Lecture Notes in Statistics 218
Proceedings

Brajendra C. Sutradhar *Editor*

# Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data

## Proceedings of the 2015 International Symposium in Statistics

Springer

# Lecture Notes in Statistics

## Proceedings

## Volume 218



More information about this series at http://www.springer.com/series/8440

Brajendra C. Sutradhar
Editor

# Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data

## Proceedings of the 2015 International Symposium in Statistics

Springer

*Editor*
Brajendra C. Sutradhar
Department of Mathematics and Statistics
Memorial University of Newfoundland
St. John's, NL, Canada

ISSN 0930-0325 ISSN 2197-7186 (electronic)
Lecture Notes in Statistics
ISBN 978-3-319-31258-3 ISBN 978-3-319-31260-6 (eBook)
DOI 10.1007/978-3-319-31260-6

Library of Congress Control Number: 2016940011



Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG Switzerland

*To*
*Bhagawan Sri Sathya Sai Baba*

# Preface

This special proceedings volume contains eight selected papers that were presented in the International Symposium in Statistics (ISS) 2015 on Advances in Parametric and Semi-parametric Analysis of Multivariate, Time Series, Spatial-Temporal, and Familial-Longitudinal Data, held in St. John's, Canada, from July 6 to 8, 2015. The main objective of the ISS-2015 was the discussion on advances and challenges in parametric and semi-parametric analysis for correlated data in both continuous and discrete setups. Thus, as a reflection of the theme of the symposium, the eight papers of this proceedings volume are presented in four parts: Part I—Elliptical *t* Distribution Theory; Part II—Spatial and/or Time Series Volatility Models with Applications; Part III—Longitudinal Multinomial Models in Parametric and Semi-parametric Setups; Part IV—An Extension of the GQL Estimation Approach for Longitudinal Data Analysis. The ISS-2015 was the continuation of ISS-2009 and ISS-2012 held in Memorial University. More specifically, the ISS-2009 was organized focussing *on inferences in generalized linear longitudinal mixed models (GLLMMs)*, and a special issue of the *Canadian Journal of Statistics* (2010, Vol. 38, June issue, John Wiley) was published with seven selected papers from this symposium. These seven papers from ISS-2009 dealt with progress and challenges in the areas of longitudinal and/or time series data analysis. As compared to ISS-2009, the papers in the ISS-2012 proceedings volume dealt with inferences for *longitudinal data with additional practical issues such as measurement errors, missing values, and/or outliers.* This proceedings volume was published as the Lecture Notes in Statistics (2013, Vol. 211, Springer) with nine selected papers from the symposium.

It is understood that the elliptical distributions have densities with equiprobable surfaces constant on homothetic ellipsoids, a property possessed in particular by the well-known and widely used multivariate normal distribution. Multivariate *t* distribution also belongs to this elliptic class of distributions, which however has symmetric but fatter tails as compared to the multivariate normal distributions. This additional tail characteristic makes the multivariate *t* distribution useful to analyze the heavy tailed such as stock return data one encounters under volatile financial markets, for example. However, unlike the multivariate normal sampling theory, the elliptical $t$ theories are not adequately discussed in the literature. In fact, unlike the normal sampling theory, the sampling theories for multivariate $t$ responses can be quite different depending on whether the responses in a sample are independent or uncorrelated but dependent. The first paper in Part I, by B.C. Sutradhar, provides an insight on the advances and challenges in inferences for the heavy-tailed data that follow either independent or uncorrelated multivariate $t$ distributions. The paper also proposes a clustered regression model where the multivariate $t$ responses in a cluster are uncorrelated, but such clustered responses are collected from a large number of independent individuals. In the second paper in Part I, R. Prabhakar Rao, B.C. Sutradhar, and V.N. Pandit deal with correlated count and binary data in mixed model setup, where invisible random effects of the individuals are considered to follow independent $t$ distributions. A part of this second paper was presented in the symposium by B.C. Sutradhar in his keynote address because of the common use of $t$ distributions, but the materials of the paper were not included in the first paper in order to show the difference between the analysis of continuous $t$ responses and discrete such as count and binary responses but influenced by continuous $t$ random effects. The main challenge in the second paper is the estimation of the regression parameters when it is known that the $t$ random effects with unknown degrees of freedom parameter cannot be integrated out from the model for any marginal estimation of such regression effects. In fact this problem is also encountered in binary and count data analysis with Gaussian random effects where suitable simulation techniques are used by generating standard Gaussian random effects for the estimation of the parameters including the variance of the normal Gaussian random effects. In the current paper, a similar simulation technique is used where $t$ random effects are generated with unit scale and four degrees of freedom, and a transformation is proposed for the purpose of estimation of the unknown degrees of freedom parameter along with regression and other such as correlation parameters.

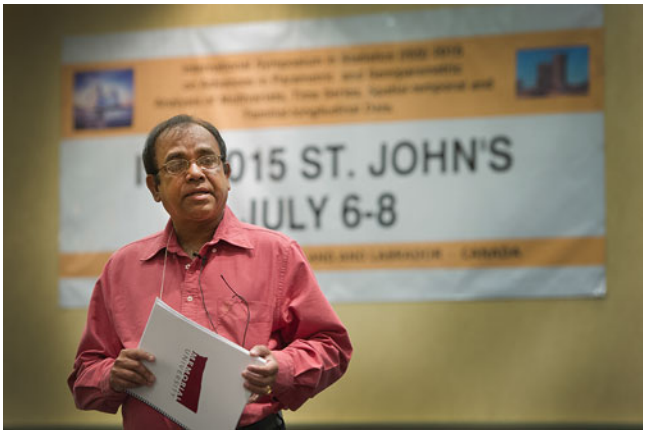
The Part II of the volume contains two papers on spatial and temporal data analysis. The first paper by L.M. Ainsworth, C.B. Dean, and R. Joy is an application paper analyzing spatial counts with a significant portion of responses being zero. This type of zero-inflated count data can provide important clues to physical characteristics associated with, for example, habitat suitability or resistance to disease or pest infestations. However, the probability modeling for this type of spatial bimodal data especially after accommodating spatial correlations is not easy. The authors have considered various existing models and used their expert knowledge to examine what and how these models are doing in order to understand, for example, the white pine weevil infestations data, where many trees did not exhibit any weevil attack or infestation. These models can be grouped into two categories. Some models are appropriate for independent spatial counts with over-dispersion generated by inflated zeros. The rest of the zero-inflated probability models accommodate spatial correlations among counts through correlated random effects. The second paper in Part II contributed by V. Tagore, N. Zheng, and B.C. Sutradhar deals with a special temporal data where the variance of the response data appears to change over time. These heteroscedastic variances are then explained through a suitable dynamic model, and it is of interest to obtain consistent estimates of the parameters involved in such a dynamic model along with consistent estimates for any regression parameters relating time-dependent covariates and the responses. This type of time series models produces larger kurtosis for the data as compared to the usual Gaussian time series with constant variance over time. Consequently these models are found to be suitable to explain the volatility in the data which is, quite often, exhibited in financial markets such as stock return data. The main contribution of the paper is the development of a simpler method of moments technique for consistent estimation of the parameters of the model, as compared to the existing QML (quasi-maximum likelihood) and the so-called popular but very lengthy and complex GMM (generalized method of moments) approaches. The authors have demonstrated the advantage of their estimation technique by reanalyzing the well-known US dollar and Swiss franc exchange rate data to understand the presence of any possible volatility.

Binary dynamic mixed logit (BDML) models are used to fit longitudinal binary data collected from the members of a large number of independent families. In the first paper of Part III, B.C. Sutradhar, R. Viveros-Aguilera, and T. Mallick have provided a generalization of the BDML model to the categorical data setup, binary case being a special case with two categories. Similar to the BDML model, this MDML (multinomial dynamic mixed logit) model uses parametric correlation structure for repeated multinomial data. More specifically, the authors have used dynamic dependence of the current multinomial response on a past response to model the correlations. The regression and dynamic dependence parameters of the model have been estimated by using the likelihood estimation approach. The variance component of the random effects is also estimated by using the likelihood approach. For the binary case, these three parameters, namely, the regression effects, dynamic dependence, and variance component parameters, may be estimated conveniently by using the GQL (generalized quasi-likelihood) approach. The authors have conducted an empirical study to examine the relative performance of the GQL and likelihood estimates for the parameters of a BDML model. The MDL (multinomial dynamic logit) model has been applied as an illustration to analyze a real-life longitudinal categorical data with three categories. The second paper in Part III deals with ordinal categorical data, whereas the first paper was confined to the nominal categorical/multinomial data. This paper by B.C. Sutradhar and N. Dasgupta discusses two correlation models for the ordinal multinomial data. The first model is constructed by cumulating the probabilities for the nominal repeated multinomial responses. The well-known likelihood estimation approach is used to compute the estimates of the parameters of the cumulative model. The second model is constructed in a completely different way than the first model. This model is developed by assuming that the responses at a given time point are available in a cumulative form so that they follow a binary distribution with so-called binary logistic probabilities. To accommodate the correlations for these repeated cumulative responses, a BDL (a binary dynamic logit) model is written involving dynamic dependence between two binary responses. Next, the parameters of this BDL model are estimated by forming a pseudo-likelihood for all possible lag 1 transitional binary probabilities. The pseudo-likelihood estimating equations are provided for all regression and binary dynamic dependence parameters. In the third paper, B.C. Sutradhar considers nominal multinomial variables which is similar to that of the first paper. However, unlike the first paper, this paper developed a semi-parametric probability model for longitudinal multinomial responses. To be specific, to construct such a semi-parametric model, first it is assumed that the traditional specified parametric regression function is not enough to explain the multinomial probabilities. Consequently a non-parametric function is added to the specified regression function which yields a semi-parametric regression function for the construction of the desired multinomial probabilities. The dynamic dependence part remains the same as in the first paper by Sutradhar, Viveros-Aguilera, and Mallick. This type of longitudinal semi-parametric model for multinomial responses is not adequately addressed in the literature. To make this semi-parametric model easily understandable, the author presents a sequence of longitudinal semi-parametric models for repeated linear and count data which are already discussed in the literature to a reasonably good extent. For all these models including the proposed semi-parametric models for the repeated multinomial data, the paper developed suitable estimating equations for the non-parametric function and all other parameters, namely, the regression and the dynamic dependence parameters. These estimation formulas should be useful for any empirical study to analyze repeated multinomial data in a semi-parametric setup.

Part IV of the volume contains one paper by T. Nadarajah, A.M. Variyath, and J.C. Loredo-Osti on the inferences for longitudinal data subject to a challenge of important covariates selection from a set of large number of covariates available for the individuals in the study. The inference technique uses an idea of penalization and estimates the regression parameters corresponding to all covariates involved in a related penalized generalized estimating function, where this later function is constructed by modifying a generalized quasi-likelihood (GQL) estimating function. Any regression parameter estimate close to zero obtained by solving the penalized GQL (PGQL) estimating equation automatically removes the unimportant covariates from the model, leading to a reduced model with important covariates for interpretation. The authors of the paper have verified the performance of this PGQL inference technique through an intensive simulation study. A data analysis is also provided justifying the technique.

St. John's, NL, Canada Brajendra C. Sutradhar

ISS-2015 delegates




ISS-2015 welcome address by Brajendra Sutradhar (Organizer)

Further to the welcome by Professor Charmaine Dean [former president of the SSC (Statistical Society of Canada) and the current dean of science of the University of Western Ontario] and Dr. Alwell Oyet [deputy head of the Department of Mathematics and Statistics at Memorial University], once again I welcome all of you with the name of the Lord to this International Symposium in Statistics 2015 (ISS-2015) on Advances in Parametric and Semi-parametric Analysis of Multivariate, Time Series, Spatial-Temporal, and Familial-Longitudinal Data. As noted in the symposium web site (http://www.iss-2015-stjohns.ca./), the ISS-2015 is the continuation of ISS-2009 and ISS-2012. Both of the last two symposiums were held in Memorial University, Canada, and they were devoted to the discussion of progresses and challenges in the analysis of longitudinal data subject to measurement errors, missing values, and/or outliers. These two symposiums were highly successful with two high-quality proceedings volumes, one in the form of a special issue of the *Canadian Journal of Statistics* in 2010 and the other as a Springer Lecture Notes in Statistics in 2013. It is my pleasure to note that we have been able to keep up the spirit of the last two symposiums in organizing the discussion topics of the present symposium covering the progress and advances in correlated data analysis in a variety of setups, such as spatial and/or temporal setup, semi-parametric setup for discrete longitudinal data, and multivariate setups for discrete familial-longitudinal and continuous non-Gaussian elliptical data.

I am very grateful to Bhagawan Sri Sathya Sai Baba, my guru, the universal spiritual master, for his blessings and inspirations in organizing these international community services. I am also thankful to all of you for your interest and response to this 2015 symposium that has attempted to attract the noble group of researchers including graduate students. I hope that you will find the symposium stimulating and will derive spirits for doing more quality research in these challenging areas as a service to the society and mankind at large.

As far as the presentation structure of this symposium is concerned, four keynote speeches are organized in four different areas to be delivered by three speakers. Professor Anthony C. Davison from EPFL, Switzerland, will give his keynote address on max-stable processes on river networks, under the theme of spatial-temporal data analysis. Professor Brajendra C. Sutradhar from Memorial University, Canada, will deliver part 1 of his keynote presentation on advances and challenges in correlated data analysis in non-Gaussian multivariate setup and part 2 of the presentation on advances and challenges in analyzing ordinal categorical data in semi-parametric setup. Part 3 of the keynote address will be given by Professor Andrew Harvey from Cambridge University, UK, on new developments in modeling dynamic volatility. Nine special invited talks over 3 days of the symposium will be given by Professors Paul D. Sampson, University of Washington; Grace Y. Yi, University of Waterloo; Nairanjana Dasgupta, Washington State University; Roman Viveros-Aguilera, McMaster University; Julio M. Singer, Universidade de Sao Paulo; David E. Tyler, Rutgers—the State University of New Jersey; Refiq Soyer, George Washington University, Charmaine Dean; University of Western Ontario; and Richard J. Cook, University of Waterloo. The symposium has another two invited speakers, Dr. Alwell Oyet from Memorial University and Dr. Ashis SenGupta from Indian Statistical Institute. Also, contributed papers will be presented by seven speakers including four graduate students. Furthermore, it is planned that a selected number of papers presented in the symposium will be published in the near future as lecture notes in the Springer's Lecture Note Series.

It is also a pleasure to note that we have 46 delegates in this specialized symposium from many countries such as Brazil, France, India, Switzerland, the USA, and Canada, covering a large part of the globe. The organizing committees would like to extend a hearty welcome to all of you including all graduate students.

We also welcome you to St. John's, the oldest city of North America, known as the City of Legends, where you can view icebergs, watch whales, and experience Newfoundland and Labrador's unique culture. It is a progressive city and is the site of many world-class facilities including an international center in marine science and technology. A mosaic of fishing villages, cultural festivals, and wildlife tours bring variety to the city. Also, Cape Spear, the most easterly point of North America, is not far from the city, where one can experience the unique beauty of sunrise. We hope that you have planned for an extended stay in St. John's following the symposium to enjoy these and other endless options!

# Acknowledgments

This proceedings volume (lecture note) is a collection of selected papers that were presented in ISS-2015 (International Symposium in Statistics 2015) held in St. John's, Canada, from July 6 to 8, 2015. All papers in this volume were refereed. Prior to the symposium, the papers were sent to the referees who were also supposed to be present during the presentation. The authors also benefitted from the warm discussion by the audience of the symposium and prepared the revision of the paper by addressing all suggestions and comments from the referees and the audience. My sincere thanks go to the delegates and referees for making the symposium and this volume a grand success. Some of the contributed papers were also considered for their publication in this volume. Special thanks go to Dr. J.C. Loredo-Osti for his warm service in processing all contributed papers for the symposium.

Organizing this symposium would not have been possible without the generous contributions from Memorial University, the Atlantic Association for Research in Mathematical Sciences (AARMS), and the Canadian Statistical Sciences Institute (CANSSI). I wish to express my special thanks to these three institutes for their support.

It has been a pleasure to work with Jon Gurstelle, Hannah Bracken, Christine Crigler, and Matthew Amboy of Springer-Verlag in preparing this volume. I also wish to thank the project manager Ms. S. Bharathi, and the Production team at Springer Spi-Global, India, for this superb production job.

# Contents

# List of Contributors

**L.M. Ainsworth** Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC, Canada

**J. Concepción Loredo-Osti** Memorial University, St. John's, NL, Canada

**Nairanjana Dasgupta** Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA

**C.B. Dean** Department of Statistical and Actuarial Science, University of Western Ontario, London, ON, Canada

**R. Joy** SMRU Consulting Ltd., Vancouver, BC, Canada

**Taslim S. Mallick** Department of Statistics, Biostatistics and Informatics, University of Dhaka, Dhaka, Bangladesh

**Tharshanna Nadarajah** Memorial University, St. John's, NL, Canada

**V.N. Pandit** Department of Economics, Sri Sathya Sai Institute of Higher Learning, Puttaparthi, Andhra Pradesh, India

**R. Prabhakar Rao** Department of Economics, Sri Sathya Sai Institute of Higher Learning, Puttaparthi, Andhra Pradesh, India

**Brajendra C. Sutradhar** Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada

**Vickneswary Tagore** Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada

**Asokan Mulayath Variyath** Memorial University, St. John's, NL, Canada

**Roman Viveros-Aguilera** Department of Mathematics and Statistics, McMaster University, Hamilton, ON, Canada

**Nan Zheng** Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada

# Part I
# Elliptical *t* Distribution Theory

# Advances and Challenges in Inferences for Elliptically Contoured *t* Distributions

**Brajendra C. Sutradhar**

**Abstract** When a multivariate elliptical such as $t$ response is taken from each of $n$ individuals, the inference for the parameters of the $t$ distribution including the location (or regression effects), scale and degrees of freedom (or shape) depends on the assumption whether $n$ multi-dimensional responses are independent or uncorrelated but dependent. In the former case, that is, when responses are independent, the exact sampling theory based inference is extremely complicated, whereas in the later case the derivation of the exact sampling distributions for the standard statistics is manageable but the estimators based on certain standard statistics such as sample covariance matrix may be inconsistent for the respective parameters. In this paper we provide a detailed discussion on the advances and challenges in inferences using uncorrelated but dependent $t$ samples. We then propose a clustered regression model where the multivariate $t$ responses in the cluster are uncorrelated but such clustered responses are taken from a large number of independent individuals. The inference including the consistent estimation of the parameters of this proposed model is also presented.

**Keywords** Clustered regression model with uncorrelated $t$ errors • Consistent estimation • Elliptically contoured distribution • Multivariate $t$ and normal as special cases • Normality based testing yielding degrees of freedom based power property • Regression effects • Scale matrix • Shape or degrees of freedom parameter

B.C. Sutradhar (✉)
Department of Mathematics and Statistics, Memorial University,
St. John's, NL, Canada A1C5S7
e-mail: bsutradh@mun.ca



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_1

# 1 Introduction

Since its inception in early twentieth century, the so-called univariate $t$-distribution proposed by Student (1908) [see also Fisher (1925)] and its multivariate generalization, namely multivariate $t$-distribution (Cornish 1954; Dunnett and Sobel 1954) have become widely popular tools to make inferences about Gaussian (normal) means. These $t$-distributions are derived from the respective normal distributions as follows. Suppose that $Y^* = (Y_1^*, \ldots, Y_p^*)'$ is a $p$-dimensional normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_p)'$ and $p \times p$ covariance matrix $\Lambda$, that is

$$Y^* \sim N_p(\mu, \Lambda), \tag{1}$$

where $\Lambda$ is the fixed $p \times p$ positive definite covariance matrix. Next, suppose that $\xi^2$ a scalar random variable which follows the well known $\chi^2_\nu$ distribution with $\nu$ degrees of freedom, that is, $\xi^2 \sim \chi^2_\nu$, and $Y^*$ and $\xi^2$ are independent. Then, for $Z^* = \Lambda^{-\frac{1}{2}}(Y^* - \mu) \sim N_p(0, I)$, the ratio variable $Y$ defined as

$$Y = \frac{1}{\sqrt{(\xi^2/\nu)}} \Lambda^{\frac{1}{2}} Z^* \tag{2}$$

has the multivariate $t$-distribution given by

$$f(y|\Lambda, \nu) = c(\nu, p)|\Lambda|^{-\frac{1}{2}} \frac{1}{[\nu + y'\Lambda^{-1}y]^{\frac{\nu+p}{2}}}, \tag{3}$$

yielding the distribution for

$$Y = \mu + \frac{1}{\sqrt{(\xi^2/\nu)}} \Lambda^{\frac{1}{2}} Z^* \tag{4}$$

as

$$f(y|\mu, \Lambda, \nu) = c(\nu, p)|\Lambda|^{-\frac{1}{2}} \frac{1}{[\nu + (y-\mu)'\Lambda^{-1}(y-\mu)]^{\frac{\nu+p}{2}}}, \tag{5}$$

(Cornish 1954; Dunnett and Sobel 1954) where the normalizing constant $c(\nu, p)$ has the formula

$$c(\nu, p) = \frac{\nu^{\nu/2}\Gamma[\frac{1}{2}(\nu + p)]}{(\sqrt{\pi})^p \Gamma[\frac{1}{2}\nu]}.$$

This distribution in (5) will be denoted as

$$Y \sim t_p(\mu, \Lambda, \nu). \tag{6}$$

For $p = 1$ with $\mu = \mu_1$ and $\Lambda = \lambda^2$, the multivariate $t$-density in (5) reduces to the univariate $t$-density

$$f(y_1|\mu_1, \lambda^2, \nu) = c(\nu, 1)|\lambda^2|^{-\frac{1}{2}} \frac{1}{[\nu + \frac{(y_1-\mu_1)^2}{\lambda^2}]^{\frac{\nu+1}{2}}}, \tag{7}$$

developed by Student (1908). For details on the applications of univariate (7) and multivariate (5) $t$-distributions, one may be referred to Kotz and Nadarajah (2004), for example.

Note that there has been many applications of $t$ distributions for inferences about the fitting of a normal distribution to a data set (Cornish 1954; Dunnett and Sobel 1954). But, fitting of a $t$ distribution to a data set with symmetric heavy tails has not been adequately discussed, especially when the data are independent. In this setup, the joint distribution of independent $t$ responses does not have a closed form and hence the theory of inference gets complicated. For some empirical studies using independent $t$ distributions, one may refer to Chib et al. (1988, 1991). However, as a generalization of joint normal distribution, there exist a joint multivariate $t$ distribution where the observations are uncorrelated but not necessarily independent. This class of $t$ distributions was proposed by Sutradhar and Ali (1986) for multi-dimensional uncorrelated but dependent observations (see also Sutradhar 1988, 1990, 1993) as a generalization of a multivariate $t$ distribution for uncorrelated but dependent scalar observations (Zellner 1976). For convenience, we show below how these three types of $t$ distributions may arise in practice.

## *Case 1. Inference for Normal Models Using t Distribution*

Note that in statistics and econometrics literature, the multivariate t-distribution (5) has been mainly used to study various inferential issues for multivariate normal data. For example, consider following case. Suppose that

$$Y_i \stackrel{iid}{\sim} N_p(\mu, \Lambda), \tag{8}$$

implying that

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \sim N_p(\mu, \frac{1}{n}\Lambda)$$
$$\xi^2 = \sum_{i=1}^{n}(Y_i - \bar{Y})'\Lambda^{-1}(Y_i - \bar{Y}) \sim \chi^2_{(n-1)p}, \tag{9}$$

and $\bar{Y}$ and $\xi^2$ are independent. Also,

$$Z = (\frac{1}{n}\Lambda)^{-\frac{1}{2}}(\bar{Y} - \mu) \sim N_p(0, I_p).$$

For known $\Lambda$, it then follows by (2) and (3) that

$$\frac{1}{\sqrt{\xi^2/(n-1)p}}(\frac{1}{n}\Lambda)^{\frac{1}{2}}Z = \frac{1}{\sqrt{\xi^2/(n-1)p}}[\bar{Y} - \mu] \sim t_p(0, \Lambda, (n-1)p), \tag{10}$$

which for the univariate case with $p = 1$ and $\xi^2 = \sum_{i=1}^{n}[\frac{(y_i - \bar{y})}{\lambda}]^2 = (n-1)\frac{s^2}{\lambda^2}$, reduces to

$$\frac{\bar{y} - \mu}{s/\sqrt{n}} \sim t_1(0, \lambda^2, n-1). \tag{11}$$

One may then construct confidence interval for $\mu$ in the multivariate case by using $t_p(0, \Lambda, (n-1)p)$ distribution from (10), and in the univariate case by using $t_1(0, \lambda^2, n-1)$ distribution from (11).

## *Case 2. Independent t Models*

Suppose that $Y_1, \ldots, Y_i, \ldots, Y_n$, are independently distributed (id) [as opposed to independently and identically distributed (iid)] as

$$Y_i \overset{id}{\sim} N_p(\mu, \psi_i^2 \Lambda), \tag{12}$$

where $\psi_1, \ldots, \psi_i, \ldots, \psi_n$ are random and independent scales each with a one-parameter ($\nu$) based inverted gamma distribution given as

$$g(\psi_i) = \frac{2(\nu/2)^{-1/2}}{\Gamma(\nu/2)} \exp\left\{-\frac{1}{2}(\nu/\psi_i^2)\right\}\left[\frac{\nu}{2\psi_i^2}\right]^{(\nu+1)/2}, \quad \psi_i > 0, \tag{13}$$

yielding the $t$-distribution for $Y_i$ as

$$\begin{aligned} f(y_i) &= \int_0^\infty \frac{1}{(\sqrt{2\pi})^p|\psi_i^2\Lambda|^{\frac{1}{2}}} \exp[-\frac{1}{2\psi_i^2}\{(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}]g(\psi_i)d\psi_i \\ &= \frac{\Gamma(\frac{\nu+p}{2})}{\pi^{p/2}\nu^{1/2}\Gamma(\frac{\nu}{2})}\Lambda^{-\frac{1}{2}}\left[1 + \frac{\{(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}}{\nu}\right]^{-(\nu+p)/2}, \end{aligned} \tag{14}$$

[see Sutradhar (1988, p. 176), e.g.] with $\nu > 0$ degrees of freedom, for all $i = 1, \ldots, n$. This $t$-distribution in (14) is similar to that of (3). It is then clear that $Y_i \overset{iid}{\sim} t_p(\mu, \Lambda, \nu)$, yielding the joint distribution as

$$f(y_1, \ldots, y_i, \ldots, y_n) = [c(\nu, p)]^n |\Lambda|^{-\frac{n}{2}} \Pi_{i=1}^n [\nu + (y_i - \mu)' \Lambda^{-1} (y_i - \mu)]^{-\frac{\nu+p}{2}}, \tag{15}$$

which is also known as independent $t$ model. Here $c(\nu, p)$ is the normalizing constant as in (3). Note that by maximizing (15), one may obtain the likelihood estimates of $\mu$, $\Lambda$, and $\nu$, but studying the finite sample properties of these estimators is extremely complicated (Walker and Saw 1978).

## *Case 3. Uncorrelated but Dependent t Models*

As opposed to the independent $t$ model (15), Sutradhar and Ali (1986) [see also Sutradhar (1988, 1990)] have proposed a dependent $t$ model which is obtained as follows. Assume that $Y_1, \ldots, Y_i, \ldots, Y_n$ are iid normal, that is,

$$Y_i \overset{iid}{\sim} N_p(\mu, \psi^2 \Lambda),$$

with $\psi$ following the same inverted gamma distribution as in (13). It then follows that

$$\begin{aligned} f(y_1, \ldots, y_i, \ldots, y_n) &= \int_0^\infty \Pi_{i=1}^n [N(y_i; \mu, \psi^2 \Lambda)] g(\psi) d\psi \\ &= \int_0^\infty [\frac{1}{(\sqrt{2\pi})^p |\psi^2 \Lambda|^{\frac{1}{2}}}]^n \exp[-\frac{1}{2\psi^2} \sum_{i=1}^n \{(y_i - \mu)' \Lambda^{-1} (y_i - \mu)\}] g(\psi_i) d\psi_i \\ &= c(\nu, np) |\Lambda|^{-\frac{n}{2}} [\nu + \sum_{i=1}^n \{(y_i - \mu)' \Lambda^{-1} (y_i - \mu)\}]^{-\frac{\nu+np}{2}}, \end{aligned} \tag{16}$$

where $c(n, np)$ is obtained from $c(\nu, p)$ in (3) by replacing $p$ with $np$. Unlike (15), this joint distribution in (16) has the $t$ distribution form and is denoted by $t_{np}(\mu, \Lambda, \nu)$. For $\nu > 2$, it may be shown under (16) that

$$\begin{aligned} E[Y_i] &= \mu, \text{ for all } i = 1, \ldots, n; \\ \text{var}[Y_i] &= \Sigma = \frac{\nu}{\nu - 2} \Lambda, \text{ for all } i = 1, \ldots, n; \\ \text{cov}[Y_i, Y_j] &= 0, \text{for } i \neq j, i, j = 1, \ldots, n. \end{aligned} \tag{17}$$

Thus, the $p$-dimensional observations $y_1, \ldots, y_i, \ldots, y_n$, are pair-wise uncorrelated, but, it follows from (16) that they are not independent. Sutradhar (1988, 1990, 1993) and Sutradhar and Ali (1986, 1989) have exploited this class of uncorrelated but dependent $t$-distributions (16) for inferences about $\mu$ and $\Sigma$, assuming that $\nu$ is known. Sutradhar (1994) and Sutradhar (2004) have dealt with estimation of $\nu$ as well.

A brief plan of the paper is as follows. The independent $t$ models are discussed in Sect. 2.1. The exact distribution theory and estimation methods under uncorrelated but dependent elliptical $t$ models are discussed in details in Sects. 2.2 and 3, respectively. In Sect. 4, we provide several new familial and longitudinal models for $t$ data, where, more specifically, each of the independent clusters contain uncorrelated or correlated multi-dimensional elliptical $t$ responses. Some results on null and null robust properties for $t$ samples based test statistics are reviewed in Sect. 5. The paper concludes in Sect. 6.

## 2 Exact Versus Asymptotic Sampling Distribution Theory

### *2.1 Distribution Theory for Independent t Sample*

When $p$-dimensional responses $\{y_i,\ i = 1, \ldots, n\}$ are independent following the multivariate $t$ distribution (14), their joint distribution is written as (15). In general it is of interest to estimate $\mu,\ \Lambda,$ and $\nu$. However as $\Lambda$ is the scale matrix, it is convenient to estimate the covariance matrix $\Sigma = \frac{\nu}{\nu-2}\Lambda$ by using the corresponding sample moments. For example, the moment estimates for $\mu,\ \Sigma,$ and $\nu$ can be obtained as

**Moment estimates:**

First we write the moment estimates for $\mu$ and $\Sigma$ as

$$\hat{\mu}_M = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\hat{\Sigma}_M = S = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(y_i - \bar{y})', \qquad (18)$$

and then obtain the moment estimate of $\nu$ by equating the multivariate measure of kurtosis of the $t$ distribution $f(y_i)$ in (14), namely

$$\beta_2^* = \int [(y_i - \mu)'\Sigma^{-1}(y_i - \mu)]^2 f(y_i)dy_i, \qquad (19)$$

(e.g., Mardia 1970) to its sample counterpart

$$\beta_{2,M}^* = \frac{1}{n}\sum_{i=1}^{n}[(y_i - \bar{y})'S^{-1}(y_i - \bar{y})]^2.$$

Next, for $\nu > 4$, it can be shown that $\beta_2^*$ in (19) has the formula

$$\beta_2^* = \left(\frac{\nu-2}{\nu-4}\right) f^*(\sigma), \tag{20}$$

where

$$f^*(\sigma) = \left[ 3\sum_{h=1}^{p} (\sigma^{hh})^2 (\sigma_{hh})^2 \right.$$
$$\left. + \sum_{h \neq h'}^{p} (\sigma_{h'h'})^2 \{\sigma^{hh}\sigma^{h'h'} + (\sigma^{hh'})^2\} \right], \tag{21}$$

with $\Sigma = (\sigma_{hu}),\ \Sigma^{-1} = (\sigma^{hu})$. Now by using

$$f^*(s) = f^*(\sigma)_{|\sigma_{hu}=s_{hu},\sigma^{hu}=s^{hu}}, \tag{22}$$

where $S = (s_{hu}),\ S^{-1} = (s^{hu})$, one obtains the moment estimator of $\nu$ as

$$\hat{\nu}_M = \frac{2[2\hat{\beta^*}_{2,M} - f^*(s)]}{[\hat{\beta^*}_{2,M} - f(s)]}. \tag{23}$$

### 2.1.1 Asymptotic Properties

All three estimators given in (18) and (23) are consistent for their respective parameter. For example, because $Y_i \overset{iid}{\sim} t_p(\mu, \Lambda, \nu)$ as in (16), it follows that $\bar{Y}$ in (18) has

$$E[\bar{Y}] = \mu,\ \mathrm{Lt}_{n\to\infty}\mathrm{var}[\bar{Y}] = \mathrm{Lt}_{n\to\infty}\frac{1}{n}[\frac{\nu}{\nu-2}\Lambda] = \mathrm{Lt}_{n\to\infty}\frac{1}{n}\Sigma \to 0, \tag{24}$$

showing that $\bar{Y}$ is consistent for $\mu$. Further, by standard central limit theorem, it follows that asymptotically (as $n \to \infty$)

$$\bar{Y} \sim N_p(\mu, \frac{1}{n}\Sigma). \tag{25}$$

Consequently, one can easily construct the desired confidence interval for $\mu$, provided $n$ is large.

As far as the as the asymptotic properties of the sample covariance matrix $S$ in (18) is concerned, one may first show that

$$\begin{aligned} E[S] &= \frac{1}{n-1}E[\sum_{i=1}^{n}(y_i-\bar{y})(y_i-\bar{y})'] \\ &= \frac{1}{n-1}E[\sum_{i=1}^{n}(y_i-\mu+\mu-\bar{y})(y_i-\mu+\mu-\bar{y})'] \\ &= \frac{1}{n-1}[\sum_{i=1}^{n}E\{(y_i-\mu)(y_i-\mu)'\}-nE\{(\bar{y}-\mu)(\bar{y}-\mu)'\}] \\ &= \frac{1}{n-1}[n\Sigma-n(\Sigma/n)] \\ &= \Sigma=(\sigma_{uv})=\frac{\nu}{\nu-2}\Lambda=\frac{\nu}{\nu-2}(\lambda_{uv}). \end{aligned} \tag{26}$$

Next, because $S=(s_{uv})$ in (18) is the sample covariance matrix constructed from a sample of $n$ independent and identically distributed vectors $\{Y_i \overset{iid}{\sim} t_p(\mu,\Lambda,\nu),\ i=1,\ldots,n\}$, it then follows, for example, from Muirhead (1982, Eq. (3), p. 42; Exercise 1.33 (b), p. 49) that

$$\begin{aligned} \mathrm{cov}[s_{uv},s_{h\ell}] = \frac{1}{n-1}&\left[\kappa\{\sigma_{uv}\sigma_{h\ell}+\sigma_{uh}\sigma_{v\ell}+\sigma_{u\ell}\sigma_{vh}\}\right. \\ &\left.+\{\sigma_{uh}\sigma_{v\ell}+\sigma_{u\ell}\sigma_{vh}\}\right], \end{aligned} \tag{27}$$

where for the $t$ distribution (14), the kurtosis parameter $\kappa=\frac{2}{\nu-4}$. Consequently,

$$\mathrm{Lt}_{n\to\infty}\mathrm{cov}[s_{uv},s_{h\ell}]\to 0, \tag{28}$$

indicating that $S$ is a consistent estimator of $\Sigma$. Furthermore, by Corollary 1.2.18 in Muirhead (1982), it follows that

$$(n-1)^{\frac{1}{2}}[\mathrm{vector}(S)-\mathrm{vector}(\Sigma)]\sim N_{n^2}(0,V), \tag{29}$$

where $V=(\mathrm{cov}[s_{uv},s_{h\ell}])\,:\ n^2\times n^2$, by (27).

### 2.1.2 Exact Sampling Distribution Theory Challenge

The multivariate $t$ distribution (14) provides a useful extension of the normal distribution for statistical modeling of data with longer than normal tails. Despite this fact, the use of independently chosen samples from the parent $t$-distribution in modeling long tailed symmetric data has been hampered by the complexity of the exact sampling distribution theory. For a discussion on this, see for example Sutradhar (1998, pp. 440–448). To illustrate the complexity with the exact sampling distribution theory, consider for example the univariate case with $p = 1$. In this case, using $\Lambda = \lambda^2$, the joint distribution (15) may be written as

$$f(y_1, \ldots, y_i, \ldots, y_n) = [c(\nu, 1)]^n (\lambda^2)^{-\frac{n}{2}} \times \Pi_{i=1}^n [\nu + (y_i - \mu)' \{\lambda^2\}^{-1} (y_i - \mu)]^{-\frac{\nu+1}{2}}, \quad (30)$$

where $c(\nu, 1) = ]\nu^{\nu/2} \Gamma\{(\nu + 1)/2\}]/[\pi^{1/2} \Gamma(\nu/2)]$. Now finding the distribution of a linear combination of the form $\sum_{j=1}^n a_j y_j$ is very complicated. For $n = 2$, Ghosh (1975) gave an explicit formula for the distribution function of $U = y_1 + y_2$ (or $V = y_1 - y_2$) in terms of hypergeometric functions for $\nu \leq 4$. For $4 < \nu < \infty$, the author used the numerical integration methods to tabulate several values of its distribution function. Walker and Saw (1978) have used the characteristic function of the $t$-distribution with odd degrees of freedom (Fisher and Healy 1956) to obtain the distribution of a special linear combination of $t$ random variables, when these variables are chosen from $t$-distribution with odd degrees of freedom only. By exploiting the general characteristic function available in Sutradhar (1986, 1988), one may similarly obtain the distribution function of a general linear combination of $t$ random variables with any type of degrees of freedom (odd, even or fraction). This is, however, still an open problem. Also note that unlike the normal distribution, the characteristic function of a linear combination of $t$ independent variables does not have the same form as that of the characteristic function of a single $t$ variable. This makes the use of the characteristic function of the $t$ distribution difficult to find the distribution of a linear combination or similar statistics.

Finding the exact sampling distribution of more complex test statistics than the linear combination will naturally be more complicated. The exact distribution of the $t$-statistic, for example, under the parent Student's $t$-distribution (7), that is, using (30), is extremely difficult to obtain analytically except for sample size 2 or 3. For a Monte Carlo method to obtain the percentage points of the distribution of the $t$-statistic, see Yuen and Murthy (1974). Similarly, in the multivariate case Hotelling's $T^2$ test in the independence setup is not robust against departure from normality, because it is not only that finding its null and non-null distributions is extremely difficult but it appears that its null distribution does not follow the central $F$-distribution when one tests the equality of the means of two independent $t$-distribution.

## 2.2 Distribution Theory for Uncorrelated but Dependent t Sample

### 2.2.1 Marginal Distribution

Suppose that $Y_i \sim t_p(\mu, \Lambda, \nu)$, that is, the pdf of $y_i$ is given by (5). Next partition $y_i$ as

$$y_i = \begin{bmatrix} y_{1i} \\ y_{2i} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \lambda_{21} & \Lambda_{22} \end{bmatrix}, \tag{31}$$

where $y_{1i} : m \times 1, \mu_1 : m \times 1$ and $\Lambda_{11} : m \times m$. Then $y_{1i} \sim t_m(\mu_1, \Lambda_{11}, \nu)$. That is,

$$f(y_{1i}) = c(\nu, m)|\Lambda_{11}|^{-\frac{1}{2}} \frac{1}{[\nu + (y_{1i} - \mu_1)'\Lambda_{11}^{-1}(y_{1i} - \mu_1)]^{\frac{\nu+m}{2}}} \tag{32}$$

[see for example Muirhead (1982, exercise 1.29 (b))]. This marginal distribution is derived in Raiffa and Schlaifer (1961) (see also Johnson and Kotz 1972) by using the gamma mixture of the multivariate normal density similar to that of (4). This marginal distribution also may be verified using the characteristic function of the $t$ distribution (5) from Sutradhar (1986, Theorem 1, p. 330) denoted by $\phi_{y_i}(s; \nu, \mu, \Lambda)$, where $s \equiv [s_1, \ldots, s_m, \ldots, s_p]'$ is a vector of $p$ auxiliary parameters. Putting $s_{m+1}, \ldots, s_p = 0$ in the characteristic function $\phi_{y_i}(s; \nu, \mu, \Lambda)$, one obtains the characteristic function of $y_{1i}$ with the same form as that of $y_i$, implying that $y_{1i} \sim (\mu_1, \Lambda_{11}, \nu)$.

A Generalization to the General Elliptical Distribution

For an individual $i$, the elliptical distribution of $y_i$ has the pdf of the form

$$c^*(\nu_0, p)|\Lambda|^{-\frac{1}{2}} g\{\nu_0(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}, \tag{33}$$

(Kelker 1970) with $g(\cdot)$ a known form and $c^*(\nu_0, p)$ as the normalizing constant. Then for $y_i$ partitioned as in (31), $y_{1i}$ has the elliptic distribution given by

$$f(y_{1i}) = c^*(\nu_0, m)|\Lambda_{11}|^{-\frac{1}{2}} g_m\{\nu_0(y_{1i} - \mu_1)'\Lambda_{11}{}^{-1}(y_{1i} - \mu_1)\}, \tag{34}$$

(Kariya 1981; Kelker 1970; Sutradhar and Ali 1989) where $g_m(\cdot)$ is determined only by the form of $g$ in (33) and by the number of components in $y_i$.

### An Application of Marginal Property: t Distribution Case

Suppose that $y_1, \ldots, y_i, \ldots, y_n$ are $n$ $p$-dimensional uncorrelated responses and they jointly follow the $np$-dimensional $t$ distribution as in (16), that is,

$$f(y_1, \ldots, y_i, \ldots, y_n) = c(\nu, np)|\Lambda|^{-\frac{n}{2}}[\nu + \sum_{i=1}^{n}\{(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}]^{-\frac{\nu+np}{2}}. \quad (35)$$

It then follows from the marginal property (31)–(32) that $y_i$ follows the $p$-dimensional $t$ distribution as $y_i \sim t_p(\mu, \Lambda, \nu)$. This is because, by writing a $np$-dimensional observation

$$y^* = [y_1', \ldots, y_i', \ldots, y_n']',$$

one can write

$$\sum_{i=1}^{n}\{(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\} = (y^* - \mu^*)'[I_n \otimes \Lambda](y^* - \mu^*), \quad (36)$$

where $\otimes$ denotes the Kronecker or direct product, and $\mu^* = 1_n \otimes \mu$, $1_n$ being the $n$-dimensional unit vector. Consequently, the joint pdf (35) may be re-expressed as

$$f(y^*|\mu^*, I_n \otimes \Lambda, \nu) = c(\nu, np)|In \otimes \Lambda|^{-\frac{1}{2}}[\nu + \{(y^* - \mu^*)'[I_n \otimes \Lambda]^{-1}(y^* - \mu^*)\}]^{-\frac{\nu+np}{2}}, \quad (37)$$

which is the same form as that of the $t$ distribution in (5) but it is a $np$-dimensional distribution.

### An Application of Marginal Property: Elliptical Distribution Case

By similar arguments as for the $t$ distribution case, suppose that $y_1, \ldots, y_i, \ldots, y_n$ are $n$ $p$-dimensional uncorrelated responses and they jointly follow the $np$-dimensional elliptical distribution given by

$$f(y_1, \ldots, y_i, \ldots, y_n) = c^*(\nu_0, np)|\Lambda|^{-\frac{n}{2}} g(\nu_0 \sum_{i=1}^{n}\{(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}). \quad (38)$$

Because, using the same notation as in (37), this density in (38) may be expressed as

$$f(y^*) = c^*(\nu_0, np)|I_n \otimes \Lambda|^{-\frac{1}{2}} g(\nu_0\{(y^* - \mu^*)'[I_n \otimes \Lambda]^{-1}(y^* - \mu^*)\}), \quad (39)$$

$y_i$ involved in $y^*$, by (33)–(34), has the marginal elliptical distribution given by

$$f(y_i) = c^*(\nu_0, p)|\Lambda|^{-\frac{1}{2}} g_p\{\nu_0(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}. \quad (40)$$

### 2.2.2 Distribution of Linear Combination of Elliptical *t* Variables

We first provide this distribution for general elliptic variables. Let $X = DY$, where $D$ is an $m \times p$ matrix of rank $m,\ m \leq p$, and $Y$ is a $p$-dimensional random variable following the general elliptical distribution (33), that is,

$$f(y) = c^*(\nu_0, p)|\Lambda|^{-\frac{1}{2}} g\{\nu_0(y-\mu)'\Lambda^{-1}(y-\mu)\}.$$

One may then show that the $m$-dimensional variable $X$ has the p.d.f. given by

$$f(x) = c^*(\nu_0, m)|D\Lambda D'|^{-\frac{1}{2}} g_m\{\nu_0(x-D\mu)'(D\Lambda D')^{-1}(x-D\mu)\}. \tag{41}$$

This p.d.f. in (41) may be derived from the p.d.f. of

$$U = \begin{pmatrix} X \\ W \end{pmatrix} = \begin{pmatrix} D \\ Q \end{pmatrix} Y = BY,$$

where $B = \begin{pmatrix} D \\ Q \end{pmatrix}$ is a $p \times p$ nonsingular matrix. Because $B$ is a nonsingular matrix, similar to the multivariate normal case, one may use the aforementioned p.d.f. of $Y$ and write the p.d.f. of $U = BY$ as

$$f(u) = c^*(\nu_0, p)|B\Lambda B'|^{-\frac{1}{2}} g\{\nu_0(u-B\mu)'(B\Lambda B')^{-1}(u-B\mu)\}. \tag{42}$$

The p.d.f. of $X$ in (41) then follows from (42) by using the marginal distribution property given in (34).

#### Distribution of the Sample Mean Vector $\bar{Y}$ Under Elliptical Distribution

Notice that $y_1, \ldots, y_i, \ldots, y_n$ have the $np$-dimensional joint distribution given by (38), yielding the distribution of $y^* = [y_1', \ldots, y_i', \ldots, y_n']'$ as given by (39). Now write

$$\bar{y} = [\bar{y}_1, \ldots, \bar{y}_p]' = x^* = D^* y^* = \frac{1}{n}[I_p, \ldots, I_p]y^*, \tag{43}$$

where

$$D^* = \frac{1}{n}[I_p, \ldots, I_p]$$

is the $p \times np$ matrix with rank $p$, which is similar to the $D$ matrix in (41). Hence, by (39), the distribution of $\bar{Y}$ follows from (41) and is given by

$$
\begin{aligned}
f(\bar{y}) &= c^*(\nu_0,p)|D^*(I_n\otimes\Lambda)D^{*\prime}|^{-\frac{1}{2}}g_p\{\nu_0(\bar{y}-D^*\mu^*)'(D^*(I_n\otimes\Lambda)D^{*\prime})^{-1}(\bar{y}-D^*\mu^*)\}\\
&= c^*(\nu_0,p)|\frac{1}{n}\Lambda|^{-\frac{1}{2}}g_p\{\nu_0(\bar{y}-\mu)'(\frac{1}{n}\Lambda)^{-1}(\bar{y}-\mu)\} \qquad (44)
\end{aligned}
$$

[see also Theorem 2.1 in Sutradhar and Ali (1989)].

Distribution of the Sample Mean Vector $\bar{Y}$ Under Elliptical $t$ Distribution

Note that the $np$-dimensional joint $t$ distribution of $y_1,\ldots,y_i,\ldots,y_n$ is given by (35) which is a special case of the joint elliptical distribution (38). By the same token, the $np$-dimensional $t$ distribution of $y^* = [y_1',\ldots,y_i',\ldots,y_n']'$ given by (37) is a special case of the elliptical distribution given in (39). Now because the elliptical distribution of $x^* = D^*y^* = \bar{y}$ in (44) is derived from (39), by similar calculation, one would obtain the $t$ distribution for $x^* = \bar{y}$ from (37). Thus, when $y_1,\ldots,y_i,\ldots,y_n$ have the joint $t$ distribution as in (35), $\bar{y}$ follows the $t_p$ distribution, namely $\bar{y}\sim t_p(\mu,\frac{1}{n}\Lambda,\nu)$, that is,

$$
f(\bar{y}) = c^*(\nu_0,p)|\frac{1}{n}\Lambda|^{-\frac{1}{2}}\left[\nu+\{(\bar{y}-\mu)'(\frac{1}{n}\Lambda)^{-1}(\bar{y}-\mu)\}\right]^{-\frac{\nu+p}{2}}. \qquad (45)
$$

### 2.2.3 Distribution of the Sample Covariance Matrix Under Elliptical Distribution and Its Special Form Under Elliptical *t* Distribution

Distribution of $A=\sum_{j=1}^{n}(y_j-\bar{y})(y_j-\bar{y})'$ Under Joint Elliptical Distribution (38)

The distribution of this sample sum of products matrix $A=\sum_{j=1}^{n}(y_j-\bar{y})(y_j-\bar{y})'$ under the general elliptical distribution (38) was derived by Sutradhar and Ali (1986, Theorem 2.1). This distribution has the form

$$
\begin{aligned}
f(A) &= K(n-1,p)g_{(n-1,p)}(\mathrm{trace}\Lambda^{-1}A)\\
&\quad\times|\Lambda|^{-(n-1)/2}|A|^{((n-1)-p-1)/2}, \qquad (46)
\end{aligned}
$$

where $K(n-1,p)$ is the normalising constant and $g_{(n-1,p)}(\cdot)$ is determined by the form of $g$ in (38) and by the number of components in $\sum_{j=1}^{n}y_jy_j'$. For a proof of this result, we refer to the Theorems 2.2 and 2.3 in Sutradhar and Ali (1989, pp. 157–159), among others.

Distribution of $A = \sum_{j=1}^{n}(y_j - \bar{y})(y_j - \bar{y})'$ Under Joint Elliptical $t$ Distribution (35)

By matching the form of $g(\cdot)$ between (35) and (38), and using $n_1 = n - 1$, the pdf of the elements of $A$ under (35) follows from (46) and it is given by

$$f(A) = K(n_1, p)|\Lambda|^{-\frac{n_1}{2}}|A|^{-\frac{n_1-p-1}{2}}[\nu + tr\Lambda^{-1}A]^{-\frac{\nu+n_1p}{2}} \tag{47}$$

[see also Sutradhar and Ali (1989, JMVA, Eq. (3.3))] where

$$K(n_1, p) = [\nu^{\frac{\nu}{2}}\Gamma\{(\nu + n_1p)/2\}]/\left[\pi^{\frac{p(p-1)}{4}}\Gamma(\nu/2)\Pi_{i=1}^{p}\Gamma\{(n_1 - i + 1)/2\}\right].$$

Further Special Case When Sample Follows Multi-Normal Distribution

When

$$Y_i \overset{iid}{\sim} N_p(\mu, \Lambda),$$

the joint distribution of $y_1, \ldots, y_i, \ldots, y_n$ has the form

$$f(y_1, \ldots, y_i, \ldots, y_n) = \frac{1}{\sqrt{2\pi}^{np}}|\Lambda|^{-\frac{n}{2}}\exp\{-\frac{1}{2}\sum_{i=1}^{n}(y_i - \mu)'\Lambda^{-1}(y_i - \mu)\}. \tag{48}$$

Now either matching the form of the $g(\cdot)$ function between (48) and (38) or by taking the limit $\nu \to \infty$, the distribution of $A$ under normality (48) can be easily derived from (46) or (47), respectively. The distribution has the form

$$f(A) = \frac{1}{2^{\frac{n_1p}{2}}\Gamma_p(\frac{n_1}{2})|\Lambda|^{\frac{n_1}{2}}}\exp(-\frac{1}{2}\text{trace}\Lambda^{-1}A)|A|^{\frac{n_1-p-1}{2}}, \tag{49}$$

[see also Muirhead (1982, Theorem 3.2.1)] where

$$\Gamma_p(\frac{n_1}{2}) = \pi^{p(p-1)/4}\Pi_{j=1}^{p}\Gamma[\frac{1}{2}(n_1 + 1 - j)].$$

# 3 Parameter Estimation Difficulty Using Uncorrelated *t* Sample

## *3.1 Likelihood Estimator of Mean μ and Covariance Matrix Σ Under Elliptical Model* (38)

For elliptically contoured distribution (ECD) (38) for $n$ uncorrelated responses, Anderson et al. (1986, Theorem 1, p. 56) computed the MLE (maximum likelihood estimator) of $\mu$ and $\Sigma = \nu_0 \Lambda$ as

$$\hat{\mu}_{\text{MLE}} = \bar{Y}, \;\; \hat{\Sigma}_{\text{MLE}} = (np/d^*)S \tag{50}$$

where

$$S = \frac{A}{n} = \sum_{j=1}^{n} (Y_j - \bar{Y})(Y_j - \bar{Y})'/n,$$

and $d^*$ is a finite positive maximum of $d^{\frac{np}{2}} g^*_{np}(d)$, with $g^*_{np}(d)$ as the spherical joint density of $y_1, \ldots, y_j, \ldots, y_n$. For example,
**Normal case:** Under normality

$$g^*_{np}(d) = (2\pi)^{-\frac{np}{2}} \exp(-d/2),$$

and $t$ **case:** Under $t$ model with $\nu$ d.f.,

$$g^*_{np}(d) = \left[\nu^{\frac{\nu}{2}} \Gamma\{(\nu + np)/2\}/\pi^{\frac{np}{2}} \Gamma(\nu/2)\right] \{\nu + d\}^{-\frac{\nu+np}{2}}.$$

### 3.1.1 $\hat{\mu}_{\text{MLE}}$ is Consistent for $\mu$ Under ECD *t* Model

Under the ECD $t$ model (35), the distribution of $\bar{Y}$ is given as

$$\bar{Y} \sim t_p(\mu, \Lambda/n, \nu).$$

Thus, the consistency of $\bar{Y}$ follows from

$$E[\bar{Y}] = \mu, \;\; \text{Lt}_{n\to\infty}\text{var}[\bar{Y}] = \text{Lt}_{n\to\infty}[\frac{\nu_0 \Lambda}{n}] \to 0, \tag{51}$$

where $\nu_0 = \frac{\nu}{\nu - 2}$.

### 3.1.2 $\hat{\Sigma}_{\text{MLE}}$ is Inconsistent for $\Sigma$ Under ECD $t$ Model

This inconsistency is discussed in details by Sutradhar (2004, Sect. 2). Notice that under the ECD $t$ model (35), it follows from (50) that Under (9) it follows that

$$\hat{\Sigma}_{\text{MLE}} = (np/d^*)S = \nu_0 S = \frac{\nu}{\nu-2}S,$$

where $S = \frac{A}{n}$. However, it follows from Sutradhar and Ali (1989, p. 161) that for $\Sigma^{\frac{1}{2}} = (m_{uh})$, the variances of the elements of $S$ are given by

$$\begin{aligned}\text{var}(S_{uv}) = n^{-2}(\nu-2)/(\nu-4)\Bigg[(n-1)^2\left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 \\ + 2(n-1)\sum_{h=1}^{p} m_{uh}^2 m_{vh}^2 + (n-1)\sum_{h<\ell}(m_{uh}m_{v\ell} + m_{u\ell}m_{vh})^2\Bigg] \\ -(1-\frac{1}{n})^2\left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2. \qquad (52)\end{aligned}$$

It is then clear from (52) that as $n \to \infty$, the variance of the $(u, v)$th element of the $S$ matrix approaches to

$$\begin{aligned}Lt_{n\to\infty}\text{var}(S_{uv}) &= \left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 \{(\nu-2)/(\nu-4) - 1\} \\ &= \{2/(\nu-4)\}\left(\sum_{h=1}^{p} m_{uh}m_{vh}\right)^2 \\ &= [2\nu^2/\{(\nu-2)^2(\nu-4)\}]\left(\sum_{h=1}^{p} \lambda_{uh}\lambda_{vh}\right)^2, \qquad (53)\end{aligned}$$

where $\lambda_{uh}$ is the $(u, h)$th element of the $\Lambda^{\frac{1}{2}}$ matrix. It is then clear that the limiting variances (and also covariances) are free from $n$, and they approach to certain finite quantities based on the elements of the $\Lambda$ matrix and the d.f. of the $t$-distribution. Consequently, the $\hat{\Sigma}_{\text{MLE}}$ derived in Anderson et al. (1986) is not a consistent estimator for $\Sigma$ under the elliptic $t$ model (35).

Remark that as the sample covariance matrix $S$ is not consistent for $\Sigma$, the James-Stein and Stein's orthogonality invariant estimators (of $\Sigma$) constructed in Kubokawa and Srivastava (1977) are in fact function of the elements of the inconsistent sample covariance matrix. The inconsistency of $S$ for $\Sigma$ also implies that there may not exist any consistent estimators for the shape parameter $\nu$ or the kurtosis parameter $\kappa(\nu)$ under the ECD $t$ distribution (35).

### Understanding Inconsistency of $S$ for Simpler Univariate Case

Consider the univariate response case where $p = 1$. Also consider a general regression case with $E[Y_i] = x_i'\beta$ where $x_i$ is a $q \times 1$ covariate vector and $\beta$ is the $q$-dimensional regression effect, whereas in the joint $t$ model (35) $E[Y_i] = \mu$ is a constant scalar quantity. Next by using $\Lambda = \lambda^2$ for the univariate case, the uncorrelated sample based joint pdf (35) reduces to $n$-dimensional $t$ distribution as

$$f(y_1, \ldots, y_n) = c(\nu, n)(\lambda^2)^{-\frac{n}{2}}[\nu + \sum_{j=1}^{n}\{(\lambda^2)^{-1}(y_j - x_j'\beta)^2\}]^{-\frac{\nu+n}{2}}. \tag{54}$$

The objective of this section is to examine whether $s^2 = \frac{1}{n}\sum_{j=1}^{n}(y_j - x_j'\beta)^2$ is consistent estimator for $\sigma^2 = \frac{\nu}{\nu-2}\lambda^2$.

Suppose that $d_j = (y_j - x_j'\beta)$. It then follows that $d_j \sim t(0, \Lambda = \lambda^2, \nu)$. Consequently,

$$\begin{aligned}\mu_2 = Ed_j^2 &= E[Y_j - x_j'\beta]^2 \\ &= \frac{\nu}{\nu-2}\lambda^2,\end{aligned} \tag{55}$$

yielding $E[s^2] = \frac{1}{n}[n\nu\lambda^2/(\nu-2)] = \frac{\nu}{\nu-2}\lambda^2$. So $s^2$ is unbiased for $\sigma^2 = \frac{\nu}{\nu-2}\lambda^2$.

Next we compute $\text{var}(s^2)$ as follows.

$$\begin{aligned}\text{var}[s^2] &= \frac{1}{n^2}\left[\sum_{j=1}^{n}\text{var}[d_j^2] + \sum_{j\neq k}^{n}\text{cov}[d_j^2, d_k^2]\right] \\ &= \frac{1}{n^2}\left[\sum_{j=1}^{n}\{Ed_j^4 - (Ed_j^2)^2\} + \sum_{j\neq k}^{n}\{E[d_j^2 d_k^2] - Ed_j^2 Ed_k^2\}\right],\end{aligned} \tag{56}$$

because $d_j$ and $d_k$ are uncorrelated ($d_j^2$ and $d_k^2$ are correlated) but not independent. Now under $t$ distribution with $\lambda^2 = 1$, say, one writes

$$E[d_j^2] = \frac{\nu}{\nu-2},\ \mu_4 = E[d_j^4] = \frac{3\nu^2}{(\nu-2)(\nu-4)},$$

and

$$E[d_j^2 d_k^2] = \mu_{2,2} = \frac{\nu^2}{(\nu-2)(\nu-4)}.$$

It then follows from (56) that

$$
\begin{aligned}
Lt_{n\to\infty}\text{var}[s^2] &= E[d_j^2 d_k^2] - Ed_j^2 Ed_k^2 \\
&= [\frac{\nu^2}{(\nu-2)(\nu-4)} - [\frac{\nu}{\nu-2}]^2] \\
&= \frac{\nu^2}{\nu-2}[\frac{1}{\nu-4} - \frac{1}{\nu-2}] \\
&= 2[\frac{\nu}{\nu-2}]^2 \frac{1}{\nu-4},
\end{aligned} \tag{57}
$$

a constant function of $\nu$, which is zero under the normal case only. Thus $s^2$ is not mean squared error consistent for $\sigma^2 = \frac{\nu}{\nu-2}\lambda^2$.

Understanding Inconsistency of $s^2$ in a Simpler Alternative Way

Take $\lambda^2 = 1$. In the normal case, for known $\beta$,

$$
E[s^2] = E[\frac{1}{n}\sum_{j=1}^{n}(y_j - x_j'\beta)^2] = \frac{1}{n}E[\chi_n^2] = \frac{1}{n}n = 1
$$

$$
\text{var}[s^2] = \frac{1}{n^2}\text{var}[\chi_n^2] = \frac{1}{n^2}2n = \frac{2}{n} \to 0, \text{as } n \to \infty.
$$

But, if $y_j \sim t(x_j'\beta, 1, \nu)$, then $(y_j - x_j'\beta)^2 \sim F_{1,\nu}$. It implies that

$$
E[s^2] = E[\frac{1}{n}\sum_{j=1}^{n}(y_j - x_j'\beta)^2] = \frac{1}{n}nE[F_{1,\nu}] = \frac{1}{n}n\frac{\nu}{\nu-2} = \frac{\nu}{\nu-2}.
$$

Hence $s^2$ is an unbiased estimator of $\sigma^2 = \frac{\nu}{\nu-2}$.

Next we compute the variance of $s^2$ as follows.

$$
\begin{aligned}
\text{var}[s^2] &= \frac{1}{n^2}\text{var}[\sum_{j=1}^{n}(y_j - x_j'\beta)^2]] \\
&= \frac{1}{n^2}\left[n\text{var}\{(y_j - x_j'\beta)^2\}\right. \\
&\qquad \left. + n(n-1)\text{cov}\{(y_j - x_j'\beta)^2, (y_k - x_k'\beta)^2\}\right] \\
&= \frac{1}{n}[\frac{2\nu^2(\nu = 1-2)}{1(\nu-2)^2(\nu-4)}] + (1-\frac{1}{n})c(\nu)
\end{aligned}
$$

$$= c(\nu) + \frac{1}{n}(v(\nu) - c(\nu))$$
$$\to c(\nu), \text{as } n \to \infty.$$

Thus, $s^2$ is not consistent for $\sigma^2 = \frac{\nu}{\nu-2}$.

## 3.2 $\hat{\nu}_{MLE}$ Does Not Exist

For convenience, without any loss of generality, consider $\mu = 0$ and $\Lambda = I_p$ in (35), and attempt to maximize the likelihood function

$$L(\nu) = c(\nu, np)\left\{1 + \sum_{j=1}^{n} y_j' y_j/\nu\right\}^{-\frac{\nu+np}{2}}, \tag{58}$$

with respect to $\nu$. Note that in (58),

$$c(\nu, np) = \left[\nu^{\frac{\nu}{2}}\Gamma\{(\nu+np)/2\}/\pi^{\frac{np}{2}}\Gamma(\nu/2)\right],$$

which is an increasing function of $\nu$ for fixed $n$ and $p$. Furthermore, as $\sum_{j=1}^{n} y_j' y_j$ and $np$ are fixed for a given data set, it also follows that the spherical function in (58) is an increasing function of $\nu$. Thus, it is intuitively clear that the joint density function in (58) is maximized at $\nu = \infty$. Consequently there does not exist any MLE for $\nu$ for the ECD $t$ distribution (35) or spherical $t$ distribution (58), as $\nu = \infty$ is the normal case.

### 3.2.1 Moment Estimator of $\nu$, Say $\hat{\nu}_M$ Is Not Consistent for $\nu$

The moment estimator of $\nu$ has the same formula (23) as under the independent case. The estimator is given by

$$\hat{\nu}_M = \frac{2[2\hat{\beta}^*_{2,M} - f^*(s)]}{[\hat{\beta}^*_{2,M} - f(s)]} \tag{59}$$

where for $S = \frac{A}{n-1} = \sum_{j=1}^{n}(Y_j - \bar{Y})(Y_j - \bar{Y})'/(n-1)$,

$$\hat{\beta}^*_{2,M} = \frac{1}{n}\sum_{j=1}^{n}[(y_j - \bar{y})' S^{-1}(y_j - \bar{y})]^2, \tag{60}$$

and for $S = (s_{hu}),\ S^{-1} = (s^{hu})$,

$$f(s) = \left[ 3 \sum_{h=1}^{p} (s^{hh})^2 (s_{hh})^2 \right. \\ \left. + \sum_{h \neq h'}^{p} (s_{h'h'})^2 \{ s^{hh} s^{h'h'} + (s^{hh'})^2 \} \right]. \tag{61}$$

Now recall from (19) and (20) that

$$\nu = \frac{2[2\beta_2^* - f^*(\sigma)]}{[\beta_2^* - f^*(\sigma)]}, \tag{62}$$

where $\beta_2^*$ and $f^*(\sigma)$ are given by (19) and (21), respectively. Next because under the ECD $t$ model it is shown in Sect. 3.1.2 that $S$ is inconsistent for $\Sigma$, it then follows that $f^*(s)$ and $\hat{\beta}_{2,M}^*$ are bound to be inconsistent for $f^*(\sigma)$ and $\beta_2^*$, respectively. Consequently, $\hat{\nu}_M$ in (59) would be inconsistent for $\nu$ defined by (62).

## 4 Estimation of Parameters for Clustered (Familial or Longitudinal) Regression Models with *t* Data

It was demonstrated in Sect. 3 that when $n$ uncorrelated multi-dimensional ($p$-dimensional) $t$ responses are collected, they can not be used for consistent estimation for the scale (covariance matrix) and shape (degrees of freedom) parameters of the distribution. However, as discussed in Sect. 2.1, this estimation difficulty does not arise for the cases when the $n$ responses are independent. Now because there are situations in practice such as in a clustered regression setup where $n$ uncorrelated or correlated observations (say $p$-dimensional) are collected from $n$ members of a large number of independent groups/families (say $K$ clusters), one may therefore be able to combine the aforementioned two models (uncorrelated responses from independent clusters) to form a clustered regression model in order to analyze such clustered $t$ data. The purpose of this section is to deal with three clustered models (CM), namely,

**CM1. Elliptical *t* model for uncorrelated familial data:** Under this model, multi-dimensional responses (those follow a multi-variate $t$ distribution) along with a set of multidimensional covariates will be collected from the members of a large number of independent clusters/groups (such as foot ball teams), where the responses from the members in a group may be uncorrelated.

**CM2. Elliptical *t* model for correlated familial data:** This model is similar to CM1 but the responses from the members of a family (such as a family with parents, brother and sister) will be correlated.

**CM3. Elliptical *t* model for correlated longitudinal data:** Under this model, a univariate or multivariate *t* responses are repeatedly collected over a small period of time from a large number of independent individuals.

These three models along with suitable estimation techniques for the consistent estimation of the parameters of these models are discussed in the following three subsections.

## 4.1 Elliptical t Model for Uncorrelated Clustered (Familial) Responses

Let there be $K$ independent national football teams with $n$ members selected from each team. Suppose that a $p$ dimensional response from each member is collected, and $y_{ij} = [y_{ij1}, \ldots, y_{iju}, \ldots, y_{ijp}]'$ denotes this response from the $j$th ($j = 1, \ldots, n$) member of the $i$th ($i = 1, \ldots, K$) team/family/cluster. Also suppose that a $c$ dimensional covariate vector is collected from each member and $x_{iju} = [x_{iju1}, \ldots, x_{ijuc}]'$ represents this $c$-dimensional covariate vector. For convenience, construct the covariate matrix $x_{ij} : p \times pc$; and the $pc$-dimensional vector of regression effects $\beta = [\beta_1', \ldots, \beta_u', \ldots, \beta_p']'$, with $\beta_u = [\beta_{u1}, \ldots, \beta_{uc}]'$.

We now construct the elliptical $t$ model for the uncorrelated responses from the members of the $i$th team/cluster, as

$$
\begin{aligned}
f(y_{i1}', \ldots, y_{ij}', \ldots, y_{in}') &= c^*(\nu, n, p)|\Lambda|^{-\frac{n}{2}} \\
&\times[\nu + \sum_{j=1}^{n}\{(y_{ij} - x_{ij}\beta)'\Lambda^{-1}(y_{ij} - x_{ij}\beta)\}]^{-\frac{\nu+np}{2}} \qquad (63) \\
&\equiv c^{**}(\nu, n, p)|\Sigma|^{-\frac{n}{2}} \\
&\times\left[1 + (\nu-2)^{-1}\sum_{j=1}^{n}\{(y_{ij} - x_{ij}\beta)'\Sigma^{-1}(y_{ij} - x_{ij}\beta)\}\right]^{-\frac{\nu+np}{2}}, \qquad (64)
\end{aligned}
$$

where $\text{cov}[Y_{ij}] = \frac{\nu}{\nu-2}\Lambda = \Sigma$.

It is of interest to obtain the consistent estimates for the parameters, namely $\beta$, $\Sigma$, and $\nu$.

### 4.1.1 Consistent Estimator of $\boldsymbol{\Sigma}$

In this section, for known $\beta$ we construct a moment estimator of *Sigma* and show that this estimator is consistent.

Notice from (64) that $\Sigma$ is a common covariance matrix for $y_{ij}$ for all $i = 1, \ldots, K$, and $j = 1, \ldots, n$. Thus, we construct a moment estimator for $\Sigma$ as

$$\hat{\Sigma}_M = (s^*_{uv}) = \frac{1}{K}\sum_{i=1}^{K} S_i = \frac{1}{K}\sum_{i=1}^{K}[\sum_{j=1}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ij} - x_{ij}\beta)'\}/n], \tag{65}$$

where $x_{ij} = \begin{bmatrix} x'_{ij1} & 0' & \ldots & 0' \\ 0' & x'_{ij2} & \ldots & 0' \\ \vdots & \vdots & & \vdots \\ 0' & 0' & \ldots & x'_{ijp} \end{bmatrix}'$. Now to show that $\hat{\Sigma}_M$, one needs to show that

$$Lt_{K\to\infty}\text{cov}(s^*_{uv}, s^*_{h\ell}) = 0. \tag{66}$$

For the purpose, we consider two general elements of $S_i$ as

$$s_{i,uv} = \sum_{j=1}^{n}\{(y_{iju} - x'_{iju}\beta_u)(y_{ijv} - x'_{ijv}\beta_v)'\}/n \tag{67}$$

$$s_{i,h\ell} = \sum_{j=1}^{n}\{(y_{ijh} - x'_{ijh}\beta_h)(y_{ij\ell} - x'_{ij\ell}\beta_\ell)'\}/n, \tag{68}$$

Next similar to (27), by using Muirhead (1982, Eq. (3), p. 42; Exercise 1.33(b), p. 49), and also by using the notation $\kappa = \frac{2}{\nu-4}$ as the kurtosis parameter of the $t$ distribution, and $\Sigma = (\sigma_{uv}) : p \times p$, we obtain

$$\text{cov}[s_{i,uv}, s_{i,h\ell}] = \frac{1}{n^2}[n(\kappa+1)\{\sigma_{uv}\sigma_{h\ell} + \sigma_{uh}\sigma_{v\ell} + \sigma_{u\ell}\sigma_{vh}\} + n(n-1)(\kappa+1)\sigma_{uv}\sigma_{h\ell}] - \sigma_{uv}\sigma_{h\ell}, \tag{69}$$

yielding (for all $p^2 \times p^2$ elements)

$$Lt_{K\to\infty}\text{cov}(s^*_{uv}, s^*_{h\ell}) = Lt_{K\to\infty}\frac{1}{K^2}[K\kappa\sigma_{uv}\sigma_{h\ell} + K(\kappa+1)\{\sigma_{uh}\sigma_{v\ell} + \sigma_{u\ell}\sigma_{vh}\}/n] = 0, \tag{70}$$

which is (66). Thus, $\hat{\Sigma}_M = (s^*_{uv}) = \frac{1}{K}\sum_{i=1}^{K} S_i$ is a consistent estimator of $\Sigma$.

### 4.1.2 Consistent Estimator of the Kurtosis Parameter $\kappa$

It is of interest to obtain a consistent estimator for the kurtosis parameter $\kappa = 2/(\nu - 4)$, where $\nu$ is the common shape parameter of the distribution of $y_{ij}$ for all $i = 1, \ldots, K$ and $j = 1, \ldots, n$. We will derive a method of moments estimator to achieve this goal.

For the purpose, we first compute the formula for a measure of kurtosis, say $\beta^*_{2i}$ (Mardia 1970) using the data from the $i$th cluster/team/group/family. Note that the responses from the $i$th cluster is represented by

$$y_i = [y'_{i1}, \ldots, y'_{ij}, \ldots, y'_{in}]' : np \times 1,$$

and this response vector has the $np$-dimensional elliptical $t$ distribution given by

$$f(y_i) = c^{**}(\nu, n, p)|I_n \otimes \Sigma|^{-\frac{1}{2}} \times \left[1 + (\nu - 2)^{-1}\{(y_i - X_i\beta)'[I_n \otimes \Sigma^{-1}](y_i - X_i\beta)\}\right]^{-\frac{\nu+np}{2}}, \tag{71}$$

where $X_i = \begin{bmatrix} x_{i1} : p \times pc \\ \vdots \\ x_{ij} : p \times pc \\ \vdots \\ x_{in} : p \times pc \end{bmatrix}' : np \times pc.$

Following Mardia (1970) one may then write the formula for the so-called measure of kurtosis as

$$\beta^*_{2i} = E[(Y_i - X_i\beta)'(I_n \otimes \Sigma^{-1})(Y_i - X_i\beta)]^2. \tag{72}$$

Next, using a transformation from the elliptic $t$ to a spherical $t$, namely

$$Z_i = (I_n \otimes \Sigma^{-\frac{1}{2}})(Y_i - X_i\beta),$$

$\beta^*_{2i}$ in (72) reduces to

$$\beta^*_{2i} = E[Z'_i Z_i]^2 = E[\sum_{h=1}^{np} Z^2_{ih}]^2 = E[\sum_{h=1}^{np} Z^4_{ih} + \sum_{h \neq \ell}^{np} Z^2_{ih} Z^2_{i\ell}]. \tag{73}$$

Now because

$$E[Z^4_{ih}] = 3(\kappa + 1) \text{ and } E[Z^2_{ih}] = 1, \tag{74}$$

(Muirhead 1982, p. 41), one obtains $\beta_{2i}$ by (73) as

$$\begin{aligned}\beta^*_{2i} &= 3np(\kappa + 1) + np(np - 1)\\ &= np[3\kappa + np + 2] = \beta^*_2, \text{ say.}\end{aligned} \tag{75}$$

Recall that $\hat{\Sigma}_M$ (65) is a consistent estimator of $\Sigma$. Now, as $y_1, \ldots, y_i, \ldots, y_K$ are independent elliptical observations, it follows that

$$\hat{\beta}^*_2 = \frac{1}{K}\sum_{i=1}^{K}[(Y_i - X_i\beta)'(I_n \otimes \hat{\Sigma}_M^{-1})(Y_i - X_i\beta)]^2 \tag{76}$$

converges in probability to $\beta^*_2 = np(3\kappa + np + 2)$ defined in (75), yielding the unbiased estimator

$$\hat{\kappa} = \frac{1}{3np}\left[\hat{\beta}^*_2 - n^2p^2 - 2np\right], \tag{77}$$

for $\kappa$. This is a consistent estimator under some mild condition on the variance of the estimator.

### 4.1.3 Consistent Estimator of the Kurtosis Parameter $\beta$

In last two sections, we derived the formulas for the consistent estimators of $\Sigma$ and $\nu$, under the assumption that $\beta$ is known. We now obtain a consistent estimator for this regression parameter vector.

GLS Estimate of $\beta$

The GLS (generalized least squared) estimate of $\beta$ is obtained by solving the estimating equation

$$\sum_{i=1}^{K} X_i'(I_n \otimes \Sigma^{-1})(y_i - X_i\beta) = 0, \tag{78}$$

which has the formula given by

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i'(I_n \otimes \Sigma^{-1})X_i\right]^{-1} \sum_{i=1}^{K} X_i'(I_n \otimes \Sigma^{-1})y_i \tag{79}$$

$$= (\sum_{i=1}^{K} P_i)^{-1} \sum_{i=1}^{K} Q_i y_i, \text{ (say)}. \tag{80}$$

By using the Lindeberg-Feller central limit theorem (Amemiya 1985, Theorem 3.3.6), this GLS estimator can be shown consistent for $\beta$.

### ML (Maximum Likelihood) Estimate of $\beta$

The likelihood function by (71) has the form

$$\begin{aligned} L(\beta) &= \tilde{c}\Pi_{i=1}^{K} \left[1 + (\nu-2)^{-1}\left\{(y_i - X_i\beta)'[I_n \otimes \Sigma^{-1}]\right.\right. \\ &\times \left.\left.(y_i - X_i\beta)\right\}\right]^{-\frac{\nu+np}{2}} \\ &= \tilde{c}\Pi_{i=1}^{K}\left[q_i^{-\frac{\nu+np}{2}}\right], \end{aligned} \tag{81}$$

yielding the log likelihood equation for $\beta$ as

$$\frac{\partial \log L}{\partial \beta} = \frac{\nu + np}{2} \sum_{i=1}^{K} \frac{\partial}{\partial \beta} \log q_i = 0. \tag{82}$$

The ML estimate, the solution of (82), is then given by

$$\hat{\beta}_{ML} = \left(\sum_{i=1}^{K} q_i^{-1} P_i\right)^{-1} \sum_{i=1}^{K} q_i^{-1} Q_i y_i. \tag{83}$$

### An Efficiency Comparison Between the GLS and ML Estimates of $\beta$

Both $\hat{\beta}_{GLS}$ and $\hat{\beta}_{ML}$ are unbiased for $\beta$. However, it can be shown that the asymptotic relative efficiency of $\hat{\beta}_{ML}$ to $\hat{\beta}_{GLS}$ is: $\frac{\nu}{\nu-2}$. This is because,

$$V_2 = \text{cov}[\hat{\beta}_{GLS}] = \left(\sum_{i=1}^{K} P_i\right)^{-1}, \tag{84}$$

and

$$V_1 = \text{cov}[\hat{\beta}_{ML}] = -E\left[\frac{\partial^2 \log L}{\partial\beta\partial\beta'}\right]$$

$$= \frac{\nu(\nu + np)}{(\nu - 2)(\nu + np + 2)} \sum_{i=1}^{K} P_i. \tag{85}$$

Hence for large $n$ so that $[\nu + np] \approx [\nu + np + 2]$, $V_1^{-1} V_2 = \frac{\nu}{\nu - 2} I_{pc}$, showing the aforementioned efficiency.

Thus, for small $\nu$, the maximum likelihood estimator of $\beta$ will be highly more efficient than the generalized least squared estimate. For large $\nu$, $\hat{\beta}_{GLS}$ and $\hat{\beta}_{ML}$ are identical, which is obvious.

## *4.2 Elliptical t Model for Correlated Clustered (Familial) Responses*

Recall from (64) or more specifically from (71) that the linear regression for the uncorrelated clustered data has the form

$$y_i = X_i\beta + \epsilon_i, \ \epsilon_i \sim t(0, I_n \otimes \Sigma, \nu).$$

However it can happen that the multivariate $t$ responses are collected from $n$ members of a family (household unit) instead of a foot ball team. The $n$ responses in such cases will be correlated. This type of multivariate familial data may be modeled as

$$y_i = X_i\beta + 1_{np}\gamma_i + \epsilon_i, \tag{86}$$

where $\gamma_i$ is the $i$th family effect common to the $p$-dimensional responses of all $n$ members. Further assume that $\gamma_i \overset{iid}{\sim} (0, \sigma_\gamma^2)$ and $\gamma_i$ and $\epsilon_i$ are independent. Then, for $U_k$ as an $k \times k$ unit matrix, one obtains

$$\begin{aligned} \text{cov}[Y_i] &= \sigma_\gamma^2 U_{np} + [I_n \otimes \Sigma] \\ &= \Sigma^* \text{ (say)}, \end{aligned} \tag{87}$$

yielding the variance-covariance breakdown for all $n$ members as

$$\text{cov}[Y_{ij}] = \sigma_\gamma^2 U_p + \Sigma, \text{for } j = 1, \ldots, n \tag{88}$$

$$\text{cov}[Y_{ij}, Y_{ik}] = \sigma_\gamma^2 U_p, \ \text{ for } j \neq k = 1, \ldots, n. \tag{89}$$

Thus the pair-wise members are not uncorrelated under the present familial model.

### 4.2.1 Estimation of the Regression Effects $\boldsymbol{\beta}$

Following (78)–(79), the GLS estimate for $\beta$ under the present model has the formula

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i'(\Sigma^{*-1})X_i\right]^{-1} \sum_{i=1}^{K} X_i'(\Sigma^{*-1})y_i \tag{90}$$

$$= (\sum_{i=1}^{K} P_i^*)^{-1} \sum_{i=1}^{K} Q_i^* y_i, \text{ (say)}. \tag{91}$$

Notice that $\Sigma^{*-1}$ in (90) may be directly computed as

$$\Sigma^{*-1} = [I_n \otimes \Sigma^{-1}] - \sigma_\gamma^2 \left[\frac{[I_n \otimes \Sigma^{-1}]U_{np}[I_n \otimes \Sigma^{-1}]}{1 + \sigma_\gamma^2 1_{np}'[I_n \otimes \Sigma^{-1}]1_{np}}\right]. \tag{92}$$

Also it follows that the GLS estimator in (91) has the covariance matrix given by

$$\text{cov}[\hat{\beta}_{GLS}] = (\sum_{i=1}^{K} P_i^*)^{-1}. \tag{93}$$

### 4.2.2 Estimation of the Kurtosis Parameter $\boldsymbol{\kappa}$

By similar calculations as in Sect. 4.1.2, we obtain the formula for the moment estimator of $\kappa$ as

$$\hat{\kappa} = \frac{1}{3np}\left[\hat{\hat{\beta}}_2 - n^2p^2 - 2np\right], \tag{94}$$

where the sample measure of kurtosis $\hat{\hat{\beta}}_2$ is computed by

$$\hat{\hat{\beta}}_2 = \frac{1}{K}\sum_{i=1}^{K}[(Y_i - X_i\hat{\beta})'\{\hat{\Sigma^*}^{-1}\}(Y_i - X_i\hat{\beta})]^2. \tag{95}$$

Notice that the formula for $\hat{\hat{\beta}}_2$ in (95) is similar to that of $\hat{\beta^*}_2$ in (76). They are, however, different. This is because the var$(Y_i)$ has the form $I_n \otimes \Sigma$ in $\hat{\beta^*}_2$ in (76), whereas it has a different form, namely $\Sigma^* = \sigma_\gamma^2 U_{np} + [I_n \otimes \Sigma]$, in (95).

### 4.2.3 Estimation of the Variance Component Parameter $\sigma_\gamma^2$

First we write the moment estimate for $\sigma_\gamma^2 U_p + \Sigma$ as

$$\widehat{\sigma_\gamma^2 U_p + \Sigma} = \frac{1}{K}\sum_{i=1}^{K} S_i$$

$$= \frac{1}{K}\sum_{i=1}^{K}[\sum_{j=1}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ij} - x_{ij}\beta)'\}/n]. \tag{96}$$

where $x_{ij}$ is given in (65). However, notice from (88) and (89) that

$$\text{cov}[Y_{ij}] - \text{cov}[Y_{ij}, Y_{ik}] = \Sigma, \text{ and} \tag{97}$$

$$\text{cov}[Y_{ij}] + \text{cov}[Y_{ij}, Y_{ik}] = 2\sigma_\gamma^2 U_p + \Sigma. \tag{98}$$

Hence, one obtains

$$\widehat{2\sigma_\gamma^2 U_p + \Sigma} = \frac{1}{K}\sum_{i=1}^{K}[\sum_{j=1}^{n}\sum_{k=1}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ik} - x_{ik}\beta)'\}/n^2]. \tag{99}$$

and

$$\hat{\Sigma} = \frac{1}{K}\sum_{i=1}^{K}\left[\sum_{j=1}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ij} - x_{ij}\beta)'\}/n \right.$$

$$\left. - \sum_{j\neq k}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ik} - x_{ik}\beta)'\}/\{n(n-1)\}\right]. \tag{100}$$

yielding

$$\widehat{2\sigma_\gamma^2 U_p} = \frac{1}{K}\sum_{i=1}^{K}[\sum_{j=1}^{n}\sum_{k=1}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ik} - x_{ik}\beta)'\}/n^2]$$

$$- \frac{1}{K}\sum_{i=1}^{K}\left[\sum_{j=1}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ij} - x_{ij}\beta)'\}/n \right.$$

$$\left. - \sum_{j\neq k}^{n}\{(y_{ij} - x_{ij}\beta)(y_{ik} - x_{ik}\beta)'\}/\{n(n-1)\}\right] \tag{101}$$

$$= \tilde{S} = (\tilde{s}_{uv}) : p \times p. \tag{102}$$

It then follows that

$$\hat{\sigma}^2_{\gamma} = \frac{1}{2p^2} \sum_{u=1}^{p} \sum_{v=1}^{p} \tilde{s}_{uv}. \tag{103}$$

## 4.3 Longitudinal Elliptical t Model with Correlated Repeated Observations

Suppose that the $i$th individual provides $T$ repeated multivariate ($p$-dimensional) responses. Let

$$y_{iu} = [y_{i1u}, \ldots, y_{itu}, \ldots, y_{iTu}]'$$

denotes this $T$-repeated responses for $u$th variable collected at time $t = 1, \ldots, T$. We write the means of $y_{iu}$ as

$$E[Y_{iu}] = x_{iu}\beta_u, \tag{104}$$

where $x_{iu} = \begin{bmatrix} x'_{i1u} \\ x'_{i2u} \\ \vdots \\ x'_{iTu} \end{bmatrix} : T \times c$, and $\beta_u = [\beta_{u1}, \ldots, \beta_{uc}]'$.

