\theta}}' = (0.19725,\ -0.00858,\ 0.01456,\ 0.089217)$, with respective standard errors $SE(\hat{\boldsymbol{\beta}}') \equiv (0.26896,\ 0.27167,\ 0.26180,\ 0.26389,\ 0.27296,\ 0.25171)$ and $SE(\hat{\boldsymbol{\theta}}') \equiv (0.30111,\ 0.30654,\ 0.29626,\ 0.28396)$. Because the biases are quite small along with not too big standard errors, the estimates are consistent. However, as expected, when the element of $\boldsymbol{\theta}$ are relatively large indicating large dynamic dependence, the estimates of the regression parameters become slightly biased in some cases.

3.1.4 An Illustration for the History Based Data

In this section, we provide a numerical illustration for proposed MDFL model (3.2)-(3.5) by re-analyzing the Three Mile Island Stress-Level data (Fienberg et al., 1985), collected from a psychological study of the mental health effects of the accident at the Three Mile Island nuclear power plant in central Pennsylvania began on March 28, 1979. This data set was analyzed by Fienberg et al. (1985). However, these authors have used a dichotomized stress responses instead of trichotomous responses. This makes the use of binary dynamic model only [see Fienberg et al. (1985, Eqn. (8))] which is the same as the binary dynamic model considered by Sutradhar and Farrell (2007).

Note that later on, Conaway (1989) has re-analyzed the same data using these three stress levels. However, Conaway used time/wave as a deterministic factor and found the time effects regressing time on the responses. This is quite different from our MDFL model, where we consider time as a stochastic factor and develop correlation model among multinomial responses over time. Also note that Conaway (1989) analyzed the data for a given level of the covariate, whereas we consider this distance covariate (greater or less than 5 miles from the plant) as a dichotomous variable. Thus, with regard to the covariates, Conaway has done marginal analysis, whereas our joint analysis is more appropriate for understanding the effects of the covariate due to its levels.

For the purpose of the application of our methodology we present the same set of data in the Table 3.3. The study focuses on the changes in the stress level of mothers of young children living within 10 miles of the nuclear plant. The accident was followed by four interviews; winter 1979 (wave 1), spring 1980 (wave 2), fall 1981 (wave 3), and fall 1982 (wave 4). In this study, the subject were classified into one of the three response categories namely, low, medium and high stress level, based on a composite score from a 90-items checklist. There were 267 subjects who completed all four interviews. Respondents were stratified into two groups, those living within 5 miles of the plant (LT5) and those lives within 5 to 10 miles from the plant (GT5).

Table 3.3: *Three Mile Island Stress-Level Data.*

For less than 5 miles distance (LT5)

				Wave 4		
Distance	Wave 1	Wave 2	Wave 3	Low4	Med4	High4
LT5	Low1	Low2	Low3	2	0	0
			Med3	2	3	0
			High3	0	0	0
		Med2	Low3	0	1	0
			Med3	2	4	0
			High3	0	0	0
		High2	Low3	0	0	0
			Med3	0	0	0
			High3	0	0	0
	Med1	Low2	Low3	5	1	0
			Med3	1	4	0
			High3	0	0	0
		Med2	Low3	3	2	0
			Med3	2	38	4
			High3	0	2	3
		High2	Low3	0	0	0
			Med3	0	2	0
			High3	0	1	1
	High1	Low2	Low3	0	0	0
			Med3	0	0	0
			High3	0	0	0
		Med2	Low3	0	0	0
			Med3	0	4	3
			High3	0	1	4
		High2	Low3	0	0	0
			Med3	1	2	0
			High3	0	5	12

Table 3.3 (Continued): *Three Mile IslandStress-LevelData.*

For greater than 5 miles distance (GT5)

Distance	Wave 1	Wave 2	Wave 3	Wave 4 Low4	Med4	High4
GT5	Low1	Low2	Low3	1	2	0
			Med3	2	0	0
			High3	0	0	0
		Med2	Low3	1	0	0
			Med3	0	3	0
			High3	0	0	0
		High2	Low3	0	0	0
			Med3	0	0	0
			High3	0	0	0
	Med1	Low2	Low3	4	4	0
			Med3	5	15	1
			High3	0	0	0
		Med2	Low3	2	2	0
			Med3	6	53	6
			High3	0	5	1
		High2	Low3	0	0	0
			Med3	0	1	1
			High3	0	3	1
	High1	Low2	Low3	0	0	1
			Med3	0	0	0
			High3	0	0	0
		Med2	Low3	0	0	0
			Med3	1	13	0
			High3	0	0	0
		High2	Low3	0	0	0
			Med3	0	7	2
			High3	0	2	7

Note that our methodology is based on lag 1 time dependence. For this reason, for preliminary understanding of the data set, we also provide all possible lag 1 transition counts over time in Table 3.4.

Table 3.4: *All possible Lag 1 Transition Counts Over Time for Different Covariate Levels from the Three Mile Island Stress-Level Data Set.*

Time $t = 2$

		LT5				GT5			
	Category	Low	Med	High	Total	Low	Med	High	Total
Time $t = 1$	Low	7	7	0	14	5	4	0	9
	Med	11	54	4	69	29	75	6	110
	High	0	12	20	32	1	14	18	33
	Total	18	73	24	115	35	93	24	152

Time $t = 3$

		LT5				GT5			
	Category	Low	Med	High	Total	Low	Med	High	Total
Time $t = 2$	Low	8	10	0	18	12	23	0	35
	Med	6	57	10	73	5	82	6	93
	High	0	5	19	24	0	11	13	24
	Total	14	72	29	115	35	116	19	152

Time $t = 4$

		LT5				GT5			
	Category	Low	Med	High	Total	Low	Med	High	Total
Time $t = 3$	Low	10	4	0	14	8	8	1	17
	Med	8	57	7	72	14	92	10	116
	High	0	9	20	29	0	10	9	19
	Total	18	70	27	115	22	110	20	152

Next, we also provide a summary statistics in Table 3.5 for the distribution of individuals under three stress-level categories verses the covariate levels for all time points $t = 1, 2, 3, 4$.

Table 3.5: *Distribution of Individuals Under Three Stress-Level Versus the Covariate Levels from the Three Mile Island Stress-Level Data Set.*

Time $t = 1$

Category	Low	Med	High	Total
GT5	9	110	33	152
LT5	14	69	32	115
Total	23	179	65	267

Time $t = 2$

Category	Low	Med	High	Total
GT5	35	93	24	152
LT5	18	73	24	115
Total	53	166	48	267

Time $t = 3$

Category	Low	Med	High	Total
GT5	17	116	19	152
LT5	14	72	29	115
Total	31	188	48	267

Time $t = 4$

Category	Low	Med	High	Total
GT5	22	110	20	152
LT5	18	70	27	115
Total	40	180	47	267

The exploratory data in Table 3.4 indicate that irrespective of time, the transition from low to high or high to low is a rare event. It happens only once in each of the cases. But the transition from low to medium is more common and almost half of the times responses belong to the low level transit to the medium level irrespective of the time.

In the whole data set, 62 (166 out of 267) to 70 (188 out of 267) percent of the individuals always belong to the medium stress level and 72 to 84 percent [see Table 3.5] of the individuals under this level remain in same level. Individuals belong to medium level, approximately 7 to 22 percent transit to low level and 5 to 10 percent [see Table 3.5] transit to high level. Approximately 35 to 40 percent [see Table 3.5] of the individuals transit from high to medium level irrespective of time.

We now apply the proposed MDFL model (3.2)-(3.5) to the Three Miles Island Stress-Level data and provide the MLE for $\boldsymbol{\beta} = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21})'$ and $\boldsymbol{\theta} = (\theta_{11}, \theta_{12}, \theta_{21}, \theta_{22})'$ as in Table 3.6.

Table 3.6: *The Likelihood Estimates for Regression and Dynamic Dependence Parameters for the Three Mile Island Stress-Level Data.*

	Category 1 vs 3		Category 2 vs 3	
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$
SM	-2.0373	-1.6618	0.1035	0.6233
SSE	0.1748	0.1636	0.1210	0.1201
	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	5.7594	2.3666	3.6497	1.8824
SSE	1.0101	0.1924	1.0092	0.1606

The estimates of β_{j0} $(j = 1, 2)$ from Table 3.6 indicate that when other variables are fixed, an individual has higher probability (with $\hat{\beta}_{20} = 0.1035$) to have medium stress-level and smaller probability to have low stress-level (with $\hat{\beta}_{10} = -2.03732$) as compared to the high stress-level.

The value of $\hat{\beta}_{11} = -1.6618$ indicates that an individual belong to GT5 has smaller probability as compared to LT5 group to be in low stress level. This result appears to fit the raw data evident by summary statistics shown in Table 3.5. For example, irrespective of time, say for $t = 3$, the individuals has observed probability $\frac{17}{152} =$

0.112 under the GT5 group to be in the low stress-level (with high stress-level as reference), which is smaller as compared to the observed probability $\frac{14}{115} = 0.122$ under the LT5 group.

When covariate effect is examined in the medium stress-level (high stress-level as the reference), $\hat{\beta}_{21} = 0.6233$ indicates that the individual in GT5 group has higher probability to experience medium stress level as compared to an individual belongs to the LT5 group. This result (similar to that for $\hat{\beta}_{11}$) is also supported by the distribution of individuals shown in Table 3.5. To be specific, for example, say for $t = 4$, an individual has observed probability $\frac{110}{152} = 0.724$ under the GT5 group to be in the medium stress-level (with high stress-level as reference), which is higher as compared to the observed probability $\frac{70}{115} = 0.609$ under the LT5 group.

As far as the dynamic dependence is concerned, $\hat{\theta}_{11} = 5.7594$ and $\hat{\theta}_{12} = 2.3666$ indicate that an individual has higher probability for remaining in the low stress-level as compared to transiting from medium to low stress-level (with transiting from high stress-level as reference). This appears to explain the observed counts well as shown in Table 3.4. This is because, transiting from time 2 to 3, an individual in low stress-level at time 2 has probability $\frac{20}{53} = 0.377$ to remain in the same level at time 3 as opposed to an individual transiting from medium to low stress-level with observed probability $\frac{11}{166} = 0.066$

One may similarly interpret the estimates $\hat{\theta}_{21} = 3.6497$ and $\hat{\theta}_{22} = 1.8824$.

3.2 Ordinal Multinomial Dynamic Fixed Logit Model

Note that in Section 3.1, we have considered nominal categories for the response variable in a longitudinal set up. Also, the Three Mile Island Stress-Level data set was reanalyzed in Section 3.1.4 under the assumption that stress levels are nominal. However, in some situations the categorical variables such as in the Three Mile Island Stress-level example, can be considered as ordinal. For this type of ordinal multinomial analysis the likelihood methodology remains almost the same except that this ordinal nature should be taken into account which would reduce the number of parameters involved in the model.

To accommodate the ordinal nature of the responses in the present set up, it is appropriate to change its past response effects θ_j to $\theta(v_j - \bar{v})$ and effect of covariates x_{it}, β_j to $\beta(w_j - \bar{w})$ in the nominal response based model (3.2)-(3.5), where v_j and w_j are suitable ordinal scorer for the jth ordinal category. For example, $v_j = j$ and $w_j = j$ indicate the standard ordinal score. Thus, for the ordinal responses, the probability model (3.3) reduces to

$$\begin{aligned} \tilde{\eta}^{(j)}_{it|t-1}(l) &= P(Y_{it} = y^{(j)}_{it} \,|\, Y_{i,t-1} = y^{(l)}_{i,t-1}) \\ &= \frac{\exp\left\{(w_j - \bar{w})\, x'_{it}\beta \,+\, (v_j - \bar{v})\,\theta' y^{(l)}_{i,t-1}\right\}}{1 + \sum_{k=1}^{K} \exp\left\{(w_k - \bar{w})x'_{it}\beta + (v_k - \bar{v})\theta' y^{(l)}_{i,t-1}\right\}}; \qquad (3.36) \\ & \qquad\qquad j = 1,\cdots,K; \;\; l = 1,\cdots,K+1, \end{aligned}$$

with,

$$\tilde{\eta}^{(K+1)}_{it|t-1}(l) = 1 - \sum_{j=1}^{K} \tilde{\eta}^{(j)}_{it|t-1}(l) = \frac{1}{1 + \sum_{k=1}^{K} \exp\left\{(w_k - \bar{w})x'_{it}\beta + (v_k - \bar{v})\theta' y^{(l)}_{i,t-1}\right\}}$$

If the response is at $(K+1)th$ category at $(t-1)$ time point then we can write

$$\begin{aligned} \tilde{\eta}^{(j)}_{it|t-1}(K+1) &= P(Y_{it} = y^{(j)}_{it} \,|\, Y_{i,t-1} = y^{(K+1)}_{i,t-1}) \\ &= \frac{\exp\{(w_j - \bar{w})x'_{it}\beta\}}{1 + \sum_{k=1}^{K} \exp\{(w_k - \bar{w})x'_{it}\beta\}}; \; j = 1, \cdots, K. \end{aligned} \quad (3.37)$$

3.2.1 Likelihood Estimation of β and θ

The likelihood function is

$$L(\beta, \theta) = \prod_{i=1}^{I} \prod_{t=1}^{T} \frac{1}{y_{it1}! ... y_{itk}! \, (1 - \sum_{k=1}^{K} y_{itk})!} \prod_{k=1}^{K+1} \left\{\tilde{\eta}^{(k)}_{it|t-1}\right\}^{y_{itk}} \quad (3.38)$$

where $y_{i0} = 0$, $y_{it,K+1} = (1 - \sum_{u=1}^{K} y_{itu})$, and $\tilde{\eta}^{(k)}_{it|t-1}$ is in (3.36)-(3.37).

We now have $\beta : p \times 1$ and $\theta : K \times 1$ parameters to estimate, whereas in the nominal model (3.2)-(3.5) we had Kp regression parameters and K^2 dynamic dependence parameters. Note that these new parameters $\beta : p \times 1$ and $\theta : K \times 1$ can easily be estimated by applying the likelihood method discussed in Section 3.1.2. Thus, we do not provide any further details for their estimation.

3.3 MDFL Model for the Contingency Table Based Data

Note that in Section 3.1, for a given individual we have considered non-stationary multinomial models with time dependent covariates. There may be some situations in practice where these covariates are time independent and as a result for a given combination level of multiple covariates, the same stationary multinomial response probability will be obtained for a group of individuals. Similar to Chapter 2, here we again consider only one covariate with $p+1$ levels, and use $x_{i(1)}, \cdots, x_{i(p)}$ as in (1.11) to represent the $p+1$ levels of the covariate for the ith individual. Suppose that under the uth level of the covariate there are $I_{(u,1)}^{(j)}$ individuals within the jth category at time $t=1$ with the same probability $\tilde{\pi}_{(u)}^{(j)}$ for each individual. Also suppose that under the uth level of the covariate for $t=2,3,4$, there are $I_{lj(u,t|t-1)}$ individuals each with the same conditional probability $\tilde{\eta}_{(u,t|t-1)}^{(j)}(l)$ for transiting to the jth category at time t from the lth category at time t-1. These common marginal and conditional probabilities under the present non-linear dynamic model have the formulas given by

$$\tilde{\pi}_{(u)}^{(j)} = P(\boldsymbol{Y}_{i1} = \boldsymbol{y}_{i1}^{(j)}) = \frac{\exp\left(\beta_{j0} + \beta_{ju}\right)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + \beta_{ku}\right)}, \tag{3.39}$$

and

$$\begin{aligned}\tilde{\eta}^{(j)}_{(u,it|t-1)}(l) &= P\Big(\boldsymbol{Y}_{it} = \boldsymbol{y}^{(j)}_{it} \,|\, \boldsymbol{Y}_{i,t-1} = \boldsymbol{y}^{(l)}_{i,t-1}\Big) \\ &= \frac{\exp\Big(\beta_{j0} + \beta_{ju} + \boldsymbol{\theta}'_j\, y^{(l)}_{i,t-1}\Big)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + \beta_{ku} + \boldsymbol{\theta}'_k y^{(l)}_{i,t-1}\right)};\ \text{for } j, l = 1, \cdots, K, \qquad (3.40)\end{aligned}$$

respectively. For the remaining cases, i.e., when $l = K + 1$, the above conditional probability reduces to

$$\begin{aligned}\tilde{\eta}^{(j)}_{(u,it|t-1)}(K+1) &= P\Big(\boldsymbol{Y}_{it} = \boldsymbol{y}^{(j)}_{it} \,|\, \boldsymbol{Y}_{i,t-1} = \boldsymbol{y}^{(K+1)}_{i,t-1}\Big) \\ &= \frac{\exp\Big(\beta_{j0} + \beta_{ju}\Big)}{1 + \sum_{k=1}^{K} \exp\Big(\beta_{k0} + \beta_{ku}\Big)};\ \text{for } j = 1, \cdots, K, \qquad (3.41)\end{aligned}$$

and for any $l = 1, \cdots, K$, the probability for the response to be in the last category $K + 1$ (i.e., for $j = K + 1$) at time t is given by

$$\tilde{\eta}^{(K+1)}_{(u,it|t-1)}(l) = 1 - \sum_{j=1}^{K} \tilde{\eta}^{(j)}_{(u,it|t-1)}(l) = \frac{1}{1 + \sum_{k=1}^{K} \exp\Big(\beta_{k0} + \beta_{ku} + \boldsymbol{\theta}'_k y^{(l)}_{i,t-1}\Big)}. \qquad (3.42)$$

Note that the marginal and conditional counts $I^{(j)}_{(u,1)}$ and $I_{lj(u,t|t-1)}$, respectively, may be summarized in tabular form as in Tables 2.2(a)-(b). Similarly, one may form tables for the marginal and conditional probabilities as in Table 2.1(a)-(b), with a difference that in Table 2.1, the marginal and conditional probabilities are given by (2.55) and (2.56), respectively, whereas under the non-linear dynamic model these formulas are now given by (3.39) and (3.40), respectively.