Further suppose that $y_{iu}$ follows a general auto-correlation structure given by

$$\text{corr}(y_{iu}) = \begin{bmatrix} 1 & \rho_1 & \ldots & \rho_{T-1} \\ \rho_1 & 1 & \ldots & \rho_{T-2} \\ \vdots & \vdots & & \vdots \\ \rho_{T-1} & \rho_{T-2} & \ldots & 1 \end{bmatrix} = C_{iu}(\rho), \tag{105}$$

(Sutradhar 2011, Sect. 2.2) for all $u = 1, \ldots, p$. It then follows that

$$\text{cov}[Y_{iu}] = \tilde{\Sigma}_{uu} = \sigma_{uu} C_{iu}(\rho), \tag{106}$$

where $\sigma_{uu}$ is the variance of $y_{itu}$ for all individuals at any time point $t$. Also suppose that the structural covariances are the same irrespective of time points. Thus, we write

$$\begin{aligned} \text{cov}[Y_{iu}, Y'_{iv}] &= \sigma_{uv} U_T \\ &= \tilde{\Sigma}_{uv}. \end{aligned} \tag{107}$$

Next, define $X_i = \begin{bmatrix} x_{i1} & 0 & \dots & 0 \\ 0 & x_{i2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & x_{ip} \end{bmatrix} : pT \times pc$, $\beta = [\beta_1', \dots, \beta_u', \dots, \beta_p']'$. It then follows that

$$Y_i = [y_{i1}', \dots, y_{iu}', \dots, y_{ip}']' \sim t_{pT}(X_i\beta, \tilde{\Sigma}, \nu), \tag{108}$$

where $\tilde{\Sigma} = \begin{bmatrix} \tilde{\Sigma}_{11} & \tilde{\Sigma}_{12} & \dots & \tilde{\Sigma}_{1p} \\ \tilde{\Sigma}_{21} & \tilde{\Sigma}_{22} & \dots & \tilde{\Sigma}_{2p} \\ \vdots & \vdots & & \vdots \\ \tilde{\Sigma}_{p1} & \tilde{\Sigma}_{p2} & \dots & \tilde{\Sigma}_{pp} \end{bmatrix} : Tp \times Tp$. That is,

$$\tilde{\Sigma} = \begin{bmatrix} \sigma_{11}C_{i1}(\rho) & \sigma_{12}U_T & \dots & \sigma_{1p}U_T \\ \sigma_{21}U_T & \sigma_{22}C_{i2}(\rho) & \dots & \sigma_{2p}U_T \\ \vdots & \vdots & & \vdots \\ \sigma_{p1}U_T & \sigma_{p2}U_T & \dots & \sigma_{pp}C_{ip}(\rho) \end{bmatrix}. \tag{109}$$

### 4.3.1 GLS Estimation for $\beta$

Using $\tilde{\Sigma}$ from (109) and following (90)–(91), the GLS estimate of $\beta$ is obtained as

$$\hat{\beta}_{GLS} = \left[\sum_{i=1}^{K} X_i'(\tilde{\Sigma}^{-1})X_i\right]^{-1} \sum_{i=1}^{K} X_i'(\tilde{\Sigma}^{-1})y_i \tag{110}$$

$$= (\sum_{i=1}^{K} \tilde{P}_i)^{-1} \sum_{i=1}^{K} \tilde{Q}_i y_i, \text{ (say)}. \tag{111}$$

### 4.3.2 Moment Estimation for Kurtosis Parameter $\kappa$

Under the present longitudinal model, we first obtain a sample measure of kurtosis as

$$\hat{\beta}_2^{\dagger} = \frac{1}{K} \sum_{i=1}^{K} [(Y_i - X_i\hat{\beta})'(\hat{\tilde{\Sigma}}^{-1})(Y_i - X_i\hat{\beta})]^2 \tag{112}$$

which converges in probability to $\beta_2^{\dagger} = np(3\kappa + np + 2)$, yielding

$$\hat{\kappa} = \frac{1}{3np}\left[\hat{\beta}_2^{\dagger} - n^2p^2 - 2np\right]. \tag{113}$$

This formula is similar to (94) with a difference in the formula for the sample measure of kurtosis.

### 4.3.3 Moment Estimation for Lag $\ell$ Autocorrelation

Using $E[Y_{itu}] = x'_{itu}\beta_u$, and $\text{var}[Y_{itu}] = \sigma_{uu}$, the moment estimate of lag $\ell$ correlation may be computed as

$$\hat{\rho}_\ell = \frac{[\sum_{i=1}^{K}\sum_{u=1}^{p}\sum_{t=1}^{T-\ell}\tilde{y}_{itu}\tilde{y}_{i,t+\ell,u}]/Kp(T-\ell)}{[\sum_{i=1}^{K}\sum_{u=1}^{p}\sum_{t=1}^{T}\tilde{y}_{itu}^2]/KpT}, \tag{114}$$

(Sutradhar 2011, Sect. 2.2.2) where

$$\tilde{y}_{itu} = \frac{y_{itu} - x'_{itu}\beta_u}{\sqrt{\sigma_{uu}}}.$$

### 4.3.4 Moment Estimation for $\Sigma$=($\sigma_{uv}$) : $p \times p$

For $\Sigma$ estimation, re-organize $y_i$ as follows:

$$y_{it} = [y_{it1}, \ldots, y_{itu}, \ldots, y_{itp}]' \;\rightarrow y_i = [y'_{i1}, \ldots, y'_{it}, \ldots, y'_{iT}]'.$$

Then $\tilde{\Sigma}$ in (109) takes the form

$$\text{cov}[Y_i] = \Omega = \begin{bmatrix} \Sigma & \rho\Sigma & \ldots & \rho^{T-1}\Sigma \\ \rho\Sigma & \Sigma & \ldots & \rho^{T-2}\Sigma \\ \vdots & \vdots & & \vdots \\ \rho^{T-1}\Sigma & \rho^{T-2}\Sigma & \ldots & \Sigma \end{bmatrix}$$
$$= C(\rho) \otimes \Sigma. \tag{115}$$

Next write

$$E[Y_{it}] = x_{it}\beta, \text{ with } \beta = [\beta'_1, \ldots, \beta'_u, \ldots, \beta'_p]'.$$

Also, use $x_{it} = \begin{bmatrix} x'_{it1} & 0' & \ldots & 0' \\ 0' & x'_{it2} & \ldots & 0' \\ \vdots & \vdots & & \vdots \\ 0' & 0' & \ldots & x'_{itp} \end{bmatrix} : p \times pc$, and

$$X_i = \begin{bmatrix} x_{i1} : p \times pc \\ \vdots \\ x_{it} : p \times pc \\ \vdots \\ x_{iT} : p \times pc \end{bmatrix} : Tp \times pc.$$

We may then write

$$E[Y_i] = X_i\beta,$$

and following (65), for example, obtain the moment estimate of $\Sigma$ as

$$\hat{\Sigma} = \frac{1}{K}\sum_{i=1}^{K}\left[\sum_{t=1}^{T}\{(y_{it} - x_{it}\beta)(y_{it} - x_{it}\beta)'\}/T\right]. \tag{116}$$

## 5 Testing for Linear Regression in Uncorrelated *t* Models

In Sect. 2.2, we have discussed the distribution theory for various familiar statistics such as marginal properties and linear combination of uncorrelated elliptical $t$ variables. In this section, we continue discussing the distribution theory for some other classical statistics which are often used for testing useful hypothesis in linear regression setup. One of these statistics is the normality based $F$-test statistic used for testing linear hypothesis in terms of regression parameters. We remark here that almost 3 decades ago the null and non-null distributional aspects of the normality based $F$-statistic were discussed by Sutradhar (1988). For the sake of completeness, we discuss below these distribution theories.

### *5.1 Classical F-Statistic Is Null Robust*

Consider the linear regression

$$Y = X\beta + \epsilon, \tag{117}$$

for the purpose of testing the hypothesis

$$H_0 : C\beta = 0 \text{ versus } H_1 : \text{the negation of } H_0,$$

where $Y$ is a $n \times 1$ response variable, $X$ is a known design matrix of order $n \times m$, $\beta$ is a $m \times 1$ vector of unknown parameters, $\epsilon$ is a $n \times 1$ error variable, and $C$ is a $r \times m$ matrix of known coefficients with rank $(C) = q$, say.

Under normality, that is, when $\epsilon \sim N(0, \sigma^2 I_n)$, it is well known that this null hypothesis may be tested by using the classical $F$-statistic

$$F^* = (W_2 - W_1)/W_1, \tag{118}$$

where $W_1 = Y'[I_n - X(X'X)^{-1}X']Y/\sigma^2$ is the residual sum of squares of the full model, and $W_2 = Y'[I_n - Z(Z'Z)^{-1}Z']Y/\sigma^2$ is the residual sum of squares of the reduced model $E[Y] = Z\alpha$ under the $H_0$. It is also well known that under $H_0$, $qF^*/(n-m)$ follows the standard $F$ distribution with degrees of freedom parameters $q$ and $n - m$, that is,

$$qF^*/(n - m) \sim F(q, n - m), \tag{119}$$

whereas under the $H_1$, $qF^*/(n - m)$ follows the non-central $F'$ distribution, that is,

$$qF^*/(n - m) \sim F'(q, n - m;\ \delta^*), \tag{120}$$

with $\delta^*$ as the non-centrality parameter defined as

$$\delta^* = \beta'[X'X - X'Z(Z'Z)^{-1}Z'X]\beta/\sigma^2.$$

Now following (16) for $p = 1$, consider uncorrelated $t$ joint distribution for the errors in (117), that is,

$$f(y_1, \ldots, y_n) = c(\nu, n)(\lambda^2)^{-\frac{n}{2}} \times [\nu + \sum_{j=1}^{n} \{(y_j - x_j'\beta)'(\lambda^2)^{-1}(y_j - x_j'\beta)\}]^{-\frac{\nu+n}{2}}. \tag{121}$$

It was demonstrated by Sutradhar (1988) that when $Y$ follows this $t$ distribution (121), the $F^*$-statistic in (118) still follows the $F$ distribution as in (119) under the $H_0$, but, under the $H_1$, this statistic $F^*$ follows a distribution which is different than the non-central $F$ distribution given by (120). We discuss this non-null distribution in the next section.

## 5.2 Classical F-Statistic Is Not Non-null Robust

To be specific, the distribution of $F^*$ under $H_1$ has the form

$$g(f^*) = \frac{2}{\nu - 2} \sum_{r=0}^{\infty} \beta_{(\delta^*/(\nu-2))} \left(r + 1, \frac{\nu}{2} - 1\right) \beta_{f^*} \left(r + \frac{q}{2}, \frac{n-m}{2}\right), \tag{122}$$

(Sutradhar 1988) where

$$\beta_x(\ell, u) = \frac{\Gamma(\ell + u)x^{\ell-1}}{\Gamma(\ell)\Gamma(u)(1+x)^{\ell+u}},$$

and $\delta^*$ is the non-centrality parameter given by $\delta^* = \beta'[X'X - X'Z(Z'Z)^{-1}Z'X]\beta/\sigma^2$, with $\sigma^2 = \frac{\nu}{\nu-2}\lambda^2 = \text{var}[Y_j]$ for $j = 1, \ldots, n$, and $Z$ is the design matrix of the reduced model.

Note that the distribution of $F^*$ in (122) reduces to the non-central $F'$ distribution (120) for $\nu \to \infty$. Thus, in general, the power of the $F$ test for testing the $H_0$ under the $t$ distribution (121) depends on the degrees of freedom. Thus the test is null robust but not non-null robust.

### 5.2.1 Power Computation

The power of the $F^*$ statistic under the non-null distribution (122) may be computed as

$$P(\nu, \delta^*, \alpha) = \frac{2}{\nu - 2} \sum_{r=0}^{\infty} \beta_{(\delta^*/(\nu-2))} \left(r + 1, \frac{\nu}{2} - 1\right) \times \int_{F_0/(2r+1)}^{\infty} g(f^*; 2r+1, n-m) df^*, \tag{123}$$

where

$$F_0 \equiv F(q, n-m, \alpha), \text{ and } g(f^*; 2r+1, n-m)$$

is the usual central $(n-m)F$ distribution with d.f. $2r+1$, and $n-m$. Next, substitute $(2r+1)f^*/(n-m) = x$ and $1/(1+x) = u$ in (123). It then follows that the power of the test has the formula

$$P(\nu, \delta^*, \alpha) = \frac{2}{\nu - 2} \sum_{r=0}^{\infty} \beta_{(\delta^*/(\nu-2))} I_{u_0}\left(\frac{n-m}{2}, r + \frac{1}{2}\right), \tag{124}$$

where $u_0 = 1/[1 + \{q/(n-m)\}F_0]$ and $I_u(a, b)$ is the well known Karl Pearson (1934) incomplete $\beta$-function.

For a numerical example on power computation, consider, for example, $\beta = (\beta_1, \beta_2)'$, $C = [0, 1]$, $q = \text{rank}(C) = 1$, and $m = 2$; $X = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix}'$. Then for $\delta^* = \beta_1^2 \sum_{j=1}^{n}(x_j - \bar{x})^2/\sigma^2$, using $\phi = (\delta^*/2)^{\frac{1}{2}} = 3.0$, $n = 16$, one obtains the powers by (124) for $\nu \equiv 5, 8, 10, 12, 15, 20, 25$, and $\infty$ as

$$0.87, 0.91, 0.92, 0.94, 0.94, 0.95, \text{ and } 0.97,$$

respectively. It is clear that the $t$ distribution based powers are smaller than the normal based power.

## 6 Concluding Remarks

When a multivariate response follows an elliptic $t$ distribution, the joint distribution of independent multivariate responses do not, however, follow an elliptic $t$ distribution. Unlike the normal distribution theory, this difference in distributions between marginal and joint responses makes the exact distribution theory using independent $t$ samples extremely complex. This difficulty has been discussed in brief in Sect. 2.1. However, when multivariate responses are uncorrelated but dependent, one may construct a joint distribution which has the same form as that of the marginal elliptical $t$ distribution. This provides closed form exact distribution theories for many $t$ samples based statistics. The distribution theories for some of these statistics under uncorrelated elliptical $t$ model are discussed in Sect. 2.2. But, the parameter estimation under such uncorrelated $t$ model encounters some difficulties (see Sect. 3), specially for the scale and shape parameters. This difficulty is removed by proposing several familial longitudinal regression models with $t$ errors, where each of the independent families may contain uncorrelated or correlated multivariate $t$ responses. These new models are discussed in details in Sect. 4. It is expected that the materials of this paper, specially the new models discussed in Sect. 4, would be useful in fitting heavy tailed data following elliptical $t$ distributions.

**Acknowledgements** The author thanks the audience for their comments and suggestions.

## References


Amemiya, T.: Advanced Econometrics. Harvard University Press, Cambridge, MA (1985)

Anderson, T.W., Fang, K.T., Hsu, H.: Maximum likelihood estimates and likelihood ratio criteria for multivariate elliptically contoured distributions. Can. J. Stat. **14**, 55–59 (1986)

Chib, S., Tiwari, R.C., Jammalamadaka, S.R.: Bayes prediction in regressions with elliptical errors. J. Econ. **38**, 349–360 (1988)

Chib, S., Osiewalski, J., Steel, M.F.J.: Posterior inference on the degrees of freedom parameter in multivariate-$t$ regression model. Econ. Lett. **37**, 391–397 (1991)

Cornish, E.A.: The multivariate $t$ distribution associated with a set of normal sample deviates. Aust. J. Phys. **7**, 531–542 (1954)

Dunnett, C.W., Sobel, M.: A bivariate generalization of Student's $t$-distribution with tables for certain special cases. Biometrika **41**, 153–169 (1954)

Fisher, R.A.: Applications of "Student's" distribution. Metron **5**, 90–104 (1925)

Fisher, R.A., Healy, M.J.R.: New tables of Behren's test of significance. J. R. Stat. Soc. B **18**, 212–216 (1956)

Ghosh, B.K.: On the distribution of the difference of two $t$-variables. J. Am. Stat. Assoc. **70**, 463–467 (1975)

Johnson, N.L., Kotz, S.: Continuous Multivariate Distributions. Wiley, New York (1972)

Kariya, T.: Robustness of multivariate tests. Ann. Stat. **9**, 1267–1275 (1981)

Kelker, D.: Distribution theory of spherical distribution and a location scale parameter generalization. Sankhya A **32** , 419–430 (1970)

Kotz, S., Nadarajah, S.: Multivariate T-Distributions and Their Applications. Cambridge University Press, Cambridge (2004)

Kubokawa, T., Srivastava, M.S.: Robust improvements in estimation of mean and covariance matrices in elliptically contoured distribution. Technical Report No. 9714. Department of Statistics, University of Toronto (1977)

Mardia, K.V.: Measures of multivariate skewness and kurtosis with applications. Biometrika **57**, 47–55 (1970)

Muirhead, R.J.: Aspects of Multivariate Statistical Theory. Wiley, New York (1982)

Pearson, K.: Tables of the Incomplete $\beta$-Function. University Press, London (1934)

Raiffa, H., Schlaifer, R.: Applied Statistical Decision Theory. Harvard University Press,Cambridge, MA (1961)

Student: Probable error of a correlation coefficient. Biometrika **6**, 302–310 (1908)

Sutradhar, B.C.: On the characteristic function of multivariate Student $t$-distribution. Can. J. Stat. **14**, 329–337 (1986)

Sutradhar, B.C.: Testing linear hypothesis with $t$ error variable. Sankhya B **40**, 175–180 (1988)

Sutradhar, B.C.: Discrimination of observations into one of two $t$ populations. Biometrics **46**, 827–835 (1990)

Sutradhar, B.C.: Score test for the covariance matrix of the elliptical $t$-distribution. J. Multivar. Anal. **46**, 1–12 (1993)

Sutradhar, B.C.: On cluster regression and factor analysis models with elliptic t errors. In: Anderson, T.W., Fang, K.T., Olkin, I. (eds.) IMS Lecture Notes-Monograph Series, Institute of Mathematical Statistics, Hayward, California, vol. 24, pp. 369–383 (1994)

Sutradhar, B.C.: Multivariate $t$ distribution. In: update volume 2 of Encyclopedia of Statistical Sciences. Wiley, New York, pp. 440–448 (1998)

Sutradhar, B.C.: On the consistency of the estimators of the parameters of elliptically contoured distributions. In the special issue of Journal of Statistical Sciences published in honour of Professor M. S. Haq, U. W. O., ed. B.K. Sinha, Indian Statistical Institute, **3**, 91–103 (2004)

Sutradhar, B.C.: Dynamic Mixed Models for Familial Longitudinal Data. Springer Series in Statistics, 512 pp. Springer, New York (2011)

Sutradhar, B.C., Ali, M.M.: Estimation of the parameters of a regression model with a multivariate $t$ error variable. Commun. Stat. Theory Methods **15**, 429–450 (1986)

Sutradhar, B.C., Ali, M.M.: A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate *t* model. J. Multivar. Anal. **29**, 155–162 (1989)
Walker, G.A., Saw, J.G.: The distribution of linear combinations of *t* variables. J. Am. Stat. Assoc. **73**, 876–878 (1978)
Yuen, K.K., Murthy, V.K.: Percentage points of the *t*-distribution of the *t* statistic when the parent is Student's *t*. Technometrics **16**, 495–497 (1974)
Zellner, A.: Bayesian and non-Bayesian analysis of the regression model with multivariate Student-*t* error terms. J. Am. Stat. Assoc. **71**, 400–405 (1976)

# Longitudinal Mixed Models with *t* Random Effects for Repeated Count and Binary Data

**R. Prabhakar Rao, Brajendra C. Sutradhar, and V.N. Pandit**

**Abstract** Unlike the estimation for the parameters in a linear longitudinal mixed model with independent *t* errors, the estimation of parameters of a generalized linear longitudinal mixed model (GLLMM) for discrete such as count and binary data with independent *t* random effects involved in the linear predictor of the model, may be challenging. The main difficulty arises in the estimation of the degrees of freedom parameter of the *t* distribution of the random effects involved in such models for discrete data. This is because, when the random effects follow a heavy tailed t-distribution, one can no longer compute the basic properties analytically, because of the fact that moment generating function of the *t* random variable is unknown or can not be computed, even though characteristic function exists and can be computed. In this paper, we develop a simulations based numerical approach to resolve this issue. The parameters involved in the numerically computed unconditional mean, variance and correlations are estimated by using the well known generalized quasi-likelihood (GQL) and method of moments approach. It is demonstrated that the marginal GQL estimator for the regression effects asymptotically follow a multivariate Gaussian distribution. The asymptotic properties of the estimators for the rest of the parameters are also indicated.

**Keywords** Asymptotic normal distribution • Consistent estimation • Count and binary panel data • Generalized quasi-likelihood • Regression effects • *t* random effects • Simulating t observations • Stationary and non-stationary covariates • Unconditional mean • Variance and correlations

R. Prabhakar Rao (✉) • V.N. Pandit
Department of Economics, Sri Sathya Sai Institute of Higher Learning,
Prasanthi Nilayam, Andhra Pradesh, India
e-mail: rprabhakarrao@gmail.com; vnpandit@gmail.com

B.C. Sutradhar
Department of Mathematics and Statistics, Memorial University,
St. John's, NL, Canada A1C5S7
e-mail: bsutradh@mun.ca



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_2

# 1 Introduction

Let $(y_{i1}, \ldots, y_{it}, \ldots, y_{iT})$ denote the $T$ repeated count or binary responses for the *ith* subject, $i = 1, \ldots, K$. Also, let $x_{it}$ be the $p \times 1$ vector of covariates corresponding to $y_{it}$, and $\beta$ is the $p \times 1$ regression effects of $x_{it}$ on $y_{it}$. Next suppose that in addition to $x_{it}$, the repeated responses of the $i$th individual are also influenced by one random effect $\gamma_i^*$. Conditional on this random effect $\gamma_i^*$, some authors have modeled the longitudinal correlations of the repeated counts and binary data by using lag 1 dynamic relationships. More specifically, Sutradhar and Bari (2007) have used an AR(1) (auto-regressive order 1) type dynamic relationship to model the longitudinal correlations for repeated count data. Similarly, Sutradhar et al. (2008) [see also Amemiya (1985, p. 353), Manski (1987), and Honore and Kyriazidou (2000, p. 84)] have used a lag 1 dynamic binary mixed logit (BDML) model to accommodate the correlations of the repeated binary data. The unconditional correlation structures in both of these papers have been computed under the normality assumption for the random effects, specifically correlations are obtained by assuming that $\gamma_i^* \overset{iid}{\sim} N(0, \sigma_\gamma^2)$. For convenience, we provide these correlation structures in brief for count and binary data as follows.

## *1.1 Conditional and Unconditional (Normality Based) Correlation Structures for Repeated Count Data*

Suppose that

$$\begin{aligned} y_{i1}|\gamma_i^* &\sim \text{Poi}(\mu_{i1}^*) \;\text{ with }\; \mu_{i1}^* = \exp(x_{i1}'\beta + \gamma_i^*) \\ y_{it}|\gamma_i^* &= \rho \circ [y_{i,t-1}|\gamma_i^*] + [d_{it}|\gamma_i^*], \;\text{for } t = 2, \ldots, T, \end{aligned} \tag{1}$$

where $\text{Poi}(\mu_{it}^*)$ refers to the Poisson distribution with mean parameter $\mu_{it}^*$, and $\rho \circ y_{i,t-1} = \sum_{s=1}^{y_{i,t-1}} b_s(\rho)$ with $Pr[b_s(\rho) = 1] = \rho$, $Pr[b_s(\rho) = 0] = 1 - \rho$, and $[d_{it}|\gamma_i^*] \sim Poi(\mu_{it}^* - \rho\mu_{i,t-1}^*)$, with $\mu_{it}^* = \exp(x_{it}'\beta + \sigma_\gamma \gamma_i^*)$. This model in (1) is referred to as the Poisson AR(1) model which produces the correlation between $y_{iu}$ and $y_{it}$ as

$$\text{corr}(Y_{iu}, Y_{it}|\gamma_i^*) = \rho^{|t-u|}\left[\frac{\mu_{iu}^*}{\mu_{it}^*}\right]^{\frac{1}{2}}, \tag{2}$$

which is free from $\gamma_i^*$, but depends on the time dependent covariates and on $\rho$, a correlation index parameter.

Note that the likelihood inference for the AR(1) model (1) is extremely complicated. This is because under this model, one writes

$$f((y_{i1}, \ldots, y_{it}, \ldots, y_{iT})|\gamma_i^*) = f(y_{i1}|\gamma_i^*)\Pi_{t=2}^T[f_{it|t-1}(y_{it}|y_{i,t-1}, \gamma_i^*)] \tag{3}$$

where the conditional distribution, namely $f_{it|t-1}(y_{it}|y_{i,t-1},\gamma_i^*)$ has a complicated form given by

$$f_{it|t-1}(y_{it}|y_{i,t-1},\gamma_i^*) = \exp[-(\mu_{it}^* - \rho\mu_{i,t-1}^*)]$$
$$\times \sum_{s_{it}=0}^{\min(y_{it},y_{i,t-1})} \frac{y_{i,t-1}!\rho^{s_{it}}(1-\rho)^{y_{it}-s_{it}}(\mu_{it}^* - \rho\mu_{i,t-1}^*)^{y_{it}-s_{it}}}{s_{it}!(y_{i,t-1}-s_{it})!(y_{it}-s_{it})!} \quad (4)$$

(Freeland and McCabe 2004). Furthermore, the integration of the conditional likelihood function (3) over the Gaussian distribution of the random effects, i.e., $\gamma_i^* \overset{iid}{\sim} N(0,\sigma_\gamma^2)$, is an additional complex problem. As opposed to the generalized linear longitudinal mixed model (GLLMM) setup, Over the last two decades many researchers, for example, Breslow and Clayton (1993), Lee and Nelder (1996), Jiang (1998), Sutradhar (2004), among others have used the normality assumption for the random effects in a generalized linear mixed model (GLMM) setup, and discussed the estimation of $\beta$ and $\sigma_\gamma^2$. In the present GLLMM setup (1)–(2), there is an additional correlation index parameter $\rho$ to estimate.

When the normality assumption for the random effect $\gamma_i^*$ is used in the count panel data setup, the mean, variance and correlations of the repeated counts contain three unknown parameters, namely $\beta$, $\sigma_\gamma^2$, and $\rho$. To be specific, by using the moment generating function (mgf) of $\gamma_i^* \overset{iid}{\sim} N(0,\sigma_\gamma^2)$, that is, $E_{\gamma_i^*}(\exp(a\gamma_i^*)) = \exp[\frac{1}{2}a^2\sigma_\gamma^2]$, $a$ being an auxiliary parameter, one obtains the three basic properties of the count panel data as follows (see Sutradhar 2011, Sect. 8.1.1):

$$\mu_{it} = E[Y_{it}] = E_{\gamma_i^*}E[Y_{it}|\gamma_i^*] = \exp(x_{it}'\beta)E_{\gamma_i^*}(\exp(\gamma_i^*)) = \exp[x_{it}'\beta + \frac{1}{2}\sigma_\gamma^2] \quad (5)$$

$$\begin{aligned}\sigma_{itt} &= \text{var}[Y_{it}] = E_{\gamma_i^*}\text{var}[Y_{it}|\gamma_i^*] + \text{var}_{\gamma_i^*}E[Y_{it}|\gamma_i^*] = E_{\gamma_i^*}\mu_{it}^* + \text{var}_{\gamma_i^*}(\mu_{it}^*)\\ &= \exp(x_{it}'\beta)E_{\gamma_i^*}\exp(\gamma_i^*) + \exp(2x_{it}'\beta)\text{var}_{\gamma_i^*}(\exp(\gamma_i^*))\\ &= \mu_{it} + \exp(2x_{it}'\beta)[\exp(2\sigma_\gamma^2) - \exp(\sigma_\gamma^2)]\\ &= \mu_{it} + [\exp(\sigma_\gamma^2) - 1]\mu_{it}^2 \end{aligned} \quad (6)$$

and for $u < t$, the unconditional covariance between $y_{iu}$ and $y_{it}$, is given by

$$\begin{aligned}\sigma_{iut} &= \text{cov}[Y_{iu}, Y_{it}] = E_{\gamma_i^*}[\text{cov}\{(Y_{iu}, Y_{it})|\gamma_i^*\}] + \text{cov}_{\gamma_i^*}[\mu_{iu}^*, \mu_{it}^*]\\ &= \rho^{t-u}\exp(x_{iu}'\beta)E_{\gamma_i^*}[\exp(\gamma_i^*)] + \exp([x_{iu} + x_{it}]'\beta)\text{var}_{\gamma_i^*}\{\exp(\gamma_i^*)\}\\ &= \rho^{t-u}\mu_{iu} + [\exp(\sigma_\gamma^2) - 1]\mu_{iu}\mu_{it}, \end{aligned} \quad (7)$$

yielding the lag $t-u$ correlation

$$\text{corr}(Y_{iu}, Y_{it}) = \frac{\rho^{t-u}\mu_{iu} + [\exp(\sigma_\gamma^2) - 1]\mu_{iu}\mu_{it}}{[\{\mu_{iu} + [\exp(\sigma_\gamma^2) - 1]\mu_{iu}^2\}\{\mu_{it} + [\exp(\sigma_\gamma^2) - 1]\mu_{it}^2\}]^{\frac{1}{2}}}. \tag{8}$$

Notice that the unconditional mean (5) and the unconditional variance (6) are functions in $\beta$ and $\sigma_\gamma^2$, whereas the unconditional covariances (7) and correlations (8) are functions in $\beta,\ \sigma_\gamma^2$, as well as the dynamic dependence or correlation index parameter $\rho$. Remark that Sutradhar and Bari (2007), among others, have exploited the aforementioned moments (5)–(8) to develop a four-moments based generalized quasi-likelihood (GQL) approach for the estimation of these parameters $\beta,\ \sigma_\gamma^2$, and $\rho$.

## *1.2 Conditional and Unconditional (Normality Based) Correlation Structures for Repeated Binary Data*

As indicated earlier, over the last three decades, many econometricians such as Heckman (1981), Amemiya (1985, p. 353), Manski (1987), and Honore and Kyriazidou (2000, p. 844) have made attempts to accommodate the dynamic nature of the repeated binary responses by using a binary dynamic mixed logit (BDML) model given by

$$Pr(y_{it} = 1|\gamma_i, y_{i,t-1}) = \begin{cases} \frac{\exp(x_{i1}'\beta + \sigma_\gamma\gamma_i)}{1+\exp(x_{i1}'\beta + \sigma_\gamma\gamma i)} & \text{for } t = 1 \\ \frac{\exp(x_{it}'\beta + \theta y_{i,t-1} + \sigma_\gamma\gamma_i)}{1+\exp(x_{it}'\beta + \theta y_{i,t-1} + \sigma_\gamma\gamma_i)} & \text{for } t = 2, \ldots, T, \end{cases} \tag{9}$$

where $\beta$ is the effect of the covariates similar to the Poisson model, $\theta$ is referred to as the dynamic dependence parameter, and $\gamma_i = [\gamma_i^*/\sigma_\gamma] \overset{iid}{\sim} (0,1)$. Note that the distribution of $\gamma_i$ is unknown. Also note that even if it is assumed that $\gamma_i$ follows the Gaussian distribution, that is, $\gamma_i \overset{iid}{\sim} N(0,1)$, obtaining the likelihood estimates for $\beta,\ \theta$, and $\sigma_\gamma^2$ is complicated. Honore and Kyriazidou (2000, p. 844) attempted to avoid the estimation difficulty by estimating the $\beta$ and $\theta$ parameters based on the transformed observations, such as the first differences of the responses $y_{i1} - y_{i0},\ y_{i2} - y_{i1}, \ldots$, which are approximately independent of $\gamma_i$. They have used an approximate weighted log likelihood estimation approach, which however puts some impractical restrictions on covariates such as assuming $x_{i3} = x_{i4}$, for the $T = 4$ case.

Remark that recently Bartolucci and Nigro (2010, Eq. (5), Sect. 3) have constructed a random effects free conditional likelihood for a binary model which is different from (9). More specifically, they exploited the conditional approach for a quadratic exponential type model (Cox 1972; Zhao and Prentice 1990) given by

$$Pr(y_{i1}\ldots,y_{iT}|\gamma_i^*,x_{i1}\ldots,x_{iT}) = \Delta_i^{-1}\exp[y_i'\xi_i + \theta g_i'(y_i)1_{T(T-1)/2} + c_i(y_i) + \gamma_i y_i' 1_T] \tag{10}$$

where $y_i = [y_{i1},\ldots,y_{iT}]'$, $g_i(y_i) = [y_{i1}y_{i2},\ldots,y_{iT-1}y_{iT}]'$, and $\xi_i = [\xi_{i1},\ldots,\xi_{iT}]'$, with $\xi_{it} = x_{it}'\beta$. In (10), $1_n$, for example, is an n-dimensional unit vector, $\Delta_i$ is a normalizing constant defined as

$$\Delta_i = \sum \exp[y_i'\xi_i + \theta g_i'(y_i)1_{T(T-1)/2} + c_i(y_i) + \gamma_i y_i' 1_T],$$

with summation overall $2^T$ possible values of $y_i$. Also in (10), $c_i(y_i)$ is referred to as a shape function that can be expressed as a linear combination of products of three or more of the elements of $y_i$. By ignoring $c_i(y_i)$, i.e., $c_i(y_i) = 0$, it can be shown that for a given total score $\sum_{t=1}^{T} y_{it} = y_{i+}$, the conditional distribution of $y_{i1},\ldots,y_{iT}$ may be written as

$$Pr(y_{i1},\ldots,y_{iT}|y_{i+},\gamma_i^*,x_{i1},\ldots,x_{iT}) = \Delta_i^{*-1}\exp\left[y_i'\xi_i + \theta\sum_{t=2}^{T} y_{i,t-1}y_{it}\right] \tag{11}$$

where $\Delta_i^* = \Delta_i$ evaluated at $\sum_{t=1}^{T} y_{it} = y_{i+}$, i.e., $y_{iT} = y_{i+} - \sum_{t=1}^{T-1} y_{it}$. Because the conditional distribution in (11) is free from $\gamma_i$, Bartolucci and Nigro (2010) used this conditional distribution to estimate the main parameters $\beta$ and $\theta$.

We now turn back to the desired binary dynamic mixed model (9). It is clear that even if one is interested to estimate $\beta$ and $\theta$, neither the aforementioned weighted likelihood approach of Honore and Kyriazidou (2000), nor the conditional likelihood approach of Bartolucci and Nigro (2010) can be used to remove the random effects from dynamic mixed model (9) for easier estimation of $\beta$ and $\theta$. Moreover, for binary panel data analysis following (9), one, in fact, is interested to understand the mean and variance of the data, which, however, can not be computed by removing the random effects $\gamma_i$ from the model. In stead, the computation of the moments require averaging over certain functions in $\gamma_i$ over its distribution. Thus, rather than making any attempt to remove $\gamma_i$ from (9), many authors such as Breslow and Clayton (1993), Lee and Nelder (1996), Jiang (1998), and Sutradhar (2004) have studied the inferences for the model (9) under the assumption that $\gamma_i^* \overset{iid}{\sim} N(0,\sigma_\gamma^2)$.

Under this normality assumption, one may obtain the conditional and unconditional means, variance and covariances as follows (see Sutradhar 2011, Sect. 9.2.1). First, conditional on $\gamma_i$, the means of the repeated binary responses under model (9) are given by

$$\pi^*_{it}(\gamma_i) = E[Y_{it}|\gamma_i] = \begin{cases} \frac{\exp(x'_{i1}\beta+\sigma_\gamma\gamma_i)}{1+\exp(x'_{i1}\beta+\sigma_\gamma\gamma_i)}, & \text{for } i = 1, \ldots, K;\ t = 1 \\ p_{it0} + \pi^*_{i,t-1}(p_{it1} - p_{it0}), & \text{for } i = 1, \ldots, k;\ t = 2, \ldots, T \end{cases} \tag{12}$$

where

$$p_{it1} = \frac{\exp(x'_{it}\beta + \theta + \sigma_\gamma\gamma_i)}{[1 + \exp(x'_{it}\beta + \theta + \sigma_\gamma\gamma_i)]} \text{ and } p_{it0} = \frac{\exp(x'_{it}\beta + \sigma_\gamma\gamma_i)}{[1 + \exp(x'_{it}\beta + \sigma_\gamma\gamma_i)]}.$$

Subsequently, one obtains the unconditional means as

$$\begin{aligned} \mu_{it} &= E(Y_{it}) = Pr(y_{it} = 1) \\ &= M^{-1}\sum_{w=1}^{M}\pi^*_{it}(\gamma_{iw}) \\ &= M^{-1}\sum_{w=1}^{M}[p_{it0} + \pi^*_{i,t-1}(p_{it1} - p_{it0})]_{|\gamma_i=\gamma_{iw}} \end{aligned} \tag{13}$$

(Jiang 1998; Sutradhar 2004) where $\gamma_{iw}$ is the $w$th ($w = 1, \ldots, M$) realized value of $\gamma_i$ generated from the standard normal distribution. Here $M$ is a sufficiently large number, such as $M = 5000$. By (12), the $p_{it1,w}$ involved in (13), for example, is written as

$$p_{it1,w} = \frac{\exp(x'_{it}\beta + \theta + \sigma_\gamma\gamma_{iw})}{[1 + \exp(x'_{it}\beta + \theta + \sigma_\gamma\gamma_{iw})]}.$$

Next, conditional on $\gamma_i$, for $u < t$, the second-order expectation may be written as

$$E(Y_{iu}Y_{it}|\gamma_i) = \lambda^*_{iut}(\gamma_i) = \text{cov}(Y_{iu}, Y_{it}|\gamma_i) + \pi_{iu}\pi_{it} = \sigma^*_{iut} + \pi^*_{iu}\pi^*_{it}, \tag{14}$$

where the conditional covariance between $y_{iu}$ and $y_{it}$, conditional on $\gamma_i$, has the formula

$$\sigma^*_{iut} = \text{cov}(Y_{iu}, Y_{it}|\gamma_i) = \pi^*_{iu}(\gamma_i)(1 - \pi^*_{iu}(\gamma_i))\Pi^t_{j=u+1}(p_{ij1} - p_{ij0}). \tag{15}$$

It then follows that the unconditional second-order raw moments have the formula

$$\phi_{iut} = E(Y_{iu}Y_{it}) = M^{-1}\sum_{w=1}^{M}\left[\pi^*_{iu}(\gamma_{iw})(1-\pi^*_{iu}(\gamma_{iw})) \right.$$
$$\left. \times\, \Pi^{t}_{j=u+1}(p_{ij1,w}-p_{ij0,w}) + \pi^*_{iu}(\gamma_{iw})\pi^*_{it}(\gamma_{iw})\right], \quad (16)$$

yielding the unconditional covariance as

$$\sigma_{iut} = \phi_{iut} - \mu_{iu}\mu_{it}, \quad (17)$$

with $\mu_{it}$ is the unconditional mean given by (13).

## 1.3 Plan of the Paper Under the Proposed t Random Effects with Unknown Degrees of Freedom ν

In this paper, as opposed to the Gaussian distribution, we consider a wider class of $t$ distributions for the random effects $\{\gamma_i^*\}$, with mean 0, a scale parameter $\lambda^2_\gamma$, and shape or degrees of freedom parameter $\nu$, i.e, $\gamma_i^* \overset{iid}{\sim} t_\nu(0,\lambda^2_\gamma,\nu)$, with its probability density given by

$$f(\gamma_i^*) = \frac{\nu^{\frac{1}{2}}\Gamma\frac{\nu+1}{2}}{\Gamma\frac{\nu}{2}}(\lambda^2_\gamma)^{\frac{-1}{2}}\left[\nu+\frac{\gamma_i^{*2}}{\lambda^2_\gamma}\right]^{-\frac{\nu+1}{2}}. \quad (18)$$

This $t$ distribution exhibits heavy symmetric tails when $\nu$ is small, and it reduces to the normal distribution $N(0,\sigma^2_\gamma)$ for $\nu\to\infty$. Note, however, that one can not compute the mgf, that is, $E_{\gamma_i^*}(\exp(a\gamma_i^*))$ under this $t$ distribution (18). As a remedy, the moments of this $t$ distribution (18) are computed either from the characteristic function (cf) (Sutradhar 1986) or by direct integrations over the distribution. For $\nu>4$, the first four moments, for example, are given by

$$E(\gamma_i^*) = 0,\ \text{var}(\gamma_i^*) = \frac{\nu}{\nu-2}\lambda^2_\gamma = \sigma^2_\gamma$$
$$E(\gamma_i^{*3}) = 0,\ E(\gamma_i^{*4}) = \frac{3\lambda^4{}_\gamma\nu^2}{(\nu-2)(\nu-4)} = 3\sigma^4_\gamma[\frac{\nu-2}{\nu-4}]. \quad (19)$$

But, it follows from (5)–(8) that in the present longitudinal mixed model setup for count data with $\gamma_i^* \overset{iid}{\sim} t_\nu(0,\lambda^2_\gamma,\nu)$, one requires the result for the mgf $E_{\gamma_i^*}(\exp(a\gamma_i^*))$, which however can not be computed analytically under the $t_\nu$ distribution (18). A similar but different problem arises in the longitudinal mixed model setup for binary data, where for (13)–(15), one needs to generate random effect values $\gamma_{iw}$ from standard $t$ distribution $t(0, 1, \nu)$ with $\nu$ degrees of freedom which is however unknown in practice.

As a remedy, in this paper, we offer a simulation-based numerical approach to compute the mgf, and develop a GQL estimation approach for the estimation of all parameters of the models including the degrees of freedom parameter $\nu > 4$. More specifically, in Sects. 2 and 3, we discuss the Poisson mixed model with $t_\nu$ random effects and the desired inferences. The binary model and the inferences with $t_\nu$ random effects are provided in Sects. 4 and 5. Some concluding remarks are given in Sect. 6.

## 2 Poisson Mixed Model with $t_\nu$ Random Effects

### 2.1 Basic Properties of the Poisson Mixed Model: Unconditional Mean and Variance

In the present setup, $\gamma_i^* \overset{iid}{\sim} t_\nu(0, \lambda_\gamma^2, \nu)$. Now because, similar to (5)–(6), the unconditional mean and variance have the formulas

$$\mu_{it} = E[Y_{it}] = E_{\gamma_i^*} E[Y_{it}|\gamma_i^*] = E_{\gamma_i^*} \mu_{it}^* = \exp(x_{it}'\beta) E_{\gamma_i^*}\left\{\exp(\gamma_i^*)\right\} \tag{20}$$

$$\begin{aligned}\sigma_{itt} &= \text{var}[Y_{it}] = E_{\gamma_i^*} \text{var}[Y_{it}|\gamma_i^*] + \text{var}_{\gamma_i^*} E[Y_{it}|\gamma_i^*] \qquad (21)\\ &= \exp(x_{it}'\beta) E_{\gamma_i^*}\left\{\exp(\gamma_i^*)\right\} + \exp(2x_{it}'\beta)\left[E_{\gamma_i^*}\left\{\exp(2\gamma_i^*)\right\} - \left[E_{\gamma_i^*}\left\{\exp(\gamma_i^*)\right\}\right]^2\right],\end{aligned}$$

they could be evaluated numerically by simulating $\gamma_{iw},\ w = 1, \ldots, W$, for a large $W$ such as $W = 5000$, from $\gamma_{iw} \overset{iid}{\sim} t_\nu(0, 1, \nu)$, and using

$$E_{\gamma_i^*}\left\{\exp(a\gamma_i^*)\right\} = E_{\gamma_i}\left\{\exp(a\lambda_\gamma \gamma_i)\right\} \approx \frac{1}{W}\sum_{w=1}^{W}\left[\exp(a\lambda_\gamma \gamma_{iw})\right], \tag{22}$$

in (20)–(21) for $a = 1, 2$, provided $\nu$ were known. Note that for known $\nu$, this simulated approximation in (22) is quite similar to the simulation approximation used by Sutradhar (2008, Sect. 3) [see also Sutradhar et al. (2008, Eq. (2.6))] for the binary case with random effects generated from $N(0, 1)$ distribution. However, because in the present case, $\nu$ is unknown and requires to be estimated, we resolve this simulation issue by generating $\gamma_{iw},\ w = 1, \ldots, W$ first from a reference $t_4(0, 1, 4)$ distribution (equivalent to standard normal reference distribution) and using the transformation from following Lemma 2.3 so that these $\gamma_{iw}, w = 1, \ldots, W$ subsequently follow the $t_\nu(0, 1, \nu)$ distribution as desired. Lemmas 2.1 and 2.2 below are needed to write the Lemma 2.3.

**Lemma 2.1.** *Suppose that $\psi_i^* \sim N_p(0, \lambda_\gamma^2)$. Next, suppose that $\xi_i^{*2}$ a scalar random variable which follows the well known $\chi_\nu^2$ distribution with $\nu$ degrees of freedom, that is, $\xi_i^{*2} \sim \chi_\nu^2$, and $\psi_i^*$ and $\xi_i^{*2}$ are independent. Then, for $\psi_i = \frac{\psi_i^*}{\lambda_\gamma}$, the ratio variable $\gamma_i^*$ defined as*

$$\gamma_i^* = \lambda_\gamma \psi_i / [\sqrt{(\xi_i^{*2}/\nu)}] = \lambda_\gamma \gamma_i \tag{23}$$

*has the $t_\nu(0, \lambda_\gamma^2, \nu)$ distribution given by (18).*

However, even though $\psi_i$ in (23) is a parameter free normal variable, an observation $\gamma_{iw}^*$ following the *t*-distribution (18) for $\gamma_i^*$ (20)–(22), can not be drawn yet, because the distribution of $\xi_i^{*2}$ is parameter $\nu$ dependent. Because $\nu > 4$ in (18), to resolve this issue, we suggest to use a *t*-distribution with 4 degrees of freedom as a reference distribution. Suppose that $\xi_i^2$ is generated from this $\chi_4^2$ distribution. One may then generate a $\xi_i^{*2}$ from $\chi_\nu^2$ approximately for any $\nu > 4$, by using the relation between $\xi_i^2$ and $\xi_i^{*2}$ as in Lemma 2.2 below.

**Lemma 2.2.** *If $\xi_i^2$ is generated from the $\chi_4^2$ distribution, one may then generate $\xi_i^{*2}$ by using the relationship*

$$\xi_i^{*2} = \sqrt{2\nu}\left[\frac{\xi_i^2 - 4}{\sqrt{8}}\right] + \nu = \frac{1}{2}\sqrt{\nu}\left[\xi_i^2 - 4\right] + \nu, \tag{24}$$

*which has the same first two moments as that of $\chi_\nu^2$.*

One may then generate an observation from a $t_\nu$ distribution as in Lemma 2.3.

**Lemma 2.3.** *For $w = 1, \ldots, W$, with $W = 5000$ (say), the w-th observation $\gamma_{iw}^*$ from the $t_\nu$ distribution may be generated by applying Lemma 2.2 to Lemma 2.1. That is,*

$$\gamma_{iw}^* = \lambda_\gamma \{2\nu\}^{\frac{1}{2}} \frac{\psi_{iw}}{\left[\sqrt{\nu}\left(\xi_{iw}^2 - 4\right) + 2\nu\right]^{\frac{1}{2}}} = \lambda_\gamma \gamma_{iw}, \tag{25}$$

*where $\psi_{iw}$ and $\xi_{iw}^2$ are observations from the standard normal $N(0,1)$ and $\chi_4^2$ distributions, respectively.*

Consequently, by applying (25) under Lemma 2.3, to (22) and (20), one computes the unconditional mean as

$$\begin{aligned}\mu_{it}(\beta, \lambda_\gamma, \nu) = E[Y_{it}] &= \exp(x_{it}'\beta) E_{\gamma_i^*}\left\{\exp(\gamma_i^*)\right\} \\ &= \exp(x_{it}'\beta) E_{\gamma_i^*}\left\{\exp(\lambda_\gamma \gamma_i)\right\}\end{aligned}$$

$$
\begin{aligned}
&= \exp(x'_{it}\beta)\frac{1}{W}\sum_{w=1}^{W}\{\exp(\lambda_\gamma \gamma_{iw})\} \\
&= \frac{1}{W}\exp(x'_{it}\beta)\sum_{w=1}^{W}\exp\left[\frac{\lambda_\gamma \psi_{iw}}{\left\{\frac{1}{2\sqrt{(\nu)}}(\xi^2_{iw}-4)+1\right\}}\right] \\
&= \frac{1}{W}\exp(x'_{it}\beta)\sum_{w=1}^{W}\exp\left\{R(\psi_{iw},\xi^2_{iw};\lambda_\gamma,\nu)\right\}, \qquad (26)
\end{aligned}
$$

where for $w = 1, \ldots, W$, $\psi_{iw}$ are generated from standard normal $N(0, 1)$ distribution, and $\xi^2_{iw}$ are generated from $\chi^2_4$ distributions. Furthermore, $\psi_{iw}$ and $\xi^2_{iw}$ are independent.

In order to compute the unconditional variance, use $\mu^*_{it} = E_{\gamma^*_i}\left[\exp(x'_{it}\beta + \gamma^*_i)\right]$ from (20), and first compute $\phi_{i,tt} = E[Y^2_{it}]$ as follows:

$$
\begin{aligned}
\phi_{i,tt}(\beta,\lambda_\gamma,\nu) = \; & E[Y^2_{it}] = E_{\gamma^*_i}\left[\mu^*_{it} + \mu^{*2}_{it}\right] \\
= \; & \frac{1}{W}\sum_{w=1}^{W}\left[\exp[x'_{it}\beta + \left\{R(\psi_{iw},\xi^2_{iw};\lambda_\gamma,\nu)\right\}]\right. \\
& \left. + \exp[2x'_{it}\beta + \left\{2R(\psi_{iw},\xi^2_{iw};\lambda_\gamma,\nu)\right\}]\right], \qquad (27)
\end{aligned}
$$

where $R(\psi_{iw}, \xi^2_{iw}; \lambda_\gamma, \nu)$ is defined in (26). Hence, the unconditional variance has the formula

$$
\sigma_{i,tt}(\beta,\lambda_\gamma,\nu) = \phi_{i,tt}(\beta,\lambda_\gamma,\nu) - \mu^2_{it}(\beta,\lambda_\gamma,\nu). \qquad (28)
$$

Note that this variance formula can be obtained from (21) as well. We remark that unlike for the Poisson-normal mixed model, the mean and variance under the Poisson-$t_\nu$ mixed model are functions of the regression effects $\beta$, and variance parameter $\lambda_\gamma$ and shape parameter $\nu$ of the random effect distribution.

### *2.2 Correlation Properties of the Poisson Mixed Model: Unconditional Covariances*

To compute the unconditional covariance between $y_{iu}$ and $y_{it}$ $(u < t)$, we first observe from (1)–(2) that their covariance conditional on the random effects $\gamma^*_i$ is not zero. Specifically, by (2), the conditional covariance is given by

$$
\mathrm{cov}[(Y_{iu}, Y_{it})|\gamma^*_i] = \rho^{t-u}\mu^*_{iu}, \qquad (29)
$$

implying that

$$E[Y_{iu}Y_{it}|\gamma_i^*] = \rho^{t-u}\mu_{iu}^* + \mu_{iu}^*\mu_{it}^*. \tag{30}$$

Consequently,

$$\begin{aligned}\delta_{i,ut}(\beta,\lambda_\gamma,\nu,\rho) &= E[Y_{iu}Y_{it}] \\ &= E_{\gamma_i^*}E[Y_{iu}Y_{it}|\gamma_i^*] \\ &= E_{\gamma_i^*}\left[\rho^{t-u}\mu_{iu}^* + \mu_{iu}^*\mu_{it}^*\right] \\ &= \rho^{t-u}\mu_{iu}(\beta,\lambda_\gamma,\nu) + E_{\gamma_i^*}\left[\mu_{iu}^*\mu_{it}^*\right],\end{aligned} \tag{31}$$

where the unconditional mean $\mu_{iu}(\beta, \lambda_\gamma, \nu)$ has the formula similar to that of (26). Next, by similar computation as in (27), one obtains

$$\begin{aligned}\delta_{i,ut}(\beta,\lambda_\gamma,\nu,\rho) &= \rho^{t-u}\mu_{iu}(\beta,\lambda_\gamma,\nu) + E_{\gamma_i^*}\left[\exp\{x_{iu}+x_{it}\}'\beta + 2\gamma_i^*\right] \\ = \rho^{t-u}\mu_{iu}(\beta,\lambda_\gamma,\nu) &+ \exp\left[\{x_{iu}+x_{it}\}'\beta\right]\frac{1}{W}\sum_{w=1}^{W}\exp\left[2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right].\end{aligned} \tag{32}$$

Hence, for $u < t$, the unconditional covariance between $y_{iu}$ and $y_{it}$ is given by

$$\sigma_{i,ut}(\beta,\lambda_\gamma,\nu,\rho) = \delta_{i,ut}(\beta,\lambda_\gamma,\nu,\rho) - \mu_{iu}(\beta,\lambda_\gamma,\nu)\mu_{it}(\beta,\lambda_\gamma,\nu), \tag{33}$$

where $\delta_{i,ut}(\beta, \lambda_\gamma, \nu, \rho)$ has the formula given by (32) and $\mu_{it}(\beta, \lambda_\gamma, \nu)$ is given by (26).

## 3 GQL Estimation for the Parameters of the Poisson Mixed Model

The estimation of the parameters of the model will be done in cycle of iterations. In Sect. 3.1, we discuss a generalized quasi-likelihood (GQL) (Sutradhar 2003, Sect. 3) estimation approach for the estimation of the main regression parameter $\beta$ under the assumption that other parameters $(\rho, \lambda_\gamma, \nu)$ are known or their consistent estimates are available. In subsequent sections, we discuss their consistent estimation.

### 3.1 GQL Estimation for the Regression Effects $\boldsymbol{\beta}$

For $\beta$ estimation, we exploit the first order responses, namely $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT}]'$. Suppose that $\mu_i = E[Y_i]$. This mean vector is given by $\mu_i(\beta, \lambda_\gamma, \nu) = [\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT}]'$, where, by (26), $\mu_{it}(\beta, \lambda_\gamma, \nu)$ has the formula

$$\mu_{it}(\beta, \lambda_\gamma, \nu) = \exp(x_{it}'\beta)\frac{1}{W}\sum_{w=1}^{W}\exp\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\} : T \times 1.$$

Next by using the formulas for the variances $\sigma_{i,tt}(\beta, \lambda_\gamma, \nu)$ from (28), and the covariances $\sigma_{i,ut}(\beta, \lambda_\gamma, \nu, \rho)$ from (33), we construct the $T \times T$ covariance matrix as

$$\Sigma_i(\beta, \lambda_\gamma, \nu, \rho) = (\sigma_{i,ut}(\beta, \lambda_\gamma, \nu, \rho)) : T \times T, \text{ for } u = t; \text{and } u \neq t.$$

Note that under the present model $\sigma_{i,tt}(\cdot)$ does not follow from $\sigma_{i,ut}(\cdot)$ as a special case. More specifically, $\sigma_{i,ut}(\cdot)$'s are constructed for $u < t$. The GQL estimating equation for $\beta$ is then given by

$$\sum_{i=1}^{K}\frac{\partial \mu_i'(\beta, \lambda_\gamma, \nu)}{\partial \beta}\Sigma_i^{-1}(\beta, \lambda_\gamma, \nu, \rho)\left(y_i - \mu_i(\beta, \lambda_\gamma, \nu)\right) = 0, \tag{34}$$

(Sutradhar 2003, 2004) where $\frac{\partial \mu_i'(\beta,\lambda_\gamma,\nu)}{\partial \beta}$ may be computed by using the formula for $\frac{\partial \mu_{it}(\beta,\lambda_\gamma,\nu)}{\partial \beta}$ for all $t = 1, \ldots, T$. This derivative follows from (26), and is given by

$$\frac{\partial \mu_{it}(\beta, \lambda_\gamma, \nu)}{\partial \beta} = \mu_{it}(\beta, \lambda_\gamma, \nu)x_{it}.$$

Consequently, the GQL estimating equation in (34) reduces to

$$\sum_{i=1}^{K} X_i'A_i\Sigma_i^{-1}(\beta, \lambda_\gamma, \nu, \rho)\left(y_i - \mu_i(\beta, \lambda_\gamma, \nu)\right) = 0, \tag{35}$$

where $X_i' = (x_{i1}, \ldots, x_{it}, \ldots, x_{iT})$ is the $p \times T$ covariate matrix for the $i$th individual, and

$$A_i = \text{diag}[\mu_{i1}(\beta, \lambda_\gamma, \nu), \ldots, \mu_{it}(\beta, \lambda_\gamma, \nu), \ldots, \mu_{iT}(\beta, \lambda_\gamma, \nu)] : \ T \times T.$$

### 3.1.1 Asymptotic Properties of the GQL Estimator of $\beta$

For true $\beta$, define

$$\bar{f}_K(\beta) = \frac{1}{K}\sum_{i=1}^{K} f_i(\beta) = \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \mu_i'(\beta, \lambda_\gamma, \nu)}{\partial \beta} \Sigma_i^{-1}(\beta, \lambda_\gamma, \nu, \rho)\left(y_i - \mu_i(\beta, \lambda_\gamma, \nu)\right), \tag{36}$$

where $y_1, \ldots, y_i, \ldots, y_K$ are independent to each other as they are collected from $K$ independent individuals, but they are not identically distributed because

$$Y_i \sim (\mu_i(\beta, \lambda_\gamma, \nu), \Sigma_i(\beta, \lambda_\gamma, \nu, \rho)), \tag{37}$$

where the mean vectors and covariance matrices vary for the individuals $i = 1, \ldots, K$. By (37), it follows from (36) that

$$\begin{aligned}
E[\bar{f}_K(\beta)] &= 0 \\
\text{cov}[\bar{f}_K(\beta)] &= \frac{1}{K^2}\sum_{i=1}^{K} \text{cov}[f_i(\beta)] \\
&= \frac{1}{K^2}\sum_{i=1}^{K} \frac{\partial \mu_i'(\beta, \lambda_\gamma, \nu)}{\partial \beta} \Sigma_i^{-1}(\beta, \lambda_\gamma, \nu, \rho) \frac{\partial \mu_i(\beta, \lambda_\gamma, \nu)}{\partial \beta'} \\
&= \frac{1}{K^2}\sum_{i=1}^{K} V_i(\beta, \lambda_\gamma, \nu, \rho) = \frac{1}{K^2} V_K^*(\beta, \lambda_\gamma, \nu, \rho).
\end{aligned} \tag{38}$$

Next if the multivariate version of Lindeberg's condition holds, that is,

$$\lim_{K\to\infty} {V^*}_K^{-1} \sum_{i=1}^{K} \sum_{(f_i' {V^*}_K^{-1} f_i) > \epsilon} f_i f_i' g(f_i) = 0 \tag{39}$$

for all $\epsilon > 0$, $g(\cdot)$ being the probability distribution of $f_i$, then Lindeberg-Feller central limit theorem (Amemiya 1985, Theorem 3.3.6; McDonald 2005, Theorem 2.2) imply that

$$Z_K = K[V_K^*]^{-\frac{1}{2}} \bar{f}_K(\beta) \to N_p(0, I_p). \tag{40}$$

Next because $\hat{\beta}_{GQL}$ is a solution of (34), one writes by (36) that

$$\sum_{i=1}^{K} f_i(\hat{\beta}_{GQL}) = 0, \tag{41}$$

which by first order Taylor's series expansion produces

$$\sum_{i=1}^{K} f_i(\beta) + (\hat{\beta}_{GQL} - \beta) \sum_{i=1}^{K} f_i'(\beta) = 0. \tag{42}$$

That is,

$$\begin{aligned}
\hat{\beta}_{GQL} - \beta &= - \left[ \sum_{i=1}^{K} f_i'(\beta) \right]^{-1} \sum_{i=1}^{K} f_i(\beta) \\
&= - \left[ - \sum_{i=1}^{K} \frac{\partial \mu_i'(\beta, \lambda_\gamma, \nu)}{\partial \beta} \Sigma_i^{-1}(\beta, \lambda_\gamma, \nu, \rho) \frac{\partial \mu_i(\beta, \lambda_\gamma, \nu)}{\partial \beta'} \right]^{-1} \sum_{i=1}^{K} f_i(\beta) \\
&= \left[ V_K^*(\beta, \lambda_\gamma, \nu, \rho) \right]^{-1} K \bar{f}(\beta) \\
&= \left[ V_K^*(\beta, \lambda_\gamma, \nu, \rho) \right]^{-\frac{1}{2}} Z_K \to N(0, {V^*}_K^{-1}(\beta, \lambda_\gamma, \nu, \rho)),
\end{aligned} \tag{43}$$

by (40). It then follows that

$$\lim_{K \to \infty} \hat{\beta}_{GQL} \to N(\beta, {V^*}_K^{-1}(\beta, \lambda_\gamma, \nu, \rho)). \tag{44}$$

Also it follows that

$$||[V_K^*(\beta, \lambda_\gamma, \nu, \rho)]^{\frac{1}{2}} [\hat{\beta}_{GQL} - \beta]|| = O_p(\sqrt{p}). \tag{45}$$

## *3.2 GQL Estimation for the Scale and Shape Parameters*

Notice from Sect. 2.1 that all three basic moment properties, namely the mean function (26), variances in (28), and the covariances given by (33) contain the scale parameter $\lambda_\gamma$ and the shape parameter $\nu$. Thus, it is sensible to exploit all first and second order responses to estimate these parameters. Note that the second order responses consist of both squared ($ss$) and pair-wise products ($pp$) of all repeated observations. Consequently, we consider a vector $g_i$ consisting of all first order and second order responses. In notation, $g_i$ has the form

$$g_i = [y_i', \; y_{iss}', \; y_{ipp}']' : \frac{T(T+3)}{2} \times 1, \tag{46}$$

where $y_i = [y_{i1}, \ldots, y_{it}, \ldots, y_{iT}]' : T \times 1$, as in (34), and

$$\begin{aligned}
y_{iss} &= [y_{i1}^2, \ldots, y_{it}^2, \ldots, y_{iT}^2]' : T \times 1, \text{ and} \\
y_{ipp} &= [y_{i1} y_{i2}, \ldots, y_{iu} y_{it}, \ldots, y_{i,T-1} y_{iT}]' : \frac{T(T-1)}{2} \times 1.
\end{aligned}$$

Let

$$E[g_i] = [\mu_i'(\beta, \lambda_\gamma, \nu), \phi_i'(\beta, \lambda_\gamma, \nu), \delta_i'(\beta, \lambda_\gamma, \nu, \rho)]'$$
$$= \eta_i(\beta, \lambda_\gamma, \nu, \rho) \text{ (say)}, \quad (47)$$

where

$$\mu_i(\beta, \lambda_\gamma, \nu) = [\mu_{i1}(\beta, \lambda_\gamma, \nu), \ldots, \mu_{it}(\beta, \lambda_\gamma, \nu), \ldots, \mu_{iT}(\beta, \lambda_\gamma, \nu)]'$$
$$\phi_i(\beta, \lambda_\gamma, \nu) = [\phi_{i,11}(\beta, \lambda_\gamma, \nu), \ldots, \phi_{i,tt}(\beta, \lambda_\gamma, \nu), \ldots, \phi_{i,TT}(\beta, \lambda_\gamma, \nu)]'$$
$$\delta_i(\beta, \lambda_\gamma, \nu, \rho) = [\delta_{i,12}(\beta, \lambda_\gamma, \nu, \rho), \ldots, \delta_{i,ut}(\beta, \lambda_\gamma, \nu, \rho), \ldots, \delta_{i,T-1,T}(\beta, \lambda_\gamma, \nu, \rho)]',$$

with $\mu_{it}(\beta, \lambda_\gamma, \nu)$ and $\phi_{i,tt}(\beta, \lambda_\gamma, \nu)$ for all $t = 1, \ldots, T$, are given by (26) and (27), respectively, and $\delta_{i,ut}(\beta, \lambda_\gamma, \nu, \rho)$ for $u < t$, are defined as in (32). Further, let

$$\Omega_i(\beta, \lambda_\gamma, \nu, \rho) = \text{cov}[g_i]$$
$$= \begin{bmatrix} \Sigma_i & \Omega_{i,ss} & \Omega_{i,pp} \\ \Omega_{i,ss}' & \Sigma_{i,ss} & \Omega_{i,sp} \\ \Omega_{i,pp}' & \Omega_{i,sp}' & \Sigma_{i,pp} \end{bmatrix}, \quad (48)$$

where

$$\Sigma_i = \text{cov}[Y_i],\ \Sigma_{i,ss} = \text{cov}[Y_{iss}],\ \Sigma_{i,pp} = \text{cov}[Y_{ipp}]$$
$$\Omega_{i,ss} = \text{cov}[Y_i, Y_{iss}'],\ \Omega_{i,pp} = \text{cov}[Y_i, Y_{ipp}'],\ \Omega_{i,sp} = \text{cov}[Y_{iss}, Y_{ipp}'].$$

Further, let $\pi = [\lambda_\gamma, \nu]' : 2 \times 1$, be a vector of the scale and shape parameters of the random effects distribution. Similar to (34), for known $\beta$ and $\rho$, one may then estimate the $\pi$ vector by solving the GQL estimation equation given by

$$\sum_{i=1}^{K} \frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi} \Omega_i^{-1}(\beta, \lambda_\gamma, \nu, \rho) \left(g_i - \eta_i(\beta, \lambda_\gamma, \nu, \rho)\right) = 0, \quad (49)$$

where

$$\frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi} = \begin{pmatrix} \frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \lambda_\gamma} \\ \frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \nu} \end{pmatrix}. \quad (50)$$

Note that the derivatives in (50) may be computed by using the following general derivatives with respect to $\lambda_\gamma$ and $\nu$:

$$\frac{\partial \mu_{it}}{\partial \lambda_\gamma} = \exp(x_{it}'\beta)\frac{1}{W}\sum_{w=1}^{W}\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \lambda_\gamma}\exp\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\},$$

$$\frac{\partial \mu_{it}}{\partial \nu} = \exp(x_{it}'\beta)\frac{1}{W}\sum_{w=1}^{W}\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \nu}\exp\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\},$$

$$\frac{\partial \phi_{i,tt}}{\partial \lambda_\gamma} = \frac{1}{W}\sum_{w=1}^{W}\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \lambda_\gamma}\Big[\exp[x_{it}'\beta+\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}] + 2\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \lambda_\gamma}\exp[2x_{it}'\beta+\left\{2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}]\Big],$$

$$\frac{\partial \phi_{i,tt}}{\partial \nu} = \frac{1}{W}\sum_{w=1}^{W}\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \nu}\Big[\exp[x_{it}'\beta+\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}] + 2\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \nu}\exp[2x_{it}'\beta+\left\{2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}]\Big],$$

$$\frac{\partial \delta_{i,ut}}{\partial \lambda_\gamma} = \rho^{t-u}\frac{\partial \mu_{iu}(\beta,\lambda_\gamma,\nu)}{\partial \lambda_\gamma} + \exp\left[\{x_{iu}+x_{it}\}'\beta\right]\frac{1}{W}\sum_{w=1}^{W}2\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \lambda_\gamma} \times\exp\left[2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right] - \left\{\frac{\partial \mu_{iu}(\beta,\lambda_\gamma,\nu)}{\partial \lambda_\gamma}\mu_{it} + \mu_{iu}\frac{\partial \mu_{it}(\beta,\lambda_\gamma,\nu)}{\partial \lambda_\gamma}\right\}$$

$$\frac{\partial \delta_{i,ut}}{\partial \nu} = \rho^{t-u}\frac{\partial \mu_{iu}(\beta,\lambda_\gamma,\nu)}{\partial \nu} + \exp\left[\{x_{iu}+x_{it}\}'\beta\right]\frac{1}{W}\sum_{w=1}^{W}2\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \nu} \times\exp\left[2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right] - \left\{\frac{\partial \mu_{iu}(\beta,\lambda_\gamma,\nu)}{\partial \nu}\mu_{it} + \mu_{iu}\frac{\partial \mu_{it}(\beta,\lambda_\gamma,\nu)}{\partial \nu}\right\},$$

where

$$\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \lambda_\gamma} = \left[\frac{\psi_{iw}}{\left\{\frac{1}{2\sqrt{(\nu)}}(\xi_{iw}^2-4)+1\right\}}\right]$$

$$\frac{\partial R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)}{\partial \nu} = \left[\frac{\lambda_\gamma\psi_{iw}\left\{\frac{1}{4\{\sqrt{(\nu)}\}^3}(\xi_{iw}^2-4)\right\}}{\left\{\frac{1}{2\sqrt{(\nu)}}(\xi_{iw}^2-4)+1\right\}^2}\right]. \tag{51}$$

The construction of the GQL estimating equation (49) still requires the computational formula for the weight matrix $\Omega_i(\beta,\lambda_\gamma,\nu,\rho)$. Now because this weight matrix requires the computation of second, third and fourth order moments for the repeated count data, unlike for the Gaussian data the computation of these moments are complicated. Some of the fourth order moments may not be computable without further joint distributional assumption for these repeated counts. Thus, for simplicity and because the consistent estimation of the parameters in $\pi$ does not require the use of exact weight matrix, in the next section, we provide an approximation for the computation of the elements of the weight matrix $\Omega_i(\beta, \lambda_\gamma, \nu, \rho)$ by pretending that the correlation index is zero, that is, $\rho = 0$ in (2). This assumption is equivalent to say that the repeated counts are assumed to conditionally (conditional on the random effects) independent (CI).

### 3.2.1 Computation of $\Omega_i(CI) \equiv \Omega_i^*(\beta, \lambda_\gamma, \nu)$

Note that as outlined above, the $\Omega_i(\beta, \lambda_\gamma, \nu, \rho) = \text{cov}(g_i)$ matrix in (49) will be replaced by

$$\text{cov}(g_i|\rho = 0) = \Omega_i^*(\beta, \lambda_\gamma, \nu), \tag{52}$$

which contains moments up to order four under conditionally independence (CI) assumption. More specifically, we compute the $\Omega_i(\cdot)$ matrix in (48), but, under the assumption that $\rho = 0$, that is,

$$\begin{aligned} \Omega_i^*(\beta, \lambda_\gamma, \nu) &= \text{cov}[g_i|\rho = 0] \\ &= \begin{bmatrix} \Sigma_i^* & \Omega_{i,ss}^* & \Omega_{i,pp}^* \\ {\Omega^*}'_{i,ss} & \Sigma_{i,ss}^* & \Omega_{i,sp}^* \\ {\Omega^*}'_{i,pp} & {\Omega^*}'_{i,sp} & \Sigma_{i,pp}^* \end{bmatrix}, \end{aligned} \tag{53}$$

where

$$\begin{aligned} \Sigma_i^* &= \text{cov}[Y_i|\rho = 0], \ \ \Sigma_{i,ss}^* = \text{cov}[Y_{iss}|\rho = 0], \ \ \Sigma_{i,pp}^* = \text{cov}[Y_{ipp}|\rho = 0] \\ \Omega_{i,ss}^* &= \text{cov}[(Y_i, Y_{iss}')|\rho = 0], \ \ \Omega_{i,pp}^* = \text{cov}[(Y_i, Y_{ipp}')|\rho = 0], \\ \Omega_{i,sp}^* &= \text{cov}[(Y_{iss}, Y_{ipp}')|\rho = 0]. \end{aligned}$$

#### (a) Computation of the Second Order Moments Matrix $\Sigma_i^*$ :

Because the variances are not affected by the correlation index parameter, their formulas remain the same as in (28). However, the covariances under $\rho = 0$ will be different than (33). More specifically the formulas for the variances and covariances under the assumption $\rho = 0$ are given by

$$\begin{aligned} \text{var}[Y_{it}|\rho] = &\ \sigma_{i,tt}^*(\beta, \sigma_\gamma^2, \nu) \\ = &\ \sigma_{i,tt}(\beta, \lambda_\gamma, \nu) = \phi_{i,tt}(\beta, \lambda_\gamma, \nu) - \mu_{it}^2(\beta, \lambda_\gamma, \nu), \ \text{by (28)}; \end{aligned} \tag{54}$$

$$\begin{aligned} \text{cov}[(Y_{iu}, Y_{it})|\rho = 0] &= \sigma^*_{i,ut}(\beta, \sigma^2_\gamma, \nu) \\ &= \exp\left[\{x_{iu} + x_{it}\}'\beta\right] \frac{1}{W} \sum_{w=1}^{W} \exp\left[2R(\psi_{iw}, \xi^2_{iw}; \lambda_\gamma, \nu)\right] \\ &\quad -\mu_{iu}(\beta, \lambda_\gamma, \nu)\mu_{it}(\beta, \lambda_\gamma, \nu), \text{ by (32)–(33).} \end{aligned} \tag{55}$$

**(b) Computation of the Third Order Moments Matrix $\Omega^*_{i,ss}$ :**

To compute this matrix $\Omega^*_{i,ss} = \text{cov}[\{Y_i, Y'_{iss}\}|\rho = 0]$, it is sufficient to compute the elements (i) $\text{cov}[\{Y_{it}, Y^2_{it}\}|\rho = 0]$, and (ii) $\text{cov}[\{Y_{iu}, Y^2_{it}\}|\rho = 0]$, for $u < t$.

(i) Formula for $\text{cov}[\{Y_{it}, Y^2_{it}\}|\rho = 0]$ :

Notice that this formula does not depend on $\rho$, and the conditioning on $\rho = 0$ is not needed. Thus

$$\begin{aligned} \text{cov}[\{Y_{it}, Y^2_{it}\}|\rho = 0] &= E[Y^3_{it}] - E[Y_{it}]E[Y^2_{it}] \\ &= E[Y^3_{it}] - \mu_{it}(\beta, \lambda_\gamma, \nu)\phi_{i,tt}(\beta, \lambda_\gamma, \nu), \text{ by (54),} \end{aligned} \tag{56}$$

where

$$\begin{aligned} E[Y^3_{it}] &= E_{\gamma^*_i}\left[\mu^*_{it} + 3\mu^{*2}_{it} + \mu^{*3}_{it}\right] \\ &= \frac{1}{W}\sum_{w=1}^{W}\left[\exp[x'_{it}\beta + \left\{R(\psi_{iw}, \xi^2_{iw}; \lambda_\gamma, \nu)\right\}]\right. \\ &\quad + 3\exp[2x'_{it}\beta + \left\{2R(\psi_{iw}, \xi^2_{iw}; \lambda_\gamma, \nu)\right\}] \\ &\quad \left.+ \exp[3x'_{it}\beta + \left\{3R(\psi_{iw}, \xi^2_{iw}; \lambda_\gamma, \nu)\right\}]\right], \end{aligned} \tag{57}$$

with $R(\psi_{iw}, \xi^2_{iw}; \lambda_\gamma, \nu)$ as defined in (26).

(ii) Formula for $\text{cov}[\{Y_{iu}, Y^2_{it}\}|\rho = 0]$, for $u < t$ :

Because $u$ and $t$ denote two different times points, the covariance between $y_{iu}$ and $y^2_{it}$ is a function of the correlation index parameter $\rho$. However, we now simplify this covariance formula as follows under the assumption that $\rho = 0$.

$$\begin{aligned} \text{cov}[\{Y_{iu}, Y^2_{it}\}|\rho = 0] &= E[Y_{iu}Y^2_{it}|\rho = 0] - E[Y_{iu}]E[Y^2_{it}] \\ &= E[Y_{iu}Y^2_{it}|\rho = 0] - \mu_{iu}(\beta, \lambda_\gamma, \nu)\phi_{i,tt}(\beta, \lambda_\gamma, \nu), \text{ by (54),} \end{aligned} \tag{58}$$

where, by (30) and (26)–(27), one writes

$$\begin{aligned} E[Y_{iu}Y^2_{it}|\rho = 0] &= E_{\gamma^*_i}[E\{Y_{iu}|\gamma^*_i\}E\{Y^2_{it}|\gamma^*_i\}] \\ &= E_{\gamma^*_i}[\mu^*_{iu}\{\mu^*_{it} + \mu^{*2}_{it}\}] \end{aligned} \tag{59}$$

$$= \frac{1}{W}\sum_{w=1}^{W}\Big[\exp[\{x_{iu}+x_{it}\}'\beta + 2\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}] + \exp[\{x_{iu}+2x_{it}\}'\beta + 3\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}]\Big]. \quad (60)$$

**(c) Computation of the Third Order Moments Matrix $\Omega^*_{i,pp}$ :**

To compute this matrix $\Omega^*_{i,pp} = \mathrm{cov}[\{Y_i, Y'_{ipp}\}|\rho = 0]$, it is sufficient to compute the elements (i) $\mathrm{cov}[\{Y_{iu}, Y_{iu}Y_{it}\}|\rho = 0]$ for $u < t$ or $u > t$, and (ii) $\mathrm{cov}[\{Y_{iu}, Y_{it}Y_{im}\}|\rho = 0]$, for $u \neq t, u \neq m, t < m$.

(i) Formula for $\mathrm{cov}[\{Y_{iu}, Y_{iu}Y_{it}\}|\rho = 0]$ :
By similar calculations as in (59)–(60), we write

$$\begin{aligned}
\mathrm{cov}[\{Y_{iu}, Y_{iu}Y_{it}\}|\rho = 0] &= E_{\gamma_i^*}[E\{Y_{iu}^2Y_{it}\}|\gamma_i^*] - E[Y_{iu}]E[\{Y_{iu}Y_{it}\}|\rho = 0] \\
&= E_{\gamma_i^*}[\{\mu^*_{iu} + \mu^{*\,2}_{iu}\}\mu^*_{it}] - \mu_{iu}(\beta,\lambda_\gamma,\nu)E_{\gamma_i^*}[\mu^*_{iu}\mu^*_{it}] \\
&= \frac{1}{W}\sum_{w=1}^{W}\Big[\exp[\{x_{iu}+x_{it}\}'\beta + 2\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}] \\
&\quad + \exp[\{2x_{iu}+x_{it}\}'\beta + 3\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}] \\
&\quad - \mu_{iu}(\beta,\lambda_\gamma,\nu)\exp[\{x_{iu}+x_{it}\}'\beta + 2\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}]\Big]. \quad (61)
\end{aligned}$$

(ii) Formula for $\mathrm{cov}[\{Y_{iu}, Y_{it}Y_{im}\}|\rho = 0]$ :
By similar calculations as in (i), we write the formula for this covariance as

$$\begin{aligned}
&\mathrm{cov}[\{Y_{iu}, Y_{it}Y_{im}\}|\rho = 0] \\
&= \frac{1}{W}\sum_{w=1}^{W}\Big[\exp[\{x_{iu}+x_{it}+x_{im}\}'\beta + 3\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}] \\
&\quad - \mu_{iu}(\beta,\lambda_\gamma,\nu)\exp[\{x_{it}+x_{im}\}'\beta + 2\left\{R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right\}]\Big]. \quad (62)
\end{aligned}$$

**(d) Computation of the Fourth Order Moments Matrix $\Sigma^*_{i,ss}$ :**

To compute this fourth order matrix, one needs the formulas for two general elements, namely (i) $\mathrm{var}[Y_{it}^2]$, and (ii) $\mathrm{cov}[\{Y_{iu}^2, Y_{it}^2\}|\rho = 0]$. These formulas are developed as follows:

(i) Recall from (27) that $E[Y_{it}^2] = \phi_{i,tt}(\beta,\lambda_\gamma,\nu)$. Next because

$$E[Y_{it}^4|\gamma_i^*] = \left[\mu^*_{it} + 7\mu^{*2}_{it} + 6\mu^{*3}_{it} + \mu^{*4}_{it}\right],$$

it then follows that

$$\begin{aligned}\mathrm{var}(Y_{it})^2 &= E[Y_{it}^4] - [E\{Y_{it}^2\}]^2 \\ &= E_{\gamma_i^*} E[Y_{it}^4|\gamma_i^*] - [\phi_{i,tt}(\beta, \lambda_\gamma, \nu)]^2 \\ &= \frac{1}{W}\sum_{w=1}^{W}\left[\exp\left[x_{it}'\beta + \left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}\right] + 7\exp\left[2x_{it}'\beta + \left\{2R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}\right]\right. \\ &+ 6\exp[3x_{it}'\beta + \left\{3R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}] + \exp[4x_{it}'\beta + \left\{4R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\Big], \end{aligned} \tag{63}$$

with $R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)$ as defined in (26).

(ii) Formula for $\mathrm{cov}[\{Y_{iu}^2, Y_{it}^2\}|\rho = 0]$, for $u < t$ :

By similar calculations, this covariance has the computing formula given by

$$\mathrm{cov}[\{Y_{iu}^2, Y_{it}^2\}|\rho = 0] = E[\{Y_{iu}^2 Y_{it}^2\}|\rho = 0] - \phi_{i,uu}(\beta, \lambda_\gamma, \nu)\phi_{i,tt}(\beta, \lambda_\gamma, \nu), \tag{64}$$

where

$$\begin{aligned}E[\{Y_{iu}^2 Y_{it}^2\}|\rho = \ 0] &= E_{\gamma_i^*}\left[\{\mu_{iu}^* + \mu_{iu}^{*\,2}\}\{\mu_{it}^* + \mu_{it}^{*\,2}\}\right] \\ &= \frac{1}{W}\sum_{w=1}^{W}\Big[\exp[\{x_{iu} + x_{it}\}'\beta + 2\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}] \\ &+ \exp\left[\{x_{iu} + 2x_{it}\}'\beta + \left\{3R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}\right] \\ &+ \exp\left[\{2x_{iu} + x_{it}\}'\beta + \left\{3R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}\right] \\ &+ \exp[2\{x_{iu} + x_{it}\}'\beta + \left\{4R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}\Big]. \end{aligned} \tag{65}$$

**(e) Computation of the Fourth Order Moments Matrix $\Omega_{i,sp}^*$:**

To compute this fourth order matrix, one needs the formulas for two general covariance elements, namely (i) $\mathrm{cov}[\{Y_{iu}^2, Y_{iu}Y_{it}\}|\rho = 0]$, and (ii) $\mathrm{cov}[\{Y_{iu}^2, Y_{it}Y_{im}\}|\rho = 0]$. These formulas are developed as follows:

(i) Formula for $\mathrm{cov}[\{Y_{iu}^2, Y_{iu}Y_{it}\}|\rho = 0]$ :

$$\begin{aligned}\mathrm{cov}[\{Y_{iu}^2, Y_{iu}Y_{it}\}|\rho = 0] &= E[\{Y_{iu}^3 Y_{it}\}|\rho = 0] \\ &-\phi_{i,uu}(\beta, \lambda_\gamma, \nu)\frac{1}{W}\sum_{w=1}^{W}\Big[\exp[\{x_{iu} + x_{it}\}'\beta \\ &+2\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\Big], \text{ by (27) and (32)}, \end{aligned} \tag{66}$$

where

$$\begin{aligned}
E[\{Y_{iu}^3 Y_{it}\}|\rho = \ & 0] = E_{\gamma_i^*}\left[E[Y_{iu}^3|\gamma_i^*]E[Y_{it}|\gamma_i^*]\right] \\
= & \ E_{\gamma_i^*}\left[\{\mu_{iu}^* + 3\mu_{iu}^{*\,2} + \mu_{iu}^{*\,3}\}\{\mu_{it}^*\}\right] \\
= & \ \frac{1}{W}\sum_{w=1}^{W}\left[\exp[\{x_{iu} + x_{it}\}'\beta + 2\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right. \\
& + 3\exp[\{2x_{iu} + x_{it}\}'\beta + 3\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}] \\
& \left. + \ \exp[\{3x_{iu} + x_{it}\}'\beta + 4\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right]. \qquad (67)
\end{aligned}$$

(ii) $\mathrm{cov}[\{Y_{iu}^2, Y_{it}Y_{im}\}|\rho = 0]$ :

$$\begin{aligned}
\mathrm{cov}[\{Y_{iu}^2, Y_{it}Y_{im}\}|\rho = 0] &= E_{\gamma_i^*}\left[\{\mu_{iu}^* + \mu_{iu}^{*\,2}\}\{\mu_{it}^*\mu_{im}^*\}\right] \\
-\phi_{i,uu}(\beta, \lambda_\gamma, \nu)\frac{1}{W}\sum_{w=1}^{W}&\left[\exp[\{x_{it} + x_{im}\}'\beta + 2\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right], \qquad (68)
\end{aligned}$$

where the first term in the right hand side of (68) has the formula

$$\begin{aligned}
& E_{\gamma_i^*}\left[\{\mu_{iu}^* + \mu_{iu}^{*\,2}\}\{\mu_{it}^*\mu_{im}^*\}\right] \\
&= \frac{1}{W}\sum_{w=1}^{W}\left[\exp[\{x_{iu} + x_{it} + x_{im}\}'\beta + 3\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right. \\
&\quad \left. + \exp[\{2x_{iu} + x_{it} + x_{im}\}'\beta + 4\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right]. \qquad (69)
\end{aligned}$$

**(f) Computation of the Fourth Order Moments Matrix $\Omega_{i,pp}^*$ :**

The computation for this matrix requires the formulas for (i) $\mathrm{cov}[\{Y_{iu}Y_{it}, Y_{iu}Y_{it}\}|\rho = 0]$, (ii) $\mathrm{cov}[\{Y_{iu}Y_{it}, Y_{iu}Y_{im}\}|\rho = 0]$, (iii) $\mathrm{cov}[\{Y_{iu}Y_{it}, Y_{iv}Y_{it}\}|\rho = 0]$, and (iv) $\mathrm{cov}[\{Y_{iu}Y_{it}, Y_{iv}Y_{im}\}|\rho = 0]$. The computations for all these four covariances are similar. We, for example, give the formulas for covariances in (i) and (iv).