3.3.1 Product Multinomial Likelihood Inferences

For the estimation of the parameters involved in the present non-linear dynamic model (3.39)-(3.40), one may write the likelihood function as

$$\tilde{L}(\boldsymbol{\beta}, \boldsymbol{\theta}) = \prod_{u=1}^{p+1} \tilde{L}_{(u)}, \tag{3.43}$$

where

$$\tilde{L}_{(u)} = \tilde{f}_{(u,1)} \prod_{t=2}^{T} \prod_{l=1}^{K+1} \left[\tilde{f}_{(u,t|t-1)}(l) \right], \tag{3.44}$$

with $\tilde{f}_{(u,1)}$ as the marginal multinomial probability at time $t = 1$ given by

$$\tilde{f}_{(u,1)} = \prod_{j=1}^{K+1} \frac{I_{(u)}!}{I_{(u,1)}^{(j)}!} \left\{ \tilde{\pi}_{(u)}^{(j)} \right\}^{I_{(u,1)}^{(j)}}, \tag{3.45}$$

and $\tilde{f}_{(u,t|t-1)}(l)$ as the conditional multinomial probability at time t given that the response was in the lth category at time $t-1$. This conditional distribution has the formula

$$\tilde{f}_{(u,t|t-1)}(l) = \prod_{j=1}^{K+1} \frac{I_{(u,t-1)}^{(l)}!}{I_{lj(u,t|t-1)}!} \left\{ \tilde{\eta}_{(u,t|t-1)}^{(j)}(l) \right\}^{I_{lj(u,t|t-1)}}. \tag{3.46}$$

Next by using the formulas for $\tilde{\pi}_{(u)}^{(j)}$ from (3.39) and for $\tilde{\eta}_{(u,t|t-1)}^{(j)}(l)$ from (3.40), after some algebra we write the log likelihood function as

$$\ln L(\boldsymbol{\beta}, \boldsymbol{\theta}) = C + \sum_{u=1}^{p+1} \left[\tilde{\tilde{g}}_u(\boldsymbol{\beta}) + \tilde{\tilde{h}}_u(\boldsymbol{\beta}, \boldsymbol{\theta}) \right] \tag{3.47}$$

where

$$\tilde{\tilde{g}}_u(\boldsymbol{\beta}) = \sum_{j=1}^{K} I_{(u,1)}^{(j)} (\beta_{j0} + \beta_{ju}) - I_{(u)} \ln \left\{ 1 + \sum_{k=1}^{K} \exp(\beta_{k0} + \beta_{ku}) \right\} \tag{3.48}$$

and

$$\begin{aligned} \tilde{\tilde{h}}_u(\boldsymbol{\beta}, \boldsymbol{\theta}) = & \sum_{t=2}^{T} \sum_{l=1}^{K+1} \left[\sum_{j=1}^{K+1} I_{lj(u,t|t-1)} (\beta_{j0} + \beta_{ju} + \theta_{jl}) \right. \\ & \left. - I_{(u,t-1)}^{(l)} \ln \left\{ 1 + \sum_{k=1}^{K} \exp(\beta_{k0} + \beta_{ku} + \theta_{kl}) \right\} \right] \end{aligned} \tag{3.49}$$

with C is a normalizing constant.

Note that the log likelihood function given by (3.47) can be maximized with respect to β_{j0}, β_{ju} and θ_{jl} ($j, l = 1, \cdots, K$ and $u = 1, \cdots, p$) in the manner similar to that of Section 2.2.1. The first and second order derivatives can easily be calculated following Lemma 3.3 and Lemma 3.4 and hence details are not shown.

Chapter 4

Multinomial Dynamic Mixed Logit Models

In Chapter 3, we have generalized the non-linear binary longitudinal model discussed in Chapter 1 (Section 1.1.2.3) to the multinomial case. As discussed, this multinomial fixed logit model introduced in Chapter 3, allows pairwise lag correlations to be in the range from -1 to 1. Also, the dynamic dependence parameter in this multinomial dynamic model was estimated by likelihood approach in the same way the regression parameters were estimated. Thus, there was no necessity of using any extra-equations for the estimation of the correlations of the data. Note however that, there may be situations in practice where the mean, variance and correlations of the data may not be fully explained through regression and dynamic dependence parameters. This may happen mainly due to certain latent factors those are unobserved but perhaps influential. This situation is usually accommodated by using suitable random effects for the individuals so that the variation of the random effects may provide additional information to understand the perfect nature of the observed multinomial data. For the purpose, in this chapter, we generalize the multinomial dynamic fixed (MDFL) logit models of Chapter 3 to the multinomial dynamic mixed logit (MDML) models. This generalization is provided in Section 4.1 below.

Note that the non-linear binary longitudinal fixed models in Chapter 1 (Section 1.1.2.3) has also been generalized to the mixed models case. For example, we refer to Sutradhar et al. (2008). These author have used generalized quasi likelihood (GQL) inference for the estimation of the parameters of such binary dynamic mixed models. Also, this was applied to analyze a well-known SLID (Survey of Labor and Income Dynamics) data set from Statistics Canada. However, it was limited to the binary cases as opposed to the multinomial case. The proposed model in Section 4.1 may therefore be treated as the generalization of such binary longitudinal mixed model.

Following the model given in Section 4.1, in Section 4.2, we provide the likelihood estimation for the parameters of the model including the random effects variances under all possible categories. Because the exact likelihood computation under the proposed mixed model is difficult, we use a simulation based approximation for such likelihood computation. In Section 4.3, we conduct an extensive simulation study to examine the finite sample performance of the simulated likelihood approach.

Note that the random effects variances in multinomial dynamic models cause overdispersion which also affects the correlations of the repeated responses. Some authors, such as Wilson and Koehler (1991) discussed overdispersed model for multinomial data at cross-sectional level. Neerchal and Morel (2005) also dealt with overdispersed multinomial data in multivariate setup at cross-sectional level. Thus they model the multivariate multinomial correlations as opposed to longitudinal correlations. Mixture multinomial models, similar to Morel and Nagaraj (1993) also has been used by Cruz-Medina et al. (2004) for repeated data. Multivariate multinomial analysis at cross-sectional level has also been done by Chen and Kuo (2001) through random effects, whereas we deal with univariate multinomial data but in longitudinal setup, also affected by random effects.

4.1 MDML Model For the History Based Data

Recall from Section 3.1 of Chapter 3 that the marginal multinomial probability $\tilde{\pi}_{i1}^{(j)}$ for $t = 1$ and $i = 1, \cdots, I$, under the MDFL model, has the formula

$$\tilde{\pi}_{i1}^{(j)} = P\Big(\mathbf{Y}_{i1} = \mathbf{y}_{i1}^{(j)}\Big) = \frac{\exp\left(\beta_{j0} + \mathbf{x}_{i1}'\boldsymbol{\beta}_j\right)}{1 + \sum_{k=1}^{K} \exp\left(\beta_{k0} + \mathbf{x}_{i1}'\boldsymbol{\beta}_k\right)} \tag{4.1}$$

[see also (3.2)], whereas the conditional probability for $\mathbf{y}_{it} = (y_{it1}, \cdots, y_{itj}, \cdots, y_{itK})'$

given $\mathbf{y}_{i,t-1}$ [using $\mathbf{y}_{i,t-1}$ for $\mathbf{y}_{i,t-1}^{(l)}$] was modeled by the non-linear dynamic relationship (3.3) denoted by $\tilde{\eta}_{it|t-1}^{(j)}(l)$, where $\boldsymbol{\theta}_j = (\theta_{j1}, \cdots, \theta_{jl}, \cdots, \theta_{jK})'$ involved in $\tilde{\eta}_{it|t-1}^{(j)}(l)$ is a vector of dynamic dependence parameters.

As opposed to the linear predictor $\mathbf{x}'_{it}\boldsymbol{\beta}_j$ $(t = 1, \cdots, T)$ defined corresponding to the jth category under the MDFL model, in the MDML model a random effect γ_{ij} is added to $\mathbf{x}'_{it}\boldsymbol{\beta}_j$ in order to reflect the ith individual latent effect that may influence the probability under the jth category. We assume that $\gamma_{ij}^* \overset{iid}{\sim} N(0, \sigma_j^2)$ for $j = 1, \cdots, K$. For $\gamma_{ij} = \gamma_{ij}^*/\sigma_j \overset{iid}{\sim} N(0, 1)$ we may then write the marginal probability at $t = 1$ under the MDML model as

$$\tilde{\pi}_{i1}^{*(j)} = P\Big(\mathbf{Y}_{i1} = \mathbf{y}_{i1}^{(j)}\Big) = \frac{\exp\Big(\beta_{j0} + \mathbf{x}'_{it}\boldsymbol{\beta}_j + \sigma_j\gamma_{ij}\Big)}{1 + \sum_{k=1}^{K}\exp\Big(\beta_{k0} + \mathbf{x}'_{it}\boldsymbol{\beta}_k + \sigma_k\gamma_{ik}\Big)}. \qquad (4.2)$$

Similarly, by adding the random effects to the linear predictor in the non-linear conditional model (3.3) we now write the non-linear conditional probability under the MDML model as

$$\begin{aligned} P\Big(\mathbf{Y}_{it} = \mathbf{y}_{it}^{(j)} \,\big|\, \mathbf{Y}_{i,t-1} = \mathbf{y}_{i,t-1}^{(l)}\Big) &= \frac{\exp\left(\beta_{j0} + \sum_{u=1}^{p}\beta_{ju}x_{it(u)} + \sigma_j\gamma_{ij} + \sum_{c=1}^{K}\theta_{jc}\,y_{i,t-1,c}^{(l)}\right)}{\sum_{k=1}^{K+1}\exp\left(\beta_{k0} + \sum_{u=1}^{p}\beta_{ku}x_{it(u)} + \sigma_k\gamma_{ik} + \sum_{c=1}^{K}\theta_{kc}\,y_{i,t-1,c}^{(l)}\right)} \\ &= \frac{\exp\Big(\beta_{j0} + \mathbf{x}'_{it}\boldsymbol{\beta}_j + \sigma_j\gamma_{ij} + \boldsymbol{\theta}'_j\,\mathbf{y}_{i,t-1}^{(l)}\Big)}{1 + \sum_{k=1}^{K}\exp\Big(\beta_{k0} + \mathbf{x}'_{it}\boldsymbol{\beta}_k + \sigma_k\gamma_{ik} + \boldsymbol{\theta}'_k\mathbf{y}_{i,t-1}^{(l)}\Big)} \\ &= \tilde{\eta}_{it|t-1}^{*(j)}(l), \text{ say}, \quad \text{for } j, l = 1, \cdots, K. \end{aligned} \qquad (4.3)$$

For the remaining cases, i.e., when $l = K+1$, the conditional probability in (4.3) reduces to

$$P\Big(\mathbf{Y}_{it} = \mathbf{y}_{it}^{(j)} \,|\, \mathbf{Y}_{i,t-1} = \mathbf{y}_{i,t-1}^{(K+1)}\Big) = \frac{\exp\Big(\beta_{j0} + \mathbf{x}_{it}'\boldsymbol{\beta} + \sigma_j\gamma_{ij}\Big)}{1 + \sum_{k=1}^{K}\exp\Big(\beta_{k0} + \mathbf{x}_{it}'\boldsymbol{\beta}_k + \sigma_k\gamma_{ik}\Big)}$$
$$= \tilde{\eta}_{it|t-1}^{*(j)}(K+1), \text{ say}, \quad \text{for } j = 1, \cdots, K, \qquad (4.4)$$

and for any $l = 1, \cdots, K$, the probability for the response to be in the last category $K+1$ (i.e., for $j = K+1$) at time t is given by

$$\tilde{\eta}_{it|t-1}^{*(K+1)}(l) = 1 - \sum_{j=1}^{K}\tilde{\eta}_{it|t-1}^{*(j)}(l) = \frac{1}{1 + \sum_{k=1}^{K}\exp\Big(\beta_{k0} + \mathbf{x}_{it}'\boldsymbol{\beta}_k + \sigma_k\gamma_{ik} + \boldsymbol{\theta}_k'\mathbf{y}_{i,t-1}^{(l)}\Big)} \qquad (4.5)$$

4.1.1 Basic Properties of the MDML Model

Note that the MDML model given by (4.2)-(4.5) is referred to as a history based model. This is because, this model accommodates all responses and covariates information for all individuals ($i = 1, \cdots, I$) over the whole duration of the study for $t = 1, \cdots, T$. We now provide the means, variances and covariances under this MDML model.

Following the Lemma 3.1, the conditional mean vector and the conditional covariance matrix of the multinomial response vector $\mathbf{Y}_{it} = (Y_{it1}, \cdots, Y_{itj}, \cdots, Y_{itK})$ have the forms

$$\begin{aligned} E\Big(\mathbf{Y}_{it}|\gamma_{ij}\Big) &= \boldsymbol{\eta}^*{}_{it|t-1} + [\tilde{\mathbf{A}}_{it} - \boldsymbol{\eta}^*{}_{it|t-1} 1'] \tilde{\mathbf{\Pi}}^*_{i,t-1} \\ &= \Big(\tilde{\pi}^{*(1)}_{it}, \cdots, \tilde{\pi}^{*(j)}_{it}, \cdots, \tilde{\pi}^{*(K)}_{it}\Big)' \\ &= \tilde{\mathbf{\Pi}}^*_{it} \end{aligned} \quad (4.6)$$

and

$$Var\Big(\mathbf{Y}_{it}|\gamma_{ij}\Big) = diag\left[\tilde{\pi}^{*(1)}_{it}, \cdots, \tilde{\pi}^{*(j)}_{it}, \cdots, \tilde{\pi}^{*(K)}_{it}\right] - \tilde{\mathbf{\Pi}}^*_{it} \tilde{\mathbf{\Pi}}^{*\prime}_{it}, \quad (4.7)$$

for all $j = 1, \cdots, K$ and $t = 1, \cdots, T$.

Similarly, following Lemma 3.2, the conditional covariance between the multinomial response vector $\mathbf{Y}_{it}$ and $\mathbf{Y}_{it'}$ at two different time points t and t', $t < t'$, is given by

$$Cov\Big(\mathbf{Y}_{it}, \mathbf{Y}_{it'}|\gamma_{ij}\Big) = Var(\mathbf{Y}_{it}) \prod_{s=t+1}^{t'} [\tilde{\mathbf{A}}_{is} - \boldsymbol{\eta}^*{}_{is|s-1} \mathbf{1}'], \quad (4.8)$$

where

$$\tilde{\mathbf{A}}_{is} = \begin{bmatrix} \tilde{\eta}^{*(1)}_{is|s-1}(1) & \cdots & \tilde{\eta}^{*(1)}_{is|s-1}(l) & \cdots & \tilde{\eta}^{*(1)}_{is|s-1}(K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\eta}^{*(j)}_{is|s-1}(1) & \cdots & \tilde{\eta}^{*(j)}_{is|s-1}(l) & \cdots & \tilde{\eta}^{*(j)}_{is|s-1}(K) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \tilde{\eta}^{*(K)}_{is|s-1}(1) & \cdots & \tilde{\eta}^{*(K)}_{is|s-1}(l) & \cdots & \tilde{\eta}^{*(K)}_{is|s-1}(K) \end{bmatrix}, \quad \boldsymbol{\eta}^*{}_{is|s-1} = \begin{bmatrix} \tilde{\eta}^{*(1)}_{is|s-1}(K+1) \\ \vdots \\ \tilde{\eta}^{*(j)}_{is|s-1}(K+1) \\ \vdots \\ \tilde{\eta}^{*(K)}_{is|s-1}(K+1) \end{bmatrix}$$

Next, by averaging over the distributions of γ_{ij} we can determine the unconditional means, variances and covariances of the MDML model. However, we don't compute these unconditional properties because the conditional properties retain the similar patterns as the unconditional properties under the MDML model. We rather concentrate to the unconditional likelihood inferences in Section 4.2.

4.2 Unconditional Likelihood Estimation for the MDFL Model

Let $\boldsymbol{\beta} \equiv (\beta_{j0}, \beta_{ju})$, $\boldsymbol{\theta} \equiv (\theta_{jc})$ and $\sigma \equiv (\sigma_j)$ for $j = 1, \cdots, K$, $c = 1, \cdots, K$, $u = 1, \cdots, p$. As opposed to the likelihood function (3.14) under the MDFL model, we may derive the unconditional likelihood under the MDML model as

$$L\Big(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma\Big) = \prod_{i=1}^{I} \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} f^*(y_{i1}) f^*(y_{i2}|\, y_{i1}) \cdots f^*(y_{it}|\, y_{i,t-1}) \times \phi(\gamma_{i1}) \cdots \phi(\gamma_{iK})\, d\gamma_{i1} \cdots d\gamma_{iK}, \quad (4.9)$$

where $\phi(\gamma_{ij})$ is the standard normal density, and σ_j^2's are additional random effects variance parameters. In (4.9),

$$f^*\Big(y_{i1}\Big) = \frac{1}{y_{i11}! \cdots y_{i1K}!, (1 - \sum_{q=1}^{K} y_{i1q})!} \prod_{q=1}^{K} \left\{\tilde{\pi}_{i1}^{*(q)}\right\}^{y_{i1q}} \times \left\{1 - \sum_{q=1}^{K} \tilde{\pi}_{i1}^{*(q)}\right\}^{(1-\sum_{q=1}^{K} y_{i1q})}, \quad (4.10)$$

and

$$f^*\left(y_{it} \mid y_{i,t-1}\right) = \frac{1}{y_{it1}! \cdots y_{itk}!, \, (1 - \sum_{q=1}^{K} y_{itq})!} \prod_{q=1}^{K} \left\{\tilde{\eta}_{it|t-1}^{*(q)}\right\}^{y_{itq}} \times \left\{1 - \sum_{q=1}^{K} \tilde{\eta}_{it|t-1}^{*(q)}\right\}^{\left(1-\sum_{q=1}^{K} y_{itq}\right)}, \quad (4.11)$$

where $\tilde{\pi}_{i1}^{*(q)}$ in (4.10) and $\tilde{\eta}_{it|t-1}^{*(q)}$ in (4.11) are given by (4.2) and (4.3), respectively. After some algebra, by using (4.10) and (4.11) in (4.9), we obtain the likelihood function as

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma) &= C\left[\exp\left\{\sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{K} y_{itj}\left(x_{it}^{*\prime}\boldsymbol{\beta}_j^* + \boldsymbol{\theta}_j' y_{i,t-1}\right)\right\}\right] \\ &\quad \times \left[\prod_{i=1}^{I} \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \frac{\exp\left\{\sum_{j=1}^{K} \sigma_j \gamma_{ij}\left(\sum_{t=1}^{T} y_{itj}\right)\right\}}{\prod_{t=1}^{T}\left\{1 + \sum_{k=1}^{K} \exp\left(x_{it}^{*\prime}\boldsymbol{\beta}_k^* + \sigma_k\gamma_{ik} + \boldsymbol{\theta}_k' y_{i,t-1}\right)\right\}} \times \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}\right], \quad (4.12) \end{aligned}$$

where $\boldsymbol{\beta}_j^*$ and $\mathbf{x}_{it}^*$ are defined as in (2.36)-(2.37).

Note that the integrations in (4.12) makes the likelihood computation complicated, whereas under the MDFL model the likelihood computation by (3.14) is much simpler. Consequently, the log-likelihood function under the present MDML model also becomes complicated. More specifically, as opposed to the log-likelihood function (3.17), we now write the unconditional log-likelihood function as

$$\ln\left(\boldsymbol{\beta},\boldsymbol{\theta},\sigma\right) = \ln C + \sum_{i=1}^{I}\sum_{t=1}^{T}\sum_{j=1}^{K} y_{itj}\,(x_{it}^{*\prime}\boldsymbol{\beta}_j^* + \boldsymbol{\theta}_j' y_{i,t-1}) + \sum_{i}^{I}\ln V_i\,, \tag{4.13}$$

where

$$V_i = \int_{\gamma_{i1}}\cdots\int_{\gamma_{iK}} \exp\{\delta_i\}\,U_i\,\phi(\gamma_{i1})\cdots\phi(\gamma_{iK})\,d\gamma_{i1}\cdots d\gamma_{iK} \tag{4.14}$$

with $\quad \boldsymbol{\gamma}_i = \left(\gamma_{i1}\cdots\gamma_{iK}\right)'$,

$$U_i(\boldsymbol{\gamma}_i) = \left[\prod_{t=1}^{T}\left\{1+\sum_{k=1}^{K}\exp\left(x_{it}^{*\prime}\boldsymbol{\beta}_k^* + \sigma_k\gamma_{ik} + \theta_k' y_{i,t-1}\right)\right\}\right]^{-1} = \prod_{t=1}^{T}\tilde{\eta}_{it|t-1}^{*(K+1)} \tag{4.15}$$

and

$$\delta_i(\boldsymbol{\gamma}_i) = \sum_{j=1}^{K}\sigma_j\gamma_{ij}\left(\sum_{t=1}^{T} y_{itj}\right). \tag{4.16}$$

4.2.1 First and Second Order Derivatives

4.2.1.1 Derivatives of the log-likelihood function with respect to $\boldsymbol{\beta}_j^*$

Note that even though $\boldsymbol{\beta}_j$ is the same regression effects of $\mathbf{x}_{it}$ on $\mathbf{y}_{it}^{(j)}$ as under the MDFL model, the likelihood estimating equations for $\boldsymbol{\beta}_j^*$ under the MDML model are different than those under the MDFL model [(3.25) and (3.30)]. For $j = 1,\cdots,K$, we now write the first and second order derivatives of the log-likelihood function with respect to $\boldsymbol{\beta}_j^*$ as follows:

$$\frac{\partial \ln\Big(\boldsymbol{\beta},\boldsymbol{\theta},\sigma\Big)}{\partial \boldsymbol{\beta}_j^*} \;=\; \sum_{i,t} y_{itj} x_{it}^* - \sum_i \frac{M_{ij}}{V_i} \tag{4.17}$$

$$\frac{\partial^2 \ln L\Big(\boldsymbol{\beta},\boldsymbol{\theta},\sigma\Big)}{\partial \boldsymbol{\beta}_j^* \, \partial \boldsymbol{\beta}_k^{*\prime}} = \begin{cases} -\sum_i \frac{1}{V_i^2}\left[V_i M_{ij\boldsymbol{\beta}_j} + M_{ij} M_{ij}^\prime\right]; & \text{for} \quad j = k \\ -\sum_i \frac{1}{V_i^2}\left[-V_i M_{ij\boldsymbol{\beta}_k} + M_{ij} M_{ik}^\prime\right]; & \text{for} \quad j \neq k \end{cases} \tag{4.18}$$

where

$$M_{ij} \;=\; -\frac{\partial V_i}{\partial \boldsymbol{\beta}_j^*} = -\int_{\gamma_{i1}} \cdots \int_{\gamma_{ik}} \exp\{\delta_i\} \left(\frac{\partial U_i}{\partial \boldsymbol{\beta}_j^*}\right) \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}\,, \tag{4.19}$$

$$\begin{aligned} \frac{\partial U_i}{\partial \boldsymbol{\beta}_j^*} \;&=\; U_i \left[\frac{\partial \ln U_i}{\partial \boldsymbol{\beta}_j^*}\right] = -U_i \left[\frac{\partial}{\partial \boldsymbol{\beta}_j^*} \sum_{t=1}^{T} \ln\left\{1 + \sum_{k=1}^{K} \exp\left(x_{it}^{*\prime}\boldsymbol{\beta}_k^* + \sigma_k\gamma_{ik} + \theta_k^\prime y_{i,t-1}\right)\right\}\right] \\ &=\; -U_i \sum_{t=1}^{T} \left[\frac{\exp\left(x_{it}^{*\prime}\boldsymbol{\beta}_j^* + \sigma_j\gamma_{ij} + \theta_j^\prime y_{i,t-1}\right) x_{it}^*}{\left\{1 + \sum_{k=1}^{K} \exp\left(x_{it}^{*\prime}\boldsymbol{\beta}_k^* + \sigma_k\gamma_{ik} + \theta_k^\prime y_{i,t-1}\right)\right\}}\right] \\ &=\; -U_i \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)} x_{it}^*\,, \end{aligned} \tag{4.20}$$

and

$$M_{ij\beta_l} \;=\; \frac{\partial M_{ij}}{\partial \boldsymbol{\beta}_l^*}$$

$$= \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \left\{ \sum_{t=1}^{T} \left(\frac{\partial \tilde{\eta}^{*(j)}_{it|t-1}}{\partial \beta^*_l} \right) x^*_{it} \right\} \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}, \quad (4.21)$$

with

$$\frac{\partial \tilde{\eta}^{*(j)}_{it|t-1}}{\partial \beta^*_l} = \begin{cases} \tilde{\eta}^{*(j)}_{it|t-1} \left(1 - \tilde{\eta}^{*(j)}_{it|t-1}\right) x^*_{it}\,; & \text{for } l = j \\ \tilde{\eta}^{*(j)}_{it|t-1}\, \tilde{\eta}^{*(k)}_{it|t-1}\, x^*_{it}\,; & \text{for } l = k \end{cases} \quad (4.22)$$

4.2.1.2 Derivatives of the log-likelihood function with respect to θ_j

Further Note that even though θ_j is the same dynamic dependence parameters as under the MDFL model, the likelihood estimating equations for θ_j under the MDML model are complicated as compared to those under the MDFL model [(3.26) and (3.30)]. For $j = 1, \cdots, K$, we now write the first and second order derivatives of the log-likelihood function with respect to θ_j as follows:

$$\frac{\partial \ln L\big(\beta, \theta, \sigma\big)}{\partial \theta_j} = \sum_{i,t} y_{itj} y_{i,t-1} - \sum_{i} \frac{N_{ij}}{V_i} \quad (4.23)$$

$$\frac{\partial^2 \ln L\big(\beta, \theta, \sigma\big)}{\partial \theta_j\, \partial \theta'_k} = \begin{cases} -\sum_i \frac{1}{V_i^2} \left[V_i\, N_{ij\theta_j} + N_{ij} N'_{ij} \right]; & \text{for } j = k \\ -\sum_i \frac{1}{V_i^2} \left[- V_i\, N_{ij\theta_k} + N_{ij} N'_{ik} \right]; & \text{for } j \neq k \end{cases} \quad (4.24)$$

where

$$N_{ij} = -\frac{\partial V_i}{\partial \theta_j} = -\int_{\gamma_{i1}} .... \int_{\gamma_{ik}} \exp\{\delta_i\} \left(\frac{\partial U_i}{\partial \theta_j}\right) \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}\,, \tag{4.25}$$

$$\begin{aligned}
\frac{\partial U_i}{\partial \theta_j} &= U_i \left[\frac{\partial \ln U_i}{\partial \theta_j}\right] = -U_i \left[\frac{\partial}{\partial \theta_j} \sum_{t=1}^{T} \ln\left\{1 + \sum_{k=1}^{K} \exp\left(x_{it}^{*\prime} \boldsymbol{\beta}_k^* + \sigma_k \gamma_{ik} + \theta_k' y_{i,t-1}\right)\right\}\right] \\
&= -U_i \sum_{t=1}^{T} \left[\frac{\exp\left(x_{it}^{*\prime} \boldsymbol{\beta}_j^* + \sigma_j \gamma_{ij} + \theta_j' y_{i,t-1}\right) y_{i,t-1}}{\left\{1 + \sum_{k=1}^{K} \exp\left(x_{it}^{*\prime} \boldsymbol{\beta}_k^* + \sigma_k \gamma_{ik} + \theta_k' y_{i,t-1}\right)\right\}}\right] \\
&= -U_i \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)}\, y_{i,t-1}\,,
\end{aligned} \tag{4.26}$$

and

$$\begin{aligned}
N_{ij\theta_l} &= \frac{\partial N_{ij}}{\partial \theta_l} \\
&= \int_{\gamma_{i1}} .... \int_{\gamma_{iK}} \exp\{\delta_i\}\, U_i \left\{\sum_{t=1}^{T} \left(\frac{\partial \tilde{\eta}_{it|t-1}^{*(j)}}{\partial \theta_l}\right) y_{i,t-1}\right\} \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}\,,
\end{aligned} \tag{4.27}$$

with

$$\frac{\partial \tilde{\eta}_{it|t-1}^{*(j)}}{\partial \theta_l} = \begin{cases} \tilde{\eta}_{it|t-1}^{*(j)} \left(1 - \tilde{\eta}_{it|t-1}^{*(j)}\right) y_{i,t-1}\,; & \text{for } l = j \\ \tilde{\eta}_{it|t-1}^{*(j)}\, \tilde{\eta}_{it|t-1}^{*(k)}\, y_{i,t-1}\,; & \text{for } l = k \end{cases} \tag{4.28}$$

4.2.1.3 Derivatives of the log-likelihood function with respect to σ_j

As opposed to the MDFL model, the MDML model contain a third parameter σ_j^2 which represent the random effects variances. For $j = 1, \cdots, K$, we now write the first and second order derivatives of the log-likelihood function with respect to σ_j as follows:

$$\frac{\partial \ln L\Big(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma\Big)}{\partial \sigma_j} = \sum_i \frac{O_{ij}}{V_i} \tag{4.29}$$

$$\frac{\partial^2 \ln L\Big(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma\Big)}{\partial \sigma_j\, \partial \sigma_k} = \sum_{i,t} \frac{1}{V_i^2}\left[V_i O_{ij\sigma_k} - O_{ij} O_{ik}\right] \tag{4.30}$$

where

$$O_{ij} = \frac{\partial V_i}{\partial \sigma_j} = \int_{\gamma_{i1}} \cdots \int_{\gamma_{ik}} \left[\frac{\partial \exp\{\delta_i\}}{\partial \sigma_j} U_i + \exp\{\delta_i\} \frac{\partial U_i}{\partial \sigma_j}\right] \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}, \tag{4.31}$$

with

$$\frac{\partial \exp\{\delta_i\}}{\partial \sigma_j} = \exp\{\delta_i\}\, \gamma_{ij} \left(\sum_{t=1}^{T} y_{itj}\right)$$

and

$$\frac{\partial U_i}{\partial \sigma_j} = -U_i \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)}\, \gamma_{ij}.$$