(i) Formula for $\mathrm{cov}[\{Y_{iu}Y_{it}, Y_{iu}Y_{it}\}|\rho = 0]$ :

$$\begin{aligned}
\mathrm{cov}[\{Y_{iu}Y_{it}, Y_{iu}Y_{it}\}|\rho = 0] &= E[\{Y_{iu}^2 Y_{it}^2\}|\rho = 0] \\
-\left[\frac{1}{W}\sum_{w=1}^{W}\right.&\left.\left(\exp[\{x_{it} + x_{im}\}'\beta + 2\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right)\right]^2, \qquad (70)
\end{aligned}$$

where $E[\{Y_{iu}^2 Y_{it}^2\}|\rho = 0]$ is computed in (65).

(iv) Formula for $\text{cov}[\{Y_{iu}Y_{it}, Y_{iv}Y_{im}\}|\rho = 0]$ :

$$\text{cov}[\{Y_{iu}Y_{it}, Y_{iv}Y_{im}\}|\rho = 0]$$
$$= \frac{1}{W}\sum_{w=1}^{W}\left[\exp[\{x_{iu} + x_{it} + x_{iv} + x_{im}\}'\beta + 4\left\{R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right\}]\right]$$
$$-\mu_{iu}(\beta, \lambda_\gamma, \nu)\mu_{it}(\beta, \lambda_\gamma, \nu)\mu_{iv}(\beta, \lambda_\gamma, \nu)\mu_{im}(\beta, \lambda_\gamma, \nu). \tag{71}$$

### 3.2.2 Asymptotic Properties of the GQL Estimator $\hat{\pi}_{GQL} = [\hat{\lambda}_{\gamma,GQL}, \hat{\nu}_{GQL}]' : 2 \times 1$

Notice that when $\Omega_i^*(\beta, \lambda_\gamma, \nu)$ from Sect. 3.2.1 is used in (49) for $\Omega_i(\beta, \lambda_\gamma, \nu, \rho)$, one solves the approximate GQL estimating equation

$$\sum_{i=1}^{K}\frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi}\Omega_i^{*-1}(\beta, \lambda_\gamma, \nu)\left(g_i - \eta_i(\beta, \lambda_\gamma, \nu, \rho)\right) = 0, \tag{72}$$

for $\pi = (\lambda_\gamma, \nu)'$.

Let $\hat{\pi}_{GQL} = [\hat{\lambda}_{\gamma,GQL}, \hat{\nu}_{GQL}]' : 2 \times 1$ be the solution of (72). By similar calculations as in Sect. 3.1.1 (see (44)), it can be shown that

$$\lim_{K\to\infty} \hat{\pi}_{GQL} \to N(\pi, {Q^*}_K^{-1}(\beta, \lambda_\gamma, \nu, \rho)), \tag{73}$$

where

$$Q_K^* = \left[\sum_{i=1}^{K}\frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi}\Omega_i^{*-1}(\beta, \lambda_\gamma, \nu)\frac{\partial \eta_i(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi'}\right]^{-1}$$
$$\times \sum_{i=1}^{K}\frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi}\Omega_i^{*-1}(\beta, \lambda_\gamma, \nu)\Omega_i(\beta, \lambda_\gamma, \nu, \rho)\Omega_i^{*-1}(\beta, \lambda_\gamma, \nu)\frac{\partial \eta_i(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi'}$$
$$\times \left[\sum_{i=1}^{K}\frac{\partial \eta_i'(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi}\Omega_i^{*-1}(\beta, \lambda_\gamma, \nu)\frac{\partial \eta_i(\beta, \lambda_\gamma, \nu, \rho)}{\partial \pi'}\right]^{-1}. \tag{74}$$

## 3.3 *Moment Estimation of Correlation Index Parameter $\rho$*

Recall from (33) that

$$E[(Y_{iu} - \mu_{iu}(\cdot))(Y_{it} - \mu_{it}(\cdot))] = \rho^{t-u}\mu_{iu}(\beta, \lambda_\gamma, \nu)$$
$$+ \exp\left[\{x_{iu} + x_{it}\}'\beta\right]\frac{1}{W}\sum_{w=1}^{W}\exp\left[2R(\psi_{iw}, \xi_{iw}^2; \lambda_\gamma, \nu)\right] - \mu_{iu}(\beta, \lambda_\gamma, \nu)\mu_{it}(\beta, \lambda_\gamma, \nu).$$

Consequently, by using lag 1 based pair-wise product responses, one obtains

$$E\sum_{i=1}^{K}\sum_{t=1}^{T-1}\frac{\{(Y_{it}-\mu_{it}(\cdot))(Y_{i,t+1}-\mu_{i,t+1}(\cdot))\}}{K(T-1)} = \rho\sum_{i=1}^{K}\sum_{t=1}^{T-1}\frac{\mu_{it}(\cdot)}{K(T-1)}$$
$$+\frac{1}{KW}\sum_{i=1}^{K}\sum_{w=1}^{W}\{\exp\left[2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right]\}\frac{\{\sum_{t=1}^{T-1}\exp\left[\{x_{iu}+x_{it}\}'\beta\right]\}}{T-1}$$
$$-\sum_{i=1}^{K}\sum_{t=1}^{T-1}\frac{\mu_{it}(\cdot)\mu_{i,t+1}(\cdot)}{K(T-1)}. \tag{75}$$

Further, one writes

$$E\sum_{i=1}^{K}\sum_{t=1}^{T}\frac{\{(Y_{it}-\mu_{it}(\cdot))^2\}}{KT} = \sum_{i=1}^{K}\sum_{t=1}^{T}\frac{\sigma_{i,tt}(\cdot)}{KT}, \tag{76}$$

where the variance $\sigma_{i,tt}(\cdot)$ has the formula given by (28).

Now by dividing (75) by (76), and using first order approximation, one obtains the unbiased moment estimator $\hat{\rho}_M$ for $\rho$ as

$$\hat{\rho}_M \simeq \left[\frac{\sum_{i=1}^{K}\sum_{t=1}^{T-1}\{(Y_{it}-\mu_{it}(\cdot))(Y_{i,t+1}-\mu_{i,t+1}(\cdot))\}/\{K(T-1)\}}{\sum_{i=1}^{K}\sum_{t=1}^{T}\{(Y_{it}-\mu_{it}(\cdot))^2\}/\{KT\}}\right]$$
$$\div\left[\frac{\sum_{i=1}^{K}\sum_{t=1}^{T-1}\mu_{it}(\cdot)/\{K(T-1)\}}{\sum_{i=1}^{K}\sum_{t=1}^{T}\sigma_{i,tt}(\cdot)/\{KT\}}\right]$$
$$-\left[\frac{1}{KW}\sum_{i=1}^{K}\sum_{w=1}^{W}\{\exp\left[2R(\psi_{iw},\xi_{iw}^2;\lambda_\gamma,\nu)\right]\}\frac{\{\sum_{t=1}^{T-1}\exp\left[\{x_{iu}+x_{it}\}'\beta\right]\}}{T-1}\right]$$
$$\div\left[\sum_{i=1}^{K}\sum_{t=1}^{T-1}\frac{\mu_{it}(\cdot)}{K(T-1)}\right]+\left[\sum_{i=1}^{K}\sum_{t=1}^{T-1}\frac{\mu_{it}(\cdot)\mu_{i,t+1}(\cdot)}{K(T-1)}\right]\div\left[\sum_{i=1}^{K}\sum_{t=1}^{T-1}\frac{\mu_{it}(\cdot)}{K(T-1)}\right]. \tag{77}$$

Under some regularity conditions on the covariates so that var$[\hat{\rho}_M]$ is bounded by a finite quantity, it follows that the moment estimator $\hat{\rho}_M$ is consistent for $\rho$. This is mainly because $\hat{\rho}_M$ given by (77) is approximately unbiased for $\rho$.

# 4 Binary Dynamic Mixed Logit Model with $t_\nu$ Random Effects

Recall the binary dynamic mixed logit (BDML) model given in (9), that is,

$$Pr(y_{it} = 1|\gamma_i, y_{i,t-1}) = \begin{cases} \frac{\exp(x_{i1}^{'}\beta+\sigma_\gamma\gamma_i)}{1+\exp(x_{i1}^{'}\beta+\sigma_\gamma\gamma i)} & \text{for } t = 1 \\ \frac{\exp(x_{it}^{'}\beta+\theta y_{i,t-1}+\sigma_\gamma\gamma_i)}{1+\exp(x_{it}^{'}\beta+\theta y_{i,t-1}+\sigma_\gamma\gamma_i)} & \text{for } t = 2, \ldots, T, \end{cases}$$

Under the normality assumption for the random effects, i.e., when $\gamma_i \overset{iid}{\sim} N(0, 1)$, the basic properties such as unconditional mean, variance and correlations under such BDML model is given by (13)–(17). In the following subsection, we provide these properties for the BDML model under the assumption that the random effects now follow a $t$-distribution with $\nu$ degrees of freedom.

## *4.1 Basic Properties of the Binary Mixed Model: Unconditional Mean and Variance*

By similar calculations as in the normal case (13), one obtains an approximate unconditional mean based on $t_\nu$ random effects, as

$$\begin{aligned} E[Y_{it}] &= \mu_{it}(\beta,\ \theta,\ \lambda_\gamma,\ \nu) = W^{-1}\sum_{w}^{W}\pi_{it}^{*}(\psi_{iw}, \xi_{iw}^{2}) \\ &= W^{-1}\sum_{w=1}^{W}\Big[p_{it0}(\psi_{iw},\ \xi_{iw}^{2}) + \pi_{i,t-1}^{*}(\psi_{iw}, \xi_{iw}^{2})\left\{p_{it1}(\psi_{iw}, \xi_{iw}^{2}) - p_{it0}(\psi_{iw}, \xi_{iw}^{2})\right\}\Big], \end{aligned} \quad (78)$$

where

$$\pi_{i1}^{*}(\psi_{iw}, \xi_{iw}^{2}) = p_{i10}(\psi_{iw}, \xi_{iw}^{2}) = \frac{\exp(x_{i1}^{'}\beta + \gamma_{iw}^{*})}{1 + \exp(x_{i1}^{'}\beta + \gamma_{iw}^{*})}, \text{ and}$$

$$p_{ity_{i,t-1}}(\psi_{iw}, \xi_{iw}^{2}) = \frac{\exp(x_{it}^{'}\beta + \theta y_{i,t-1} + \gamma_{iw}^{*})}{1 + \exp(x_{it}^{'}\beta + \theta y_{i,t-1} + \gamma_{iw}^{*})},$$

with $\gamma_{iw}^{*}$ as the $t_\nu(0, \lambda_\gamma^2, \nu)$ random effect given by (25). Next because $y_{it}$ is a binary observation, it follows that

$$\text{var}[Y_{it}] = \sigma_{i,tt}(\beta,\ \theta,\ \lambda_\gamma,\ \nu) = \mu_{it}(\beta,\ \theta,\ \lambda_\gamma,\ \nu)[1 - \mu_{it}(\beta,\ \theta,\ \lambda_\gamma,\ \nu)], \quad (79)$$

where the unconditional mean $\mu_{it}(\beta,\ \theta,\ \lambda_\gamma,\ \nu)$ has the recursive type formula as in (78).

### 4.2 Computation of Unconditional Covariances for BDML Model with $t_\nu$ Random Effects

To compute the covariance between $y_{iu}$ and $y_{it}$ $(u < t)$, we note that under the present dynamic model (9), conditional on the random effects $\gamma_i^*$ defined by (25), $y_{iu}$ and $y_{it}$ are not independent. This is because conditional on $\gamma_i^*$, $y_{it}$ and $y_{i,t-1}$, for example, satisfy the dynamic dependence relationship (9). Next because

$$\begin{aligned} E[\{Y_{iu}Y_{it}\}|\gamma_i^*] &= \text{cov}[\{Y_{iu}, Y_{it}\}|\gamma_i^*] + E[Y_{iu}|\gamma_i^*]E[Y_{iT}|\gamma_i^*] \\ &= \sigma^*_{i,ut}(\psi_i, \xi_i^2) + \pi^*_{iu}(\psi_i, \xi_i^2)\pi^*_{it}(\psi_i, \xi_i^2), \end{aligned}$$

with

$$\sigma^*_{i,ut}(\psi_i, \xi_i^2) = \pi^*_{iu}(\psi_i, \xi_i^2)[1 - \pi^*_{iu}(\psi_i, \xi_i^2)]\Pi_{j=u+1}^t \left[p_{ij1}(\psi_i, \xi_i^2) - p_{ij0}(\psi_i, \xi_i^2)\right], \tag{80}$$

(Sutradhar and Farrell 2007), one may compute the covariance between $y_{iu}$ and $y_{it}$, first by computing $E[Y_{iu}Y_{it}]$ using

$$E[Y_{iu}Y_{it}] = W^{-1}\sum_{w=1}^{W}\left[\sigma^*_{i,ut}(\psi_{iw}, \xi_{iw}^2) + \pi^*_{iu}(\psi_{iw}, \xi_{iw}^2)\pi^*_{it}(\psi_{iw}, \xi_{iw}^2)\right] = \tau_{i,ut}, \text{ (say)}, \tag{81}$$

where $\psi_{iw}$ and $\xi_{iw}^2$ are generated from $N(0, 1)$ and $\chi_4^2$, respectively, in order to compute $\gamma_{iw}^*$ by (25), and $\pi_{it}^*(\psi_{iw}, \xi_{iw}^2)$ is computed by (78).

## 5 GQL Estimation for the Parameters of the BDML Model with $t_\nu$ Random Effects

It is clear from Sects. 4.1 and 4.2 that the basic properties of the BDML (binary dynamic mixed logit) model (9), that is, the first and second order moments of the repeated binary responses contain all four parameters, namely $\beta$, $\theta$, $\lambda_\gamma$, and $\nu$, of the model. Consequently, we exploit all first and second order observations, and minimize their generalized distance from their corresponding means to construct a GQL estimating equations [Sutradhar (2010, Sect. 5.4); see also Sutradhar (2011, Sect. 9.2)] for these desired parameters. Note that for the binary data, $y_{it}^2 \equiv y_{it}$. One may, thus, consider a vector of first and second order responses given by

$$v_i = (y_{i1}, \ldots, y_{it}, \ldots, y_{iT}, y_{i1}y_{i2}, \ldots, y_{iu}y_{it}, \ldots, y_{i(T-1)}y_{iT})',$$

for the purpose of constructing the desired estimating equation. Now, denote the $E[V_i]$ by

$$\zeta_i = E[V_i] = \eta_i(\beta, \theta, \lambda_\gamma, \nu) = [\mu_{i1}, \ldots, \mu_{it}, \ldots, \mu_{iT}, \tau_{i,12}, \ldots, \tau_{i,ut}, \ldots, \tau_{i,(T-1)T}]' \\ = [\mu_i', \tau_i']', \tag{82}$$

where the formula for the unconditional mean $\mu_{it}$ for all $t = 1, \ldots, T$, is given by (78), and $\tau_{i,ut} = E[Y_{iu}Y_{it}]$ for all $u < t$, may be computed by (81). Further let $\alpha = (\beta, \theta, \lambda_\gamma, \nu)'$, and $\Omega_i$ denote the $T(T+2)/2 \times T(T+1)/2$ covariance of $v_i$. Following Sutradhar (2010) [see also Sutradhar (2004)], one may then write the GQL estimating equation for $\alpha$ as

$$\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \alpha} \Omega_i^{-1}(v_i - \zeta_i) = 0, \tag{83}$$

which may be solved by using the iterative equation

$$\hat{\alpha}_{GQL}(r+1) = \hat{\alpha}_{GQL}(r) + \left[ \left\{ \sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \alpha} \Omega_i^{-1} \frac{\partial \zeta_i}{\partial \alpha'} \right\}^{-1} \sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \alpha} \Omega_i^{-1}(v_i - \zeta_i) \right]_{|\alpha = \hat{\alpha}_{GQL}(r)} . \tag{84}$$

Note that to compute the $\Omega_i$ matrix for (83) and (84) , one needs to compute the following elements: **(a)** var$[Y_{it}]$; **(b)** cov$[Y_{iu}, Y_{it}]$; **(c)** var$[Y_{iu}Y_{it}]$; **(d)** cov$[Y_{iu}Y_{it}, Y_{i\ell}Y_{im}]$; and **(e)** cov$[Y_{iu}, Y_{im}Y_{it}]$. However, all these elements through **(a)**–**(e)**, may be computed by using the moments up to order four given in Sects. 4.1, 4.2, and 5.1 below. For example,

$$\text{cov}[Y_{iu}Y_{it}, Y_{i\ell}Y_{im}] = E[Y_{iu}Y_{it}Y_{i\ell}Y_{im}] - E[Y_{iu}Y_{it}]E[Y_{i\ell}Y_{im}] \\ = \tilde{\phi}_{i,ut\ell m} - \tau_{i,ut}\tau_{i,\ell m}, \tag{85}$$

where the formula for $\tilde{\phi}_{i,ut\ell m}$ is given in (96), and the formula for $\tau_{i,ut}$, for example, is given in (81).

**Computation of $\frac{\partial \zeta_i'}{\partial \alpha}$** for (84):

Because $\zeta_i = [\mu_i', \tau_i']'$, the gradients for searching for the estimate of $\alpha = (\beta', \theta, \lambda_\gamma, \nu)'$ can be computed by by using the formulas for $\frac{\partial \mu_{it}}{\partial \alpha}$ and $\frac{\partial \tau_{i,ut}}{\partial \alpha}$, where $\mu_{it}$ and $\tau_{i,ut}$ are given by (78) and (81), respectively. For the purpose, we derive these formulas by using

$$\frac{\partial \mu_{it}}{\partial \alpha} = W^{-1} \sum_{w=1}^{W} \Bigg[ \frac{\partial p_{it0}(\psi_{iw}, \xi_{iw}^2)}{\partial \alpha} + \{\frac{\partial \pi_{i,t-1}^*(\psi_{iw}, \xi_{iw}^2)}{\partial \alpha}\} \{p_{it1}(\psi_{iw}, \xi_{iw}^2) - p_{it0}(\psi_{iw}, \xi_{iw}^2)\} \\ + \pi_{i,t-1}^*(\psi_{iw}, \xi_{iw}^2) \left\{ \frac{\partial p_{it1}(\psi_{iw}, \xi_{iw}^2)}{\partial \alpha} - \frac{\partial p_{it0}(\psi_{iw}, \xi_{iw}^2)}{\partial \alpha} \right\} \Bigg], \tag{86}$$

$$\frac{\partial \tau_{i,ut}}{\partial \alpha} = W^{-1} \sum_{w=1}^{W} \frac{\partial \left[\sigma_{i,ut}^*(\psi_{iw}, \xi_{iw}^2) + \pi_{iu}^*(\psi_{iw}, \xi_{iw}^2)\pi_{it}^*(\psi_{iw}, \xi_{iw}^2)\right]}{\partial \alpha}, \tag{87}$$

where

$$\sigma^*_{i,ut}(\psi_{iw}, \xi^2_{iw}) = \pi^*_{iu}(\psi_{iw}, \xi^2_{iw})[1 - \pi^*_{iu}(\psi_{iw}, \xi^2_{iw})]\Pi^t_{j=u+1}\left[p_{ij1}(\psi_{iw}, \xi^2_{iw}) - p_{ij0}(\psi_{iw}, \xi^2_{iw})\right].$$

Note that to compute the derivative of the product factor involved in $\sigma^*_{i,ut}(\cdot)$, one can use the formula

$$\begin{aligned} & \frac{\partial \Pi^t_{j=u+1}\left[p_{ij1}(\psi_{iw}, \xi^2_{iw}) - p_{ij0}(\psi_{iw}, \xi^2_{iw})\right]}{\partial \alpha} \\ = \; & \Pi^t_{j=u+1}\left[p_{ij1}(\psi_{iw}, \xi^2_{iw}) - p_{ij0}(\psi_{iw}, \xi^2_{iw})\right] \\ & \times \sum_{j=u+1}^{t} \frac{\partial log\;\left[p_{ij1}(\psi_{iw}, \xi^2_{iw}) - p_{ij0}(\psi_{iw}, \xi^2_{iw})\right]\}}{\partial \alpha}. \end{aligned} \quad (88)$$

To complete the formulation of the above derivatives, we now give the derivatives for one term, namely $p_{ij1}(\psi_{iw}, \xi^2_{iw})$, with respect to each element of $\alpha = (\beta', \theta, \lambda_\gamma, \nu)'$. To be specific,

$$\frac{\partial p_{ij1}(\psi_{iw}, \xi^2_{iw})}{\partial \beta} = p_{ij1}(\psi_{iw}, \xi^2_{iw})(1 - p_{ij1}(\psi_{iw}, \xi^2_{iw}))x_{ij}; \quad (89)$$

$$\frac{\partial p_{ij1}(\psi_{iw}, \xi^2_{iw})}{\partial \theta} = p_{ij1}(\psi_{iw}, \xi^2_{iw})(1 - p_{ij1}(\psi_{iw}, \xi^2_{iw})); \quad (90)$$

$$\begin{aligned} \frac{\partial p_{ij1}(\psi_{iw}, \xi^2_{iw})}{\partial \lambda_\gamma} = \; & p_{ij1}(\psi_{iw}, \xi^2_{iw})(1 - p_{ij1}(\psi_{iw}, \xi^2_{iw})) \\ & \times \left[\psi_{iw}\{2\nu\}^{\frac{1}{2}}\left[\sqrt{\nu}\left(\xi^2_{iw} - 4\right) + 2\nu\right]^{-\frac{1}{2}}\right];\text{ and} \end{aligned} \quad (91)$$

$$\begin{aligned} \frac{\partial p_{ij1}(\psi_{iw}, \xi^2_{iw})}{\partial \nu} = \; & p_{ij1}(\psi_{iw}, \xi^2_{iw})(1 - p_{ij1}(\psi_{iw}, \xi^2_{iw}))\sqrt{2}\lambda_\psi\psi_{iw} \\ & \times \left[\frac{1}{2}\{\nu\}^{-\frac{1}{2}}\left[\sqrt{\nu}\left(\xi^2_{iw} - 4\right) + 2\nu\right]^{-\frac{1}{2}}\right. \\ & \left. - \; \frac{1}{2}\{\nu\}^{\frac{1}{2}}\left[\sqrt{\nu}\left(\xi^2_{iw} - 4\right) + 2\nu\right]^{-\frac{3}{2}}[\frac{1}{2}\{\nu\}^{-\frac{1}{2}}(\xi^2_{iw} - 4) + 2]\right]. \end{aligned} \quad (92)$$

### 5.1 *Computation Higher Order Moments to Construct $\Omega_i$ in (84)*

Note that when the first and second order responses are used to construct distance functions for the estimation of the parameters $\beta,\ \theta,\ \lambda_\gamma,$ and $\nu$, one requires the third and fourth order moments which are used in the weight matrix to develop the estimating equations. The first (mean) and the second order moments are computed in (78), (79) and (81) using suitable close form expressions. For higher order such as the third and fourth order moments, it is convenient to compute them numerically (Sutradhar et al. 2008). To be specific, for the computation of the third order moments, let $\sum_{(y_{iu},y_{i\ell},y_{it})\ni s}$ indicates the summation over all binary variables in the sample space s that contain $T-3$ elements out of T elements except $y_{iu}, y_{il}, y_{it}$. one may then compute the third order moments as

$$E[Y_{iu}Y_{i\ell}Y_{it}] = P[y_{iu}=1, y_{i\ell}=1, y_{it}=1] \equiv \tilde{\delta}_{i,u\ell t} \tag{93}$$
$$= W^{-1}\sum_{w=1}^{W}\sum_{y_{iu},y_{il},y_{it}\ni s}\left[f(y_{i1}|\gamma^*_{iw})\Pi_{j=2}^{T}f(y_{ij}|y_{i,j-1},\gamma^*_{iw})\right]_{y_{iu}=1,y_{i\ell}=1,y_{it}=1}$$

where by (78)

$$f(y_{i1}|\gamma^*_{iw}) = [p_{i10}(\gamma^*_{iw})]^{y_{i1}}[1-p_{i10}(\gamma^*_{iw})]^{1-y_{i1}}$$
$$f(y_{ij}|y_{i,j-1},\gamma^*_{iw}) = [p_{ijy_{i,j-1}}(\gamma^*_{iw})]^{y_{ij}}[1-p_{ijy_{i,j-1}}(\gamma^*_{iw})]^{1-y_{ij}}. \tag{94}$$

After an algebra, one may simplify the third order moments in (93) as

$$\tilde{\delta}_{i,u\ell t} = W^{-1}\sum_{w=1}^{W}\sum_{y_{iu},y_{i\ell},y_{it}\ni s}\left[\tilde{p}_{i10}(y_{i1},\gamma^*_{iw})\Pi_{j=2}^{T}\tilde{p}_{ij1}(y_{ij},y_{i,j-1},\gamma^*_{iw})\right]_{y_{iu}=1,y_{i\ell}=1,y_{it}=1}, \tag{95}$$

with $\tilde{p}_{i10}(y_{i1},\gamma^*_{iw}) = \frac{\exp\{y_{i1}(x'_{i1}\beta+\gamma^*_{iw})\}}{1+\exp(x'_{i1}\beta+\gamma^*_{iw})}$, and $\tilde{p}_{ij1}(y_{ij},y_{i,j-1},\gamma^*_{iw}) = \frac{\exp\{y_{ij}(x'_{ij}\beta+\theta y_{i,j-1}+\gamma^*_{iw})\}}{1+\exp(x'_{ij}\beta+\theta y_{i,j-1}+\gamma^*_{iw})}$,

where $\gamma^*_{iw} \equiv \gamma^*_{iw}(\lambda_\gamma, \nu; \psi_{iw}, \xi^2_{iw})$ as defined by (25).

The computation for the fourth order moments is similar to that of the third order moments. Let $\sum_{(y_{iu},y_{i\ell},y_{im},y_{it})\ni s^*}$ indicates the summation over all binary variables in the sample space $s^*$ that contain $T-4$ elements out of T elements except $y_{iu}, y_{il}, y_{im}, y_{it}$. Now by implementing this summation, following (95), one writes the formula for the fourth order moments as

$$\begin{aligned} E[Y_{iu}Y_{i\ell}Y_{im}Y_{it}] = &\ W^{-1}\sum_{w=1}^{W}\sum_{y_{iu},y_{i\ell},y_{im},y_{it}\ni s^*}\left[\tilde{p}_{i10}(y_{i1},\gamma^*_{iw})\right. \\ &\times \left.\Pi_{j=2}^{T}\tilde{p}_{ij1}(y_{ij},y_{i,j-1},\gamma^*_{iw})\right]_{y_{iu}=1,y_{i\ell}=1,y_{im}=1,y_{it}=1} \\ = &\ \tilde{\phi}_{i,u\ell mt}, \text{ (say).} \end{aligned} \tag{96}$$

This completes the computation of all moments up to order four. These moments were exploited to construct the GQL estimating Eq. (84) for all parameters involved in the model, namely $\beta,\ \theta,\ \lambda_\gamma,$ and $\nu$.

### 5.2 Asymptotic Normality and Consistency of $\hat{\alpha}_{GQL}$

Following (83), for true $\alpha$, define

$$\bar{g}_K(\alpha) = \frac{1}{K}\sum_{i=1}^{K} g_i(\alpha) = \frac{1}{K}\sum_{i=1}^{K} \frac{\partial \zeta_i'}{\partial \alpha} \Omega_i^{-1}(v_i - \zeta_i), \tag{97}$$

where $v_1, \ldots, v_i, \ldots, v_K$ are independent to each other as they are collected from $K$ independent individuals, but they are not identically distributed because

$$v_i \sim (\zeta_i(\beta, \theta, \lambda_\gamma, \nu), \Omega_i(\beta, \theta, \lambda_\gamma, \nu)), \tag{98}$$

where the mean vectors in (82) and also the covariance matrices in (83) are different for different individuals.

Now one may derive the asymptotic distribution of $\hat{\alpha}_{GQL}$ by using the same technique as for the derivation of the asymptotic distribution of $\hat{\beta}_{GQL}$ given in Sect. 3.1.1 for the Poisson mixed model. Thus, it can be shown that

$$\lim_{K\to\infty} \hat{\alpha}_{GQL} \to N(\alpha, \tilde{V}_K^{-1}(\beta, \theta, \lambda_\gamma, \nu)), \tag{99}$$

or equivalently

$$||[\tilde{V}_K(\beta, \theta, \lambda_\gamma, \nu)]^{\frac{1}{2}}[\hat{\alpha}_{GQL} - \alpha]|| = O_p(\sqrt{p+3}), \tag{100}$$

where

$$\tilde{V}_K(\beta, \theta, \lambda_\gamma, \nu) = \sum_{i=1}^{K} \frac{\partial(\zeta_i)'}{\partial \alpha} [\Omega_i(\beta, \theta, \lambda_\gamma, \nu)]^{-1} \frac{\partial(\zeta_i)}{\partial \alpha'}.$$

This establishes the consistency of $\hat{\alpha}_{GQL}$ for $\alpha$.

## 6 Discussion

It has been assumed in various econometric studies for count and binary panel data that the distribution of the random effects involved in the model is unknown. This makes the estimation of the regression effects $\beta$ and dynamic dependence parameter $\rho$ under the Poisson dynamic mixed model, and the estimation of $\beta$ and the dynamic dependence parameter $\theta$ under the binary dynamic mixed model, very difficult. As a remedy, some authors such as Wooldridge (1999) and Montalvo (1997) developed certain estimation techniques those automatically remove the random effects from the model and estimate rest of the parameters, $\beta$ and $\rho$ in the Poisson case. However as demonstrated by Sutradhar et al. (2014), these estimation approaches have two drawbacks. First, the conditional maximum likelihood (CML) method used by Wooldridge (1999) and the instrumental variables based GMM (IVGMM) method used by Montalvo (1997) become useless for the estimation of the regression effects $\beta$ when covariates are stationary (time independent), even though they are able to remove the random effects. Second, when the random effects $\gamma_i$ or $\gamma_i^*$ are removed technically, the estimates of $\beta$ and the dynamic dependence parameter ($\rho$) alone are not sufficient to compute the mean, variance and correlations of the data, which is a major drawback from the view point of data understanding/analysis. One encounters similar problems with the weighted kernel likelihood approach of Honore and Kyriazidou (2000, p. 84) for the inferences in binary dynamic mixed logit models.

The aforementioned inference issues do not arise when one can assume a suitable distribution for the random effects. Because the random effects appear in the linear predictive function of the generalized linear model, many studies mainly in statistics literature have considered normality as a reasonable assumption for the distribution of the random effects. Thus, under the assumption that the random effects involved in the longitudinal mixed models follow $N(0, \sigma_\gamma^2)$, the GQL estimation of the regression effects $\beta$ and $\sigma_\gamma^2$ and moment estimation of the longitudinal correlation index parameter $\rho$ were developed by Sutradhar and Bari (2007), for example, for longitudinal count data, and by Sutradhar (2008) for longitudinal binary data. See also Breslow and Clayton (1993), Breslow and Lin (1995), Lin and Breslow (1996), Jiang (1998), and Sutradhar and Qu (1998).

However, in this paper we have provided an extension of the normal latent effects based longitudinal mixed models for count and binary data to the $t_\nu$ latent effects based models. The inference for these extended models have been complex not only because of an additional degrees of freedom parameter but also for the difficulty that unlike simulation of $N(0, 1)$ random effects in the Gaussian case, the simulation of $t_\nu(0, 1)$ is not possible when $\nu$ is unknown. In this paper we have resolved this issue through a new transformation which helps to generate data from a $t_4(0, 1)$ distribution for the purpose and then proceed for estimation of the $\nu$ parameter. In summary, we have developed a GQL estimation technique for the estimation of all parameters involved in the models including the degrees of freedom parameter.

**Acknowledgements** The authors thank a referee for comments and suggestions. The second author presented a part of this paper in the symposium as a part of his Key Note address part I. Special thanks go to the audience of the symposium for their feedback.

## References

Amemiya, T.: Advanced Econometrics. Harvard University Press, Cambridge, MA (1985)

Bartolucci, F., Nigro, V.: A dynamic model for binary panel data with unobserved heterogenity admitting a $\sqrt{n}$-consistent conditional estimator. Econometrica **78**, 719–733 (2010)

Breslow, N.E., Clayton, D.G.: Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. **88**, 9–25 (1993)

Breslow, N.E., Lin, X.: Bias correction in generalized linear models with a single component of dispersion. J. Am. Stat. Assoc. **82**, 81–92 (1995)

Cox, D.R.: The analysis of mulitivariate binary data. Appl. Stat. **21**, 113–120 (1972)

Freeland, R.K., McCabe, B.P.M.: Forecasting discrete valued low count time series. Int. J. Forecast. **20**, 427–434 (2004)

Heckman, J.J.: Statistical models for discrete panel data. In: McFadden, D., Manski, C.F. (eds.) Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge (1981)

Honore, B.E., Kyriazidou, E.: Panel data discrete choice models with lagged dependent variables. Econometrica **68**, 839–874 (2000)

Jiang, J.: Consistent estimators in generalized linear mixed models. J. Am. Stat. Assoc. **93**, 720–729 (1998)

Lee, Y., Nelder, J.A.: Hierarchical generalized linear models. J. R. Stat. Soc. B **58**, 619–678 (1996)

Lin, X., Breslow, N.E.: Bias correction in generalized linear mixed models with multiple components of dispersion. J. Am. Stat. Assoc. **91**, 1007–1016 (1996)

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

McDonald, D.R.: The local limit theorem: a historical perspective. J. Iran. Stat. Soc. **4**, 73–86 (2005)

Montalvo, J.G.: GMM estimation of count-panel data models with fixed effects and predetermined instruments. J. Bus. Econ. Stat. **15**, 82–89 (1997)

Sutradhar, B.C.: On the Characteristics function of multivariate student t-distribution. Can. J. Stat. **14**, 329–337 (1986)

Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. Stat. Sci. **18**, 377–393 (2003)

Sutradhar, B.C.: On exact quasilikelihood inference in generalized linear mixed models. Sankhya: Indian J. Stat. **66**, 261–289 (2004)

Sutradhar, B.C.: On auto-regression type dynamic mixed models for binary panel data. Metron **56**, 205–217 (2008)

Sutradhar, B.C.: Inferences in generalized linear longitudinal mixed models. Can. J. Stat. **38**, 174–196 (2010) [Special issue Ed. B.C. Sutradhar]

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

Sutradhar, B.C., Bari, W.: On generalized quasilikelihood inference in longitudinal mixed models for count data. Sankhya: Indian J. Stat. **69**, 671–699 (2007)

Sutradhar, B.C., Farrell, P.J.: On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. Sankhya B: Indian J. Stat. **69**, 448–467 (2007)

Sutradhar, B.C., Qu, Z.: On approximate likelihood inference in Poisson mixed model. Can. J. Stat. **26**, 169–186 (1998)

Sutradhar, B.C., Prabhakar Rao, R., Pandit, V.N.: Generalized method of moments versus generalized quasilikelihood inferences in binary panel data. Sankhya B: Indian J. Stat. **70**, 34–62 (2008)

Sutradhar, B.C., Jowaheer, V., Rao, R.P.: Remarks on asymptotic efficient estimation for regression effects in stationary and non-stationary models for panel count data. Br. J. Probab. Stat. **28**, 241–254 (2014)

Wooldridge, J.: Distribution-free estimation of some non-linear panel data models. J. Econ. **90**, 77–97 (1999)

Zhao, L.P., Prentice, R.L.: Correlated binary regression using a quadratic exponential model. Biometrika **77**, 642–648 (1990)

# Part II
# Spatial and/or Time Series Volatility Models with Applications

# Zero-Inflated Spatial Models: Application and Interpretation

**L.M. Ainsworth, C.B. Dean, and R. Joy**

**Abstract** Many environmental applications, such as species abundance studies, rainfall monitoring or tornado count reports, yield data with a preponderance of zero counts. Although standard statistical distributions may not fit these data, a large body of literature has been dedicated to methods for modeling zero-inflated data. One type of regression model for zero-inflated data is categorized as a mixture model. Mixture models postulate two types of zeros, represented using a latent variable, and model their probabilities separately. The latent classification of zeros may be of particular interest as it can provide important clues to physical characteristics associated with, for example, habitat suitability or resistance to disease or pest infestations. Different zero-inflated models can be developed depending on the biological and physical characteristics of the application at hand. Here, several zero-inflated spatial models are applied to a case study of spruce weevil (*Pissodes strobi*) infestations in a Sitka spruce tree plantation. The data illustrate the unique features distinguished by various models and show the importance of using expert knowledge to inform model structures that in turn provide insight into underlying biological processes driving the probability of belonging to the zero, *resistant*, component. For instance, one model focuses on individually resistant trees located among infested trees. Another focuses on clusters of resistant trees which are likely located in unsuitable habitats. We apply six models: a standard generalized linear model (GLM); an overdispersion model; a random effects zero-inflated model; a conditional autoregressive random effects model (CAR); a multivariate CAR (MCAR) model; and a model developed using discrete random effects to

L.M. Ainsworth (✉)
Department of Statistics and Actuarial Science, Simon Fraser University, 8888 University Drive, Burnaby, BC, Canada V5A 1S6
e-mail: lmainswo@stat.sfu.ca

C.B. Dean
Department of Statistical and Actuarial Science, University of Western Ontario, Western Science Centre - Room 262, 1151 Richmond Street, London, ON, Canada N6A 5B7
e-mail: cbdean@stats.uwo.ca

R. Joy
SMRU Consulting Ltd., Suite 510, 1529 West 6th Ave., Vancouver, BC, Canada V6J 1R1
e-mail: rj@smruconsulting.com



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_3

accommodate spatial outliers. We discuss the distinct features identified by the zero-inflated spatial models and make recommendations regarding their application in general.

**Keywords** Autocovariate • Discrete mixture model • Hierarchical Bayesian model • Mixed binomial model • Spatial autocorrelation • Spruce weevil

## 1 Introduction

Biological phenomena and physical characteristics in the environment often give rise to ecological and environmental data with a large number of zeros. Thus the data may not fit standard statistical distributions such as the Poisson or binomial distribution. Further, the data often exhibit spatial correlation such as with animal abundance studies carried out over a range of habitat suitability types. Spatial correlation may arise in the positive counts, in the zero counts, or between the two. The underlying biology, physical characteristics of the sample space and the study design will all influence the correlations in the data. For instance, areas of unsuitable habitat are expected to create spatial clusters of zero counts. On the other hand, spatial outliers may occur when disease resistance gives rise to zero counts in the vicinity of very large counts termed *spatial outliers* (Ainsworth and Dean 2008, see).

Zero-inflated models were originally developed as a means to address the statistical problem of overdispersion when it arises from many zero counts. However, their structure can be utilized to gain biological insight and to address practical problems that impact parameter estimation. There are a wide variety of zero-inflated models to choose from. For any given application, expert knowledge is essential to inform the best model choice. If the spatial correlation is expected to be smooth over the space, a CAR model may be the most appropriate. On the other hand, isolated zero counts may create a large negative spike in an otherwise smooth spatial surface. In this case, adding a discrete random effect can accommodate the spatial outliers. Identification of potentially unsuitable habitat and spatial outliers is useful for postulating biological influences on abundance and/or resistance.

Here we briefly review zero-inflation models as they relate to a case study that is selected to demonstrate how different structures on the zero counts can highlight distinctive features of the data. The dataset selected for analysis is one of spruce weevil (*Pissodes strobi*) infestations monitored over several years in a spruce tree plantation. The spruce weevil produces one generation of larval offspring per year. The larvae cause extensive damage by feeding as a group on the tree top, forming a ring around the principal stem. This feeding action cuts off the water flow, causing the deformity and death of the stem. Repeated infestations in successive years may stunt the growth of trees and may also cause deformed or forked trunks, a serious problem for Christmas tree growers as well as for silvicultural objectives. However, in a spruce tree plantation not all trees are affected; some trees simply don't have eggs hatch on them and therefore experience no larvae damage, while others exhibit resistance to larvae infestation. Therefore, trees without infestations are a mixture of two sorts of trees with zero infestations observed because of either mechanism.

We apply six regression models to the spruce weevil dataset and compare biological insights and model performance of each. These models are: (1) a standard generalized linear model (GLM); (2) an overdispersion generalized linear model (Beta-binomial); (3) a zero-inflated model (ZIB); (4) a spatial zero-inflated conditional autoregressive random effects model (ZIB-CAR); (5) a zero-inflated multivariate CAR random effects model (ZIB-MCAR); and (6) a zero-inflated model with a discrete random effect to accommodate spatial outliers (ZIB-discrete). A particular focus is to compare methods. This case study provides a focal point for a discussion of the unique features of the zero-inflated spatial models developed here, provides a context for recommendations regarding applications, as well as ideas for future research.

## 2 Zero-Inflation Models

Overdispersion occurs when the observed variance of a random variable is larger than the variance of the theoretical distribution used to model the data. It is important to distinguish between true overdispersion and apparent overdispersion which may arise when the data contains outliers, the model is misspecified (using a linear relationship to model curvilinear trends) or the model does not include important covariates and/or interaction terms (Hardin et al. 2007). Apparent overdispersion can be handled by addressing the shortcomings in the model. True overdispersion can be handled in a variety of ways. These include the use of a scale parameter to inflate standard errors (McCullagh and Nelder 1989) or specialized distributions, for example the generalized Poisson distribution (Consul and Jain 1973) for the analysis of count data. Other approaches incorporate random effects (Wang et al. 2002; Breslow and Clayton 1993; Lawless 1987; Williams 1982), such as through the use of generalized linear mixed models (Breslow and Clayton 1993). However, these methods are generally only suitable for unimodal data.

When overdispersion is induced by an extra mass at zero, as often arises in ecological abundance studies which contain unsuitable habitat or survey rare species, a bimodal distribution may be created. In this situation, traditional methods for handling overdispersion may not be satisfactory. Martin et al. (2005) discuss the types of zeros arising in ecological studies and show that failing to account for zero-inflation may lead to underestimation of variance parameters. It seems prudent to exploit the special structure of a point mass at zero by modelling it explicitly. This can be done using a hurdle model or a mixture model. Hurdle models, also referred to as conditional or two-part models, use a single probability to model all the zeros. A zero mass is coupled with a truncated form of a standard distribution such as the binomial, Poisson or negative binomial (Heilbron 1994; Mullahy 1986). In the ecological setting, one interpretation is that the zeros represent unsuitable habitat while the conditional mean represents mean abundance given suitable habitat. The orthogonality of this parameterization allows for simple computation and interpretation of covariate effects. However, the hurdle model does not provide a convenient way to estimate separately the probability of a zero arising from the zero mass and the probability of the zero arising under the Poisson or binomial rate. The focus of the discussion herein is on mixture models.

Ridout et al. (1998), Martin et al. (2005) and Kuhnert et al. (2005) make the distinction between different types of zeros in the ecological setting: structural zeros which arise due to individual immunity or unsuitable habitat, and random zeros which arise by chance when populations are small or when sighting probabilities are low. Thus, mixture models provide a means for exploring the underlying biological mechanisms associated with the zeros. The zero-inflated mixture model (Lambert 1992), uses a latent indicator variable to mix a zero-mass with a standard distribution and provides a convenient way to model the two types of zeros separately.

Many authors extend zero-inflated models to account for correlation structures arising in, for example, longitudinal, clustered or spatial data (Dobbie and Welsh 2001; Hall 2000; Hall and Zhang 2004; Liang and Zeger 1986). In many applications, zero-inflated count data are spatially oriented. Rathbun and Fei (2006) develop a zero-inflated Bayesian Poisson model with excess zeros generated by a spatial probit model. Their model uses a single surface to model the spatially correlated zero component and the non-zero component of the model. On the other hand, Lawson and Clark (2002) recognize the limitations of using smooth surfaces to model spatial correlation. They develop a method that uses spatial mixtures of components to accommodate discontinuities in the spatial surface.

Agarwal et al. (2002) provide a comprehensive discussion of Bayesian methods for zero-inflated Poisson regression models. They discuss the issues of posterior propriety, informative prior specification, well behaved simulation-based model fitting and handling data with a large proportion of zeros, and devise techniques for proper prior specification of regression parameters.

## 3 Case Study: Pine Weevil Infestations

White pine weevil infestations in a 21,960 $m^2$ spruce tree plantation in British Columbia serve as a case study. The spruce trees in this plantation were part of an experiment to reduce pine weevil infestations. Therefore investigators were particularly interested in the areas with the highest resistance and trees that could remain resistant in highly infested areas. They inspected the $N = 2662$ susceptible trees for the presence of a weevil attack annually over 7 years [see Ainsworth and Dean (2007), for the data]. However, data were not necessarily available for all trees in all years. The counts have an upper bound of seven and are not rare enough to justify the Poisson approximation to the binomial distribution. Longitudinal trends in the infestation data are considered elsewhere (Nathoo and Dean 2007).

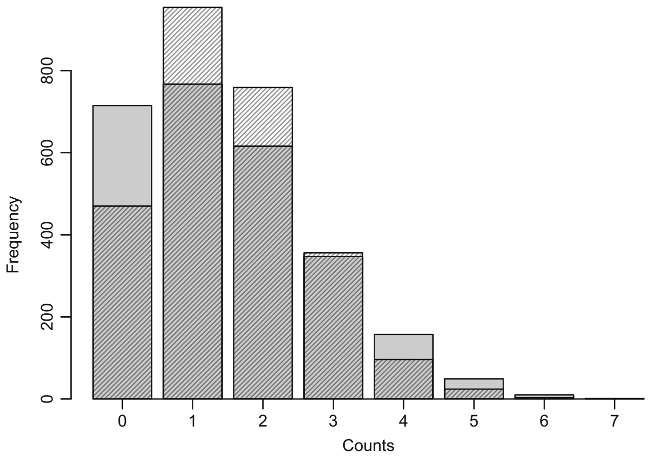



**Fig. 1** Histogram of infestation counts in *grey*. A binomial distribution with mean equal to the average infestation count is superimposed as *clear boxes with diagonal lines* (Color figure online)

Here the analysis focuses on the spatial patterns in the binomial data and interpretations derived from different mechanisms to handle zero-inflation.

The infestation counts are shown in Fig. 1. The data have a unimodal distribution with many trees having no infestations (715/2662 = 0.269) over the study period. The mechanisms giving rise to these zeros was of particular interest. Thus, mixture models were a natural parameterization for exploring the data. The zeros were modelled as arising from either (a) resistant trees or unsuitable habitat, or (b) susceptible trees. The first type of zero was assumed to arise from a distinct zero mass associated with resistance; the second type was assumed to arise from a binomial distribution. Zero-inflated binomial models developed here were applied to demonstrate the features distinguished by different models in the context of this study.

## 4 Model Specification

Here we investigate six models for count and proportion data: two non-inflated models, and four zero-inflated models. The GLM model and its basic overdispersed counterpart, the Beta-binomial model, do not incorporate zero-inflation or spatial correlation. The ZIB model, incorporates zero inflation but does not accommodate spatial correlation. The ZIB-CAR, ZIB-MCAR, and ZIB-discrete models are ZI spatial models that accommodate a variety of local spatial effects through the addition of random effects. The ZIB-CAR and ZIB-MCAR models assume a smooth spatial surface while the ZIB-discrete model uses discrete random effects to accommodate spatial clustering and isolated spikes on the spatial surface. Below, we briefly present the six models as they are applied to the case study.

1. *Standard Binomial Generalized Linear Model (GLM)*

Let $m_i$ be the number of years tree $i$ was observed, let $y_i$ be the number of years tree $i$ was infested, and let $N$ be the number of trees in the sample. Let $y_i, \ldots, y_N$ denote $N$ independent observations of the random variable $Y$, the number of years of larval infestation. The generalized linear model, assumes $Y_i \sim Bin(m_i, \mu_i)$ with expectation $E(Y_i) = m_i\mu_i$, variance $Var(Y_i) = m_i\mu_i(1 - \mu_i)$ and proportion of infestations $\mu_i$.

We further assume that the expected value $\mu_i$ relates to a set of $p$ covariates that take values $x_i' = (x_{i1}, ..., x_{ip})$ through a logit link function

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha_\mu + x_i'\beta$$

where $\beta$ is a vector of fixed, but unknown, regression parameters.

2. *Overdispersed Beta-Binomial Model (Beta-Binomial)*

When applying generalized linear models for the binomial distribution as in model 1, we expect the residual deviance to be approximately the same as the residual degrees of freedom. When there are many zeros observed, the residual deviance can be much greater and cause the estimates of $\beta$ to be less precise than expected. Thus, when we have overdispersion, reported standard errors may be too small (Stroup 2013).

To accommodate this extra variability, we allow for general overdispersion. Models that allow for overdispersion typically replace the mean-variance function of the original model by a more general form involving additional parameters. One way to allow for overdispersion, is to adopt a two-stage model as follows.

Assume $Y_i \sim Bin(m_i, P_i)$, where the $P_i$'s are random variables from a Beta distribution with mean $\mu_i$. Then, unconditionally, we have

$$E(Y_i) = m_i\mu_i, \quad \text{and}$$
$$Var(Y_i) = m_i\mu_i(1 - \mu_i)[1 + \phi(m_i - 1)]$$

where $\phi$ is an overdispersion parameter. This gives the beta-binomial distribution which assumes the response to be binomially distributed, and the $P_i$'s to be from a $Beta(\alpha^*, \beta^*)$ distribution with $\phi = \alpha^* + \beta^*$. The expected mean $\mu_i$ of the Beta-binomial model is related, as in the GLM model, to a set of covariates $x_i$ through a logit link function

$$g(\mu_i) = \log\left(\frac{\mu_i}{1 - \mu_i}\right) = \alpha_\mu + x_i'\beta.$$

These models accommodate general overdispersion but do not necessarily fit overdispersion induced by a point mass.

*3. Zero-Inflated Binomial Model (ZIB)*

Zero-inflated models explicitly handle overdispersion that arises from a point mass at zero. Mixture model formulations are useful when zeros arise from two processes; one binary process that describes a mechanism for observing zeros (e.g., resistance or unsuitable habitat), and another process that generates both zero and non-zero data such as counts, or in this case study, the proportion of years a tree is observed to have a larval infestation. It is important to realize that this model formulation mixes two components that represent the separate processes generating the zeros.

The zero-inflated model formulation is facilitated by the introduction of a latent variable, $Z_i$, which in the present context, measures resistance. Let $Z_i = 1$ if the observation arises from the zero mass, a distribution which generates zero with probability 1, and let $Z_i = 0$ if the observation arises from the alternate count or proportion distribution $f()$. In other words, if $Z_i = 0$, tree $i$ is considered non-resistant or susceptible and if $Z_i = 1$, tree $i$ is considered to be either individually resistant or to reside in unsuitable habitat.

A zero-inflated binomial regression model is given by

$$Y_i|Z_i = \begin{cases} Bin(m_i, \mu_i) & \text{if} \quad Z_i = 0 \\ 0 & \text{if} \quad Z_i = 1, \end{cases}$$

Thus, under this parameterization, $Z_i$ models the probability that a tree is resistant or in unsuitable habitat, and has a Bernoulli distribution $Z_i \sim Bern(\theta_i)$, and $E(Y_i \,|\, Z_i = 0) = m_i\mu_i$, and $Var(Y_i \,|\, Z_i = 0) = m_i\mu_i(1 - \mu_i)$. As in models 1 and 2, we use a link function, to relate the conditional means of these distributions to linear models as follows:

$$\text{logit}(\mu_i) = \alpha_\mu + x'_{\mu i}\beta_\mu$$

$$\text{logit}(\theta_i) = \alpha_\theta + x'_{\theta i}\beta_\theta,$$

where $x'_{\mu i}$, $x'_{\theta i}$ are vectors of covariates associated with the fixed effects, and $\beta_\mu$, $\beta_\theta$ their regression coefficients. This model explicitly accommodates a point mass at zero, but it does not accommodate spatial correlation amongst observations. The extensions we consider in the next three models account for spatial correlation and are implemented through the addition of random effects.

*4. Zero-inflated Conditional Autoregressive Model (ZIB-CAR)*

The CAR model specifies a set of spatially correlated random effects $b = (b_1, \ldots, b_N)$ where each $b_i$ is associated with a particular location (tree) in $\mathscr{R}^2$. A convenient distributional form for the random effects is the multivariate normal distribution with mean 0, and spatially structured covariance matrix, $\Sigma$.

The CAR model specifies $\Sigma$ indirectly through a set of conditional distributions. The $i = 1, \ldots, N$ conditional distributions are assumed to be univariate normal

$$b_i | b_{j \neq i} \sim N(u_i, \sigma_i^2) \quad i, j = 1, \ldots, N,$$

where $\sigma_i^2 > 0$ is the conditional variance, $u_i = \sum_{j \in \delta_i} w_{ij} b_j$, and $\delta_i$ is a set of neighbours for location $i$. The weights, $w_{ij} \geq 0$, $w_{ii} = 0$, can be based on the distance between locations $i$ and $j$. The unconditional distribution is,

$$b \sim N(0, (I - W)^{-1} M)$$

where $W = (w_{ij})$ and $M = diag(\sigma_1^2, \ldots, \sigma_N^2)$. In order to ensure symmetry of $\Sigma$, we require $w_{ij}\sigma_j^2 = w_{ji}\sigma_i^2$. Further, (I-W) must be invertible and $(I - W)^{-1}M$ must be positive definite. Besag et al. (1991) propose an intrinsic CAR model in which the covariance matrix $\Sigma$ is not positive definite. The weighting choice used with intrinsic autoregression is $w_{ij} = c_{ij}/c_i$, and $\sigma_i^2 = \sigma^2/c_i$ where $c_{ij}$ are user defined weights set so that $c_i = \sum_j c_{ij}$ with $c_{ij} = 1$ for neighbours, and 0 otherwise. In this case we can write the variance of the joint distribution of the random effects as $(I - W)^{-1}\sigma^2$.

The conditional specification of the CAR model facilitates Markov Chain Monte Carlo (MCMC) estimation. In particular, Gibbs sampling requires sampling from the full conditional distributions which are obtained from the joint posterior distribution. Full conditional distributions may not take a standard form, however log-concave distributions may be sampled via adaptive rejection sampling (Gilks and Wild 1992).

A zero-inflated spatial model is specified by introducing two independent CAR random effects into the two components of the ZIB model, i.e.,

$$\text{logit}(\mu_i) = \alpha_\mu + x'_{\mu i}\beta_\mu + b_{\mu i}$$
$$\text{logit}(\theta_i) = \alpha_\theta + x'_{\theta i}\beta_\theta + b_{\theta i},$$

where $b_{\mu i}$, $b_{\theta i}$ are random effects with variance $\sigma_{b_\mu}^2$ and $\sigma_{b_\theta}^2$. Note that we use two sets of random effects; one set is associated with the probability of membership in the resistant group, $\theta_i$, the other with the mean proportion for the non-resistant group $\mu_i$.

The intrinsic autocorrelation in this model imposes two independent smooth spatial structures, one for each model component. This model is similar to the model described in Agarwal et al. (2002). They use a Poisson distribution but provide a thorough discussion of these types of models and their theoretical properties.

*5. Zero-Inflated Multivariate CAR Model (ZIB-MCAR)*

An MCAR distribution (Jin et al. 2005) may be used to model the correlation between the random effects for the two components of the model ($b_\mu^m$ and $b_\theta^m$). This permits correlation between the two spatial processes.

Let $B_i = (b^m_{\mu i}, b^m_{\theta i})$ be a 2-dimensional vector of spatially correlated Gaussian random effects at site $i$. Here, the conditional MCAR distribution is:

$$B_i | B_{j \neq i} \sim N_2(\bar{B}_i, \Sigma / c_i) \tag{1}$$

where $N_2$ denotes the bivariate normal distribution, $\bar{B}_i$ is the mean of the spatial random effects corresponding to trees in the neighbourhood of the *ith* tree, and $\Sigma$ is the $2 \times 2$ positive definite, symmetric matrix that represents the conditional within-region covariance of the random effects. The diagonal elements of $\Sigma$, $\sigma^2_{b_\mu}$ and $\sigma^2_{b_\theta}$, represent the conditional variance parameters corresponding to $b^m_\mu$ and $b^m_\theta$ respectively, while the unrestricted off-diagonal elements represent the conditional correlation.

Similar to the CAR model, the ZIB model components are specified as follows:

$$\text{logit}(\mu_i) = \alpha_\mu + x'_{\mu i}\beta_\mu + b^m_{\mu i}$$
$$\text{logit}(\theta_i) = \alpha_\theta + x'_{\theta i}\beta_\theta + b^m_{\theta i},$$

The MCAR model, like the CAR model imposes a smooth spatial surface.

6. *Zero-Inflated Discrete Random Effects Model (ZIB-Discrete)*

Discrete random effects allow for discontinuities in the spatial surface and tend to be more flexible than those constrained to be normally distributed. They are particularly useful for accommodating spatial outliers (Ainsworth and Dean 2008). Model 6 uses discrete random effects for the zero component of the model and thus allows individual trees to have excessively elevated or deflated posterior probabilities of resistance. Let the discrete random effect, $d_i$, take one of $k$ values, $\log(R^1), \log(R^2), \ldots, \log(R^k)$, with probability $\gamma^1, \gamma^2, \ldots, \gamma^k$, respectively, where $\sum_{j=1}^k \gamma^j = 1$; define $\gamma = (\gamma^1, \gamma^2, \ldots, \gamma^k)$. Define $b = (b_1, \ldots, b_N)$ to be CAR random effects as in model 4 (ZIB-CAR). The ZI-discrete model is specified as:

$$\text{logit}(\mu_i) = \alpha_\mu + x'_{\mu i}\beta_\mu + b_i$$
$$\text{logit}(\theta_i) = \alpha_\theta + x'_{\theta i}\beta_\theta + d_i,$$

Under this model, the random effects associated with resistance are not necessarily spatially correlated. Rather they classify trees into resistance groups of, say, low, moderate, and high probability of resistance.

## 5 Model Implementation

He and Alfaro (1997) discuss neighbourhood definitions in the context of weevil infestations. These authors found spatial correlations attenuated at different distances depending on the stage of a weevil infection cycle. Exploratory analysis, including variograms for the total number of infestations, the number of infestations given at least one, and whether or not the tree was ever infested, as well as biological expertise, led to defining a neighbourhood as all trees within a radius of 6 m. Two trees were more than 6 m away from any neighbours; the closest neighbour was located at a distance of 6.1 and 11.3 m for these trees. These nearest neighbours were defined as the single neighbour for each of these trees. The number of neighbours ranges from 1 to 37 with a median of 15.

Model specification is completed by assigning prior distributions. A brief study of sensitivity to priors is discussed at the end of this section. We used uninformative priors throughout. The intercept terms, $\alpha_\mu$ and $\alpha_\theta$ were given flat priors, Uniform$(-\infty, \infty)$. For the Beta-binomial model, the parameters of the beta distribution were given Gamma (1,0.001) priors. Here, Gamma $(a, b)$ denotes the gamma probability density function with mean $a/b$ and variance $a/b^2$. For the CAR distribution, the precision parameters $1/\sigma^2_{b_\mu}$ and $1/\sigma^2_{b_\theta}$ were assigned a conjugate Gamma $(0.05, 0.05)$ prior. The precision matrix for the MCAR distribution was given a Wishart prior with matrix elements $(0.001, 0, 0, 0.05)$ and 2 degrees of freedom. For a detailed discussion of implementation issues related to priors, identifiability and convergence for spatial models, see Eberly and Carlin (2000) and Agarwal et al. (2002). The ZIB-discrete model, $\gamma$ was assigned a Dirichlet$(1, 1, \ldots 1)$ prior distribution. $R^1$ was assigned a $N(1, 10)$ prior, suitably standardized for truncation at zero. For computational stability, we used $R^{j+1} = R^j + \delta^j$ with $\delta^j, j = 1, 2, \ldots, k-1$, having zero truncated $N(0, 10)$ prior distributions. Here, $k = 3$ provided sufficient flexibility, and $R^1, R^2$, and $R^3$ reflect low, medium and high probabilities of resistance. See Chen and Knalili (2008) for a detailed discussion of the order of finite mixture models.

Posterior distributions were computed using the default Markov chain Monte Carlo algorithms within WINBUGS 1.4 (Spiegelhalter et al. 2003). All analyses were run using two parallel Markov chains with dispersed starting values. CAR prior distributions were fit using the Winbugs function car.nomral, and MCAR prior distributions were fit using the mv.car function. Due to the large lag in the autocorrelation of the parameter estimates in these sorts of studies, we thinned the samples to no less than every 20*th* sample. We used a burn-in of between 10,000 and 500,000 iterations with more complex models requiring longer burn-in and thinning. Posterior summaries were based on 2000 iterations, once convergence had been confirmed. Convergence was assessed via trace plots, Gelman and Rubin plots (Gelman and Rubin 1992) and Monte Carlo errors.

In order to fit the ZIB-CAR and ZIB-MCAR models we utilized the WinBUGs *zeros trick* (Lunn et al. 2012). This trick is useful for specifying specific sampling distributions and, in our experience, is useful for fitting complex zero-inflated models. However, it was difficult to obtain convergence for ZIB-MCAR model parameters. The choice of starting values for the intercept and the choice of priors placed on the variance components impacted model convergence. Thus, we carried out a sensitivity analysis to assess the stability of parameter estimates, as discussed in the following section.

# 6 Results

Table 1 presents transformed estimates of the intercept terms associated with the linear predictors. These yield mean estimates of the proportion of years a tree is infested and of the probability of resistance. The standard binomial model and the four models which account for spatial correlation provide very similar estimates of the proportion of years infested. Standard errors are smaller under the GLM model, as expected. The ZIB-CAR and ZIB-MCAR models account for spatial correlation in the zero classification through a CAR distribution; they provide similar estimates and standard errors of the proportion of zeros arising from the resistant group. Under the ZIB model, both the estimated proportion of infestations and especially the estimated proportion of resistant trees are larger than those obtained from other zero-inflated models. However, as we see later, neither the GLM nor the ZIB model has sufficient flexibility to fit this data well. Under the ZIB-discrete model, the odds of resistance takes values ($R^j, j = 1, 2, 3$), 0.024, 0.76 and 2.22, with probabilities ($\gamma^j, j = 1, 2, 3$), 0.77, 0.18 and 0.04, respectively. This model allows a few trees (4%) to have a very large probability (0.69) of resistance ($R^j/(1 + R^j), j = 1, 2, 3$).

Table 1 also shows that the estimated variance of the random effects associated with the non-resistant trees, $\sigma^2_{b_\mu}$, is quite similar for the ZIB-CAR and ZIB-MCAR models. The unique flexibility of the ZIB-discrete model to accommodate isolated resistant trees by a discrete random effect results in an estimated variability in the random effects associated with the non-resistant component being much smaller than that estimated under the ZIB-CAR and ZIB-MCAR models.

The estimates of the variability in the random effects associated with the resistant component, $\sigma^2_{b_\theta}$, (Table 1) differ across models. The estimated random effects $b_{\mu i}$ from the ZIB-CAR and ZIB-MCAR models are highly correlated ($r = 0.99$). On the other hand, estimated random effects $b_{\theta i}$ are only moderately correlated ($r = 0.58$) across models. Under the ZIB-CAR model, the estimates of the two sets of random effects, $b_{\mu i}$ and $b_{\theta i}$, have a moderate, negative correlation (r = –0.48). The ZIB-MCAR model, imposes strong correlation structure on the two sets of random effects and they exhibit an almost perfect negative correlation (r = –0.995).

**Table 1** Posterior mean estimates of proportion of years infested, $\frac{e^{\hat{\alpha_\mu}}}{1+e^{\hat{\alpha_\mu}}}$, proportion of resistant trees, $\frac{e^{\hat{\alpha}_\theta}}{1+e^{\hat{\alpha}_\theta}}$, and variability of random effects

| Model | $\frac{e^{\hat{\alpha_\mu}}}{1+e^{\hat{\alpha_\mu}}}$ (95 % CI) | $\frac{e^{\hat{\alpha}_\theta}}{1+e^{\hat{\alpha}_\theta}}$ (95 % CI) | $\hat{\sigma}^2_{b_\mu}$ (95 % CI) | $\hat{\sigma}^2_{b_\theta}$ (95 % CI) |
|---|---|---|---|---|
| GLM | 0.22 (0.209,0.221) | – | – | – |
| ZIB | 0.26 (0.250,0.265) | 0.16 (0.141,0.182) | – | – |
| ZIB-CAR | 0.22 (0.204,0.231) | 0.05 (0.017,0.095) | 1.98 (1.28, 2.89) | 3.03 (0.18, 9.67) |
| ZIB-MCAR | 0.22 (0.203,0.233) | 0.06 (0.019,0.112) | 1.95 (1.03, 2.93) | 0.43 (0.05, 0.93) |
| ZIB-discrete | 0.21 (0.204,0.226) | – | 1.46 (1.19, 1.73) | – |

CI is the credibility interval

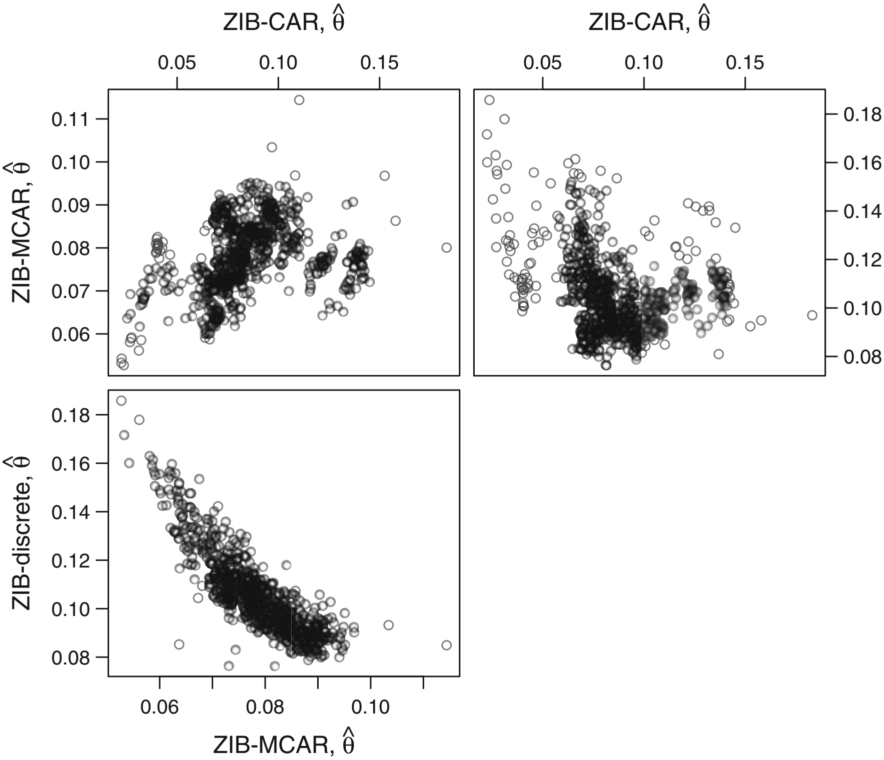



**Fig. 2** Correspondence of the posterior probability of resistance among the spatial models

Forcing such a correlation drastically reduces the variability in that component of the model for which less information is available for estimation. Here, and typically for zero-inflated models this is the resistant component (as reflected in $\sigma^2_{b_\theta}$).

Figure 2 highlights the relationship among the estimates of the posterior probabilities of resistance obtained from the ZIB-CAR, ZIB-MCAR and ZIB-discrete models. The ZIB-discrete model clearly identifies different features of the data. Further insight into the unique characteristics of ZIB-discrete model is obtained by considering the relationship between the posterior probability of resistance for infestation-free trees and two measures on neighbouring trees: mean proportion of years of infestations observed, and the proportion of trees which were never infested (Fig. 3). The posterior probability of resistance is weakly associated with these two measures under the ZIB-CAR model, but strongly associated under the ZIB-MCAR and ZIB-discrete models. Large posterior probabilities of resistance are associated with small infestation rates in neighbouring trees and having many non-infested neighbours under the ZIB-MCAR model.

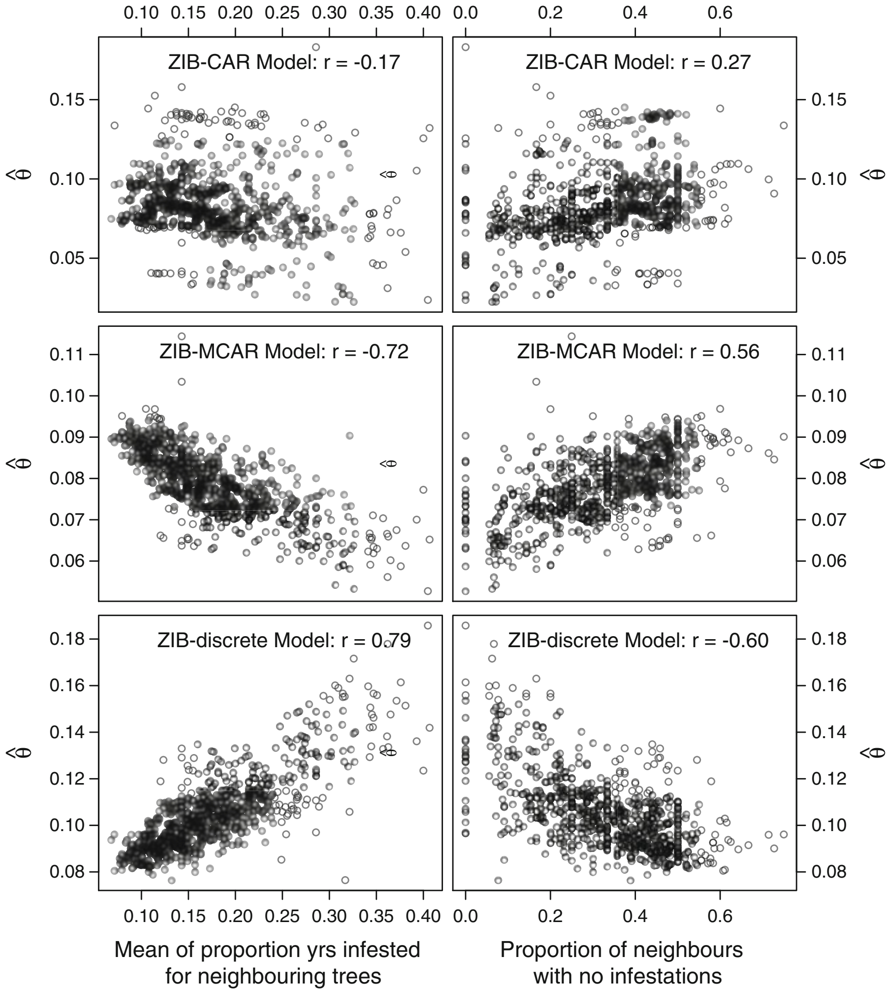



**Fig. 3** The relationship between the posterior estimate of resistance for trees with no infestations and measurements on neighbouring trees: mean proportion of infestations observed for neighbouring trees and the proportion of neighbouring trees which were never infested

In contrast, the ZIB-discrete model assigns large posterior probabilities of resistance to isolated resistant trees; those trees which are surrounded by many highly infested trees. For this model, spatial correlation in resistance is captured through the random effects in the non-resistant component, $b_{\mu i}$, which permit spatial clusters with very low means.

Figure 4a and 4b highlight the 100 trees with the largest estimated probabilities of resistance in light; the remaining non-infested trees are indicated in dark; and infested trees are indicated in black with larger circles indicating a larger proportion

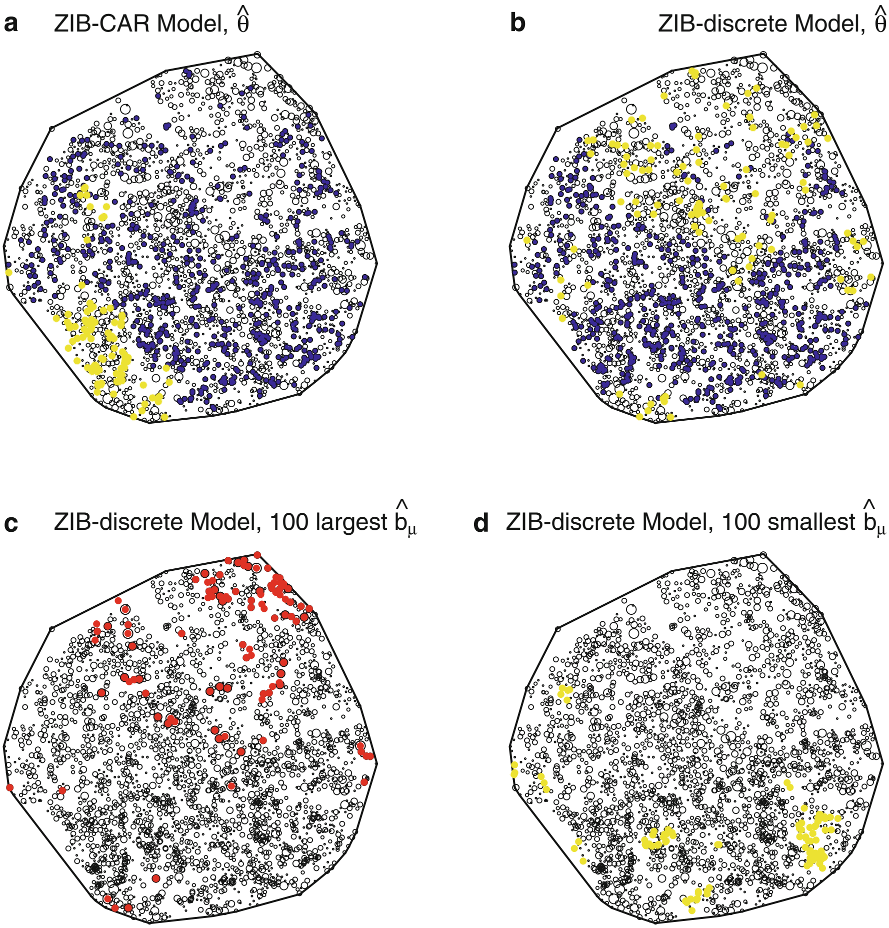



**Fig. 4** Largest and smallest posterior estimates of the probability of resistance (**a** and **b**) and the random effect associated with the non-resistant component (**c** and **d**). Infested trees are indicated in *black* with *larger circles* indicating a larger proportion of years infested. In **a** and **b**, trees with the largest probability of resistance are indicated in *yellow*; remaining non-infested trees are indicated in *blue*. The 100 trees with the largest (*red*) and smallest (*yellow*) estimates of the random effect associated with the non-resistant component are indicated in **c** and **d** (Color figure online)

of years infested. The ZIB-CAR model locates spatial clusters of resistant trees in the south-west corner of the plantation. The ZIB-discrete model identifies isolated resistant trees which are surrounded by trees with many infestations, especially in the north. This is seen more clearly by comparison with Fig. 4c (and Fig. 4d). In Fig. 4c red highlights indicate the 100 trees with the largest posterior estimates of the random effect associated with the non-resistant component—indicating sites with the largest estimated probability of infestation. In Fig. 4d yellow highlights indicate the 100 trees with the smallest estimates. The trees with the largest estimated random effects tend to be located in the northern corner of the plantation. Note that Fig. 4c, d present $\hat{b}_\mu$ from the ZIB-discrete model; those from the ZIB-CAR model are almost perfectly correlated with the estimates from the ZIB-discrete model and are not shown here.

Goodness-of-fit is assessed in a variety of ways. Standardized residuals (not shown here) for each model, defined as the posterior mean of $r_i = (y_i - E(y))/\sqrt{V(y)}$, indicate that the GLM and ZIB models tend to underestimate the largest infestation rates. Histograms of the posterior predictive values, $P(y_i < {y_i}^{rep})$ (Gelman et al. 1996), for each model (not shown here) reveal that the Beta-binomial model has very few extreme $p$-values. However, the $p$-values associated with the ZIB-CAR, ZIB-MCAR and ZIB-discrete models show a more uniform distribution as is expected with a good fit.

Table 2 displays goodness-of-fit measures obtained from posterior predictive distributions. The first column displays the number of trees observed to have 0, 1, 2, ... 7, infestations, $O_j$, $j = 0, 1, \ldots, 7$. Columns 3–8 display the posterior median number of trees, $E_j$, with $j$ infestations, $j = 0, 1, \ldots, 7$. The bottom of the table presents the GOF $= \sum_{j=1}^{7}(O_j - E_j)^2/E_j$, the $p$-value corresponding to a $\chi^2$ discrepancy measure based on the statistic, $T = (y_i - E(y_i|\mathbf{b}))^2/Var(y_i|\mathbf{b})$, (Gelman et al. 1996), and Johnson's (2004) goodness-of-fit statistic, $R^B$, which has a $\chi^2$ distribution if the model is accurate. The posterior mean of $R^B$, as well as the proportion of times $R^B$ exceeds the critical value of the $\chi^2$ distribution, are reported in the last two rows of Table 2.

*Goodness-of-fit* measures indicate that the simplest models, the GLM and ZIB models, provide poor fits to this data. Although the ZIB model regenerates approximately the correct number of zeros, it does not provide a good fit to the non-zero counts. Accounting for overdispersion beyond that arising from zero-inflation is necessary. The Beta-binomial model, accounts for overdispersion in general, but does not account for zero-inflation specifically. Surprisingly, it provides a reasonable fit to the data. This is likely due to the fact that the data are unimodal. If the mean of the non-zero counts was slightly larger or the proportion of zeros was slightly greater, the resulting bimodal distribution would not be accommodated by the Beta-binomial model. All of the spatial models are able to accommodate both the excess zeros and the overdispersion in the counts; they provide a reasonable fit to the data. In particular, the ZIB-CAR and ZIB-discrete models provide a good fit. The best choice amongst these models depends on the focus of the application. The ZIB-CAR model is well suited to identifying unsuitable habitat which manifests as clusters of zeros. The ZIB-discrete model is more flexible and can identify isolated zeros which may represent individual resistance or immunity.

*Sensitivity to prior distributions* are a key consideration. Gelman (2006) and Gelman and Hill (2007) note that the prior distribution used for the scale parameter needs to be selected carefully. A typical choice is an inverse-Gamma($\epsilon, \epsilon$) distribution for the variance parameters. Table 3 displays parameter estimates from a $2 \times 2$ factorial design used to test sensitivity to variance priors for the ZIB-CAR and ZIB-discrete models. For the ZIB-CAR model, the precision parameter, $\sigma_{b_\mu}^{-2}$

**Table 2** Classification of observed data and posterior median number of values generated for each category under each model

| Infestations | Observed | GLM | Beta-binomial | ZIB | ZIB-CAR | ZIB-MCAR | ZIB-discrete |
|---|---|---|---|---|---|---|---|
| 0 | 715 | 496 | 695 | 715 | 704 | 697 | 697 |
| 1 | 767 | 941 | 810 | 681 | 775 | 785 | 793 |
| 2 | 616 | 765 | 600 | 698 | 629 | 629 | 625 |
| 3 | 347 | 348 | 340 | 398 | 349 | 346 | 343 |
| 4 | 157 | 95 | 151 | 137 | 146 | 144 | 144 |
| 5 | 49 | 15 | 52 | 28 | 47 | 46 | 47 |
| 6 | 10 | 1 | 12 | 3 | 11 | 11 | 11 |
| 7 | 1 | 0 | 1 | 0 | 1 | 1 | 1 |
| $\text{GOF} = \sum_{j=1}^{7} \frac{(O_j - E_j)^2}{E_j}$ | | 375 | 4.17 | 67 | 1.54 | 2.61 | 2.84 |
| $p$-value for $\chi^2_{disc}$ | | 0.0000 | 0.44 | 0.0005 | 0.32 | 0.25 | 0.23 |
| $R^B$ | | 310.6 | 5.7 | 58.3 | 4.0 | 5.6 | 4.6 |
| $R^B$ exceeding $\chi^2_{6,0.95}$ | | 1.0 | 0.009 | 1.0 | 0.021 | 0.054 | 0.019 |

*Note:* GLM and ZIB Models have $E_7 = 0$. To avoid division by zero, categories 6 and 7 are collapsed

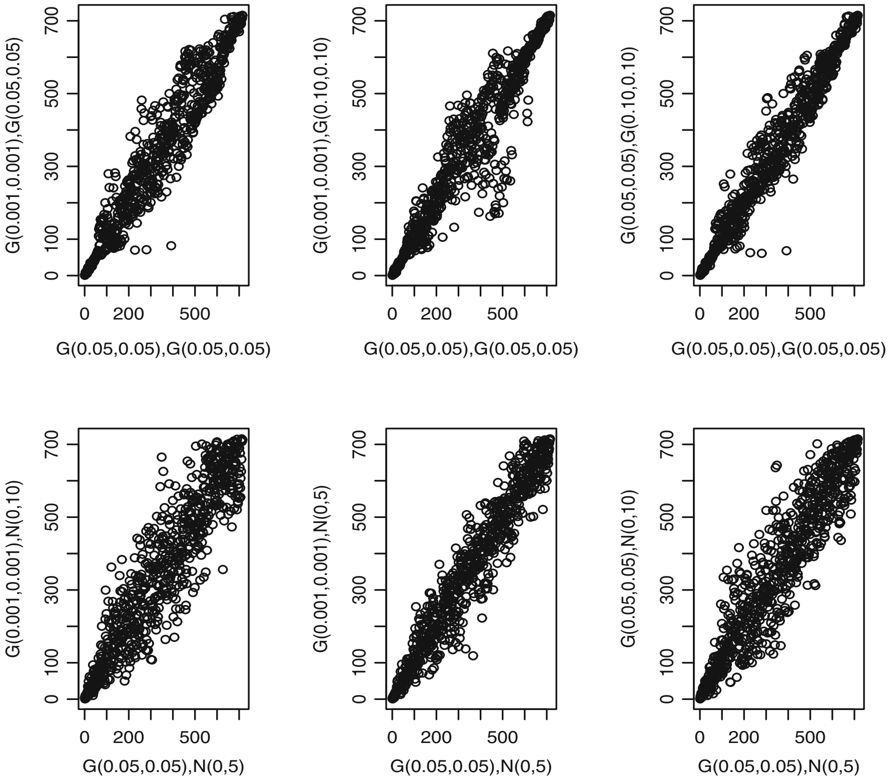



**Fig. 5** Comparison of estimated posterior probability of resistance for various prior specifications; ZIB-CAR model is in the *top panel*, and ZIB-discrete model is in the *bottom panel*

is assigned a Gamma$(0.001, 0.001)$ or Gamma$(0.05, 0.05)$ distribution while $\sigma_{b_\theta}^{-2}$ is assigned a Gamma$(0.05, 0.05)$ or Gamma$(0.10, 0.10)$ distribution. For the ZIB-discrete model, $\sigma_{b_\mu}^{-2}$ is assigned the same priors as above; the priors for $R^1$ and $\delta^j, j = 1, 2, \ldots, k-1$ are $N(1, 5)$ and $N(0, 5)$ or $N(1, 10)$ and $N(0, 10)$, suitably standardized for zero truncation. Estimates are quite stable with those of $\sigma_{b_\theta}^2$ displaying slightly higher variability across priors. As well, goodness of fit measures were consistent across prior specifications. Figure 5 presents the correspondence of posterior probabilities of belonging to the resistant group across prior specifications. Both models exhibit strong correlations.

Although not discussed here, we also conducted a sensitivity analysis for the ZIB-MCAR model. Currently there is little literature to guide the choice of Dirichlet priors. We found the choice of starting values for the intercept term, $\alpha_\theta$, and the random effects, $b_\theta$, was important for convergence. Regardless of prior choice, with poor starting values, the intercept term $\alpha_\theta$ often drifted toward large negative values. It was difficult to obtain convergence for MCAR model parameters. We recommend caution when fitting this model. It is important to do a thorough sensitivity analysis to determine parameter stability before making inference.

**Table 3** Study of sensitivity to priors: posterior mean estimates

| Model | Priors | Parameters | | | |
|---|---|---|---|---|---|
| | $\sigma_{b_\mu}^{-2}, \sigma_{b_\theta}^{-2}$ | $\alpha_\mu$ | $\alpha_\theta$ | $\sigma^2_{b_\mu}$ | $\sigma^2_{b_\theta}$ |
| ZIB-CAR | G(0.001, 0.001), G(0.05, 0.05) | −1.272 | −2.83 | 1.98 | 2.81 |
| | G(0.001, 0.001), G(0.10, 0.10) | −1.271 | −2.86 | 1.96 | 3.69 |
| | G(0.05, 0.05), G(0.05, 0.05) | −1.271 | −2.84 | 1.96 | 3.00 |
| | G(0.05, 0.05), G(0.10, 0.10) | −1.277 | −2.87 | 2.04 | 2.88 |
| | $\sigma_{b_\mu}^{-2}, R^1, R^j, j = 2, 3$ | $\alpha_\mu$ | $\sigma^2_{b_\mu}$ | R/(1+R) | $\theta$ |
| ZIB-discrete | G(0.001, 0.001), $N(1, 5)$, $N(0, 5)$ | −1.30 | 1.46 | (0.023, 0.47, 0.70) | (0.84, 0.12, 0.05) |
| | G(0.001, 0.001), $N(1, 10)$, $N(0, 10)$ | −1.30 | 1.47 | (0.024, 0.62, 0.80) | (0.90, 0.07, 0.03) |
| | G(0.05, 0.05), $N(1, 5)$, $N(0, 5)$ | −1.30 | 1.46 | (0.024, 0.41, 0.67) | (0.73, 0.21, 0.06) |
| | G(0.05, 0.05), $N(1, 10)$, $N(0, 10)$ | −1.30 | 1.46 | (0.023, 0.54, 0.77) | (0.77, 0.19, 0.04) |

## 7 Discussion

This paper considers a rich and flexible class of zero-inflated mixture models which incorporate random effects for modelling spatial correlation, clusters of resistance and isolated resistance to weevil infestation. These complementary models provide a variety of structures for modelling the probability of belonging to a resistant group and accommodating spatial correlation in both the zero and non-zero components of the model. The mixture models estimate the probability of belonging to a resistant group using an unobserved latent variable. Estimation of this latent variable is sensitive to the structure imposed on the probability of belonging to the zero component. In order to gain meaningful insight, it is essential to choose the model structure wisely, based on expert knowledge and the key biological questions at hand.