In (4.30),

$$
O_{ij\sigma_l} = \frac{\partial O_{ij}}{\partial \sigma_l}
$$

$$
= \begin{cases} \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \gamma_{ij}^2 \Bigg[\left(\sum_{t=1}^{T} y_{itj} - \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)} \right)^2 \\ \quad - \left\{ \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)} \left(1 - \tilde{\eta}_{it|t-1}^{*(j)}\right) \right\} \Bigg] \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}; \quad \text{for} \quad l = j \\ \\ \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \gamma_{ij} \gamma_{ik} \Bigg[\left(\sum_{t=1}^{T} y_{itj} - \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)} \right) \left(\sum_{t=1}^{T} y_{itk} - \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(k)} \right) \\ \quad + \left\{ \sum_{t=1}^{T} \tilde{\eta}_{it|t-1}^{*(j)} \tilde{\eta}_{it|t-1}^{*(m)} \right\} \Bigg] \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}; \quad \text{for} \quad l = k \end{cases} \tag{4.32}
$$

4.2.1.4 Cross Derivatives of the log-likelihood function with respect to $\boldsymbol{\beta}_j$, $\boldsymbol{\theta}_j$ and σ_j

The cross derivatives of the log-likelihood function with respect to $\boldsymbol{\beta}_j$, $\boldsymbol{\theta}_j$ and σ_j have the following forms:

$$
\frac{\partial^2 \ln L\left(\boldsymbol{\beta}, \boldsymbol{\theta}, \sigma\right)}{\partial Q_j\, \partial R_k'} = \begin{cases} -\sum_{i,t} \frac{1}{V_i^2} \left[V_i\, M_{ij\theta_k} + M_{ij} N_{ik}' \right]; & \text{for} \quad Q_j = \beta_j^*,\ R_k = \theta_k \\ -\sum_{i,t} \frac{1}{V_i^2} \left[V_i\, M_{ij\sigma_k} + M_{ij} O_{ik}' \right]; & \text{for} \quad Q_j = \beta_j^*,\ R_k = \sigma_k \\ -\sum_{i,t} \frac{1}{V_i^2} \left[V_i\, N_{ij\sigma_k} + N_{ij} O_{ik}' \right]; & \text{for} \quad Q_j = \theta_j,\ R_k = \sigma_k \end{cases} \tag{4.33}
$$

where

$$
\begin{aligned}
M_{ij\theta_j} &= \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \left[\left\{ \sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} \left(1 - \tilde{\eta}^{*(j)}_{it|t-1}\right) x^*_{it} y'_{i,t-1} \right\} \right. \\
&\quad \left. - \left\{ \sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} x^*_{it} \sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} y'_{i,t-1} \right\} \right] \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik},
\end{aligned}
$$

$$
\begin{aligned}
M_{ij\sigma_j} &= \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \gamma_{ij} \left[\sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} \left(1 - \tilde{\eta}^{*(j)}_{it|t-1}\right) x^*_{it} + \left(\sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} x^*_{it} \right) \right. \\
&\quad \left. \times \left(\sum_{t=1}^{T} y_{itj} - \sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} \right) \right] \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik},
\end{aligned}
$$

and

$$
\begin{aligned}
N_{ij\sigma_j} &= \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \gamma_{ij} \left[\sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} \left(1 - \tilde{\eta}^{*(j)}_{it|t-1}\right) y_{i,t-1} \right. \\
&\quad \left. + \left(\sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} y_{i,t-1} \right) \left(\sum_{t=1}^{T} y_{itj} - \sum_{t=1}^{T} \tilde{\eta}^{*(j)}_{it|t-1} \right) \right] \prod_{k=1}^{K} \phi(\gamma_{ik})\, d\gamma_{ik}.
\end{aligned}
$$

4.2.2 Computation Aspect

Note that first and second order derivatives shown in Section 4.2.1 contain multiple integrations over random effects. Because the exact likelihood computation is almost impossible due to integration problem, we approximate the integrations over the random effects by using a simulation approach [Fahrmeier and Tutz (1994), Jiang (1998), Sutradhar (2011, Section 5.1.1)]. For example, consider the computation of

V_i which requires multiple integrations as shown in (4.14). To exploit the simulation approach, we first generate 5000 standard normal vectors

$$\gamma_i^{(w)} = \left(\gamma_{i1}^{(w)} \cdots \gamma_{iK}^{(w)}\right)'; \quad \text{for } w = 1, \cdots, 5000, \quad \text{say}$$

and then compute approximate V_i as

$$\begin{aligned} V_i &= \int_{\gamma_{i1}} \cdots \int_{\gamma_{iK}} \exp\{\delta_i\} U_i \, \phi(\gamma_{i1}) \cdots \phi(\gamma_{iK}) \, d\gamma_{i1} \cdots d\gamma_{iK} \\ &\simeq \frac{1}{5000} \sum_{w=1}^{5000} \exp\left\{\delta_i\left(\gamma_i^{(w)}\right)\right\} U_i\left(\gamma_i^{(w)}\right) \end{aligned} \qquad (4.34)$$

4.3 A Simulation Study

4.3.1 Simulation Design

Recall that in Section 3.1.3 we examined the finite sample performance of the likelihood estimation approach in estimating the parameters of the MDFL model through a simulation study. In this section, we conduct a simulation study for the MDML model. Thus, in addition to the estimation performance for $\boldsymbol{\beta}_j : p \times 1$ and $\boldsymbol{\theta}_j : K \times 1$ $(j = 1, \cdots, K)$, we now also examine the performance for the likelihood estimation of σ_j^2, the variance component of the random effects γ_{ij}. Unlike Chapter 3, for the simulation purpose, by using a true set of values for the components of $\boldsymbol{\beta}$, $\boldsymbol{\theta}$ and σ, we generate the initial multinomial response $\mathbf{y}_{i1} = (y_{i11}, \cdots, y_{i1K})'$ $(i = 1, \cdots, I)$ following (4.2), and $\mathbf{y}_{it} = (y_{it1}, \cdots, y_{itK})'$ for $t = 2, \cdots, T$, following (4.3)-(4.5). As far as the covariate design is concerned, we use the same covariate $x_{it(1)}$ and $x_{it(2)}$ as in Section 3.1.3.1 under the MDFL model. The true values for the regression $(\boldsymbol{\beta}_j)$ and dynamic dependence $(\boldsymbol{\theta}_j)$ parameters are also chosen to be the same as in Section 3.1.3.1 under the MDFL model. Furthermore, we choose the same parameter dimension and sample size as the MDFL model. Thus, we retain $I = 100$ individuals, $T = 4$ time points, $p = 2$ covariates, and $K = 2$ (that is $K + 1 = 3$ categories).

With regard to the selection of the additional variance components under various categories of the MDML model, we now choose the true values of σ_j as

$$\sigma' = (\sigma_1, \sigma_2) = \begin{cases} (0.5, & 0.6) \\ (1.0, & 0.9) \\ (1.2, & 1.75) \end{cases}$$

These values of σ_j^2 are chosen to reflect small (such as $\sigma_1^2 = 0.25$) and large (such as σ_2^2=3.0625) random effects variances.

4.3.2 Simulation Results

Using the covariates and parameters as mentioned in Section 4.3.1, we generate the multinomial responses under the present MDML model following the marginal probability (4.2) and conditional probability (4.3). Next we solve the likelihood estimating equations (4.17), (4.23) and (4.29) for $\boldsymbol{\beta}_j$, $\boldsymbol{\theta}_j$ and σ_j, respectively. The data generation and estimation are repeated for 500 times. The simulated mean (SM) and simulated standard error (SSE) of the estimates are shown in Tables 4.1 through 4.6. More specifically, corresponding to the Table 3.1 of Chapter 3, we now present the simulation results in Tables 4.1, 4.2 and 4.3 for the selection of 3 sets of true values of σ_j parameters. Similarly, corresponding to Table 3.2 of Chapter 3, we exhibit the present simulation results in Tables 4.4, 4,5 and 4.6.

The results in the Tables 4.1, 4.2 and 4.3 show that the likelihood approach performs well in estimating the regression effects and dynamic dependence parameters even if the random effects variances are large. For example, for $\sigma_1 = 0.5$ and $\sigma_2 = 0.6$, the results in Table 4.1 show that $\beta_{11} = 0.0$, $\beta_{12} = 0.0$, $\beta_{21} = 0.0$, $\beta_{22} = 0.0$, and $\theta_{11} = 0.3$, $\theta_{12} = 0.7$, $\theta_{21} = 0.8$, $\theta_{22} = 0.5$ are estimated as

$$\hat{\beta}_{11} = 0.015, \ \hat{\beta}_{12} = -0.008, \ \hat{\beta}_{21} = 0.007, \ \hat{\beta}_{22} = -0.006$$

and

$$\hat{\theta}_{11} = 0.245,\ \hat{\theta}_{12} = 0.687,\ \hat{\theta}_{21} = 0.800,\ \hat{\theta}_{22} = 0.462,$$

respectively, showing good agrement between estimates and parameters. In this case, $\sigma_1 = 0.5$ and $\sigma_2 = 0.6$ are estimated as $\hat{\sigma}_1 = 0.466$ and $\hat{\sigma}_2 = 0.560$, respectively, showing slightly underestimation. For large random effects variances i.e., $\sigma_1 = 1.2$ and $\sigma_2 = 1.75$, Table 4.3 shows that the estimates for the same regression and dynamic dependence parameters are as follows:

$$\hat{\beta}_{11} = 0.000,\ \hat{\beta}_{12} = -0.012,\ \hat{\beta}_{21} = -0.010,\ \hat{\beta}_{22} = -0.019$$

and

$$\hat{\theta}_{11} = 0.316,\ \hat{\theta}_{12} = 0.717,\ \hat{\theta}_{21} = 0.829,\ \hat{\theta}_{22} = 0.545,$$

also showing good agrement between estimates and parameters. In this case σ_1 and σ_2 are estimated as $\hat{\sigma}_1 = 1.185$ and $\hat{\sigma}_2 = 1.769$, respectively. As far as the estimation of the intercepts under the categories are concerned, they appear to be estimated well when the variance components are large.

Furthermore, the results in these three tables indicate that as the values of σ_1 and σ_2 get larger, the standard errors of the estimates for all parameters also get large, as expected.

Table 4.1: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* $\boldsymbol{\theta}$ *and the True Value of* $\boldsymbol{\beta} \equiv (\beta_{10} = 0.1, \beta_{11} = 0.0, \beta_{12} = 0.0, \beta_{20} = 0.2, \beta_{21} = 0.0, \beta_{22} = 0.0)$ *and* $\sigma \equiv (\sigma_1 = 0.5, \sigma_2 = 0.6)$*, Under the MDML Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2,\ \theta_{12} = 0.0$; $\theta_{21} = 0.0,\ \theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.10363	0.03256	-0.00178	0.19006	0.02657	0.00924
SSE	0.29809	0.28645	0.28501	0.28754	0.27405	0.30164
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.47692	0.12837	-0.02670	0.54607	-0.02537	0.06701
SSE	0.30933	0.35560	0.32677	0.31631	0.32704	0.33688

True	$\theta_{11} = 0.8,\ \theta_{12} = 0.3$; $\theta_{21} = 0.3,\ \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.13777	0.00976	0.00045	0.23534	0.03344	0.00900
SSE	0.31739	0.31178	0.31697	0.32033	0.31078	0.32799
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.50048	0.70830	0.26930	0.59087	0.25251	0.71463
SSE	0.34490	0.36446	0.35134	0.32659	0.38399	0.35333

True	$\theta_{11} = 0.3,\ \theta_{12} = 0.7$; $\theta_{21} = 0.8,\ \theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.14582	0.01492	-0.00828	0.26312	0.00700	-0.00629
SSE	0.29852	0.31315	0.30235	0.28481	0.29231	0.29329
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.46633	0.24486	0.68738	0.55967	0.79998	0.46236
SSE	0.31381	0.39242	0.34911	0.29625	0.37290	0.37073

True	$\theta_{11} = 0.8,\ \theta_{12} = -0.5$; $\theta_{21} = -0.5,\ \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12136	0.04289	-0.01321	0.25479	0.02150	-0.02059
SSE	0.30215	0.32144	0.31154	0.30597	0.30877	0.31802
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.50738	0.69618	-0.55242	0.56418	-0.56697	0.73605
SSE	0.37023	0.35020	0.40835	0.36362	0.41108	0.33706

Table 4.2: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* $\boldsymbol{\theta}$ *and the True Value of* $\boldsymbol{\beta} \equiv (\beta_{10} = 0.1, \beta_{11} = 0.0, \beta_{12} = 0.0, \beta_{20} = 0.2, \beta_{21} = 0.0, \beta_{22} = 0.0)$ *and* $\sigma \equiv (\sigma_1 = 1.0, \sigma_2 = 0.9)$*, Under the MDML Model.*

	Category 1 vs 3			**Category 2 vs 3**		
True	$\theta_{11} = 0.2, \theta_{12} = 0.0$; $\theta_{21} = 0.0, \theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11328	-0.01389	-0.01013	0.18593	-0.00146	0.01343
SSE	0.33838	0.30143	0.36727	0.32431	0.30654	0.33636
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.94639	0.21316	-0.00139	0.84404	0.00466	0.10633
SSE	0.33425	0.38765	0.36730	0.32485	0.36618	0.37827
True	$\theta_{11} = 0.8, \theta_{12} = 0.3$; $\theta_{21} = 0.3, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12043	-0.00228	-0.00656	0.23485	0.00873	-0.00939
SSE	0.35486	0.32091	0.39558	0.32429	0.33636	0.34599
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.95077	0.81424	0.25935	0.86523	0.28420	0.76166
SSE	0.36430	0.40539	0.36712	0.33894	0.40915	0.36935
True	$\theta_{11} = 0.3, \theta_{12} = 0.7$; $\theta_{21} = 0.8, \theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11072	0.02053	-0.00026	0.21294	0.03159	0.01799
SSE	0.33224	0.32068	0.35647	0.32516	0.31950	0.34641
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.95097	0.29976	0.68266	0.88419	0.77818	0.45466
SSE	0.30267	0.41894	0.36644	0.31672	0.37439	0.37802
True	$\theta_{11} = 0.8, \theta_{12} = -0.5$; $\theta_{21} = -0.5, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11451	0.01850	-0.01449	0.21345	0.01067	0.00661
SSE	0.35498	0.33556	0.37708	0.33965	0.31250	0.35241
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.94604	0.80465	-0.55410	0.86399	-0.55415	0.77572
SSE	0.39291	0.38365	0.44344	0.40684	0.45211	0.36983

Table 4.3: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* $\boldsymbol{\theta}$ *and the True Value of* $\boldsymbol{\beta} \equiv (\beta_{10} = 0.1, \beta_{11} = 0.0, \beta_{12} = 0.0, \beta_{20} = 0.2, \beta_{21} = 0.0, \beta_{22} = 0.0)$ *and* $\sigma \equiv (\sigma_1 = 1.2, \sigma_2 = 1.75)$*, Under the MDML Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2, \theta_{12} = 0.0$; $\theta_{21} = 0.0, \theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.06165	-0.01325	-0.00671	0.18157	-0.05334	0.00132
SSE	0.36304	0.33532	0.40198	0.40940	0.35925	0.46001
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.13808	0.27906	0.03467	1.69751	0.04942	0.20019
SSE	0.35822	0.41974	0.40381	0.38975	0.45111	0.40868
True	$\theta_{11} = 0.8, \theta_{12} = 0.3$; $\theta_{21} = 0.3, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.10860	-0.04231	0.03201	0.20482	-0.02913	0.01971
SSE	0.39002	0.36411	0.41867	0.47381	0.39923	0.49440
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.15418	0.83250	0.30659	1.69552	0.31071	0.87368
SSE	0.38643	0.43022	0.44603	0.42729	0.49292	0.44747
True	$\theta_{11} = 0.3, \theta_{12} = 0.7$; $\theta_{21} = 0.8, \theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11054	-0.00005	-0.01191	0.19438	-0.00990	-0.01958
SSE	0.37421	0.35677	0.39914	0.44936	0.37879	0.48087
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.18512	0.31596	0.71710	1.76878	0.82906	0.54510
SSE	0.36551	0.41818	0.44952	0.38576	0.43466	0.44380
True	$\theta_{11} = 0.8, \theta_{12} = -0.5$; $\theta_{21} = -0.5, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.10160	-0.02237	0.00163	0.20342	-0.04749	0.00794
SSE	0.37699	0.36650	0.40995	0.45208	0.39301	0.50287
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.13405	0.84476	-0.52439	1.68468	-0.56548	0.88184
SSE	0.46333	0.42717	0.51656	0.48421	0.54774	0.46408

Table 4.4: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* $\boldsymbol{\theta}$ *and the True Value of* $\boldsymbol{\beta} \equiv (\beta_{10} = 0.1, \beta_{11} = 0.3, \beta_{12} = 0.1, \beta_{20} = 0.2, \beta_{21} = -0.2, \beta_{22} = 0.0)$ *and* $\sigma \equiv (\sigma_1 = 0.5, \sigma_2 = 0.6)$*, Under the MDML Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2, \theta_{12} = 0.0$; $\theta_{21} = 0.0, \theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.09229	0.33861	0.10992	0.21642	-0.20478	-0.00575
SSE	0.29392	0.29021	0.27435	0.29721	0.31258	0.30831
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.47199	0.14044	-0.02873	0.56499	-0.03908	0.02664
SSE	0.30229	0.32754	0.32874	0.32182	0.33124	0.39392
True	$\theta_{11} = 0.8, \theta_{12} = 0.3$; $\theta_{21} = 0.3, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.14937	0.35859	0.08503	0.26592	-0.17081	-0.01980
SSE	0.29977	0.31039	0.32069	0.30091	0.31102	0.32843
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.51474	0.67828	0.24433	0.55733	0.23457	0.68028
SSE	0.32137	0.33264	0.37454	0.34651	0.36139	0.38281
True	$\theta_{11} = 0.3, \theta_{12} = 0.7$; $\theta_{21} = 0.8, \theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12657	0.36112	0.09132	0.24065	-0.15696	-0.01262
SSE	0.30393	0.29079	0.30686	0.30113	0.31454	0.31905
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.46683	0.22714	0.66891	0.54167	0.76959	0.42435
SSE	0.30742	0.36235	0.37966	0.30943	0.35239	0.41359
True	$\theta_{11} = 0.8, \theta_{12} = -0.5$; $\theta_{21} = -0.5, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.13066	0.34871	0.09528	0.22675	-0.18622	0.00589
SSE	0.29630	0.30348	0.28275	0.31359	0.33487	0.31936
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.49433	0.68665	-0.54706	0.58885	-0.56853	0.71103
SSE	0.35108	0.32719	0.37658	0.37930	0.41948	0.36276

Table 4.5: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* $\boldsymbol{\theta}$ *and the True Value of* $\boldsymbol{\beta} \equiv (\beta_{10} = 0.1, \beta_{11} = 0.3, \beta_{12} = 0.1, \beta_{20} = 0.2, \beta_{21} = -0.2, \beta_{22} = 0.0)$ *and* $\sigma \equiv (\sigma_1 = 1.0, \sigma_2 = 0.9)$*, Under the MDML Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2,\ \theta_{12} = 0.0$; $\theta_{21} = 0.0,\ \theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.10062	0.29985	0.11140	0.19856	-0.19983	0.02073
SSE	0.34318	0.32436	0.36053	0.32646	0.31309	0.33169
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.97994	0.19667	-0.01847	0.84238	-0.01227	0.09836
SSE	0.30155	0.34811	0.40638	0.34139	0.37498	0.42302

True	$\theta_{11} = 0.8,\ \theta_{12} = 0.3$; $\theta_{21} = 0.3,\ \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.09792	0.34835	0.08586	0.22661	-0.18866	0.00389
SSE	0.34820	0.32386	0.35823	0.33118	0.321667	0.33831
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.98934	0.74352	0.29722	0.84268	0.27998	0.78617
SSE	0.35277	0.39236	0.40611	0.40193	0.41933	0.42743

True	$\theta_{11} = 0.3,\ \theta_{12} = 0.7$; $\theta_{21} = 0.8,\ \theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.14397	0.32397	0.06986	0.22881	-0.15399	-0.00569
SSE	0.35413	0.34632	0.38840	0.33517	0.323968	0.34928
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.98019	0.27360	0.70048	0.85570	0.76017	0.46678
SSE	0.29829	0.39231	0.38296	0.34199	0.36628	0.41613

True	$\theta_{11} = 0.8,\ \theta_{12} = -0.5$; $\theta_{21} = -0.5,\ \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.12761	0.30306	0.09424	0.21080	-0.20789	-0.00700
SSE	0.36576	0.33817	0.39400	0.33914	0.33983	0.38379
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	0.96499	0.78549	-0.54358	0.83783	-0.51877	0.78795
SSE	0.39205	0.37668	0.44313	0.40955	0.45718	0.39349

Table 4.6: *Simulated Means (SMs) and Simulated Standard Errors (SSEs)of the Likelihood Estimates for the Regression and Dynamic Dependence Parameters for Some Selected Values of* $\boldsymbol{\theta}$ *and the True Value of* $\boldsymbol{\beta} \equiv (\beta_{10} = 0.1, \beta_{11} = 0.3, \beta_{12} = 0.1, \beta_{20} = 0.2, \beta_{21} = -0.2, \beta_{22} = 0.0)$ *and* $\sigma \equiv (\sigma_1 = 1.2, \sigma_2 = 1.75)$*, Under the MDML Model.*

	Category 1 vs 3			Category 2 vs 3		
True	$\theta_{11} = 0.2, \theta_{12} = 0.0$; $\theta_{21} = 0.0, \theta_{22} = 0.1$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.09427	0.25060	0.11247	0.21489	-0.25123	-0.04084
SSE	0.37712	0.34119	0.38884	0.42134	0.359182	0.458291
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.15486	0.22880	0.05105	1.64657	0.01455	0.24260
SSE	0.35446	0.37856	0.41946	0.40558	0.44402	0.42179

True	$\theta_{11} = 0.8, \theta_{12} = 0.3$; $\theta_{21} = 0.3, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.11883	0.30271	0.07953	0.21211	-0.24970	0.00226
SSE	0.40518	0.39409	0.44028	0.44676	0.390132	0.49545
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.16167	0.79968	0.31012	1.70964	0.29011	0.90151
SSE	0.38159	0.41989	0.48702	0.44735	0.46396	0.47332

True	$\theta_{11} = 0.3, \theta_{12} = 0.7$; $\theta_{21} = 0.8, \theta_{22} = 0.5$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.09685	0.32649	0.08281	0.19342	-0.19262	0.00948
SSE	0.36810	0.34389	0.38153	0.44411	0.371932	0.50916
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.16615	0.29825	0.67625	1.69524	0.79591	0.53493
SSE	0.34628	0.41503	0.44498	0.40684	0.44194	0.46107

True	$\theta_{11} = 0.8, \theta_{12} = -0.5$; $\theta_{21} = -0.5, \theta_{22} = 0.8$					
	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$	$\hat{\beta}_{20}$	$\hat{\beta}_{21}$	$\hat{\beta}_{22}$
SM	0.09600	0.26230	0.13661	0.21660	-0.25836	-0.01163
SSE	0.41804	0.38373	0.42763	0.43866	0.38499	0.489767
	$\hat{\sigma}_1$	$\hat{\theta}_{11}$	$\hat{\theta}_{12}$	$\hat{\sigma}_2$	$\hat{\theta}_{21}$	$\hat{\theta}_{22}$
SM	1.16613	0.82204	-0.51500	1.68195	-0.53980	0.92939
SSE	0.41954	0.44239	0.51662	0.49074	0.56894	0.48181

Note that the simulation results for non-zero but small regression effects are shown in Tables 4.4 to 4.6. The likelihood estimates for all the parameters including regression effects, dynamic dependence and variance components parameters appear to exhibit similar pattern as that of Tables 4.1 to 4.3. Thus, no additional interpretation is given for the results from Tables 4.4 to 4.6.

Chapter 5

Concluding Remarks

Even though in many practical situations categorical responses are collected over time [e.g., Fienberg et al. (1985), Conaway (1989)], the analysis of this type of data has been hampered because of lack of proper modeling and methodological developments. Some of the existing studies (Section 1.1.2.1) have modeled the multinomial longitudinal data by treating time as a fixed categorical covariate and hence ignoring the longitudinal correlations among the responses. In the thesis, we have developed longitudinal correlations based multinomial models where time has been treated as a stochastic factor. In this new modeling, the conditional multinomial probability function plays an important role. The thesis has used two types of conditional probability models. One such model is constructed by using linear probability function conditioning on past multinomial responses. The second model is constructed by using a logistic (non-linear) probability function dynamic in multinomial responses over time. Note that these models were first developed under the assumption that the responses are influenced by fixed covariates only and they were referred to as the multinomial linear dynamic fixed probability and multinomial dynamic fixed logit (MDFL) models. Both history (complete history of the data being known) and contingency table based likelihood analysis were discussed in details. Furthermore, because the asymptotic properties of the likelihood estimators are well-known, the thesis has concentrated on the finite sample performances only.

We have also considered multinomial dynamic mixed (MDML) models under the assumption that certain extra random effects with different variances under categories may be needed in some cases to fit the data well. More specifically, these random effects are capable of accommodating latent or unobserved effects of the individuals which however remain the same over time. However, in the thesis, we have not included any contingency tables based analysis for the MDML model. This is because of the difficulty that in the mixed model random effects vary from individual to individual which does not allow any grouping of the individuals for the construction of the contingency table. It is, therefore, clear that any contingency table formation will require suitable assumption about the random effects, mainly to reduce the number of random effects and do the appropriate inferences. This is, however, beyond the scope of the present thesis.

With regard to inferences, it was demonstrated that the likelihood approach both in fixed and mixed models estimate the true parameters of the model very well. The likelihood estimation method was applied to reanalyze the Three Mile Island Stress-level data [earlier analyzed by Conaway (1989)] by emphasizing the correlations of the responses through dynamic dependence parameters. The regression effects are also much clearly interpreted as compared to the model parameters used by Conaway (1989).

Note that in the thesis we have dealt with univariate multinomial responses in longitudinal setup. However, there may be situations where several multinomial responses are collected from the same individual over a short period of time. Even though there exists some studies involving multivariate multinomial data in cross-sectional level (Neerchal and Morel, 2005), there does not appear any studies with multivariate multinomial data in longitudinal set up. One may pursue this in future research.

Further note that there are non-standard situations when longitudinal multinomial responses may be subject to outliers, missing values and measurement errors. One may exploit the longitudinal multinomial models given in the thesis to study these non-standard situations. We wish to explore them in the future.

Appendix A

Proof of Lemma 3.1: (page 73)

For simplicity, we have chosen $K = 2$ i.e., 3 categories and $T = 4$ time points in the study. So, at the initial time point $t = 1$, for W given individual i ($i = 1, \cdots, I$) the categorical response Y_{i1} marginally follows W multinomial distribution with density function

$$P(Y_{i11} = y_{i11}, Y_{i12} = y_{i12}) = \frac{\left\{\tilde{\pi}_{i1}^{(1)}\right\}^{y_{i11}} . \left\{\tilde{\pi}_{i1}^{(2)}\right\}^{y_{i12}} . (1 - \tilde{\pi}_{i1}^{(1)} - \tilde{\pi}_{i1}^{(2)})^{(1-y_{i11}-y_{i12})}}{y_{i11}!.\, y_{i12}!.\, (1 - y_{i11} - y_{i12})!}$$

yielding the marginal mean and marginal variance of Y_{i1}

$$E(Y_{i1}) = \tilde{\Pi}_{i1} = \left(\tilde{\pi}_{i1}^{(1)}, \tilde{\pi}_{i1}^{(2)}\right)' \tag{W.1}$$

$$Var(Y_{i1}) = diag\left[\tilde{\pi}_{i1}^{(1)}, \tilde{\pi}_{i1}^{(2)}\right] - \tilde{\Pi}_{i1}\,\tilde{\Pi}_{i1}'. \tag{W.2}$$

Now, in general, at time point $t = 2, 3, 4$, we can write the conditional distribution of the multinomial response vector Y_{it} given $y_{i,t-1}$ as

$$P(Y_{it1} = y_{it1}, Y_{it2} = y_{it2} | \, y_{i,t-1}^{(l)}) = \frac{\left\{\tilde{\eta}_{it|t-1}^{(1)}(l)\right\}^{y_{it1}} \cdot \left\{\tilde{\eta}_{it|t-1}^{(2)}(l)\right\}^{y_{it2}} \cdot \left(1 - \tilde{\eta}_{it|t-1}^{(1)}(l) - \tilde{\eta}_{it|t-1}^{(2)}(l)\right)^{(1-y_{it1}-y_{it2})}}{y_{it1}!.\, y_{it2}!.\, (1 - y_{it1} - y_{it2})!}$$

where the conditional probability $\tilde{\eta}_{it|t-1}^{(j)}(l)$ are defined in (3.3). It then follows that the conditional mean and conditional variance have the formulas

$$E(Y_{it} | \, y_{i,t-1}^{(l)}) \; = \; \boldsymbol{\eta}_{it|t-1} \; = \; \left(\tilde{\eta}_{it|t-1}^{(1)}(l), \, \tilde{\eta}_{it|t-1}^{(2)}(l)\right)' \qquad \text{(W.3)}$$

$$Var(Y_{it} | \, y_{i,t-1}^{(l)}) \; = \; diag\left[\tilde{\eta}_{it|t-1}^{(1)}(l), \, \tilde{\eta}_{it|t-1}^{(2)}(l)\right] \; - \; \boldsymbol{\eta}_{it|t-1} \, \boldsymbol{\eta}'_{it|t-1} \qquad \text{(W.4)}$$

Note that in what follows we use $\pi_{itj(l)}$ instead of $\tilde{\eta}_{it|t-1}^{(j)}(l)$ and π_{itj} instead of $\tilde{\pi}_{it}^{(j)}$.