For the weevil infestation data, the simple Beta-binomial model, although accommodating the zero counts well, did not provide enough flexibility for modelling the overdispersed counts. Although the Beta-binomial model did not provide insight regarding resistant trees, it, surprisingly, provided a great deal of flexibility in modelling zero-inflation relative to the GLM model. In situations where the number of zeros is not extreme and tree resistance to weevil infestation is not of biological interest, the GLM may provide a reasonable overall fit to data.

The zero-inflated spatial models (ZIB-CAR and ZIB-discrete) fit the data well. They also provided insight into possible underlying mixture mechanisms related to resistance. The ZIB-CAR model specified zeros to be clustered while the ZIB-discrete model was more flexible and accommodated spatially isolated zeros (spatial outliers). Both models were relatively robust to choice of priors, especially with regard to the estimated probability of resistance. The ZIB-MCAR model was quite constrained, and as implemented here, it was difficult to gain convergence for this model. For a discussion of alternative formulations of MCAR models see Jin et al. (2005).

We note that extensions of Rathbun and Fei's (2006) zero-inflated probit model to accommodate spatial correlation in the non-resistant group may provide another competing alternative model. Computational challenges for large datasets arise due to the size of the covariance matrix. However, this may be handled by using an approximation to the Matern process, for example using spatial models based on kernel convolution (Nathoo 2010). A natural extension to the zero-inflated models outlined here, is to use splines to account for temporal relationships. Splines could be used to model trends in both the probability of resistance and the mean count. For binomial spatio-temporal analyses where the upper bound on counts is not too large, the use of a dynamic state-space structure where certain sites are in an absorbing state may also provide a useful modelling approach. Although the temporal features of the case study were not considered here, in other instances, models where a temporal trend is of key interest may be required. Various models for longitudinal analysis of ZI data can be found in, for example, Alfo and Maruotti (2010), Hasan and Sneddon (2009) and Hasan et al. (2009).

There is also a large body of literature on modelling spatio-temporal trends in zero-inflated data (Tzala and Best 2008; Velarde et al. 2004; Ver Hoef and Jansen 2007; Wikle and Anderson 2003). For instance, Ver Hoef and Jansen (2007) use both mixture and hurdle models for zero-inflated seal haul out analyses. More recent zero-inflation research focuses on the joint modelling of multiple outcomes (Diao et al. 2013; Feng and Dean 2012; Hatfield et al. 2012; Rodrigues-Motta et al. 2013). For instance Feng and Dean (2012) use a latent random risk term to link the spatial component across outcomes and Dean and Lundy (pers.comm., Lundy's thesis dissertation, in preparation, 2015) model juvenile delinquent behaviour using a longitudinal joint model.

The weevil infestation case study demonstrates that a variety of models can accommodate the large number of zeros in this data. However, models with similar measures of goodness of fit can vary in utility. The Beta-binomial model fit the data fairly well but did not provide foresters with insight into tree resistance to weevil infestations. The zero-inflated spatial models provided a good fit with the advantage of providing biological insight. The ZIB-CAR model provided insight into the locations with the highest resistance while the ZIB-discrete mixture model provided a means of identifying the trees with the largest probability of resistance. This analysis highlights the importance of choosing a statistical model that not only fits the data, but reflects the underlying biological processes and provides the investigator with biologically meaningful interpretations and thus insight into key scientific questions.

**Acknowledgements** We would like to thank the Natural Sciences and Engineering Research Council of Canada for research funding. We thank Erin Lundy and Alisha Albert-Green for their assistance with the literature review for this paper. We would also like to thank all those who have provided valuable feedback on this work: Giovani da Silva for his review and helpful comments, the ISS-2015 Symposium audience for their thoughtful questions, and the ISS reviewer and Proceedings editor for their useful comments and suggestions.

## References

Agarwal, D.K., Gelfand, A.E., Citron-Pousty, S.: Zero-inflated models with application to spatial count data. Environ. Ecol. Stat. **9**, 341–355 (2002)

Ainsworth, L.M., Dean, C.B.: Zero-inflated spatial models: web supplement. http://www.stat.sfu.ca/~dean/students/ainsworth.html#nav (2007)

Ainsworth, L.M., Dean, C.B.: Detection of local and global outliers in mapping studies. Environmetrics **19**, 21–37 (2008)

Alfo, M., Maruotti, A.: Two-part regression models for longitudinal zero-inflated count data. Can. J. Stat. **38**(2), 197–216 (2010)

Besag, J., York, J., Mollié, A.: Bayesian image restoration with two applications in spatial statistics. Ann. Inst. Stat. Math. **43**(1), 1–59 (1991)

Breslow, N.E., Clayton, D.G.: Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. **88**, 9–25 (1993)

Chen, J., Knalili, A.: Order selection in finite mixture models. with a nonsmooth penalty, J. Am. Stat. Assoc. **103**, 1674–1683 (2008)

Consul, P.C., Jain, G.C.: A generalization of the Poisson distribution. Technometrics **15**(4), 791–799 (1973)
Diao, L., Cook, R., Lee, K.: A copula model for marked point processes. Lifetime Data Anal. **19**, 463–489 (2013)
Dobbie, M.J., Welsh, A.H.: Modelling correlated zero-inflated count data. Aust. N. Z. J. Stat. **43**, 431–444 (2001)
Eberly, L.E., Carlin, B.P.: Identifiability and convergence issues for Markov chain Monte Carlo fitting of spatial models. Stat. Med. **19**, 2279–2294 (2000)
Feng, C.X., Dean, C.B.: Joint analysis of multivariate spatial count and zero-heavy count outcomes using common spatial factor models. Environmetrics **23**(6), 493–508 (2012)
Gelman, A.: Prior distributions for variance parameters in hierarchical models. Bayesian Anal. **1**, 515–533 (2006)
Gelman, A., Hill, J.: Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York (2007)
Gelman, A., Rubin, D.: Inference from iterative simulation using multiple sequences (with discussion). Stat. Sci. **7**, 457–511 (1992)
Gelman, A., Meng, X., Stern, H.: Posterior predictive assessment of model fitness via realized discrepancies (with discussion). Stat. Sin. **6**, 733–807 (1996)
Gilks, W.R., Wild, P.: Adaptive rejection sampling for Gibbs sampling. Appl. Stat. **41**(2), 337–348 (1992)
Hall, D.B.: Zero-inflated Poisson and binomial regression with random effects: a case study. Biometrics **56**, 1030–1039 (2000)
Hall, D.B., Zhang, Z.: Marginal models for zero-inflated clustered data. Stat. Model. **4**, 161–180 (2004)
Hardin, J.W., Hilbe, J.M., Hible, J.: Generalized Linear Models and Extensions, 2nd edn. Stata Press, Texas (2007)
Hasan, M.T., Sneddon, G.: Zero-inflated Poisson regression for longitudinal data. Commun. Stat. Simul. Comput. **38**(3), 638–653 (2009)
Hasan, T., Sneddon, G., Ma, R.: Pattern-mixture zero-inflated mixed models for longitudinal unbalanced count data with excessive zeros. Biom. J. **51**(6), 946–960 (2009)
Hatfield, L., Boye, M., Hackshaw, M., Carlin, B.: Multilevel Bayesian models for survival times and longitudinal patient-reported outcomes with many zeros. J. Am. Stat. Assoc. **107**(499), 875–885 (2012)
He, F., Alfaro, R.: White pine Weevil (Coleoptera: Curculionidae) attack on white spruce: spatial and temporal patterns. Environ. Entomol. **26**(4), 888–895 (1997)
Heilbron, D.: Zero-altered and other regression models for count data with added zeros. Biom. J. **36**(5), 531–547 (1994)
Jin, X., Carlin, B.P., Banerjee, S.: Generalized hierarchical multivariate CAR models for areal data. Biometrics **61**, 950–961 (2005). doi:10.1111/j.1541–0420.2005.00359.x
Johnson, V.: A Bayesian $\chi^2$ test for goodness of fit. Ann. Stat. **32**(6), 2361–2384 (2004)
Kuhnert, P.M., Martin, T.G., Mengersen, K., Possingham, H.P.: Assessing the impacts of grazing levels on bird density in woodland habitat: a Bayesian approach using expert opinion. Environmetrics **16**, 717–747 (2005)
Lambert, D.: Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics **34**, 1–14 (1992)
Lawless, J.F.: Negative binomial and mixed Poisson regression. Commun. Stat. **15**, 209–225 (1987)
Lawson, A.B., Clark, A.: Spatial mixture relative risk models applied to disease mapping. Stat. Med. **21**(3), 359–370 (2002)
Liang, K., Zeger, S.: Longitudinal data analysis using generalized linear models. Biometrika **73**, 13–22 (1986)
Lunn, D., Jackson, C., Best, N., Thomas, A., Speigelhalter, D.: The Bugs Book - A Practical Introduction to Bayesian Analysis. CRC Press, Chapman and Hall, Boca Raton (2012)

Martin, T.G., Wintle, B.A., Rhodes, J.R., Kuhnert, P.M., Field, S.A., Low-Choy, S.J., Tyre, A.J., Possingham, H.P.: Zero tolerance ecology: improving ecological inference by modelling the source of zero observations. Ecol. Lett. **8**, 1235–1246 (2005)

McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn. Chapman and Hall, New York (1989)

Mullahy, J.: Specification and testing of some modified count data models. J. Econ. **33**, 341–365 (1986)

Nathoo, F.: Joint spatial modeling of recurrent infection and growth with processes under intermittent observation. Biometrics **66**, 336–346 (2010). doi:10.1111/j.1541-0420.2009.01305.x

Nathoo, F., Dean, C.B.: A mixed mover-stayer model for spatiotemporal two-state processes. Biometrics **63**, 881–891 (2007). doi:10.1111/j.1541-0420.2007.00752.x

Rathbun, S.L., Fei, S.: A spatial zero-inflated Poisson regression model for oak regeneration. Environ. Ecol. Stat. **13**, 409–426 (2006)

Ridout, M., Demetrio, C.G.B., Hinde, J.: Models for count data with many zeros. In: International Biometric Conference, Cape Town (1998)

Rodrigues-Motta, M., Pinheiro, H.P., Martins, E.G., Araujo, M.S., dos Reis, S.F.: Multivariate models for correlated count data. J. Appl. Stat. **40**(7), 1586–1596 (2013)

Spiegelhalter, D., Thomas, A., Best, N., Lunn, D.: WinBUGS User Manual Version 1.4. Medical Research Council Biostatistics Unit, Cambridge (2003)

Stroup, W.W.: Generalized linear mixed models, modern concepts, methods and applications. CRC Press, Taylor & Francis Group, New York (2013)

Tzala, E., Best, N.: Bayesian latent variable modelling of multivariate spatio-temporal variation in cancer mortality. Stat. Methods Med. Res. **17**, 97–118 (2008)

Velarde, L.G.C., Migon, H.S., Pereira, B.B.: Spate-time modeling of rainfall data. Environmetrics **15**, 561–576 (2004)

Ver Hoef, J.M., Jansen, J.K.: Space-time zero-inflated count models of Harbour Seals. Environmetrics **18**(7), 697–712 (2007)

Wang, K., Yau, K.K.W., Lee, A.H.: A zero-inflated Poisson mixed model to analyze diagnosis related groups with majority of same-day hospital stays. Comput. Methods Programs Biomed. **68**, 195–203 (2002)

Wikle, C.K., Anderson, C.J.: Climatological analysis of tornado report counts using a hierarchical Bayesian spatio-temporal model. J. Geophys. Res. Atmos. **108**, 9005 (2003). doi:10.1029/2002JD002806

Williams, D.A. (1982). Extra-binomial variation in logistic linear models. Appl. Stat. **31**(2), 144–148 (1982)

# Inferences in Stochastic Volatility Models: A New Simpler Way

**Vickneswary Tagore, Nan Zheng, and Brajendra C. Sutradhar**

**Abstract** Two competitive analytical approaches, namely, the generalized method of moments (GMM) and quasi-maximum likelihood (QML) are widely used in statistics and econometrics literature for inferences in stochastic volatility models (SVMs). Alternative numerical approaches such as monte carlo markov chain (MCMC), simulated maximum likelihood (SML) and Bayesian approaches are also available. All these later approaches are, however, based on simulations. In this paper, we revisit the analytical estimation approaches and briefly demonstrate that the existing GMM approach is unnecessarily complicated. Also, the asymptotic properties of the likelihood approximation based QML approach are unknown and the finite sample based QML estimators can be inefficient. We then develop a precise set of moment estimating equations and demonstrate that the proposed method of moments (MM) estimators are easy to compute and they perform well in estimating the parameters of the SVMs in both small and large time series set up. A 'working' generalized quasi-likelihood (WGQL) estimation approach is also considered. Estimation methods are illustrated by reanalyzing a part of the Swiss-Franc and U.S. dollar exchange rates data.

**Keywords** Correlated squared observations • Consistent estimation • Generalized method of moments and complexity • Kurtosis estimation • Large sample properties • Quasi-maximum likelihood estimation • Simpler method of moments using fewer unbiased estimating equations • Small sample comparison • Time dependent variances • Volatility parameters

## 1 Introduction

In economic such as in financial time series, one frequently encounters certain uncorrelated but dependent responses, for example, stock returns and exchange rates. Also, the responses may contain more extreme values than usual indicating

V. Tagore (✉) • N. Zheng • B.C. Sutradhar
Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada A1C5S7
e-mail: vtagore@mun.ca; k33nz@mun.ca; bsutradh@mun.ca



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_4

higher kurtosis; and they may exhibit non-stationary variation over time. This type of data are commonly analyzed by using the so called stochastic volatility (SV) models, introduced by Taylor (1982), among others. The SV models are developed such that the responses conditional on their latent heteroscedastic variances behave like a white noise series, where the variances maintain a dynamic relationship over time ensuring higher kurtosis than a Gaussian series. In notation, let $y_t$ be the response on a continuous scale recorded at time t (t $= 1, \ldots, T$) with zero mean (or shifted to zero) and a time dependent unobservable variance $\sigma_t^2$. To reflect the aforementioned characteristics of the responses $\{y_t\}$, one uses the volatility model

$$y_t|\sigma_t = \sigma_t\,\epsilon_t \qquad t = 1, \ldots, T \tag{1}$$

$$\ln(\sigma_t^2) \equiv h_t = \gamma_0 + \gamma_1\,h_{t-1} + \eta_t; \quad t = 2, \ldots, T \tag{2}$$

(Andersen and Sorensen 1996; Harvey et al. 1994; Ruiz 1994;Mills 1999, pp.127–128), where error variables $\{\epsilon_t, t = 1, \ldots, T\}$ are independently and identically (iid) distributed with mean zero and variance 1, that is, $\epsilon_t \overset{iid}{\sim} (0, 1)$. Also, $\epsilon_t$ and $\sigma_t$ are assumed to be independent. In non-stationary variance model (2), $\gamma_0$ is the intercept parameter, $\gamma_1$ is the volatility persistence parameter and $\eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2)$ with $\sigma_\eta^2$ as the measure of uncertainty about future volatility. One may then obtain the conditional mean and variance of $Y_t$, and the conditional lag $\ell$ auto-covariances between $Y_t$ and $Y_{t-\ell}$ as

$$\begin{aligned} E(Y_t|\sigma_t) = \sigma_t\,E[\epsilon_t] = 0,\ \ \text{var}(Y_t|\sigma_t) = \sigma_t^2\,\text{var}(\epsilon_t) = \sigma_t^2,\ \ \text{and} \\ \text{Cov}[(Y_t, Y_{t-\ell})|\sigma_t, \ldots, \sigma_{t-\ell}] = \sigma_t\sigma_{t-\ell}\text{Cov}[\epsilon_t, \epsilon_{t-\ell}] = 0, \end{aligned} \tag{3}$$

respectively, yielding further the corresponding unconditional mean, variance and covariances as

$$\begin{aligned} E[Y_t] = 0,\ \text{var}[Y_t] = E[\sigma_t^2],\ \text{and} \\ \text{Cov}[Y_t, Y_{t-\ell}] = 0. \end{aligned} \tag{4}$$

Note that even though the lag covariances are zero, that is, the responses are uncorrelated, they are, however, dependent to each other. For example,

$$\text{Cov}[Y_t^2, Y_{t-\ell}^2] = E[\sigma_t^2\sigma_{t-\ell}^2] - E[\sigma_t^2]E[\sigma_{t-\ell}^2] \neq 0. \tag{5}$$

This is because in the present set up, heteroscedastic latent variances over time are not independent, instead they follow, say, the dynamic relationship given in (2). Note that for model (2), it is necessary to know the initial variance $\sigma_1^2$ at time $t = 1$. It is reasonable to assume that

$$\ln(\sigma_1^2) = h_1 \overset{iid}{\sim} N\left(\frac{\gamma_0}{1-\gamma_1}, \frac{\sigma_\eta^2}{1-\gamma_1^2}\right), \tag{6}$$

(Lee and Koopman 2004, Eq. (1.1c)).

As far as the fitting of the volatility model (1)–(2) to the data $\{y_t\}$ is concerned, it is important to estimate the parameters of the model, namely, $\gamma_0, \gamma_1$ and $\sigma^2_\eta$, consistently and as efficiently as possible. It is also of interest to obtain a consistent estimate for the kurtosis of the responses, namely

$$\kappa_t(\gamma_0,\ \gamma_1,\ \sigma^2_\eta) = \frac{E(Y_t^4)}{[E(Y_t^2)]^2}. \tag{7}$$

It can be shown under the model (1)–(2) that this kurtosis is larger than 3, its counterpart value under the Gaussian responses. Thus this SV model can fit fatter than normal tailed values which causes volatility in the data. However the estimation of the parameters $\gamma_0,\ \gamma_1,$ and $\sigma^2_\eta$ for this SV model is complex because of the fact that $\sigma_t^2$ $(t = 1, \ldots, T)$ in (2) are unobserved and their correlation structure affect the correlation structure of $\{y_t^2\}$ in a complicated way, even though $y_t$'s $(t = 1, \ldots, T)$ are uncorrelated. Nevertheless, there exist some widely used analytical estimation approaches, namely, the quasi-maximum likelihood (QML) and the so-called generalized method of moments (GMM). The advantages and disadvantages of these estimation methods are reviewed in brief in Sect. 2.

Note that to obtain consistent estimates in a finite sample set up (i.e. for a time series with moderate length), as opposed to the GMM and QML approaches, there exists several numerical approaches such as Bayesian approach by Jacquier et al. (1994) and the simulated ML (SML) approach which is considered to be an improvement over the so-called monte carlo markov chain (MCMC) approach. For SML approach, we refer to Danielsson (1994), Shephard and Pitt (1997), Durbin and Koopman (1997), Liesenfeld and Richard (2003) and Lee and Koopman (2004). It is, however, recognized that these numerical techniques are computationally intensive. For this reason, and also because in practice such as in financial or environmental analysis one may encounter a large time series, similar to Andersen and Sorensen (1996, 1997), in this paper, we concentrate on the use of the moments to obtain efficient estimates in a large time series set up. But, in contrary to Andersen and Sorensen (1996, 1997) who used 24 moment functions for three parameters, we use only three moments equations for three parameters, and we demonstrate that this simpler MM surprisingly performs very well in estimating the desired parameters. The proposed moment equations along with an algorithm for this simpler MM approach is given in Sect. 3. We discuss the asymptotic properties of the proposed estimators in Sect. 5. In Sect. 3, we also provide an alternative 'weighted' GQL (WGQL) [or approximate GQL (AGQL)] estimation approach for the estimation of the same volatility parameters.

To examine the performance of the proposed MM in estimating the volatility parameters, in Sect. 4, we conduct a finite samples based simulation study. For the purpose, we consider both small and large time series. Note that as it will be discussed in Sect. 2, the computation for the existing QML estimators is manageable, but it is extremely difficult or impossible to compute the exact formulas for the asymptotic variances of the existing QML estimators. Thus, to compare the proposed MM method we have included the QML method in the finite samples based simulation study. By the same token, as it is also demonstrated in Sect. 2 that the existing GMM approach is unnecessarily complicated and cumbersome, we do not include this method in the simulation study. Rather, we include the aforementioned new alternative moments based WGQL approach in the comparison study.

In Sect. 6, we estimate the kurtosis using the proposed simpler MM approach. Section 7 presents an empirical example in which the model is fitted to a series of daily U.S. Dollar and Swiss-Franc exchange rates. The paper concludes in Sect. 8.

## 2 GMM Versus QML Estimation for Volatility Models

### 2.1 Existing GMM Estimation and Complexity

In a linear dynamic mixed model set up for panel data, many econometricians such as Arellano and Bond (1991) and Ahn and Schmidt (1995) [see also Chamberlain (1992), and Keane and Runkle (1992)] have applied the so-called generalized method of moments (GMM) (Hansen 1982) to estimate the parameters of the model. Let $\alpha$ denote the $p$-dimensional vector of parameters of the model. Let $\psi_i(z_{i1},\ldots,z_{iT};\alpha)$ be a $p$-dimensional unbiased moment function constructed based on the data $z_{i1},\ldots,z_{iT}$ from the $i$th $(i=1,\ldots,I)$ individual over $T$ time points, so that

$$E[\psi_i(z_{i1},\ldots,z_{iT};\alpha)]=0.$$

In the GMM approach, one then obtains consistent estimate of $\alpha$ by minimizing the quadratic form

$$I^{-1}\left[\sum_{i=1}^{I}\psi_i(z_{i1},\ldots,z_{iT},\alpha)\right]' C^{-1}\left[\sum_{i=1}^{I}\psi_i(z_{i1},\ldots,z_{iT},\alpha)\right], \tag{8}$$

(Hansen 1982) for some positive definite $p\times p$ symmetric matrix $C$. The GMM estimating equation (8) provides optimal estimates when $C$ is chosen as the covariance matrix of the unbiased function $\sum_{i=1}^{I}\psi_i(z_{i1},\ldots,z_{iT},\alpha)$.

Following the aforementioned GMM approach, Andersen and Sorensen (1996, pp. 350–351) [see also Andersen and Sorensen (1997, Sect. 3, pp. 399–400)], for example, have constructed the GMM estimating equations for the parameters of the volatility model (1)–(2). To be specific, for the estimation of the main parameters, namely $\gamma_0,\gamma_1,$ and$\sigma^2_\eta$, these authors have first chosen 24 moment functions as

$$g_{t1}=|y_t|-E|y_t|,\qquad g_{t2}=y_t^2-E[y_t^2]$$

$$g_{t3} = |y_t|^3 - E|y_t|^3, \qquad g_{t4} = y_t^4 - E[y_t^4]$$
$$g_{t,4+l} = |y_t y_{t-l}| - E|y_t y_{t-l}| \qquad l = 1, \ldots, 10$$
$$g_{t,14+l} = y_t^2 y_{t-l}^2 - E[y_t^2 y_{t-l}^2], \qquad l = 1, \ldots, 10. \tag{9}$$

Note that these moment functions are unbiased because

$$E[g_{tu}] = 0, \text{ for } u = 1, \ldots, 24.$$

Next by using these unbiased moment functions, for $\alpha = (\gamma_0,\ \gamma_1,\ \sigma_\eta^2)'$, they have constructed the $24 \times 1$ vector of unbiased functions as

$$g(\alpha) = \frac{1}{T}\sum_{t=1}^{T} g_t(\alpha) \qquad \text{with} \qquad g_t(\alpha) = [g_{t1}(\alpha), \ldots, g_{tu}(\alpha), \ldots, g_{t,24}(\alpha)]'. \tag{10}$$

These unbiased functions are exploited to construct the GMM estimating equation for $\alpha = (\gamma_0, \gamma_1, \sigma_\eta^2)'$ as given by

$$\frac{\partial g'(\alpha)}{\partial \alpha} \Lambda^{-1}(\alpha) g(\alpha) = 0, \tag{11}$$

where $\Lambda(\alpha) = \text{Cov}(g(\alpha))$ is the covariance matrix of $g(\alpha)$ as an optimal choice, and $\frac{\partial g'(\alpha)}{\partial \alpha}$ is the $3 \times 24$ first derivative matrix.

Note however that the construction of the GMM estimating equation (11) is quite cumbersome. This is because, it requires the computation of the following expectations:

$$E|y_t|,\ E[y_t^2],\ E|y_t|^3,\ E[y_t^4],\ E|y_t y_{t-l}|,\ E[y_t^2 y_{t-l}^2], \qquad l = 1, \ldots, 10,$$

and their derivatives with respect to $\alpha = (\gamma_0, \gamma_1, \sigma_\eta^2)'$. Also it requires the computation for the $24 \times 24$ matrix $\Lambda(\alpha)$. It is, thus, clear that the over all computation for the construction of the estimating equation (11) is cumbersome. In particular, the computation for the $\Lambda(\alpha)$ matrix is extremely complicated which requires the formulas for the moments up to order 8. Furthermore, there is no guidelines available how these 24 functions were chosen, when in fact, one can think of infinite number of such functions (Melino and Turnbull 1990, p. 250). This raises a concern about such an estimation procedure where an arbitrary large number of functions are needed to estimate a small number of parameters. A further drawback of this GMM procedure is that even if one pursues this algebraically painstaking procedure, it does not, however, show any substantial efficiency gain in estimation (Andersen and Sorensen 1997; Ruiz 1997) over other competitive approaches such as the quasi-maximum likelihood (QML) estimation approach discussed below. Consequently, we will not follow this GMM estimation approach any more in the present paper.

## 2.2 QML Estimation

Note that the exact likelihood estimation for the volatility parameters $\gamma_0$, $\gamma_1$ and $\sigma_\eta^2$, under the volatility model (1)–(2), is extremely complex. To understand this complexity, we follow (1) and first express $\log y_t^2$ as

$$\begin{aligned} z_t = \log y_t^2 &= \log \sigma_t^2 + \log \epsilon_t^2 \\ &= E[\log \epsilon_t^2] + \log \sigma_t^2 + u_t \\ &\equiv \kappa_1 + \log \sigma_t^2 + u_t \qquad\qquad t = 1, \ldots, T, \end{aligned} \tag{12}$$

where $\epsilon_t \overset{iid}{\sim} N(0, 1)$ and $\kappa_1 = -1.270363$. In (12), $u_t$ follows the log chi-square distribution with mean zero and variance $\kappa_2 = \pi^2/2$ (Abramowitz and Stegun 1970, p. 943). It then follows that the exact likelihood function is given by

$$\begin{aligned} L(\gamma_0, \gamma_1, \sigma_\eta^2 | z_1, z_2, \ldots, z_T) &= \int_{\sigma_1^2,\ldots,\sigma_T^2} \prod_{t=1}^{T} f(z_t | \log \sigma_t^2) dz_t \;\; d(\log \sigma_t^2) \\ &= \int_{\sigma_1^2,\ldots,\sigma_T^2} \prod_{t=1}^{T} g^*(u_t) du_t \;\; d(\log \sigma_t^2) \end{aligned} \tag{13}$$

where $g^*(u_t)$ represents the $\log \chi^2(0, \kappa_2)$ distribution. The computation of the likelihood function (13) is very complicated as it requires the multi-dimensional integrations of the product chi-square distributions with regard to the random variances $\sigma_1^2, \ldots, \sigma_T^2$.

To avoid this complex integration problem, some authors have suggested to use the normal approximation to the log chi-square distribution. See, for example, Ruiz (1994), Harvey et al. (1994), Koopman et al. (1995, Chap. 7.5) and Mills (1999, pp. 130–131). This is equivalent to approximate the likelihood function (13) by pretending that

$$\mathbf{z} = (z_1, z_2, \ldots, z_T)'$$

follows a MVN (multivariate normal) distribution with true mean vector and true covariance matrix under the model. Let $\mathbf{m} = (m_1, \ldots, m_t, \ldots, m_T)' = E[\mathbf{Z}]$ and the covariance matrix $\Phi = \text{Cov}(\mathbf{Z}) = (v_{ut})$ of the response vector $\mathbf{z}$. It can be shown that

$$m_t = \begin{cases} \kappa_1 + \log \sigma_1^2 & \text{for t = 1} \\ \kappa_1 + \gamma_1^{t-1} \;\log \sigma_1^2 + \frac{\gamma_0(1-\gamma_1^{t-1})}{1-\gamma_1} & \text{for t=2}, \ldots, T \end{cases}$$

and for u<t,

$$v_{ut} = \begin{cases} \kappa_2 & \text{for u=t=1} \\ 0 & \text{for u=1,t=}2,\ldots,T \\ \sigma_\eta^2 \ \sum_{i=0}^{t-2} \gamma_1^{2i} \ + \kappa_2 & \text{for u=t=}2,\ldots,T \\ \sigma_\eta^2 \ \gamma_1^{t-u} \sum_{i=0}^{u-2} \gamma_1^{2i} & \text{for u} < \text{t} \quad \text{u=}2\ ,\ldots,T \end{cases}$$

and $v_{tu} = v_{ut}$.

One may then write the normality based pseudo-likelihood, which however has been referred to as the quasi-likelihood (QL), with log quasi-likelihood function given by

$$\log L_Q^* = c_0 \ \ -\frac{1}{2}\log|\Phi| \ \ -\frac{1}{2}[(\mathbf{z}-\mathbf{m})' \ \Phi^{-1} \ (\mathbf{z}-\mathbf{m})], \tag{14}$$

(Shephard 1996, Eq. 1.17). It then follows from (14) that for known $\gamma_0$, the quasi maximum likelihood (QML) estimates for $\gamma_1$ and $\sigma_\eta^2$ can be obtained by solving

$$\frac{\partial \log L_Q^*}{\partial \boldsymbol{\gamma_1}} = 0 \qquad\qquad \text{and} \qquad\qquad \frac{\partial \log L_Q^*}{\partial \boldsymbol{\sigma_\eta^2}} = 0 \tag{15}$$

For stationary case, $m_t = \kappa_1 + \gamma_0/(1-\gamma_1)$ for $t = 2, \cdots, T$, then the QML estimator for $\kappa_1 + \gamma_0/(1-\gamma_1)$ is the sample mean of $z_t$, which is used to estimate $\gamma_0$ in Ruiz (1994) and Harvey et al. (1994). Let the final QML estimates from (15) are denoted by $\hat{\gamma}_{1,QML}$ and $\widehat{\sigma_\eta^2}_{,QML}$, respectively, and that for $\gamma_0$ by $\hat{\gamma}_{0,QML}$.

Note that because $T$ is usually large, the computation for the inversion of the $T \times T$ covariance matrix $\Phi$ may be time intensive. Nevertheless, this QML approach is computationally feasible, especially when it is compared to the cumbersome GMM approach discussed in the last section. For this reason, we include the QML approach in Sect. 4 in a simulation study to examine its finite sample relative performance as compared to the proposed simpler MM approach to be discussed in Sect. 3.

Further note that the approximate QML approach has, however, some analytical drawbacks. First, because the true distribution of $u_t$, namely $g^*(u_t)$ ($\log \chi^2$ distribution) is extremely left skewed, conditional on $\log \sigma_t^2$, $z_t$ follows the $\log \chi^2$ distribution. Consequently, this normality based QML approximation can be inefficient. Furthermore, it is analytically extremely difficult to compute the asymptotic properties such as asymptotic variances of the QML estimators. This is because $L_Q^*$ in (14) is the pseudo or quasi-likelihood but the computation of the covariance matrix of $\hat{\alpha}_{QML}^* = (\hat{\gamma}_{1,QML}, \hat{\sigma}_{\eta,QML}^2)'$ requires the expectation of the moments up to order four over the true model (TM). More specifically, the covariance of $\hat{\alpha}_{QML}^*$ obtained by solving the likelihood equations (15), has the form

$$\lim_{T\to\infty} \mathrm{Cov}(\hat{\alpha}^*_{QML}) = \lim_{T\to\infty} - \begin{pmatrix} E_{TM}\left[\dfrac{\partial^2 \log L_Q^*}{\partial\gamma_1^2}\right] & E_{TM}\left[\dfrac{\partial^2 \log L_Q^*}{\partial\gamma_1 \partial\sigma_\eta^2}\right] \\ & E_{TM}\left[\dfrac{\partial^2 \log L_Q^*}{\partial(\sigma_\eta^2)^2}\right] \end{pmatrix}^{-1} \mathrm{cov}_{TM}\begin{pmatrix} \frac{\partial \log L_Q^*}{\partial \boldsymbol{\gamma_1}} \\ \frac{\partial \log L_Q^*}{\partial \boldsymbol{\sigma_\eta^2}} \end{pmatrix}$$

$$\times \begin{pmatrix} E_{TM}\left[\dfrac{\partial^2 \log L_Q^*}{\partial\gamma_1^2}\right] & E_{TM}\left[\dfrac{\partial^2 \log L_Q^*}{\partial\gamma_1 \partial\sigma_\eta^2}\right] \\ & E_{TM}\left[\dfrac{\partial^2 \log L_Q^*}{\partial(\sigma_\eta^2)^2}\right] \end{pmatrix}^{-1}. \tag{16}$$

Note that the score functions in (15) have the formulas

$$\begin{aligned} \frac{\partial \log L_Q^*}{\partial\gamma_1} &= -\frac{1}{2}\,\frac{\partial \log|\Phi|}{\partial\gamma_1} - \frac{\partial(\mathbf{Z}-\mathbf{m})'}{\partial\gamma_1}\,\Phi^{-1}(\mathbf{Z}-\mathbf{m}) - \frac{1}{2}\,(\mathbf{Z}-\mathbf{m})'\,\frac{\partial\Phi^{-1}}{\partial\gamma_1}(\mathbf{Z}-\mathbf{m}) \\ &= -\frac{1}{2}\,trace[\Phi^{-1}\,\frac{\partial\Phi}{\partial\gamma_1}] + \frac{\partial\mathbf{m}'}{\partial\gamma_1}\,\Phi^{-1}\,(\mathbf{Z}-\mathbf{m}) \\ &\quad + \frac{1}{2}\,(\mathbf{Z}-\mathbf{m})'\Phi^{-1}\,\frac{\partial\Phi}{\partial\gamma_1}\,\Phi^{-1}\,(\mathbf{Z}-\mathbf{m}), \end{aligned} \tag{17}$$

and

$$\begin{aligned} \frac{\partial \log L_Q^*}{\partial\sigma_\eta^2} &= -\frac{1}{2}\,\frac{\partial \log|\Phi|}{\partial\sigma_\eta^2} - \frac{1}{2}\,(\mathbf{Z}-\mathbf{m})'\,\frac{\partial\Phi^{-1}}{\partial\sigma_\eta^2}(\mathbf{Z}-\mathbf{m}) \\ &= -\frac{1}{2}\,trace[\Phi^{-1}\,\frac{\partial\Phi}{\partial\sigma_\eta^2}] + \frac{1}{2}\,(\mathbf{Z}-\mathbf{m})'\,\Phi^{-1}\,\frac{\partial\Phi}{\partial\sigma_\eta^2}\,\Phi^{-1}\,(\mathbf{Z}-\mathbf{m}), \end{aligned} \tag{18}$$

respectively. Consequently, even though we can compute the expectations for the second order derivatives over the true model as

$$\begin{aligned} E_{TM}\left[\frac{\partial^2 \log L_Q^*}{\partial\gamma_1^2}\right] &= -\frac{1}{2}\,\frac{\partial[trace(\Phi^{-1}\frac{\partial\Phi}{\partial\gamma_1})]}{\partial\gamma_1} - \frac{\partial\mathbf{m}'}{\partial\gamma_1}\,\Phi^{-1}\frac{\partial\mathbf{m}}{\partial\gamma_1} - \frac{1}{2}trace\left[\frac{\partial^2\Phi^{-1}}{\partial\gamma_1^2}\Phi\right] \\ E_{TM}\left[\frac{\partial^2 \log L_Q^*}{\partial\gamma_1\,\partial\sigma_\eta^2}\right] &= -\frac{1}{2}\,\frac{\partial[trace(\Phi^{-1}\,\frac{\partial\Phi}{\partial\gamma_1})]}{\partial\sigma_\eta^2} - \frac{1}{2}trace\left[\frac{\partial^2\Phi^{-1}}{\partial\gamma_1\,\partial\sigma_\eta^2}\Phi\right] \\ E_{TM}\left[\frac{\partial^2 \log L_Q^*}{\partial(\sigma_\eta^2)^2}\right] &= -\frac{1}{2}\,\frac{\partial[trace(\Phi^{-1}\frac{\partial\Phi}{\partial\sigma_\eta^2})]}{\partial\sigma_\eta^2} - \frac{1}{2}\,trace\left[\frac{\partial^2\Phi^{-1}}{\partial(\sigma_\eta^2)^2}\Phi\right], \end{aligned} \tag{19}$$

but it is clear from (17) and (18) that the computation of $\mathrm{cov}_{TM}\begin{pmatrix} \frac{\partial \log L_Q^*}{\partial \boldsymbol{\gamma_1}} \\ \frac{\partial \log L_Q^*}{\partial \boldsymbol{\sigma_\eta^2}} \end{pmatrix}$ for (16) requires the formulas for the fourth order moments of the elements of $\mathbf{Z}$ under the true model, which are however not available or extremely difficult to compute. For this reason one is unable to examine the asymptotic properties of the QML estimator of $\alpha$. We however include this QML approach in Sect. 4 in a simulation study to examine its relative finite sample performance as compared to the proposed MM approach.

# 3 Proposed Estimation

## *3.1 Unbiased Moment Equations*

We suggest choosing only three moment equations for consistent estimation of three volatility parameters $\gamma_0$, $\gamma_1$, and $\sigma_\eta^2$. These three equations along with rational behind their choice are given below. In the next section, we write the computational steps in an algorithm form.

Because $\sigma_\eta^2$ in theory can take any value from 0 to $\infty$, it is understandable from the volatility model (1)–(2) that a wrong estimate of $\sigma_\eta^2$ can more adversely affect the estimate of $\gamma_1$, than what a wrong estimate of $\gamma_1$ can do to the estimation of $\sigma_\eta^2$. For this reason, it is important to search for a reliable estimate of $\sigma_\eta^2$, first.

**(a) Unbiased Estimating Equation for $\sigma_\eta^2$**

In order to develop an estimating equation for $\sigma_\eta^2$, we observe from (2) that $\sigma_t^2$'s maintain a non-stationary dynamic relationship, stationarity being a special case. Now if $\sigma_t^2$ were stationary, that is $E[\sigma_t^2] = h^*(\sigma_\eta^2)$, a suitable constant function of $\sigma_\eta^2$, then one would have estimated $h^*(\sigma_\eta^2)$ consistently by using $S_1 = \frac{1}{T}\sum_{t=1}^{T} y_t^2$ because of the fact that $E[Y_t|\sigma_t^2] = 0$. However, in the present case, $\sigma_t^2$'s are unobservable and their log values satisfy a non stationary Gaussian AR(1) type relationship given by (2), with errors $\eta_t \overset{iid}{\sim} N(0, \sigma_\eta^2)$. This leads to the expected value of $S_1$ as

$$\begin{aligned} E[S_1] &= E_{\sigma_t^2} E[\frac{1}{T}\sum_{t=1}^{T} y_t^2] \\ &= \frac{1}{T}\left[\sigma_1^2 + \sum_{t=2}^{T} \exp\left(\gamma_1^{t-1}\log\sigma_1^2 + \gamma_0\sum_{i=0}^{t-2}\gamma_1^i + \frac{\sigma_\eta^2}{2}\sum_{r=0}^{t-2}\gamma_1^{2r}\right)\right] \\ &= g_1(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2), \text{ say.} \end{aligned} \tag{20}$$

Note that we assume all through that $\sigma_1^2$ is known and by (6) we replace $\log\sigma_1^2$ with its mean value $\frac{\gamma_0}{1-\gamma_1}$. Thus, for given values of $\gamma_0$ and $\gamma_1$, by (20), one may solve the unbiased estimating equation

$$S_1 - g_1(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2) = 0 \tag{21}$$

to obtain a consistent estimate for $\sigma_\eta^2$. Further note that the solution of (21), however, requires a good initial value for $\sigma_\eta^2$, which we suggest to obtain by solving an asymptotic unbiased estimating equation. It is clear that for a suitable large $T_0$, for any $t > T_0$, $\gamma_1^{t-1} \to 0$ for $|\gamma_1| < 1$. Further, the expectation of $y_t^2$ for $t > T_0$ becomes stationary, producing

$$\lim_{t\to\infty} E[Y_t^2] = \exp\left[\frac{\gamma_0}{1-\gamma_1} + \frac{\sigma_\eta^2}{2}\left(\frac{1}{1-\gamma_1^2}\right)\right] \tag{22}$$
$$= g_{10}(\gamma_0, \gamma_1, \sigma_\eta^2), say.$$

Thus, if the series under consideration is large, i.e., $T \to \infty$, replacing the exact $E[Y_t^2]$ for $t > T_0$ by $\lim_{t\to\infty} E[Y_t^2] = g_{10}(\gamma_0, \gamma_1, \sigma_\eta^2)$, one can consistently estimate the stationary mean function, namely $g_{10}(.)$ by using $S_{10} = \frac{1}{T-T_0}\sum_{t=T_0+1}^{T} y_t^2$. Consequently, for known $\gamma_0$ and $\gamma_1$, we may obtain a very reasonable initial value for $\sigma_\eta^2$ by solving

$$S_{10} - g_{10}(\gamma_0, \gamma_1, \sigma_\eta^2) = 0 \tag{23}$$

We denote this initial value of $\sigma_\eta^2$ by $\sigma_\eta^2(0)$.

**(b) Unbiased Estimating Equation for $\gamma_1$**

Next, to construct an unbiased estimating equation for $\gamma_1$, we first observe that $\gamma_1$ is the lag 1 dependence parameter in the Gaussian AR(1) model (2). We therefore choose a lag 1 based function given by $S_2 = \frac{1}{T-1}\sum_{t=2}^{T} y_{t-1}^2 y_t^2$ to construct the moment equation for $\gamma_1$. For the purpose, for known $\gamma_0$ and using the estimate of $\sigma_\eta^2$ from (21), we compute the expected value of $S_2$ as

$$\begin{aligned}
E[S_2] &= E_{\sigma_t^2} E[\frac{1}{T-1}\sum_{t=2}^{T} y_{t-1}^2 y_t^2] \\
&= \frac{e^{\gamma_0}}{T-1}\Bigg[\sigma_1^2 \ \exp\left(\gamma_1 \log\sigma_1^2 + \frac{\sigma_\eta^2}{2}\right) \\
&\quad + \sum_{t=3}^{T} \exp\Bigg(\gamma_0(1+\gamma_1)\sum_{l=0}^{t-3}\gamma_1^l + \gamma_1^{t-1}\log\sigma_1^2 + \gamma_1^{t-2}\log\sigma_1^2 \\
&\quad + \frac{\sigma_\eta^2}{2}\left\{(1+\gamma_1)^2 \sum_{l=0}^{t-3}\gamma_1^{2l} + 1\right\}\Bigg)\Bigg]
\end{aligned}$$

$$= g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2), \quad \text{say}, \tag{24}$$

and solve the unbiased estimating equation

$$S_2 - g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2) = 0, \tag{25}$$

for $\gamma_1$ iteratively.

**(c) Unbiased Estimating Equation for $\gamma_0$**

Note that $\gamma_0$ is the intercept in the dynamic function for $\log \sigma_t^2$ with $\sigma_t^2$ as the conditional expectation of $Y_t^2$. When the regression (slope) parameter $\gamma_1$ is known, $\gamma_0$ may simply be estimated by exploiting the linear relationship (12) for $\log y_t^2$. Thus, for known $\gamma_1$, we solve the moment equation

$$S_3 - \mathrm{E}(S_3) = 0 \tag{26}$$

for $\gamma_0$, where

$$S_3 = \frac{1}{T} \sum_{t=1}^{T} \log \left(y_t^2\right).$$

This yields a closed form moment estimating equation for $\gamma_0$ given by

$$\gamma_0 = \frac{\left\{(S_3 - \kappa_1)(1 - \gamma_1) - \frac{1}{T} \log(\sigma_1^2) \left(1 - \gamma_1^T\right)\right\} (1 - \gamma_1)}{1 - \gamma_1 - \frac{1}{T}\left(1 - \gamma_1^T\right)},$$

where $\kappa_1 = E[\log \epsilon_t^2] = -1.270363$ as in (12).

### 3.1.1 Algorithm

For convenience, we provide an algorithm indicating how to solve the estimating equations (21) for $\sigma_\eta^2$ and (25) for $\gamma_1$, while $\gamma_0$ would be estimated by using the closed form formula in (27).

**Step 1**: For small initial values $\gamma_1 = \gamma_{10}$ and $\sigma_1^2 = \sigma_{10}^2$, we calculate an initial value for $\gamma_0$, say $\gamma_{00}$, by (26) and then choose the initial value of $\sigma_\eta^2 = \sigma_{\eta 0}^2$ by solving the asymptotic unbiased estimating equation (23). To be specific,

$$\sigma_{\eta 0}^2 = 2\left(\ln(S_{10}) - \frac{\gamma_{00}}{1 - \gamma_{10}}\right)(1 - \gamma_{10}^2). \tag{27}$$

**Step 2**: Once the initial values are chosen or computed as in Step 1, we solve $S_2 - g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2) = 0$ (25) iteratively to obtain an improved value for $\gamma_1$. The iterative equation has the form

$$\hat{\gamma}_1(r+1) = \hat{\gamma}_1(r) + \left[\left(\frac{\partial g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2)}{\partial \gamma_1}\right)^{-1}\left(S_2 - g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2)\right)\right]_{[r]} \quad (28)$$

where $\widehat{\gamma_1}(r)$ is a value of $\gamma_1$ at $r$th iteration, and $[.]_{\hat{\gamma}_1(r)}$ is the value of the expression in the square bracket evaluated at $\gamma_1 = \hat{\gamma}_1(r)$. Then by using $\log \sigma_1^2 = \frac{\gamma_0}{1-\hat{\gamma}_1}$, $\gamma_0$ is updated by (3.1).

**Step 3**: The estimate of $\gamma_0$ and $\gamma_1$ obtained from Step 2 is then used to solve $S_1 - g_1(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2) = 0$ (21) iteratively to obtain an improvement over $\sigma_{\eta 0}^2$ . The iterative equation has the form

$$\hat{\sigma}_\eta^2(r+1) = \hat{\sigma}_\eta^2(r) + \left[\left(\frac{\partial g_1(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2)}{\partial \sigma_\eta^2}\right)^{-1}\left(S_1 - g_1(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2)\right)\right]_{[r]} \quad (29)$$

where $\widehat{\sigma_\eta^2}(r)$ is the value of $\sigma_\eta^2$ at $r$th iteration, and $[.]_{\hat{\sigma}_\eta^2(r)}$ is the value of the expression in the square bracket evaluated at $\sigma_\eta^2 = \hat{\sigma}_\eta^2(r)$.

This three step cycle of iteration continues until convergence. Let the final estimates obtained from (27), (28) and (29) are denoted by $\hat{\gamma}_{0,MM}$, $\hat{\gamma}_{1,MM}$ and $\hat{\sigma}^2_{\eta,MM}$ respectively.

### 3.2 Remarks on Large Sample Moment Estimation

In Sect. 3.1, we have introduced a moment estimation technique for the estimation of the parameters involved in the volatile time series (1)–(2) with finite length. In practice such as in financial time series analysis, one may deal with a large time series such as a series with $T = 5000, 10,000$ or more. To analyze this type of volatile time series with infinite length, one may use the asymptotic moment properties of the selected moment functions for $\gamma_1$ and $\sigma_\eta^2$ and develop much more simpler moment estimating equations for the desired parameters. The intercept parameter $\gamma_0$ in the volatility model (2), is still estimated by (26), because it is already a closed-form solution for $\gamma_0$.

To be specific, for given $\gamma_0 = \gamma_{00}$ and $\gamma_1 = \gamma_{10}$, we solve the asymptotic property based moment equation $S_{10} - g_{10}(.) = 0$ (27), for $\sigma_\eta^2$.

Note that, to construct an asymptotic moment function based estimating equation for $\gamma_1$, we use $S_{20} = \dfrac{1}{T - T_0 - 1}\displaystyle\sum_{t=T_0+1}^{T} y_{t-1}^2 y_t^2$ and similar to (22) compute the asymptotic expectation of $S_{20}$ by using

$$\lim_{t\to\infty} E[y_{t-1}^2 y_t^2] = \lim_{t\to\infty} g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2) = \exp\left[\frac{2\gamma_0 + \sigma_\eta^2}{1-\gamma_1}\right]$$
$$= g_{20}(\gamma_0, \gamma_1, \sigma_\eta^2), \quad say, \tag{30}$$

where the formula for $g_2(\gamma_0, \gamma_1, \sigma_\eta^2, \sigma_1^2)$ is given in (24). Hence the asymptotic moment based unbiased estimating equation is written as

$$S_{20} - g_{20}(\gamma_0, \gamma_1, \sigma_\eta^2) = 0. \tag{31}$$

which yields a closed-form estimating equation for $\gamma_1$ as

$$\gamma_1 = 1 - \frac{2\gamma_0 + \sigma_\eta^2}{\ln(S_{20})}. \tag{32}$$

Next, the estimate of $\gamma_1$ obtained from (33) is used in (27) to compute an improved estimate for $\gamma_0$. This improved value of $\gamma_0$ and the improved value of $\gamma_1$ from (32), are then used in (27) for improved estimation of $\sigma_\eta^2$. This cycle of iterations continues until convergence, yielding large sample based moment estimators for all three parameters.

### *3.3 A GQL (Generalized Quasi-Likelihood) Approximation*

Note that there exists a generalized quasi-likelihood (GQL) approach which always produces more efficient estimates than the MM and GMM approaches. For example, one may refer to Rao et al. (2012) for an efficiency comparison between these two approaches in a linear panel data setup. Following Sutradhar (2004), for known $\gamma_0$, the GQL estimating equations for the main volatility parameters $\gamma_1$ and $\sigma_\eta^2$, may be constructed as follows. The $\gamma_0$ parameter may still be estimated using the moment equation (27). For the GQL estimation of $\sigma_\eta^2$, one may minimize a generalized distance function in $y_t^2$ $(t = 1, \ldots, T)$, whereas $\sum_{t=1}^T y_t^2$ was equated to its expectation to obtain its estimate by the MM approach. Similarly, a generalized distance function in $y_{t-1}^2 y_t^2$ $(t = 2, \ldots, T)$ may be minimized to obtain the GQL estimate of $\gamma_1$, whereas $\sum_{t=2}^T y_{t-1}^2 y_t^2$ was equated to its mean to obtain its MM estimate. To be specific, for

$$\mathrm{u} = [y_1^2, \ldots, y_t^2, \ldots, y_T^2]', \quad \text{and} \quad \mathrm{v} = [y_1^2 y_2^2, \ldots, y_{t-1}^2 y_t^2, \ldots, y_{T-1}^2 y_T^2]', \tag{33}$$

and their corresponding means

$$\lambda = E[U] = [\lambda_1, \ldots, \lambda_t, \ldots, \lambda_T]', \quad \text{and}$$
$$\psi = E[V] = [\psi_{12}, \ldots, \psi_{t-1,t}, \ldots, \psi_{T-1,T}]', \tag{34}$$

and covariances

$$\text{and} \qquad \Sigma = \text{Cov}(U), \qquad \Omega = \text{Cov}(V), \tag{35}$$

the GQL estimating equations (Sutradhar 2004) for $\sigma_\eta^2$ and $\gamma_1$ have the forms

$$\frac{\partial \lambda'}{\partial \sigma_\eta^2} \Sigma^{-1}(\text{u} - \lambda) = 0, \tag{36}$$

and

$$\frac{\partial \psi'}{\partial \gamma_1} \Omega^{-1}(\text{v} - \psi) = 0, \tag{37}$$

respectively. For the computation of (37) and (38), the general elements of $\lambda$ and $\psi$ have the formulas

$$\begin{aligned} \lambda_t &= E[Y_t^2] = E_{\sigma_t^2} E[Y_t^2 | \sigma_t^2] \\ &= \begin{cases} \sigma_1^2 & \text{for } t = 1 \\ \exp\left[\frac{\gamma_0}{1-\gamma_1} + \gamma_1^{t-1}(\log \sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}) + \frac{\sigma_\eta^2}{2} \sum_{r=0}^{t-2} \gamma_1^{2r}\right] & \text{for } t = 2, \ldots, T, \end{cases} \end{aligned} \tag{38}$$

[see also (20)] and

$$\begin{aligned} \psi_{t-1,t} &= E[Y_{t-1}^2 Y_t^2] = E_{\sigma_{t-1}^2, \sigma_t^2} E[Y_{t-1}^2 Y_t^2 | \sigma_{t-1}^2 \sigma_t^2] \\ &= \begin{cases} \sigma_1^2 \;\; \exp\left[\gamma_1 \log \sigma_1^2 + \frac{\sigma_\eta^2}{2}\right] & \text{for } t = 2 \\ \exp\left[\frac{2\gamma_0}{1-\gamma_1} + \gamma_1^{t-2}\left(\log \sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}\right) + \gamma_1^{t-1}\left(\log \sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}\right) \right. \\ \left. + \frac{\sigma_\eta^2}{2}\left((1+\gamma_1)^2 \sum_{l=0}^{t-3} \gamma_1^{2l} + 1\right)\right] & \text{for } t = 3, \ldots, T, \end{cases} \end{aligned} \tag{39}$$

respectively.

Note that even though the elements of the covariance matrix $\Sigma$ in (37) may be computed easily, the computation for the elements of $\Omega$ matrix in (38) is, however, difficult. We therefore propose an approximation to the GQL estimating equations (37) and (38) by replacing $\Sigma$ and $\Omega$ matrices with their corresponding diagonal forms, namely $\Sigma_d$, and $\Omega_d$, and solve the approximate GQL equations

$$\frac{\partial \lambda'}{\partial \sigma_\eta^2} \Sigma_d^{-1}(\mathrm{u} - \lambda) = 0, \tag{40}$$

and

$$\frac{\partial \psi'}{\partial \gamma_1} \Omega_d^{-1}(\mathrm{v} - \psi) = 0, \tag{41}$$

for $\sigma_\eta^2$ and $\gamma_1$, respectively, where

$$\Sigma_d = \mathrm{diag}[\mathrm{Var}(Y_1^2), \ldots, \mathrm{Var}(Y_t^2), \ldots, \mathrm{Var}(Y_T^2)], \tag{42}$$

with

$$\begin{aligned} \sigma_{tt} &= \mathrm{Var}[Y_t^2] \\ &= \begin{cases} 3\sigma_1^4 - \lambda_1^2 & \text{for } t = 1 \\ 3\exp\left[\frac{2\gamma_0}{1-\gamma_1} + 2\gamma_1^{t-1}\left(\log\sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}\right) + 2\,\sigma_\eta^2[\sum_{r=0}^{t-2}\gamma_1^{2r}]\right] & \\ -\lambda_t^2 & \text{for } t = 2, \ldots, T; \end{cases} \end{aligned} \tag{43}$$

and

$$\Omega_d = \mathrm{diag}[\mathrm{Var}(Y_1^2 Y_2^2), \ldots, \mathrm{Var}(Y_{t-1}^2 Y_t^2), \ldots, \mathrm{Var}(Y_{T-1}^2 Y_T^2)], \tag{44}$$

with

$$\begin{aligned} &Var[Y_{t-1}^2 Y_t^2] \\ &= \begin{cases} 9\,\sigma_1^4 \exp\left[\frac{2\gamma_0}{1-\gamma_1} + 2\gamma_1\left(\log\sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}\right) + 2\,\sigma_\eta^2\right] - \psi_{12}^2 & \text{for } t = 2 \\ 9\ \exp\left[\frac{4\gamma_0}{1-\gamma_1} + 2\left(\gamma_1^{t-2}\left(\log\sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}\right) + \gamma_1^{t-1}\left(\log\sigma_1^2 - \frac{\gamma_0}{1-\gamma_1}\right)\right)\right. & \\ \left. +2\,\sigma_\eta^2\left((1+\gamma_1)^2\ \sum_{l=0}^{t-3}\gamma_1^{2l} + 1\right)\right] & \\ -\psi_{t-1,t}^2 & \text{for } t = 3, \ldots, T, \end{cases} \end{aligned} \tag{45}$$

where $\psi_{t-1,t}$ is given in (39).

Remark that $\gamma_0$ is still updated by (3.1) after each estimation of $\gamma_1$. Also remark that the approximate GQL estimating equations for $\gamma_1$ (41) and $\sigma_\eta^2$ (40) are similar to the well known weighted least square (WLS) equations for the corresponding parameters.

In (41), for $t = 2$, the first derivative of $\psi_{t-1,t}$ w.r.t $\gamma_1$ is given by

$$\frac{\partial \psi_{1,2}}{\partial \gamma_1} = \psi_{1,2}\ \log\sigma_1^2,$$

and for a general t $= 3, \ldots, T$, the derivative has the expression as

$$\frac{\partial \psi_{t-1,t}}{\partial \gamma_1} = \psi_{t-1,t}\Bigg[\frac{\gamma_0}{(1-\gamma_1)^2}\left[2-(1+\gamma_1)\gamma_1^{t-2}\right] - \frac{\gamma_0}{1-\gamma_1}\left[(t-2)(1+\gamma_1)\gamma_1^{t-3}+\gamma_1^{t-2}\right]$$
$$(t-2)\gamma_1^{t-3}\log\sigma_1^2 + (t-1)\gamma_1^{t-2}\log\sigma_1^2$$
$$+\frac{\sigma_\eta^2}{2}\left(2\,(1+\gamma_1)\sum_{l=0}^{t-3}\gamma_1^{2l} + (1+\gamma_1)^2\sum_{l=0}^{t-3}(2l)\,\gamma_1^{(2l-1)}\right)\Bigg].$$

In (40), the derivative of $\lambda_t$ w.r.t $\sigma_\eta^2$ has the formula

$$\frac{\partial \lambda_t}{\partial \sigma_\eta^2} = \begin{cases} 0 & \text{for } t = 1 \\ \frac{1}{2}\lambda_t \sum_{r=0}^{t-2}\gamma_1^{2r} & \text{for } t = 2, \ldots, T. \end{cases}$$

## 3.4 A Modified QML Approach

Recall from Sect. 2.2 that for known $\gamma_0$, the QML (quasi maximum likelihood) estimates for the volatility parameters $\gamma_1$ and $\sigma_1^2$ are obtained by solving their quasi score equations given in (15), where $\Phi$ is the covariance matrix of $\mathbf{Z} = (Z_1, \ldots, Z_t, \ldots, Z_T)'$, with $z_t = \kappa_1 + \log\sigma_t^2 + u_t$ as in (12), where

$$u_t \sim \log\chi^2(0, \kappa_2),$$

with $\kappa_2 = \pi^2/2$. Note that the score equations in (15) involves $\Phi^{-1}$. See also (17) and (18) for the exact formulas of the first order derivatives. Because $T$ is large, even though the elements $\nu_{ut}$ of the $\Phi$ matrix are known, the computation of the inverse matrix, namely $\Phi^{-1}$ is, however, cumbersome and time consuming. We suggest to approximate the covariance matrix $\Phi$ in (15) by a tridiagonal matrix as

$$\Phi = (\nu_{ut}) \approx \begin{pmatrix} \Phi_{11} & \Phi_{12} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ \Phi_{12} & \Phi_{22} & \Phi_{23} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \Phi_{23} & \Phi_{33} & \Phi_{34} & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \Phi_{34} & \Phi_{44} & \Phi_{45} & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \Phi_{T-2,T-1} & \Phi_{T-1,T-1} & \Phi_{T-1,T} \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \Phi_{T-1,T} & \Phi_{TT} \end{pmatrix}. \quad (46)$$

This makes the computation of $\Phi^{-1}$ and other multiplication of large dimensional matrices, necessary to compute the first and second order derivatives with respect to $\gamma_1$ and $\sigma_\eta^2$, quite manageable. In fact following Usmani (1994), one can analytically derive the formulas for the determinant and inverse of the tridiagonal matrix $\Phi$ as follows.

Note that the tridiagonal matrix $\Phi$ in (47) is symmetric. For $k = 1, \ldots, T$, let $\det[\Phi]_{\{1,\cdots,k\}}$ denotes the $k$th principal minor, that is, $[\Phi]_{\{1,\cdots,k\}}$ is the submatrix formed by the first $k$ rows and columns of $\Phi$. By using conventional notation $\det[\Phi]_{-1} = 0$ and $\det[\Phi]_0 = 1$, one may compute the determinant of $\Phi$ by using the recursive formula

$$\begin{aligned}\det[\Phi]_{\{1,\cdots,k\}} &= \Phi_{kk} \det[\Phi]_{\{1,\cdots,k-1\}} - \Phi_{k,k-1}\Phi_{k-1,k} \det[\Phi]_{\{1,\cdots,k-2\}} \\ &= \Phi_{kk} \det[\Phi]_{\{1,\cdots,k-1\}} - \Phi^2_{k-1,k} \det[\Phi]_{\{1,\cdots,k-2\}}. \end{aligned} \tag{47}$$

Now by writing $\theta_i = \det[\Phi]_{\{1,\cdots,i\}}$, $i = 1, \cdots, T$, and using the sequence $\{\phi_i\}$ defined by the recurrence formula

$$\phi_i = \Phi_{ii}\phi_{i+1} - \Phi_{i,i+1}\Phi_{i+1,i}\phi_{i+2}, \;\; \phi_{T+1} = 1, \;\; \phi_{T+2} = 0, (i = T, T-1, \cdots, 3, 2, 1), \tag{48}$$

and following Usmani (1994), for example, one derives the inverse of the tridiagonal matrix $\Phi$ as

$$\left(\Phi^{-1}\right)_{ij} = \begin{cases} (-1)^{i+j}\Phi_{i,i+1}\Phi_{i+1,i+2}\cdots\Phi_{j-1,j}\theta_{i-1}\phi_{j+1}/\theta_T, & i < j, \\ \theta_{i-1}\phi_{i+1}/\theta_T, & i = j, \\ (-1)^{i+j}\Phi_{j+1,j}\Phi_{j+2,j+1}\cdots\Phi_{i,i-1}\theta_{j-1}\phi_{i+1}/\theta_T, & i > j. \end{cases} \tag{49}$$

However, there is still some computational burden with the formula (50) for computing the inverse. This is because, in financial time series, for a $T$ as large as several hundred or thousand, the computation of the determinants $|\Phi|$, which appears in the formula for every element of $\Phi^{-1}$, can be cumbersome and time consuming. To address this computational difficulty, we make further simplifications as follows.

Using the notation $\theta_i$, we rewrite the determinant in (48) as

$$\theta_i = \Phi_{ii}\theta_{i-1} - \Phi_{i,i-1}\Phi_{i-1,i}\theta_{i-2}, \; \theta_{-1} = 0, \; \theta_0 = 1, \; i = 1, \cdots, T, \tag{50}$$

which by division of both sides by $\theta_{i-1}$, yields

$$\frac{\theta_i}{\theta_{i-1}} = \Phi_{ii} - \Phi_{i,i-1}\Phi_{i-1,i}\left(\frac{\theta_{i-1}}{\theta_{i-2}}\right)^{-1}. \tag{51}$$

For convenience, we use $u_i = \frac{\theta_i}{\theta_{i-1}}$ and re-express (52) as

$$u_i = \Phi_{ii} - \Phi_{i,i-1}\Phi_{i-1,i}u^{-1}_{i-1}, \; u_1 = \frac{\theta_1}{\theta_0} = \theta_1 = \Phi_{11}, \; i = 2, \cdots, T. \tag{52}$$

Similarly, dividing both sides of (48) by $\phi_{i+1}$, we obtain

$$\frac{\phi_i}{\phi_{i+1}} = \Phi_{ii} - \Phi_{i,i+1}\Phi_{i+1,i}\frac{\phi_{i+2}}{\phi_{i+1}} \Rightarrow \frac{\phi_{i+1}}{\phi_i} = \frac{1}{\Phi_{ii} - \Phi_{i,i+1}\Phi_{i+1,i}\frac{\phi_{i+2}}{\phi_{i+1}}}, \tag{53}$$

which by using $v_i = \frac{\phi_{i+2}}{\phi_{i+1}}$, we re-express as

$$v_{i-1} = \frac{1}{\Phi_{ii} - \Phi_{i,i+1}\Phi_{i+1,i}v_i}, \quad v_{T-1} = \frac{1}{\Phi_{TT}}, \quad v_T = \frac{\phi_{T+2}}{\phi_{T+1}} = 0,$$

for $i = 2, \cdots, T-1$.

Note that $u_i$'s are the ratios of the neighbouring principal minors, and $v_i$'s have similar interpretation for the submatrices starting from the opposite corner of the matrix $\Phi$, so they are easy to compute. Next because $\Phi$ is a tridiagonal matrix, its determinant is the summation of a series of two types of terms, where the first type of terms will contain $\Phi_{j,j+1}$ and $\Phi_{j+1,j}$ [or symmetrically $\Phi_{j-1,j}$ and $\Phi_{j,j-1}$] appearing together, and the second type of terms can not contain $\Phi_{j,j+1}$ and $\Phi_{j+1,j}$ together. Also note that $\Phi_{j,j}$ can not appear together with any of $\Phi_{j-1,j}$ and $\Phi_{j,j-1}$, or any of $\Phi_{j,j+1}$ and $\Phi_{j+1,j}$. Thus the determinant can be obtained as

$$\begin{aligned} |\Phi| &= \text{summation of terms with } \Phi_{j,j+1} \text{ and } \Phi_{j+1,j} \\ &\quad + \text{summation of terms without } \Phi_{j,j+1} \text{ and } \Phi_{j+1,j} \\ &= \theta_{j-1}[-\Phi_{j+1,j}\Phi_{j,j+1}]\phi_{j+2} + \theta_j\phi_{j+1}, \end{aligned}$$

yielding

$$\theta_T = \theta_j\phi_{j+1} - \Phi_{j+1,j}\Phi_{j,j+1}\theta_{j-1}\phi_{j+2},$$

for any $j = T, T-1, \cdots, 2, 1$ [ see also Usmani (1994)]. Now by applying (56), (55), and (53), into (50), one obtains the diagonal elements of the inverse matrix as

$$\begin{aligned} \left(\Phi^{-1}\right)_{ii} &= \frac{\theta_{i-1}\phi_{i+1}}{\theta_i\phi_{i+1} - \Phi_{i+1,i}\Phi_{i,i+1}\theta_{i-1}\phi_{i+2}} = \frac{1}{\frac{\theta_i}{\theta_{i-1}} - \Phi_{i+1,i}\Phi_{i,i+1}\frac{\phi_{i+2}}{\phi_{i+1}}} \\ &= \frac{1}{u_i - \Phi_{i+1,i}\Phi_{i,i+1}v_i}, \text{ for } i < T, \end{aligned}$$

and

$$\left(\Phi^{-1}\right)_{TT} = \frac{\theta_{T-1}\phi_{T+1}}{\theta_T} = \frac{1}{u_T}$$

Similarly, the off-diagonal elements are computed as

$$\left(\Phi^{-1}\right)_{ij} = \frac{(-1)^{i+j}\Phi_{i,i+1}\Phi_{i+1,i+2}\cdots\Phi_{j-1,j}\theta_{i-1}\phi_{j+1}}{\theta_j\phi_{j+1} - \Phi_{j+1,j}\Phi_{j,j+1}\theta_{j-1}\phi_{j+2}} = \frac{(-1)^{i+j}\Phi_{i,i+1}\Phi_{i+1,i+2}\cdots\Phi_{j-1,j}}{\frac{\theta_j}{\theta_{i-1}} - \Phi_{j+1,j}\Phi_{j,j+1}\frac{\theta_{j-1}}{\theta_{i-1}}\frac{\phi_{j+2}}{\phi_{j+1}}}$$

$$= \frac{(-1)^{i+j}\Phi_{i,i+1}\Phi_{i+1,i+2}\cdots\Phi_{j-1,j}}{u_i\cdots u_j - \Phi_{j+1,j}\Phi_{j,j+1}u_i\cdots u_{j-1}v_j}, \text{ for } i < j < T,$$

and

$$\left(\Phi^{-1}\right)_{iT} = \frac{(-1)^{i+T}\Phi_{i,i+1}\Phi_{i+1,i+2}\cdots\Phi_{T-1,T}}{u_i\cdots u_T}.$$

Because $\Phi$ and $\Phi^{-1}$ are symmetric matrices, the above formulas define the whole $\Phi^{-1}$.

# 4 Estimation Performance: A Simulations Based Empirical Study

Recall from Sect. 2.1 that the existing GMM is a highly cumbersome approach and it is not practically worth using specially when an alternative simpler approach becomes available which produces consistent estimates for the parameters of the stochastic volatility model. Furthermore as discussed in Sect. 2.2 [see also Sect. 3.4], it is well known that the QML approach provides approximate estimates, but it is highly competitive to the GMM approach and much simpler technically. For these reasons in the proposed simulation study, we do not include the GMM approach, rather make a comparative study between the QML and the proposed MM approach (Sect. 3.1) for a selected set of parameter values under the SV model. This we do in Sect. 4.2. The proposed MM approach appears to work better than the QML approach. We then continue to examine the performance of the MM approach for more sets of parameter values of the SV model. The results of this intensive simulation study are reported in Sect. 4.3. In the same simulation study, we also include the approximate GQL approach discussed in Sect. 3.3. The simulation design for both of these aforementioned studies is given in Sect. 4.1.

## *4.1 Simulation Design*

**(a)** We use volatility parameters $\gamma_0 = 0.05, 0.2$; $\gamma_1 = 0.5$; and $\sigma_\eta^2 = 0.25, 1.0$ to generate the data by model (1)–(2) for the comparative study in Sect. 4.2. For time series length we consider $T = 1000$ and 3000.

**(b)** To examine the performance of the MM approach through an additional intensive simulation study in Sect. 4.3, we consider more sets of parameter values as: $\gamma_0 = 0.05, 0.2$; $\gamma_1 = 0.25, 0.5$; and $\sigma_\eta^2 = 0.25, 0.5,$ and 1.0. For time series length we consider $T = 200, 500, 1000, 3000, 5000,$ and 10,000.

**(c)** For both studies we use 500 simulations.

## *4.2 Relative Performance of the MM and QML Approaches*

In a financial time series, $T$ is usually large such as $T = 1000$ or more. To examine the relative performance of the existing QML approach with the proposed MM approach we consider large time series with length $T = 1000$ and 3000. Recall that the GMM approach is not included in the simulation because of its cumbersomeness. Moreover, to obtain reasonably good estimates for the volatility parameters, the GMM approach requires the time series length infinitely large such as $T = 10{,}000$ or 15,000, which is another major limitation.

To generate the desired time series satisfying the volatility model (1)–(2), one starts with a value for $\sigma_1^2$. Note that as $\log \sigma_1^2$ is assumed to have a normal distribution with mean $\gamma_0/(1 - \gamma_1)$ as shown in (6), we choose the value of $\gamma_0/(1 - \gamma_1)$ for $\log(\sigma_1^2)$ in our simulation study. Now to obtain the MM estimates for the large sample case ($T = 1000$ or more), we solve the large sample based estimating equations (31) and (27) for $\gamma_1$ and $\sigma_\eta^2$, respectively, and still update $\gamma_0$ by (3.1) after obtaining new $\gamma_1$. For selected values of volatility parameters, namely for $\gamma_0 = 0.05, 0.2$; $\gamma_1 = 0.5$; and $\sigma_\eta^2 = 0.25$ and 1.0, the simulated means (SM), simulated standard errors (SSE), along with mean squared errors (MSE), for the MM estimates for these three parameters, based on 500 simulations, are reported in Table 1.

Next, the QML estimates for $\gamma_1$ and $\sigma_\eta^2$ are obtained by solving the two equations in (15), whereas the QML estimate of $\gamma_0$ is obtained from (27) by using the QML estimate for $\gamma_1$. The SM, SSE, along with MSE of QML estimates for the same selected parameter values as in the aforementioned MM approach, based on 500 simulations, are given also in Table 1. The estimation results based on $T = 1000$ and 3000, are not as good as those from the MM approach. For example, for $\gamma_0 = 0.05$, $\gamma_1 = 0.5$, $\sigma_\eta^2 = 0.25$, and $T = 1000$ case, the QML estimates give $\hat{\gamma}_{0,QML} = 0.0491$, $\hat{\gamma}_{1,QML} = 0.433$ and $\hat{\sigma}^2_{\eta,QML} = 0.324$, which are all farther away from the true parameter values than the MM estimates of $\hat{\gamma}_{0,MM} = 0.0517$, $\hat{\gamma}_{1,MM} = 0.441$ and $\hat{\sigma}^2_{\eta,MM} = 0.260$, while the simulated MSEs for $\hat{\gamma}_{0,QML}$, $\hat{\gamma}_{1,QML}$ and $\hat{\sigma}^2_{\eta,QML}$ in QML approach are respectively 0.0032, 0.1648 and 0.0815, which are all greater than the corresponding MSEs in MM approach, which are 0.0027, 0.0846 and 0.0183, respectively. Notice that the MSEs for the estimates of $\gamma_1$ and $\sigma_\eta^2$ are almost double under the QML approach as compared to the MM approach. When $\sigma_\eta^2$ gets larger such as $\sigma_\eta^2 = 1.0$, the QML approach appears to produce better estimate for $\gamma_1$ as compared to the small $\sigma_\eta^2 = 0.25$ case. For example, when $\sigma_\eta^2 = 1.0$, the QML approach, when $T = 1000$, gives almost unbiased estimate for $\gamma_1 = 0.5$ as $\hat{\gamma}_{1,QML} = 0.493$ with MSE as 0.0407, whereas the MM yielded

**Table 1** Comparison of MM and QML estimates for selected parameter values based on 500 simulations

| $\gamma_0$ | $\gamma_1$ | $\sigma_\eta^2$ | Estimate | Method | Quantity | Time series length (T) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1000 | 3000 | 5000 | 10,000 |
| 0.05 | 0.5 | 0.25 | $\hat{\gamma}_0$ | QML | SM | 0.0491 | 0.0433 | | |
| | | | | | SSE | 0.0569 | 0.0368 | | |
| | | | | | MSE | 0.0032 | 0.0014 | | |
| | | | | MM | SM | 0.0517 | 0.0495 | 0.0489 | 0.0492 |
| | | | | | SSE | 0.0524 | 0.0281 | 0.0237 | 0.0149 |
| | | | | | MSE | 0.0027 | 0.0008 | 0.0006 | 0.0002 |
| | | | $\hat{\gamma}_1$ | QML | SM | 0.4329 | 0.5171 | | |
| | | | | | SSE | 0.4059 | 0.3368 | | |
| | | | | | MSE | 0.1648 | 0.1134 | | |
| | | | | MM | SM | 0.4409 | 0.4826 | 0.4974 | 0.5058 |
| | | | | | SSE | 0.2909 | 0.2055 | 0.1768 | 0.1208 |
| | | | | | MSE | 0.0846 | 0.0422 | 0.0313 | 0.0146 |
| | | | $\hat{\sigma}_\eta^2$ | QML | SM | 0.3239 | 0.2616 | | |
| | | | | | SSE | 0.2855 | 0.2193 | | |
| | | | | | MSE | 0.0815 | 0.0481 | | |
| | | | | MM | SM | 0.2597 | 0.2456 | 0.2435 | 0.2457 |
| | | | | | SSE | 0.1353 | 0.0926 | 0.0822 | 0.0586 |
| | | | | | MSE | 0.0183 | 0.0086 | 0.0068 | 0.0034 |
| | | 1.0 | $\hat{\gamma}_0$ | QML | SM | 0.0451 | 0.0508 | | |
| | | | | | SSE | 0.0504 | 0.0282 | | |
| | | | | | MSE | 0.0025 | 0.0008 | | |
| | | | | MM | SM | 0.0530 | 0.0513 | 0.0538 | 0.0518 |
| | | | | | SSE | 0.0631 | 0.0332 | 0.0265 | 0.0182 |
| | | | | | MSE | 0.0040 | 0.0011 | 0.0007 | 0.0003 |
| | | | $\hat{\gamma}_1$ | QML | SM | 0.4929 | 0.4689 | | |
| | | | | | SSE | 0.2018 | 0.1125 | | |
| | | | | | MSE | 0.0407 | 0.0127 | | |
| | | | | MM | SM | 0.4187 | 0.4560 | 0.4705 | 0.4859 |
| | | | | | SSE | 0.2007 | 0.1387 | 0.1371 | 0.1120 |
| | | | | | MSE | 0.0403 | 0.0192 | 0.0188 | 0.0125 |
| | | | $\hat{\sigma}_\eta^2$ | QML | SM | 0.9960 | 1.0318 | | |
| | | | | | SSE | 0.4845 | 0.2974 | | |
| | | | | | MSE | 0.2347 | 0.0884 | | |
| | | | | MM | SM | 1.0280 | 1.0264 | 1.0088 | 0.9962 |
| | | | | | SSE | 0.3110 | 0.2063 | 0.2176 | 0.1738 |
| | | | | | MSE | 0.0967 | 0.0426 | 0.0473 | 0.0302 |

(continued)

**Table 1** (continued)

| $\gamma_0$ | $\gamma_1$ | $\sigma_\eta^2$ | Estimate | Method | Quantity | Time series length (T) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | 1000 | 3000 | 5000 | 10,000 |
| 0.20 | 0.5 | 0.25 | $\hat{\gamma}_0$ | QML | SM | 0.2340 | 0.1979 | | |
| | | | | | SSE | 0.1629 | 0.1285 | | |
| | | | | | MSE | 0.0265 | 0.0165 | | |
| | | | | MM | SM | 0.2224 | 0.2036 | 0.1978 | 0.1970 |
| | | | | | SSE | 0.1155 | 0.0764 | 0.0681 | 0.0471 |
| | | | | | MSE | 0.0133 | 0.0058 | 0.0046 | 0.0022 |
| | | | $\hat{\gamma}_1$ | QML | SM | 0.3794 | 0.4911 | | |
| | | | | | SSE | 0.4094 | 0.3299 | | |
| | | | | | MSE | 0.1676 | 0.1088 | | |
| | | | | MM | SM | 0.4250 | 0.4830 | 0.4994 | 0.5053 |
| | | | | | SSE | 0.2853 | 0.1945 | 0.1760 | 0.1223 |
| | | | | | MSE | 0.0814 | 0.0378 | 0.0310 | 0.0150 |
| | | | $\hat{\sigma}_\eta^2$ | QML | SM | 0.3697 | 0.2756 | | |
| | | | | | SSE | 0.3046 | 0.2263 | | |
| | | | | | MSE | 0.0928 | 0.0512 | | |
| | | | | MM | SM | 0.2631 | 0.2484 | 0.2452 | 0.2444 |
| | | | | | SSE | 0.1360 | 0.0946 | 0.0830 | 0.0590 |
| | | | | | MSE | 0.0185 | 0.0089 | 0.0069 | 0.0035 |
| | | 1.0 | $\hat{\gamma}_0$ | QML | SM | 0.1962 | 0.2102 | | |
| | | | | | SSE | 0.0857 | 0.0455 | | |
| | | | | | MSE | 0.0073 | 0.0021 | | |
| | | | | MM | SM | 0.2241 | 0.2071 | 0.2063 | 0.2044 |
| | | | | | SSE | 0.0942 | 0.0606 | 0.0566 | 0.0458 |
| | | | | | MSE | 0.0089 | 0.0037 | 0.0032 | 0.0021 |
| | | | $\hat{\gamma}_1$ | QML | SM | 0.4937 | 0.4689 | | |
| | | | | | SSE | 0.2043 | 0.1123 | | |
| | | | | | MSE | 0.0417 | 0.0126 | | |
| | | | | MM | SM | 0.4253 | 0.4768 | 0.4848 | 0.4907 |
| | | | | | SSE | 0.1997 | 0.1418 | 0.1361 | 0.1102 |
| | | | | | MSE | 0.0399 | 0.0201 | 0.0185 | 0.0121 |
| | | | $\hat{\sigma}_\eta^2$ | QML | SM | 0.9960 | 1.0318 | | |
| | | | | | SSE | 0.4900 | 0.2972 | | |
| | | | | | MSE | 0.2401 | 0.0883 | | |
| | | | | MM | SM | 1.0334 | 1.0004 | 0.9897 | 0.9945 |
| | | | | | SSE | 0.3054 | 0.2200 | 0.2113 | 0.1779 |
| | | | | | MSE | 0.0933 | 0.0484 | 0.0446 | 0.0316 |

$\hat{\gamma}_{1,MM} = 0.4187$ with MSE as 0.0403. But, the QML approach appears to produce an estimate 0.4689 for $\gamma_1$ with a larger bias when $T$ increased to 3000. This result is counter intuitive and indicates an convergence problem. This appears to hold for the $\gamma_0 = 0.2$ case as well. For this reason we did not run the QML approach for larger $T$ cases such as for $T = 5000$ and 10,000. However, to examine the continued performance of the MM approach we have run only the MM approach for the larger $T$ cases, and the results are reported in the same Table 1. The results of the table show that as the time series length $T$ increases, the MM approach, unlike the QML approach, consistently produces better estimates.

### *4.3 Further Simulations for the MM Versus Approximate GQL (AGQL) Approach*

The results from the simulation study conducted in the last section revealed the superiority of the proposed MM approach to the QML approach for the estimation of the parameters of the SV model (1)–(2). Note that this relative performance was examined based on the time series length $T = 1000,\ 3000$, and in addition further individual performance of the MM approach was studied for much larger time series with $T = 5000,$ and 10,000. In this section, we carry out another simulation study mainly by selecting more combinations of parameter values in order to examine whether the MM approach really produces almost unbiased estimates for a wide range of parameter values. In this simulation study we also consider small time series with length $T = 200,$ and 500. Further recall that in Sect. 3.3 we have discussed an approximation to the so-called GQL approach (AGQL) which is expected to be highly competitive to the MM approach. For the sake of completeness, we include this AGQL approach in the present simulation study but examine its relative performance for time series of length $T = 200, 500, 1000,$ and 3000. This we have done because similar to the QML approach, the AGQL approach also takes longer computing times as compared to the MM approach. Under the AGQL approach, the estimates of $\sigma_\eta^2$ and $\gamma_1$ are obtained as the solutions of (41) and (38), respectively, whereas $\gamma_0$ is still estimated by moment equation (27).

With regard to the performance of the MM approach, the results of Table 2 are similar to those in Table 1, indicating that the MM approach performs well in large sample case in estimating all parameters of the SV model (1)–(2) for all reasonable range for parameter values. For example, when $\gamma_0 = 0.10,\ \gamma_1 = 0.25,\ \sigma_\eta^2 = 1.0$, the results of Table 2 show that the MM approach provides almost unbiased estimates $\hat{\gamma}_0 = 0.1067$ for $\gamma_0 = 010$ and $\hat{\sigma}_\eta^2 = 0.971$ for $\sigma_\eta^2 = 1.0$ with time series length $T = 1000$. These estimates improve significantly when $T$ gets larger such as when $T = 10{,}000$, the estimates are $\hat{\gamma}_0 = 0.1012$ and $\hat{\sigma}_\eta^2 = 0.9955$ which are not so different than the corresponding true values. The estimate of $\gamma_1$ (0.2081) is not so good when $T = 1000$, but improves highly to 0.2428 for larger $T = 10{,}000$. For small sample cases such as when $T = 200,\ 500$, the MM estimates are in general slightly biased, as expected. When the MM approach is compared to the AGQL approach, they appear to be highly competitive to each other, the AGQL approach is being slightly better as it produces almost the same estimates as that of the MM approach but with slightly smaller standard errors. For example, for the above parameter values, the AGQL approach produces $\hat{\gamma}_1 = 0.2309$ with standard error $SSE = 0.1190$ when $T = 3000$, whereas the MM approach produced $\hat{\gamma}_1 = 0.2221$ with $SSE = 0.1207$. The AGQL approach is, however, slightly more involved computationally as compared to the MM approach. Thus, over all the MM approach performs much better than the existing MQL approach, and it is highly competitive to the AGQL approach. Also as demonstrated in Sect. 2.1, the existing GMM approach is extremely cumbersome . This makes the proposed MM approach as the best possible approach for the estimation of all parameters of the standard volatility models.

## 5 Asymptotic Properties of the MM Estimators

Because the MM and AGQL estimators of $\gamma_0$, $\gamma_1$, and $\sigma_\eta^2$ are obtained by solving unbiased moment equations, these estimators are asymptotically unbiased to their corresponding true parameter values. This property is also evident from the simulation results in Tables 1 and 2. For example, for $T = 10{,}000$, the MM estimates for all three parameters are very close to the corresponding true parameter values. The AGQL estimates in Table 2 for $T = 3000$ also appear to be close to their true parameter values. For $T = 10{,}000$, the estimates will improve naturally but not shown to save computational time.

In Tables 1 and 2 we have also given the simulated standard errors (SSE) of the estimators. In practice, one may however need to compute the estimate for the variance of the estimator, specially for large size financial time series. For the purpose, we show below how to compute the formulas for the asymptotic variances of the MM estimators of $\gamma_1$ and $\sigma_\eta^2$, for example. These variances may then be estimated simply by replacing the parameters with their corresponding MM estimates obtained from Sects. 3.1 and 3.2.

### 5.1 Asymptotic Variance of the Estimator of $\gamma_1$

For known $\gamma_0$, it follows from (32) that the asymptotic variance of the MM estimator of $\gamma_1$ is given by

$$\lim_{T_0\to\infty} \text{Var}(\widehat{\gamma}_{1MM}) = \lim_{T_0\to\infty} \left[ \left( \frac{\partial g_{20}(\gamma_0, \gamma_1, \sigma_\eta^2)}{\partial \gamma_1} \right)^{-2} \text{Var}(S_{20}) \right], \tag{54}$$

where

**Table 2** Comparison of MM and AGQL estimates for selected parameter values based on 500 simulations

| | | | | | | Time series length (T) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\gamma_0$ | $\gamma_1$ | $\sigma_\eta^2$ | Estimate | Method | Quantity | 200 | 500 | 1000 | 3000 | 5000 | 10,000 |
| 0.05 | 0.25 | 0.50 | $\hat{\gamma}_0$ | AGQL | SM | 0.0309 | 0.0454 | 0.0514 | 0.0507 | | |
| | | | | | SSE | 0.1154 | 0.0846 | 0.0604 | 0.0352 | | |
| | | | | MM | SM | 0.0361 | 0.0434 | 0.0486 | 0.0508 | 0.0503 | 0.0498 |
| | | | | | SSE | 0.1201 | 0.0835 | 0.0434 | 0.0351 | 0.0241 | 0.0188 |
| | | | $\hat{\gamma}_1$ | AGQL | SM | 0.2273 | 0.2176 | 0.2240 | 0.2493 | | |
| | | | | | SSE | 0.3205 | 0.2762 | 0.2347 | 0.1485 | | |
| | | | | MM | SM | 0.2180 | 0.2415 | 0.2282 | 0.2505 | 0.2462 | 0.2485 |
| | | | | | SSE | 0.3501 | 0.3133 | 0.2463 | 0.1478 | 0.1288 | 0.0921 |
| | | | $\hat{\sigma}_\eta^2$ | AGQL | SM | 0.5285 | 0.4846 | 0.4680 | 0.4813 | | |
| | | | | | SSE | 0.2849 | 0.2046 | 0.1485 | 0.0883 | | |
| | | | | MM | SM | 0.5195 | 0.4706 | 0.4644 | 0.4802 | 0.4880 | 0.4934 |
| | | | | | SSE | 0.2937 | 0.2209 | 0.1552 | 0.0873 | 0.0708 | 0.0493 |
| | | 1.0 | $\hat{\gamma}_0$ | AGQL | SM | 0.0343 | 0.0506 | 0.0495 | 0.0514 | | |
| | | | | | SSE | 0.1390 | 0.0972 | 0.0660 | 0.0359 | | |
| | | | | MM | SM | 0.0425 | 0.0479 | 0.0486 | 0.0508 | 0.0500 | 0.0511 |
| | | | | | SSE | 0.1527 | 0.0989 | 0.0658 | 0.0351 | 0.0278 | 0.0198 |
| | | | $\hat{\gamma}_1$ | AGQL | SM | 0.2273 | 0.2176 | 0.2240 | 0.2493 | | |
| | | | | | SSE | 0.2878 | 0.2386 | 0.2032 | 0.1245 | | |
| | | | | MM | SM | 0.1664 | 0.1978 | 0.2087 | 0.2279 | 0.2376 | 0.2502 |
| | | | | | SSE | 0.3000 | 0.2491 | 0.2067 | 0.1236 | 0.1133 | 0.0996 |
| | | | $\hat{\sigma}_\eta^2$ | AGQL | SM | 0.9621 | 0.9607 | 0.9718 | 1.0010 | | |
| | | | | | SSE | 0.3747 | 0.2769 | 0.2014 | 0.1198 | | |
| | | | | MM | SM | 0.9458 | 0.9485 | 0.9676 | 0.9947 | 0.9970 | 0.9841 |
| | | | | | SSE | 0.3770 | 0.2889 | 0.2066 | 0.1186 | 0.1111 | 0.1008 |
| 0.10 | 0.25 | 0.50 | $\hat{\gamma}_0$ | AGQL | SM | 0.0767 | 0.0921 | 0.1006 | 0.1006 | | |
| | | | | | SSE | 0.1215 | 0.0953 | 0.0693 | 0.0382 | | |
| | | | | MM | SM | 0.0911 | 0.0932 | 0.0969 | 0.0995 | 0.1006 | 0.1000 |
| | | | | | SSE | 0.1354 | 0.0979 | 0.0664 | 0.0386 | 0.0293 | 0.0201 |
| | | | $\hat{\gamma}_1$ | AGQL | SM | 0.2345 | 0.1933 | 0.2226 | 0.2395 | | |
| | | | | | SSE | 0.3205 | 0.2721 | 0.2329 | 0.1461 | | |
| | | | | MM | SM | 0.2289 | 0.2117 | 0.2438 | 0.2473 | 0.2438 | 0.2474 |
| | | | | | SSE | 0.3400 | 0.3053 | 0.2455 | 0.1524 | 0.1247 | 0.0859 |
| | | | $\hat{\sigma}_\eta^2$ | AGQL | SM | 0.5086 | 0.4919 | 0.4796 | 0.4905 | | |
| | | | | | SSE | 0.2868 | 0.1997 | 0.1418 | 0.0870 | | |
| | | | | MM | SM | 0.4937 | 0.4755 | 0.4726 | 0.4877 | 0.4921 | 0.4964 |
| | | | | | SSE | 0.2903 | 0.2044 | 0.1551 | 0.0903 | 0.0756 | 0.0487 |

(continued)

**Table 2** (continued)

| $\gamma_0$ | $\gamma_1$ | $\sigma_\eta^2$ | Estimate | Method | Quantity | Time series length (T) 200 | 500 | 1000 | 3000 | 5000 | 10,000 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.10 | 0.25 | 1.0 | $\hat{\gamma}_0$ | AGQL | SM | 0.0883 | 0.1193 | 0.1047 | 0.1031 | | |
| | | | | | SSE | 0.1442 | 0.0965 | 0.0731 | 0.0398 | | |
| | | | | MM | SM | 0.0978 | 0.1219 | 0.1067 | 0.1031 | 0.1007 | 0.1012 |
| | | | | | SSE | 0.1551 | 0.1042 | 0.0729 | 0.0401 | 0.0321 | 0.0232 |
| | | | $\hat{\gamma}_1$ | AGQL | SM | 0.1929 | 0.1726 | 0.2156 | 0.2309 | | |
| | | | | | SSE | 0.2797 | 0.2306 | 0.1947 | 0.1190 | | |
| | | | | MM | SM | 0.1731 | 0.1768 | 0.2081 | 0.2221 | 0.2426 | 0.2428 |
| | | | | | SSE | 0.3031 | 0.2420 | 0.1966 | 0.1207 | 0.1251 | 0.0881 |
| | | | $\hat{\sigma}_\eta^2$ | AGQL | SM | 0.9338 | 0.9465 | 0.9635 | 0.9926 | | |
| | | | | | SSE | 0.3840 | 0.2529 | 0.2081 | 0.1138 | | |
| | | | | MM | SM | 0.9223 | 0.9364 | 0.9710 | 0.9946 | 0.9865 | 0.9955 |
| | | | | | SSE | 0.3892 | 0.2716 | 0.2058 | 0.1130 | 0.1221 | 0.0737 |
| 0.20 | 0.25 | 0.50 | $\hat{\gamma}_0$ | AGQL | SM | 0.1656 | 0.1881 | 0.2070 | 0.1983 | | |
| | | | | | SSE | 0.1344 | 0.0968 | 0.0860 | 0.0491 | | |
| | | | | MM | SM | 0.1858 | 0.1988 | 0.2061 | 0.2027 | 0.1979 | 0.1997 |
| | | | | | SSE | 0.1562 | 0.1210 | 0.0879 | 0.0511 | 0.0392 | 0.0289 |
| | | | $\hat{\gamma}_1$ | AGQL | SM | 0.2097 | 0.2221 | 0.2053 | 0.2402 | | |
| | | | | | SSE | 0.3135 | 0.2614 | 0.2286 | 0.1478 | | |
| | | | | MM | SM | 0.2016 | 0.2151 | 0.2184 | 0.2318 | 0.2506 | 0.2513 |
| | | | | | SSE | 0.3433 | 0.3210 | 0.2421 | 0.1489 | 0.1348 | 0.0921 |
| | | | $\hat{\sigma}_\eta^2$ | AGQL | SM | 0.5377 | 0.4971 | 0.4907 | 0.4972 | | |
| | | | | | SSE | 0.2735 | 0.1885 | 0.1536 | 0.0905 | | |
| | | | | MM | SM | 0.5092 | 0.4685 | 0.4827 | 0.4953 | 0.4918 | 0.4981 |
| | | | | | SSE | 0.2779 | 0.2138 | 0.1625 | 0.0932 | 0.0791 | 0.0518 |
| | | 1.0 | $\hat{\gamma}_0$ | AGQL | SM | 0.1815 | 0.2108 | 0.2108 | 0.2054 | | |
| | | | | | SSE | 0.1558 | 0.1090 | 0.0802 | 0.0483 | | |
| | | | | MM | SM | 0.2144 | 0.2159 | 0.2072 | 0.2039 | 0.2016 | 0.2037 |
| | | | | | SSE | 0.1882 | 0.1144 | 0.0815 | 0.0476 | 0.0422 | 0.0285 |
| | | | $\hat{\gamma}_1$ | AGQL | SM | 0.1830 | 0.1868 | 0.1908 | 0.2294 | | |
| | | | | | SSE | 0.2695 | 0.2190 | 0.1796 | 0.1253 | | |
| | | | | MM | SM | 0.1635 | 0.1927 | 0.1957 | 0.2354 | 0.2489 | 0.2410 |
| | | | | | SSE | 0.3036 | 0.2322 | 0.1854 | 0.1315 | 0.1189 | 0.0795 |
| | | | $\hat{\sigma}_\eta^2$ | AGQL | SM | 0.9768 | 0.9747 | 0.9882 | 0.9921 | | |
| | | | | | SSE | 0.3823 | 0.2711 | 0.1914 | 0.1273 | | |
| | | | | MM | SM | 0.9455 | 0.9599 | 0.9900 | 0.9873 | 0.9746 | 0.9947 |
| | | | | | SSE | 0.4156 | 0.2741 | 0.2000 | 0.1309 | 0.1061 | 0.0708 |

$$S_{20} = \sum_{t=T_0+1}^{T} y_{t-1}^2 y_t^2/(T - T_0 - 1),$$

$$g_{20}(\gamma_0,\ \gamma_1,\ \sigma_\eta^2) = \exp\left[\frac{2\gamma_0 + \sigma_\eta^2}{1-\gamma_1}\right].$$

Next because

$$\lim_{T_0\to\infty}\left(\frac{\partial g_{20}(\gamma_0, \gamma_1, \sigma_\eta^2)}{\partial \gamma_1}\right) = \left(\frac{2\gamma_0 + \sigma_\eta^2}{(1-\gamma_1)^2}\right) g_{20}(\gamma_0, \gamma_1, \sigma_\eta^2) = \delta_{20} \text{ (say)}, \quad (55)$$

and

$$\lim_{T_0\to\infty} \text{Var}(S_{20}) = \lim_{T_0\to\infty}\left[\frac{1}{(T-T_0-1)^2}\left(\sum_{t=T_0+1}^{T} \text{Var}(Y_{t-1}^2 Y_t^2) + \sum_{t=T_0+1}^{T-1} \text{Cov}(Y_{t-1}^2 Y_t^2, Y_t^2 Y_{t+1}^2)\right)\right]$$
$$= \frac{1}{(T-T_0-1)^2}\left[(T-T_0-1)\xi_{20} + 2(T-T_0-2)\xi_{20}^*\right], \quad (56)$$

with

$$\xi_{20} = g_{20}^2(\gamma_0, \gamma_1, \sigma_\eta^2)\left[9\exp\{\frac{2\sigma_\eta^2}{1-\gamma_1}\} - 1\right]$$
$$\xi_{20}^* = 3\ \exp\left[\frac{4\gamma_0}{1-\gamma_1} + \sigma_\eta^2\left(\frac{3+\gamma_1}{1-\gamma_1}\right)\right] - g_{20}^2(\gamma_0, \gamma_1, \sigma_\eta^2), \quad (57)$$

it then follows that the asymptotic variance of $\widehat{\gamma}_{1MM}$ has the formula

$$\lim_{T_0\to\infty} \text{Var}(\widehat{\gamma}_{1MM}) = \delta_{20}\left[\frac{1}{T-T_0-1}\xi_{20} + \frac{2(T-T_0-2)}{(T-T_0-1)^2}\xi_{20}^*\right]. \quad (58)$$

## 5.2 Asymptotic Variance of the Estimator of $\sigma_\eta^2$

Recall that for given $\gamma_0$ and $\gamma_1$, one may obtain the asymptotically unbiased estimator of $\sigma_\eta^2$ by solving the unbiased estimating equation (23). For $S_{10} = \sum_{t=T_0+1}^{T} y_t^2/(T-T_0)$, (23) implies by (22) that

$$\hat{\sigma}_\eta^2 = \lim_{T_0\to\infty}\left[2\left(ln(S_{10}) - \frac{\gamma_0}{1-\gamma_1}\right)(1-\gamma_1^2)\right] = f(S_{10}) \text{ (say)}. \quad (59)$$

One may then computes the asymptotic variance of the MM estimator of $\sigma_\eta^2$ as

$$\begin{aligned}\lim_{T_0\to\infty} \mathrm{Var}(\hat{\sigma}^2_{\eta,MM}) &= \lim_{T_0\to\infty} E\Big[f(S_{10}) - E[f(S_{10})]\Big]^2 \\ &= \lim_{T_0\to\infty} E\Bigg[\left(\frac{\partial f(S_{10})}{\partial S_{10}}|_{S_{10}=g_{10}}\right)^2 (S_{10} - g_{10}(\gamma_0, \gamma_1, \sigma^2_\eta))^2\Bigg] \\ &= \lim_{T_0\to\infty} \Bigg[\left(\frac{\partial f(S_{10})}{\partial S_{10}}|_{S_{10}=g_{10}}\right)^2 \mathrm{Var}(S_{10})\Bigg], \qquad (60)\end{aligned}$$

where

$$\begin{aligned}\frac{\partial f(S_{10})}{\partial S_{10}}|_{s_{10}=g_{10}} &= 2(1-\gamma_1^2)S_{10}^{-1}|_{s_{10}=g_{10}} \\ \frac{\partial f(S_{10})}{\partial S_{10}} &= 2\,(1-\gamma_1^2)\, g_{10}^{-1}(\gamma_1, \sigma^2_\eta),\end{aligned}$$

and

$$\begin{aligned}\lim_{T_0\to\infty} \mathrm{Var}(S_{10}) &= \lim_{T_0\to\infty} \Bigg[\frac{1}{(T-T_0)^2}\Bigg(\sum_{t=T_0+1}^{T} \mathrm{Var}(Y_t^2) + \sum_{s\neq t,s,t=T_0+1}^{T} \mathrm{Cov}(Y_s^2, Y_t^2)\Bigg)\Bigg] \\ &= \frac{1}{(T-T_0)^2}\Bigg[(T-T_0)\xi_{10} + \sum_{s\neq t,s,t=T_0+1}^{T} \xi^*_{10}(s,t)\Bigg]\end{aligned}$$

with

$$\begin{aligned}\xi_{10} &= g^2_{10}(\gamma_0, \gamma_1, \sigma^2_\eta)\left[3\exp\{\frac{2\sigma^2_\eta}{1-\gamma_1}\} - 1\right] \\ \xi^*_{10}(s,t) &= \exp\left((1+\gamma_1^{|t-s|})\frac{\sigma^2_\eta}{1-\gamma_1^2}\right) - g^2_{10}(\gamma_0, \gamma_1, \sigma^2_\eta). \qquad (61)\end{aligned}$$

To illustrate the computations of the asymptotic variances of $\hat{\gamma}_{1,MM}$ and $\hat{\sigma}^2_{\eta,MM}$, we have considered, for example, $T_0 = 500$ when $T = 10{,}000$, and computed these variances for $\gamma_0 = 0,\ \gamma_1 = 0.50,$ and $\sigma^2_\eta = 0.25$, by (61) and (63), respectively. The variances were found to be 0.0136 and 0.0029, which are very close to the corresponding simulated variances $SSE^2 = 0.1208^2$ and $SSE^2 = 0.0586^2$, respectively, given in Table 1 for $\gamma_0 = 0.05,\ \gamma_1 = 0.5,$ and $\sigma^2_\eta = 0.25$, with $T = 10{,}000$.

## 6 Understanding Volatility Through Kurtosis of the Data

To understand the volatility, that is, to understand the changes in variance pattern in the time series, it is recommended to examine the kurtosis of the data over time. See, for example, Jacquier et al. (1994, p. 387), Shephard (1996, p. 23), Mills (1999, p. 129) and Tsay (2005, p. 134). First, computing $E[Y_t^2]$ by (38) and $E[Y_t^4]$ by (43), and then using the formula

$$\kappa_t(\gamma_0,\ \gamma_1,\ \sigma_\eta^2) = \frac{E(Y_t^4)}{[E(Y_t^2)]^2},$$

from (7), one obtains the kurtosis of the $y_t$ variable under the volatility model (1)–(2), as

$$\kappa_t(\gamma_1, \sigma_\eta^2) = \begin{cases} 3 & \text{for } t = 1 \\ 3\exp\left(\sigma_\eta^2 \sum_{r=0}^{t-2} \gamma_1^{2r}\right) & \text{for } t = 2, \ldots, T \end{cases} \tag{62}$$

Note that the kurtosis does not depend on $\gamma_0$ and $\sigma_1^2$. Also, in the limiting case, i.e, when $t \to \infty$, the kurtosis in (62) reduces to

$$\lim_{t\to\infty} \kappa_t(\gamma_1, \sigma_\eta^2) = 3\ \exp\left\{\frac{\sigma_\eta^2}{1-\gamma_1^2}\right\}, \tag{63}$$

which agrees with the formula for kurtosis studied by Harvey et al. (1994, p. 249), Mills (1999, p. 249) and Broto and Ruiz (2004, p. 615), among others. Further note that, the formula for the kurtosis given in (63) is independent of time, whereas kurtosis at a finite time point given by (62) is dependent on first few times. Now to understand the effects of the parameters $\gamma_1$ and $\sigma_\eta^2$ on the kurtosis, we, for example, display the true kurtosis computed by (62) in Figs. 1 and 2 for selected values of the parameters. In the same figures we also display the estimated kurtosis computed by using $\hat{\gamma}_{1,MM}$ and $\hat{\sigma}^2_{\eta,MM}$ for $\gamma_1$ and $\sigma_\eta^2$ respectively. For the estimation of these two parameters, without any loss of generality, we have chosen $\gamma_0 = 0$ and $\sigma_1^2 = 0.5$.

It is clear from Figs. 1 and 2 that the kurtosis under the present volatility model (1)–(2) is much larger than the Gaussian based kurtosis ($= 3$). These figures also exhibit that the kurtosis gets stabilized quickly after an initial short period. To be specific, Fig. 1 shows that when $\gamma_1 = 0.5$ and $\sigma_\eta^2 = 0.5$, the kurtosis gets stabilized at $\kappa_t = 5.8432$ for any t $> 4$. Similarly, Fig. 2 shows that when $\gamma_1 = 0.5$ and $\sigma_\eta^2 = 1.0$, the kurtosis gets stabilized at $\kappa_t = 11.3810$ for any t $> 4$.

Note that in both Figs. 1 and 2, the estimated kurtosis appears to be very close to the corresponding true value of kurtosis, indicating that the proposed MM technique performs very well in estimating the parameters of the volatility model.

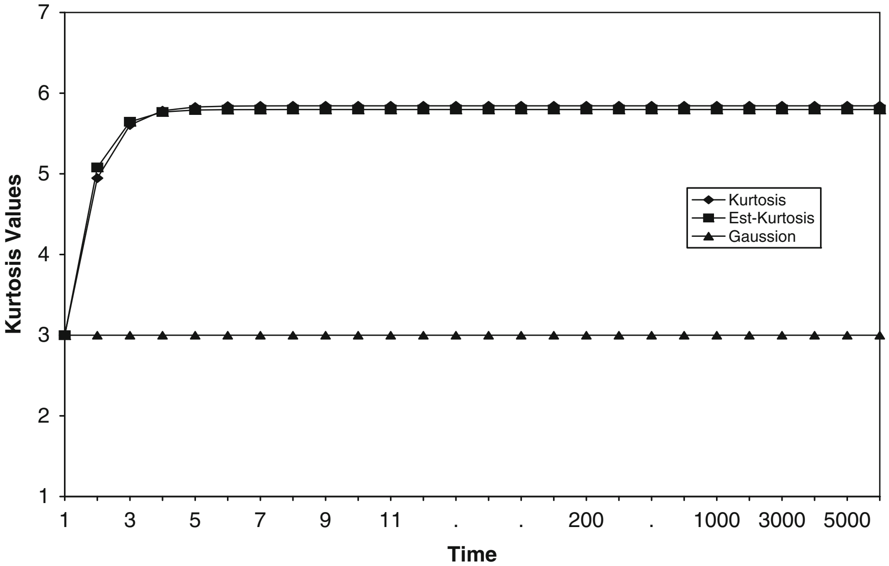



**Fig. 1** True and estimated kurtosis with volatility parameters $\gamma_1 = 0.5$, $\sigma_\eta^2 = 0.5$

## 7 Volatility in US-Dollar and Swiss-Franc Exchange Rate: A Numerical Illustration

Ruiz (1994) applied the QML estimation method to fit a stochastic volatility (SV) model to the log return of Swiss-Franc and US-Dollar exchange rate from 1/10/81 to 28/6/85. By using the QML estimation method, Harvey et al. (1994) illustrated the fitting of a multivariate stochastic volatility model to the log return of Pound and Dollar, Deutschmark and Dollar, Yen and Dollar and Swiss-Franc and Dollar exchange rates from 1/10/81 to 28/6/85. It was however demonstrated in Sects. 2 and 4 that the proposed MM approach performs better than the QML and GMM estimation approaches. It was also shown in Sect. 4 that the AGQL (approximate GQL) is a highly competitive approach to the MM approach. Note that the AGQL approach may also be referred to as the weighted GQL (WGQL) approach. In this section, we revisit the Swiss-Franc and US-Dollar exchange rate data and apply the MM, QML, and WGQL estimation approaches in order to fit the SV model to the log return of the exchange rates. To be specific, we consider a recent series of $T = 1000$ observations on daily log returns of Swiss-Franc and US-Dollar from July 24, 2007 to July 24, 2012. Let $P_t$ denote the exchange rate for a day indexed by time $t$. We then consider the response $y_t$ as the log return of the exchange rates with mean subtracted, that is

$$y_t = \log P_{t+1} - \log P_t - \left(\sum \triangle \log P_t\right)/T, \qquad t = 1, \cdots, T,$$

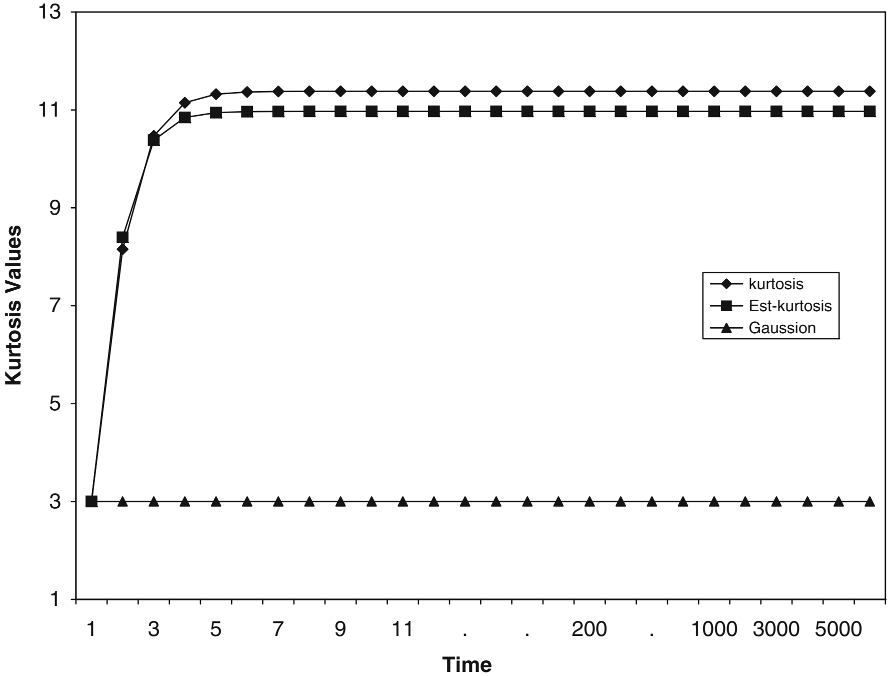



**Fig. 2** True and estimated kurtosis with volatility parameters $\gamma_1 = 0.5$, $\sigma_\eta^2 = 1.0$

where $\triangle \log P_t = \log P_{t+1} - \log P_t$. To understand the sample data, we computed the sample means and variances for every group of 50 consecutive $y_t^2$ values. Thus there are 20 groups for 1000 observations. These 20 means and variances are displayed in Figs. 3 and 4, respectively. There appears to be an example of volatile activities in log returns around at $t = 400$. It is of interest to see the relative performance of the aforementioned three methods in fitting this volatile series.

The parameters of the SV model were estimated by the MM, WGQL and QML approaches presented in Sects. 3.1, 3.3, and 3.4, respectively. These estimates are given in Table 3, which are then used in the formulas for $\mathrm{E}[Y_t^2]$ and $\mathrm{Var}[Y_t^2]$ given by (38) [see also (20)] and (43), respectively. Note that the initial variance parameter $\sigma_1^2$ was estimated using 100 daily observations recorded just before July 24, 2007. The estimates for $\mathrm{E}[Y_t^2]$ under three methods are displayed in Fig. 3, and similarly the estimates for $\mathrm{Var}[Y_t^2]$ are displayed in Fig. 4.

It is observed from Table 3 that the MM and WGQL (AGQL) estimates for $\gamma_0$, $\gamma_1$ and $\sigma_\eta^2$ are quite close to each other, yielding almost the same curves for the means as shown in Fig. 3 and for the variances as shown in Fig. 4. Note that these estimation results agree with the simulation results discussed in Sect. 4 that the MM and WGQL methods produce almost the same estimates for all these three volatility parameters. In contrast, the QML estimation results are quite different from those by MM and WGQL approaches. Figures 3 and 4 indicate that the MM and WGQL

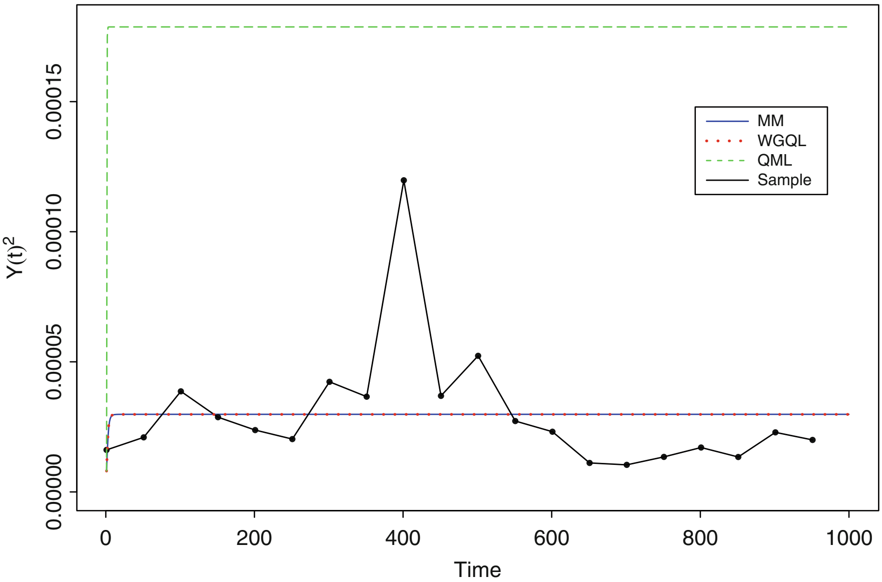



**Fig. 3** The estimated mean of $Y_t^2$ for the 1000 data from the log return with mean subtracted of the US-Dollar/Swiss-Franc daily exchange rates from July 24, 2007 to July 24, 2012 (color online)

**Table 3** The estimated parameter values for fitting the stochastic volatility model to the 1000 observations of the log return of the daily US-Dollar and Swiss-Franc exchange rates from July 24, 2007 to July 24, 2012

| Method | $\hat{\gamma}_0$ | $\hat{\gamma}_1$ | $\hat{\sigma}_\eta^2$ |
|---|---|---|---|
| MM | −3.585664 | 0.6897393 | 1.191857 |
| WGQL | −3.880882 | 0.6641959 | 1.269635 |
| QML | −10.87192 | 0.05930448 | 5.834532 |

estimates are considerably better than the QML estimates. This is because the MM and WGQL estimates for means and variances are much closer to the sample means and variances, as compared to that of the QML approach.

Further note that in Fig. 3, the sample means of $y_t^2$ show a positive correlation, that is, the neighboring sample means tend to trace each other. It can be shown that

$$\text{Cov}\left(Y_t^2, Y_{t+k}^2\right) = \lambda_t \lambda_{t+k} \left\{ \exp\left[ 2\gamma_1^k \left( \frac{1-\gamma_1^{2(t-1)}}{1-\gamma_1^2} \right) \right] - 1 \right\} \tag{64}$$

Since $|\gamma_1| < 1$, it is obvious from (64) that for $0 < \gamma_1 < 1$, $\text{Cov}\left(Y_t^2, Y_{t+k}^2\right) > 0$ and $\to 0$ as $\gamma_1 \to 0$ for all positive integer $k$. The MM and WGQL approaches produce $\gamma_1$ estimates of $\hat{\gamma}_{1,MM} = 0.690$ and $\hat{\gamma}_{1,WGQL} = 0.664$, which are all positive

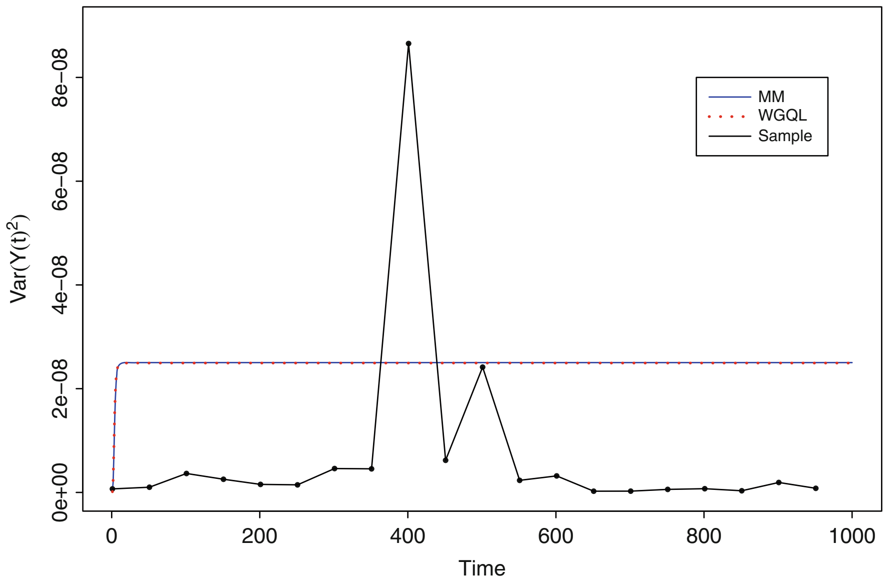



**Fig. 4** The estimated variance of $Y_t^2$ for the 1000 data from the log return with mean subtracted of the US-Dollar/Swiss-Franc daily exchange rates from July 24, 2007 to July 24, 2012 (color online)

and can account for the positive correlation in Fig. 3, while the QML estimate of $\hat{\gamma}_{1,QML} = 0.0593$ can be too small to explain this positive correlation. In addition, in Fig. 3, the values for sample means oscillate around the curves for the estimated mean by the MM and WGQL approaches, indicating a reasonable fitting, while the curve for the estimated mean by the QML method is far above the points of sample means, implying some inaccuracy in the QML approach, which may be due to the large standard errors of the QML estimators. In Fig. 4, once again the curves for the MM and WGQL approaches almost overlap each other, indicating the similarity of the estimates by the two approaches. Except the first several time points, the estimated variance of $y_t^2$ by the QML approach is on the order of $10^{-5}$, which is considerably larger than the sample variances and the estimated values produced by the MM and WGQL approaches, so the curve for QML estimation is not shown in Fig. 4.

## 8 Concluding Remarks

To fit the volatility model, the existing GMM approach uses a large number of moments (9) to construct the GMM estimating equations (11) for the consistent estimation of the volatility parameters, whereas the QML approach uses a normal approximation to a log chi-square distribution that arises in the construction of the so-called likelihood estimating equations. In this paper, it is demonstrated that unlike the GMM approach, the moment estimating equations for three volatility parameters can be constructed by using only three unbiased moment functions selected carefully following the nature or definition of the parameters. The later approach is simply refereed to as the MM (method of moments) approach. As the GMM approach is complex, it is not included for any comparison in the present paper. We also provided a modification to the QML approach so that the associated covariance matrix involved in the QML estimating equation can be computed easily. The finite sample behavior of the proposed MM estimation approach is studied intensively, and it is found that the MM approach works very well in estimating all volatility parameters for time series size as small as 1000. The drawbacks of the QML approach is discussed. The AGQL (WGQL) approach performs similarly to the MM approach, however, it is computationally more involved than the MM approach. The asymptotic variances of the proposed simpler MM, and WGQL estimators are computed and these variances agree quite well with the large finite sample based results. Remark that as the QML estimates are not obtained using the true likelihood of the data, it is not possible to compute the asymptotic variances of the QML estimators. All three methods, MM. WGQL, and QML were applied to fit the SV model (1)–(2) to a real life financial time series with length $T = 1000$, and it was found that the MM and WGQL approaches provide relatively much better fitting than the QML approach.

**Acknowledgements** The authors thank the audience of the symposium and a referee for their comments and suggestions.

## References

Abramowitz, M., Stegun, N.C.: Handbook of Mathematical Functions. Dover Publications, New York (1970)
Ahn, S., Schmidt, P.: Efficient estimation of models for dynamic panel data. J. Econ. **68**, 5–28 (1995)
Andersen, T.G., Sorensen, B.E.: GMM estimation of a stochastic volatility model: a Monte Carlo study. J. Bus. Econ. Stat. **14**, 328–352 (1996)
Andersen, T.G., Sorensen, B.E.: GMM and QML asymptotic standard deviations in stochastic volatility models: comments on Ruiz (1994). J. Econ. **76**, 397–403 (1997)
Arellano, M., Bond, S.: Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. Rev. Econ. Stat. **58**, 277–298 (1991)
Broto, C., Ruiz, E.: Estimation methods for stochastic volatility models: a survey. J. Econ. Surv. **18**, 613–649 (2004)
Chamberlain, G.: Sequential moment restrictions in panel data-comment. J. Bus. Econ. Stat. **10**, 20–26 (1992)
Danielsson, J.: Stochastic volatility in asset prices: estimation with simulated maximum likelihood. J. Econ. **64**, 375–400 (1994)
Durbin, J., Koopman, S.J.: Monte Carlo maximum likelihood estimation for non-Gaussian state space models. Biometrika **84**, 669–684 (1997)

Hansen, L.P.: Large sample properties of generalized method of moments estimators. Econometrica **50**, 1029–1054 (1982)
Harvey, A.C., Ruiz, E., Shephard, N.: Multivariate stochastic variance models. Rev. Econ. **61**, 247–264 (1994)
Jacquier, E., Polson, N.G., Rossi, P.E.: Bayesian analysis of stochastic volatility models (with discussion). J. Bus. Econ. Stat. **12**, 371–417 (1994)
Keane, M., Runkle, D.: On the estimation of panel data models with serial correlation when instruments are not strictly exogenous. J. Bus. Econ. Stat. **10**, 1–9 (1992)
Koopman, S.J., Harvey, A.C., Doornik, J.A., Shephard, N.: Stamp 5.0: Structural Time Series Analyser, Modeller, and Predictor. Chapman and Hall, London (1995)
Lee, K.M., Koopman, S.J.: Estimating stochastic volatility models: a comparison of two importance samplers. Stud. Nonlin. Dyn. Econ. **8**, 1–15 (2004)
Liesenfeld, R., Richard, J.F.: Univariate and multivariate stochastic volatility models: estimation and diagnostics. J. Empir. Finan. **207**, 1–27 (2003)
Melino, A., Turnbull, S.M.: Pricing foreign currency options with stochastic volatility. J. Econ. **45**, 239–265 (1990)
Mills, T.C.: The Econometric Modelling of Finance Time Series. Cambridge University Press, Cambridge (1999)
Rao, R.P., Sutradhar, B.C., Pandit, V.N.: GMM versus GQL inferences in semiparametric linear dynamic mixed models. Braz. J. Probab. Stat. **26**, 167–177 (2012)
Ruiz, E.: Quasi-maximum likelihood estimation of stochastic volatility models. J. Econ. **63**, 289–306 (1994)
Ruiz, E.: QML and GMM estimators of stochastic volatility models: response to Anderson and Sorensen. J. Econ. **76**, 405 (1997)
Shephard, N.: Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen,O.E. (eds.) Time Series Models in Econometrics, Finance and Other Fields, pp. 1–67. Chapman & Hall, London (1996)
Shephard, N., Pitt, M.K.: Likelihood analysis of non-Gaussian parameter-driven models. Biometrica **84**, 653–667 (1997)
Sutradhar, B.C.: On exact quasilikelihood inference in generalized linear mixed models. Sankhya B: Indian J. Stat. **66**, 261–289 (2004)
Taylor, S.J.: Financial returns modelled by the product of two stochastic processes - a study of daily sugar prices, 1961–79. In: Anderson, O.D. (ed.) Time Series Analysis: Theory and Practice, vol. 1. North-Holland Publishing Company, Amsterdam (1982)
Tsay, R.S.: Analysis of Financial Time Series. Wiley Interscience, New York (2005)
Usmani, R.: Inversion of a tridiagonal Jacobi matrix. Linear Algebra Appl. **212/213**, 413–414 (1994)

# Part III
# Longitudinal Multinomial Models in Parametric and Semi-parametric Setups

# A Generalization of the Familial Longitudinal Binary Model to the Multinomial Setup

**Brajendra C. Sutradhar, Roman Viveros-Aguilera, and Taslim S. Mallick**

**Abstract** When repeated binary responses along with time dependent covariates are collected over a short period of time from the members of a large number of independent families, there exits a well developed binary dynamic mixed logit (BDML) model to analyze such familial longitudinal binary data. As far as the inferences are concerned, this BDML model has been fitted by using the generalized quasi-likelihood (GQL) and the well known maximum likelihood (ML) methods. There are however situations in practice where categorical/multinomial responses with more than two categories are repeatedly collected from all members of the family. However, the analysis for this type of familial longitudinal multinomial data is not adequately addressed in the literature. We offer two main contributions in this paper. First, for the analysis of familial longitudinal multinomial data, we propose a multinomial dynamic mixed logit (MDML) model as a generalization of the BDML model and derive the basic properties such as non-stationary mean, variance and correlations for the repeated multinomial responses. Next, to understand these basic properties, we develop step by step likelihood estimating equations for the parameters involved in these properties. The relative asymptotic efficiency performance of the ML and GQL approaches is examined through a simulation study based on repeated binary responses, for example, from a large number of independent families each consisting of two members, causing both familial and longitudinal correlations. Also, a real life example on repeated multinomial data analysis is considered as an illustration.

---

B.C. Sutradhar (✉)
Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada A1C5S7
e-mail: bsutradh@mun.ca

R. Viveros-Aguilera
Department of Mathematics and Statistics, McMaster University,
Hamilton, ON, Canada L8S 4K1
e-mail: rviveros@math.mcmaster.ca

T.S. Mallick
Department of Statistics, Biostatistics and Informatics, University of Dhaka, Dhaka, Bangladesh
e-mail: taslim@statdu.ac.bd



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_5

**Keywords** Categorial responses • Dynamic mixed models • Familial correlations through random effects • Longitudinal correlations through dynamic relationships • Multinomial logit model

## 1 Introduction

In practice, there are situations where repeated multinomial responses along with time dependent covariates may be collected from the members of a large number of independent families. For example, we first refer to the health care utilization data discussed by Sutradhar (2011, Sect. 10.3), where number of yearly visits to the physician were collected for 6 years from the members of 48 families, along with the covariate information from each member. This familial longitudinal count data were analyzed by using the so-called GLLMM (generalized linear longitudinal mixed model). Now suppose that one is not interested to the number of specific visits, instead, the data in category form such as low, medium and high visits, are of interest. This will lead to a familial longitudinal multinomial data set. Note that even though this type of multinomial data can be treated as a generalization of the familial longitudinal binary data and there exists inferences for such binary data using BDML (binary dynamic mixed model) [see, for example, Sutradhar (2011, Chap. 11)], the analysis of familial longitudinal multinomial data is, however, not adequately addressed in the literature. The purpose of this paper is to develop a suitable model for such repeated multinomial data and provide inferences for the proposed model. However, for convenience, we begin the discussion with a review of the familial longitudinal binary data analysis.

Let $y_{ijt}$ denote the binary response collected at $t$-th ($t = 1, \ldots, T$) time point from the $j$-th ($j = 1, \ldots, n_i$) member of the $i$-th ($i = 1, \ldots, K$) family and $x_{ijt} = (x_{ijt1}, \ldots, x_{ijtu}, \ldots, x_{ijtp})'$ be the $p \times 1$ vector of covariates corresponding to $y_{ijt}$. In this set up, $T$ and $n_i$ are considered to be small, whereas $K$, the number of independent families, is considered to be quite large. Suppose that irrespective of the time points, the responses of all members of the $i$th family are influenced by a common unobservable family effect $\tau_i^*$. As far as the time effects on the responses are concerned, it is sensible to assume that conditional on the common random family effect $\tau_i^*$, two responses, say $y_{ijt}$ and $y_{ijv}$, for the $j$th individual of the $i$th family collected at any two time points $t$ and $v$ ($t \neq v; t, v = 1, \ldots, T$) will be longitudinally correlated, whereas the responses of any two members, say $y_{ijt}$ and $y_{ikv}$ for the $j$th and $k$th members ($j \neq k$) of the $i$th family will be independent. This type of familial longitudinal data has recently been analyzed by Sutradhar (2011, Sect. 11.3.1) using a non-linear binary dynamic mixed logit (BDML) model given by

$$Pr(Y_{ij1} = 1|x_{ij1}; \tau_i^*) = \frac{\exp(x'_{ij1}\beta + \tau_i^*)}{[1 + \exp(x'_{ij1}\beta + \tau_i^*)]}, \quad (1)$$

$$Pr(Y_{ijt} = 1|x_{ijt}, y_{ij,t-1}; \tau_i^*) = \frac{\exp\{x'_{ijt}\beta + \gamma^* y_{ij,t-1} + \tau_i^*\}}{[1 + \exp\{x'_{ijt}\beta + \gamma^* y_{ij,t-1} + \tau_i^*\}]}$$
$$= p^*_{ijty_{ij,t-1}}(\tau_i^*), \quad (2)$$

[see also Fienberg et al. (1985)] for $t = 2, \ldots, T$. In (1)–(2), $\beta$ is the $p$-dimensional vector of regression effects, $\gamma^*$ is the lag 1 dynamic dependence parameter, and $\tau_i^*$ is the unobservable random effect for the $i$th family. As far as the distribution of $\tau_i^*$ is concerned, we assume that $\tau_i^* \overset{iid}{\sim} N(0, \sigma_\tau^2)$ (Breslow and Clayton 1993; Sutradhar et al. 2008) so that for $\tau_i = \tau_i^*/\sigma_\tau$, $\tau_i \overset{iid}{\sim} N(0, 1)$. Note that the familial-longitudinal model defined by (1)–(2) may be treated as a generalization of the binary longitudinal mixed model for an individual considered by Sutradhar et al. (2008). Further note that the binary dynamic mixed model defined by (1) and (2) appears to be quite suitable to interpret the data for many health problems. For example, this model produces the mean (also the variance) at a given time point for an individual member of a given family as a function of the covariate history of the individual up to the present time. This history based mean function appears to be useful to interpret the current asthma status (yes or no) of an individual as a function of the related covariates such as smoking habits and cleanliness over a suitable past period.

As far as the inference for the parameters $\beta, \gamma^*,$ and $\sigma_\tau$ involved in the dynamic mixed model (1)–(2) is concerned, they may be estimated, for example, by exploiting the well known likelihood approach. To be specific, because the responses of any two members of the same family at any two time points, say $y_{iju}$ and $y_{ikt}$, conditional on the family effect $\tau_i$, are assumed to be independent, by using (1)–(2), one may write the likelihood function for $\beta,\ \gamma^*,$ and $\sigma_\tau$ as

$$L(\beta, \gamma^*, \sigma_\tau) = \Pi_{i=1}^K \int_{-\infty}^{\infty} \Pi_{j=1}^{n_i} \left[\tilde{p}^*_{ij10}(\tau_i) \Pi_{t=2}^T \tilde{p}^*_{ijty_{ij,t-1}}(\tau_i)\right] f_N(\tau_i) d\tau_i, \quad (3)$$

where $f_N(\tau_i)$ is the standard normal density of $\tau_i$. Next, by using $y_{ij0} = 0$ for all $j = 1, \ldots, n_i$, and $i = 1, \ldots, K$, for technical convenience, and following the likelihood function (3) one may write the log-likelihood function as

$$\log L(\beta, \gamma^*, \sigma_\tau) = \sum_{i=1}^K \sum_{j=1}^{n_i} \sum_{t=1}^T \left[y_{ijt} x'_{ijt}\beta + \gamma^* y_{ijt} y_{ij,t-1}\right]$$
$$+ \sum_{i=1}^K \log \int_{-\infty}^{\infty} exp(d_i s_i) \Delta_i \phi(\tau_i) d\tau_i, \quad (4)$$

where $s_i = \sum_{j=1}^{n_i} \sum_{t=1}^T y_{ijt},\ d_i = \sigma_\tau \tau_i$, and

$$\Delta_i = \left[\Pi_{j=1}^{n_i}\Pi_{t=1}^{T}\{1 + exp(x'_{ijt}\beta + \gamma^* y_{ij,t-1} + \sigma_\tau \tau_i)\}\right]^{-1}.$$

This log likelihood function (4) may then be exploited to obtain the likelihood estimates for all three parameters $\beta, \gamma^*,$ and $\sigma_\tau$.

There are however situations in practice where a multinomial response is recorded from the $j$th member of the $i$th family at a given time $t$. We denote such a multinomial response variable with $C > 2$ categories as $y_{ijt} = (y_{ijt1}, \ldots, y_{ijtc}, \ldots, y_{ijt,C-1})'$, where for $c = 1, \ldots, C-1$,

$$y_{ijt}^{(c)} = (y_{ijt1}^{(c)}, \ldots, y_{ijtc}^{(c)}, \ldots, y_{ijt,C-1}^{(c)})' = (01'_{c-1}, 1, 01'_{C-1-c})' \equiv \delta_{ijtc} \quad (5)$$

indicates that the multinomial response of $j$th individual of the $i$th family belongs to $c$th category at time t. For $c = C$, one writes $y_{ijt}^{(C)} = \delta_{ijtC} = 01_{C-1}$. Here $1_m$, for example, denotes the $m$-dimensional unit vector. Similar to the binary case, further suppose that $x_{ijt}$ be the $p$-dimensional covariate vector recorded at time $t$ corresponding to $y_{ijt}$. It is of interest to find the effects of $x_{ijt}$ on $y_{ijt}$, when it is known that $C-1$ dimensional multinomial responses $y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT}$ will be correlated. The modeling and analysis of this type of familial longitudinal multinomial data are, however, not adequately addressed in the literature.

For the case when $n_i = 1$, i.e., the $i$th family refers to the $i$th ($i = 1, \ldots, K$) individual only, there exist some models for the analysis of repeated multinomial data collected from $K$ independent individuals. For example, Fienberg et al. (1985) and Conaway (1989) have modeled repeated multinomial such as stress level (Low, Medium, High) data for 4 years 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. The study focuses on the changes in the post accident stress level of mothers of young children living within 10 miles of the nuclear plant. Note however that Fienberg et al. (1985) have collapsed the trinomial (3 category) data into 2 category based dichotomized data and modeled the correlations among such repeated binary responses through a binary dynamic logit model [see also Sutradhar and Farrell (2007)]. Thus their model does not deal with correlations of repeated multinomial (trinomial in this case to be specific) responses and do not make the desired comparison among three original stress levels. To remedy this modeling problem, Conaway (1989) has attempted to model the multinomial correlations but has used a random effect, that is, mixed model approach to compute the correlations. Note however that the random effects based correlations are not able to address the dynamic dependence among repeated responses.

Some authors such as Lipsitz et al. (1994), Williamson et al. (1995), and Chen et al. (2009) have used the so-called multinomial dynamic fixed logit (MDFL) model to study the aforementioned longitudinal categorical data. To be specific, their model for repeated univariate multinomial data has the form

$$P[y_{i1} = y_{i1}^{(c)} = \delta_{ij1c}] = \pi_{(i1)c}$$

$$= \begin{cases} \frac{\exp(x'_{i1}\beta^*_c)}{1+\sum_{g=1}^{C-1}\exp(x'_{ij1}\beta^*_g)} \text{ for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1}\exp(x'_{ij1}\beta^*_g)} \text{ for } c = C, \end{cases} \tag{6}$$

where

$$\beta^*_c = (\beta_{c0}, \beta'_c)' = (\beta_{c0}, \beta_{c1}, \ldots, \beta_{c,p-1})'$$

is the effect of $x_{ij1}$, leading to the regression parameters set as

$$\beta \equiv (\beta^{*\prime}_1, \ldots, \beta^{*\prime}_c, \ldots, \beta^{*}_{C-1}{}')' : (C-1)p \times 1.$$

Further for $t = 2, \ldots, T$, the non-linear conditional multinomial dynamic fixed logit (MDFL) probabilities are given by

$$\eta^{(c)}_{it|t-1}(g) = P\Big(Y_{it} = y^{(c)}_{it} \,\Big|\, Y_{i,t-1} = y^{(g)}_{i,t-1},\Big)$$
$$= \begin{cases} \frac{\exp\left[x'_{it}\beta^*_c + \gamma'_c y^{(g)}_{i,t-1}\right]}{1+\sum_{v=1}^{C-1}\exp\left[x'_{it}\beta^*_v + \gamma'_v y^{(g)}_{i,t-1}\right]}, \text{ for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{v=1}^{C-1}\exp\left[x'_{it}\beta^*_v + \gamma'_v y^{(g)}_{i,t-1}\right]}, \text{ for } c = C, \end{cases} \tag{7}$$

where $g = 1, \ldots, C$, and $\gamma_c = (\gamma_{c1}, \ldots, \gamma_{cv}, \ldots, \gamma_{c,C-1})'$ denotes the dynamic dependence parameters, which may be referred to as the correlation index or a particular type of odds ratio parameters. More specifically, the correlations of the repeated multinomial responses will be functions of these $\gamma_c (c = 1, \ldots, C-1)$ parameters. As far as the estimation of the model parameters is concerned, these authors first use an arbitrary log-linear model for odds ratios and then use the estimated odds ratios for the estimation of the regression parameters. This type of 'working' odds ratios or correlations approach may, however, fail to produce consistent estimates for true odds ratios and such inconsistent estimates may lead to breakdown for the estimation of the regression parameters. See for example, Crowder (1995) and Sutradhar and Das (1999) for a discussion with regard to such breakdown difficulties in the context of longitudinal correlation models for repeated binary data. For an alternative exact likelihood approach for the estimation of the parameters of the MDFL model (6)–(7), we, for example, refer to Sutradhar (2014, Chap. 3).

In this paper, we, however, plan to deal with an extension of the longitudinal model (6)–(7) to the familial longitudinal setup involving familial longitudinal categorical data indicated by (5). To be specific, in Sect. 2, we provide a multinomial dynamic mixed logit (MDML) model to fit familial longitudinal categorical data. Basic properties of the proposed multinomial model are derived in the same section. With regard to estimation of the parameters involved in the model and hence in basic properties, because the inferences for such complex high dimensional data model are naturally complex, in Sect. 3, we provide all necessary likelihood estimating equations in simpler possible forms, for the benefit of the practitioners. In Sect. 4, we examine the relative asymptotic efficiency performance of the ML and GQL approaches through a simulation study based on repeated binary responses, for example, from a large number of independent families each consisting of two ($n_i = 2$) members. In Sect. 5, we provide an example by illustrating the application of the proposed model to the analysis of the so-called Three Miles Island Stress Level (TMISL) data containing repeated multinomial responses with three categories.

## 2 Proposed Familial Longitudinal Multinomial Model

Refer to the multinomial variable $y_{ijt}$ with $C \geq 2$ categories defined by (5) for the $j$th individual of the $i$th family, recorded at time point $t$. To be specific, $y_{ijt}$ is the $C-1$-dimensional multinomial response variable. The repeated multinomial responses of an individual member are correlated. Also, the two multinomial responses at a given time collected from two members of the same family are correlated. In this section, we propose a multinomial dynamic mixed logit (MDML) model that exhibits both of these familial and longitudinal correlations. More specifically, we generalize the BDML model (1)–(2) to the multinomial setup, and write the desired MDML model as follows. First, for $t = 1, \ldots, T$, let

$$\pi^*_{(ijt)c} = \begin{cases} \frac{\exp(x'_{ijt}\beta^*_c + \sigma_\tau \tau_i)}{1+\sum_{g=1}^{C-1} \exp(x'_{ijt}\beta^*_g + \sigma_\tau \tau_i)} & \text{for } c = 1, \ldots, C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1} \exp(x'_{ijt}\beta^*_g + \sigma_\tau \tau_i)} & \text{for } c = C, \end{cases} \tag{8}$$

where $\beta^*_c = (\beta_{c0}, \beta'_c)' = (\beta_{c0}, \beta_{c1}, \ldots, \beta_{c,p-1})'$. Now use this form (8) for $t = 1$ only, and define the marginal probability for the multinomial variable $y_{ij1}$ as

$$P[\{y_{ij1} = y^{(c)}_{ij1} = \delta_{ij1c}\}|\tau_i] = \pi^*_{(ij1)c}. \tag{9}$$

Next, for $t = 2, \ldots, T$, as a generalization of the binary transitional probability (2), write the multinomial transitional probability as

$$\begin{aligned} \eta^{*(c)}_{ijt|t-1}(g) &= P\Big(Y_{ijt} = y^{(c)}_{ijt} \;\Big|\; Y_{ij,t-1} = y^{(g)}_{ij,t-1}, \tau_i\Big) \\ &= \begin{cases} \frac{\exp\left[x'_{ijt}\beta^*_c + \gamma'_c y^{(g)}_{ij,t-1} + \sigma_\tau \tau_i\right]}{1+\sum_{v=1}^{C-1} \exp\left[x'_{ijt}\beta^*_v + \gamma'_v y^{(g)}_{ij,t-1} + \sigma_\tau \tau_i\right]}, & \text{for } c = 1, \ldots, C-1 \\ \frac{1}{1+\sum_{v=1}^{C-1} \exp\left[x'_{ijt}\beta^*_v + \gamma'_v y^{(g)}_{ij,t-1} + \sigma_\tau \tau_i\right]}, & \text{for } c = C, \end{cases} \end{aligned} \tag{10}$$

where for known response category $g = 1,\ldots,C$, at time $t-1$, $\gamma_c = (\gamma_{c1},\ldots,\gamma_{cv},\ldots,\gamma_{c,C-1})'$, for $c = 1,\ldots,C-1$, denotes the dynamic dependence parameters for the response to be in the $c$th category at time $t$. Note that the marginal and transitional probabilities in (9)–(10) combining together may be referred to as the multinomial dynamic mixed logit (MDML) model. This MDML model (9)–(10) may also be treated as a generalization of the MDFL (multinomial dynamic fixed logit) model (6)–(7) appropriate for the longitudinal setup to the familial longitudinal setup. It is clear that unlike the fixed model (6)–(7), the mixed model (9)–(10) is written for all members of the same family by accommodating the familial correlations through the introduction of the common random family effect $\tau_i^* = \sigma_\tau \tau_i$. This additional random effect, however, makes the inference for the regression effects ($\beta$) and dynamic dependence parameters $\gamma_c (c = 1,\ldots,C-1)$ extremely complicated. Furthermore, when in practice, one is interested to understand the data through computing the mean, variance and correlations, it becomes necessary to estimate the random effects variance parameter $\sigma_\tau$ as well.

Now toward the computation of the mean, variance and correlations of the repeated multinomial responses $y_{ijt}$, $t = 1,\ldots,T$, under the MDML model (9)–(10), we first provide their moments conditional on $\tau_i$, as follows. Specifically, it follows from (9) and (10) that

$$E[Y_{ijt}|\tau_i] = \tilde{\pi}_{(ijt)}(\tau_i) \tag{11}$$
$$= \begin{cases} \pi^*_{(ij1)}(\tau_i) & \text{for } t = 1 \\ \pi^*_{(ijt)}(\tau_i) + \left[\eta^*_{ijt|t-1,M} - \pi^*_{(ijt)}(\tau_i)1'_{C-1}\right]\tilde{\pi}_{(ij,t-1)}(\tau_i) & \text{for } t = 2,\ldots,T, \end{cases}$$

where $\pi^*_{(ijt)}(\tau_i)$ and $\tilde{\pi}_{(ijt)}(\tau_i)$ are $(C-1)$-dimensional vectors written as

$$\pi^*_{(ijt)}(\tau_i) = [\pi^*_{(ijt)1},\ldots,\pi^*_{(ijt)c},\ldots,\pi^*_{(ijt)(C-1)}]'$$
$$\tilde{\pi}_{(ijt)}(\tau_i) = [\tilde{\pi}_{(ijt)1},\ldots,\tilde{\pi}_{(ijt)c},\ldots,\tilde{\pi}_{(ijt)(C-1)}]'$$

and $\eta^*_{ijt|t-1,M}$ is the $(C-1)\times(C-1)$ matrix of transitional probabilities given by

$$\eta^*_{ijt|t-1,M} = \begin{pmatrix} \eta^{*(1)}_{ijt|t-1}(1) & \cdots & \eta^{*(1)}_{ijt|t-1}(g) & \cdots & \eta^{*(1)}_{ijt|t-1}(C-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta^{*(j)}_{ijt|t-1}(1) & \cdots & \eta^{*(j)}_{ijt|t-1}(g) & \cdots & \eta^{*(j)}_{ijt|t-1}(C-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta^{*(C-1)}_{ijt|t-1}(1) & \cdots & \eta^{*(C-1)}_{ijt|t-1}(g) & \cdots & \eta^{*(C-1)}_{ijt|t-1}(C-1) \end{pmatrix}. \tag{12}$$

Notice that in (11), $\pi^*_{(ijt)}(\tau_i)$ for $t = 2, \ldots, T$, is the same as the transitional probability vector with response transiting from $C$th category. That is,

$$\begin{aligned}
\pi^*_{(ijt)}(\tau_i) &= [\pi^*_{(ijt)1}, \ldots, \pi^*_{(ijt)c}, \ldots, \pi^*_{(ijt)(C-1)}]' \\
&= [\eta^{*(1)}_{ijt|t-1}(C), \ldots, \eta^{*(c)}_{ijt|t-1}(C) \ldots, \eta^{*(C-1)}_{ijt|t-1}(C)]' \\
&= \eta^*_{ijt|t-1}(C). \qquad (13)
\end{aligned}$$

It then follows that

$$\begin{aligned}
\text{var}[Y_{ijt}|\tau_i] &= \text{diag}[\tilde{\pi}_{(ijt)1}, \ldots, \tilde{\pi}_{(ijt)c}, \ldots, \tilde{\pi}_{(ijt)(C-1)}] - \tilde{\pi}_{(ijt)}\tilde{\pi}'_{(ijt)} \\
&= \Sigma^*_{(ijj,tt)}(\tau_i) \\
&= (\text{cov}\{(Y_{ijtr}, Y_{ijtc})|\tau_i\}) = (\sigma^*_{(ijj,tt)rc}(\tau_i)),\ r, c = 1, \ldots, C-1 \\
\text{cov}[(Y_{iju}, Y_{ijt})|\tau_i] &= \text{var}[Y_{iju}|\tau_i][\eta^*_{ijt|t-1,M} - \eta^*_{ijt|t-1}(C)1'_{C-1}]^{t-u},\ \text{for } u < t \\
&= \Sigma^*_{(ijj,ut)}(\tau_i) = (\text{cov}\{(Y_{ijur}, Y_{ijtc})|\tau_i\}) \\
&= (\sigma^*_{(ijj,ut)rc}(\tau_i)),\ r, c = 1, \ldots, C-1, \qquad (14)
\end{aligned}$$

with $\tilde{\pi}_{(ij1)} = \pi^*_{ij1} : (C-1) \times 1$, and where , for example,

$$\begin{aligned}
[\eta^*_{ijt|t-1,M} - \eta^*_{ijt|t-1}(C)1'_{C-1}]^3 &= [\eta^*_{ijt|t-1,M} - \eta^*_{ijt|t-1}(C)1'_{C-1}] \\
\times [\eta^*_{ijt|t-1,M} - \eta^*_{ijt|t-1}(C)1'_{C-1}]&[\eta^*_{ijt|t-1,M} - \eta^*_{ijt|t-1}(C)1'_{C-1}].
\end{aligned}$$

Further, we assume that irrespective of time points, the responses from two individuals from the same family, conditional on $\tau_i$, will be uncorrelated. That is, for $j \neq k$,

$$\text{cov}[\{Y_{iju}, Y_{ikt}\}|\tau_i] = 0,\ \text{for all } u, t = 1, \ldots, T. \qquad (15)$$

## 2.1 Basic Properties of the Model

Similar to the BDML model (1)–(2), we assume that $\tau_i^* \overset{iid}{\sim} N(0, \sigma_\tau^2)$ (Breslow and Clayton 1993; Sutradhar et al. 2008) so that for $\tau_i = \tau_i^*/\sigma_\tau$, $\tau_i \overset{iid}{\sim} N(0, 1)$. By averaging over the distribution of $\tau_i$ involved in the conditional expectation in (11), one obtains the unconditional expectation, i.e., the mean of the responses as

$$
\begin{aligned}
E[Y_{ijt}] &= E_{\tau_i} E[Y_{ijt}|\tau_i] \\
&= \begin{cases} \int_{-\infty}^{\infty} \pi^*_{(ij1)}(\tau_i) f_N(\tau_i) d\tau_i & \text{for } t = 1 \\ \int_{-\infty}^{\infty} \Big[\pi^*_{(ijt)}(\tau_i) + \Big\{\eta^*_{ijt|t-1,M}(\tau_i) - \pi^*_{(ijt)}(\tau_i) 1'_{C-1}\Big\} & \\ \times\, \tilde{\pi}_{(ij,t-1)}(\tau_i)\Big] f_N(\tau_i) d\tau_i & \text{for } t = 2, \ldots, T \end{cases} \\
&= \pi_{(ijt)}(\beta, \gamma, \sigma_\tau), \text{ (say).}
\end{aligned} \tag{16}
$$

where

$$
\beta = [\beta_1^{*'}, \ldots, \beta_c^{*'}, \ldots, \beta_{C-1}^{*'}]'; \; \gamma = [\gamma_1', \ldots, \gamma_c', \ldots, \gamma_{C-1}']',
$$

and $f_N(\tau_i)$ is the standard normal density. Next by using the conditioning and un-conditioning principle over (11) and (14), one may obtain the unconditional covariance matrix of $y_{ijt} : (C-1) \times 1$, as

$$
\begin{aligned}
\text{var}[Y_{ijt}] &= E_{\tau_i} \text{var}[Y_{ijt}|\tau_i] + \text{var}_{\tau_i} E[Y_{ijt}|\tau_i] \\
&= \int_{-\infty}^{\infty} \Big\{\text{diag}[\tilde{\pi}_{(ijt)1}, \ldots, \tilde{\pi}_{(ijt)c}, \ldots, \tilde{\pi}_{(ijt)(C-1)}] - \tilde{\pi}_{(ijt)} \tilde{\pi}'_{(ijt)}\Big\} f_N(\tau_i) d\tau_i \\
&\quad + \int_{-\infty}^{\infty} [\tilde{\pi}_{(ijt)}(\tau_i) \tilde{\pi}'_{(ijt)}(\tau_i)] f_N(\tau_i) d\tau_i \\
&\quad - \{\int_{-\infty}^{\infty} \tilde{\pi}_{(ijt)}(\tau_i) f_N(\tau_i) d\tau_i\}\{\int_{-\infty}^{\infty} \tilde{\pi}'_{(ijt)}(\tau_i) f_N(\tau_i) d\tau_i\} \\
&= \int_{-\infty}^{\infty} \text{diag}[\tilde{\pi}_{(ijt)1}, \ldots, \tilde{\pi}_{(ijt)c}, \ldots, \tilde{\pi}_{(ijt)(C-1)}] f_N(\tau_i) d\tau_i \\
&\quad - \{\int_{-\infty}^{\infty} \tilde{\pi}_{(ijt)}(\tau_i) f_N(\tau_i) d\tau_i\}\{\int_{-\infty}^{\infty} \tilde{\pi}'_{(ijt)}(\tau_i) f_N(\tau_i) d\tau_i\} \\
&= \Sigma_{(ijj,tt)}(\beta, \gamma, \sigma_\tau) = (\sigma_{(ijj,tt),rc}(\beta, \gamma, \sigma_\tau)), \text{ (say),}
\end{aligned} \tag{17}
$$

where, for $r, c = 1, \ldots, C-1$; $\sigma_{(ijj,tt),rc}(\beta, \gamma, \sigma_\tau)$ is the $(r,c)$th element of the $(C-1) \times (C-1)$ covariance matrix $\Sigma_{(ijj,tt)}(\beta, \gamma, \sigma_\tau)$.

Next for the same family member, the covariance of the responses collected from two different time points, by exploiting (11), (14), and (16), may be computed as

$$
\begin{aligned}
\text{cov}[Y_{iju}, Y_{ijt}] &= E_{\tau_i} \text{cov}[\{Y_{iju}, Y_{ijt}\}|\tau_i] + \text{cov}_{\tau_i}[E[Y_{iju}|\tau_i, E[Y_{ijt}|\tau_i] \\
&= \int_{-\infty}^{\infty} \Sigma^*_{(ijj,ut)}(\tau_i) f_N(\tau_i) d\tau_i + \int_{-\infty}^{\infty} [\tilde{\pi}_{(ijt)}(\tau_i) \tilde{\pi}'_{(ijt)}(\tau_i)] f_N(\tau_i) d\tau_i \\
&\quad - \pi_{(iju)}(\beta, \gamma, \sigma_\tau) \pi'_{(ijt)}(\beta, \gamma, \sigma_\tau) \\
&= \Sigma_{(ijj,ut)}(\beta, \gamma, \sigma_\tau) = (\sigma_{(ijj,ut),rc}(\beta, \gamma, \sigma_\tau)).
\end{aligned} \tag{18}
$$

However, when responses are collected from two different members of the same family, by (11), (15), and (16), one writes the covariances as

$$\begin{aligned}\text{cov}[Y_{iju}, Y_{ikt}] &= E_{\tau_i}\text{cov}[\{Y_{iju}, Y_{ikt}\}|\tau_i] + \text{cov}_{\tau_i}[E[Y_{iju}|\tau_i, E[Y_{ikt}|\tau_i]\\ &= \int_{-\infty}^{\infty}[\tilde{\pi}_{(iju)}(\tau_i)\tilde{\pi}'_{(ikt)}(\tau_i)]f_N(\tau_i)d\tau_i - \pi_{(iju)}(\beta,\gamma,\sigma_\tau)\pi'_{(ikt)}(\beta,\gamma,\sigma_\tau)\\ &= \Sigma_{(ijk,ut)}(\beta,\gamma,\sigma_\tau) = (\sigma_{(ijk,ut),rc}(\beta,\gamma,\sigma_\tau)).\end{aligned} \tag{19}$$

Note that it is of primary objective to understand the multinomial mean vector (16), variance matrix (17), and the covariance matrices (18)–(19) of two multinomial response vectors. This analysis requires the estimates for the parameters involved in all these mean, variance and covariance functions. In the next section, we develop the likelihood estimating equations for the parameters involved which provide consistent and highly efficient estimates. We also provide an outline for the GQL estimating equations for the same parameters.

# 3 Likelihood Estimation for the MDML Model Parameters

## *3.1 Construction of the Likelihood Function*

Recall that under the MDML model, the $j$th member of the $i$th family provides a multinomial response with $C \geq 2$ categories, with (a) marginal probability given as in (8)–(9), and (b) the lag 1 type conditional probability as in (10). Further it is assumed that conditional on the random family effect $\tau_i$, the multinomial responses from any two members of the same family will be independent. Consequently, by using (9), $y_{ij1}$, conditional on $\tau_i$, has the marginal multinomial distribution given by

$$f(y_{ij1}|\tau_i) = \frac{1!}{y_{ij11}!\ldots y_{ij1c}!\ldots y_{ij1,C-1}!y_{ij1C}!}\Pi_{c=1}^{C}\left[\pi^{*}_{(ij1)c}\right]^{y_{ij1c}}, \tag{20}$$

where the formula for the probability $\pi^{*}_{(ij1)c}$, is given by (8)–(9). Next, for $t = 2,\ldots,T$, conditional on the random family effect $\tau_i$ and the past response $Y_{ij,t-1} = y^{(g)}_{ij,t-1}$, one may write the conditional distribution of $y_{ijt}$ as

$$\begin{aligned}f(y_{ijt}|y^{(g)}_{ij,t-1},\tau_i) &= \frac{1!}{y_{ijt1}!\ldots y_{ijtc}!\ldots y_{ijt,C-1}!y_{ijtC}!}\\ &\quad\times\Pi_{c=1}^{C}\left[\eta^{*(c)}_{ijt|t-1}(g)\right]^{y_{ijtc}},\ g = 1,\ldots,C,\end{aligned} \tag{21}$$

where $\eta^{*(c)}_{ijt|t-1}(g)$ for $c = 1, \ldots, C$, are given by (10). In (20)–(21), by convention,

$$y_{ijtC} = 1 - \sum_{c=1}^{C-1} y_{ijtc}, \text{ for all } i = 1, \ldots, K,\ j = 1, \ldots, n_i,\ t = 1, \ldots, T.$$

Let

$$\beta \equiv (\beta_1^{*\prime}, \ldots, \beta_c^{*\prime}, \ldots, \beta_{C-1}^{*}{}^{\prime})' : (C-1)p \times 1,$$

where $\beta_c^* = (\beta_{c0}, \beta_c')'$, with $\beta_c = [\beta_{c1}, \ldots, \beta_{cs}, \ldots, \beta_{c,p-1}]'$, and

$$\gamma_M = \begin{pmatrix} \gamma_1' \\ \vdots \\ \gamma_c' \\ \vdots \\ \gamma_{C-1}' \end{pmatrix} : (C-1) \times (C-1). \tag{22}$$

where $\gamma_c = (\gamma_{c1}, \ldots, \gamma_{cv}, \ldots, \gamma_{c,C-1})'$ denotes the dynamic dependence parameters relating $y_{ijt}^{(c)}$ with $y_{ij,t-1}^{(g)}$, for any given category $g = 1, \ldots, C$. By using (20) and (21), one may then write the likelihood function for $\beta$, $\gamma_M$, and $\sigma_\tau$, under the present MDML model (9)–(10), as

$$\begin{aligned} L(\beta, \gamma_M, \sigma_\tau) &= \Pi_{i=1}^K \int_{-\infty}^{\infty} \Pi_{j=1}^{n_i} \left\{ f(y_{ij1}|\tau_i) \Pi_{t=2}^T f(y_{ijt}|y_{ji,t-1}, \tau_i) \right\} \phi(\tau_i) d\tau_i \\ &= c_0^* \Pi_{i=1}^K \int_{-\infty}^{\infty} \Pi_{j=1}^{n_i} \left[ \left\{ \Pi_{c=1}^C \left( \pi^*_{(ij1)c} \right)^{y_{ij1c}} \right\} \right. \\ &\quad \times \left. \left\{ \Pi_{t=2}^T \Pi_{c=1}^C \Pi_{g=1}^C \left( \eta^{*(c)}_{ijt|t-1}(g) \right)^{y_{ijtc}} \right\} \right] f_N(\tau_i) d\tau_i, \end{aligned} \tag{23}$$

where $c_0^*$ is the normalizing constant free from any parameters, $\pi^*_{(ij1)c}$ is a marginal probability at initial time $t = 1$ as in (9), and $\eta^{*(c)}_{ijt|t-1}(g)$ are the conditional probabilities at times $t = 2, \ldots, T$, defined in (10). For convenience, we now re-express the likelihood function in (23) as

$$\begin{aligned} L(\beta, \gamma_M, \sigma_\tau) &= c_0^* \Pi_{i=1}^K \int_{-\infty}^{\infty} L_i\{(\beta, \gamma_M, \sigma_\tau)|\tau_i\} f_N(\tau_i) d\tau_i \\ &= c_0^* \Pi_{i=1}^K J_i(\beta, \gamma_M, \sigma_\tau), \end{aligned} \tag{24}$$

where

$$L_i\{(\beta,\gamma_M,\sigma_\tau)|\tau_i\} = \Pi_{j=1}^{n_i}\left[\left\{\Pi_{c=1}^{C}\left(\pi^*_{(ij1)c}\right)^{y_{ij1c}}\right\}\left\{\Pi_{t=2}^{T}\Pi_{c=1}^{C}\Pi_{g=1}^{C}\left(\eta^{*(c)}_{ijt|t-1}(g)\right)^{y_{ijtc}}\right\}\right].$$

It then follows that the log likelihood function is given by

$$logL(\beta,\gamma_M,\sigma_\tau) = log\ c_0^* + \sum_{i=1}^{K} log\ J_i(\beta,\gamma_M,\sigma_\tau). \tag{25}$$

Note that all integrations over the standard normal distribution of the random effects $\tau_i$ may be computed by using, for example, the so-called Binomial approximation (Sutradhar 2011, Eqs. (5.24)–(5.27); Ten Have and Morabia 1999, Eq. (7)). Thus, $J_i$ in (24) or (25) may be computed as

$$\begin{aligned} J_i(\beta,\gamma_M,\sigma_\tau) &= \int_{-\infty}^{\infty} L_i(\tau_i) f_N(\tau_i) d\tau_i \\ &= \sum_{v=0}^{V} L_i(\tau_i(v)) \binom{V}{v} (1/2)^v (1/2)^{V-v}, \end{aligned} \tag{26}$$

where $V$ is a user's choice large selected number of Bernoulli trials such as $V = 10$, and

$$\tau_i(v) = \frac{v - V(\frac{1}{2})}{\sqrt{V(\frac{1}{2})(\frac{1}{2})}}.$$

## 3.2 Likelihood Estimating Equations

Suppose that $\phi$ represents any one of the three parameters $\beta, \gamma_M,$ or $\sigma_\tau$. Then in general the score function for $\phi$ has the form

$$\frac{\partial logL(\beta,\gamma_M,\sigma_\tau)}{\partial\phi} = \sum_{i=1}^{K} \frac{1}{J_i}\frac{\partial J_i}{\partial\phi}, \tag{27}$$

where

$$\begin{aligned} \frac{\partial J_i}{\partial\phi} &= \int_{-\infty}^{\infty} \frac{\partial L_i\{(\beta,\gamma_M,\sigma_\tau)|\tau_i\}}{\partial\phi} f_N(\tau_i) d\tau_i \\ &\equiv \int_{-\infty}^{\infty} \frac{\partial L_i(\tau_i)}{\partial\phi} f_N(\tau_i) d\tau_i, \end{aligned} \tag{28}$$

with

$$\begin{aligned}\frac{\partial L_i(\tau_i)}{\partial \phi} &= L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial \phi} \\ &= L_i(\tau_i)\left[\frac{\partial}{\partial \phi}\left\{\sum_{j=1}^{n_i}\sum_{c=1}^{C} y_{ij1c} log\ \pi^*_{(ij1)c}\right\}\right. \\ &\quad \left. + \frac{\partial}{\partial \phi}\left\{\sum_{t=2}^{T}\sum_{j=1}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C} y_{ijtc} log\ \eta^{*(c)}_{ijt|t-1}(g)\right\}\right]. \end{aligned} \tag{29}$$

Note that similar to the computation of $J_i(\cdot)$ by (26), the derivative $\frac{\partial J_i}{\partial \phi}$ in (28) may be computed as

$$\begin{aligned}\frac{\partial J_i}{\partial \phi} &= \int_{-\infty}^{\infty} \frac{\partial L_i(\tau_i)}{\partial \phi} f_N(\tau_i) d\tau_i \\ &= \sum_{v=0}^{V} \frac{\partial L_i(\tau_i(v))}{\partial \phi}\binom{V}{v}(1/2)^v(1/2)^{V-v}. \end{aligned} \tag{30}$$

### 3.2.1 Likelihood Estimating Equation for $\beta$

Recall that

$$\beta = [\beta_1^{*'}, \ldots, \beta_c^{*'}, \ldots, \beta_{C-1}^{*'}]' : (C-1)p \times 1$$

is a global regression parameter vector, $\beta_c^* : p \times 1$ being the local regression parameter vector under the $c$ category for $c = 1, \ldots, C-1$, and $\beta_C^* = 0$ for the reference category. These local regression parameter vectors are involved in the marginal probabilities at initial time $t = 1$ and in dynamic relationships, that is, in conditional probabilities, for $t = 2, \ldots, T$, as shown by (9) and (10), respectively. Now to compute the score function in $\beta$ by (27)–(28), we need to compute $\frac{\partial L_i(\tau_i)}{\partial \beta}$ by using (29). Thus, we write

$$\begin{aligned}\frac{\partial L_i(\tau_i)}{\partial \beta} &= L_i(\tau_i)\frac{\partial Log\ L_i(\beta, \gamma_M, \sigma_\tau)}{\partial \beta} \\ &= L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C}\frac{y_{ij1c}}{\pi^*_{(ij1)c}}\frac{\partial \pi^*_{(ij1)c}}{\partial \beta}\right. \\ &\quad \left. + \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\left\{\frac{y_{ijtc}}{\eta^{*(c)}_{ijt|t-1}(g)}\frac{\partial \eta^{*(c)}_{ijt|t-1}(g)}{\partial \beta}\right\}\right]. \end{aligned} \tag{31}$$

Next, putting this derivative in (28), after some algebras as shown in the Appendix, one obtains the likelihood estimating equation for $\beta$, by (27) as

$$\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta} = \sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty} L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial\beta} f_N(\tau_i)d\tau_i$$

$$= \sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty} L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C}\frac{y_{ij1c}}{\pi^*_{(ij1)c}}\left[\left\{\pi^*_{(ij1)c}(\delta_{(ij1)c}-\pi^*_{(ij1)})\right\}\otimes x_{ij1}\right]\right.$$

$$\left. + \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\frac{y_{ijtc}}{\eta^{*(c)}_{ijt|t-1}(g)}\left[\left\{\eta^{*(c)}_{ijt|t-1}(g)(\delta_{(ij,t-1)c}-\eta^*_{ijt|t-1}(g))\right\}\otimes x_{ijt}\right]\right] f_N(\tau_i)d\tau_i$$

$$= 0, \quad (32)$$

where $\otimes$ denotes the Kronecker product, and

$$\delta_{(ijt)c} = \begin{cases} [01'_{c-1}, 1, 01'_{C-1-c}]' & \text{for } c = 1,\ldots,C-1 \\ 01_{C-1} & \text{for } c = C, \end{cases}$$

for all $j = 1,\ldots,n_i,\ t = 1,\ldots,T,\ i = 1,\ldots,K;$

$$\pi^*_{(ij1)} \equiv \pi^*_{(ij1)}(\tau_i) = [\pi^*_{(ij1)1},\ldots,\pi^*_{(ij1)c},\ldots,\pi^*_{(ij1)(C-1)}]' : (C-1),$$

by (11), and

$$\eta^*_{ijt|t-1}(g) = [\eta^{*(1)}_{ijt|t-1}(g),\ldots,\eta^{*(c)}_{ijt|t-1}(g),\ldots,\eta^{*(C-1)}_{ijt|t-1}(g)]'.$$

Now, for given values for $\gamma_M$ and $\sigma_\tau$, the likelihood equations in (32) may be solved iteratively by using the iterative equations for $\beta$ given by

$$\hat{\beta}(r+1) = \hat{\beta}(r) - \left[\left\{\frac{\partial^2 log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta'\partial\beta}\right\}^{-1}\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta}\right]_{|\beta=\hat{\beta}(r)};\ (J-1)p\times 1, \quad (33)$$

where the formula for the second order derivative matrix $\frac{\partial^2 log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta'\partial\beta}$ may be derived by taking the derivative of the $(J-1)p\times 1$ vector with respect to $\beta'$. The exact second order derivative has a complicated formula. We provide an approximation for this second order derivative matrix as follows.

**An approximation for $\frac{\partial^2 Log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta'\partial\beta}$ based on iteration principle:**

In this approach, one assumes that $\beta$ in the derivatives in (31), that is, $\beta$ involved in $\frac{\partial\pi^*_{(ij1)c}}{\partial\beta}$ and $\frac{\partial\eta^{*(c)}_{ijt|t-1}(g)}{\partial\beta}$ are known from a previous iteration, and then take the derivative of $\frac{\partial Log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta}$ in (31) or (32), with respect to $\beta'$. This provides a simpler formula for the second order derivative as

$$
\begin{aligned}
&\frac{\partial^2 log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta'\partial\beta} \\
&= -\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty}L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C}\frac{y_{ij1c}}{[\pi^*_{(ij1)c}]^2}\right. \\
&\times\left[\left\{\pi^*_{(ij1)c}(\delta_{(ij1)c}-\pi^*_{(ij1)})\right\}\otimes x_{ij1}\right]\left[\left\{\pi^*_{(ij1)c}(\delta_{(ij1)c}-\pi^*_{(ij1)})\right\}\otimes x_{ij1}\right]' \\
&+\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\frac{y_{ijtc}}{[\eta^{*(c)}_{ijt|t-1}(g)]^2}\left[\left\{\eta^{*(c)}_{ijt|t-1}(g)(\delta_{(ij,t-1)c}-\eta^*_{ijt|t-1}(g))\right\}\otimes x_{ijt}\right] \\
&\times\left.\left[\left\{\eta^{*(c)}_{ijt|t-1}(g)(\delta_{(ij,t-1)c}-\eta^*_{ijt|t-1}(g))\right\}\otimes x_{ijt}\right]'\right]f_N(\tau_i)d\tau_i \\
&+\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty}L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C}\frac{y_{ij1c}}{\pi^*_{(ij1)c}}\left[\left\{\pi^*_{(ij1)c}(\delta_{(ij1)c}-\pi^*_{(ij1)})\right\}\otimes x_{ij1}\right]\right. \\
&+\left.\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\frac{y_{ijtc}}{\eta^{*(c)}_{ijt|t-1}(g)}\left[\left\{\eta^{*(c)}_{ijt|t-1}(g)(\delta_{(ij,t-1)c}-\eta^*_{ijt|t-1}(g))\right\}\otimes x_{ijt}\right]\right] \\
&\times\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta'}f_N(\tau_i)d\tau_i \\
&-\sum_{i=1}^{K}\frac{1}{J_i^2}\int_{-\infty}^{\infty}L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C}\frac{y_{ij1c}}{\pi^*_{(ij1)c}}\left[\left\{\pi^*_{(ij1)c}(\delta_{(ij1)c}-\pi^*_{(ij1)})\right\}\otimes x_{ij1}\right]\right. \\
&+\left.\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\frac{y_{ijtc}}{\eta^{*(c)}_{ijt|t-1}(g)}\left[\left\{\eta^{*(c)}_{ijt|t-1}(g)(\delta_{(ij,t-1)c}-\eta^*_{ijt|t-1}(g))\right\}\otimes x_{ijt}\right]\right]f_N(\tau_i)d\tau_i \\
&\times\frac{\partial J_i}{\partial\beta'},
\end{aligned}
\tag{34}
$$

where $\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta'}$ is the transpose of $\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\beta}$ easily computed from (32), and

$$
\frac{\partial J_i}{\partial\beta'}=\int_{-\infty}^{\infty}L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial\beta'}f_N(\tau_i)d\tau_i \tag{35}
$$

which also can easily be extracted from (32).