Derivation of Mean:

For $t = 2$, it follows that

$$E(Y_{i2}) \; = \; E_{Y_{i1}} E(Y_{i2} \, | \, y_{i1}^{(l)}) \; = \; E_{Y_{i1}}[\Pi_{i2(l)}] \; = \; E_{Y_{i1}} \begin{bmatrix} \pi_{i21(l)} \\ \\ \pi_{i22(l)} \end{bmatrix} \; = \; \Pi_{i2}$$

$$E_{Y_{i1}}\left[\pi_{i21(l)}\right] \; = \; \sum_{Y_{i1}} \pi_{i21(l)} \, f(y_{i1})$$

$$= \pi_{i21(1)}\,\pi_{i11} + \pi_{i21(2)}\,\pi_{i12} + \pi_{i21(3)}\,(1 - \pi_{i11} - \pi_{i12})$$

$$= \pi_{i21(3)} + \left(\pi_{i21(1)} - \pi_{i21(3)}\right)\pi_{i11} + \left(\pi_{i21(2)} - \pi_{i21(3)}\right)\pi_{i12}$$

$$E_{Y_{i1}}\left[\pi_{i22(l)}\right] = \pi_{i22(3)} + \left(\pi_{i22(1)} - \pi_{i22(3)}\right)\pi_{i11} + \left(\pi_{i22(2)} - \pi_{i22(3)}\right)\pi_{i12}$$

We can write in matrix notation

$$\begin{aligned} E(Y_{i2}) &= \begin{pmatrix} \pi_{i21(3)} \\ \pi_{i22(3)} \end{pmatrix} + \begin{bmatrix} \pi_{i21(1)} - \pi_{i21(3)} & \pi_{i21(2)} - \pi_{i21(3)} \\ \pi_{i22(1)} - \pi_{i22(3)} & \pi_{i22(2)} - \pi_{i22(3)} \end{bmatrix} \begin{pmatrix} \pi_{i11} \\ \pi_{i12} \end{pmatrix} \\ &= \begin{pmatrix} \pi_{i21(3)} \\ \pi_{i22(3)} \end{pmatrix} + \left\{ \begin{bmatrix} \pi_{i21(1)} & \pi_{i21(2)} \\ \pi_{i22(1)} & \pi_{i22(2)} \end{bmatrix} - \begin{pmatrix} \pi_{i21(3)} \\ \pi_{i22(3)} \end{pmatrix} \begin{pmatrix} 1\,,1 \end{pmatrix} \right\} \begin{pmatrix} \pi_{i11} \\ \pi_{i12} \end{pmatrix} \end{aligned}$$

$$\Pi_{i2} = \Pi_{i2(3)} + \{W_{i2} - \Pi_{i2(3)}\,1'\}\,\Pi_{i1}$$

Similarly, for $t = 3$, we can write

$$E(Y_{i3}) = E_{Y_{i1}}\,E_{Y_{i2}}\,E[Y_{i3}\,|\,y_{i2}^{(m)},\,y_{i1}^{(l)}] = E_{Y_{i1}}\,E_{Y_{i2}}\,[\Pi_{i3(m)}] = E_{Y_{i1}}\,E_{Y_{i2}} \begin{bmatrix} \pi_{i31(m)} \\ \pi_{i32(m)} \end{bmatrix}$$

$$E_{Y_{i1}}\,E_{Y_{i2}}\,[\pi_{i31(m)}] = E_{Y_{i1}}\{\pi_{i31(1)}\pi_{i21(l)} + \pi_{i31(2)}\pi_{i22(l)} + \pi_{i31(3)}(1 - \pi_{i21(l)} - \pi_{i22(l)})\}$$

$$
\begin{aligned}
&= E_{Y_{i1}}\{\pi_{i31(3)} + (\pi_{i31(1)} - \pi_{i31(3)})\pi_{i21(l)} + (\pi_{i31(2)} - \pi_{i31(3)})\pi_{i22(l)}\} \\
&= \pi_{i31(3)} + \big(\pi_{i31(1)} - \pi_{i31(3)}\big)\big\{\pi_{i21(1)}\,\pi_{i11} + \pi_{i21(2)}\,\pi_{i12} \\
&\quad + \pi_{i21(3)}(1 - \pi_{i11} - \pi_{i12})\big\} + \big(\pi_{i31(2)} - \pi_{i31(3)}\big)\big\{\pi_{i22(1)}\pi_{i11} \\
&\quad + \pi_{i22(2)}\,\pi_{i12} + \pi_{i22(3)}(1 - \pi_{i11} - \pi_{i12})\big\} \\
&= \pi_{i31(3)} + \big(\pi_{i31(1)} - \pi_{i31(3)}\big)\big\{\pi_{i21(3)} + \big(\pi_{i21(1)} - \pi_{i21(3)}\big)\,\pi_{i11} \\
&\quad + (\pi_{i21(2)} - \pi_{i21(3)})\pi_{i12}\big\} + \big(\pi_{i31(2)} - \pi_{i31(3)}\big)\big\{\pi_{i22(3)} \\
&\quad + \big(\pi_{i22(1)} - \pi_{i22(3)}\big)\,\pi_{i11} + \big(\pi_{i22(2)} - \pi_{i22(3)}\big)\,\pi_{i12}\big\}
\end{aligned}
$$

$$
\begin{aligned}
E_{Y_{i1}}\,E_{Y_{i2}}\left[\pi_{i32(m)}\right] &= \pi_{i32(3)} + \big(\pi_{i32(1)} - \pi_{i32(3)}\big)\big\{\pi_{i21(3)} + \big(\pi_{i21(1)} - \pi_{i21(3)}\big)\,\pi_{i11} \\
&\quad + (\pi_{i21(2)} - \pi_{i21(3)})\pi_{i12}\big\} + \big(\pi_{i32(2)} - \pi_{i32(3)}\big)\big\{\pi_{i22(3)} \\
&\quad + \big(\pi_{i22(1)} - \pi_{i22(3)}\big)\,\pi_{i11} + \big(\pi_{i22(2)} - \pi_{i22(3)}\big)\,\pi_{i12}\big\}
\end{aligned}
$$

We can write in matrix notation

$$
\begin{aligned}
E(Y_{i3}) &= \begin{pmatrix} \pi_{i31(3)} \\ \pi_{i32(3)} \end{pmatrix} + \begin{bmatrix} \pi_{i31(1)} - \pi_{i31(3)} & \pi_{i31(2)} - \pi_{i31(3)} \\ \pi_{i32(1)} - \pi_{i32(3)} & \pi_{i32(2)} - \pi_{i32(3)} \end{bmatrix} \\
&\quad \times \left\{ \begin{pmatrix} \pi_{i21(3)} \\ \pi_{i22(3)} \end{pmatrix} + \begin{bmatrix} \pi_{i21(1)} - \pi_{i21(3)} & \pi_{i21(2)} - \pi_{i21(3)} \\ \pi_{i22(1)} - \pi_{i22(3)} & \pi_{i22(2)} - \pi_{i22(3)} \end{bmatrix} \begin{pmatrix} \pi_{i11} \\ \pi_{i12} \end{pmatrix} \right\}
\end{aligned}
$$

$$= \begin{pmatrix} \pi_{i31(3)} \\ \pi_{i32(3)} \end{pmatrix} + \begin{bmatrix} \pi_{i31(1)} - \pi_{i31(3)} & \pi_{i31(2)} - \pi_{i31(3)} \\ \pi_{i32(1)} - \pi_{i32(3)} & \pi_{i32(2)} - \pi_{i32(3)} \end{bmatrix} E(Y_{i2})$$

$$= \begin{pmatrix} \pi_{i31(3)} \\ \pi_{i32(3)} \end{pmatrix} + \left\{ \begin{bmatrix} \pi_{i31(1)} & \pi_{i31(2)} \\ \pi_{i32(1)} & \pi_{i32(2)} \end{bmatrix} - \begin{pmatrix} \pi_{i31(3)} \\ \pi_{i32(3)} \end{pmatrix} \begin{pmatrix} 1\,,1 \end{pmatrix} \right\} \begin{pmatrix} \pi_{i21} \\ \pi_{i22} \end{pmatrix}$$

$$\Pi_{i3} = \Pi_{i3(3)} + \{W_{i3} - \Pi_{i3(3)}\, 1'\}\, \Pi_{i2}$$

And, for $t = 4$, we can write

$$\begin{aligned} E(Y_{i4}) &= E_{Y_{i1}}\, E_{Y_{i2}}\, E_{Y_{i3}}\, E[Y_{i3} \mid y_{i3}^{(n)},\, y_{i2}^{(m)},\, y_{i1}^{(l)}] = E_{Y_{i1}}\, E_{Y_{i2}}\, E_{Y_{i3}}\, [\Pi_{i4(n)}] \\ &= E_{Y_{i1}}\, E_{Y_{i2}}\, E_{Y_{i3}} \begin{bmatrix} \pi_{i41(n)} \\ \pi_{i42(n)} \end{bmatrix} = \begin{bmatrix} \pi_{i41} \\ \pi_{i42} \end{bmatrix} = \Pi_{i4} \end{aligned}$$

$$\begin{aligned} E_{Y_{i1}} E_{Y_{i2}} E_{Y_{i3}} [\pi_{i41(n)}] &= E_{Y_{i1}}\, E_{Y_{i2}}\, \{\pi_{i41(1)}\, \pi_{i31(m)} + \pi_{i41(2)}\, \pi_{i32(m)} + \pi_{i41(3)} \\ &\qquad (1 - \pi_{i31(m)} - \pi_{i32(m)})\} \\ &= E_{Y_{i1}} E_{Y_{i2}} \{\pi_{i41(3)} + (\pi_{i41(1)} - \pi_{i41(3)})\pi_{i31(m)} \\ &\qquad + (\pi_{i41(2)} - \pi_{i41(3)})\pi_{i32(m)}\} \\ &= \pi_{i41(3)} + (\pi_{i41(1)} - \pi_{i41(3)})\, [\, \pi_{i31(3)} + (\pi_{i31(1)} - \pi_{i31(3)}) \\ &\qquad \{\pi_{i21(3)} + (\pi_{i21(1)} - \pi_{i21(3)})\, \pi_{i11} + (\pi_{i21(2)} - \pi_{i21(3)})\pi_{i12}\} \end{aligned}$$

$$+\left(\pi_{i31(2)}-\pi_{i31(3)}\right)\left\{\pi_{i22(3)}+\left(\pi_{i22(1)}-\pi_{i22(3)}\right)\pi_{i11}\right.$$

$$\left.+\left(\pi_{i22(2)}-\pi_{i22(3)}\right)\pi_{i12}\right\}\Big]$$

$$+\left(\pi_{i41(2)}-\pi_{i41(3)}\right)\left[\pi_{i32(3)}+\left(\pi_{i32(1)}-\pi_{i32(3)}\right)\right.$$

$$\times\left\{\pi_{i21(3)}+(\pi_{i21(1)}-\pi_{i21(3)})\pi_{i11}+(\pi_{i21(2)}-\pi_{i21(3)})\pi_{i12}\right\}$$

$$+\left(\pi_{i32(2)}-\pi_{i32(3)}\right)\left\{\pi_{i22(3)}+\left(\pi_{i22(1)}-\pi_{i22(3)}\right)\pi_{i11}\right.$$

$$\left.\left.+\left(\pi_{i22(2)}-\pi_{i22(3)}\right)\pi_{i12}\right\}\right]$$

We can write in matrix notation

$$\begin{aligned}
E(Y_{i4}) &= \begin{pmatrix}\pi_{i41(3)}\\ \pi_{i42(3)}\end{pmatrix}+\begin{bmatrix}\pi_{i41(1)}-\pi_{i41(3)} & \pi_{i41(2)}-\pi_{i41(3)}\\ \pi_{i42(1)}-\pi_{i42(3)} & \pi_{i42(2)}-\pi_{i42(3)}\end{bmatrix}\\
&\quad\times\left[\begin{pmatrix}\pi_{i31(3)}\\ \pi_{i32(3)}\end{pmatrix}+\begin{pmatrix}\pi_{i31(1)}-\pi_{i31(3)} & \pi_{i31(2)}-\pi_{i31(3)}\\ \pi_{i32(1)}-\pi_{i32(3)} & \pi_{i32(2)}-\pi_{i32(3)}\end{pmatrix}\right.\\
&\quad\left.\times\left\{\begin{pmatrix}\pi_{i21(3)}\\ \pi_{i22(3)}\end{pmatrix}+\begin{pmatrix}\pi_{i21(1)}-\pi_{i21(3)} & \pi_{i21(2)}-\pi_{i21(3)}\\ \pi_{i22(1)}-\pi_{i22(3)} & \pi_{i22(2)}-\pi_{i22(3)}\end{pmatrix}\begin{pmatrix}\pi_{i11}\\ \pi_{i12}\end{pmatrix}\right\}\right]\\
&= \begin{pmatrix}\pi_{i41(3)}\\ \pi_{i42(3)}\end{pmatrix}+\begin{bmatrix}\pi_{i41(1)}-\pi_{i41(3)} & \pi_{i41(2)}-\pi_{i41(3)}\\ \pi_{i42(1)}-\pi_{i42(3)} & \pi_{i42(2)}-\pi_{i42(3)}\end{bmatrix}E(Y_{i3})\\
&= \begin{pmatrix}\pi_{i41(3)}\\ \pi_{i42(3)}\end{pmatrix}+\left\{\begin{bmatrix}\pi_{i41(1)} & \pi_{i41(2)}\\ \pi_{i42(1)} & \pi_{i42(2)}\end{bmatrix}-\begin{pmatrix}\pi_{i41(3)}\\ \pi_{i42(3)}\end{pmatrix}\begin{pmatrix}1\,,1\end{pmatrix}\right\}\begin{pmatrix}\pi_{i31}\\ \pi_{i32}\end{pmatrix}
\end{aligned}$$

$$\Pi_{i4} = \Pi_{i4(3)} + \{W_{i4} - \Pi_{i4(3)}\, 1'\}\, \Pi_{i3}$$

Thus, for any $t = 1, 2, 3, 4$; we can write

$$E(Y_{it}) = \Pi_{it} = \begin{bmatrix} \pi_{it1} \\ \pi_{it2} \end{bmatrix} = \Pi_{it(3)} + [W_{it} - \Pi_{it(3)}\, 1']\, \Pi_{i,t-1}$$

Derivation of Variance:

Next to derive the unconditional covariance matrix of Y_{it}, we will use the same conditioning and unconditioning properties of expectations as we have used for the derivation of the unconditional mean vector. To be specific, we first write

$$\begin{aligned} Var(Y_{i1}) &= E[Y_{i1}\, Y_{i1}'] - E(Y_{i1})\{E(Y_{i1})\}' \\ &= E\begin{bmatrix} y_{i11}^2 & y_{i11}y_{i12} \\ y_{i11}y_{i12} & y_{i12}^2 \end{bmatrix} - \begin{bmatrix} \pi_{i11} \\ \pi_{i12} \end{bmatrix} \begin{bmatrix} \pi_{i11} & \pi_{i12} \end{bmatrix} \\ &= \begin{bmatrix} \pi_{i11} & 0 \\ 0 & \pi_{i12} \end{bmatrix} - \Pi_{i1}\, \Pi_{i1}' \\ &= diag\,[\pi_{i11}, \pi_{i12}] - \Pi_{i1}\, \Pi_{i1}' \end{aligned}$$

$$\begin{aligned} Var(Y_{i2}) &= E[Y_{i2}\, Y_{i2}'] - E(Y_{i2})\{E(Y_{i2})\}' \\ &= E_{Y_{i1}}\, E[Y_{i2}\, Y_{i2}' \,|\, y_{i1}] - E(Y_{i2})\{E(Y_{i2})\}' \end{aligned}$$

$$= E_{Y_{i1}} \{ diag\, [\pi_{i21(l)},\, \pi_{i22(l)}] \} \; - \; \Pi_{i2}\, \Pi'_{i2}$$

$$= \begin{bmatrix} \pi_{i21} & 0 \\ 0 & \pi_{i22} \end{bmatrix} \; - \; \Pi_{i2}\, \Pi'_{i2}$$

$$= diag\, [\pi_{i21},\, \pi_{i22}] \; - \; \Pi_{i2}\, \Pi'_{i2}$$

Thus, for any $t \; = \; 1,\, 2,\, 3,\, 4$; we can write

$$Var(Y_{it}) \quad = \quad diag\, [\pi_{it1},\, \pi_{it2}] \; - \; \Pi_{it}\, \Pi'_{it}$$

Proof of Lemma 3.2: (page 74)