### 3.2.2 Likelihood Estimating Equation for $\gamma_M$

The estimation of the $(C-1)\times(C-1)$ matrix $\gamma_M$ is equivalent to estimate $\gamma$ in (16), where

$$\gamma = (\gamma'_1,\ldots,\gamma'_c,\ldots,\gamma'_{C-1})' : (C-1)^2 \times 1; \text{ with } \gamma_c = (\gamma_{c1},\ldots,\gamma_{ch},\ldots,\gamma_{c,C-1})' \tag{36}$$

as the $(C-1)\times 1$ vector of dynamic dependence parameters involved in the conditional multinomial logit function in (10). Similar to (32), one writes

$$\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma} = \sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty} L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial\gamma} f_N(\tau_i)d\tau_i, \tag{37}$$

where

$$\begin{aligned}\frac{\partial log\ L_i(\tau_i)}{\partial\gamma} &\equiv \frac{\partial log\ L_i(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma}\\ &= \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\left\{\frac{y_{ijtc}}{\eta^{*(c)}_{ijt|t-1}(g)}\frac{\partial\eta^{*(c)}_{ijt|t-1}(g)}{\partial\gamma}\right\}.\end{aligned} \tag{38}$$

Next, as shown in the Appendix, because

$$\frac{\partial log\ L_i(\tau_i)}{\partial\gamma_c} = \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{g=1}^{C}\delta_{(ij,t-1)g}\left[y_{ijtc}-\left(\eta^{*(c)}_{ijt|t-1}(g)\right)\right], \tag{39}$$

we obtain the likelihood estimating equation for $\gamma$ as

$$\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma} = \begin{pmatrix}\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma_1}\\ \vdots\\ \frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma_c}\\ \vdots\\ \frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma_{C-1}}\end{pmatrix} = 0 : (C-1)^2\times 1, \tag{40}$$

where

$$\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\gamma_c} = \sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty} L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial\gamma_c} f_N(\tau_i)d\tau_i$$

$$= \sum_{i=1}^{K} \frac{1}{J_i} \int_{-\infty}^{\infty} L_i(\tau_i) \sum_{j=1}^{n_i} \sum_{t=2}^{T} \sum_{g=1}^{C} \delta_{(ij,t-1)g} \times \left[ y_{ijtc} - \left( \eta^{*(c)}_{ijt|t-1}(g) \right) \right] f_N(\tau_i) d\tau_i. \quad (41)$$

One may now solve this likelihood equation (40) for $\gamma$ by using the iterative equation

$$\hat{\gamma}(r+1) = \hat{\gamma}(r) - \left[ \left\{ \frac{\partial^2 \log L(\beta, \gamma_M, \sigma_\tau)}{\partial \gamma \partial \gamma'} \right\}^{-1} \frac{\partial \log L(\beta, \gamma_M, \sigma_\tau)}{\partial \gamma} \right]_{|\gamma = \hat{\gamma}(r)}, \quad (42)$$

where the $(C-1)^2 \times (C-1)^2$ second derivative matrix is computed by using the formula

$$\begin{aligned}
&\frac{\partial^2 \log L(\beta, \gamma_M, \sigma_\tau)}{\partial \gamma_c \partial \gamma_k'} \\
&= -\sum_{i=1}^{K} \frac{1}{J_i} \int_{-\infty}^{\infty} L_i(\tau_i) \sum_{j=1}^{n_i} \sum_{t=2}^{T} \sum_{g=1}^{C} \delta_{(ij,t-1)g} \delta'_{(ij,t-1)g} \eta^{*(c)}_{ijt|t-1}(g) \left[ \delta^*_{ck} - \eta^{*(k)}_{ijt|t-1}(g) \right] f_N(\tau_i) d\tau_i \\
&+ \sum_{i=1}^{K} \frac{1}{J_i} \int_{-\infty}^{\infty} L_i(\tau_i) \sum_{j=1}^{n_i} \sum_{t=2}^{T} \sum_{g=1}^{C} \delta_{(ij,t-1)g} \left[ y_{ijtc} - \left( \eta^{*(c)}_{ijt|t-1}(g) \right) \right] \\
&\times \sum_{j=1}^{n_i} \sum_{t=2}^{T} \sum_{g=1}^{C} \delta'_{(ij,t-1)g} \left[ y_{ijtk} - \left( \eta^{*(k)}_{ijt|t-1}(g) \right) \right] f_N(\tau_i) d\tau_i \\
&- \sum_{i=1}^{K} \frac{1}{J_i^2} \int_{-\infty}^{\infty} L_i(\tau_i) \sum_{j=1}^{n_i} \sum_{t=2}^{T} \sum_{g=1}^{C} \delta_{(ij,t-1)g} \left[ y_{ijtc} - \left( \eta^{*(c)}_{ijt|t-1}(g) \right) \right] f_N(\tau_i) d\tau_i \frac{\partial J_i}{\partial \gamma_k'}, \quad (43)
\end{aligned}$$

where

$$\delta^*_{ck} = \begin{cases} 1 \text{ for } c = k, c, k = 1, \ldots, C-1 \\ 0 \text{ for } c \neq k, \ c, k = 1, \ldots, C-1, \end{cases}$$

and

$$\begin{aligned}
\frac{\partial J_i}{\partial \gamma_k'} &= \int_{-\infty}^{\infty} L_i(\tau_i) \frac{\partial log\ L_i(\tau_i)}{\partial \gamma_k'} f_N(\tau_i) d\tau_i \\
&= \int_{-\infty}^{\infty} L_i(\tau_i) \sum_{j=1}^{n_i} \sum_{t=2}^{T} \sum_{g=1}^{C} \delta'_{(ij,t-1)g} \left[ y_{ijtk} - \left( \eta^{*(k)}_{ijt|t-1}(g) \right) \right] f_N(\tau_i) d\tau_i. \quad (44)
\end{aligned}$$

### 3.2.3 Likelihood Estimating Equation for $\sigma_\tau$

Similar to (37)–(38) [see also (32)], one obtains the score function for $\sigma_\tau$, as

$$\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma_\tau} = \sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty} L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial\sigma_\tau} f_N(\tau_i)d\tau_i, \tag{45}$$

where

$$\frac{\partial log\ L_i(\tau_i)}{\partial\sigma_\tau} \equiv \frac{\partial log\ L_i(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma_\tau}$$
$$= \sum_{j=1}^{n_i}\sum_{c=1}^{C}\frac{y_{ij1c}}{\pi^*_{(ij1)c}}\frac{\partial\pi^*_{(ij1)c}}{\partial\sigma_\tau} + \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C}\left\{\frac{y_{ijtc}}{\eta^{*(c)}_{ijt|t-1}(g)}\frac{\partial\eta^{*(c)}_{ijt|t-1}(g)}{\partial\sigma_\tau}\right\}. \tag{46}$$

Next, by using the formulas for the derivatives from section "Algebras for the Likelihood Estimating Equation (47) for $\sigma_\tau$" in Appendix in (46), the likelihood estimating equation for $\sigma_\tau$ is obtained by (45) as

$$\begin{aligned}\frac{\partial log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma_\tau} &= -\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty}\tau_i L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C} y_{ij1c}\left(\delta^*_{cC}-\pi^*_{(ij1)C}\right)\right.\\ &\quad\left.+\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C} y_{ijtc}\left(\delta^*_{cC}-\eta^{*(C)}_{ijt|t-1}(g)\right)\right] f_N(\tau_i)d\tau_i\\ &= 0,\end{aligned} \tag{47}$$

where

$$\delta^*_{cC} = \begin{cases} 1 \text{ for } c = C \\ 0 \text{ for } c \neq C,\ c = 1,\ldots,C-1. \end{cases}$$

This likelihood equation (47) for $\sigma_\tau$ may be solved iteratively, by using the formula

$$\widehat{\sigma_\tau}(r+1) = \widehat{\sigma_\tau}(r) - \left[\left\{\frac{\partial^2 \log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma^2_\tau}\right\}^{-1}\frac{\partial \log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma_\tau}\right]_{|\gamma=\hat{\gamma}(r)}, \tag{48}$$

where the formula for the second order derivative $\frac{\partial^2 \log\ L(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma^2_\tau}$ is computed from (47) as

$$
\begin{aligned}
\frac{\partial^2 \log L(\beta,\gamma_M,\sigma_\tau)}{\partial\sigma_\tau^2} &= -\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty}\tau_i^2 L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C} y_{ij1c}\pi^*_{(ij1)C}\left(1-\pi^*_{(ij1)C}\right)\right.\\
&\quad\left.+\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C} y_{ijtc}\eta^{*(C)}_{ijt|t-1}(g)\left(1-\eta^{*(C)}_{ijt|t-1}(g)\right)\right] f_N(\tau_i)d\tau_i\\
&\quad+\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty}\tau_i^2 L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C} y_{ij1c}\left(\delta^*_{cC}-\pi^*_{(ij1)C}\right)\right.\\
&\quad\left.+\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C} y_{ijtc}\left(\delta^*_{cC}-\eta^{*(C)}_{ijt|t-1}(g)\right)\right]^2 f_N(\tau_i)d\tau_i\\
&\quad+\sum_{i=1}^{K}\frac{1}{J_i^2}\int_{-\infty}^{\infty}\tau_i L_i(\tau_i)\left[\sum_{j=1}^{n_i}\sum_{c=1}^{C} y_{ij1c}\left(\delta^*_{cC}-\pi^*_{(ij1)C}\right)\right.\\
&\quad\left.+\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{c=1}^{C}\sum_{g=1}^{C} y_{ijtc}\left(\delta^*_{cC}-\eta^{*(C)}_{ijt|t-1}(g)\right)\right] f_N(\tau_i)d\tau_i\frac{\partial J_i}{\partial\sigma_\tau},
\end{aligned}
\tag{49}
$$

with

$$
\frac{\partial J_i}{\partial\sigma_\tau} = \int_{-\infty}^{\infty} L_i(\tau_i)\frac{\partial log\ L_i(\tau_i)}{\partial\sigma_\tau} f_N(\tau_i)d\tau_i.
$$

## 4 A Remark on Likelihood Versus GQL Estimation for the Binary Case

An analytical efficiency comparison between the likelihood and the so-called GQL (generalized quasi-likelihood) estimation approaches is extremely difficult because these approaches produce complicated formulas for the asymptotic variances. Thus, to shed some lights on their relative asymptotic efficiency performances, in this section, we conduct an empirical study. For further convenience, we choose $C = 2$ categories, that is, the familial longitudinal binary case. The formulas for the asymptotic variances of the likelihood estimators of $\beta$, $\gamma$, and $\sigma_\tau$ are given in Sect. 4.1 below. The formulas for the corresponding asymptotic variances of the GQL estimators are computed in Sect. 4.2. In Sect. 4.3, a numerical efficiency comparison is given by using the formulas from Sects. 4.2 and 4.3.

### *4.1 Computation of Asymptotic Variance of the Likelihood Estimators*

For the familial longitudinal binary ($C = 2$) case, let $\alpha = (\beta', \gamma, \sigma_\tau^2)'$ be the $p + 2$-dimensional vector of parameters, and $\hat{\alpha}_{ML}$ denote the ML estimator of $\alpha$. Next suppose that $Q$ is the $(p + 2) \times (p + 2)$ Hessian matrix, which is computed as

$$Q = -\begin{bmatrix} \frac{\partial^2 log\ L}{\partial\beta\partial\beta'} & \frac{\partial^2 log\ L}{\partial\beta\partial\gamma} & \frac{\partial^2 log\ L}{\partial\beta\partial\sigma_\tau} \\ \cdot & \frac{\partial^2 log\ L}{\partial\gamma^2} & \frac{\partial^2 log\ L}{\partial\gamma\partial\sigma_\tau} \\ \cdot & \cdot & \frac{\partial^2 log\ L}{\partial\sigma_\tau^4} \end{bmatrix}. \tag{50}$$

It then follows that as $K \to \infty$, $\hat{\alpha}_{ML}$ follows the $(p+2)$-dimensional Gaussian distribution with mean $\alpha$ and covariance matrix

$$\text{cov}(\hat{\alpha}_{ML}) = I^{-1}(\alpha) = -[E_y Q]^{-1}, \tag{51}$$

with its diagonal elements as the asymptotic variances of the ML estimators. The elements of the expected Hessian matrix, i.e. of $E_y Q$ may be computed by taking expectation numerically over the elements of the $Q$ matrix. For example, to compute $E[\frac{\partial^2 log\ L}{\partial\beta\partial\beta'}]$, we first re-express the second order derivative matrix from (34) as

$$\begin{aligned} \frac{\partial^2 log\ L}{\partial\beta\partial\beta'} = &-\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty} f_{i1}(\tau_i,\beta,\gamma,\sigma_\tau) f_N(\tau_i) d\tau_i \\ &+\sum_{i=1}^{K}\frac{1}{J_i}\int_{-\infty}^{\infty}\left[f_{i2}(\tau_i,\beta,\gamma,\sigma_\tau)\frac{\partial logL(\beta,\gamma,\sigma_\tau)}{\partial\beta'}\right] f_N(\tau_i) d\tau_i \\ &-\sum_{i=1}^{K}\frac{1}{J_i^2}\frac{\partial J_i}{\partial\beta'}\int_{-\infty}^{\infty} f_{i3}(\tau_i,\beta,\gamma,\sigma_\tau) f_N(\tau_i) d\tau_i. \end{aligned} \tag{52}$$

Now the expectation is computed by summing (52) over the values 0 and 1, for all $\{y_{ijt1}\}$. Thus,

$$\begin{aligned} E\left[\frac{\partial^2 \text{log} L}{\partial\beta\partial\beta'}\right] = &\sum_{i=1}^{K}\sum_{y_{i111}=0}^{1}\sum_{y_{i121}=0}^{1}\cdots\sum_{y_{ijt1}=0}^{1}\cdots\sum_{y_{in_iT1}=0}^{1} \\ &\times\left[-\frac{1}{J_i}\sum_{v=0}^{V} f_{i1}(\tau_{iv},\beta,\gamma,\sigma_\tau)\binom{V}{v}(1/2)^v(1/2)^{V-v}\right. \\ &+\frac{1}{J_i}\sum_{v=0}^{V}\left\{f_{i2}(\tau_{iv},\beta,\gamma,\sigma_\tau)\frac{\partial logL(\beta,\gamma,\sigma_\tau)}{\partial\beta'}\right\}\binom{V}{v}(1/2)^v(1/2)^{V-v} \\ &\left.-\frac{1}{J_i^2}\frac{\partial J_i}{\partial\beta'}\sum_{v=0}^{V} f_{i3}(\tau_{iv},\beta,\gamma,\sigma_\tau)\binom{V}{v}(1/2)^v(1/2)^{V-v}\right], \end{aligned} \tag{53}$$

where, similar to (26), $V$ is a user's choice large selected number of Bernoulli trials such as $V = 10$, and

$$\tau_{iv} = \frac{v - V(\frac{1}{2})}{\sqrt{V(\frac{1}{2})(\frac{1}{2})}}.$$

## *4.2 Computation of Asymptotic Variance of the GQL Estimators*

As shown in Sect. 2.1, the first and all second order moments, namely the squared and pairwise product moments for all members of a given family contain the parameters $\beta, \gamma$, and $\sigma_\tau^2$. Let the first and second order response vectors under the $i$th family ($i = 1, \ldots, K$) be written as

$$\begin{aligned}
y_i &= [y_{i11}, \ldots, y_{ijt}, \ldots, y_{in_iT}]' : n_iT \times 1, \\
s_i^* &= [y_{i11}^2, \ldots, y_{ijt}^2, \ldots, y_{in_iT}^2]' : n_iT \times 1, \\
s_{i1} &= [y_{i11}y_{i12}, \ldots, y_{iju}y_{ijt}, \ldots, y_{in_i(T-1)}y_{in_iT}]' : n_iT(T-1)/2 \times 1, \text{ and} \\
s_{i2} &= [y_{i11}y_{i21}, \ldots, y_{iju}y_{ikt}, \ldots, y_{i(n_i-1)T}y_{in_iT}]' : \frac{n_i(n_i-1)}{2}T^2 \times 1
\end{aligned} \tag{54}$$

Next, consider a vector of distinct first and second order responses, given by

$$f_i = [y_i', s_{i1}', s_{i2}']'. \tag{55}$$

For $\alpha = [\beta', \gamma, \sigma_\tau^2]$, suppose that

$$E[f_i] = \xi_i(\alpha), \text{ and } \mathrm{cov}[f_i, f_i'] = \Omega_i(\alpha). \tag{56}$$

One may then write a GQL estimating equation for $\alpha$ as

$$\sum_{i=1}^{K} \frac{\partial \xi_i'}{\partial \alpha} \Omega_i^{-1}(\alpha) \left(f_i - \xi_i(\alpha)\right) = 0 \tag{57}$$

(Sutradhar 2003, Sect. 3; Sutradhar 2004; Sutradhar et al. 2008), where $\frac{\partial \xi_i'(\alpha)}{\partial \alpha}$ denotes the matrix of the first derivative of $\xi_i(\alpha)$ with respect to the components of $\alpha$.

Let $\hat{\alpha}_{GQL}$ denote the solution of (57). Since the expectation of the GQL estimating function in the left hand side of this equation is zero, $\hat{\alpha}_{GQL}$ is consistent for $\alpha$. The GQL estimator $\hat{\alpha}_{GQL}$ is also expected to be highly efficient because of the fact that the GQL estimating equation (57) is constructed by using the inverse of the covariance matrix $\Omega_i$ as the weight matrix. Furthermore, under some mild regularity conditions it may be shown that $\hat{\alpha}_{GQL}$ asymptotically ($K \rightarrow \infty$) follows a Gaussian distribution with mean $\alpha$ and covariance matrix

$$\left[\sum_{i=1}^{K} \frac{\partial \xi_i'(\alpha)}{\partial \alpha} \Omega_i^{-1}(\alpha) \frac{\partial \xi_i(\alpha)}{\partial \alpha'}\right]^{-1}. \tag{58}$$

For computational convenience, the mean vector and the covariance matrix used in the GQL estimating equation (57), may be re-expressed as

$$\begin{aligned} \xi_i(\alpha) = E(f_i) &= [E(Y_i'), E(S_{i1}'), E(S_{i2}')]' \\ &= [\mu_i', \lambda_{i1}', \lambda_{i2}']', \text{ (say)} \end{aligned} \tag{59}$$

$$\begin{aligned} \Omega_i(\alpha) = \text{cov}(f_i) &= \begin{bmatrix} \text{cov}(Y_i) & \text{cov}(Y_i, S_{i1}') & \text{cov}(Y_i, S_{i2}') \\ \cdot & \text{cov}(S_{i1}) & \text{cov}(S_{i1}, S_{i2}') \\ \cdot & \cdot & \text{cov}(S_{i2}) \end{bmatrix} \\ &= \begin{bmatrix} \Sigma_i & \Delta_{i11} & \Delta_{i12} \\ \cdot & \Omega_{i11} & \Omega_{i12} \\ \cdot & \cdot & \Omega_{i22} \end{bmatrix}, \text{ (say)}. \end{aligned} \tag{60}$$

All these first and second order moments may be easily computed empirically. We demonstrate below how the general elements of $\mu_i$, $\lambda_{i2}$, and $\Omega_i$, for example, may be computed. Suppose that for given $i$ and $j$, $s_0 \equiv \{y_{ij1}, \ldots, y_{ijt}, \ldots, y_{ijT}\}$ be the whole sample space. Also, for $r = 1, \ldots, 4$, let $S_r$ denote a fixed set with $r$ selected binary components. This implies that for given $i$ and $j$, $s_r \equiv s_0 - S_r$ denotes the sample space with $T - r$ binary components. One may then compute the general element of the $\mu_i$ vector, say $\mu_{ijt} = E[Y_{ijt}]$ by using $S_1 \equiv y_{ijt}$ and summing the likelihood function over the ranges of all binary elements contained in $s_1 = s_0 - S_1$. Thus,

$$\begin{aligned} \mu_{ijt} = E[Y_{ijt}] = \sum_{v=1}^{V} \sum_{\{s_1\}}^{1} \Big[\Big[\{(p_{ij10}^*(\tau_{iv}))^{y_{ij1}} (1 - p_{ij10}^*(\tau_{iv}))^{1-y_{ij1}}\} \\ \times \Pi_{h=2}^{T} \{(p_{ijhy_{ij,h-1}}^*(\tau_{iv}))^{y_{ijh}} (1 - p_{ijhy_{ij,h-1}}^*(\tau_{iv}))^{1-y_{ijh}}\}\Big] \binom{V}{v} (1/2)^V\Big]_{|y_{ijt}=1}, \end{aligned} \tag{61}$$

where the marginal probability $p^*_{ij10}(\tau_{iv})$ and the conditional probability $p^*_{ijhy_{ij,h-1}}(\tau_{iv})$ are given by (1) and (2), respectively. Similarly, by writing $S_2 \equiv [y_{iju}, y_{ikt}]$, and $s_2 = s_0 - S_2$, one may compute the general element of $\lambda_{i2}$ as

$$\begin{aligned}\lambda_{i2,ju,kt} = E[Y_{iju}Y_{ikt}] = \sum_{v=1}^{V}\sum_{\{s_2\}}^{1} & \Big[\Big[\{(p^*_{ij10}(\tau_{iv}))^{y_{ij1}}(1-p^*_{ij10}(\tau_{iv}))^{1-y_{ij1}}\} \\ & \times \Pi_{h=2}^{T}\{(p^*_{ijhy_{ij,h-1}}(\tau_{iv}))^{y_{ijh}}(1-p^*_{ijhy_{ij,h-1}}(\tau_{iv}))^{1-y_{ijh}}\}\Big] \\ & \times \binom{V}{v}(1/2)^V\Big]_{|y_{iju}=1,y_{ikt}=1}. \end{aligned} \tag{62}$$

Next we show how a general element of the $\Omega_i$ matrix can be calculated. Consider

$$S_4 = [Y_{iju}, Y_{ij\ell}, Y_{ijm}, Y_{ijt}]$$

with four elements for given $i$ and $j$, and write $s_4 = s_0 - S_4$. Then a general element of $\Omega_i$ matrix, namely, $E(Y_{iju}Y_{ij\ell}Y_{ijm}Y_{ijt})$, may be computed as

$$\begin{aligned}E(Y_{iju}Y_{ij\ell}Y_{ijm}Y_{ijt}) = \sum_{v=1}^{V}\sum_{\{s_4\}}^{1} & \Big[\Big[\{(p^*_{ij10}(\tau_{iv}))^{y_{ij1}}(1-p^*_{ij10}(\tau_{iv}))^{1-y_{ij1}}\} \\ & \times \Pi_{h=2}^{T}\{(p^*_{ijhy_{ij,h-1}}(\tau_{iv}))^{y_{ijh}}(1-p^*_{ijhy_{ij,h-1}}(\tau_{iv}))^{1-y_{ijh}}\}\Big] \\ & \times \binom{V}{v}(1/2)^V\Big]_{|y_{iju}=1,y_{ij\ell}=1,y_{ijm}=1,y_{ijt}=1}. \end{aligned} \tag{63}$$

### 4.3 An Empirical Asymptotic Efficiency Comparison Between GQL and ML Estimates

Notice that both GQL estimating equation (57) and the likelihood estimating equations in (32), (40), and (47) produce consistent estimators because of the fact that the GQL estimating function (left hand side of (57)) as well as the likelihood estimating functions [left hand side of (32), (40), and (47)], are all unbiased for zero. However, as the analytical efficiency comparison between these GQL and likelihood estimators, is complicated, to examine the relative performance of these two competitive approaches, in this section, we make an empirical efficiency comparison by comparing the asymptotic variances in (58) for the GQL estimators with those of the ML estimators given in the diagonal elements of (51).

As far as the true parameters are concerned, we choose three sets of 2-dimensional main regression parameters, namely,

$$\beta_1 = \beta_2 = 0.0;\ \beta_1 = \beta_2 = 0.5;\ \text{and}\ \beta_1 = \beta_2 = 1.0.$$

For a given set of regression parameters, we then choose three values for the dynamic dependence parameter, namely, $\gamma = -1.0,\ 0.5,\ 1.0$, covering both negative and positive dependence, and three values for the variance of the random family effects, namely, $\sigma_\tau^2 = 0.5,\ 0.8,\ 1.2$, covering small and large familial variation. We consider two covariates from $n_i = 2$ members of $K = 100$ families, over a period of $T = 3$ time points. These time dependent covariates were chosen as:

$$x_{i1t1} = \begin{cases} 1/2 & \text{for}\ i = 1,\ldots,K/4;\ t = 1,2 \\ 0 & \text{for} i = 1,\ldots,K/4;\ t = 3 \\ -1/2 & \text{for}\ i = K/4+1,\ldots,3K/4;\ t = 1 \\ 0 & \text{for}\ i = K/4+1,\ldots,3K/4;\ t = 2,3 \\ t/6 & \text{for}\ i = 3K/4+1,\ldots,K;\ t = 1,\ldots,3, \end{cases},$$

$$x_{i1t2} = \begin{cases} (t-2.0)/2 & \text{for}\ i = 1,\ldots,K/2;\ t = 1,\ldots,3 \\ 0 & \text{for}\ i = K/2+1,\ldots,K;\ t = 1,2 \\ 1/2 & \text{for}\ i = K/2+1,\ldots,K;\ t = 3. \end{cases}$$

For the second member of the families we consider the first covariate as binary variable following the distribution

$$Pr(x_{i2t1} = 1) = \begin{cases} 0.3 & \text{for}\ t = 1 \\ 0.8 & \text{for}\ t = 2,3, \end{cases},$$

and the second covariate is chosen as:

$$x_{i2t2} = \begin{cases} (t-2.0)/2 & \text{for}\ i = 1,\ldots,K/2;\ t = 1,\ldots,3 \\ t/2 & \text{for}\ i = K/2+1,\ldots,K;\ t = 1,\ldots,3. \end{cases}$$

The variances for selected parameter values obtained from (51) and (58), are reported in Tables 1, 2, and 3, for three sets of regression parameter values.

The results of these three tables show that (1) the ML approach always produce smaller variances than the GQL approach for the estimation of the dynamic dependence parameter, $\gamma$, indicating that ML estimator of $\gamma$ is more efficient than the GQL estimator; (2) the ML estimators of the regression parameters $\beta_1$ and $\beta_2$ have smaller variances than the corresponding GQL estimators when dynamic dependence between the binary responses is negative, but the pattern reverses when responses exhibit positive dynamic dependence; (3) and for the estimation of the random effects variance $\sigma_\tau^2$, the ML approach appears to be more efficient when true $\sigma_\tau^2$ is small, but, GQL estimator becomes more efficient for true large values of $\sigma_\tau^2$. Thus, the ML and GQL approaches appears to be highly competitive, ML approach being slightly better specially for the estimation of the dynamic dependence parameter. Remark that the aforementioned relative efficiency comparison indicates that one should be able to use either of the GQL and ML approaches for the inferences for dynamic mixed models for categorical data, binary models being special cases with two categories.

In the next section, we consider a real life data analysis illustration for a categorical data set with $C = 3$ categories collected longitudinally over a period of $T = 4$.

**Table 1** Comparison of asymptotic variances (V) of the GQL and ML estimators for the estimation of the regression parameters ($\beta_1$ and $\beta_2$), lag 1 dynamic dependence parameter ($\gamma$), and the standard deviation of random family effects ($\sigma_\tau$), of a familial-longitudinal model for binary data, with $K = 100$ families of size $n_i = 2$ each, $T = 3$ time points, and $\beta_1 = \beta_2 = 0$

| | | | $\sigma_\tau^2$ | | | | |
|---|---|---|---|---|---|---|---|
| $\gamma$ | Method | Variance | 0.5 | 0.8 | 1.2 | 1.5 | 2.0 |
| −1.0 | ML | $V(\hat{\beta}_1)$ | 0.034 | 0.037 | 0.041 | 0.043 | 0.046 |
| | | $V(\hat{\beta}_2)$ | 0.034 | 0.037 | 0.040 | 0.043 | 0.046 |
| | | $V(\hat{\gamma})$ | 0.017 | 0.012 | 0.009 | 0.008 | 0.006 |
| | | $V(\widehat{\sigma_\tau})$ | 0.046 | 0.043 | 0.046 | 0.051 | 0.061 |
| | GQL | $V(\hat{\beta}_1)$ | 0.123 | 0.123 | 0.123 | 0.123 | 0.122 |
| | | $V(\hat{\beta}_2)$ | 0.081 | 0.077 | 0.072 | 0.070 | 0.067 |
| | | $V(\hat{\gamma})$ | 0.267 | 0.248 | 0.228 | 0.214 | 0.193 |
| | | $V(\widehat{\sigma_\tau})$ | 0.103 | 0.070 | 0.053 | 0.046 | 0.039 |
| 0.5 | ML | $V(\hat{\beta}_1)$ | 0.033 | 0.037 | 0.042 | 0.044 | 0.049 |
| | | $V(\hat{\beta}_2)$ | 0.034 | 0.037 | 0.041 | 0.044 | 0.048 |
| | | $V(\hat{\gamma})$ | 0.002 | 0.002 | 0.001 | 0.001 | 0.001 |
| | | $V(\widehat{\sigma_\tau})$ | 0.024 | 0.025 | 0.029 | 0.033 | 0.039 |
| | GQL | $V(\hat{\beta}_1)$ | 0.046 | 0.047 | 0.049 | 0.050 | 0.053 |
| | | $V(\hat{\beta}_2)$ | 0.033 | 0.033 | 0.034 | 0.035 | 0.036 |
| | | $V(\hat{\gamma})$ | 0.089 | 0.083 | 0.075 | 0.069 | 0.061 |
| | | $V(\widehat{\sigma_\tau})$ | 0.079 | 0.056 | 0.043 | 0.037 | 0.031 |
| 1.0 | ML | $V(\hat{\beta}_1)$ | 0.035 | 0.039 | 0.044 | 0.047 | 0.051 |
| | | $V(\hat{\beta}_2)$ | 0.036 | 0.040 | 0.045 | 0.048 | 0.052 |
| | | $V(\hat{\gamma})$ | 0.002 | 0.002 | 0.002 | 0.001 | 0.001 |
| | | $V(\widehat{\sigma_\tau})$ | 0.040 | 0.039 | 0.043 | 0.047 | 0.056 |
| | GQL | $V(\hat{\beta}_1)$ | 0.032 | 0.033 | 0.035 | 0.037 | 0.039 |
| | | $V(\hat{\beta}_2)$ | 0.026 | 0.027 | 0.028 | 0.029 | 0.031 |
| | | $V(\hat{\gamma})$ | 0.052 | 0.049 | 0.046 | 0.043 | 0.040 |
| | | $V(\widehat{\sigma_\tau})$ | 0.061 | 0.044 | 0.035 | 0.030 | 0.026 |

**Table 2** Comparison of asymptotic variances (V) of the GQL and ML estimators for the estimation of the regression parameters ($\beta_1$ and $\beta_2$), lag 1 dynamic dependence parameter ($\gamma$), and the standard deviation of random family effects ($\sigma_\tau$), of a familial-longitudinal model for binary data, with $K = 100$ families of size $n_i = 2$ each, $T = 3$ time points, and $\beta_1 = \beta_2 = 0.5$

| | | | $\sigma_\tau^2$ | | | | |
|---|---|---|---|---|---|---|---|
| $\gamma$ | Method | Variance | 0.5 | 0.8 | 1.2 | 1.5 | 2.0 |
| −1.0 | ML | $V(\hat{\beta}_1)$ | 0.035 | 0.039 | 0.042 | 0.045 | 0.049 |
| | | $V(\hat{\beta}_2)$ | 0.035 | 0.038 | 0.041 | 0.044 | 0.047 |
| | | $V(\hat{\gamma})$ | 0.005 | 0.004 | 0.003 | 0.003 | 0.003 |
| | | $V(\hat{\sigma}_\tau)$ | 0.030 | 0.029 | 0.032 | 0.036 | 0.043 |
| | GQL | $V(\hat{\beta}_1)$ | 0.092 | 0.094 | 0.096 | 0.097 | 0.099 |
| | | $V(\hat{\beta}_2)$ | 0.071 | 0.069 | 0.067 | 0.066 | 0.063 |
| | | $V(\hat{\gamma})$ | 0.226 | 0.218 | 0.202 | 0.189 | 0.167 |
| | | $V(\hat{\sigma}_\tau)$ | 0.091 | 0.066 | 0.052 | 0.046 | 0.040 |
| 0.5 | ML | $V(\hat{\beta}_1)$ | 0.043 | 0.046 | 0.050 | 0.053 | 0.058 |
| | | $V(\hat{\beta}_2)$ | 0.043 | 0.046 | 0.050 | 0.052 | 0.056 |
| | | $V(\hat{\gamma})$ | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | $V(\hat{\sigma}_\tau)$ | 0.030 | 0.030 | 0.034 | 0.037 | 0.044 |
| | GQL | $V(\hat{\beta}_1)$ | 0.051 | 0.050 | 0.050 | 0.051 | 0.052 |
| | | $V(\hat{\beta}_2)$ | 0.055 | 0.053 | 0.052 | 0.051 | 0.050 |
| | | $V(\hat{\gamma})$ | 0.076 | 0.071 | 0.064 | 0.060 | 0.053 |
| | | $V(\hat{\sigma}_\tau)$ | 0.072 | 0.053 | 0.041 | 0.036 | 0.030 |
| 1.0 | ML | $V(\hat{\beta}_1)$ | 0.049 | 0.052 | 0.056 | 0.059 | 0.064 |
| | | $V(\hat{\beta}_2)$ | 0.050 | 0.053 | 0.057 | 0.060 | 0.064 |
| | | $V(\hat{\gamma})$ | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | $V(\hat{\sigma}_\tau)$ | 0.048 | 0.045 | 0.049 | 0.053 | 0.062 |
| | GQL | $V(\hat{\beta}_1)$ | 0.040 | 0.040 | 0.041 | 0.042 | 0.043 |
| | | $V(\hat{\beta}_2)$ | 0.050 | 0.049 | 0.048 | 0.048 | 0.047 |
| | | $V(\hat{\gamma})$ | 0.047 | 0.046 | 0.043 | 0.041 | 0.038 |
| | | $V(\hat{\sigma}_\tau)$ | 0.061 | 0.045 | 0.036 | 0.032 | 0.027 |

## 5 An Illustration: Fitting MDL Model to the TMISL Data

Consider the Three Mile Island Stress-Level (TMISL) data (Conaway 1989; 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. The study focuses on the changes in the post accident 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). The subjects were classified into one of the three response categories namely, low, medium and high stress level,

**Table 3** Comparison of asymptotic variances (V) of the GQL and ML estimators for the estimation of the regression parameters ($\beta_1$ and $\beta_2$), lag 1 dynamic dependence parameter ($\gamma$), and the standard deviation of random family effects ($\sigma_\tau$), of a familial-longitudinal model for binary data, with $K = 100$ families of size $n_i = 2$ each, $T = 3$ time points, and $\beta_1 = \beta_2 = 1.0$

| | | | $\sigma_\tau^2$ | | | | |
|---|---|---|---|---|---|---|---|
| $\gamma$ | Method | Variance | 0.5 | 0.8 | 1.2 | 1.5 | 2.0 |
| −1.0 | ML | $V(\hat{\beta}_1)$ | 0.036 | 0.040 | 0.043 | 0.046 | 0.050 |
| | | $V(\hat{\beta}_2)$ | 0.038 | 0.041 | 0.044 | 0.046 | 0.050 |
| | | $V(\hat{\gamma})$ | 0.003 | 0.003 | 0.003 | 0.002 | 0.002 |
| | | $V(\hat{\sigma}_\tau)$ | 0.029 | 0.028 | 0.031 | 0.034 | 0.041 |
| | GQL | $V(\hat{\beta}_1)$ | 0.094 | 0.096 | 0.098 | 0.098 | 0.099 |
| | | $V(\hat{\beta}_2)$ | 0.095 | 0.093 | 0.089 | 0.086 | 0.082 |
| | | $V(\hat{\gamma})$ | 0.240 | 0.234 | 0.219 | 0.205 | 0.181 |
| | | $V(\hat{\sigma}_\tau)$ | 0.090 | 0.067 | 0.054 | 0.048 | 0.042 |
| 0.5 | ML | $V(\hat{\beta}_1)$ | 0.045 | 0.048 | 0.053 | 0.056 | 0.060 |
| | | $V(\hat{\beta}_2)$ | 0.052 | 0.055 | 0.058 | 0.061 | 0.065 |
| | | $V(\hat{\gamma})$ | 0.001 | 0.001 | 0.001 | 0.001 | 0.000 |
| | | $V(\hat{\sigma}_\tau)$ | 0.033 | 0.033 | 0.036 | 0.039 | 0.046 |
| | GQL | $V(\hat{\beta}_1)$ | 0.055 | 0.055 | 0.054 | 0.054 | 0.054 |
| | | $V(\hat{\beta}_2)$ | 0.089 | 0.084 | 0.079 | 0.076 | 0.072 |
| | | $V(\hat{\gamma})$ | 0.092 | 0.087 | 0.079 | 0.073 | 0.065 |
| | | $V(\hat{\sigma}_\tau)$ | 0.079 | 0.059 | 0.046 | 0.041 | 0.034 |
| 1.0 | ML | $V(\hat{\beta}_1)$ | 0.051 | 0.054 | 0.059 | 0.062 | 0.066 |
| | | $V(\hat{\beta}_2)$ | 0.063 | 0.066 | 0.070 | 0.072 | 0.076 |
| | | $V(\hat{\gamma})$ | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 |
| | | $V(\hat{\sigma}_\tau)$ | 0.051 | 0.048 | 0.051 | 0.055 | 0.065 |
| | GQL | $V(\hat{\beta}_1)$ | 0.046 | 0.046 | 0.046 | 0.046 | 0.047 |
| | | $V(\hat{\beta}_2)$ | 0.080 | 0.077 | 0.073 | 0.071 | 0.067 |
| | | $V(\hat{\gamma})$ | 0.062 | 0.060 | 0.056 | 0.053 | 0.049 |
| | | $V(\hat{\sigma}_\tau)$ | 0.068 | 0.051 | 0.041 | 0.036 | 0.031 |

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 5–10 miles from the plant (GT5). It was of interest to compare the distribution of individuals under three stress levels collected over four different time points under both distance groups. The observed distributions of the individuals under three stress levels in each distance group at the initial point of time ($t = 1$) are shown in Table 4, and the transitional counts from time $t - 1$ to $t$ for $t = 2, 3, 4$, are shown in Tables 5, 6 and 7.

Note that when compared with familial longitudinal multinomial dynamic mixed logit (MDML) model (8)–(10), in this TMISL data set $n_i = 1$ and $i = 1, \ldots, K$,

**Table 4** Contingency table for TMISL data with $C = 3$ categories at initial time $t = 1$ for individuals under 2 distance (covariate) groups

| | t ($t = 1$) | | | |
|---|---|---|---|---|
| | Stress level (c) | | | |
| Distance covariate | Low (1) | Medium (2) | High (3) | Total subjects |
| Distance $\leq$ 5 miles | 14 | 69 | 32 | 115 |
| Distance $>$ 5 miles | 9 | 110 | 33 | 152 |
| Total | 23 | 179 | 65 | 267 |

**Table 5** Transitional counts for the TMISL data from time 1 to 2 for individuals under both distance groups

| | | Distance $\leq$ 5 miles | | | |
|---|---|---|---|---|---|
| Time | | t=2 | | | |
| | | Stress level ($c$) | | | |
| Time | Stress level ($g$) | Low (1) | Medium (2) | High (3) | Total subjects |
| $[t-1] = 1$ | Low (1) | 7 | 7 | 0 | 14 |
| | Medium (2) | 11 | 54 | 4 | 69 |
| | High (3) | 0 | 12 | 20 | 32 |
| | Total | 18 | 73 | 24 | 115 |
| | | Distance $>$ 5 miles | | | |
| Time | | t=2 | | | |
| | | Stress level ($j$) | | | |
| Time | Stress level ($g$) | Low (1) | Medium (2) | High (3) | Total subjects |
| $[t-1] = 1$ | Low (1) | 5 | 4 | 0 | 9 |
| | Medium (2) | 29 | 75 | 6 | 110 |
| | High (3) | 1 | 14 | 18 | 33 |
| | Total | 35 | 93 | 24 | 152 |

with $K = 267$. Thus, there will be no family effect $\tau_i$. Also $C = 3$ here, because of 3 stress levels: low, medium and high. Consequently, as opposed to the MDML model (8)–(10), we fit the MDL model (6)–(7) to this data set with $C = 3$. As far as the past studies are concerned, this TMISL data set was analyzed by Fienberg et al. (1985) and reanalyzed by Conaway (1989), among others. Fienberg et al. (1985) have collapsed the trinomial (3 category) data into 2 category based dichotomized data and modeled the correlations among such repeated binary responses through a binary dynamic logit model [see also Sutradhar and Farrell (2007)]. Thus their model does not deal with correlations of repeated multinomial (trinomial in this case to be specific) responses and do not make the desired comparison among three original stress levels. Conaway (1989) has however attempted to model the multinomial correlations but has used a random effects, that is, mixed model approach to compute the correlations. More specifically, Conaway (1989) fitted a modified shorter version of the MDML model (8)–(10) with $\gamma_c = 0$. Thus, this author used $\tau_i$ (individual random effect as opposed to random family effect) to

**Table 6** Transitional counts for the TMISL data from time 2 to 3 for individuals under both distance groups

| | | Distance $\leq$ 5 miles | | | |
|---|---|---|---|---|---|
| Time | | t=3 | | | |
| | | Stress level ($c$) | | | |
| Time | Stress level ($g$) | Low (1) | Medium (2) | High (3) | Total subjects |
| $[t-1]=2$ | Low (1) | 8 | 10 | 0 | 18 |
| | Medium (2) | 6 | 57 | 10 | 73 |
| | High (3) | 0 | 5 | 19 | 24 |
| | Total | 14 | 72 | 29 | 115 |
| | | Distance $>$ 5 miles | | | |
| Time | | t=3 | | | |
| | | Stress level ($c$) | | | |
| Time | Stress level ($g$) | Low (1) | Medium (2) | High (3) | Total subjects |
| $[t-1]=2$ | Low (1) | 12 | 23 | 0 | 35 |
| | Medium (2) | 5 | 82 | 6 | 93 |
| | High (3) | 0 | 11 | 13 | 24 |
| | Total | 17 | 116 | 19 | 152 |

**Table 7** Transitional counts for the TMISL data from time 3 to 4 for individuals under both distance groups

| | | Distance $\leq$ 5 miles | | | |
|---|---|---|---|---|---|
| Time | | t=4 | | | |
| | | Stress level ($c$) | | | |
| Time | Stress level ($g$) | Low (1) | Medium (2) | High (3) | Total subjects |
| $[t-1]=3$ | Low (1) | 10 | 4 | 0 | 14 |
| | Medium (2) | 8 | 57 | 7 | 72 |
| | High (3) | 0 | 9 | 20 | 29 |
| | Total | 18 | 70 | 27 | 115 |
| | | Distance $>$ 5 miles | | | |
| Time | | t=4 | | | |
| | | Stress level ($c$) | | | |
| Time | Stress level ($g$) | Low (1) | Medium (2) | High (3) | Total subjects |
| $[t-1]=3$ | Low (1) | 8 | 8 | 1 | 17 |
| | Medium (2) | 14 | 92 | 10 | 116 |
| | High (3) | 0 | 10 | 9 | 19 |
| | Total | 22 | 110 | 20 | 152 |

model the correlations of the repeated multinomial responses which is, however, appropriate only to model equi-correlations among responses.

Recently, this TMISL data was analyzed by Sutradhar (2014, Sect. 3.5.2.2) by fitting the MDL model (6)–(7), whereas in this paper we have developed the likelihood estimating equations for $\beta$ (33), $\gamma$ (42) and $\sigma_\tau$ (48) for the general MDML model (8)–(10). Now because when $\sigma_\tau = 0$, the likelihood equations for $\beta$ and $\gamma$ developed in (33) and (42), reduce to the likelihood estimating equations developed in Sutradhar (2014, Sect. 3.5.2.1) for the parameters corresponding to $\beta$ and $\gamma$, there is no need of any additional computations. For the sake of completeness, we provide the following estimates from Sutradhar (2014, Sect. 3.5.2.2). To be specific, the maximum likelihood (ML) estimates for the category intercept and regression parameters (see model (8)–(10) with $C = 3,\ p = 2$) where found to be

$$\hat{\beta}_{10,ML} = -1.6962,\ \hat{\beta}_{11,ML} = -0.3745;\ \hat{\beta}_{20,ML} = 0.6265, \hat{\beta}_{21,ML} = -0.5196,$$

and for the lag 1 dynamic dependence parameters (for $c = 1, \ldots, C-1$, with $C = 3$), the ML estimates were

$$\hat{\gamma}_{11,ML} = 5.7934,\ \hat{\gamma}_{12,ML} = 2.4007;\ \hat{\gamma}_{21,ML} = 3.6463,\ \hat{\gamma}_{22,ML} = 1.8790.$$

Next, these regression and dynamic dependence parameter estimates may be used to compute the recursive means (multinomial probabilities) over time under both distance groups following (11), that is, using the formula

$$E[Y_{ijt}|\tau_i = 0] = \tilde{\pi}_{(ijt)}(\tau_i = 0) \tag{64}$$

$$= \begin{cases} \pi^*_{(ij1)}(\tau_i = 0) & \text{for } t = 1 \\ \left[\pi^*_{(ijt)}(\tau_i) + \left\{\eta^*_{ijt|t-1,M}(\tau_i) - \pi^*_{(ijt)}(\tau_i)1'_{C-1}\right\} \tilde{\pi}_{(ij,t-1)}(\tau_i)\right]_{|\tau_i=0} & \text{for } t = 2, \ldots, T. \end{cases}$$

The fitted marginal probabilities (FMP) for all time $t = 1, \ldots, 4$, and all categories $c = 1, \ldots, 3$, under both distance groups are shown in Table 8. Note that the covariates representing two distance groups was coded as $x_{i1t} = (1, 1)'$ for the group with distance $\leq 5$ miles and $x_{i1t} = (1, 0)'$ for the group with distance $> 5$ miles.

These probabilities clearly reveal differences between the stress levels under two distance groups. For example, even though the probabilities in the high level stress group decrease as time progress under both distance groups, these probabilities remain much higher in the short distance group as expected. As far as the other two stress levels are concerned, the probabilities increase as time progress under both distance groups. Furthermore, relatively more workers appear to have medium stress

**Table 8** MDL model based fitted marginal probabilities (FMP) by using (64) under all three stress levels for the TMISL data

| | FMP at time $t$ [distance $\leq$ 5 miles] | | | | FMP at time $t$ [distance $>$ 5 miles] | | | |
|---|---|---|---|---|---|---|---|---|
| Stress level | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| Low | 0.0563 | 0.1232 | 0.1628 | 0.1826 | 0.0600 | 0.1270 | 0.1612 | 0.1757 |
| Medium | 0.4970 | 0.6253 | 0.6587 | 0.6676 | 0.6126 | 0.7252 | 0.7419 | 0.7427 |
| High | 0.4466 | 0.2515 | 0.1785 | 0.1498 | 0.3274 | 0.1478 | 0.0969 | 0.0816 |

under both distance groups, with smaller proportion in the short distance group as compared to the long distance group.

## Appendix

### *Algebras for the Likelihood Estimating Equation* (32) *for* $\boldsymbol{\beta}$

To obtain the formula for $\frac{\partial \pi^*_{(ij1)c}}{\partial \beta}$ in (31), we first compute the derivatives with respect to category based location regression parameters. Thus,

$$\begin{aligned}\frac{\partial \pi^*_{(ij1)c}}{\partial \beta^*_c} &= \pi^*_{(ij1)c}[1 - \pi^*_{(ij1)c}]x_{ij1}\\ \frac{\partial \pi^*_{(ij1)c}}{\partial \beta^*_k} &= -[\pi^*_{(ij1)c}\pi^*_{(ij1)k}]x_{ij1},\end{aligned} \tag{65}$$

yielding

$$\begin{aligned}\frac{\partial \pi^*_{(ij1)c}}{\partial \beta} &= \begin{pmatrix} -\pi^*_{(ij1)1}\pi^*_{(ij1)c} \\ \vdots \\ \pi^*_{(ij1)c}[1 - \pi^*_{(ij1)c}] \\ \vdots \\ -\pi^*_{(ij1)(C-1)}\pi^*_{(ij1)c} \end{pmatrix} \otimes x_{ij1} : (J-1)p \times 1 \\ &= \left[\pi^*_{(ij1)c}(\delta_{(ij1)c} - \pi^*_{(ij1)})\right] \otimes x_{ij1},\end{aligned} \tag{66}$$

where $\otimes$ stands for the Kronecker product, and the formulas for $\delta_{(ij1)c}$ and $\pi^*_{(ij1)}$ vectors are as in (32). Similarly, to compute $\frac{\partial \eta^{*(c)}_{ijt|t-1}(g)}{\partial \beta}$ for (31), we first obtain the local derivatives of the conditional probability function $\eta^{*(c)}_{ijt|t-1}(g)$ given in (10) for $t = 2, \ldots, T$, as

$$\begin{aligned}\frac{\partial \eta^{*(c)}_{ijt|t-1}(g)}{\partial \beta^*_c} &= \eta^{*(c)}_{ijt|t-1}(g)[1 - \eta^{*(c)}_{ijt|t-1}(g)]x_{ijt}\\ \frac{\partial \eta^{*(c)}_{ijt|t-1}(g)}{\partial \beta^*_k} &= -[\eta^{*(c)}_{ijt|t-1}(g)\eta^{*(k)}_{ijt|t-1}(g)]x_{ijt},\end{aligned} \tag{67}$$

yielding the derivative with respect to the global parameter vector $\beta$ as

$$\frac{\partial \eta^{*(c)}_{ijt|t-1}(g)}{\partial \beta} = \begin{pmatrix} -\eta^{*(1)}_{ijt|t-1}(g)\eta^{*(c)}_{ijt|t-1}(g) \\ \vdots \\ \eta^{*(c)}_{ijt|t-1}(g)[1-\eta^{*(c)}_{ijt|t-1}(g)] \\ \vdots \\ -\eta^{*(C-1)}_{ijt|t-1}(g)\eta^{*(c)}_{ijt|t-1}(g) \end{pmatrix} \otimes x_{ijt} : (J-1)p \times 1$$

$$= \left[\eta^{*(c)}_{ijt|t-1}(g)(\delta_{(ij,t-1)c} - \eta^{*}_{ijt|t-1}(g))\right] \otimes x_{ijt}, \quad (68)$$

where $\eta^{*}_{ijt|t-1}(g)$ is the $(C-1)\times 1$ vector of transitional probabilities as in (32).

## *Algebras for the Likelihood Estimating Equation* (39) *for* $\gamma \equiv \gamma_M$

To compute the derivative in (37) through (38), it is convenient to find the derivative with respect to the local dynamic dependence parameter vector $\gamma_c$ for $c = 1,\ldots,C-1$. Thus, we write

$$\frac{\partial \eta^{*(h)}_{ijt|t-1}(g)}{\partial \gamma_c} = \begin{cases} \delta_{(ij,t-1)g}\eta^{*(c)}_{ijt|t-1}(g)[1-\eta^{*(c)}_{ijt|t-1}(g)] & \text{for } h=c; h,c=1,\ldots,C-1 \\ -\delta_{(ij,t-1)g}\eta^{*(c)}_{ijt|t-1}(g)\eta^{*(h)}_{ijt|t-1}(g) & \text{for } h\neq c; h,c=1,\ldots,C-1 \\ nonumber\ -\delta_{(ij,t-1)g}\eta^{*(c)}_{ijt|t-1}(g)\eta^{*(C)}_{ijt|t-1}(g) & \text{for } h=C;\ c=1,\ldots,C-1, \end{cases} \quad (69)$$

for all $g = 1,\ldots,C$, with

$$\delta_{(ijt)g} = \begin{cases} [01'_{g-1}, 1, 01'_{C-1-g}]' & \text{for } g=1,\ldots,C-1 \\ 01_{C-1} & \text{for } g=C, \end{cases}$$

as in (32), leading to

$$\frac{\partial log\ L_i(\tau_i)}{\partial \gamma_c} \equiv \frac{\partial log\ L_i(\beta,\gamma_M,\sigma_\tau)}{\partial \gamma_c}$$

$$= \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{h=1}^{C}\sum_{g=1}^{C}\left\{\frac{y_{ijth}}{\eta^{*(h)}_{ijt|t-1}(g)}\frac{\partial \eta^{*(h)}_{ijt|t-1}(g)}{\partial \gamma}\right\}$$

$$= \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{g=1}^{C} y_{ijtc}\delta_{(ij,t-1)g}[1-\eta^{*(c)}_{ijt|t-1}(g)]$$

$$-\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{g=1}^{C}\sum_{h\neq c}^{C}\frac{y_{ijth}}{\eta^{*(h)}_{ijt|t-1}(g)}\delta_{(ij,t-1)g}\left(\eta^{*(c)}_{ijt|t-1}(g)\eta^{*(h)}_{ijt|t-1}(g)\right)$$

$$= \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{g=1}^{C} y_{ijtc}\delta_{(ij,t-1)g}$$

$$-\sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{g=1}^{C}\sum_{h=1}^{C} y_{ijth}\delta_{(ij,t-1)g}\left(\eta^{*(c)}_{ijt|t-1}(g)\right)$$

$$= \sum_{j=1}^{n_i}\sum_{t=2}^{T}\sum_{g=1}^{C} \delta_{(ij,t-1)g}\left[y_{ijtc} - \left(\eta^{*(c)}_{ijt|t-1}(g)\right)\right], \tag{70}$$

for $c = 1, \ldots, C-1$.

### *Algebras for the Likelihood Estimating Equation* (47) *for* $\sigma_\tau$

The derivative $\frac{\partial \pi^*_{(ij1)c}}{\partial \sigma_\tau}$ in (46) follows from (9) and is given by

$$\frac{\partial \pi^*_{(ij1)c}}{\partial \sigma_\tau} = \frac{\partial}{\partial \sigma_\tau}\begin{cases} \frac{\exp(x'_{ijt}\beta^*_c + \sigma_\tau\tau_i)}{1+\sum_{g=1}^{C-1}\exp(x'_{ijt}\beta^*_g+\sigma_\tau\tau_i)} & \text{for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1}\exp(x'_{ijt}\beta^*_g+\sigma_\tau\tau_i)} & \text{for } c = C, \end{cases}$$

$$= \begin{cases} \tau_i\pi^*_{(ij1)c}\pi^*_{(ij1)C} & \text{for } c = 1,\ldots,C-1 \\ -\tau_i\pi^*_{(ij1)C}[1-\pi^*_{(ij1)C}] & \text{for } h = C; \end{cases} \tag{71}$$

and similarly, the derivative $\frac{\partial \eta^{*(c)}_{ijt|t-1}(g)}{\partial \sigma_\tau}$ in (46) follows from (10), and it has the formula given by

$$\frac{\partial}{\partial \sigma_\tau}[\eta^{*(c)}_{ijt|t-1}(g)] = \frac{\partial}{\partial \sigma_\tau}\begin{cases} \frac{\exp\left[x'_{ijt}\beta^*_c+\gamma'_c y^{(g)}_{ij,t-1}+\sigma_\tau\tau_i\right]}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ijt}\beta^*_v+\gamma'_v y^{(g)}_{ij,t-1}+\sigma_\tau\tau_i\right]}, & \text{for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ijt}\beta^*_v+\gamma'_v y^{(g)}_{ij,t-1}+\sigma_\tau\tau_i\right]}, & \text{for } c = C, \end{cases}$$

$$= \begin{cases} \tau_i\eta^{*(c)}_{ijt|t-1}(g)\eta^{*(C)}_{ijt|t-1}(g) & \text{for } c = 1,\ldots,C-1 \\ -\tau_i\eta^{*(C)}_{ijt|t-1}(g)[1-\eta^{*(C)}_{ijt|t-1}(g)] & \text{for } c = C. \end{cases} \tag{72}$$

**Acknowledgements** The authors thank the audience of the symposium and a referee for their comments and suggestions.

## References

Breslow, N., Clayton, D.G.: Approximate inference in generalized linear mixed models. J. Am. Stat. Assoc. **88**, 9–25 (1993)

Chen, B., Yi, G.Y., Cook, R.: Likelihood analysis of joint marginal and conditional models for longitudinal categorical data. Can. J. Stat. **37**, 182–205 (2009)

Conaway, M.R.: Analysis of repeated categorical measurements with conditional likelihood methods. J. Am. Stat. Assoc. **84**, 53–62 (1989)

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

Fienberg, S.E., Bromet, E.J., Follmann, D., Lambert, D., May, S.M.: Longitudinal analysis of categorical epidemiological data: a study of three mile island. Environ. Health Perspect. **63**, 241–248 (1985)

Lipsitz, S.R., Kim, K., Zhao, L.: Analysis of repeated categorical data using generalized estimating equations. Stat. Med. **13**, 1149–1163 (1994)

Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. Stat. Sci. **18**, 377–393 (2003)

Sutradhar, B.C.: On exact quasilikelihood inference in generalized linear mixed models. Sankhya: Indian J. Stat. **66**, 261–289 (2004)

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

Sutradhar, B.C.: Longitudinal Categorical Data Analysis. Springer, New York (2014)

Sutradhar, B.C., Das, K.: On the efficiency of regression estimators in generalized linear models for longitudinal data. Biometrika **86**, 459–465 (1999)

Sutradhar, B.C., Farrell, P.J.: On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. Sankhya B **69**, 448–467 (2007)

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

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

Williamson, J.M., Kim, K.M., Lipsitz, S.R.: Analyzing bivariate ordinal data using a global odds ratio. J. Am. Stat. Assoc. **90**, 1432–1437 (1995)

# Dynamic Models for Longitudinal Ordinal Non-stationary Categorical Data

**Brajendra C. Sutradhar and Nairanjana Dasgupta**

**Abstract** When a nominal categorical response along with a multidimensional covariate are repeatedly collected over a small period of time from a large number of independent individuals, it is standard to use the so-called multinomial logits to model the marginal means or probabilities of such responses in terms of time dependent covariates. However, because of the difficulties in modeling the correlations of the repeated multinomial responses, some researchers over the last two decades have analyzed this type of repeated multinomial data by using certain 'working' odds ratio, or 'working' correlation structures. But, in the context of longitudinal binary data analysis, these 'working' correlations based approaches were shown to produce inefficient regression estimates as compared to simpler moment and/or quasi-likelihood approaches. The situation can be worse for longitudinal nominal multinomial/categorical data. In this paper, following the recent correlation models proposed by Sutradhar (Longitudinal Categorical Data Analysis. Springer, New York, 2014, Chap. 3) for the stationary nominal and ordinal categorical data, we discuss similar correlation models but for non-stationary ordinal categorical data. More specifically we exploit the multinomial dynamic logits (MDL) in two different ways to develop correlation models for ordinal categorical data. Under model 1, the ordinal responses are first treated to be nominal and a multinominal dynamic logits model is written for correlations between repeated nominal categorical responses. The regression and correlation parameters involved in the dynamic logits are, however, computed from an updated cumulative dynamic logits model which accommodates the actual ordinal nature of the data. Under model 2, a direct dynamic logits relationship is developed linking the cumulative multinomial responses over time. A lag 1 conditional likelihood is then exploited to estimate the desired regression and correlation parameters. Some asymptotic properties of the estimators under both models are also discussed.

B.C. Sutradhar (✉)
Department of Mathematics and Statistics, Memorial University, St. John's, NL, Canada A1C5S7
e-mail: bsutradh@mun.ca

N. Dasgupta
Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA
e-mail: dasgupta@wsu.edu



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_6

**Keywords** Binary models for cumulative MDLs • Bivariate binary models for cumulative logits • Likelihood inferences • Ordinal categories • Repeated categorical responses • Repeated ordinal categorical responses

## 1 Introduction

There are situations in practice where a univariate multinomial response, for example, (1) physician visit status of an individual such as none, low, medium, or high, may be recorded over a small number of years along with known covariates such as gender, age, education level, and chronic disease status; (2) TMISL (Three Miles Island Stress Level) data containing stress level such as low, medium or high, of an individual worker recorded in four waves of time, following the 1979 accident in Three Mile Island nuclear power plant in central Pennsylvania, along with a living distance (less than or greater than 5 miles from the plant) covariate. In these examples, it is clear that the multinomial responses are ordinal, and also it is likely that these repeated categorical responses collected from an individual over the years will be correlated. Also notice that even though the distance covariate involved in the TMISL data is time independent, there are other problems similar to the aforementioned physician visits status study where the covariates can be time dependent causing non-stationary means, variances and correlations. It is of interest to know both (1) the effects of the covariates which may be time dependent or time independent, on the responses, and (2) the dynamic relationship among the responses over the years. This type of longitudinal multinomial data, in particular the TMISL data have been analyzed by some authors treating the stress level as a nominal categorical response variable. See for example, Fienberg et al. (1985), Conaway (1989) and the recent study by Sutradhar (2014, Sect. 3.5), among others.

Fienberg et al. (1985) have collapsed the multinomial (with more than two categories) data into two category based dichotomized data and modeled the correlations among such repeated binary responses through a binary dynamic logit model [see also Sutradhar and Farrell (2007)]. Thus their model does not deal with correlations of repeated multinomial responses and do not make the desired comparison among all categories. Conaway (1989) has however attempted to model the multinomial correlations but has used a random effects, that is, a mixed model approach to compute the correlations. There are, however, at least two major drawbacks with this mixed model approach used in Conaway (1989). First, the random effects based correlations are not able to address the correlations among repeated responses. Second, it is difficult to estimate the parameters involved in a multinomial mixed model. As a remedy, Conaway (1989) has used a conditioning to remove the random effects from the model and estimated the regression parameters exploiting the so-called conditional likelihood. But a close look at this conditioning show that there was a mistake in constructing the conditional likelihood and this approach does not in fact remove the random effects. Remark that the aforementioned approaches will encounter further complexity for the repeated ordinal multinomial data. Sutradhar

(2014, Sect. 3.5) has provided a multinomial dynamic logits (MDL) model based analysis for this stationary TMISL data which accommodates the longitudinal correlations of the categorical responses, but these categorical responses were treated to be nominal similar to Fienberg et al. (1985) and Conaway (1989). An analytical extension of this MDL model to the stationary (time independent covariates based) ordinal categorical setup is also given in Sutradhar (2014, Sect. 3.6), whereas our objective in this paper is to exploit this MDL model for the non-stationary (time dependent covariates based) ordinal categorical data.

Agresti and Natarajan (2001, Sect. 2.2) [see also Agresti (1989)] described the so-called 'working' correlations based GEE (generalized estimating equation) approach to accommodate the correlations of the repeated multinomial responses. This marginal approach has some severe drawbacks. First of all, it is not known which type of 'working' correlations can be used for ordinal as opposed to nominal multinomial responses. Secondly, even if the ordinal nature is compromised, the use of their time lag free constant correlation (equi-correlations over time) is hypothetical and may cause inconsistency and lack of efficiency problems (Crowder 1995; Sutradhar and Das 1999) in desired regression estimation. Agresti and Natarajan (2001, Sect. 3) also have random effects approach to accommodate correlations among cumulative responses, but this type of individual or cluster specific random effects do not take care of longitudinal correlations. For details on this issue in nominal data analysis setup, we refer to Sutradhar (2011, Sect. 1.2, p. 6; Sect. 2.4), for example.

There exists some studies where an odds ratio approach is used to model the association of nominal categories in longitudinal setup. For example, one may refer to Lipsitz et al. (1991, 1994), Williamson et al. (1995), and Yi and Cook (2002). In this approach, the odds ratios are estimated using a 'working' log linear model. This may produce inconsistent estimates for odds ratios and subsequently regression estimates may or may not be consistent but will be inefficient. See Sutradhar (2014, Sect. 4.2.1) for some discussions on this inconsistency issue. The inference would be more complex in ordinal categorical setup. Instead of using odds ratios, following Ekholm et al. (2003) [see also Ekholm et al. (2002)], Jokinen (2006) defines dependence ratios of various order but common for all individuals to understand the time effects on the longitudinal ordinal responses. That is, the individual joint probabilities for categorical responses over selected time points are formulated in terms of common dependence ratios. In the next stage, these dependence ratios are further formulated in terms of fewer parameters mainly by using a 'weighting' technique constructed based on number of repetitions of response categories. Also, the formulations would vary depending on the situations in the field. For example, in a situation with longitudinal data, in general, some Markovian type formulas are used. This 'working weighting' technique appears to be arbitrary. Moreover, this dependence ratio approach still requires the estimation of time dependent parameters, which is computationally expensive.

In this paper, unlike the existing 'working' models, we extend the stationary covariates based parametric correlation models for repeated ordinal responses suggested by Sutradhar (2014, Sect. 3.6) to the non-stationary covariates setup. These models are similar to the correlation models considered by Sutradhar and Kovacevic

(2000), but they are simpler. To be specific, similar to Sutradhar (2014) we show two ways to develop the correlations for repeated ordinal categorical data by using the MDL model. In Sect. 2, we present the MDL model which is appropriate for repeated nominal categorical data. In Sect. 3, we introduce a cumulative multinomial dynamic logit (MDL) model, where (1) longitudinal correlations are first modeled for nominal repeated data as in Sect. 2; (2) but then apply the cumulative principle to reflect the ordinal nature of the correlated data. A cut-points based lag 1 conditional likelihood function is developed and exploited to obtain likelihood estimates for both regression and correlation parameters. In Sect. 4, we first write a binary model for the cumulative of nominal responses at a given time and use a binary dynamic logits (BDL) model to relate the cumulative of the consecutive (lag 1) responses over time. This model is also analyzed by using the well known likelihood estimation approach. The paper concludes in Sect. 5.

## 2 MDL Model for Repeated Nominal Categorical Data

Suppose that $y_{it} = (y_{it1}, \ldots, y_{itj}, \ldots, y_{it,J-1})'$ denotes the $(J-1)$-dimensional multinomial response variable for the $i$th $(i = 1, \ldots, K)$ at time $t$ $(t = 1, \ldots, T)$, and for $j = 1, \ldots, J-1$,

$$y_{it}^{(j)} = (y_{it1}^{(j)}, \ldots, y_{itj}^{(j)}, \ldots, y_{it,J-1}^{(j)})' = (01'_{j-1}, 1, 01'_{J-1-j})' \equiv \delta_{itj} \tag{1}$$

indicates that the multinomial response from the $i$th individual recorded at time t belongs to the $j$th category. For $j = J$, one writes $y_{it}^{(J)} = \delta_{itJ} = 01_{J-1}$. Note that in the non-stationary case, that is, when covariates are time dependent, one uses the time dependent marginal probabilities. Specifically, suppose that at time point $t$ $(t = 1, \ldots, T)$, $x_{it} = (x_{it1}, \ldots, x_{it\ell}, \ldots, x_{it,p+1})'$ denotes the $(p+1)$-dimensional covariate vector and $\beta_j = (\beta_{j0}, \beta_{j1}, \ldots, \beta_{jp})'$ denotes the effect of $x_{it}$ on $y_{it}^{(j)}$ for $j = 1, \ldots, J-1$, $i = 1, \ldots, K$, and all $t = 1, \ldots, T$. In such cases, the multinomial probability at time $t$, has the form

$$P[y_{it} = y_{it}^{(j)}] = \pi_{(it)j} = \begin{cases} \frac{\exp(x'_{it}\beta_j)}{1+\sum_{g=1}^{J-1}\exp(x'_{it}\beta_g)} & \text{for } j = 1, \ldots, J-1;\ t = 1, \ldots, T \\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(x'_{it}\beta_g)} & \text{for } j = J;\ t = 1, \ldots, T, \end{cases} \tag{2}$$

and the elements of $y_{it} = (y_{it1}, \ldots, y_{itj}, \ldots, y_{it,J-1})'$ at time $t$ follow the multinomial probability distribution given by

$$P[y_{it1}, \ldots, y_{itj}, \ldots, y_{it,J-1}] = \Pi_{j=1}^{J} \pi_{(it)j}^{y_{itj}}, \tag{3}$$

for all $t = 1, \ldots, T$. In (3), $y_{itJ} = 1 - \sum_{j=1}^{J-1} y_{itj}$, and $\pi_{itJ} = 1 - \sum_{j=1}^{J-1} \pi_{itj}$.

Note that the marginal multinomial probability in (2) has the multinomial logit form. Many existing studies use this multinomial logit model (2) as the marginal model at a given time $t$. As far as the correlations between repeated responses are concerned, some authors such as Agresti (1989), Lipsitz et al. (1994), Agresti and Natarajan (2001) do not model them, rather they use 'working' correlations to construct the so-called generalized estimating equations and solve them to obtain the estimates for regression parameters involved in the marginal multinomial logits model (2). These estimates however may not be reliable as they can be inefficient as compared to the 'working' independence assumption based estimates [see Sutradhar and Das (1999), Sutradhar (2011, Chap. 7) in the context of binary longitudinal data analysis]. Some other authors such as Williamson et al. (1995), and Yi and Cook (2002) use the same marginal model (2) but use odds ratios to model the association for repeated responses which are, however, estimated using arbitrary log linear type models for odds ratios. This 'working' model based approach for the odds ratios may not provide consistent estimate for the true odds ratios which subsequently may lead to convergence problems for the regression parameter estimation. See Sutradhar (2014, Sect. 4.2) for a discussion on this inconsistency issue.

Further note that unlike the 'working' correlations and odds ratios based correlation models, some authors have used a multinomial dynamic logits (MDL) model for correlations among repeated nominal multinomial responses. See, for example, Sutradhar (2014, Sect. 3.4.2). For a similar binary dynamic logit (BDL) model, one may refer to Fienberg et al. (1985), Conaway (1989), Sutradhar and Farrell (2007), for example. Following Sutradhar (2014), we now explain the MDL model below that generates auto-correlations among the repeated nominal multinominal responses. Remark that this MDL model does not yield the marginal multinomial logit model as in (2) at times $t = 2, \ldots, T$. In stead, it provides a recursive relationship among means over time, indicating that the margin mean at time $t$ say is a function of all previous means computed over time up to $t-1$.

To be specific, for $t = 2, \ldots, T$, we assume for the $i$th individual that the transitional probability from the $g$th ($g = 1, \ldots, J$) category at time $t-1$ to the $j$th category at time $t$, has the MDL form given by

$$\eta^{(j)}_{it|t-1}(g) = P\Big(Y_{it} = y^{(j)}_{it} \,\Big|\, Y_{i,t-1} = y^{(g)}_{i,t-1}\Big)$$
$$= \begin{cases} \dfrac{\exp\left[x'_{it}\beta_j + \gamma'_j y^{(g)}_{i,t-1}\right]}{1 + \sum_{v=1}^{J-1} \exp\left[x'_{it}\beta_v + \gamma'_v y^{(g)}_{i,t-1}\right]}, \text{ for } j = 1, \ldots, J-1 \\[2ex] \dfrac{1}{1 + \sum_{v=1}^{J-1} \exp\left[x'_{it}\beta_v + \gamma'_v y^{(g)}_{i,t-1}\right]}, \text{ for } j = J, \end{cases} \tag{4}$$

where $\gamma_j = (\gamma_{j1}, \ldots, \gamma_{jv}, \ldots, \gamma_{j,J-1})'$ denotes the dynamic dependence parameters. Note that unlike the stationary covariates based MDL model discussed in Sutradhar (2014, Eq. (3.381), Sect. 3.6.2.2), the MDL model in (4) involves the time dependent, i.e., non-stationary covariate vector $x_{it}$. As far as the marginal probability at time $t = 1$ is concerned, it has the same formula as in (2) for $t = 1$, that is,

$$P[y_{i1}=y_{i1}^{(j)}]=\pi_{(i1)j}=\begin{cases}\frac{\exp(x'_{i1}\beta_j)}{1+\sum_{g=1}^{J-1}\exp(x'_{i1}\beta_g)} & \text{for } j=1,\ldots,J-1\\ \frac{1}{1+\sum_{g=1}^{J-1}\exp(x'_{i1}\beta_g)} & \text{for } j=J.\end{cases} \tag{5}$$

For further notational convenience, we re-express the conditional probabilities in (4) as

$$\eta_{it|t-1}^{(j)}(g)=\begin{cases}\frac{\exp\left[x'_{it}\beta_j+\gamma'_j\delta_{i(t-1)g}\right]}{1+\sum_{v=1}^{J-1}\exp\left[x'_{it}\beta_v+\gamma'_v\delta_{i(t-1)g}\right]}, & \text{for } j=1,\ldots,J-1\\ \frac{1}{1+\sum_{v=1}^{J-1}\exp\left[x'_{it}\beta_v+\gamma'_v\delta_{i(t-1)g}\right]}, & \text{for } j=J,\end{cases} \tag{6}$$

where for $t=2,\ldots,T$, $\delta_{i(t-1)g}$, by (1), has the formula

$$\delta_{i(t-1)g}=\begin{cases}[01'_{g-1},1,01'_{J-1-g}]' & \text{for } g=1,\ldots,J-1\\ 01_{J-1} & \text{for } g=J.\end{cases}$$

Let $\beta=(\beta'_1,\ldots,\beta'_j,\ldots,\beta'_{J-1})' : (p+1)(J-1)\times 1$, and $\gamma=(\gamma'_1,\ldots,\gamma'_j,\ldots,\gamma'_{J-1})' : (J-1)^2\times 1$. These parameters are involved in the unconditional mean, variance and covariances of the responses. More specifically one may show (Loredo and Sutradhar 2012) that

$$\begin{aligned}E[Y_{it}] &= \tilde{\pi}_{(it)}(\beta,\gamma)=(\tilde{\pi}_{(it)1},\ldots,\tilde{\pi}_{(it)j},\ldots,\tilde{\pi}_{(it)(J-1)})' : (J-1)\times 1\\ &= \begin{cases}[\pi_{(i1)1},\ldots,\pi_{(i1)j},\ldots,\pi_{(i1)(J-1)}]' & \text{for } t=1\\ \eta_{(it|t-1)}(J)+\left[\eta_{(it|t-1),M}-\eta_{(it|t-1)}(J)1'_{J-1}\right]\tilde{\pi}_{i(t-1)} & \text{for } t=2,\ldots,T-1\end{cases} \qquad (7)\\ \text{var}[Y_{it}] &= \text{diag}[\tilde{\pi}_{(it)1},\ldots,\tilde{\pi}_{(it)j},\ldots,\tilde{\pi}_{(it)(J-1)}]-\tilde{\pi}_{(it)}\tilde{\pi}'_{(it)}\\ &= (\text{cov}(Y_{itj},Y_{itk}))=(\tilde{\sigma}_{i(tt)jk}),\ j,k=1,\ldots,J-1\\ &= \tilde{\Sigma}_{i(tt)}(\beta,\gamma),\ \text{for } t=1,\ldots,T \qquad (8)\\ \text{cov}[Y_{iu},Y_{it}] &= \Pi_{s=u+1}^{t}\left[\eta_{(is|s-1),M}-\eta_{(is|s-1)}(J)1'_{J-1}\right]\text{var}[Y_{iu}],\ \text{for } u<t, t=2,\ldots,T\\ &= (\text{cov}(Y_{iuj},Y_{itk}))=(\tilde{\sigma}_{i(ut)jk}),\ j,k=1,\ldots,J-1\\ &= \tilde{\Sigma}_{i(ut)}(\beta,\gamma), \qquad (9)\end{aligned}$$

where

$$\eta_{(is|s-1)}(J) = [\eta^{(1)}_{is|s-1}(J), \ldots, \eta^{(j)}_{is|s-1}(J) \ldots, \eta^{(J-1)}_{is|s-1}(J)]' = \pi_{(is)} : (J-1) \times 1$$

$$\eta_{(is|s-1),M} = \begin{pmatrix} \eta^{(1)}_{is|s-1}(1) & \cdots & \eta^{(1)}_{is|s-1}(g) & \cdots & \eta^{(1)}_{is|s-1}(J-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta^{(j)}_{is|s-1}(1) & \cdots & \eta^{(j)}_{is|s-1}(g) & \cdots & \eta^{(j)}_{is|s-1}(J-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta^{(J-1)}_{is|s-1}(1) & \cdots & \eta^{(J-1)}_{is|s-1}(g) & \cdots & \eta^{(J-1)}_{is|s-1}(J-1) \end{pmatrix} : (J-1) \times (J-1).$$

It is of importance to estimate $\beta$ and $\gamma$ parameters mainly to understand the aforementioned basic properties including the pair-wise correlations of the responses.

Note however that the longitudinal multinomial model (4)–(5) and its basic moment properties shown in (7)–(9) are derived without any order restrictions of the categories of the responses. The purpose of this paper, however, is the estimation of the parameters $\beta$ and $\gamma$ under an ordinal categorical response model. In Sect. 3, we introduce a cumulative multinomial dynamic logit (MDL) model, where (1) longitudinal correlations are first modeled for nominal repeated data; (2) and then apply the cumulative principle to reflect the ordinal nature of the correlated data. In Sect. 3.3, we demonstrate the application of the well known likelihood approach for the estimation for the parameters involved in the proposed model including the correlation index parameters.

## 3 Cumulative MDL Model for Ordinal Categorical Data

Notice from (1)–(4) that in the nominal setup, $i$th individual response in the $j$th category at time $t$ is denoted by $y^{(j)}_{it} = (01'_{j-1}, 1, 01'_{J-1-j})'$. However, when the categorical data is of ordinal nature, $i$th individual's response belong to, say, in the low group identified by categories from 1 to $j$ inclusive, or in the high group beyond category $j$. Basically, it is a binary type response, which for convenience may be defined as

$$b^{(j)}_i(t) = \begin{cases} 1 \text{ for the response of the } i\text{th individual in a category } c_{it} > j \text{ at time } t \\ 0 \text{ for the response of the } i\text{th individual in a category } c_{it} \leq j \text{ at time } t. \end{cases} \quad (10)$$

Here $c_{it} = 1, \ldots, J$, but $j$ as a cut point has the range $j = 1, \ldots, J-1$. Consequently, at $t = 1$, one needs a probability model for $b^{(j)}_i(1)$, for all $j = 1, \ldots, J-1$. Next suppose that for $t = 2, \ldots, T$, the $i$th individual provides the cut point based transitional responses $b^{(g)}_i(t-1)$ to $b^{(j)}_i(t)$, from time $t-1$ to $t$. Here $g$ also ranges from 1 to $J-1$. We now exploit the marginal probabilities from (5) for $t = 1$ to develop a marginal cumulative model for the ordinal responses $b^{(j)}_i(1)$. Similarly, we develop a lag 1 transitional cumulative model for ordinal response $b_i^{(j)}(t)$ conditional on the past ordinal response $b_i^{(g)}(t-1)$. These marginal and transitional cumulative models are presented in next two Sects. 3.1 and 3.2, respectively.

## 3.1 Marginal Cumulative Model at Time $t = 1$

In (5), we have assumed that $\pi_{(i1)c}$ is the probability that $i$th individual's response at time $t = 1$ belongs to category $c$ for $c = 1, \ldots, J$. Because in the ordinal categories setup, one observes a cut point based binary response such as in (10) by collapsing the nominal categories into 2 groups. Let $g^*$ denote these two groups. To be specific, suppose that $g^* = 1$ will indicate the lower group with $b_i^{(j)}(t) = 0$ and $g^* = 2$ will stand for the higher group. This notation $g^*$ will be used later on for an easy construction of the likelihood function. Now, for a selected cut-point $j$, we can compute the probability for the binary variable (10) as

$$P[b_i^{(j)}(1) = 0] = \sum_{v=1}^{j} \pi_{(i1)v} = F_{(i1)j}, \text{ equivalently}$$

$$P[b_i^{(j)}(1) = 1] = \sum_{v=j+1}^{J} \pi_{(i1)v} = 1 - F_{(i1)j}. \tag{11}$$

It is convenient to put the possible cut points based responses of the $i$th individual in a vector form. Let

$$B_i^*(t) = [b_i^{(1)}(t), \ldots, b_i^{(j)}(t), \ldots, b_i^{(J-1)}(t)]' \tag{12}$$

denote a vector of responses for the $i$th individual with all such possibilities. Further because in practice, the individual provides a single response at a given time $t$, assuming that the $i$th individual at time $t$ provides a cumulative response up to or beyond the category $c_{it}$ ($c_{it} = 1, \ldots, J$), one may write the cut points based observed vector as

$$b_i^*(t) = [b_i^{(1)}(t) = 1, \ldots, b_i^{(c_{it}-1)}(t) = 1, b_i^{(c_{it})}(t) = 0, \ldots, b_i^{(J-1)}(t) = 0]', \tag{13}$$

as the realization of $B_i^*(t)$ in (12).

### 3.2 Lag 1 Transitional Cumulative Model at Time $t = 2, \ldots, T$

In order to develop a transitional model, suppose that at time $t-1$, the $i$th individual provides a $c_{i,t-1}$ category based cut points response

$$b_i^*(t-1) = [b_i^{(1)}(t-1) = 1, \ldots, b_i^{(c_{i,t-1}-1)}(t-1) = 1,$$
$$b_i^{(c_{i,t-1})}(t-1) = 0, \ldots, b_i^{(J-1)}(t-1) = 0]', \tag{14}$$

whereas the similar response for time point $t$ was shown by (13). Notice that these two responses for the $i$th individual collected at consecutive time points $t-1$ and $t$ are correlated. For this reason, in general, the $j$th cut point based response $b_i^{(j)}(t)$ is likely to depend on the $g$th cut point response $b_i^{(g)}(t-1)$ at time $t-1$. hence to develop a lag 1 transitional probability model, for a selected bivariate cut point $(g, j)$; $g, j = 1, \ldots, J-1$, one needs to develop the formulas for the conditional probability

$$P[b_i^{(j)}(t) = 1 | b_i^{(g)}(t-1)] \text{ or } P[b_i^{(j)}(t) = 0 | b_i^{(g)}(t-1)], \tag{15}$$

under the assumption that the cut point $g$ at time $t-1$ is known. Notice that the bivariate cut point $(g, j)$ collapse the $(J-1)^2$ response cells into four cells. By comparing with two categories $c_{i,t-1} \equiv c_1$ (say) and $c_{it} \equiv c_2$ (say), these four cells created by the joint cut point $(g, j)$ are:

$$\begin{array}{ll} [c_1 \le g,\ c_2 \le j] & [c_1 \le g, c_2 > j] \\ [c_1 > g,\ c_2 \le j] & [c_1 > g, c_2 > j] \end{array} \tag{16}$$

where, $[c_1 \le g,\ c_2 \le j]$, for example, represents the cell with the past response at time $t-1$ in a category $\le g$ and the present response at time $t$ in a category $\le j$. By using (10) and (15), one may then write the following four conditional probabilities corresponding to the four cells in (16) as:

$$\begin{array}{ll} P[b_i^{(j)}(t) = 0 | b_i^{(g)}(t-1) = 0] & P[b_i^{(j)}(t) = 1 | b_i^{(g)}(t-1) = 0] \\ P[b_i^{(j)}(t) = 0 | b_i^{(g)}(t-1) = 1] & P[b_i^{(j)}(t) = 1 | b_i^{(g)}(t-1) = 1]. \end{array} \tag{17}$$

Now to compute the formulas for the conditional probabilities in (17), we recall from (4) or (6) that for a given response category $g$ $(g = 1, \ldots, J)$ at time $t-1$, the conditional probabilities $\eta_{it|t-1}^{(j)}(g)$ at time $t$, satisfy the condition $\sum_{j=1}^{J} \eta_{it|t-1}^{(j)}(g) = 1$. Hence, by using (16), the conditional probabilities (6) from the nominal categories setup may be collected appropriately to obtain the formulas for the conditional probabilities in (17) under the desired ordinal categories setup. We provide these formulas as follows:

$$P[b_i^{(j)}(t) = 1|b_i^{(g)}(t-1)] = \lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1)) = \lambda_{i,gj}^{(2)}(g^*), \text{ (say)}$$

$$= \begin{cases} \lambda_{i,gj}^{(2)}(g^* = 1) \text{ for } b_i^{(g)}(t-1) = 0 \\ \lambda_{i,gj}^{(2)}(g^* = 2) \text{ for } b_i^{(g)}(t-1) = 1 \end{cases} \tag{18}$$

$$= \begin{cases} \frac{1}{g}\sum_{v_1=1}^{g}\sum_{v_2=j+1}^{J}\eta_{it|t-1}^{(v_2)}(v_1) \\ \frac{1}{J-g}\sum_{v_1=g+1}^{J}\sum_{v_2=j+1}^{J}\eta_{it|t-1}^{(v_2)}(v_1), \end{cases} \tag{19}$$

where the conditional probability $\lambda_{it|t-1}^{(v_2)}(v_1)$, has the known multinomial dynamic logit (MDL) form given by (6). For convenience, following (18)–(19), we also write

$$P[b_i^{(j)}(t) = 0|b_i^{(g)}(t-1)] = 1 - \lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))$$

$$= \begin{cases} \lambda_{i,gj}^{(1)}(g^* = 1) = 1 - \lambda_{i,gj}^{(2)}(g^* = 1) \text{ for } b_i^{(g)}(t-1) = 0 \\ \lambda_{i,gj}^{(1)}(g^* = 2) = 1 - \lambda_{i,gj}^{(2)}(g^* = 2) \text{ for } b_i^{(g)}(t-1) = 1, \end{cases} \tag{20}$$

$$= \begin{cases} \frac{1}{g}\sum_{v_1=1}^{g}\left[1 - \sum_{v_2=j+1}^{J}\eta_{it|t-1}^{(v_2)}(v_1)\right] \\ \frac{1}{J-g}\sum_{v_1=g+1}^{J}\left[1 - \sum_{v_2=j+1}^{J}\eta_{it|t-1}^{(v_2)}(v_1)\right] \end{cases} \tag{21}$$

$$= \begin{cases} \frac{1}{g}\sum_{v_1=1}^{g}\sum_{v_2=1}^{j}\eta_{it|t-1}^{(v_2)}(v_1) \\ \frac{1}{J-g}\sum_{v_1=g+1}^{J}\sum_{v_2=1}^{j}\eta_{it|t-1}^{(c_v)}(v_1). \end{cases} \tag{22}$$

The lag 1 transitional models (19) and (22) for $t = 2, \ldots, T$, along with the marginal model (11) for $t = 1$, will be used in the next section to develop a suitable likelihood function.