For simplicity, we have chosen $K = 2$ i.e., 3 categories and $T = 4$ time points in the study. Next to derive the unconditional covariance matrix between Y_{it} and $Y_{it'}$ we will use the same conditioning and unconditioning properties of expectations as we have used for the derivation of the unconditional mean vector. To be specific, we first write

$$\begin{aligned} E[Y_{i2} Y'_{i1}] \;\; &= \;\; E_{Y_{i1}}\, E\, [Y_{i2}\, Y'_{i1}\, |\, y_{i1}] \\ &= \;\; E_{Y_{i1}}\, \{ \Pi_{i2(l)}\, Y'_{i1} \} \end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} \pi_{i21(1)} & 0 \\ \pi_{i22(1)} & 0 \end{bmatrix} \pi_{i11} + \begin{bmatrix} 0 & \pi_{i21(2)} \\ 0 & \pi_{i22(2)} \end{bmatrix} \pi_{i12} \\
&= \begin{bmatrix} \pi_{i21(1)} & \pi_{i21(2)} \\ \pi_{i22(1)} & \pi_{i22(2)} \end{bmatrix} \begin{bmatrix} \pi_{i11} & 0 \\ 0 & \pi_{i12} \end{bmatrix} \\
&= W_{i2}\,.diag\,[\pi_{i11},\, \pi_{i12}]
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_{i2}, Y_{i1}) &= E\,[Y_{i2}\,Y_{i1}'] \;-\; E(Y_{i2})\,\{E(Y_{i1})\}' \\
&= W_{i2}\,.diag\,[\pi_{i11},\, \pi_{i12}] \;-\; \{\Pi_{i2(3)} \;+\; [W_{i2} \;-\; \Pi_{i2(3)}\,1']\,\Pi_{i1}\}\,\Pi_{i1}' \\
&= W_{i2}\,.diag\,[\pi_{i11},\, \pi_{i12}] \;-\; \Pi_{i2(3)}\,\Pi_{i1}' \;-\; [W_{i2} \;-\; \Pi_{i2(3)}\,1']\,\Pi_{i1}\,\Pi_{i1}' \\
&= [W_{i2} \;-\; \Pi_{i2(3)}\,1']\,.diag\,[\pi_{i11},\, \pi_{i12}] \;-\; [W_{i2} \;-\; \Pi_{i2(3)}\,1']\,\Pi_{i1}\,\Pi_{i1}' \\
&= [W_{i2} \;-\; \Pi_{i2(3)}\,1']\,\{diag\,[\pi_{i11},\, \pi_{i12}] \;-\; \Pi_{i1}\,\Pi_{i1}'\} \\
&= [W_{i2} \;-\; \Pi_{i2(3)}\,1']\,Var(Y_{i1})
\end{aligned}
$$

$$
\begin{aligned}
E[Y_{i3}Y_{i2}'] &= E_{Y_{i1}}\,E_{Y_{i2}}\,E\,[Y_{i3}\,Y_{i2}'\,|\,y_{i2}\,y_{i1}] \\
&= E_{Y_{i1}}\,E_{Y_{i2}}\,[\Pi_{i3(m)}\,Y_{i2}'\,|\,y_{i1}] \\
&= E_{Y_{i1}} \left\{ \begin{bmatrix} \pi_{i31(1)} & 0 \\ \pi_{i32(1)} & 0 \end{bmatrix} \pi_{i21(l)} + \begin{bmatrix} 0 & \pi_{i31(2)} \\ 0 & \pi_{i32(2)} \end{bmatrix} \pi_{i22(l)} \right\}
\end{aligned}
$$

$$
\begin{aligned}
&= \begin{bmatrix} \pi_{i31(1)} & 0 \\ \pi_{i32(1)} & 0 \end{bmatrix} \pi_{i21} + \begin{bmatrix} 0 & \pi_{i31(2)} \\ 0 & \pi_{i32(2)} \end{bmatrix} \pi_{i22} \\
&= \begin{bmatrix} \pi_{i31(1)} & \pi_{i31(2)} \\ \pi_{i32(1)} & \pi_{i32(2)} \end{bmatrix} \begin{bmatrix} \pi_{i21} & 0 \\ 0 & \pi_{i22} \end{bmatrix} \\
&= W_{i3}\,.diag\,[\pi_{i21},\, \pi_{i22}]
\end{aligned}
$$

$$
\begin{aligned}
Cov(Y_{i3}, Y_{i2}) &= E\,[Y_{i3}\,Y_{i2}'] - E(Y_{i3})\,\{E(Y_{i2})\}' \\
&= W_{i3}\,.diag\,[\pi_{i21},\, \pi_{i22}] - \{\Pi_{i3(3)} + [W_{i3} - \Pi_{i3(3)}\,1']\,\Pi_{i2}\}\,\Pi_{i2}' \\
&= W_{i3}\,.diag\,[\pi_{i21},\, \pi_{i22}] - \Pi_{i3(3)}\,\Pi_{i2}' - [W_{i3} - \Pi_{i3(3)}\,1']\,\Pi_{i2}\,\Pi_{i2}' \\
&= [W_{i3} - \Pi_{i3(3)}\,1']\,.diag\,[\pi_{i21},\, \pi_{i22}] - [W_{i3} - \Pi_{i3(3)}\,1']\,\Pi_{i2}\,\Pi_{i2}' \\
&= [W_{i3} - \Pi_{i3(3)}\,1']\,\{diag\,[\pi_{i21},\, \pi_{i22}] - \Pi_{i2}\,\Pi_{i2}'\} \\
&= [W_{i3} - \Pi_{i3(3)}\,1']\,Var(Y_{i2})
\end{aligned}
$$

Thus, we can write

$$
Cov(Y_{it}, Y_{i,t-1}) = [W_{it} - \Pi_{it(3)}\,1']\,Var(Y_{i,t-1}); \qquad for\ t = 2,\, 3,\, 4.
$$

$$
\begin{aligned}
E[Y_{i3}Y_{i1}'] &= E_{Y_{i1}}\,E_{Y_{i2}}\,E\,[Y_{i3}\,Y_{i1}' \mid y_{i2}\,y_{i1}] \\
&= E_{Y_{i1}}\,E_{Y_{i2}}\,[\Pi_{i3(m)}\,Y_{i1}' \mid y_{i1}]
\end{aligned}
$$

$$= E_{Y_{i1}} \left[\{ \Pi_{i3(3)} + (W_{i3} - \Pi_{i3(3)} 1') \Pi_{i2(l)} \} Y_{i1}' \right]$$

$$= \Pi_{i3(3)} \Pi_{i1}' + \{ W_{i3} - \Pi_{i3(3)} 1' \} W_{i2} \, diag \, [\pi_{i11}, \, \pi_{i12}]$$

$$\begin{aligned}
Cov(Y_{i3}, Y_{i1}) &= E \, [Y_{i3} \, Y_{i1}'] - E(Y_{i3}) \{ E(Y_{i1}) \}' \\
&= \Pi_{i3(3)} \Pi_{i1}' + \{ W_{i3} - \Pi_{i3(3)} 1' \} W_{i2} \, diag \, [\pi_{i11}, \, \pi_{i12}] \\
&\qquad - \{ \Pi_{i3(3)} + (W_{i3} - \Pi_{i3(3)} 1') \Pi_{i2} \} \Pi_{i1}' \\
&= \{ W_{i3} - \Pi_{i3(3)} 1' \} \left[W_{i2} \, diag \, [\pi_{i11}, \, \pi_{i12}] - \Pi_{i2} \Pi_{i1}' \right] \\
&= \{ W_{i3} - \Pi_{i3(3)} 1' \} \left[W_{i2} \, diag \, [\pi_{i11}, \, \pi_{i12}] - \Pi_{i2(3)} \Pi_{i1}' - \{ W_{i2} \right. \\
&\qquad \left. - \Pi_{i2(3)} 1' \} \Pi_{i1} \Pi_{i1}' \right] \\
&= \{ W_{i3} - \Pi_{i3(3)} 1' \} \{ W_{i2} - \Pi_{i2(3)} 1' \} \left[diag \, [\pi_{i11}, \, \pi_{i12}] - \Pi_{i1} \Pi_{i1}' \right] \\
&= \{ W_{i3} - \Pi_{i3(3)} 1' \} \{ W_{i2} - \Pi_{i2(3)} 1' \} \, Var(Y_{i1})
\end{aligned}$$

$$\begin{aligned}
E[Y_{i4} Y_{i1}'] &= E_{Y_{i1}} \, E_{Y_{i2}} \, E_{Y_{i3}} \, E \, [Y_{i4} \, Y_{i1}' \mid y_{i3} \, y_{i2} \, y_{i1}] \\
&= E_{Y_{i1}} \, E_{Y_{i2}} \, E_{Y_{i3}} [\, \Pi_{i4(n)} \, Y_{i1}' \mid y_{i2} \, y_{i1}] \\
&= E_{Y_{i1}} \, E_{Y_{i2}} \left[\{ \Pi_{i4(3)} + (W_{i4} - \Pi_{i4(3)} 1') \Pi_{i3(m)} \} Y_{i1}' \right] \\
&= E_{Y_{i1}} \left(\left[\Pi_{i4(3)} + \{ W_{i4} - \Pi_{i4(3)} 1' \} \{ \Pi_{i3(3)} \right. \right. \\
&\qquad \left. \left. + (W_{i3} - \Pi_{i3(3)} 1') \Pi_{i2(l)} \} \right] Y_{i1}' \right) \\
&= \Pi_{i4(3)} \Pi_{i1}' + \{ W_{i4} - \Pi_{i4(3)} 1' \} \left[\Pi_{i3(3)} \Pi_{i1}' \right.
\end{aligned}$$

$$+ \{W_{i3} - \Pi_{i3(3)} 1'\} W_{i2} \, diag \, (\pi_{i11}, \pi_{i12})]$$

$$\begin{aligned}
Cov(Y_{i4}, Y_{i1}) &= E[Y_{i4} Y'_{i1}] - E(Y_{i4}) \{E(Y_{i1})\}' \\
&= \Pi_{i4(3)} \Pi'_{i1} + \{W_{i4} - \Pi_{i4(3)} 1'\} [\Pi_{i3(3)} \Pi'_{i1} + \{W_{i3} - \Pi_{i3(3)} 1'\} \\
&\qquad \times W_{i2} \, diag(\pi_{i11}, \pi_{i12})] - [\Pi_{i4(3)} + \{W_{i4} - \Pi_{i4(3)} 1'\} \Pi_{i3}] \Pi'_{i1} \\
&= \{W_{i4} - \Pi_{i4(3)} 1'\} [\Pi_{i3(3)} \Pi'_{i1} + \{W_{i3} - \Pi_{i3(3)} 1'\} W_{i2} \\
&\qquad \times diag(\pi_{i11}, \pi_{i12}) - \{\Pi_{i3(3)} + (W_{i3} - \Pi_{i3(3)} 1') \Pi_{i2}\} \Pi'_{i1}] \\
&= \{W_{i4} - \Pi_{i4(3)} 1'\} \{W_{i3} - \Pi_{i3(3)} 1'\} [W_{i2} \, diag \, [\pi_{i11}, \pi_{i12}] \\
&\qquad - \{\Pi_{i2(3)} + (W_{i2} - \Pi_{i2(3)} 1') \Pi_{i1}\} \Pi'_{i1}] \\
&= \{W_{i4} - \Pi_{i4(3)} 1'\} \{W_{i3} - \Pi_{i3(3)} 1'\} [W_{i2} \, diag \, [\pi_{i11}, \pi_{i12}] \\
&\qquad - \Pi_{i2(3)} \Pi'_{i1} - \{W_{i2} - \Pi_{i2(3)} 1'\} \Pi_{i1} \Pi'_{i1}] \\
&= \{W_{i4} - \Pi_{i4(3)} 1'\} \{W_{i3} - \Pi_{i3(3)} 1'\} \{W_{i2} - \Pi_{i2(3)} 1'\} \\
&\qquad \times [diag \, [\pi_{i11}, \pi_{i12}] - \Pi_{i1} \Pi'_{i1}] \\
&= \{W_{i4} - \Pi_{i4(3)} 1'\} \{W_{i3} - \Pi_{i3(3)} 1'\} \{W_{i2} - \Pi_{i2(3)} 1'\} Var(Y_{i1})
\end{aligned}$$

$$\begin{aligned}
E[Y_{i4} Y'_{i2}] &= E_{Y_{i1}} E_{Y_{i2}} E_{Y_{i3}} E[Y_{i4} Y'_{i2} \mid y_{i3} \, y_{i2} \, y_{i1}] \\
&= E_{Y_{i1}} E_{Y_{i2}} E_{Y_{i3}} [\Pi_{i4(n)} Y'_{i2} \mid y_{i2} \, y_{i1}] \\
&= E_{Y_{i1}} E_{Y_{i2}} [\{\Pi_{i4(3)} + (W_{i4} - \Pi_{i4(3)} 1') \Pi_{i3(m)}\} Y'_{i2} \mid y_{i1}]
\end{aligned}$$

$$= \quad \Pi_{i4(3)}\,\Pi'_{i2} \,+\, \{W_{i4} \,-\, \Pi_{i4(3)}\,1'\}\,W_{i3}\,diag\,(\pi_{i21},\,\pi_{i22})$$

$$\begin{aligned}
Cov(Y_{i4}, Y_{i2}) &= E\,[Y_{i4}\,Y'_{i2}] \,-\, E(Y_{i4})\,\{E(Y_{i2})\}' \\
&= \Pi_{i4(3)}\Pi'_{i2} + \{W_{i4} - \Pi_{i4(3)}1'\}\,W_{i3}\,diag\,(\pi_{i21},\,\pi_{i22}) \\
&\qquad\qquad -[\,\Pi_{i4(3)} + \{W_{i4} - \Pi_{i4(3)}1'\}\Pi_{i3}\,]\Pi'_{i2} \\
&= \{W_{i4} - \Pi_{i4(3)}1'\}[\,W_{i3}\,diag(\pi_{i21},\,\pi_{i22}) \\
&\qquad\qquad - \{\Pi_{i3(3)} + (W_{i3} - \Pi_{i3(3)}1')\,\Pi_{i2}\}\,\Pi'_{i2}\,] \\
&= \{W_{i4} - \Pi_{i4(3)}1'\}\,[\,W_{i3}\,diag\,(\pi_{i21},\,\pi_{i22}) \\
&\qquad\qquad - \Pi_{i2(3)}\,\Pi'_{i1} - \{W_{i3} \,-\, \Pi_{i3(3)}\,1'\}\,\Pi_{i2}\,\Pi'_{i2}\,] \\
&= \{W_{i4} - \Pi_{i4(3)}1'\}\{W_{i2} - \Pi_{i2(3)}1'\}\,[\,diag(\pi_{i11},\,\pi_{i12}) - \Pi_{i2}\Pi'_{i2}] \\
&= \{W_{i4} - \Pi_{i4(3)}1'\}\,\{W_{i2} \,-\, \Pi_{i2(3)}\,1'\}\,Var(Y_{i2})
\end{aligned}$$

Thus, we can write

$$Cov(Y_{it}, Y_{iu}) \,=\, Var(Y_{iu}) \prod_{j=u+1}^{t} [W_{ij} \,-\, \Pi_{ij(3)}\,1']; \qquad for\; u \,<\, t.$$

where

$$W_{ij} \,=\, \begin{bmatrix} \pi_{ij1(1)} & \pi_{ij1(2)} \\ \pi_{ij2(1)} & \pi_{ij2(2)} \end{bmatrix} \qquad \Pi_{ij(3)} \,=\, \begin{bmatrix} \pi_{ij1(3)} \\ \pi_{ij2(3)} \end{bmatrix}$$

Proof of Lemma 3.5: (page 85)

We can find the expected value of the 2nd derivatives of the log-likelihood function by using following results;