### *3.3 Likelihood Construction and Estimation of Parameters Under Cumulative MDL Model*

Notice that the nominal categories based MDL model (4)–(5) has been transformed to the bivariate binary model represented by (11), (19) and (22). For $t = 1$, it follows from (11) that $E[b_i^{(j)}(1)] = 1 - F_{(i1)j}$ for all $j = 1, \ldots, J-1$. As far as the observed value for $b_i^{(j)}(1)$ is concerned, we find it from (13). Now because at $t = 1$, all expected functions involving $F_{(i1)j}$ contain $\beta$ only, one may write the partial likelihood for $\beta$ at $t = 1$ as

$$L_{i1}(\beta) = \Pi_{j=1}^{J-1}\left[\{1 - F_{(i1)j}\}^{b_i^{(j)}(1)}\right]\left[\{F_{(i1)j}\}^{1-b_i^{(j)}(1)}\right]$$

$$= \Pi_{j=1}^{J-1} \left[ \{ \sum_{v=j+1}^{J} \pi_{(i1)v} \}^{b_i^{(j)}(1)} \right] \left[ \{ \sum_{v=1}^{j} \pi_{(i1)v} \}^{1-b_i^{(j)}(1)} \right], \quad (23)$$

where the observed values for $\{b_i^{(j)}(1),\ j = 1, \ldots, J-1\}$ are determined as in (13). More specifically, assuming that the $i$th individual responded in category $c_{i1}$ at time $t = 1$, the values for $\{b_i^{(j)}(1)$ in (23) are given by

$$b_i^{(j)}(1) = \begin{cases} 1 \text{ for } c_{i1} > j \\ 0 \text{ for } c_{i1} \leq j. \end{cases} \quad (24)$$

Next, by (18) it follows that

$$b_i^{(j)}(t) | [b_i^{(g)}(t-1)] \sim \text{bin}(\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))), \quad (25)$$

where the formulas for

$$\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1) = 0) \equiv \lambda_{i,gj}^{(2)}(g^* = 1), \text{ and } \lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1) = 1) \equiv \lambda_{i,gj}^{(2)}(g^* = 2),$$

are given in (19). Note that these binary probabilities are functions of $\beta$ and $\gamma$, where $\gamma$ is the dynamic dependence parameter as defined in (6). For $t = 2, \ldots, T$, one may then exploit (25) and write the likelihood function for $\beta$ and $\gamma$ as

$$L_{it|t-1}(\beta, \gamma) = \Pi_{g=1}^{J-1} \Pi_{j=1}^{J-1} \left[ \{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))\}^{b_i^{(j)}(t)} \{\lambda_{i,gj}^{(1)}(b_i^{(g)}(t-1))\}^{1-b_i^{(j)}(t)} \right], \quad (26)$$

where, for $g^* = 1, 2$, the formulas for $\lambda_{i,gj}^{(1)}(b_i^{(g)}(t-1)) \equiv \lambda_{i,gj}^{(1)}(g^*)$ are given in (22), and the values for $\{b_i^{(j)}(t)\}$ assuming categories were collapsed at $c_{it}$ are given in (13).

Now by combining (23) and (26), one obtains the likelihood function for $\beta$ and $\gamma$ as

$$\begin{aligned} L(\beta, \gamma) &= \Pi_{i=1}^{K} \left[ L_{i1}(\beta) \Pi_{t=2}^{T} L_{it|t-1}(\beta, \gamma) \right] \\ &= \Pi_{i=1}^{K} \Pi_{j=1}^{J-1} \left[ \{F_{(i1)j}\}^{1-b_i^{(j)}(1)} \right] \left[ \{1 - F_{(i1)j}\}^{b_i^{(j)}(1)} \right] \quad (27) \\ &\times \Pi_{i=1}^{K} \Pi_{t=2}^{T} \Pi_{g=1}^{J-1} \Pi_{j=1}^{J-1} \left[ \{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))\}^{b_i^{(j)}(t)} \{\lambda_{i,gj}^{(1)}(b_i^{(g)}(t-1))\}^{1-b_i^{(j)}(t)} \right]. \end{aligned}$$

### 3.3.1 Likelihood Estimating Equation for $\beta$

Recall from (6) that $\beta = (\beta_1', \ldots, \beta_j', \ldots, \beta_{J-1}')'$ is a $(p+1)(J-1)$-dimensional vector with regression effects, and $\gamma = (\gamma_1', \ldots, \gamma_j', \ldots, \gamma_{J-1}')'$ is a $(J-1)^2$-dimensional vector of dynamic dependence parameters. We take a marginal approach to estimate these parameters. Thus, for known $\gamma$, say $\gamma = 0$, we will solve the likelihood estimating equation for $\beta$, and in the second step, this estimate of $\beta$ will be used in the likelihood estimating equation for $\gamma$ in order to obtain the first step estimate of $\gamma$. The estimate of $\gamma$ will next be used in the likelihood estimating equation for $\beta$ to obtain a better estimate. This cycle of iterations will continue until the convergence for both estimates.

We now develop the log-likelihood estimating equation for $\beta$. For this we take log of the likelihood function in (27) and equate the first derivative of this log likelihood with respect to $\beta$, to a zero vector as follows:

$$\frac{\partial \text{Log } L(\beta,\gamma)}{\partial \beta} = \sum_{i=1}^{K}\sum_{j=1}^{J-1}\left[\frac{\{1-b_i^{(j)}(1)\}}{F_{(i1)j}} - \frac{\{b_i^{(j)}(1)\}}{\{1-F_{(i1)j}\}}\right]\frac{\partial F_{(i1)j}}{\partial \beta}$$
$$+\sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\left[\frac{b_i^{(j)}(t)}{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))} - \frac{\{1-b_i^{(j)}(t)\}}{\{1-\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))\}}\right]\frac{\partial \lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial \beta}$$
$$= 0, \tag{28}$$

where

$$\frac{\partial F_{(i1)j}}{\partial \beta} = \sum_{v=1}^{j}\left[\pi_{(i1)v}(\delta_{(1)v}-\pi_{(i1)})\right]\otimes x_{i1}; \tag{29}$$

and

$$\frac{\partial \lambda_{i,gj}^{(2)}(g^*)}{\partial \beta} = \begin{cases} \frac{1}{g}\sum_{v_1=1}^{g}\sum_{v_2=j+1}^{J}\left[\eta_{it|t-1}^{(v_2)}(v_1)(\delta_{(t-1)v_2}-\eta_{it|t-1}(v_1))\right]\otimes x_t & \text{for } g^*=1 \\ \frac{1}{J-g}\sum_{v_1=g+1}^{J}\sum_{v_2=j+1}^{J}\left[\eta_{it|t-1}^{(v_2)}(v_1)(\delta_{(t-1)v_2}-\eta_{it|t-1}(v_1))\right]\otimes x_{it} & \text{for } g^*=2, \end{cases} \tag{30}$$

with

$$\pi_{(i1)} = [\pi_{(i1)1},\ldots,\pi_{(i1)v},\ldots,\pi_{(i1)(J-1)}]'$$
$$\delta_{(t-1)v} = \begin{cases} [01'_{v-1},1,01'_{J-1-v}]' & \text{for } v=1,\ldots,J-1 \\ 01_{J-1} & \text{for } v=J, \end{cases}$$
$$\eta_{it|t-1}(v_1) = [\eta_{it|t-1}^{(1)}(v_1),\ldots,\eta_{it|t-1}^{(v_2)}(v_1),\ldots,\eta_{it|t-1}^{(J-1)}(v_1)]'. \tag{31}$$

Now, for given $\gamma$, the likelihood equations in (28) may be solved iteratively by using the iterative equation for $\beta$ given by

$$\hat{\beta}(r+1) = \hat{\beta}(r) - \left[\left\{\frac{\partial^2 \text{Log } L(\beta,\gamma)}{\partial \beta' \partial \beta}\right\}^{-1}\frac{\partial \text{Log } L(\beta,\gamma)}{\partial \beta}\right]_{|\beta=\hat{\beta}(r)};\ (J-1)(p+1)\times 1, \tag{32}$$

where the formula for the second order derivative matrix $\frac{\partial^2 Log\ L(\beta,\gamma}{\partial\beta'\partial\beta}$ may be derived by taking the derivative of the $(J-1)(p+1)\times 1$ vector with respect to $\beta'$. The exact second order derivative matrix has a complicated formula. We provide an approximation as follows.

**An approximation to** $\frac{\partial^2 Log\ L(\beta,\gamma)}{\partial\beta'\partial\beta}$ **:**

Re-express the likelihood estimating equation from (28) as

$$\frac{\partial \text{Log } L(\beta,\gamma)}{\partial\beta} = \sum_{i=1}^{K}\sum_{j=1}^{J-1}\frac{\partial F_{(i1)j}}{\partial\beta}\{(1-F_{(i1)j})F_{(i1)j}\}^{-1}\left[\{1-b_i^{(j)}(1)\}-F_{(i1)j}\right]$$
$$+\sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\frac{\partial\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial\beta}\{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))(1-\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1)))\}^{-1}$$
$$\times\left[b_i^{(j)}(t)-\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))\right]$$
$$=0. \tag{33}$$

Notice that in the first term in the left hand side of (33), $\{1-b_i^{(j)}(1)\}$ is, by (11), a binary variable with

$$E\{1-b_i^{(j)}(1)\} = F_{(i1)j}$$
$$\text{var}\{1-b_i^{(j)}(1)\} = F_{(i1)j}\{1-F_{(i1)j}\}, \tag{34}$$

and similarly in the second term, by (25), $b_i^{(j)}(t)$ conditional on $b_i^{(g)}(t-1)$ is a binary variable with

$$E[b_i^{(j)}(t)|b_i^{(g)}(t-1)] = \lambda_{i,gj}^{(2)}(g^*)$$
$$\text{var}[b_i^{(j)}(t)|b_i^{(g)}(t-1)] = \lambda_{i,gj}^{(2)}(g^*)[1-\lambda_{i,gj}^{(2)}(g^*)], \tag{35}$$

with $g^* = 1,2$ for $b_i^{(g)}(t-1) = 0,1$, respectively. Thus, the likelihood estimating function in the left hand side of (33) is equivalent to a conditional quasi-likelihood (CQL) function in $\beta$ for the cut points based binary data [e.g. see Tagore and Sutradhar (2009, Eq. (27), p. 888)]. Now because the variance of the binary data is a function of the mean, the variance and gradient functions in the left hand side of (33) may be treated to be known when mean is known. Thus, when a QL estimating equation is solved iteratively, the gradient and variance functions use $\beta$ from a previous iteration (Wedderburn 1974; McCullagh 1983). Consequently, by (33), the second derivative matrix required to compute (32) has a simpler approximate formula

$$\frac{\partial^2 \text{Log } L(\beta,\gamma)}{\partial\beta\partial\beta'} = -\sum_{i=1}^{K}\sum_{j=1}^{J-1}\frac{\partial F_{(i1)j}}{\partial\beta}\{(1-F_{(i1)j})F_{(i1)j}\}^{-1}\frac{\partial F_{(i1)j}}{\partial\beta'}$$
$$-\sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\frac{\partial\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial\beta}\{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))(1-\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1)))\}^{-1}$$
$$\times\frac{\partial\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial\beta'}. \tag{36}$$

### 3.3.2 Likelihood Estimating Equation for $\gamma$

Similar to the estimating equation for $\beta$ as in (28), the likelihood function (27) provides the likelihood estimating equation for $\gamma$ as

$$\frac{\partial \text{Log } L(\beta,\gamma)}{\partial\gamma} = \sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\left[\frac{b_i^{(j)}(t)}{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))} - \frac{\{1-b_i^{(j)}(t)\}}{\{1-\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))\}}\right]$$
$$\times\frac{\partial\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial\gamma} = 0, \tag{37}$$

where for $b_i^{(g)}(t-1) = 0, 1$, that is, $g^* = 1, 2$,

$$\frac{\partial\lambda_{i,gj}^{(2)}(g^*)}{\partial\gamma}$$
$$= \begin{cases} \frac{1}{g}\sum_{v_1=1}^{g}\sum_{v_2=j+1}^{J}\left[\eta_{it|t-1}^{(v_2)}(v_1)(\delta_{(t-1)v_2}-\eta_{it|t-1}(v_1))\right]\otimes\delta_{(t-1)v_1} & \text{for } g^*=1 \\ \frac{1}{J-g}\sum_{v_1=g+1}^{J}\sum_{v_2=j+1}^{J}\left[\eta_{it|t-1}^{(v_2)}(v_1)(\delta_{(t-1)v_2}-\eta_{it|t-1}(v_1))\right]\otimes\delta_{(t-1)v_1} & \text{for } g^*=2. \end{cases} \tag{38}$$

By similar formula as in (32), one may solve the likelihood estimating equation in (37) for $\gamma$ using the iterative equation

$$\hat{\gamma}(r+1) = \hat{\gamma}(r) - \left[\left\{\frac{\partial^2 \text{Log } L(\beta,\gamma)}{\partial\gamma\partial\gamma'}\right\}^{-1}\frac{\partial \text{Log } L(\beta,\gamma)}{\partial\gamma}\right]_{|\gamma=\hat{\gamma}(r)}; \ (J-1)^2\times 1, \tag{39}$$

where the second order derivative matrix with respect to $\gamma$, by using the same argument for the derivation of (36) for $\beta$, has a simple approximate formula given by

$$\frac{\partial^2 \text{Log } L(\beta,\gamma)}{\partial\gamma\partial\gamma'}$$
$$= -\sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\frac{\partial\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial\gamma}\{\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))(1-\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1)))\}^{-1}$$
$$\times\frac{\partial\lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))}{\partial\gamma'}, \quad (40)$$

where $\frac{\partial\lambda_{i,gj}^{(2)}(g^*)}{\partial\gamma}$ is given in (38).

### 3.4 Estimation of the Basic Properties of the Model

Note that because in the longitudinal ordinal categorical setup, the longitudinal data from each individual is available in the cumulative form as in (13), the $\beta$ and $\gamma$ parameters involved in the marginal probability (11) and conditional probabilities (19), were estimated accordingly by solving the likelihood estimating equations (28) and (37), respectively. These estimates may now be used to understand the mean structure of the data, by computing

$$E[B_i^*(t)] = E[b_i^{(1)}(t),\ldots,b_i^{(j)}(t),\ldots,b_i^{(J-1)}(t)]' \quad (41)$$

for all $t=1,\ldots,T$, where $B_i^*(t)$ is defined in (12). For $t=1$, one may compute $E[b_i^{(j)}(1)]$ easily for all $j=1,\ldots,J-1$, by using $\hat{\beta}$ in (11). For other $t$, the formula for the expectation can be computed in a standard way. For example, for $t=2$, by using (25), one writes

$$\begin{aligned}
E[b_i^{(j)}(2)] &= \sum_{g=1}^{J-1} E_{t=1}E_{t=2|t-1}[b_i^{(j)}(2)|b_i^{(g)}(1)] \\
&= \sum_{g=1}^{J-1} E_{t=1}[\lambda_{i,gj}^{(2)}(b_i^{(g)}(1))], \text{ by (25)} \\
&= \sum_{g=1}^{J-1}\sum_{b_i^{(g)}(1)=0}^{1} \lambda_{i,gj}^{(2)}(b_i^{(g)}(1))[1-F_{(i1)g}]^{b_i^{(g)}(1)}[F_{(i1)g}]^{1-b_i^{(g)}(1)} \quad (42)\\
&= \sum_{g=1}^{J-1}\Big[\{\lambda_{i,gj}^{(2)}(g^*=1)[F_{(i1)g}]\}+\{\lambda_{i,gj}^{(2)}(g^*=2)(1-F_{(i1)g})\}\Big]
\end{aligned}$$

$$= \sum_{g=1}^{J-1} \left[ \lambda_{i,gj}^{(2)}(g^* = 2) + F_{(i1)g}\{\lambda_{i,gj}^{(2)}(g^* = 1) - \lambda_{i,gj}^{(2)}(g^* = 1)\} \right]$$

$$= \sum_{g=1}^{J-1} \left[ \{\frac{1}{g} \sum_{v_1=1}^{g} \sum_{v_2=j+1}^{J} \eta_{i2|1}^{(v_2)}(v_1)\} \sum_{v=1}^{g} \pi_{(i1)v} \right]$$

$$+ \sum_{g=1}^{J-1} \left[ \{\frac{1}{J-g} \sum_{v_1=g+1}^{J} \sum_{v_2=j+1}^{J} \eta_{i2|1}^{(v_2)}(v_1)\} \sum_{v=g+1}^{J} \pi_{(i1)v} \right], \tag{43}$$

where, by (5),

$$\pi_{(i1)v} = \begin{cases} \frac{\exp(x_{i1}'\beta_v)}{1+\sum_{m=1}^{J-1}\exp(x_{i1}'\beta_m)} & \text{for } j = 1, \ldots, J-1 \\ \frac{1}{1+\sum_{m=1}^{J-1}\exp(x_{i1}'\beta_m)} & \text{for } j = J, \end{cases}$$

and by (6),

$$\eta_{i2|1}^{(v_2)}(v_1) = \begin{cases} \frac{\exp\left[x_{i2}'\beta_{v_2} + \gamma_{v_2}'\delta_{i(1)v_1}\right]}{1+\sum_{v=1}^{J-1}\exp\left[x_{i2}'\beta_v + \gamma_v'\delta_{i(1)v_1}\right]}, & \text{for } j = 1, \ldots, J-1 \\ \frac{1}{1+\sum_{v=1}^{J-1}\exp\left[x_{i2}'\beta_v + \gamma_v'\delta_{i(1)v_1}\right]}, & \text{for } j = J. \end{cases}$$

This marginal mean in (43) for $t = 2$ may now be estimated by replacing $\beta$ and $\gamma$, with their likelihood estimates $\hat{\beta}$ and $\hat{\gamma}$, respectively.

Remark that the likelihood estimates for the parameters $\beta$ and $\gamma$ may also be used in the basic properties of the repeated nominal categorical data shown in (7)–(9). However, the interpretation of such results are valid under the assumption that the even though data were ordinal, responses from an individual are treated to be in the cut point category only. But if the data were truly nominal, one could easily obtain the likelihood estimates for the parameters involved in the dynamic model (5)–(6) for these repeated nominal categorical data. For details on such parameter estimation under the non-stationary MDL (multinomial dynamic logit) model, one may refer to Sutradhar (2014, Sect. 4.4).

## 3.5 Asymptotic Properties of the Regression Estimates

Because the regression parameter $\beta$ is of main interest, in this section, we study the asymptotic properties of the likelihood estimate for this parameter obtained by solving the likelihood estimating equation (28). We treat the dynamic dependence parameter $\gamma$ to be known for the purpose. The asymptotic results for $\beta$ estimate however will still be valid if a consistent estimate is used for $\gamma$.

Turning back to the likelihood estimating equation (33) for $\beta$, it is convenient to re-express this equation as a function of suitable vectors as follows. For the first term in (28) corresponding to time $t = 1$, construct a $(J-1) \times 1$ vector

$$\begin{aligned} d_i(1) &= [\{1 - b_i^{(1)}(1)\} - F_{(i1)1}, \ldots, \{1 - b_i^{(j)}(1)\} - F_{(i1)j}, \ldots, \{1 - b_i^{(J-1)}(1)\} - F_{(i1)(J-1)}]' \\ &= [\tilde{b}_i^{(1)}(1) - F_{(i1)1}, \ldots, \tilde{b}_i^{(j)}(1) - F_{(i1)j}, \ldots, \tilde{b}_i^{(J-1)}(1) - F_{(i1)(J-1)} \\ &= [\tilde{b}_i^{*} - F_{(i1)}^{*}], \end{aligned} \tag{44}$$

and express the first term in (33) without the summation over $i$, as

$$\begin{aligned} &\sum_{j=1}^{J-1} \frac{\partial F_{(i1)j}}{\partial \beta} \{(1 - F_{(i1)j}) F_{(i1)j}\}^{-1} \left[\{1 - b_i^{(j)}(1)\} - F_{(i1)j}\right] \\ &= \frac{\partial F'_{(i1)}}{\partial \beta} P_i^{-1}(1) d_i(1) : \ (p+1)(J-1) \times 1, \end{aligned} \tag{45}$$

where

$$\begin{aligned} F'_{(i1)} &= [F_{(i1)1}, \ldots, F_{(i1)j}, \ldots, F_{(i1)(J-1)}] : 1 \times (J-1) \\ P_i(1) &= \text{diag}\left[\{(1 - F_{(i1)1}) F_{(i1)1}\}, \ldots, \{(1 - F_{(i1)j}) F_{(i1)j}\}, \ldots, \right. \\ &\left. \{(1 - F_{(i1)(J-1)}) F_{(i1)(J-1)}\}\right] : (J-1) \times (J-1), \end{aligned}$$

and $d_i(1)$ is the $(J-1) \times 1$ distance vector defined in (44).

In order to express the second term in (33) in matrix vector notation, we first write $d^*_{i,gj}(t) = \left[b_i^{(j)}(t) - \lambda_{i,gj}^{(2)}(b_i^{(g)}(t-1))\right]$ and define a $(J-1)^2 \times 1$ vector at time $t = 2, \ldots, T$, given by

$$\begin{aligned} d_i^*(t) &= \left[d^*_{i,11}(t), d^*_{i,21}(t), \ldots, d^*_{i,(J-1)1}(t), \ldots, \right. \\ &\quad d^*_{i,1j}(t), d^*_{i,2j}(t), \ldots, d^*_{i,(J-1)j}(t), \ldots, \\ &\quad \left. d^*_{i,1(J-1)}(t), d^*_{i,2(J-1)}(t), \ldots, d^*_{i,(J-1)(J-1)}(t)\right]' \\ &= [b_i^*(t) - \lambda_i^{(2)*}(t-1)]. \end{aligned} \tag{46}$$

Next, for each $t = 2, \ldots, T$, write a $(J-1)^2 \times (J-1)^2$ diagonal matrix as

$$Q_i(t) = \text{diag}\left[\{\lambda_{i,11}^{(2)}(b_i^{(1)}(t-1))(1 - \lambda_{i,11}^{(2)}(b_i^{(1)}(t-1)))\}, \ldots, \right.$$

$$\{\lambda^{(2)}_{i,(J-1)1}(b_i^{(J-1)}(t-1))(1-\lambda^{(2)}_{i,(J-1)1}(b_i^{(J-1)}(t-1)))\},\ldots,$$
$$\{\lambda^{(2)}_{i,1j}(b_i^{(1)}(t-1))(1-\lambda^{(2)}_{i,1j}(b_i^{(1)}(t-1)))\},\ldots,$$
$$\{\lambda^{(2)}_{i,(J-1)j}(b_i^{(J-1)}(t-1))(1-\lambda^{(2)}_{i,(J-1)j}(b_i^{(J-1)}(t-1)))\},\ldots,$$
$$\{\lambda^{(2)}_{i,1(J-1)}(b_i^{(1)}(t-1))(1-\lambda^{(2)}_{i,1(J-1)}(b_i^{(1)}(t-1)))\},\ldots,$$
$$\{\lambda^{(2)}_{i,(J-1)(J-1)}(b_i^{(J-1)}(t-1))(1-\lambda^{(2)}_{i,(J-1)(J-1)}(b_i^{(J-1)}(t-1)))\}\Big], \tag{47}$$

and define a $1 \times (J-1)^2$ conditional mean parameter vector as

$$\begin{aligned}\Lambda_i'(t) = \Big[&\lambda^{(2)}_{i,11}(b_i^{(1)}(t-1)),\ldots,\lambda^{(2)}_{i,(J-1)1}(b_i^{(J-1)}(t-1)),\ldots,\\ &\lambda^{(2)}_{i,1j}(b_i^{(1)}(t-1)),\ldots,\lambda^{(2)}_{i,(J-1)j}(b_i^{(J-1)}(t-1)),\ldots,\\ &\lambda^{(2)}_{i,1(J-1)}(b_i^{(1)}(t-1)),\ldots,\lambda^{(2)}_{i,(J-1)(J-1)}(b_i^{(J-1)}(t-1))\Big].\end{aligned} \tag{48}$$

It then follows that the second term in (33) without the sum over $i$ can be expressed as

$$\begin{aligned}&\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\frac{\partial\lambda^{(2)}_{i,gj}(b_i^{(g)}(t-1))}{\partial\beta}\{\lambda^{(2)}_{i,gj}(b_i^{(g)}(t-1))(1-\lambda^{(2)}_{i,gj}(b_i^{(g)}(t-1)))\}^{-1}\\ &\times\Big[b_i^{(j)}(t)-\lambda^{(2)}_{i,gj}(b_i^{(g)}(t-1))\Big]\\ &=\sum_{t=2}^{T}\frac{\partial\Lambda_i'(t)}{\partial\beta}Q_i^{-1}(t)d_i^*(t).\end{aligned} \tag{49}$$

Furthermore, using the notation

$$\begin{aligned}\Lambda_i' &= [\Lambda_i'(2),\ldots,\Lambda_i'(t),\ldots,\Lambda_i'(T)] : 1\times(T-1)(J-1)^2\\ Q_i &= \begin{pmatrix} Q_i(2) & 0 & \ldots & 0\\ 0 & Q_i(3) & \ldots & 0\\ \vdots & \vdots & \vdots & \vdots\\ 0 & 0 & \ldots & Q_i(T)\end{pmatrix} : (T-1)(J-1)^2\times(T-1)(J-1)^2\\ d_i^* &= [{d^*}'_i(2),\ldots,{d^*}'_i(t),\ldots,{d^*}'_i(T)]' : (T-1)(J-1)^2\times 1,\end{aligned} \tag{50}$$

the quantity in (49) may be expressed as

$$\sum_{t=2}^{T} \frac{\partial \Lambda_i'(t)}{\partial \beta} Q_i^{-1}(t) d_i^*(t) = \frac{\partial \Lambda_i'}{\partial \beta} Q_i^{-1} d_i^*, \tag{51}$$

where $d_i^*$ can also be written as $d_i^* = [b_i^* - \Lambda_i]$. Consequently, using (45), (49), and (51) into the second term in (33), the likelihood estimating function in $\beta$ can be re-expressed as

$$\begin{aligned} f(\beta,\gamma) &= \sum_{i=1}^{K} \frac{\partial F_{(i1)}'}{\partial \beta} P_i^{-1}(1) d_i(1) + \sum_{i=1}^{K} \frac{\partial \Lambda_i'}{\partial \beta} Q_i^{-1} d_i^* \\ &= \sum_{i=1}^{K} \frac{\partial F_{(i1)}'}{\partial \beta} P_i^{-1}(1) \{\tilde{b}_i^* - F_{(i1)}^*\} + \sum_{i=1}^{K} \frac{\partial \Lambda_i'}{\partial \beta} Q_i^{-1} \{b_i^* - \Lambda_i\}, \text{ by (44)–(45) and (51)} \\ &= f_1(\beta) + f_2(\beta,\gamma), \text{ (say).} \end{aligned} \tag{52}$$

For true $\beta$ and given $\gamma$, define

$$\begin{aligned} \bar{f}_K(\beta) &= \frac{1}{K} \sum_{i=1}^{K} [f_{1i}(\beta) + f_{2i}(\beta,\gamma)] \\ &= \frac{1}{K} \sum_{i=1}^{K} \left[ \frac{\partial F_{(i1)}'}{\partial \beta} P_i^{-1}(1) \{\tilde{b}_i^* - F_{(i1)}^*\} + \frac{\partial \Lambda_i'}{\partial \beta} Q_i^{-1} \{b_i^* - \Lambda_i\} \right], \end{aligned} \tag{53}$$

where $\tilde{b}_1^*, \ldots, \tilde{b}_i^*, \ldots, \tilde{b}_K^*$ are independent to each other, and similarly $b_1^*, \ldots, b_i^*, \ldots, b_K^*$, are also independent to each other, as they are collected from $K$ independent individuals, but they are not identically distributed because

$$\tilde{b}_i^* \sim [F_{(i1)}^*, P_i], \text{ and } b_i^* \sim [\Lambda_i, Q_i] \tag{54}$$

where the mean vectors and covariance matrices are different for different individuals. By (54), it follows from (53) that

$$\begin{aligned} E[\bar{f}_K(\beta)] &= 0 \\ \text{cov}[\bar{f}_K(\beta)] &= \frac{1}{K^2} \sum_{i=1}^{K} \text{cov}[f_{1i}(\beta) + f_{2i}(\beta,\gamma)] \\ &= \frac{1}{K^2} \sum_{i=1}^{K} \left[ \frac{\partial F_{(i1)}'}{\partial \beta} P_i^{-1}(1) \frac{\partial F_{(i1)}}{\partial \beta'} + \frac{\partial \Lambda_i'}{\partial \beta} Q_i^{-1} \frac{\partial \Lambda_i}{\partial \beta'} \right] \end{aligned}$$

$$= \frac{1}{K^2} \sum_{i=1}^{K} \tilde{V}_i(\beta, \gamma)$$

$$= \frac{1}{K^2} V_K^*(\beta, \gamma). \quad (55)$$

Next if the multivariate version of Lindeberg's condition holds, that is,

$$\lim_{K \to \infty} {V^*}_K^{-1} \sum_{i=1}^{K} \sum_{\{(f_{1i}+f_{2i})' {V^*}_K^{-1} (f_{1i}+f_{2i})\} > \epsilon} \{(f_{1i}+f_{2i})(f_{1i}+f_{2i})' g\{(f_{1i}+f_{2i})\} = 0 \quad (56)$$

for all $\epsilon > 0$, $g(\cdot)$ being the probability distribution of $(f_{1i} + f_{2i})$, then Lindeberg-Feller central limit theorem (Amemiya 1985, Theorem 3.3.6; McDonald 2005, Theorem 2.2) imply that

$$Z_K = K[V_K^*]^{-\frac{1}{2}} \bar{f}_K(\beta) \to N_{(p+1)(J-1)}(0, I_{(p+1)(J-1)}). \quad (57)$$

Now because $\hat{\beta}_{ML}$ obtained by (32) is a solution of $f(\beta, \gamma) = 0$ in (52) [see also (33)], one writes

$$\sum_{i=1}^{K} [f_{1i}(\hat{\beta}_{ML}) + f_{2i}(\hat{\beta}_{ML})] = 0, \quad (58)$$

which by first order Taylor's series expansion produces

$$\sum_{i=1}^{K} \{f_{1i}(\beta) + f_{2i}(\beta)\} + (\hat{\beta}_{ML} - \beta) \sum_{i=1}^{K} \{f_{1i}(\beta) + f_{2i}(\beta)\}' = 0. \quad (59)$$

That is,

$$\hat{\beta}_{ML} - \beta = - \left[ \sum_{i=1}^{K} \{f_{1i}(\beta) + f_{2i}(\beta)\}' \right]^{-1} \sum_{i=1}^{K} \{f_{1i}(\beta) + f_{2i}(\beta)\}$$

$$= - \left[ - \sum_{i=1}^{K} \{ \frac{\partial F'_{(i1)}}{\partial \beta} P_i^{-1}(1) \frac{\partial F_{(i1)}}{\partial \beta'} + \frac{\partial \Lambda'_i}{\partial \beta} Q_i^{-1} \frac{\partial \Lambda_i}{\partial \beta'} \} \right]^{-1} \sum_{i=1}^{K} [f_{1i}(\beta) + f_{2i}(\beta)]$$

$$= \left[ V_K^*(\beta, \gamma) \right]^{-1} K \bar{f}(\beta)$$

$$= \left[ V_K^*(\beta, \gamma) \right]^{-\frac{1}{2}} Z_K \to N(0, {V^*}_K^{-1}(\beta, \gamma)), \quad (60)$$

by (57). It then follows that

$$\lim_{K \to \infty} \hat{\beta}_{ML} \to N(\beta, V^{*-1}_{K}(\beta, \gamma)). \tag{61}$$

Also it follows that

$$||[V_K^{*}(\beta, \gamma)]^{\frac{1}{2}}[\hat{\beta}_{ML} - \beta]|| = O_p(\sqrt{(p+1)(J-1)}). \tag{62}$$

## 4 BDL Model for Repeated Ordinal Responses

In Sect. 3, we have introduced a MDL (multinomial dynamic logits) model for repeated nominal categorical data which was mapped to a bivariate binary (BB) model to reflect the ordinal nature of the categorical responses. In this section, we use a direct approach where, to reflect the ordinal nature of the categories, a cut point based binary response is collected at a given time $t$ for all $t = 1, \ldots, T$. These repeated binary responses are then modeled through a BDL (binary dynamic logits) model. When the associated covariates are time dependent, the BDL model for the responses is referred to as the non-stationary BDL model, which, may be treated as a generalization of the stationary BDL model used by Sutradhar (2014, Sect. 3.6.2).

For the purpose, we consider the same binary variable as in (10). That is, for all $t(t = 1, \ldots, T)$, the binary variable is defined as

$$b_i^{(j)}(t) = \begin{cases} 1 \text{ for } i\text{th individual's response beyond category } j \\ 0 \text{ for } i\text{th individual's response belong to a category } \le j. \end{cases}$$

Under the proposed over all BDL model, we need a marginal model for binary probabilities $P[b_i^j(1) = 1]$ at time $t = 1$, and a conditional probability model to compute the conditional binary probabilities, namely $P[b_i^{(j)}(t) = 1|b_i^{(g)}(t-1)]$ for $t = 2, \ldots, T$. Recall that under the cumulative model proposed in the last section, these marginal and conditional probabilities were modeled using cumulative of marginal and conditional probabilities for nominal categories as in (11) and (19), respectively. As opposed to these cumulative models, we now define them using binary logits directly. To be specific, for $t = 1$, we model the binary probability as

$$P[b_i^j(1) = 1] = 1 - F_{(i1)j} = \frac{\exp(\alpha_{j0} + x'_{i1}\tilde{\alpha}_j)}{1 + \exp(\alpha_{j0} + x'_{i1}\tilde{\alpha}_j)}. \tag{63}$$

For example, suppose that $x'_{i1} = [x_{i11}, x_{i12}, x_{i13}]$, where $x_{i11}$ represents gender (male (1) or female (0)); $x_{i12}$ represents the education with two levels (high (1) or low (0)) for the $i$th individual at time $t = 1$; and $x_{i13}$ represents the gender and education level interaction. Then $\alpha'_j = [\alpha_{j1}, \alpha_{j2}, \alpha_{j3}]$ indicates the effect of $x_{i1}$ for putting the $i$th individual's response beyond category $j$.

Next for $t = 2, \ldots, T$, we suppose that the $i$th individual provided a cumulative response below or beyond category $g$, at time $t-1$. Notice that $g$ ranges from 1 to $J-1$. At time $t$, the response can be below or beyond category $j$, for all $j = 1, \ldots, J-1$. Suppose that the conditional probability for the binary response at time $t$, conditional on the binary response from time $t-1$, has a binary dynamic logit (BDL) form

$$\begin{aligned} P[b_i^{(j)}(t) = 1 | b_i^{(g)}(t-1)] &= \frac{\exp(\alpha_{j0} + x'_{it}\tilde{\alpha}_j + \tilde{\gamma}_{gj} b_i^g(t-1))}{1 + \exp(\alpha_{j0} + x'_{it}\tilde{\alpha}_j + \tilde{\gamma}_{gj} b_i^g(t-1))} \\ &= \tilde{\eta}_{i,gj}^{(2)}(b_i^{(g)}(t-1)), \end{aligned} \tag{64}$$

(Sutradhar 2014, Sect. 3.6.2), where $\tilde{\gamma}_{gj}$ is the dynamic dependence parameter relating $b_i^{(j)}(t)$ and $b_i^{(g)}(t-1)$. This type of BDL function was earlier used by Sutradhar and Farrell (2007), for example, in the context of a correlation model for standard longitudinal binary responses, whereas the present BDL function (64) relates two consecutive cumulative binary responses defined collapsing ordinal categories into two groups at a selected cut point. Following (64), we will write the complimentary probabilities as

$$\begin{aligned} P[b_i^{(j)}(t) = 0 | b_i^{(g)}(t-1)] &= 1 - \tilde{\eta}_{i,gj}^{(2)}(b_i^{(g)}(t-1)) \\ &= \tilde{\eta}_{i,gj}^{(1)}(b_i^{(g)}(t-1)), \end{aligned} \tag{65}$$

where $b_i^{(g)}(t-1) \equiv 1$ or 0.

Furthermore, for notational convenience, we re-express the marginal probabilities in (63) and the conditional probabilities in (64) as

$$P[b_i^{(j)}(1) = 1] = 1 - F_{(i1)j} = \frac{\exp(\tilde{x}'_{i1}\alpha_j)}{1 + \exp(\tilde{x}'_{i1}\alpha_j)}, \tag{66}$$

and

$$\begin{aligned} P[b_i^{(j)}(t) = 1 | b_i^{(g)}(t-1)] &= \frac{\exp(\tilde{x}'_{it}\alpha_j + \tilde{\gamma}_{gj} b_i^g(t-1))}{1 + \exp(\tilde{x}'_{it}\alpha_j + \tilde{\gamma}_{gj} b_i^g(t-1))} \\ &= \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1)), \end{aligned} \tag{67}$$

respectively, where

$$\tilde{x}_{it} = [1, x'_{it}]', \text{ and } \alpha_j = [\alpha_{j0}, \tilde{\alpha}'_j]'. \tag{68}$$

## 4.1 *Basic Properties of the BDL Model*

### 4.1.1 Marginal Means and Variances

It is clear from (66) that at time $t = 1$, the marginal mean and variance are given by

$$\mu_{i1}^{(j)} = E[b_i^{(j)}(1)] = 1 - F_{(i1)j} = \frac{\exp(\tilde{x}_{i1}'\alpha_j)}{1 + \exp(\tilde{x}_{i1}'\alpha_j)}$$
$$\text{var}[b_i^{(j)}(1)] = \mu_{i1}^{(j)}\{1 - \mu_{i1}^{(j)}\} = F_{(i1)j}\{1 - F_{(i1)j}\}. \quad (69)$$

As $t$ increases, computation for the marginal means and variances gets complicated. For example, for $t = 2$, the marginal may be computed, following (43), as

$$\begin{aligned} E[b_i^{(j)}(2)] &= \sum_{g=1}^{J-1} E_{t=1}E_{t=2|t-1}[b_i^{(j)}(2)|b_i^{(g)}(1)] \\ &= \sum_{g=1}^{J-1} E_{t=1}[\tilde{\lambda}_{i,gj}^{(2)}(2, b_i^{(g)}(1))], \text{ by (64)} \\ &= \sum_{g=1}^{J-1} \sum_{b_i^{(g)}(1)=0}^{1} \tilde{\lambda}_{i,gj}^{(2)}(2, b_i^{(g)}(1))[1 - F_{(i1)g}]^{b_i^{(g)}(1)}[F_{(i1)g}]^{1-b_i^{(g)}(1)} \quad (70) \\ &= \sum_{g=1}^{J-1} \left[\{\tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 1)[F_{(i1)g}]\} + \{\tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 2)(1 - F_{(i1)g})\}\right] \\ &= \sum_{g=1}^{J-1} \left[\tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 2) + F_{(i1)g}\{\tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 1) - \tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 2)\}\right] \\ &= \mu_{i2}^{(j)} \text{ (say)}, \quad (71) \end{aligned}$$

where

$$F_{(i1)g} = \sum_{v=1}^{g} \pi_{(i1)v}$$
$$\tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 1) = \frac{\exp(\tilde{x}_{i2}'\alpha_j)}{1 + \exp(\tilde{x}_{i2}'\alpha_j)}$$
$$\tilde{\lambda}_{i,gj}^{(2)}(2, g^* = 2) = \frac{\exp(\tilde{x}_{i2}'\alpha_j + \tilde{\gamma}_{gj})}{1 + \exp(\tilde{x}_{i2}'\alpha_j + \tilde{\gamma}_{gj})}. \quad (72)$$

Similarly,

$$\begin{aligned}
E[b_i^{(j)}(3)] &= \sum_{g=1}^{J-1} E_{t=1}E_{2|1}E_{3|2}[b_i^{(j)}(3)|b_i^{(g)}(2)] \\
&= \sum_{g=1}^{J-1} E_2[\tilde{\lambda}_{i,gj}^{(2)}(3, b_i^{(g)}(2))], \text{ by (64)} \\
&= \sum_{g=1}^{J-1} \sum_{b_i^{(g)}(2)=0}^{1} \tilde{\lambda}_{i,gj}^{(2)}(3, b_i^{(g)}(2))[\mu_{i2}^{(g)}]^{b_i^{(g)}(2)}[1-\mu_{i2}^{(g)}]^{1-b_i^{(g)}(1)} \\
&= \sum_{g=1}^{J-1} \left[\{\tilde{\lambda}_{i,gj}^{(2)}(3,0)[1-\mu_{i2}^{(g)}]\} + \{\tilde{\lambda}_{i,gj}^{(2)}(3,1)[\mu_{i2}^{(g)}]\}\right] \\
&= \mu_{i3}^{(j)},
\end{aligned} \tag{73}$$

where

$$\begin{aligned}
\tilde{\lambda}_{i,gj}^{(2)}(3,0) &= \frac{\exp(\tilde{x}_{i3}'\alpha_j)}{1+\exp(\tilde{x}_{i3}'\alpha_j)} \\
\tilde{\lambda}_{i,gj}^{(2)}(3,1) &= \frac{\exp(\tilde{x}_{i3}'\alpha_j + \tilde{\gamma}_{gj})}{1+\exp(\tilde{x}_{i3}'\alpha_j + \tilde{\gamma}_{gj})}.
\end{aligned} \tag{74}$$

In general,

$$\begin{aligned}
E[b_i^{(j)}(t)] &= \sum_{g=1}^{J-1} E_{t-1}[\tilde{\lambda}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1)) \\
&= \sum_{g=1}^{J-1} \left[\{\tilde{\lambda}_{i,gj}^{(2)}(t,0)[1-\mu_{i,t-1}^{(g)}]\} + \{\tilde{\lambda}_{i,gj}^{(2)}(t,1)[\mu_{i,t-1}^{(g)}]\}\right] \\
&= \mu_{it}^{(j)},
\end{aligned} \tag{75}$$

where

$$\begin{aligned}
\tilde{\lambda}_{i,gj}^{(2)}(t,0) &= \frac{\exp(\tilde{x}_{it}'\alpha_j)}{1+\exp(\tilde{x}_{it}'\alpha_j)} \\
\tilde{\lambda}_{i,gj}^{(2)}(t,1) &= \frac{\exp(\tilde{x}_{it}'\alpha_j + \tilde{\gamma}_{gj})}{1+\exp(\tilde{x}_{it}'\alpha_j + \tilde{\gamma}_{gj})}.
\end{aligned} \tag{76}$$

Therefore, a general formula for the marginal variance may now be written as

$$\text{var}[b_i^{(j)}(t)] = \mu_{it}^{(j)}[1 - \mu_{it}^{(j)}]. \quad (77)$$

where $\mu_{it}^{(j)}$ is given by (75).

These means (75) and variances (77) are important properties to understand the nature of ordinal response over time from an individual. For the purpose of estimating these means and variances, one requires to estimate the $\alpha$ and $\gamma$ parameters involved in the model. In the next section, we demonstrate how to obtain the likelihood estimates for these parameters.

## *4.2 Likelihood Estimation of the Parameters*

The $\alpha$ parameters are involved in both marginal and conditional binary probabilities as shown in (66) and (67). Let

$$\alpha = [\alpha_1', \ldots, \alpha_j', \ldots, \alpha_{J-1}']' : (p+1)(J-1) \times 1, \text{ with } \alpha_j = [\alpha_{j0}, \tilde{\alpha}_j']',$$

for $j = 1, \ldots, J-1$. Similarly, we use the notation $\tilde{\gamma}$ for all dynamic dependence parameters, namely

$$\tilde{\gamma} = [\tilde{\gamma}_1', \ldots, \tilde{\gamma}_j', \ldots, \tilde{\gamma}_{J-1}']' : (J-1)^2 \times 1, \text{ with } \tilde{\gamma}_j = [\tilde{\gamma}_{1j}, \ldots, \tilde{\gamma}_{gj}, \ldots, \tilde{\gamma}_{(J-1)j}]'.$$

Similar to (27), the likelihood function for $\alpha$ and $\tilde{\gamma}$ as

$$\begin{aligned} L(\alpha, \tilde{\gamma}) &= \Pi_{i=1}^K \left[ L_{i1}(\alpha) \Pi_{t=2}^T L_{it|t-1}(\alpha, \tilde{\gamma}) \right] \\ &= \Pi_{i=1}^K \Pi_{j=1}^{J-1} \left[ \{F_{(i1)j}\}^{1-b_i^{(j)}(1)} \right] \left[ \{1 - F_{(i1)j}\}^{b_i^{(j)}(1)} \right] \quad (78) \\ &\times \Pi_{i=1}^K \Pi_{t=2}^T \Pi_{g=1}^{J-1} \Pi_{j=1}^{J-1} \left[ \{\tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))\}^{b_i^{(j)}(t)} \{\tilde{\eta}_{i,gj}^{(1)}(t, b_i^{(g)}(t-1))\}^{1-b_i^{(j)}(t)} \right], \end{aligned}$$

where

$$\begin{aligned} 1 - F_{(i1)j} &= \frac{\exp(\tilde{x}_{i1}'\alpha_j)}{1 + \exp(\tilde{x}_{i1}'\alpha_j)}, \text{ and} \\ \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1)) &= \frac{\exp(\tilde{x}_{it}'\alpha_j + \tilde{\gamma}_{gj} b_i^g(t-1))}{1 + \exp(\tilde{x}_{it}'\alpha_j + \tilde{\gamma}_{gj} b_i^g(t-1))} \\ \tilde{\eta}_{i,gj}^{(1)}(t, b_i^{(g)}(t-1)) &= 1 - \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1)). \quad (79) \end{aligned}$$

It then follows that the likelihood estimating equations for $\alpha$ and $\tilde{\gamma}$ are given by

$$\frac{\partial \text{Log } L(\alpha, \tilde{\gamma})}{\partial \alpha} = \sum_{i=1}^{K}\sum_{j=1}^{J-1}\left[\frac{\{1-b_i^{(j)}(1)\}}{F_{(i1)j}} - \frac{\{b_i^{(j)}(1)\}}{\{1-F_{(i1)j}\}}\right]\frac{\partial F_{(i1)j}}{\partial \alpha}$$
$$+ \sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\left[\frac{b_i^{(j)}(t)}{\tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))} - \frac{\{1-b_i^{(j)}(t)\}}{\{1-\tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))\}}\right]\frac{\partial \eta_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))}{\partial \alpha}$$
$$= 0, \tag{80}$$

and

$$\frac{\partial \text{Log } L(\alpha, \tilde{\gamma})}{\partial \tilde{\gamma}} = \sum_{i=1}^{K}\sum_{t=2}^{T}\sum_{g=1}^{J-1}\sum_{j=1}^{J-1}\left[\frac{b_i^{(j)}(t)}{\tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))} - \frac{\{1-b_i^{(j)}(t)\}}{\{1-\tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))\}}\right]$$
$$\times \frac{\partial \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))}{\partial \tilde{\gamma}} = 0, \tag{81}$$

respectively. In (80),

$$\frac{\partial F_{(i1)j}}{\partial \alpha} = \begin{pmatrix} \frac{\partial F_{(i1)j}}{\partial \alpha_1} \\ \vdots \\ \frac{\partial F_{(i1)j}}{\partial \alpha_{j-1}} \\ \frac{\partial F_{(i1)j}}{\partial \alpha_j} \\ \frac{\partial F_{(i1)j}}{\partial \alpha_{j+1}} \\ \vdots \\ \frac{\partial F_{(i1)j}}{\partial \alpha_{J-1}} \end{pmatrix} = \begin{pmatrix} 0_{p+1} \\ \vdots \\ 0_{p+1} \\ -F_{(i1)j}[1-F_{(i1)j}]\tilde{x}_{i1} \\ 0_{p+1} \\ \vdots \\ 0_{p+1} \end{pmatrix} : (p+1)(J-1) \times 1, \tag{82}$$

and

$$\frac{\partial \eta_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))}{\partial \alpha} \tag{83}$$
$$= \begin{pmatrix} \frac{\partial \eta_{i,gj}^{(2)}(t,b_i^{(g)}(t-1))}{\partial \alpha_1} \\ \vdots \\ \frac{\partial \eta_{i,gj}^{(2)}(t,b_i^{(g)}(t-1))}{\partial \alpha_{j-1}} \\ \frac{\partial \eta_{i,gj}^{(2)}(t,b_i^{(g)}(t-1))}{\partial \alpha_j} \\ \frac{\partial \eta_{i,gj}^{(2)}(t,b_i^{(g)}(t-1))}{\partial \alpha_{j+1}} \\ \vdots \\ \frac{\partial \eta_{i,gj}^{(2)}(t,b_i^{(g)}(t-1))}{\partial \alpha_{J-1}} \end{pmatrix} = \begin{pmatrix} 0_{p+1} \\ \vdots \\ 0_{p+1} \\ \eta_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))\eta_{i,gj}^{(1)}(t, b_i^{(g)}(t-1))\tilde{x}_{it} \\ 0_{p+1} \\ \vdots \\ 0_{p+1} \end{pmatrix} : (p+1)(J-1) \times 1.$$

Similarly, in (81),

$$\frac{\partial \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))}{\partial \tilde{\gamma}} \tag{84}$$

$$= \begin{pmatrix} \frac{\partial \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))}{\partial \tilde{\gamma}_1} \\ \vdots \\ \frac{\partial \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))}{\partial \tilde{\gamma}_{j-1}} \\ \frac{\partial \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))}{\partial \tilde{\gamma}_j} \\ \frac{\partial \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))}{\partial \tilde{\gamma}_{j+1}} \\ \vdots \\ \frac{\partial \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))}{\partial \tilde{\gamma}_{J-1}} \end{pmatrix} = \begin{pmatrix} 0_{J-1} \\ \vdots \\ 0_{J-1} \\ \eta^{(2)}_{i,gj}(t,b^{(g)}_i(t-1))\eta^{(1)}_{i,gj}(t,b^{(g)}_i(t-1))b^*_i(t-1) \\ 0_{J-1} \\ \vdots \\ 0_{J-1} \end{pmatrix} : (J-1)^2 \times 1,$$

where $b^*_i(t-1) = [b^{(1)}_i(t-1), \ldots, b^{(g)}_i(t-1), \ldots, b^{(J-1)}_i(t-1)]' : (J-1) \times 1.$

The likelihood estimating equation in (80) may be solved iteratively by using the iterative equation for $\alpha$ given by

$$\hat{\alpha}(r+1) = \hat{\alpha}(r) - \left[ \left\{ \frac{\partial^2 \text{Log } L(\alpha, \tilde{\gamma})}{\partial \alpha' \partial \alpha} \right\}^{-1} \frac{\partial \text{Log } L(\alpha, \tilde{\gamma})}{\partial \alpha} \right]_{|\alpha=\hat{\alpha}(r)} ;\ (J-1)(p+1) \times 1, \tag{85}$$

where the formula for the second order derivative matrix $\frac{\partial^2 Log\ L(\alpha,\tilde{\gamma}}{\partial \alpha' \partial \alpha}$ is, similar to (36), given by

$$\frac{\partial^2 \text{Log } L(\alpha, \tilde{\gamma})}{\partial \alpha \partial \alpha'} = -\sum_{i=1}^{K} \sum_{j=1}^{J-1} \frac{\partial F_{(i1)j}}{\partial \alpha} \{(1 - F_{(i1)j})F_{(i1)j}\}^{-1} \frac{\partial F_{(i1)j}}{\partial \alpha'}$$
$$- \sum_{i=1}^{K} \sum_{t=2}^{T} \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \frac{\partial \tilde{\eta}^{(2)}_{i,gj}(t, b^{(g)}_i(t-1))}{\partial \alpha} \{\tilde{\eta}^{(2)}_{i,gj}(t, b^{(g)}_i(t-1))(1 - \tilde{\eta}^{(2)}_{i,gj}(t, b^{(g)}_i(t-1)))\}^{-1}$$
$$\times \frac{\partial \tilde{\eta}^{(2)}_{i,gj}(t, b^{(g)}_i(t-1))}{\partial \alpha'}. \tag{86}$$

Similarly, the likelihood estimating equation in (81) may be solved iteratively by using the iterative equation for $\tilde{\gamma}$ given by

$$\hat{\tilde{\gamma}}(r+1) = \hat{\tilde{\gamma}}(r) - \left[ \left\{ \frac{\partial^2 \text{Log } L(\alpha, \tilde{\gamma})}{\partial \tilde{\gamma} \partial \tilde{\gamma}'} \right\}^{-1} \frac{\partial \text{Log } L(\alpha, \tilde{\gamma})}{\partial \tilde{\gamma}} \right]_{|\tilde{\gamma}=\hat{\tilde{\gamma}}(r)} ;\ (J-1)^2 \times 1, \tag{87}$$

where the second order derivative matrix with respect to $\tilde{\gamma}$, by using the same argument for the derivation of (86) [see also (36)] for $\alpha$, has a simple approximate formula given by

$$\frac{\partial^2 \text{Log } L(\alpha, \tilde{\gamma})}{\partial \tilde{\gamma} \partial \tilde{\gamma}'}$$
$$= -\sum_{i=1}^{K} \sum_{t=2}^{T} \sum_{g=1}^{J-1} \sum_{j=1}^{J-1} \frac{\partial \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))}{\partial \tilde{\gamma}} \{\tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))(1 - \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1)))\}^{-1}$$
$$\times \frac{\partial \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))}{\partial \tilde{\gamma}'}, \tag{88}$$

where $\frac{\partial \tilde{\eta}_{i,gj}^{(2)}(t, b_i^{(g)}(t-1))}{\partial \tilde{\gamma}}$ is given in (84).

## 5 Concluding Remarks

Longitudinal categorical data have been modeled using several different techniques in the past. The real problem is to simplify the complex issue of adding a dependence structure to categorical data. Common methods are using random effects model, or using working correlations. Both suffer from serious drawbacks that we indicated in our paper. For this paper unlike the existing 'working' models, we develop two correlation models for repeated ordinal responses, where we directly put in a dependence structure for the categorical data.

The models, discussed in this paper, are similar to the correlation models considered by Sutradhar and Kovacevic (2000), but they are simpler. These are cumulative multinomial dynamic logit (CMDL) model (where longitudinal correlations are first modeled for nominal repeated data; and then we apply the cumulative principle to reflect the ordinal nature of the correlated data by using binary cut-points;) and the correlated binary dynamic model (CBDM) (where we first look at binary cut-points then add the correlation directly in the binary model).

Both models are analyzed by using the well known likelihood estimation approach. We have provided details about the estimation in terms of finding the Hessian and the Gradient function for the likelihood estimation. The next step would be to conduct a simulation study to compare and contrast the performance of these two methods with the existing methods. In the near future we plan to have R code freely available for practitioners to use.

**Acknowledgements** The authors thank the audience of the symposium for constructive discussion and a referee for valuable comments and suggestions.

## References

Agresti, A.: A survey of models for repeated ordered categorical response data. Stat. Med. **8**, 1209–1224 (1989)

Agresti, A., Natarajan, R.: Modeling clustered ordered categorical data: a survey. Int. Stat. Rev. **69**, 345–371 (2001)

Amemiya, T.: Advanced econometrics. Cambridge: Harvard University Press, Cambridge (1985)

Conaway, M.R.: Analysis of repeated categorical measurements with conditional likelihood methods. J. Am. Stat. Assoc. **84**, 53–62 (1989)

Crowder, M.: On the use of a working correlation matrix in using generalized linear models for repeated measure. Biometrika **82**, 407–410 (1995)

Ekholm, A., Jokinen, J., Kilpi, T.: Combining regression and association modelling for longitudinal data on bacterial carriage. Stat. Med. **21**, 773–791 (2002)

Ekholm, A., Jokinen, J., McDonald, J.W., Smith, P.W.F.: Joint regression and association modelling for longitudinal ordinal data. Biometrics **59**, 795–803 (2003)

Fienberg, S.F., Bromet, E.J., Follmann, D., Lambert, D., May, S.M.: Longitudinal analysis of categorical epidemiological data: a study of three mile island. Environ. Health Perspect. **63**, 241–248 (1985)

Jokinen, J.: Fast estimation algorithm for likelihood-based analysis of repeated categorical responses. Comput. Stat. Data Anal. **51**, 1509–1522 (2006)

Lipsitz, S.R., Laird, N.M., Harrington, D.P.: Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. Biometrika **78**, 53–160 (1991)

Lipsitz, S.R., Kim, K., Zhao, L.: Analysis of repeated categorical data using generalized estimating equations. Stat. Med. **13**, 1149–1163 (1994)

Loredo-Osti, J.C., Sutradhar, B.C.: Estimation of regression and dynamic dependence parameters for non-stationary multinomial time series. J. Time Ser. Anal. **33**, 458–467 (2012)

McCullagh, P.: Quasi-likelihood functions. Ann. Stat. **11**, 59–67 (1983)

McDonald, D.R.: The local limit theorem: a historical perspective. J. Iran. Stat. Soc. **4**, 73–86 (2005)

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

Sutradhar, B.C.: Longitudinal Categorical Data Analysis. Springer, New York (2014)

Sutradhar, B.C., Das, K.: On the efficiency of regression estimators in generalized linear models for longitudinal data. Biometrika **86**, 459–465 (1999)

Sutradhar, B.C., Farrell, P.J.: On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. Sankhya B **69**, 448–467 (2007)

Sutradhar, B.C., Kovacevic, M.: Analyzing ordinal longitudinal survey data: generalized estimating equations approach. Biomerika **87**, 837–848 (2000)

Tagore, V., Sutradhar, B.C.: Conditional inference in linear versus nonlinear models for binary time series. J. Stat. Comput. Simul. **79**, 881–897 (2009)

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

Williamson, J.M., Kim, K., Lipsitz, S.R.: Analyzing bivariate ordinal data using a global odds ratio. J. Am. Stat. Assoc. **90**, 1432–1437 (1995)

Yi, G.Y., Cook, R.: Marginal methods for incomplete longitudinal data arising in clusters. J. Am. Stat. Assoc. **97**, 1071–1080 (2002)

# Semi-parametric Models for Longitudinal Count, Binary and Multinomial Data

**Brajendra C. Sutradhar**

**Abstract** In a longitudinal setup, the semi-parametric regression model contains a specified regression function in some suitable time dependent primary covariates and a non-parametric function in some other time dependent say secondary covariates. However, the functional form for such a semi-parametric regression model depends on the nature of the repeated responses collected from a large number of independent individuals. In cross sectional setup, these functional forms represent the marginal expectations of the responses, whereas in a longitudinal setup, in general, they represent the expectation of a response at a given time conditional on past responses. More specifically, these conditional expectations are modelled through certain dynamic relationships among the repeated responses which also specify the longitudinal correlation structure among these repeated responses. In this paper, we consider a lag 1 dynamic relationship among repeated responses whether they are linear, count, binary or multinomial, and exploit the underlying correlation structure for consistent and efficient estimation of the regression parameters involved in the specified regression function in primary covariates. Because the non-parametric function in secondary covariates is not of direct interest, for simplicity, we estimate this function consistently in all cases by using 'working' independence assumption for the repeated responses.

**Keywords** Binary response • Consistency • Count response • Dynamic model for repeated responses • Multinomial/categorical response • Non-parametric function • Non-stationary correlations • Parametric regression function • Semi-parametric quasi-likelihood estimation • Semi-parametric generalized quasi-likelihood estimation

B.C. Sutradhar (✉)
Department of Mathematics and Statistics, Memorial University,
St. John's, NL, Canada A1C5S7
e-mail: bsutradh@mun.ca

 
B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_7

## 1 Introduction

In a cross sectional setup, when a discrete such as count or binary or multinomial response $y_i$ along with a $p$-dimensional covariate vector $x_i$ is collected from an individual $i(i = 1, \ldots, K)$, the effects of $x_i$ on $y_i$, say $\beta = (\beta_1, \ldots, \beta_p)'$, is customarily computed by fitting the so-called mean regression model of the form

$$E[Y_i|x_i] = m_i(\beta, x_i), \tag{1}$$

where the specific form of $m_i(\beta, x_i)$ depends on the nature of the response $y_i$. For example, it is customary to use

$$m_i(\beta, x_i) = \begin{cases} E[Y_i] = \exp(x_i'\beta) & \text{for scalar count response } y_i \\ P[y_i = 1] = \frac{\exp(x_i'\beta)}{1+\exp(x_i'\beta)} & \text{for scalar binary response } y_i, \end{cases} \tag{2}$$

where $x_i'\beta$ is referred to as a linear predictor in a generalized linear model (GLM); and

$$m_{(i)c}(\beta, x_i) = P[y_i = y_i^{(c)}] = \begin{cases} \frac{\exp(x_i'\beta_c)}{1+\sum_{g=1}^{C-1}\exp(x_i'\beta_g)} & \text{for } c = 1, \ldots, C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1}\exp(x_i'\beta_g)} & \text{for } c = C, \end{cases} \tag{3}$$

for a $C-1$ ($C \geq 3,\ C = 2$ being the binary case) dimensional multinomial variable $y_i = (y_{i1}, \ldots, y_{ic}, \ldots, y_{i,C-1})'$ to have its realization in the $c$th ($c = 1, \ldots, C-1$) category, that is

$$y_i = y_i^{(c)} = (y_{i1}^{(c)}, \ldots, y_{ij}^{(c)}, \ldots, y_{i,C-1}^{(c)})' = (01'_{c-1},\ 1,\ 01'_{C-1-c})' \equiv \delta_{ic}, \tag{4}$$

out of $C$ possible categories. In (3), $\beta_c = (\beta_{c1}, \ldots, \beta_{cp})'$ for $c = 1, \ldots, C-1$, and $\beta = (\beta_1', \ldots, \beta_c', \ldots, \beta_{C-1}')' : (C-1)p \times 1$. Notice that fitting these models (2) and (3) to the corresponding data amounts to estimating the $p$-dimensional $\beta$ vector in (2) and $(C-1)p$-dimensional $\beta$ vector in (3). Note that in (4), we have written $01'_{c-1}$ for $0 \otimes 1'_{c-1}$ for simplicity. This notation will be followed all through the paper for convenience.

In practice, there are some situations, where in addition to $x_i$, some other covariates, say $z_i$, are collected from the same individual $i$, but the effects of these additional covariates are not of direct interest. For example, in a health care utilization data, one may be interested to find the effects of the primary covariates $x_i \equiv$ [gender, education level, chronic condition]$'$ of the $i$th individual on the number of physician visits $y_i$ in a given time, by treating age of the $i$th individual $(z_i)$ as a secondary covariate. This type of secondary covariates $z_i$ may be fitted non-parametrically by adding a smooth non-parametric function $\psi(z_i)$ to the linear predictor in (2) and $\psi(z_i^{(j)})$ to the linear predictor $x_i'\beta_j$ in (3). Thus, a semi-parametric model of the form

$$\mu_i(\beta, x_i, \psi(z_i)) = \begin{cases} E[Y_i|x_i, z_i] = \exp(x_i'\beta + \psi(z_i)) & \text{for scalar count response } y_i \\ \pi_i(\beta, x_i, z_i) = P[y_i = 1|x_i, z_i] = \frac{\exp(x_i'\beta+\psi(z_i))}{1+\exp(x_i'\beta+\psi(z_i))} & \text{for scalar binary response } y_i \end{cases} \quad (5)$$

corresponding to (2) (Severini and Staniswallis 1994, Sects. 1–7; Carota and Parmigiani 2002; Horowitz 2009), or

$$\pi_{(i)c}(\beta, x_i, z_i) = P[y_i = y_i^{(c)}|x_i, z_i] = \begin{cases} \frac{\exp(x_i'\beta_c+\psi^{(c)}(z_i))}{1+\sum_{g=1}^{C-1}\exp(x_i'\beta_g+\psi^{(g)}(z_i)} & \text{for } c = 1, \ldots, C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1}\exp(x_i'\beta_g+\psi^{(c)}(z_i))} & \text{for } c = C, \end{cases} \quad (6)$$

corresponding to (3), is fitted to the responses $\{y_i, i = 1, \ldots, K\}$, but one is still interested to estimate the main regression parameters $\beta$. Remark that when $z_i'$s are assumed to influence $y_i$ through model (5) or (6), any estimate obtained for $\beta$ by ignoring $\psi(z_i)$ (or $\psi^{(c)}(z_i)$), that is, by fitting (2) or (3), would be biased and hence mean squared error inconsistent.

Note that as a generalization of the semi-parametric model (6) under the independence setup, there does not exist any studies for multinomial data in the longitudinal setup. In fact, semi-parametric model (6) is also not adequately addressed in the independence setup. A detailed discussion about this model in a wider longitudinal setup along with inferences for the model parameters is given in Sect. 5.

Turning back to the semi-parametric models given in (5) for count and binary data, some authors such as Zeger and Diggle (1994), Severini and Staniswallis (1994, Sect. 8), Lin and Carroll (2001) have extended these models to the longitudinal setup, where semi-parametric models are constructed to relate the repeated responses with two types (primary of direct interest and secondary) of multi-dimensional covariates. Suppose that $t_{ij}$ denote the time at which the $j$th ($j = 1, \ldots, n_i$) count response is recorded from the $i$th ($i = 1, \ldots, K$) individual, and $y_{ij}$ denote this count response. Next, unlike the scalar response case explained through the models (5), suppose that $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$ denotes the $n_i \times 1$ vector of repeated counts for the $i$th ($i = 1, \ldots, K$) individual. Also suppose that $y_{ij}$ is influenced by a fixed and known $p$-dimensional time dependent primary covariate vector $x_{ij}(t_{ij})$ and an additional time dependent scalar secondary covariate $z_{ij}(t_{ij})$. Note that similar to (5), the primary covariates are included in the regression model parametrically using linear predictor, whereas the covariate(s) of secondary interest are included in the model non-parametrically. Further note that in most of the longitudinal studies, the primary covariates are collected at $t_{ij} = j$, and hence $x_{ij}(t_{ij})$ can be replaced by $x_{ij}(j)$, when convenient. For example, in the aforementioned health care utilization study, $x_{ij}(j)$ may represent the education level status of the

$i$th individual in $j$th year, where detailed breakdown such as the exact month when education level is attained may not be more informative. For the same yearly data, as opposed to the primary covariates, the secondary covariates are usually collected at time $t_{ij}$. For example, even though the $i$th individual provides $x_{ij}(j)$ at $j$th year, the individual may be asked to provide, say the worst allergy status $(z_{ij})$ in a given scale, and its occurrence time $t_{ij}$ defined as

$$\begin{aligned} t_{ij} &\equiv (j-1) + [\text{numeric month for the occurrence for individual } i]_i \\ &\div 12,\; j = 1,\ldots,n_i;\; i = 1,\ldots,K, \end{aligned} \tag{7}$$

which makes both $t_{ij}$ and $z_{ij}$ quite dense. Some authors such as Severini and Staniswallis (1994, Sect. 8) have considered time independent but dense $z_{ij}$, whereas Lin and Carroll (2001), for example, have considered $t_{ij}$ as the secondary covariate, that is $z_{ij}(t_{ij}) \equiv t_{ij}$. In their numerical illustration, Lin and Carroll (2001, Sect. 8) have considered $z_{ij} \equiv t_{ij} =$ exact age of the $i$th child in $j$th quarter. In general, these authors have assumed that in a longitudinal setup, in addition to time dependent primary covariates $x_{ij}(t_{ij})$, the times $t_{ij}$ as a fixed secondary covariate also influence the response $y_{ij}$. Consequently, because of the fact that the repeated responses $\{y_{ij}, j = 1,\ldots,n_i\}$ are likely to be correlated, Lin and Carroll (2001, Eq. (10)), for example, have estimated the regression effects $\beta$ by solving the so-called 'working' correlations based GEE (generalized estimating equation)

$$\sum_{i=1}^{K} \frac{\partial \mu_i'(\beta, X_i, \hat{\psi}(\beta, z_i))}{\partial \beta} V_i^{-1}(y_i - \mu_i(\beta, X_i, \hat{\psi}(\beta, z_i))) = 0, \tag{8}$$

where $X_i' = (x_i(t_{i1}),\ldots,x_i(t_{ij}),\ldots,x_i(t_{in_i}))$ denote the $p \times n_i$ covariate matrix with $x_i(t_{ij})$ as the $p$-dimensional covariate vector for the $i$th individual at time point $t_{ij}$, $\mu_i(\beta, X_i, \hat{\psi}(\beta, z_i)$ is a mean vector as opposed to the scalar mean $\mu_i(\cdot)$ in (5), $\hat{\psi}(\beta, z_i)$ is a $n_i \times 1$ consistent estimate of the nonparametric vector function for known $\beta$ corresponding to $j = 1,\ldots,n_i$, and $V_i$ is a so-called $n_i \times n_i$ 'working' correlation matrix representing the correlations of the repeated responses which is computed by

$$V_i = A_i^{\frac{1}{2}} R_i A_i^{\frac{1}{2}}, \tag{9}$$

where $A_i = \text{diag}[\text{var}(y_{i1}),\ldots,\text{var}(y_{ij}),\ldots,\text{var}(y_{in_i})]$ with $\text{var}(y_{ij}) = \mu_{ij}(\beta, x_{ij}, \hat{\psi}(\beta, z_{ij}))$ for the Poisson panel data, for example; and $R_i$ has been computed by an unstructured (UNS) common constant correlation matrix $(R)$ as

$$R(\equiv R_i) = K^{-1} \sum_{i=1}^{K} r_i r_i', \text{ where } r_i = (r_{i1},\ldots,r_{ij},\ldots,r_{in_i})', \tag{10}$$

with $r_{ij} = \frac{(y_{ij}-\mu_{ij}(\beta,x_{ij},\hat{\psi}(\beta,z_{ij})))}{[\mu_{ij}(\beta,x_{ij},\hat{\psi}(\beta,z_{ij}))]^{\frac{1}{2}}}$, for count data, for example Severini and Staniswallis (1994, Sect. 8) have also used this 'working' correlation $R_i$ for the estimation of the correlation matrix for the repeated responses $y_{i1},\ldots,y_{in_i}$. Lin and Carroll (2001, Eqs. (6)–(7)) also estimate the non-parametric function $\psi(\beta,z_i)$ using the working correlation matrix $R_i$, whereas Severini and Staniswallis (1994, Sect. 8) (see also Zeger and Diggle 1994, Sneddon and Sutradhar 2004, and You and Chen 2007) estimated this function using 'working' independence among the repeated data. This approach of Lin and Carroll (2001) have several drawbacks as follows both for estimation of $\beta$ and the non-parametric function.

**(1)** The common matrix $R$ cannot be computed unless $n_i = n$ for all $i = 1 = 1,\ldots,K$, and it is not true that one can use this $R$ matrix for panel data especially when $n_i \times n_i$ matrix is needed for the $i$th individual (see Sutradhar 2010). Furthermore, because covariates $(x_{ij})$ of an individual $i$ are dependent on $j$, it is demonstrated by Sutradhar (2010) that the correlations of the repeated data following a sensible dynamic model also involve $x_{ij}$ [see also Eq. (37) in Sect. 3]. This, for $j < k$, for a known function $h$, produces

$$E[r_{ij}r_{ik}] = h(x_{ij}, x_{ik}, \hat{\psi}(z_{ij}), \hat{\psi}(z_{ik})) \tag{11}$$

and hence the average $K^{-1}\sum_{i=1}^{K} r_{ij}r_{ij}$ obtained from all individuals may be biased (far away from true correlation) for the true correlation element $\rho_{i,jk}$ for the $i$th individual. This will produce inefficient estimate of $\beta$, specially when the covariates are $j$ dependent.

**(2)** When secondary covariates are fixed times such as $z_{ij} \equiv t_{ij}$, obtaining an estimate of the nonparametric function $\psi(t_{ij})$ at a fixed time point by using correlation matrix $R$ corresponding to all time points $\{j, j = 1,\ldots,n_i\}$ may be counter productive. In fact, Lin and Carroll (2001, Sect. 7) found that the 'working' correlation approach using the $R_i$ matrix for the estimation of $\psi()$ produces less efficient estimate than using the independence assumption, that is, $R_i = I_{n_i}$. Besides, $\psi()$ is of secondary interest and hence it is sufficient to estimate this function consistently, whereas more efforts is needed to obtain consistent and efficient estimate for the main regression parameter $\beta$.

Recently, Sutradhar et al. (2015) have studied a semi-parametric dynamic model for repeated count data. As far as the inference is concerned, these authors, as opposed to the 'working' correlations based semi-parametric GEE (SGEE) estimation approach, have used a simpler SQL (semi-parametric quasi-likelihood) approach for consistent estimation of the non-parametric function, and a SGQL (semi-parametric generalized quasi-likelihood) approach for the estimation of the regression effects involved in the specified regression function, whereas the longitudinal correlation index parameter involved in the dynamic model was consistently estimated by using the well known method of moments. This dynamic model and the estimation approach used by these authors are explained in brief in Sect. 3. A similar but different dynamic model, namely multinomial dynamic logit (MDL) model is discussed in Sect. 4, binary dynamic logit (BDL) model being a special case. This MDL model produces recursive means and variances, whereas the dynamic model for the count data to be discussed in Sect. 3 produces fixed marginal means and variances. The correlation structures produced by these two models (dynamic models for count data and MDL model) are quite different as well. Hence, the estimating equations under the MDL or BDL model cannot be obtained from those under the dynamic model for the count data by any simple substitutions, or vice versa.

Remark that as opposed to the inferences for the aforementioned semi-parametric longitudinal models for count, binary and multinomial data, there exist some studies for the inferences in simpler linear longitudinal models. See, for example, the studies by Sneddon and Sutradhar (2004), You and Chen (2007), and Warriyar and Sutradhar (2014). For convenience and simplicity, we therefore provide a brief discussion in Sect. 2 on the inferences for the semi-parametric linear longitudinal model following the recent work of Warriyar and Sutradhar (2014).

## 2 Semi-parametric Linear Longitudinal Models and Inferences

In a semi-parametric linear model setup, the repeated continuous data measured from the $i$th individual at time point $t_{ij}$ are usually modeled as

$$\begin{aligned} y_{ij} &= x'_{ij}(t_{ij})\beta + \psi(z_{ij}(t_{ij})) + \epsilon_{ij}(t_{ij}) \\ &= \mu_{ij}(t_{ij}) + \epsilon_{ij}(t_{ij}), \end{aligned} \tag{12}$$

or equivalently

$$y_i = X_i\beta + \psi_i + \epsilon_i, \tag{13}$$

where

$$\psi_i = (\psi(z_{i1}(t_{i1})), \cdots, \psi(z_{in_i}(t_{in_i})))' \text{ and } \epsilon_i = (\epsilon_{i1}(t_{i1}), \cdots, \epsilon_{ij}(t_{ij}), \cdots, \epsilon_{in_i}(t_{in_i}))'.$$

Note that in (13), $\psi_i$ is not a subject specific non-parametric function as its construction requires only knowing $\psi(z_{ij})$ at any time $t$ (see Zeger and Diggle 1994 and Sneddon and Sutradhar 2004, for example). To be specific, $\psi_i$ is used here to represent $n_i$ components, each with the same non-parametric function but evaluated at $n_i$ different values $z_{ij}$ for the $i$th individual. Further note that in this longitudinal setup, the components of the error vector $\epsilon_i$ must be correlated. But as the correlation structure is unknown in practice, many authors such as Zeger and Diggle (1994), Severini and Staniswallis (1994), and Lin and Carroll (2001) have considered that

$$\epsilon_i \sim (0, \sigma^2 R_i(\alpha)), \tag{14}$$

where $\sigma^2 = \text{var}[\epsilon_{ij}(t_{ij})] = \sigma_{ijj}(t_{ij}) = \sigma^2(t_{ij})$, and $R_i(\alpha)$ is a 'working' correlation matrix as in (9). However, Warriyar and Sutradhar (2014) have demonstrated that in stead of $R_i(\alpha)$, one may use a true correlation structure

$$C_i(\rho) = \begin{pmatrix} 1 & \rho_1 & \rho_2 \cdots & \rho_{n_i-1} \\ \rho_1 & 1 & \rho_1 \cdots & \rho_{n_i-2} \\ \vdots & & \cdots & \vdots \\ \rho_{n_i-1} & \rho_{n_i-2} & \cdots & 1 \end{pmatrix} \quad \text{for all i} = 1, 2, \ldots, K;$$

$$\Sigma_i(\rho) = \text{var}(Y_i) = A_i^{\frac{1}{2}} C_i(\rho) A_i^{\frac{1}{2}}, \tag{15}$$

where for $\ell = 1, \ldots, n_i - 1$, $\rho_\ell$ denotes the lag $\ell$ correlation between $\epsilon_{ij}(t_{ij})$ and $\epsilon_{i,j+\ell}(t_{i,j+\ell})$, and $A_i = \sigma^2 I_{n_i}$ where $\sigma^2$ is an unknown scalar constant, and $I_{n_i}$ is the $n_i \times n_i$ identity matrix. For example, this correlation structure $C_i(\rho)$ in (15) is appropriate for the ARMA (auto-regressive moving average) type dynamic models as follows:

(1) **AR(1) model:**

$$\epsilon_{ij}(t_{ij}) = \phi\, \epsilon_{i,j-1}(t_{i,j-1}) + a_{ij}(t_{ij}),\ |\phi| < 1,$$
$$a_{ij}(t_{ij}) \stackrel{iid}{\sim} N(0, \sigma_a^2)\ \ \forall \text{ i} = 1, 2, \ldots, K; j = 1, \ldots, n_i,$$

(2) **MA(1) model:**

$$\epsilon_{ij}(t_{ij}) = \theta\, a_{i,j-1}(t_{i,j-1}) + a_{ij}(t_{ij}),\ |\theta| < 1,$$
$$a_{ij}(t_{ij}) \stackrel{iid}{\sim} N(0, \sigma_a^2)\ \ \forall \text{ i} = 1, 2, \ldots, K; j = 1, \ldots, n_i,$$

and

(3) **EQC model :**

$$\varepsilon_{ij}(t_{ij}) = \varepsilon_{i0}(t_{i0}) + a_{ij}(t_{ij}),$$
$$a_{ij}(t_{ij}) \stackrel{iid}{\sim} (0, \sigma_a^2),\ \ \varepsilon_{i0}(t_{i0}) \sim N(0, \tilde{\sigma}^2),$$

yield $\rho_\ell$, the lag $\ell$ correlations between $\epsilon_{ij}(t_{ij})$ and $\epsilon_{i,j+\ell}(t_{i,j+\ell})$, as

$$\rho_\ell = \phi^\ell;\ \rho_\ell = \begin{cases} \frac{\theta}{1+\theta^2}, & \text{for } \ell = 1 \\ 0, & \text{for } \ell = 2, 3, \ldots, n_i - 1, \end{cases} \text{ and } \rho_\ell = \zeta = \frac{\tilde{\sigma}^2}{\tilde{\sigma}^2 + \sigma_a^2},$$

respectively, and they satisfy the auto-correlation structure $C_i(\rho)$ in (15).

### 2.1 SQL Estimation of the Non-parametric Function $\psi(z_{ij})$

Even though the repeated responses $y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i}$ are correlated with correlation structure as in (15), it is, however, simpler to use an 'working' independence assumption and obtain a consistent estimator for the desired non-parametric function. To be specific, one may solve the SQL (semi-parametric quasi-likelihood) estimating equation for $\psi(z_{ij})|_{z_{ij}=z_0}$ given by

$$\sum_{h=1}^{K}\sum_{u=1}^{n_h} w_{hu}(z_0)\frac{\partial \mu_{hu}(t_{hu})}{\partial \psi(z_0)}\frac{(y_{hu}-\mu_{hu}(t_{hu}))}{\sigma^2} = 0 \tag{16}$$

(Carota and Parmigiani 2002) where

$$w_{hu}(z_0) = \frac{p_{hu}(\frac{z_0-z_{hu}}{b})}{\sum_{h=1}^{K}\sum_{u=1}^{n_h} p_{hu}(\frac{z_0-z_{hu}}{b})},$$

$p_{hu}(.)$ being a suitable kernel, for example, we choose

$$p_{hu}(\frac{z_0-z_{hu}}{b}) = \frac{1}{\sqrt{2\pi}b}\exp(\frac{-1}{2}(\frac{z_0-z_{hu}}{b})^2)$$

with a suitable bandwidth $b$.

Next, because $\mu_{ij}(t_{ij}) = x'_{ij}(t_{ij})\beta + \psi(z_{ij})$ by (12), it is convenient to express the SQL estimating equation for $\psi(z_{ij})$ (16), in terms of known $\beta$, as

$$\hat{\psi}(z_{ij}) = \hat{y}_{ij} - \hat{x'}_{ij}(t_{ij})\beta, \tag{17}$$

where

$$\hat{y}_{ij} = \sum_{h=1}^{K}\sum_{u=1}^{n_h} w_{hu}(z_{ij})y_{hu} \text{ and } \hat{x'}_{ij}(t_{ij}) = \sum_{h=1}^{K}\sum_{u=1}^{n_h} w_{hu}(z_{ij})x'_h(t_{hu})$$

with $\sum_{h=1}^{K}\sum_{u=1}^{n_h} w_{hu}(z_{ij}) = 1$.

### 2.2 SGQL Estimation of $\beta$

Notice that when the nonparametric function $\psi(z_{ij})$ is replaced in the linear model (12) with its estimate $\hat{\psi}(z_{ij})$ from (17) for known $\beta$, one obtains

$$\begin{aligned} y_{ij} &= x'_{ij}(t_{ij})\beta + \hat{\psi}(z_{ij}) + \epsilon^*_{ij}(t_{ij}) \\ &= x'_{ij}(t_{ij})\beta + \hat{y}_{ij} - \hat{x'}_{ij}(t_{ij})\beta + \epsilon^*_{ij}(t_{ij}) \end{aligned} \tag{18}$$

where $\epsilon^*_{ij}(t_{ij})$ is a new error component different from that of (12). Now for all elements of the $i$th individual we use (18) and write

$$y_i - \hat{y}_i = (X_i - \hat{X}_i)\beta + \epsilon^*_i, \tag{19}$$

where by using the formulas for $\hat{y}_{ij}$ and $\hat{x'}_{ij}(t_{ij})$ from (17), one writes

$$\begin{aligned} \hat{y}_i &= \sum_{h=1}^{K}[W_h(z_{i1}, \ldots, z_{in_i})]y_h \\ \hat{X}_i &= \sum_{h=1}^{K}[W_h(z_{i1}, \ldots, z_{in_i})]X_h \end{aligned} \tag{20}$$

with $W_h(z_{i1}, \ldots, z_{in_i})$ as the kernel weights matrix defined as

$$W_h(z_{i1}, \ldots, z_{in_i}) = \begin{pmatrix} w'_h(z_{i1}) \\ \vdots \\ w'_h(z_{ij}) \\ \vdots \\ w'_h(z_{in_i}) \end{pmatrix} : n_i \times n_h \tag{21}$$

where $w'_h(z_{ij}) = [w_{h1}(z_{ij}), \ldots, w_{hu}(z_{ij}), \ldots, w_{hn_h}(z_{ij})]$ with $w_{hu}(z)$ at a given value $z$ as given in (16), and

$$X_h = \begin{pmatrix} x'_h(t_{h1}) \\ \vdots \\ x'_h(t_{hu}) \\ \vdots \\ x'_h(t_{hn_h}) \end{pmatrix} : n_h \times p \tag{22}$$

Notice that $\epsilon^*_i$ in (19) does not have the same mean vector and covariance matrix as for $\epsilon_i$ in (13). Suppose that

$$\mu^*_i = E[\epsilon^*_i] \text{ and } \Sigma^*_i = \text{cov}[\epsilon^*_i].$$

It then follows from (19) that

$$E[Y_i - \hat{Y}_i] = [X_i - \hat{X}_i]\beta + \mu^*_i, \; \text{cov}[Y_i - \hat{Y}_i] = \Sigma^*_i(\rho). \tag{23}$$

Consequently, following Sutradhar (2003, Sect. 3), Warriyar and Sutradhar (2014) have suggested to solve the so-called fully standardized SGQL estimating equation for $\beta$ given by

$$\sum_{i=1}^{K} \frac{\partial[(X_i - \hat{X}_i)\beta + \mu_i^*]'}{\partial \beta}[\Sigma_i^*]^{-1} \{(y_i - \hat{y}_i) - (X_i - \hat{X}_i)\beta - \mu_i^*\} = 0, \quad (24)$$

which yields a closed form formula for $\beta$ estimator as

$$\hat{\beta}_{FSSGQL} = \left[ \sum_{i=1}^{K} (X_i - \hat{X}_i)' \, (\Sigma_i^*)^{-1} \, (X_i - \hat{X}_i) \right]^{-1} \times \sum_{i=1}^{K} (X_i - \hat{X}_i)' \, (\Sigma_i^*)^{-1} \, (y_i - \hat{y}_i - \hat{\mu}_i^*), \quad (25)$$

where $\hat{\mu}_i^*$ is obtained by using $\hat{\psi}_i$ for $\psi_i$ involved in the formula for $\mu_i^*$ given below.