When $k = 2$ (i.e. 3 categories);

$$\begin{aligned}
E\left[\pi_{itj(l)}\right] &= E_{Y_{i,t-1}} E\left[\pi_{itj(l)} \,|\, y_{i,t-1}\right] = E_{Y_{i,t-1}}\left[\pi_{itj(l)}\right] = \sum_{Y_{i,t-1}} \pi_{itj(l)}\, f(y_{i,t-1}) \\
&= \pi_{itj(1)}\,\pi_{i,t-1,1} + \pi_{itj(2)}\,\pi_{i,t-1,2} + \pi_{itj(3)}\left(1 - \pi_{i,t-1,1} - \pi_{i,t-1,2}\right) \\
&= \pi_{itj(3)} + \left(\pi_{itj(1)} - \pi_{itj(3)}\right)\pi_{i,t-1,1} + \left(\pi_{itj(2)} - \pi_{itj(3)}\right)\pi_{i,t-1,2} \\
&= \pi_{itj}
\end{aligned}$$

$$\begin{aligned}
E\left[\pi^2_{itj(l)}\right] &= E_{Y_{i,t-1}} E\left[\pi^2_{itj(l)} \,|\, y_{i,t-1}\right] = E_{Y_{i,t-1}}\left[\pi^2_{itj(l)}\right] \\
&= \sum_{Y_{i,t-1}} \pi^2_{itj(l)}\, f(y_{i,t-1}) = \sum_{l=1}^{k+1} \pi^2_{itj(l)}\,\pi_{i,t-1,l} \\
&= \pi^2_{itj(1)}\,\pi_{i,t-1,1} + \pi^2_{itj(2)}\,\pi_{i,t-1,2} + \pi^2_{itj(3)}\left(1 - \pi_{i,t-1,1} - \pi_{i,t-1,2}\right)
\end{aligned}$$

$$\begin{aligned}
E\left[\pi_{itj(l)}\pi_{itj'(l)}\right] &= E_{Y_{i,t-1}} E\left[\pi_{itj(l)}\pi_{itj'(l)} \,|\, y_{i,t-1}\right] = E_{Y_{i,t-1}}\left[\pi_{itj(l)}\pi_{itj'(l)}\right] \\
&= \sum_{Y_{i,t-1}} \pi_{itj(l)}\pi_{itj'(l)}\, f(y_{i,t-1}) \\
&= \sum_{l=1}^{k+1} \pi_{itj(l)}\pi_{itj'(l)}\,\pi_{i,t-1,l}
\end{aligned}$$

$$
\begin{aligned}
E\left[Y_{i,t-1}\,\pi_{itj(l)}\right] &= E_{Y_{i,t-1}}\,E\left[Y_{i,t-1}\,\pi_{itj(l)}\,|\,y_{i,t-1}\right] \\
&= E_{Y_{i,t-1}}\left[Y_{i,t-1}\,\pi_{itj(l)}\right] \\
&= \sum_{Y_{i,t-1}} y_{i,t-1}\,\pi_{itj(l)}\,f(y_{i,t-1}) \\
&= e_1\pi_{itj(1)}\,\pi_{i,t-1,1} + e_2\pi_{itj(2)}\,\pi_{i,t-1,2} + (0,0)' \\
&= \left(\,\pi_{itj(1)}\pi_{i,t-1,1},\;\pi_{itj(2)}\pi_{i,t-1,2}\,\right)'
\end{aligned}
$$

$$
\begin{aligned}
E\left[Y_{i,t-1}Y'_{i,t-1}\pi_{itj(l)}\right] &= E_{Y_{i,t-1}}\,E\left[Y_{i,t-1}Y'_{i,t-1}\,\pi_{itj(l)}\,|\,y_{i,t-1}\right] \\
&= E_{Y_{i,t-1}}\left[Y_{i,t-1}Y'_{i,t-1}\,\pi_{itj(l)}\right] \\
&= \sum_{Y_{i,t-1}} y_{i,t-1}y'_{i,t-1}\,\pi_{itj(l)}\,f(y_{i,t-1}) \\
&= (e_1e'_1)\,\pi_{itj(1)}\,\pi_{i,t-1,1} + (e_2e'_2)\,\pi_{itj(2)}\,\pi_{i,t-1,2} \\
&= diag\left(\,\pi_{itj(1)}\pi_{i,t-1,1},\;\pi_{itj(2)}\pi_{i,t-1,2}\,\right)
\end{aligned}
$$

$$
\begin{aligned}
E\left[Y_{i,t-1}\,\pi^2_{itj(l)}\right] &= E_{Y_{i,t-1}}\,E\left[Y_{i,t-1}\,\pi^2_{itj(l)}\,|\,y_{i,t-1}\right] \\
&= E_{Y_{i,t-1}}\left[Y_{i,t-1}\,\pi^2_{itj(l)}\right] \\
&= \sum_{Y_{i,t-1}} y_{i,t-1}\,\pi^2_{itj(l)}\,f(y_{i,t-1}) \\
&= e_1\pi^2_{itj(1)}\,\pi_{i,t-1,1} + e_2\pi^2_{itj(2)}\,\pi_{i,t-1,2} + (0,0)' \\
&= \left(\,\pi^2_{itj(1)}\pi_{i,t-1,1},\;\pi^2_{itj(2)}\pi_{i,t-1,2}\,\right)'
\end{aligned}
$$

$$
\begin{aligned}
E\left[Y_{i,t-1}Y'_{i,t-1}\pi^2_{itj(l)}\right] &= E_{Y_{i,t-1}}\, E\left[Y_{i,t-1}Y'_{i,t-1}\,\pi^2_{itj(l)}\,|\,y_{i,t-1}\right] \\
&= E_{Y_{i,t-1}}\left[Y_{i,t-1}Y'_{i,t-1}\,\pi^2_{itj(l)}\right] \\
&= \sum_{Y_{i,t-1}} y_{i,t-1}y'_{i,t-1}\,\pi^2_{itj(l)}\, f(y_{i,t-1}) \\
&= (e_1e'_1)\,\pi^2_{itj(1)}\,\pi_{i,t-1,1} + (e_2e'_2)\,\pi^2_{itj(2)}\,\pi_{i,t-1,2} \\
&= diag\left(\pi^2_{itj(1)}\pi_{i,t-1,1},\ \pi^2_{itj(2)}\pi_{i,t-1,2}\right)
\end{aligned}
$$

$$
\begin{aligned}
E\left[Y_{i,t-1}\pi_{itj(l)}\pi_{itj'(l)}\right] &= E_{Y_{i,t-1}}\, E\left[Y_{i,t-1}\,\pi_{itj(l)}\pi_{itj'(l)}\,|\,y_{i,t-1}\right] \\
&= E_{Y_{i,t-1}}\left[Y_{i,t-1}\,\pi_{itj(l)}\pi_{itj'(l)}\right] \\
&= \sum_{Y_{i,t-1}} y_{i,t-1}\,\pi_{itj(l)}\pi_{itj'(l)}\, f(y_{i,t-1}) \\
&= e_1\,\pi_{itj(1)}\,\pi_{itj'(1)}\,\pi_{i,t-1,1} + e_2\,\pi_{itj(2)}\,\pi_{itj'(2)}\,\pi_{i,t-1,2} \\
&= \left(\pi_{itj(1)}\pi_{itj'(1)}\pi_{i,t-1,1},\ \pi_{itj(2)}\pi_{itj'(2)}\pi_{i,t-1,2}\right)
\end{aligned}
$$

$$
\begin{aligned}
E\left[Y_{i,t-1}Y'_{i,t-1}\pi_{itj(l)}\pi_{itj'(l)}\right] &= E_{Y_{i,t-1}}\, E\left[Y_{i,t-1}Y'_{i,t-1}\,\pi_{itj(l)}\pi_{itj'(l)}\,|\,y_{i,t-1}\right] \\
&= E_{Y_{i,t-1}}\left[Y_{i,t-1}Y'_{i,t-1}\,\pi_{itj(l)}\pi_{itj'(l)}\right] \\
&= \sum_{Y_{i,t-1}} y_{i,t-1}y'_{i,t-1}\,\pi_{itj(l)}\pi_{itj'(l)}\, f(y_{i,t-1}) \\
&= diag\left(\pi_{itj(1)}\pi_{itj'(1)}\pi_{i,t-1,1},\ \pi_{itj(2)}\pi_{itj'(2)}\pi_{i,t-1,2}\right)
\end{aligned}
$$

where $e_1 = (1, 0)'$, $e_2 = (0, 1)'$ and $e_3 = (0, 0)'$

Now, the expected values of the 2nd derivatives are;

$$E\left[\frac{\partial^2 \ln L(.)}{\partial\beta\,\partial\beta'}\right] = E\left[\sum_{i,t}\{ diag(\pi_{it1(l_{t-1})}, ..., \pi_{itk(l_{t-1})}) - \Pi_{it(l_{t-1})}\Pi'_{it(l_{t-1})}\}_{k\times k} \otimes (x_{it}\, x'_{it})_{p\times p}\right]$$

$$= \sum_{i,t}\left[diag(\pi_{it1}, ..., \pi_{itk}) - \sum_{l=1}^{k+1} \pi_{i,t-1,l}\,(\Pi_{it(l)}\Pi'_{it(l)})\right] \otimes (x_{it}\, x'_{it})_{p\times p} :\longrightarrow kp \times kp$$

$$E\left[\frac{\partial^2 \ln L(.)}{\partial\theta\,\partial\theta'}\right] = E\left[\sum_{i,t}\{diag(\pi_{it1(l_{t-1})}, ..., \pi_{itk(l_{t-1})}) - \Pi_{it(l_{t-1})}\Pi'_{it(l_{t-1})}\}_{k\times k} \otimes (y_{i,t-1}\, y'_{i,t-1})_{k\times k}\right]$$

$$= \sum_{i,t}[\, D_{it} - W_{it}\,]_{k^2\times k^2} :\longrightarrow k^2 \times k^2$$

where

$$D_{it} = \bigoplus_{j=1}^{k} D_{itj} = \begin{bmatrix} D_{it1} & 0 & \cdots & 0 \\ 0 & D_{it2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{itk} \end{bmatrix}_{k^2\times k^2}$$

with $D_{itj} = diag[\,\pi_{itj(1)}\pi_{i,t-1,1}, ..., \pi_{itj(k)}\pi_{i,t-1,k}\,]_{k\times k}$

$$W_{it} = \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1k} \\ V_{21} & V_{22} & \cdots & V_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ V_{k1} & V_{k2} & \cdots & V_{kk} \end{bmatrix}_{k^2 \times k^2}$$

with $V_{jj'} = diag[\, \pi_{itj(1)}\pi_{itj'(1)}\pi_{i,t-1,1}, \ldots, \pi_{itj(k)}\pi_{itj'(k)}\pi_{i,t-1,k} \,]_{k\times k}$

$$E\left[\frac{\partial^2 \ln L(.)}{\partial\beta\,\partial\theta'}\right] = E\left[\sum_{i,t}\{\, diag(\pi_{it1(l_{t-1})}, \ldots, \pi_{itk(l_{t-1})}) - \Pi_{it(l_{t-1})}\Pi'_{it(l_{t-1})}\}_{k\times k} \otimes (x_{it}\, y'_{i,t-1})_{p\times k}\right]$$
$$= \sum_{i,t}[\, D^*_{it}(x) \;-\; W^*_{it}(x)\,]_{kp\times k^2} \;:\rightarrow kp \times k^2$$

where

$$D^*_{it}(x) = \bigoplus_{j=1}^{k} D^*_{itj} = \begin{bmatrix} D^*_{it1} & 0 & \cdots & 0 \\ 0 & D^*_{it2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D^*_{itk} \end{bmatrix}_{kp\times k^2}$$

with $D^*_{itj} = x_{it}\,[\,\Pi_{itj()} \oplus \Pi_{i,t-1}\,]'_{p\times k}$ and $W^*_{it}(x) = (V^*_{jj'})_{kp\times k^2}$ with

$V^*_{jj'} = x_{it}\,[\,\Pi_{itj()} \oplus \Pi_{itj'()} \oplus \Pi_{i,t-1}\,]'_{1\times k};\; p\times k$ and $\Pi_{itj()} = (\pi_{itj(1)}, \pi_{itj(2)}, \ldots, \pi_{itj(k)})'$

Bibliography

[1] Agresti, A. (1989). A survey of models for repeated ordered categorical response data. *Statistics in Medicine*, **8**, 1209-1224.

[2] Agresti, A. (1990). *Categorical Data Analysis*. Wiley, New York.

[3] Agresti, A. (1993). Distribution-free fitting of logit models with random effects for repeated categorical responses. *Statistics in Medicine*, **12**, 1969-1987.

[4] Agresti, A. (1997). A model for repeated measurements of a multivariate binary response. *J. Amer. Statist. Assoc.*, **92**, 315-321 .

[5] Agresti, A. (1999). Modeling ordered categorical data: recent advances and future challenges. *Statistics in Medicine*, **18**, 2191-2207.

[6] Agresti, A. (2002). *Categorical Data Analysis*. Second edition, Wiley, New York.

[7] Amemiya. T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge.

[8] Chen, Z. and Kuo, L. (2001). A note on the estimation of the multinomial logit model with random effects. *J. Amer. Statist. Assoc.*, **55, No. 2**, 89–95.

[9] Conaway, M. R. (1989). Analysis of Repeated Categorical Measurements with Conditional Likelihood Methods. *J. Amer. Statist. Assoc.*, **84, No. 405**, 53–62.

[10] Crowder, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika*, **82**, 407-410.

[11] Cruz-Medina, I. R., Hettmansperger, T. P. and Thomas, H. (2004). Semiparametric mixture models and repeated measurements: the multinomial cut point model. *Applied Statistics*, **53**, 463-474.

[12] De Rooij, M. (2011). Transitional ideal point models for longitudinal multinomial outcomes. *Statistical Modelling*, **11(2)**, 115–135.

[13] Fahrmeir, L. and Kaufmann, H. (1987). Regression models for non-stationary categorical time series. *J. of Time Series Analysis*, **8**, 147–160.

[14] Fahrmeir, L. and Tutz, G. (1994). Multivariate Statistical Modelling Based on Generalized Linear Models. Springer - Verlag, New York.

[15] Fienberg, S.E., Bromet, E. J., Follmann, D., Lambert, D., and May, S. M. (1985). Longitudinal analysis of categorical epidemiological data: A study of Three Mile Island. *Environmental Health Perspectives*. **63**, 241–248.

[16] Fokianos, K. and Kedem, B. (2003). Regression theory for categorical time series. *Statistical Science*, **18**, 357–376.

[17] Jiang, J. (1998). Consistent estimators in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **93**, 720-729.

[18] Li, Y. and Chan, W. (2006). Analysis of longitudinal multinomial outcome data. *Biometrical Journal*, **48**, 319–326.

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

[20] Loredo-Osti, J. C. and Sutradhar, B. C. (2011). Estimation of regression and dynamic dependence parameters for non-stationary multinomial time series. Revised for *J. of Time Series Analysis.*

[21] Manski, C.F. (1987). Semi-parametric analysis of random effects linear models for binary panel data. *Econometrica*, **55**, 357–362.

[22] Miller, M. E., Davis, C. S. and Landis, J. R. (1993). The analysis of longitudinal polytomous data: generalized estimating equations and connections with weighted least squares. *Biometrics*, **49**, 1033–1044.

[23] Molenberghs, G. and Lesaffre, E. (1994). Marginal modeling of correlated ordinal data using a multivariate Plackett distribution. *J. Amer. Statist. Assoc.*, *89*, 633–644.

[24] Morel, J. G. and Nagaraj, N. K. (1993). A finite mixture distribution for modelling multinomial extra variation. *Biometrika*, **80, No. 2**, 363–371.

[25] Neerchal, N. K. and Morel, J. G. (2005). An improved method for the computation of maximum likelihood estimates for multinomial overdispersion models. *Comp. Statist. & Data Analysis*, **49**, 33–43.

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

[27] Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*. John Wiley & Sons, New York.

[28] Stram, D. O., Wei, L. J. and Ware, J. H. (1988). Analysis of repeated ordered categorical outcomes with possibly missing observations and time-dependent covariates. *J. Amer. Statist. Assoc.*, **83**, 631–637.

[29] Sutradhar, B.C. (2003). An overview on Regression Models for discrete Longitudinal responses. *Statistical Science*, **18**, 377–393.

[30] Sutradhar, B.C. (2010). Inferences in generalized linear longitudinal mixed models. *Canad. J. Statist.*, **38**, 174–196.

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

[32] Sutradhar, B.C. and Farrell, P. M. (2007). On optimal Lag 1 dependence estimation for dynamic binary models with application to Asthma. *The Ind. J. of Statist.*, **69**, 448–467.

[33] Sutradhar, B.C. and Kovacevic, M. (2000). Ordinal Longitudinal Survey Data: Generalised Estimating Equations Approach. *Biometrika*, **87**, 837–848.

[34] Sutradhar, B.C., Rao, R. P. and Pandit, V. N. (2008). Generalize method of moments versus generalized quasi-likelihood inferences in binary panel data models. *Sankhya B*, **70**, 34–62.

[35] Wilson, J. R. and Koehler, K. J. (1991). Hierarchical models for cross-classified overdispersed multinomial data. *J. of Bus. and Econ. Statistics*, **9**, 103–110.