To compute the formula for $\mu_i^*$, we go back to (23) and write

$$\mu_i^* = E[Y_i - \hat{Y}_i] - [X_i - \hat{X}_i]\beta, \quad (26)$$

where by (20) one writes

$$\begin{aligned} E[\hat{Y}_i] &= \sum_{h=1}^{K} W_h(z_{i1}, \ldots, z_{in_i}) E[Y_h] \\ &= \sum_{h=1}^{K} W_h(z_{i1}, \ldots, z_{in_i})[X_h\beta + \psi_h], \text{ by (13)} \\ &= \hat{X}_i\beta + \sum_{h=1}^{K} W_h(z_{i1}, \ldots, z_{in_i})\psi_h, \quad (27) \end{aligned}$$

by (20), where $\psi_h = [\psi(z_{h1}), \ldots, \psi(z_{hu}), \ldots, \psi(z_{hn_h})]'$, yielding

$$E[Y_i - \hat{Y}_i] = [X_i - \hat{X}_i]\beta + \psi_i - \sum_{h=1}^{K} W_h(z_{i1}, \ldots, z_{in_i})\psi_h, \quad (28)$$

that is,

$$\begin{aligned} \mu_i^* = E[\epsilon_i^*] &= \psi_i - \sum_{h=1}^{K} W_h(z_{i1}, \ldots, z_{in_i})\psi_h \\ &= [\mu_{i1}^*, \ldots, \mu_{ij}^*, \ldots, \mu_{in_i}^*]', \quad (29) \end{aligned}$$

with

$$\mu^*_{ij} = \psi_i(z_{ij}) - \sum_{h=1}^{K}\sum_{u=1}^{n_h} w_{hu}(z_{ij})\psi_h(z_{hu}), \tag{30}$$

which is free from $\beta$.

As far as the formula for $\Sigma^*_i = \text{cov}[Y_i - \hat{Y}]$ in (25) is concerned, we first observe from (15) that

$$\text{cov}(Y_i) = \Sigma_i(\rho) = A_i^{-\frac{1}{2}} C_i(\rho) A_i^{-\frac{1}{2}}.$$

Next because $\hat{Y}_i = \sum_{h=1}^{K}[W_h(z_{i1},\ldots,z_{in_i})]Y_h$, it then follows that

$$\begin{aligned}
\Sigma^*_i &= \text{cov}[\epsilon^*_i] \\
&= \text{cov}(Y_i - \hat{Y}_i) \\
&= \text{cov}(Y_i) + \text{cov}(\hat{Y}_i) - 2\,\text{cov}(Y_i, \hat{Y}_i) \\
&= \Sigma_i(\rho) + \left[\sum_{h=1}^{K} W_h(z_{i1},\ldots,z_{in_i})\Sigma_h(\rho)W'_h(z_{i1},\ldots,z_{in_i})\right] \\
&\quad -2W_i(z_{i1},\ldots,z_{in_i})\Sigma_i(\rho).
\end{aligned} \tag{31}$$

### 2.3 *Moment Estimation of ρ*

For $n = \max_{1\le i\le K} n_i$, and

$$\delta_{iu} = \begin{cases} 1, \text{ if } u \le n_i \\ 0, \text{ if } n_i < u \le n, \end{cases}$$

the auto-correlation matrix $C_i(\rho)$ (15) is estimated by using the estimates of lag correlation $\rho_\ell$ given by

$$\widehat{\rho_\ell} = \frac{\sum_{i=1}^{K}\sum_{u=1}^{n-\ell}\delta_{iu}\delta_{i,u+\ell}\tilde{y}_{iu}\tilde{y}_{i,u+l}/\sum_{i=1}^{K}\sum_{u=1}^{n-\ell}\delta_{iu}\delta_{i,u+\ell}}{\sum_{i=1}^{K}\sum_{u=1}^{n_i}\delta_{iu}\tilde{y}^2_{iu}/\sum_{i=1}^{K}\sum_{u=1}^{n_i}\delta_{iu}},\ \ell = 1, 2, \ldots, n-1 \tag{32}$$

(Sutradhar 2011, Sect. 2.2.2) with $\tilde{y}_{iu} = \frac{y_{iu} - x'_{iu}\hat{\beta} - \hat{\psi}(z_{iu})}{\hat{\sigma}}$, where $\hat{\beta}$ is the SGQL estimate of $\beta$ as given in the last section, and $\hat{\psi}(z)$ is the SQL estimate of $\psi(z)$ as given in Sect. 2.1, and $\sigma^2$ for the $A_i$ matrix in (15) is estimated as

$$\hat{\sigma^2} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i}(y_{ij} - x'_{ij}\hat{\beta} - \hat{\psi}(z_{ij}))^2}{\sum_{i=1}^{K} n_i}. \tag{33}$$

For detailed finite sample performances of the aforementioned estimators discussed in Sects. 2.1–2.3, we refer to Warriyar and Sutradhar (2014).

## 3 Semi-parametric Longitudinal Models for Count Data and Inferences

It is understood that the Poisson distribution for count data belongs to the exponential family. Thus, unlike the additive model (12)–(13) in linear model setup, one may model the means and hence variances of the repeated count data by using the so-called log linear relationship. Thus, for all $j = 1, \ldots, n_i$, marginally $y_{ij}$ is a Poisson count with mean and variance given by

$$\begin{aligned} E[Y_{ij}|x_{ij}, z_{ij}] &= \mu_{ij}(\beta, x_{ij}, \psi(z_{ij})) = \exp(x'_{ij}(t_{ij})\beta + \psi(z_{ij})) \\ \text{var}[Y_{ij}|x_{ij}, z_{ij}] &= \sigma_{i,jj}(\beta, x_{ij}, \psi(z_{ij})) = \mu_{ij}(\beta, x_{ij}, \psi(z_{ij})) \\ &= \exp(x'_{ij}(t_{ij})\beta + \psi(z_{ij})). \end{aligned} \tag{34}$$

As far as the correlation model for the repeated counts is concerned, following Sutradhar (2010, Eq. (14)) (see also Sutradhar 2011, Chap. 6), we assume that for $j = 2, \ldots, n_i$, the repeated responses satisfy the lag 1 dynamic model given by

$$y_{ij} = \rho * y_{i,j-1} + d_{ij} \equiv \sum_{s=1}^{y_{i,j-1}} b_s(\rho) + d_{ij}, \tag{35}$$

where $Pr[b_s(\rho) = 1] = \rho$ and $Pr[b_s(\rho) = 0] = 1 - \rho$, with $\rho$ as the correlation index parameter; and $y_{i1} \sim$ Poisson $[\mu_{i1}(\beta, x_{i1}, \psi(z_{i1}))]$, and

$$d_{ij} \sim \text{Poisson}\left[\mu_{ij}(\beta, x_{ij}, \psi(z_{ij})) - \rho\mu_{i,j-1}(\beta, x_{i,j-1}, \psi(z_{i,j-1}))\right]$$

for $j = 2, \ldots, n_i$. Also, $d_{ij}$ and $y_{i,j-1}$ are assumed to be independent. It then follows from (35) that for $j < k$, the covariance between $y_{ij}$ and $y_{ik}$ has the formula

$$\begin{aligned} & cov(Y_{ij}, Y_{ik}|x_{ij}, x_{ik}, \psi(z_{ij}), \psi(z_{ik})) \\ &= E(Y_{ij}Y_{ik}|x_{ij}, x_{ik}, \psi(z_{ij}), \psi(z_{ik})) - E(Y_{ij}|x_{ij}, \psi(z_{ij}))E(Y_{ik}|x_{ik}, \psi(z_{ik})) \\ &= E_{Y_{ij}} Y_{ij} E_{Y_{i,j+1}} \ldots E_{Y_{i,k-1}} E[Y_{ik}|y_{i,k-1}, y_{i,k-2}, \ldots, y_{i,j+1}] \\ & \quad -\mu_{ij}(\beta, x_{ij}, \psi(z_{ij}))\mu_{ik}(\beta, x_{ik}, \psi(z_{ik})) \\ &= \sigma_{i,jk}(\beta, x_{ij}, x_{ik}, \psi(z_{ij}), \psi(z_{ik}), \rho) \\ &= \rho^{k-j}\mu_{ij}(\beta, x_{ij}, \psi(z_{ij})), \end{aligned} \tag{36}$$

yielding the correlations between $y_{ij}$ and $y_{ik}$ as

$$corr(Y_{ij}, Y_{ik}|x_{ij}, x_{ik}, \psi(z_{ij}), \psi(z_{ik})) = \begin{cases} \rho^{k-j}\sqrt{\frac{\mu_{ij}(\beta, x_{ij}; \psi(z_{ij}))}{\mu_{ik}(\beta, x_{ik}; \psi(z_{ik}))}} & j < k \\ \rho^{j-k}\sqrt{\frac{\mu_{ik}(\beta, x_{ik}; \psi(z_{ik}))}{\mu_{ij}(\beta, x_{ij}; \psi(z_{ij}))}} & j > k. \end{cases} \quad (37)$$

Remark that because the so-called error count $d_{ij}$ in model (35) has the Poisson distribution with mean $\mu_{ij}(\cdot) - \rho\mu_{i,j-1}(\cdot)$, it then follows that the correlation index parameter $\rho$ must satisfy the restriction

$$0 < \rho < \min\left[1, \frac{\mu_{ij}(\cdot)}{\mu_{i,j-1}(\cdot)}\right], \text{ for } j = 2, \ldots, n_i.$$

Thus, the dynamic model (35) allows only positive correlations between the repeated responses, and the exact correlation between any two responses can be computed by using the formula in (37).

Further remark that this formula in (37) demonstrates that the correlation values for an individual $i$ can be quite different than those of another individual. This makes the use of the common and constant UNS correlation matrix $R$ in (10) unsuitable for the $i$th individual.

As far as the inference is concerned, similar to the SQL and SGQL approaches discussed in Sects. 2.1 and 2.2, recently Sutradhar et al. (2015) have extended these estimation approaches to the repeated count data setup, and the authors have also examined their asymptotic performances analytically and their small sample performances through an intensive simulation study. In this section, we provide these estimation approaches in brief as follows.

### 3.1 SQL Estimation of the Non-parametric Function $\psi(z_{ij})$

Similar to (16) under the linear model, one may solve the SQL (semi-parametric quasi-likelihood) estimating equation for $\psi(z_{ij})|_{z_{ij}=z_0}$ as

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\,\frac{\partial \mu_{ij}}{\partial \psi(z_0)}\left(\frac{y_{ij} - \mu_{ij}}{\sigma_{i,jj}}\right) = 0, \quad (38)$$

(Carota and Parmigiani 2002) but unlike (16), we now have $\mu_{ij} \equiv \mu_{ij}(\beta, x_{ij}; \psi(z_{ij})) = \exp(x'_{ij}(t_{ij})\beta + \psi(z_{ij}))$ as in (34), and $\sigma_{i,jj} \equiv \sigma_{i,jj}(\beta, x_{ij}; \psi(z_{ij})) = \mu_{ij}$. The formula for the kernel weight is the same as in (16), that is,

$$w_{ij}(z_0) = p_{ij}\left(\frac{z_0 - z_{ij}}{b}\right)/\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}\left(\frac{z_0 - z_{ij}}{b}\right), \quad (39)$$

with $p_{ij}$ as the suitable kernel density.

Next because for this Poisson case one obtains

$$\frac{\partial \mu_{ij}(\beta, x_{ij}, \psi(z_0))}{\partial \psi(z_0)} = \frac{\partial \left[\exp(x'_{ij}(t_{ij})\beta + \psi(z_0))\right]}{\partial \psi(z_0)} = \exp(x'_{ij}(t_{ij})\beta + \psi(z_0)),$$

the SQL estimating equation (38) reduces to

$$\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)[y_{ij} - \exp(x'_{ij}(t_{ij})\beta + \psi(z_0))] = 0. \tag{40}$$

Consequently, one obtains

$$\exp(\hat{\psi}(\beta, z_0)) = \left(\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)y_{ij}\right) \Bigg/ \left[\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)\exp\{x'_{ij}(t_{ij})\beta\}\right], \tag{41}$$

yielding a closed form for the estimate of the non-parametric function given by

$$\hat{\psi}(\beta, z_0) = log\left(\frac{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)y_{ij}}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} w_{ij}(z_0)exp[x'_{ij}(t_{ij})\beta]}\right). \tag{42}$$

Under some mild conditions on the primary design covariates, Sutradhar et al. (2015) have shown that

$$\hat{\psi}(\beta, z_{ij}) = \psi(z_{ij}) + o_p(1), \tag{43}$$

justifying that $\hat{\psi}(\beta, z_{ij})$ given by (42) is a consistent estimator of $\psi(z_{ij})$.

### 3.2 SGQL Estimation of $\beta$

In Sect. 3.1, the non-parametric function $\psi(z_{ij})$ was estimated by $\hat{\psi}(\beta, z_{ij})$ (42) as a function of $\beta$. We may now replace the $\psi(z_{ij})$ function involved in the original mean and variance of the responses given in (34), and in the covariance of the repeated count responses given in (36), by this estimate $\hat{\psi}(\beta, z_{ij})$, and re-express these moments as

$$\begin{aligned}\tilde{\mu}_{ij}(\beta, x_{ij}, \hat{\psi}(\beta, z_{ij})) &= E[Y_{ij}|x_{ij}, \hat{\psi}(\cdot)] = \exp[x'_{ij}(t_{ij})\beta + \hat{\psi}(\beta, z_{ij})] \\ \tilde{\sigma}_{i,jj}(\beta, x_{ij}, \hat{\psi}(\beta, z_{ij})) &= \tilde{\mu}_{ij}(\beta, x_{ij}, \hat{\psi}(\beta, z_{ij})) = \exp[x'_{ij}(t_{ij})\beta + \hat{\psi}(\beta, z_{ij})], \text{ and} \\ \tilde{\sigma}_{i,jk}(\beta, \rho, x_{ij}, \hat{\psi}(\beta, z_{ij})) &= \rho^{k-j}\tilde{\mu}_{ij}(\beta, x_{ij}, \hat{\psi}(\beta, z_{ij})), \text{for } j < k,\end{aligned} \tag{44}$$

respectively. Now for notational simplicity, in (44), we use $\tilde{\mu}_{ij}(\beta, \hat{\psi}(\beta))$ for $\tilde{\mu}_{ij}(\beta, x_{ij}, \hat{\psi}(\beta, z_{ij}))$, and $\tilde{\sigma}_{i,jk}(\beta, \rho, \hat{\psi}(\beta))$ for $\tilde{\sigma}_{i,jk}(\beta, \rho, x_{ij}, \hat{\psi}(\beta, z_{ij}))$, and construct the mean vector and covariance matrix of the response vector $y_i = (y_{i1}, \ldots, y_{ij}, \ldots, y_{in_i})'$, as

$$\tilde{\mu}_i(\beta, \hat{\psi}(\beta)) = E[Y_i] = [\tilde{\mu}_{i1}(\beta, \hat{\psi}(\beta)), \ldots, \tilde{\mu}_{ij}(\beta, \hat{\psi}(\beta)), \ldots, \tilde{\mu}_{in_i}(\beta, \hat{\psi}(\beta))]' : n_i \times 1$$
$$\tilde{\Sigma}_i(\beta, \rho, \hat{\psi}(\beta)) = \mathrm{cov}[Y_i] = (\tilde{\sigma}_{i,jk}(\beta, \rho, \hat{\psi}(\beta))) : n_i \times n_i. \tag{45}$$

One may then follow Sutradhar (2003, 2010), for example, and construct the GQL (generalized quasi-likelihood) estimating equation for $\beta$ as

$$\sum_{i=1}^{K} \frac{\partial[\tilde{\mu}_i(\beta, \hat{\psi}(\beta))]'}{\partial\beta} \; [\tilde{\Sigma}_i(\beta, \rho, \hat{\psi}(\beta))]^{-1} \; [y_i - \tilde{\mu}_i(\beta, \hat{\psi}(\beta))] = 0, \tag{46}$$

where

$$\frac{\partial[\tilde{\mu}_i(\beta, \hat{\psi}(\beta))]'}{\partial\beta} = \frac{\partial(\tilde{\mu}_{i1}(\beta, \hat{\psi}(\beta)), \ldots, \tilde{\mu}_{ij}(\beta, \hat{\psi}(\beta)), \ldots, \tilde{\mu}_{in_i}(\beta, \hat{\psi}(\beta)))}{\partial\beta},$$

with

$$\begin{aligned}\frac{\partial\tilde{\mu}_{ij}(\beta, \hat{\psi}(\beta))}{\partial\beta} &= \frac{\partial}{\partial\beta}\left(\exp[x_{ij}'(t_{ij})\beta + \hat{\psi}(\beta, z_{ij})]\right) \\ &= \tilde{\mu}_{ij}(\beta, \hat{\psi}(\beta))\left(x_{ij}(t_{ij}) + \frac{\partial}{\partial\beta}\hat{\psi}(\beta, z_{ij})\right) \\ &= \tilde{\mu}_{ij}(\beta, \hat{\psi}(\beta))\left[x_{ij}(t_{ij}) \right.\\ &\quad \left. - \frac{\sum_{\ell=1}^{K}\sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij}) \exp[x_{\ell u}'(t_{\ell u})\beta] x_{\ell u}(t_{\ell u})}{\sum_{\ell=1}^{K}\sum_{u=1}^{n_\ell} w_{\ell u}(z_{ij}) \exp[x_{\ell u}'(t_{\ell u})\beta]}\right].\end{aligned} \tag{47}$$

Let $\hat{\beta}_{SGQL}$ be the SGQL estimator of $\beta$ obtained as a solution of the SGQL estimating equation (46). For true $\beta$, define

$$\begin{aligned}\bar{f}_K(\beta) = \frac{1}{K}\sum_{i=1}^{K} f_i(\beta) = \frac{1}{K}\sum_{i=1}^{K} \frac{\partial[\tilde{\mu}_i(\beta, \hat{\psi}(\beta))]'}{\partial\beta} \; [\tilde{\Sigma}_i(\beta, \rho, \hat{\psi}(\beta))]^{-1} \\ \times [y_i - \tilde{\mu}_i(\beta, \hat{\psi}(\beta))],\end{aligned} \tag{48}$$

where $y_1, \ldots, y_i, \ldots, y_K$ are independent to each other as they are collected from $K$ independent individuals, but they are not identically distributed because

$$Y_i \sim [\tilde{\mu}_i(\beta, \hat{\psi}(\beta)), \tilde{\Sigma}_i(\beta, \rho, \hat{\psi}(\beta))], \tag{49}$$

where the mean vectors and covariance matrices are different for different individuals. By (49), it follows from (48) that $E[\bar{f}_K(\beta)] = 0$. Also it follows that

$$\begin{aligned} \text{cov}[\bar{f}_K(\beta)] &= \frac{1}{K^2}\sum_{i=1}^{K}\text{cov}[f_i(\beta)] \\ &= \frac{1}{K^2}\sum_{i=1}^{K}\frac{\partial[\tilde{\mu}_i(\beta,\hat{\psi}(\beta))]'}{\partial\beta}\;[\tilde{\Sigma}_i(\beta,\rho,\hat{\psi}(\beta))]^{-1}\frac{\partial\tilde{\mu}_i(\beta,\hat{\psi}(\beta))}{\partial\beta'} \\ &= \frac{1}{K^2}\sum_{i=1}^{K}\tilde{V}_i(\beta,\rho) = \frac{1}{K^2}V_K^*(\beta,\rho). \end{aligned} \tag{50}$$

Next, by applying the Lindeberg-Feller central limit theorem (Amemiya 1985, Theorem 3.3.6; McDonald 2005, Theorem 2.2), one may show following Sutradhar et al. (2015) that

$$\lim_{K\to\infty}\hat{\beta}_{SGQL} \to N(\beta, {V^*}_K^{-1}(\beta,\rho)), \tag{51}$$

that is,

$$||[V_K^*(\beta,\rho)]^{\frac{1}{2}}[\hat{\beta}_{SGQL}-\beta]|| = O_p(\sqrt{p}), \tag{52}$$

where $V_K^*(\beta,\rho)$ is given in (50).

### 3.3 Moment Estimation of Correlation Index Parameter $\rho$

Notice from (37) that the lag 1 correlation of the repeated count responses has the formula

$$\text{corr}[Y_{i,j-1},Y_{ij}] = \rho\sqrt{\frac{\mu_{i,j-1}(\beta,x_{i,j-1},\psi(z_{i,j-1}))}{\mu_{ij}(\beta,x_{ij},\psi(z_{ij}))}}, \tag{53}$$

or equivalently,

$$\text{cov}[Y_{i,j-1},Y_{ij}] = \rho\mu_{i,j-1}(\beta,x_{i,j-1},\psi(z_{i,j-1})). \tag{54}$$

Consequently, to develop a moment equation for $\rho$, one may construct a moment estimating equation by equating the average sample covariance with its population counterpart in (54). Specifically, the moment estimator of $\rho$ has the formula

$$\hat{\rho} = \frac{\sum_{i=1}^{K}\sum_{j=2}^{n_i}\left[\frac{(y_{ij}-\mu_{ij}(x_{ij},\beta,\psi(z_{ij})))}{\sqrt{\mu_{ij}(\beta,x_{ij},\psi(z_{ij}))}}\right]\left[\frac{(y_{i,j-1}-\mu_{i,j-1}(x_{i,j-1},\beta,\psi(z_{i,j-1})))}{\sqrt{\mu_{i,j-1}(\beta,x_{i,j-1},\psi(z_{i,j-1}))}}\right]}{\sum_{i=1}^{K}(n_i-1)} \Bigg/ \frac{\sum_{i=1}^{K}\sum_{j=2}^{n_i}\left[\frac{\sqrt{\mu_{i,j-1}(\beta,x_{i,j-1},\psi(z_{i,j-1}))}}{\sqrt{\mu_{ij}(\beta,x_{i,j},\psi(z_{i,j}))}}\right]}{\sum_{i=1}^{K}(n_i-1)}. \quad (55)$$

Next following (44), we replace the $\psi(z_{ij})$ function involved in (55) with $\hat{\psi}(\beta, z_{ij})$ for known $\beta$, and re-express the estimator of $\rho$ from (55) as

$$\hat{\hat{\rho}} = \frac{\sum_{i=1}^{K}\sum_{j=2}^{n_i}\left[\frac{(y_{ij}-\tilde{\mu}_{ij}(x_{ij},\beta,\hat{\psi}(\beta,z_{ij})))}{\sqrt{\tilde{\mu}_{ij}(\beta,x_{ij},\hat{\psi}(\beta,z_{ij}))}}\right]\left[\frac{(y_{i,j-1}-\tilde{\mu}_{i,j-1}(x_{i,j-1},\beta,\hat{\psi}(\beta,z_{i,j-1})))}{\sqrt{\tilde{\mu}_{i,j-1}(\beta,x_{i,j-1},\hat{\psi}(\beta,z_{i,j-1}))}}\right]}{\sum_{i=1}^{K}(n_i-1)} \Bigg/ \frac{\sum_{i=1}^{K}\sum_{j=2}^{n_i}\left[\frac{\sqrt{\tilde{\mu}_{i,j-1}(\beta,x_{i,j-1},\hat{\psi}(\beta,z_{i,j-1}))}}{\sqrt{\tilde{\mu}_{ij}(\beta,x_{i,j},\hat{\psi}(z_{i,j}))}}\right]}{\sum_{i=1}^{K}(n_i-1)}. \quad (56)$$

Remark that under some mild conditions on the design covariates, one may show that this $\hat{\hat{\rho}}$ in (56) is consistent (Sutradhar et al. 2015) for $\rho$.

## 4 Semi-parametric Multinomial Dynamic Logit Models and Inferences

Suppose that $y_{ij} = (y_{ij1}, \ldots, y_{ijc}, \ldots, y_{ij,C-1})'$ denotes the $(C-1)$-dimensional multinomial response variable ($C = 2$ being the binary case) for the $i$th ($i = 1, \ldots, K$) at time $j$ ($j = 1, \ldots, n_i$), and for $c = 1, \ldots, C-1$,

$$y_{ij}^{(c)} = (y_{ij1}^{(c)}, \ldots, y_{ijc}^{(c)}, \ldots, y_{ij,C-1}^{(c)})' = (01'_{c-1}, 1, 01'_{C-1-c})' \equiv \delta_{ijc} \quad (57)$$

indicates that the multinomial response from the $i$th individual recorded at time $j$ belongs to the $c$th category. For $c = C$, one writes $y_{ij}^{(C)} = \delta_{ijC} = 01_{C-1}$. Note that in the non-stationary case, that is, when covariates are time dependent, one uses the time dependent marginal probabilities. Specifically, suppose that at time point $j$ ($j = 1, \ldots, n_i$), $x_{ij} = (x_{ij1}, \ldots, x_{ij\ell}, \ldots, x_{ij,p+1})'$ denotes the $(p+1)$-dimensional covariate vector and $\beta_c = (\beta_{c0}, \beta_{c1}, \ldots, \beta_{cp})'$ denotes the effect of $x_{ij}$ on $y_{ij}^{(c)}$ for $c = 1, \ldots, C-1$, $i = 1, \ldots, K$, and all $j = 1, \ldots, n_i$. Further suppose that $z_{ij}(j)$ is the secondary covariate collected from the $i$th individual around time $j$. For simplicity, we will denote this covariate value by $z_{ij}$. Next, let $\psi^{(c)}(z_{ij})$ be a non-parametric function which influences the $j$th response of the $i$th individual to be in $c$th category. In such cases, as an extension of the cross-sectional multinomial probability model (6) to the longitudinal case, the semi-parametric multinomial probability at time $j = 1$, may be expressed as

$$P[y_{i1} = y_{i1}^{(c)}] = \pi_{(i1)c} = \begin{cases} \frac{\exp(x'_{i1}\beta_c + \psi^{(c)}(z_{i1}))}{1+\sum_{g=1}^{C-1}\exp(x'_{i1}\beta_g + \psi^{(g)}(z_{i1}))} & \text{for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1}\exp(x'_{i1}\beta_g + \psi^{(g)}(z_{i1}))} & \text{for } c = C, \end{cases} \tag{58}$$

and the elements of $y_{i1} = (y_{i11},\ldots,y_{i1c},\ldots,y_{i1,C-1})'$ at time $j = 1$ follow the multinomial probability distribution given by

$$P[y_{i11},\ldots,y_{i1c},\ldots,y_{i1,C-1}] = \Pi_{c=1}^{C}\pi_{(i1)c}^{y_{i1c}}. \tag{59}$$

In (59), $y_{i1C} = 1 - \sum_{c=1}^{C-1} y_{i1c}$, and $\pi_{i1C} = 1 - \sum_{c=1}^{C-1}\pi_{i1c}$.

Next, for the $i$th individual, we define the transitional probability from the $g$th ($g = 1,\ldots,C$) category at time $j-1$ to the $c$th category at time $j$, given by

$$\begin{aligned} \eta_{ij|j-1}^{(c)}(g) &= P\left(Y_{ij} = y_{ij}^{(c)} \,\middle|\, Y_{i,j-1} = y_{i,j-1}^{(g)}\right) \\ &= \begin{cases} \frac{\exp\left[x'_{ij}\beta_c + \gamma'_c y_{i,j-1}^{(g)} + \psi^{(c)}(z_{ij})\right]}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ij}\beta_v + \gamma'_v y_{i,j-1}^{(g)} + \psi^{(v)}(z_{ij})\right]}, & \text{for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ij}\beta_v + \gamma'_v y_{i,j-1}^{(g)} + \psi^{(v)}(z_{ij})\right]}, & \text{for } c = C, \end{cases} \end{aligned} \tag{60}$$

where $\gamma_c = (\gamma_{c1},\ldots,\gamma_{cv},\ldots,\gamma_{c,C-1})'$ denotes the dynamic dependence parameters. For further notational convenience, we re-express the conditional probabilities in (60) as

$$\eta_{ij|j-1}^{(c)}(g) = \begin{cases} \frac{\exp\left[x'_{ij}\beta_c + \gamma'_c\delta_{i(j-1)g} + \psi^{(c)}(z_{ij})\right]}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ij}\beta_v + \gamma'_v\delta_{i(j-1)g} + \psi^{(v)}(z_{ij})\right]}, & \text{for } c = 1,\ldots,C-1 \\ \frac{1}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ij}\beta_v + \gamma'_v\delta_{i(j-1)g} + \psi^{(v)}(z_{ij})\right]}, & \text{for } c = C, \end{cases} \tag{61}$$

where for $j = 2,\ldots,n_i$, $\delta_{i(j-1)g}$, by (4), has the formula

$$\delta_{i(j-1)g} = \begin{cases} [01'_{g-1}, 1, 01'_{C-1-g}]' & \text{for } g = 1,\ldots,C-1 \\ 01_{C-1} & \text{for } g = C. \end{cases}$$

Let $\beta = (\beta'_1,\ldots,\beta'_c,\ldots,\beta'_{C-1})' : (p+1)(C-1)\times 1$, and $\gamma = (\gamma'_1,\ldots,\gamma'_c,\ldots,\gamma'_{C-1})' : (C-1)^2 \times 1$. Also let $\psi(z) = [\psi^{(1)}(z),\ldots,\psi^{(c)}(z),\ldots,\psi^{(C-1)}(z)]'$ be the $(c-1)\times 1$ vector of non-parametric functions. These parameters and the non-parametric functions are involved in the unconditional mean, variance and covariances of the responses. More specifically one may show (Loredo-Osti and Sutradhar 2012) that

$$E[Y_{ij}] = \bar{\pi}_{(ij)}(\beta,\gamma) = (\bar{\pi}_{(ij)1},\ldots,\bar{\pi}_{(ij)c},\ldots,\bar{\pi}_{(ij)(C-1)})' : (C-1)\times 1 \tag{62}$$

$$= \begin{cases} [\pi_{(i1)1},\ldots,\pi_{(i1)c},\ldots,\pi_{(i1)(C-1)}]' & \text{for } j=1 \\ \eta_{(ij|j-1)}(C) + \left[\eta_{(ij|j-1),M} - \eta_{(ij|j-1)}(C)1'_{C-1}\right]\bar{\pi}_{i(j-1)} & \text{for } j=2,\ldots,n_i-1 \end{cases}$$

$$\begin{aligned} \text{var}[Y_{ij}] &= \text{diag}[\bar{\pi}_{(ij)1},\ldots,\bar{\pi}_{(ij)c},\ldots,\bar{\pi}_{(ij)(C-1)}] - \bar{\pi}_{(ij)}\bar{\pi}'_{(ij)} \\ &= (\text{cov}(Y_{ijc},Y_{ijk})) = (\bar{\sigma}_{i(jj)ck}),\ c,k=1,\ldots,C-1 \\ &= \bar{\Sigma}_{i(jj)}(\beta,\gamma),\ \text{for } j=1,\ldots,n_i \end{aligned} \tag{63}$$

$$\begin{aligned} \text{cov}[Y_{iu},Y_{ij}] &= \Pi_{s=u+1}^{j}\left[\eta_{(is|s-1),M} - \eta_{(is|s-1)}(C)1'_{C-1}\right]\text{var}[Y_{iu}],\ \text{for } u<j, j=2,\ldots,n_i \\ &= (\text{cov}(Y_{iuc},Y_{ijk})) = (\bar{\sigma}_{i(uj)ck}),\ c,k=1,\ldots,C-1 \\ &= \bar{\Sigma}_{i(uj)}(\beta,\gamma), \end{aligned} \tag{64}$$

where

$$\eta_{(is|s-1)}(C) = [\eta^{(1)}_{is|s-1}(C),\ldots,\eta^{(c)}_{is|s-1}(C)\ldots,\eta^{(C-1)}_{is|s-1}(C)]' = \pi_{(is)} : (C-1)\times 1$$

$$\eta_{(is|s-1),M} = \begin{pmatrix} \eta^{(1)}_{is|s-1}(1) & \cdots & \eta^{(1)}_{is|s-1}(g) & \cdots & \eta^{(1)}_{is|s-1}(C-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta^{(c)}_{is|s-1}(1) & \cdots & \eta^{(c)}_{is|s-1}(g) & \cdots & \eta^{(c)}_{is|s-1}(C-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \eta^{(C-1)}_{is|s-1}(1) & \cdots & \eta^{(C-1)}_{is|s-1}(g) & \cdots & \eta^{(C-1)}_{is|s-1}(C-1) \end{pmatrix} : (C-1)\times(C-1).$$

It is of importance to estimate the vector of non-parametric function $\psi(z)$ consistently, for the consistent estimation of the main parameters $\beta$ and $\gamma$. Note that these functions and parameter estimates are needed to understand the aforementioned basic properties including the pair-wise correlations of the multinomial responses in the present semi-parametric setup. Semi-parametric conditional quasi-likelihood and semi-parametric likelihood methods are discussed in the following subsections for the estimation of the desired non-parametric functions and the parameters involved in the specified regression functions.

### 4.1 Semi-parametric Weighted Likelihood Estimation of the Non-parametric Functions $\psi^{(c)}(z)$ for $c = 1,\ldots,C-1$

If the non-parametric functions $\psi^{(c)}(z)$ in (58) and (61) were known, one would then write the likelihood function for $\beta$ and $\gamma$ as

$$\begin{aligned} L(\beta,\gamma) &= \Pi_{i=1}^{K}\left[f(y_{i1})\Pi_{j=2}^{n_i}f(y_{ij}|y_{i,j-1})\right] \\ &= \left[\Pi_{i=1}^{K}f(y_{i1})\right] \end{aligned}$$

$$\begin{aligned} & \times \Pi_{i=1}^{K}\Pi_{j=2}^{n_i}\Pi_{g=1}^{C}\left[f(y_{ij}|y_{i,j-1}^{(g)})\right] \\ = \; & c_0^* \left[\Pi_{i=1}^{K}\Pi_{c=1}^{C}\pi_{(i1)c}^{y_{i1c}}\right] \\ & \times \Pi_{i=1}^{K}\Pi_{j=2}^{n_i}\Pi_{c=1}^{C}\Pi_{g=1}^{C}\left\{\eta_{ij|j-1}^{(c)}(g)\right\}^{y_{ijc}}, \end{aligned} \tag{65}$$

where $c_0^*$ is the normalizing constant free from any parameters, and the formulas for $\pi_{(i1)c}$ and $\eta_{ij|j-1}^{(c)}(g)$ are given in (58) and (61), respectively. However, as the non-parametric functions $\psi^{(c)}(z)$ are unknown, in terms of $\beta$ and $\gamma$, one may use the kernel weights based likelihood function for their estimation. For the estimation of $\psi^{(c)}(z_0)$ for a selected value $z_0$ for the secondary covariate $z_{ij}$, we first write this weighted likelihood function as

$$\begin{aligned} L(\psi(z_0)|\beta,\gamma) = \; & \Pi_{i=1}^{K}\left[\{f(y_{i1})\}^{w_{i1}(z_0)}\Pi_{j=2}^{n_i}\{f(y_{ij}|y_{i,j-1})\}^{w_{ij}(z_0)}\right] \\ = \; & c_0^* \left[\Pi_{i=1}^{K}\Pi_{c=1}^{C}\pi_{(i1)c}^{w_{i1}(z_0)y_{i1c}}\right] \\ & \times \Pi_{i=1}^{K}\Pi_{j=2}^{n_i}\Pi_{c=1}^{C}\Pi_{g=1}^{C}\left\{\eta_{ij|j-1}^{(c)}(g)\right\}^{w_{ij}(z_0)y_{ijc}}, \end{aligned} \tag{66}$$

where the kernel weights $w_{ij}(z_0)$ have the same formula as in (16), that is,

$$w_{ij}(z_0) = \frac{p_{ij}(\frac{z_0 - z_{ij}}{b})}{\sum_{i=1}^{K}\sum_{j=1}^{n_i} p_{ij}(\frac{z_0 - z_{ij}}{b})} \tag{67}$$

with $p_{ij}$ as the kernel density, for example,

$$p_{ij}(\frac{z_0 - z_{ij}}{b}) = \frac{1}{\sqrt{2\pi} b}\exp\left(\frac{-1}{2}(\frac{z_0 - z_{ij}}{b})^2\right).$$

It then follows from (66) that the log likelihood function has the form

$$\begin{aligned} Log\; L(\psi(z_0)|\beta,\gamma) = \; & log\; c_0^* + \sum_{i=1}^{K}\sum_{c=1}^{C} w_{i1}(z_0) y_{i1c} log\; \pi_{(i1)c} \\ & + \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C} w_{ij}(z_0)\left[y_{ijc} log\; \eta_{ij|j-1}^{(c)}(g)\right], \end{aligned} \tag{68}$$

yielding the likelihood estimating equation for $\psi(z_0)$ as

$$\frac{\partial Log\ L(\psi(z_0)|\beta,\gamma)}{\partial \psi(z_0)} = \sum_{i=1}^{K}\sum_{c=1}^{C} w_{i1}(z_0)\left[\frac{y_{i1c}}{\pi_{(i1)c}}\frac{\partial \pi_{(i1)c}}{\partial \psi(z_0)}\right] + \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C} w_{ij}(z_0)\left[\frac{y_{ijc}}{\eta^{(c)}_{ij|j-1}(g)}\frac{\partial \eta^{(c)}_{ij|j-1}(g)}{\partial \psi(z_0)}\right] = 0. \quad (69)$$

We denote the solution of (69) as $\hat{\psi}(\beta,\gamma;z_0)$.

Let

$$\begin{aligned}
\tilde{\pi}_{(i1)c} &= \left[\pi_{(i1)c}\right]_{\psi^{(c)}(z_0)=\hat{\psi}^{(c)}(\beta,\gamma;z_0)}\\
\tilde{\eta}^{(c)}_{ij|j-1}(g) &= \left[\eta^{(c)}_{ij|j-1}(g)\right]_{\psi^{(c)}(z_0)=\hat{\psi}^{(c)}(\beta,\gamma;z_0)}\\
\pi^{*}_{(i1)c} &= \left[\frac{\partial \pi_{(i1)c}}{\partial \psi(z_0)}\right]_{\psi^{(c)}(z_0)=\hat{\psi}^{(c)}(\beta,\gamma;z_0)}\\
\eta^{*(c)}_{ij|j-1}(g) &= \left[\frac{\partial \eta^{(c)}_{ij|j-1}(g)}{\partial \psi(z_0)}\right]_{\psi^{(c)}(z_0)=\hat{\psi}^{(c)}(\beta,\gamma;z_0)},
\end{aligned} \quad (70)$$

for all $c = 1,\ldots,C-1$. Use them in (69) and write

$$f(\hat{\psi}(\beta,\gamma;z_0)) = 0, \quad (71)$$

which can be exploited to derive

$$\frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_0)}{\partial \beta} \text{ and } \frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_0)}{\partial \gamma}, \quad (72)$$

for all $c = 1,\ldots,C-1$, for the purpose of the construction of the estimating equations for $\beta$ and $\gamma$ as shown in the next two sections.

In (69),

$$\begin{aligned}
\frac{\partial \pi_{(i1)c}}{\partial \psi^{(c)}(z_0)} &= \pi_{(i1)c}[1-\pi_{(i1)c}]\\
\frac{\partial \pi_{(i1)c}}{\partial \psi^{(k)}(z_0)} &= -[\pi_{(i1)c}\pi_{(i1)k}],
\end{aligned} \quad (73)$$

yielding

$$\frac{\partial \pi_{(i1)c}}{\partial \psi(z_0)} = \begin{pmatrix} -\pi_{(i1)1}\pi_{(i1)c} \\ \vdots \\ \pi_{(i1)c}[1-\pi_{(i1)c}] \\ \vdots \\ -\pi_{(i1)(C-1)}\pi_{(i1)c} \end{pmatrix} : (C-1)\times 1$$
$$= \left[\pi_{(i1)c}(\delta_{i(1)c} - \pi_{(i1)})\right], \tag{74}$$

with

$$\delta_{i(1)c} = \begin{cases} [01'_{c-1}, 1, 01'_{C-1-c}]' & \text{for } c = 1, \ldots, C-1;\ i = 1, \ldots, K \\ 01_{C-1} & \text{for } c = C;\ i = 1, \ldots, K. \end{cases}$$

Similarly, for $j = 2, \ldots, n_i$, it follows from (61) that

$$\frac{\partial \eta^{(c)}_{ij|j-1}(g)}{\partial \psi^{(c)}(z_0)} = \eta^{(c)}_{ij|j-1}(g)[1 - \eta^{(c)}_{ij|j-1}(g)]$$
$$\frac{\partial \eta^{(c)}_{ij|j-1}(g)}{\partial \psi^{(k)}(z_0)} = -[\eta^{(c)}_{ij|j-1}(g)\eta^{(k)}_{ij|j-1}(g)], \tag{75}$$

yielding

$$\frac{\partial \eta^{(c)}_{ij|j-1}(g)}{\partial \psi(z_0)} = \begin{pmatrix} -\eta^{(1)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g) \\ \vdots \\ \eta^{(c)}_{ij|j-1}(g)[1-\eta^{(c)}_{ij|j-1}(g)] \\ \vdots \\ -\eta^{(C-1)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g) \end{pmatrix} : (C-1)\times 1$$
$$= \left[\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c} - \eta_{ij|j-1}(g))\right], \tag{76}$$

where

$$\eta_{ij|j-1}(g) = [\eta^{(1)}_{ij|j-1}(g), \ldots, \eta^{(c)}_{ij|j-1}(g), \ldots, \eta^{(C-1)}_{ij|j-1}(g)]'.$$

Thus, the semi-parametric likelihood equation in (69) has the computational formula

$$\frac{\partial Log\ L(\psi(z_0)|\beta,\gamma)}{\partial \psi(z_0)} = \sum_{i=1}^{K}\sum_{c=1}^{C} w_{i1}(z_0)\frac{y_{i1c}}{\pi_{(i1)c}}\left[\left\{\pi_{(i1)c}(\delta_{i(1)c} - \pi_{(i1)})\right\}\right]$$

$$+\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C} w_{ij}(z_0)\frac{y_{ijc}}{\eta^{(c)}_{ij|j-1}(g)}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]=0. \quad (77)$$

For given $\beta$ and $\gamma$, the likelihood equations in (77) may be solved iteratively by using the iterative equations for $\psi(z_0)$ given by

$$\begin{aligned}\hat{\psi}(z_0)(r+1) &= \hat{\psi}(z_0)(r) - \left[\left\{\frac{\partial^2 \text{Log } L(\psi(z_0)|\beta,\gamma)}{\partial\psi'(z_0)\partial\psi(z_0)}\right\}^{-1}\right. \\ &\quad \times \left.\frac{\partial \text{Log } L(\psi(z_0)|\beta,\gamma)}{\partial\psi(z_0)}\right]_{|\psi(z_0)=\hat{\psi}(z_0)(r)} ;\ (C-1)\times 1, \quad (78)\end{aligned}$$

where the formula for the second order derivative matrix $\frac{\partial^2 Log\ L(\psi(z_0)|\beta,\gamma)}{\partial\psi(z_0)'\partial\psi(z_0)}$ may be derived by taking the derivative of the $(C-1)\times 1$ vector with respect to $\psi'(z_0)$. The exact derivative has a complicated formula. We provide an approximation first and then give the exact formula for the sake of completeness.

**An approximation for $\frac{\partial^2 Log\ L(\psi(z_0)|\beta,\gamma)}{\partial\psi(z_0)'\partial\psi(z_0)}$ based on iteration principle:**

In this approach, one assumes that $\psi(z_0)$ in the derivatives in (69), that is, $\psi(z_0)$ involved in $\frac{\partial\pi_{(i1)c}}{\partial\psi(z_0)}$ and $\frac{\partial\eta^{(c)}_{ij|j-1}(g)}{\partial\psi(z_0)}$ are known from a previous iteration, and then take the derivative of $\frac{\partial Log\ L(\psi(z_0)|\beta,\gamma)}{\partial\psi(z_0)}$ in (69) or (77), with respect to $\psi'(z_0)$. This provides a simpler formula for the second order derivative as

$$\begin{aligned}\frac{\partial^2 Log\ L(\psi(z_0)|\beta,\gamma)}{\partial\psi(z_0)'\partial\psi(z_0)} &= -\sum_{i=1}^{K}\sum_{c=1}^{C}\frac{y_{i1c}}{[\pi_{(i1)c}]^2}\left[\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\}\right]\left[\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\}\right]' \\ &\quad -\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C}\left[\frac{y_{ijc}}{[\eta^{(c)}_{ij|j-1}(g)]^2}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]\right. \\ &\quad \times\left.\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]'\right] : (C-1)\times(C-1). \quad (79)\end{aligned}$$

**Exact formula for $\frac{\partial^2 Log\ L(\psi(z_0)|\beta,\gamma)}{\partial\psi'(z_0)\partial\psi(z_0)}$:**

Here it is assumed that $\psi(z_0)$ in the derivatives in (69), that is, $\psi(z_0)$ involved in $\frac{\partial\pi_{(i1)c}}{\partial\psi(z_0)}$ and $\frac{\partial\eta^{(c)}_{ij|j-1}(g)}{\partial\psi(z_0)}$ are unknown, implying that the second order derivatives of these quantities cannot be zero. Hence, in stead of (79), one obtains

$$\begin{aligned}\frac{\partial^2 Log\ L(\psi(z_0)|\beta,\gamma)}{\partial\psi'(z_0)\partial\psi(z_0)} &= -\sum_{i=1}^{K}\sum_{c=1}^{C}\frac{y_{i1c}}{[\pi_{(i1)c}]^2}\left[\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\}\right]\left[\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\}\right]' \\ &\quad +\sum_{i=1}^{K}\sum_{c=1}^{C}\frac{y_{i1c}}{\pi_{(i1)c}}\frac{\partial}{\partial\psi'(z_0)}\left[\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\}\right]\end{aligned}$$

$$-\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C}\left[\frac{y_{ijc}}{[\eta^{(c)}_{ij|j-1}(g)]^2}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]\right.$$
$$\left.\times\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]'\right]$$
$$+\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C}\frac{y_{ijc}}{\eta^{(c)}_{ij|j-1}(g)}\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right], \tag{80}$$

where $\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\right\}\right]$ and $\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]$ are computed as follows.

**Computation of $\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\right\}\right]$:**
By (74),

$$\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\pi_{(i1)c}(\delta_{i(1)c}-\pi_{(i1)})\right\}\right]$$
$$=\frac{\partial}{\partial\psi'(z_0)}\left[\begin{pmatrix}-\pi_{(i1)1}\pi_{(i1)c}\\ \vdots\\ \pi_{(i1)c}[1-\pi_{(i1)c}]\\ \vdots\\ -\pi_{(i1)(C-1)}\pi_{(i1)c}\end{pmatrix}\right]$$
$$=\begin{pmatrix}\left[-\pi_{(i1)1}\pi_{(i1)c}\left\{\left(\delta_{i(1)c}+\delta_{i(1)1}-2\pi_{(i1)}\right)'\right\}\right]\\ \left[-\pi_{(i1)2}\pi_{(i1)c}\left\{\left(\delta_{i(1)c}+\delta_{i(1)2}-2\pi_{(i1)}\right)'\right\}\right]\\ \vdots\\ \left[\pi_{(i1)c}(1-2\pi_{(i1)c})\left\{\left(\delta_{i(1)c}-\pi_{(i1)}\right)'\right\}\right]\\ \vdots\\ \left[-\pi_{(i1)(C-1)}\pi_{(i1)c}\left\{\left(\delta_{i(1)c}+\delta_{i(1)(C-1)}-2\pi_{(i1)}\right)'\right\}\right]\end{pmatrix}. \tag{81}$$

**Computation of $\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]$:**
By (76),

$$\frac{\partial}{\partial\psi'(z_0)}\left[\left\{\eta^{(c)}_{ij|j-1}(g)(\delta_{i(j-1)c}-\eta_{ij|j-1}(g))\right\}\right]$$

$$= \frac{\partial}{\partial \psi'(z_0)} \left[ \begin{pmatrix} -\eta^{(1)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g) \\ \vdots \\ \eta^{(c)}_{ij|j-1}(g)[1-\eta^{(c)}_{ij|j-1}(g)] \\ \vdots \\ -\eta^{(C-1)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g) \end{pmatrix} \right]$$

$$= \begin{pmatrix} \left[-\eta^{(1)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g)\left\{\left(\delta_{i(j-1)c}+\delta_{i(j-1)1}-2\eta_{ij|j-1}(g)\right)'\right\}\right] \\ \left[-\eta^{(2)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g)\left\{\left(\delta_{i(j-1)c}+\delta_{i(j-1)2}-2\eta_{ij|j-1}(g)\right)'\right\}\right] \\ \vdots \\ \left[\eta^{(c)}_{ij|j-1}(g)(1-2\eta^{(c)}_{ij|j-1}(g))\left\{\left(\delta_{i(j-1)c}-\eta_{ij|j-1}(g)\right)'\right\}\right] \\ \vdots \\ \left[-\eta^{(C-1)}_{ij|j-1}(g)\eta^{(c)}_{ij|j-1}(g)\left\{\left(\delta_{i(j-1)c}+\delta_{i(j-1)(C-1)}-2\eta_{ij|j-1}(g)\right)'\right\}\right] \end{pmatrix}. \quad (82)$$

## 4.2 Likelihood Estimation for the Regression Effects β

Recall from (70) and by (58) write

$$\begin{aligned} \tilde{\pi}_{(i1)c} &= \left[\pi_{(i1)c}\right]_{\psi^{(c)}(z_{i1})=\hat{\psi}^{(c)}(\beta,\gamma;z_{i1})} \\ &= \begin{cases} \frac{\exp(x'_{i1}\beta_c+\hat{\psi}^{(c)}(\beta,\gamma;z_{i1}))}{1+\sum_{g=1}^{C-1}\exp(x'_{i1}\beta_g+\hat{\psi}^{(g)}(\beta,\gamma;z_{i1}))} & \text{for } c=1,\ldots,C-1 \\ \frac{1}{1+\sum_{g=1}^{C-1}\exp(x'_{i1}\beta_g+\hat{\psi}^{(g)}(\beta,\gamma;z_{i1}))} & \text{for } c=C, \end{cases} \end{aligned} \quad (83)$$

Similarly, by (70) and (61), one writes

$$\begin{aligned} \tilde{\eta}^{(c)}_{ij|j-1}(g) &= \left[\eta^{(c)}_{ij|j-1}(g)\right]_{\psi^{(c)}(z_{ij})=\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})} \\ &= \begin{cases} \frac{\exp\left[x'_{ij}\beta_c+\gamma'_c\delta_{i(j-1)g}+\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})\right]}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ij}\beta_v+\gamma'_v\delta_{i(j-1)g}+\hat{\psi}^{(v)}(\beta,\gamma;z_{ij})\right]}, & \text{for } c=1,\ldots,C-1 \\ \frac{1}{1+\sum_{v=1}^{C-1}\exp\left[x'_{ij}\beta_v+\gamma'_v\delta_{i(j-1)g}+\hat{\psi}^{(v)}(\beta,\gamma;z_{ij})\right]}, & \text{for } c=C, \end{cases} \end{aligned} \quad (84)$$

It then follows from (65) that for known $\hat{\psi}^{(c)}(\beta, \gamma; z_{ij})$, the likelihood function for $\beta$ and $\gamma$ has the new formula given by

$$\begin{aligned} L(\beta, \gamma) &= \Pi_{i=1}^{K} \left[ f(y_{i1}) \Pi_{j=2}^{n_i} f(y_{ij}|y_{i,j-1}) \right] \\ &= c_0^* \left[ \Pi_{i=1}^{K} \Pi_{c=1}^{C} \tilde{\pi}_{(i1)c}^{y_{i1c}} \right] \\ &\quad \times \Pi_{i=1}^{K} \Pi_{j=2}^{n_i} \Pi_{c=1}^{C} \Pi_{g=1}^{C} \left\{ \tilde{\eta}_{ij|j-1}^{(c)}(g) \right\}^{y_{ijc}}, \end{aligned} \tag{85}$$

yielding the likelihood equation for $\beta$ as

$$\begin{aligned} \frac{\partial Log\ L(\beta, \gamma)}{\partial \beta} &= \sum_{i=1}^{K} \sum_{c=1}^{C} \frac{y_{i1c}}{\tilde{\pi}_{(i1)c}} \frac{\partial \tilde{\pi}_{(i1)c}}{\partial \beta} \\ &\quad + \sum_{i=1}^{K} \sum_{j=2}^{n_i} \sum_{c=1}^{C} \sum_{g=1}^{C} \left[ \frac{y_{ijc}}{\tilde{\eta}_{ij|j-1}^{(c)}(g)} \frac{\partial \tilde{\eta}_{ij|j-1}^{(c)}(g)}{\partial \beta} \right] = 0, \end{aligned} \tag{86}$$

where by (83)

$$\begin{aligned} \frac{\partial \pi_{(i1)c}}{\partial \beta_c} &= \tilde{\pi}_{(i1)c}[1 - \tilde{\pi}_{(i1)c}] \left[ x_{i1} + \frac{\partial \hat{\psi}^{(c)}(\beta, \gamma; z_{i1})}{\partial \beta_c} \right] \\ \frac{\partial \tilde{\pi}_{(i1)c}}{\partial \beta_k} &= -[\tilde{\pi}_{(i1)c} \tilde{\pi}_{(i1)k}] \left[ x_{i1} + \frac{\partial \hat{\psi}^{(c)}(\beta, \gamma; z_{i1})}{\partial \beta_k} \right], \end{aligned} \tag{87}$$

yielding

$$\begin{aligned} \frac{\partial \tilde{\pi}_{(i1)c}}{\partial \beta} &= \begin{pmatrix} -\tilde{\pi}_{(i1)1} \tilde{\pi}_{(i1)c} \\ \vdots \\ \tilde{\pi}_{(i1)c}[1 - \tilde{\pi}_{(i1)c}] \\ \vdots \\ -\tilde{\pi}_{(i1)(C-1)} \pi_{(i1)c} \end{pmatrix} \otimes x_{i1} \\ &\quad + \begin{pmatrix} -\tilde{\pi}_{(i1)1} \tilde{\pi}_{(i1)c} d_{c1}(1) \\ \vdots \\ \tilde{\pi}_{(i1)c}[1 - \tilde{\pi}_{(i1)c}] d_{cc}(1) \\ \vdots \\ -\tilde{\pi}_{(i1)(C-1)} \pi_{(i1)c} d_{c,C-1}(1) \end{pmatrix} : (C-1)(p+1) \times 1 \end{aligned} \tag{88}$$

where for a given $c$, $d_{ck}(1)$ is the $(p+1) \times 1$ derivative vector defined as $\frac{\partial \hat{\psi}^{(c)}(\beta, \gamma; z_{i1})}{\partial \beta_k}$ for all $k = 1, \ldots, c, \ldots, C-1$. These vectors are derived from (71) to (72). The computational formulas are skipped for convenience.

Next, for $j = 2, \ldots, n_i$, it follows from (84) that

$$\frac{\partial \tilde{\eta}^{(c)}_{ij|j-1}(g)}{\partial \beta_c} = \tilde{\eta}^{(c)}_{ij|j-1}(g)[1 - \tilde{\eta}^{(c)}_{ij|j-1}(g)] \left[ x_{ij} + + \frac{\partial \hat{\psi}^{(c)}(\beta, \gamma; z_{ij})}{\partial \beta_c} \right]$$
$$\frac{\partial \tilde{\eta}^{(c)}_{ij|j-1}(g)}{\partial \beta_k} = -[\tilde{\eta}^{(c)}_{ij|j-1}(g)\tilde{\eta}^{(k)}_{ij|j-1}(g)] \left[ x_{ij} + + \frac{\partial \hat{\psi}^{(c)}(\beta, \gamma; z_{ij})}{\partial \beta_k} \right], \quad (89)$$

yielding

$$\frac{\partial \tilde{\eta}^{(c)}_{ij|j-1}(g)}{\partial \beta} = \begin{pmatrix} -\tilde{\eta}^{(1)}_{ij|j-1}(g)\tilde{\eta}^{(c)}_{ij|j-1}(g) \\ \vdots \\ \tilde{\eta}^{(c)}_{ij|j-1}(g)[1 - \tilde{\eta}^{(c)}_{ij|j-1}(g)] \\ \vdots \\ -\tilde{\eta}^{(C-1)}_{ij|j-1}(g)\tilde{\eta}^{(c)}_{ij|j-1}(g) \end{pmatrix} \otimes x_{ij}$$
$$+ \begin{pmatrix} -\tilde{\eta}^{(1)}_{ij|j-1}(g)\tilde{\eta}^{(c)}_{ij|j-1}(g)d_{c1}(j) \\ \vdots \\ \tilde{\eta}^{(c)}_{ij|j-1}(g)[1 - \tilde{\eta}^{(c)}_{ij|j-1}(g)]d_{cc}(j) \\ \vdots \\ -\tilde{\eta}^{(C-1)}_{ij|j-1}(g)\tilde{\eta}^{(c)}_{ij|j-1}(g)d_{c,C-1}(j) \end{pmatrix} : (C-1)(p+1) \times 1 \quad (90)$$

where for a given $c$, following (71)–(72), $d_{ck}(j)(j = 2, \ldots, n_i)$ is the $(p + 1) \times 1$ derivative vector defined as $\frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial \beta_k}$ for all $k = 1, \ldots, c, \ldots, C-1$.

By using (90) and (88) into (86), for known $\gamma$, one may then solve this likelihood estimating equation (86) for $\beta$ by applying the iterative equation

$$\hat{\beta}(r+1) = \hat{\beta}(r) - \left[ \left\{ \frac{\partial^2 \text{Log } L(\beta,\gamma)}{\partial \beta' \partial \beta} \right\}^{-1} \frac{\partial \text{Log } L(\beta,\gamma_M)}{\partial \beta} \right]_{|\beta=\hat{\beta}(r)}; \ (C-1)(p+1) \times 1, \quad (91)$$

where the formula for the second order derivative matrix $\frac{\partial^2 Log\ L(\beta,\gamma)}{\partial \beta' \partial \beta}$ may be derived by taking the derivative of the $(C-1)(p+1) \times 1$ vector with respect to $\beta'$. The exact derivative has a complicated formula. Similar to (79), we provide an approximation as follows.

**An approximation for $\frac{\partial^2 Log\ L(\beta,\gamma)}{\partial \beta' \partial \beta}$ based on iteration principle:**

In this approach, one assumes that $\beta$ in the derivatives in (86), that is, $\beta$ involved in $\frac{\partial \tilde{\pi}_{(i1)c}}{\partial \beta}$ and $\frac{\partial \tilde{\eta}^{(c)}_{ij|j-1}(g)}{\partial \beta}$ are known from a previous iteration, and then take the derivative of $\frac{\partial Log\ L(\beta,\gamma)}{\partial \beta}$ in (86) or (91), with respect to $\beta'$. This provides a simpler formula for the second order derivative as

$$\frac{\partial^2 Log\ L(\beta,\gamma)}{\partial\beta'\partial\beta} = -\sum_{i=1}^{K}\sum_{c=1}^{C}\frac{y_{i1c}}{[\tilde{\pi}_{(i1)c}]^2}\left[\frac{\partial\tilde{\pi}_{(i1)c}}{\partial\beta}\right]\left[\frac{\partial\tilde{\pi}_{(i1)c}}{\partial\beta}\right]' - \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{c=1}^{C}\sum_{g=1}^{C}\left[\frac{y_{ijc}}{[\tilde{\eta}^{(c)}_{ij|j-1}(g)]^2}\left[\frac{\partial\tilde{\eta}^{(c)}_{ij|j-1}(g)}{\partial\beta}\right] \times\left[\frac{\partial\tilde{\eta}^{(c)}_{ij|j-1}(g)}{\partial\beta}\right]'\right] : (J-1)(p+1)\times(C-1)(p+1). \quad (92)$$

## 4.3 Likelihood Estimation for the Dynamic Dependence Parameters $\gamma$

Consider $\gamma^*$ as

$$\gamma^* = (\gamma_1',\ldots,\gamma_c',\ldots,\gamma_{C-1}')' : (C-1)^2\times 1;\ \text{with}\ \gamma_c = (\gamma_{c1},\ldots,\gamma_{ch},\ldots,\gamma_{c,C-1})' \quad (93)$$

as the $(C-1)\times 1$ vector of dynamic dependence parameters involved in the conditional multinomial logit function in (84). Similar to (89), one obtains

$$\frac{\partial\tilde{\eta}^{(h)}_{ij|j-1}(g)}{\partial\gamma_c} = \begin{cases} \tilde{\eta}^{(c)}_{ij|j-1}(g)[1-\tilde{\eta}^{(c)}_{ij|j-1}(g)]\left[\delta_{i(j-1)g} + +\frac{\partial\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\right] \\ \text{for } h=c; h,c=1,\ldots,C-1 \\ -\tilde{\eta}^{(c)}_{ij|j-1}(g)\tilde{\eta}^{(h)}_{ij|j-1}(g)\left[\delta_{i(j-1)g} + +\frac{\partial\hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\right] \\ \text{for } h\neq c; h,c=1,\ldots,C-1 \\ -\tilde{\eta}^{(c)}_{ij|j-1}(g)\tilde{\eta}^{(C)}_{ij|j-1}(g)\left[\delta_{i(j-1)g} + +\frac{\partial\hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\right] \\ \text{for } h=C;\ c=1,\ldots,C-1, \end{cases} \quad (94)$$

for all $g = 1,\ldots,J$. Using these derivatives, it follows from the likelihood function (85) that

$$\begin{aligned}\frac{\partial Log\ L(\beta,\gamma)}{\partial\gamma_c} &= \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{h=1}^{C}\sum_{g=1}^{C}\left[\frac{y_{ijh}}{\tilde{\eta}^{(h)}_{ij|j-1}(g)}\frac{\partial\tilde{\eta}^{(h)}_{ij|j-1}(g)}{\partial\gamma_c}\right] \\ &= \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C} y_{ijc}[1-\tilde{\eta}^{(c)}_{ij|j-1}(g)]\left[\delta_{i(j-1)g} + +\frac{\partial\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\right] \\ &\quad -\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C}\sum_{h\neq c}^{C}\frac{y_{ijh}}{\tilde{\eta}^{(h)}_{ij|j-1}(g)}\left(\tilde{\eta}^{(c)}_{ij|j-1}(g)\tilde{\eta}^{(h)}_{ij|j-1}(g)\right)\left[\delta_{i(j-1)g} + +\frac{\partial\hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\right]\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C} y_{ijc}\left[\delta_{i(j-1)g} + \frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial \gamma_c}\right] \\
&\quad - \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C}\sum_{h=1}^{C} y_{ijh}\left(\tilde{\eta}^{(c)}_{ij|j-1}(g)\right)\left[\delta_{i(j-1)g} + + \frac{\partial \hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial \gamma_c}\right] \\
&= 0, \qquad (95)
\end{aligned}
$$

for $c = 1, \ldots, C-1$, leading to the estimating equations for the elements of $\gamma^* = (\gamma_1', \ldots, \gamma_c', \ldots, \gamma_{C-1}')'$ as

$$
\frac{\partial Log\ L(\beta,\gamma)}{\partial \gamma^*} = \begin{pmatrix} \frac{\partial Log\ L(\beta,\gamma)}{\partial \gamma_1} \\ \vdots \\ \frac{\partial Log\ L(\beta,\gamma)}{\partial \gamma_c} \\ \vdots \\ \frac{\partial Log\ L(\beta,\gamma)}{\partial \gamma_{C-1}} \end{pmatrix} = 0 : (C-1)^2 \times 1. \qquad (96)
$$

One may solve this likelihood equation (96) for $\gamma^*$ by using the iterative equation

$$
\hat{\gamma}^*(r+1) = \hat{\gamma}^*(r) - \left[\left\{\frac{\partial^2 \text{Log}\ L(\beta,\gamma)}{\partial \gamma^* \partial \gamma^{*\prime}}\right\}^{-1} \frac{\partial \text{Log}\ L(\beta,\gamma)}{\partial \gamma^*}\right]_{|\gamma^* = \hat{\gamma}^*(r)}, \qquad (97)
$$

where the $(C-1)^2 \times (C-1)^2$ second derivative matrix is computed by using the formulas

$$
\begin{aligned}
\frac{\partial^2 \text{Log}\ L(\beta,\gamma)}{\partial \gamma_c \partial \gamma_c'} &= \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C} y_{ijc}\left[\frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial \gamma_c}\frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial \gamma_c'}\right] \\
&\quad - \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C}\sum_{h=1}^{C} y_{ijh}\left(\tilde{\eta}^{(c)}_{ij|j-1}(g)\{1-\tilde{\eta}^{(c)}_{ij|j-1}(g)\}\right)\left[\delta_{i(j-1)g} + \frac{\partial \hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial \gamma_c}\right] \\
&\quad \times \left[\delta_{i(j-1)g} + \frac{\partial \hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial \gamma_c}\right]' \\
&\quad - \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C}\sum_{h=1}^{C} y_{ijh}\left(\tilde{\eta}^{(c)}_{ij|j-1}(g)\right)\left[\frac{\partial \hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial \gamma_c}\frac{\partial \hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial \gamma_c'}\right] \qquad (98)
\end{aligned}
$$

for all $c = 1, \ldots, C-1$, and

$$\frac{\partial^2 \text{Log } L(\beta,\gamma)}{\partial\gamma_c\partial\gamma_k'} = \sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C} y_{ijc}\left[\frac{\partial\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\frac{\partial\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial\gamma_k'}\right]$$
$$+\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C}\sum_{h=1}^{C} y_{ijh}\left(\tilde{\eta}^{(c)}_{ij|j-1}(g)\tilde{\eta}^{(k)}_{ij|j-1}(g)\right)\left[\delta_{i(j-1)g}+\frac{\partial\hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\right]$$
$$\times\left[\delta_{i(j-1)g}+\frac{\partial\hat{\psi}^{(c)}(\beta,\gamma;z_{ij})}{\partial\gamma_k}\right]'$$
$$-\sum_{i=1}^{K}\sum_{j=2}^{n_i}\sum_{g=1}^{C}\sum_{h=1}^{C} y_{ijh}\left(\tilde{\eta}^{(c)}_{ij|j-1}(g)\right)\left[\frac{\partial\hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial\gamma_c}\frac{\partial\hat{\psi}^{(h)}(\beta,\gamma;z_{ij})}{\partial\gamma_k'}\right] \quad (99)$$

for all $c \neq k; c, k = 1, \ldots, C-1$.

## 5 Concluding Remarks

The parametric correlation models for repeated binary and count data are discussed in Sutradhar (2011) and similar parametric correlation models for categorical/multinomial data are provided in Sutradhar (2014). All of these correlation models are developed based on suitable dynamic relationship among the repeated responses. In this paper we have extended these correlation models to the longitudinal semi-parametric setup, where the new regression function involved in the mean and variance of the responses contains a non-parametric function on top of the traditional pre-specified regression function. This makes the estimation of the parameters involved in the pre-specified regression function complicated as the estimate of the non-parametric function contains these parameters. A weighted QL (WQL) approach is used for the estimation of the non-parametric functions for the linear and count models, whereas a weighted ML (WML) approach is used for the multinomial model (binary being the special case). The pre-specified regression functions involved in the linear and count models are estimated using the GQL (generalized quasi-likelihood) and MM (method of moments) approaches, whereas similar parameters involved in the binary and multinomial models are estimated using the likelihood approach. The estimating equations are developed in details which may be conveniently used by practitioners to fit longitudinal linear, count, binary and multinomial data in the semi-parametric setup. The semi-parametric model developed for the categorical data may also be extended to the ordinal categorical data setup.

**Acknowledgements** The author thanks the audience for their comments and suggestions.

## References

Amemiya, T.: Advanced Econometrics. Harvard University Press, Cambridge (1985)
Carota, C., Parmigiani, G.: Semiparametric regression for count data. Biometrika **89**, 265–281 (2002)
Horowitz, J.L.: Semiparametric and Nonparametric Methods in Econometrics. Springer, New York (2009)
Lin, X., Carroll, R.J.: Semiparametric regression for clustered data using generalized estimating equations. J. Am. Stat. Assoc. **96**, 1045–1056 (2001)
Loredo-Osti, J.C., Sutradhar, B.C.: Estimation of regression and dynamic dependence parameters for non-stationary multinomial time series. J. Time Series Anal. **33**, 458–467 (2012)
McDonald, D.R.: The local limit theorem: a historical perspective. J. Iranian Stat. Soc. **4**, 73–86 (2005)
Severini, T.A., Staniswallis, J.G.: Quasi-likelihood estimation in semiparametric models. J. Am. Stat. Assoc. **89**, 501–511 (1994)
Sneddon, G., Sutradhar, B.C.: On semiparametric familial-longitudinal models. Stat. Probab. Lett. **69**, 369–379 (2004)
Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. Stat. Sci. **18**, 377–393 (2003)
Sutradhar, B.C.: Inferences in generalized linear longitudinal mixed models. Can. J. Stat. Special issue **38**, 174–196 (2010)
Sutradhar, B.C.: Dynamic Mixed Models for Familial Longitudinal Data. Springer, New York (2011)
Sutradhar, B.C.: Longitudinal Categorical Data Analysis. Springer, New York (2014)
Sutradhar, B.C., Warriyar, K.V., Zheng, N.: Inferences in semi-parametric dynamic models for repeated count data. Aust N Z J Stat. (2016)
Warriyar, K.V., Sutradhar, B.C.: Estimation with improved efficiency in semi-parametric linear longitudinal models. Braz. J. Probab. Stat. **28**, 561–586 (2014)
You, J., Chen, G.: Semiparametric generalized least squares estimation in partially linear regression models with correlated errors. J. Stat. Plan. Infer. **137**, 117–132 (2007)
Zeger, S.L., Diggle, P.J.: Semi-parametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters. Biometrics **50**, 689–699 (1994)

# Part IV
# An Extension of the GQL Estimation Approach for Longitudinal Data Analysis

# Penalized Generalized Quasi-Likelihood Based Variable Selection for Longitudinal Data

**Tharshanna Nadarajah, Asokan Mulayath Variyath, and J. Concepción Loredo-Osti**

**Abstract** High-dimensional longitudinal data with a large number of covariates, have become increasingly common in many bio-medical applications. The identification of a sub-model that adequately represents the data is necessary for easy interpretation. Also, the inclusion of redundant variables may hinder the accuracy and efficiency of estimation and inference. The joint likelihood function for longitudinal data is challenging, particularly in correlated discrete data. To overcome this problem Wang et al. (Biometrics 68:353–360, 2012) introduced penalized GEEs (PGEEs) with a non-convex penalty function which requires only the first two marginal moments and a working correlation matrix. This method works reasonably well in high-dimensional problems; however, there is a risk of model mis-specification such as variance function and correlation structure and in such situations, we propose variable selection based on penalized generalized quasi-likelihood (PGQL). Simulation studies show that when model assumptions are true, the PGQL method has performance comparable with that of PGEEs. However, when the model is mis-specified, the PGQL method has clear advantages over the PGEEs method. We have implemented the proposed method in a real case example.

**Keywords** GEEs • Generalized quasi-likelihood • Longitudinal data • Variable selection

## 1 Introduction

Variable selection is an important topic in statistical modeling, especially for longitudinal data with high-dimensional covariates, which often arise in large-scale bio-medical studies. In practice, a large number of covariates, $(X_1, X_2, \ldots, X_p)$, are believed to have an influence on the response variable $y$ of interest. However, some covariates have no influence or a weak influence, and a regression model that includes all the covariates is not advisable. Excluding the unimportant covariates

T. Nadarajah (✉) • A.M. Variyath • J. C. Loredo-Osti
Memorial University, St. John's, NL, Canada A1C5S7
e-mail: nadarajah.tharshana@mun.ca; variyath@mun.ca; jcloredoosti@mun.ca



B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6_8

results in a simpler model with better interpretive and predictive value. The problem of identifying a sub-model that adequately represents the response is generally referred to as the variable selection problem. For example, in generalized linear models, the sub-model that relates to the random variable $y$ with the mean denoted by $\mu$ to a subset of components of $X$ in the form

$$g(x;\mu) = X(s)\beta(s),$$

where $g(*)$ is the link function, $X(s)$ is a subset of the components of $X$, $\beta(s)$ is a vector of the corresponding regression parameters, and $s \subseteq (1, 2, \ldots, p)$. The variable selection problem is to find the best subset $s$ such that the sub-model is optimal according to some criteria that give a good description of the data-generating mechanism. Statistically speaking, variable selection is a way to reduce the complexity of the model, in some cases by accepting a small amount of bias to improve the precision.

Traditionally, variable selection is achieved by evaluating all possible sub-models via information criteria such as Akaike's information criterion (AIC; Akaike 1973, 1974), Bayesian information criterion (BIC; Schwarz 1978), and Empirical-likelihood-based information-theoretic approaches (EAIC, EBIC; Variyath et al. 2010). The sub-model that minimizes the information criteria is then selected together with the corresponding covariates. For high dimensional data, which are often encountered in modern applications, the computational burden makes the direct application of these information criteria infeasible. To overcome the computational difficulties as well as to achieve some selection stability, regularization methods have drawn substantial attention. There is a large volume of literature on the penalized likelihood approach for building such models; for example, least absolute shrinkage and selection operator (LASSO; Tibshirani 1996) and the smoothly clipped absolute deviation (SCAD; Fan and Li 2001). Both approaches have many desirable properties. Other related variable selection methods include penalized empirical likelihood-based variable selection (Nadarajah 2011; Variyath 2006), adaptive LASSO (Zhang and Lu 2007; Zou 2006), least-square approximation (Wang and Leng 2007) and the folded concave penalty method (Lv and Fan 2009). However, the aforementioned variable selection methods are only applicable to generalized linear regression models.

Variable selection for longitudinal data is quite challenging due to the high dimensionality of covariates and the correlation within subject. Pan (2001) developed a quasi-likelihood information criterion (QIC) under the working independence model and the naive and robust covariance estimates of estimated regression coefficients. Cantoni et al. (2005) proposed a generalized version of Mallows's $C_p$ suitable for use with both parametric and nonparametric models. This model selection avoids a stepwise procedure and is based on a measure of predictive error rather than on significance testing. Wang and Qu (2009) introduced a novel Bayesian information criterion type model selection procedure based on the quadratic inference function, which does not require the full likelihood. The implementation of best subset type model selection procedures call for the evaluation of all possible submodels, which becomes computationally intensive when the number of covariates is moderately large.

The idea of penalization is very useful in longitudinal modeling, particularly in high dimensional variable selection. Fan and Li (2004) proposed an innovative class of variable selection procedures to select significant variables in the semiparametric models for continuous responses. Wang et al. (2008) studied regularized estimation procedures for nonparametric varying coefficient models for continuous responses that can simultaneously perform variable selection and the estimation of smooth coefficient functions. Xiao et al. (2009) recently investigated a double-penalized likelihood approach for selecting important parametric fixed effects in semiparametric mixed models for continuous responses. Dziak et al. (2009) discussed the applications of the SCAD-penalized quadratic inference function. Xu et al. (2010) investigated a GEEs-based shrinkage estimator with an artificial objective function. Xue et al. (2010) considered the model selection of a generalized additive model when responses from the same cluster are correlated. However, the aforementioned methods assume that the dimension of predictors are relatively small and some of these works are only applicable to continuous responses. The joint likelihood function for correlated discrete responses does not have a closed form when the correlations among the repeated measures are taken into account. To avoid specifying the full joint likelihood for correlated data, Wang et al. (2012) proposed penalized GEEs with a non-convex penalty function, which requires only the specification of the first two marginal moments and a working correlation matrix. This penalized GEEs method is superior to traditional methods because of their computational efficiency and stability. The true regression coefficients that are zero are automatically shrunk to zero, and the remaining coefficients are simultaneously estimated. These methods work reasonably well in high-dimensional problem; however, the GEEs estimate of $\beta$ is not necessarily consistent in some situations, as discussed by Crowder (1995) and Sutradhar and Das (1999). To overcome this problem, Sutradhar (2003) has proposed a generalization of the quasi-likelihood (GQL) approach to improve the efficiency of the parameter estimates. Utilizing this, to avoid the risk of model mis-specifications such as in variance function and correlation structure, we propose penalized generalized quasi-likelihood (PGQL) based on stationary lag correlation structure. Our simulation studies show that the proposed method works well compared to PGEEs.

The remaining part of the paper is organized as follows. In Sect. 2, we discussed the importance of the GQL approach and compared its performance with the GEEs approach. In Sect. 3, we introduced PGQL-based variable selection for longitudinal data. Its theoretical properties and numerical algorithm are also explored in this section. In Sect. 4, the performance analysis of the proposed method is assessed based on Monte Carlo simulations. The proposed method is applied to health care utilization count data in Sect. 5 and our conclusions are given in Sect. 6.

## 2 Generalized Quasi-Likelihood

The structure of the longitudinal data-set consists of an outcome random variable $y_{it}$ and $p$-dimensional vector of covariates $x_{it}$ that are observed for subjects $i = 1, \dots, k$ at a time point $t$, $t = 1, \dots, m_i$. For the $i$th subject, let $y_i = (y_{i1}, \dots, y_{im_i})^T$ be the response vector and let $X_i = (x_{i1}, x_{i2}, \dots, x_{im_i})^T$ be the $m_i \times p$ matrix of covariates. We assume that the $k$ subjects are independent while the repeated measurements $y_{it}$ taken on each subject are correlated. The marginal density of $y_{it}$ is assumed to follow a canonical exponential family (Liang and Zeger 1986) of the form

$$f(y_{it}) = \exp\left[(y_{it}\theta_{it} - a(\theta_{it}))\phi + b(y_{it}, \phi)\right], \tag{1}$$

where $\theta_{it} = g(\eta_{it})$, $g$ is the known injective function with $\eta_{it} = x_{it}\beta$, $\beta$ is a $p \times 1$ vector of the regression effects of $x_{it}$ on $y_{it}$, $a(*)$, and $b(*)$ are known functional forms. The mean and the variance of $y_{it}$ as $\mathrm{E}(y_{it}|x_{it}) = a'(\theta_{it}) = \mu_{it}$, and $\mathrm{Var}(y_{it}) = a''(\theta_{it}) = v(\mu_{it})\phi$, where $\phi$ is the unknown over-dispersion parameter for simplicity we assume $\phi = 1$ in this study and $v(*)$ is a known variance function.

Note that when there is a functional relationship between the mean and variance of the response, Wedderburn (1974) proposed a quasilikelihood (QL) approach for independence data which utilize both the mean and the variance in estimating the regression effects. When there is insufficient information about the data for us to specify a parametric model, quasi-likelihood is often used. The QL optimal estimating equation for $\beta$ is given by

$$\sum_{i=1}^{k}\sum_{t=1}^{m_i}\left[\frac{\partial a'(\theta_{it})}{\partial \beta}\frac{(y_{it} - a'(\theta_{it}))}{\mathrm{Var}(y_{it})}\right] = 0.$$

In a longitudinal setup, the components of the response vector $y_i$ are repeated, which are likely to be correlated. Let $C_i(\rho)$ be the $m_i \times m_i$ true correlation matrix of $y_i$, $i = 1, \dots, k$, which is unknown in practice. Our primary interest is to estimate $\beta$ after taking the longitudinal correlation $C_i(\rho)$ into account. For known $C_i(\rho)$, the QL estimator of $\beta$ under (1) is the solution of the score equation

$$g(y; \beta) = \sum_{i=1}^{k} X_i^T A_i \Sigma_i^{-1}(\rho)(y_i - \mu_i) = 0, \tag{2}$$

where $A_i = \mathrm{diag}\left[a''(\theta_{i1}), \dots, a''(\theta_{it}), \dots, a''(\theta_{im_i})\right]$, and $\Sigma_i(\rho) = A_i^{1/2} C_i(\rho) A_i^{1/2}$ is the true covariance of $y_i$. In real applications the true correlation structure is often unknown. Ignoring the correlation among the same individual could lead to an inefficient estimation of the regression coefficients and underestimation of standard errors.

To overcome these problems, in a seminal paper Liang and Zeger (1986) proposed the generalized estimating equations (GEEs) approach. The GEEs approach for estimating the parameter vector of the marginal regression model (1) allows the user to specify any working correlation structure for the correlation matrix of a subject's outcomes $y_i$. They developed the joint probability model by introducing a "working" correlation structure based on a generalized estimating equations approach to obtain consistent and efficient estimators for the regression parameter $\beta$, given by

$$g(\beta, \hat{\alpha}(\beta)) = \sum_{i=1}^{k} X_i^T A_i^{1/2} R_i^{-1}(\hat{\alpha}) A_i^{-1/2}(y_i - \mu_i) = 0, \tag{3}$$

where $A_i = m_i \times m_i$ diagonal matrix with $\mathrm{Var}(\mu_{it})$ as the *ith* diagonal element, $R_i(\hat{\alpha})$ is the "working" correlation matrix of the $m_i$ repeated measures used for $C_i(\rho)$ in equation (2). We can choose the form of the $m_i \times m_i$ "working" correlation matrix $R_i$ for each $y_i$, defined by the $(j, j^{'})$ element of $R_i$ is the known, hypothesized, or estimated correlation between $y_{ij}$ and $y_{ij'}$. The working correlation structure can depend on an unknown $s \times 1$ correlation parameter vector $\alpha$. The observation times and correlation matrix can differ from subject to subject, but the correlation matrix $R_i(\alpha)$ of the $i$th subject is fully specified by $\alpha$. For a given working correlation structure, $\alpha$ can be estimated using a residual-based method of moments. The GEEs estimate of $\beta$ is not necessarily consistent in some situations as discussed by Crowder (1995) and Sutradhar and Das (1999). Crowder (1995) demonstrated that there may not be any solutions for $\hat{\alpha}$ , which misleads the estimation of regression parameters. Also the GEEs approach gives a consistent estimator of $\beta$, but this estimator in some situations is less efficient than the independence estimating equation approach under an arbitrary working correlation structure as shown by Sutradhar and Das (1999).

In such situations, Sutradhar (2003) has proposed a generalization of the quasi-likelihood (GQL) approach to improve the efficiency of the parameter estimates. The estimation for $\beta$ is obtained by solving the GQL estimating equations given by

$$g(\beta, \rho) = \sum_{i=1}^{k} X_i^T A_i \Sigma_i^{-1}(\hat{\rho})(y_i - \mu_i) = 0, \tag{4}$$

where $\Sigma_i(\hat{\rho}) = A_i^{1/2} C_i^*(\rho) A_i^{1/2}$, with $C_i^*(\rho)$ as the stationary lag-correlation structure for any of the AR(1), MA(1), or EQC models, and

$$C_i^*(\rho) = \begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{m-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{m-2} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \rho_{m-1} & \rho_{m-2} & \rho_{m-3} & \cdots & 1 \end{bmatrix}. \tag{5}$$

**Table 1** A class of stationary correlation models for longitudinal count data from Sutradhar (2011, Sect. 6.3)

| Model | Dynamic relationship | Mean, variance and correlations |
|---|---|---|
| AR(1) | $y_{it} = \rho * y_{i,t-1} + d_{it},\ \ t = 2, 3, \ldots, m$ | $\mathrm{E}[y_{it}] = \mu_i$ |
| | $y_{i1} \sim \mathrm{Poi}(\mu_i = \exp[\tilde{X}_i \beta])$ | $\mathrm{var}[y_{it}] = \mu_i$ |
| | $d_{it} \sim \mathrm{Poi}(\mu_i(1-\rho)),\ \ t = 2, 3, \ldots, m$ | $\mathrm{corr}[y_{it}, y_{i,t+l}] = \rho_l = \rho^l$ |
| MA(1) | $y_{it} = \rho * d_{i,t-1} + d_{it},\ \ t = 1, 2, \ldots, m$ | $\mathrm{E}[y_{it}] = \mu_i$ |
| | $d_{i0} \sim \mathrm{Poi}(\mu_i/(1+\rho))$ | $\mathrm{var}[y_{it}] = \mu_i$ |
| | $d_{it} \sim \mathrm{Poi}(\mu_i/(1+\rho))\ t = 1, 2, \ldots, m$ | $\mathrm{corr}[y_{it}, y_{i,t+l}] = \rho_l$ |
| | | $= \dfrac{\rho}{(1+\rho)}$ for $l$=1 |
| EQC | $y_{it} = \rho * y_{i1} + d_{it},\ \ t = 2, 3, \ldots, m$ | $\mathrm{E}[y_{it}] = \mu_i$ |
| | $y_{i1} \sim \mathrm{Poi}(\mu_i)$ | $\mathrm{var}[y_{it}] = \mu_i$ |
| | $d_{it} \sim \mathrm{Poi}(\mu_i(1-\rho)),\ t = 2, 3, \ldots, m$ | $\mathrm{cor}[y_{it}, y_{i,t+l}] = \rho_l = \rho$ |

The stationary lag-correlations are estimated by the method of moments introduced by Sutradhar and Kovacevic (2000) and given by

$$\hat{\rho}_l = \frac{\displaystyle\sum_{i=1}^{k}\sum_{t=1}^{m-l} \tilde{y}_{it}\tilde{y}_{i,t+l}/k(m-l)}{\displaystyle\sum_{i=1}^{k}\sum_{t=1}^{m} \tilde{y}_{it}^2/km}, \tag{6}$$

where $l = |t - t'|,\ t \neq t',\ t, t' = 1, \ldots, m$ and $\tilde{y}_{it}$ is the standardized residual, defined as $\tilde{y}_{it} = \{y_{it} - \mu_{it}\}/\{a''(\theta_{it})\}^{1/2}$. The stationary lag correlation approach produces consistent as well as more efficient regression estimates as compared to the independence assumption-based estimating equation approaches (Sutradhar and Das 1999). We conducted a small simulation study to illustrate the comparison of GEEs with GQL under a mis-specified correlation structure. A class of stationary correlation AR(1) model for longitudinal count data are generated as per the dynamic relationship given in Table 1, which are discussed by McKenzie (1988), and Sutradhar (2011). The stationary covariates $\tilde{x}_i = (\tilde{x}_{i1}, \tilde{x}_{i2})$ are generated from the normal distribution with mean 0, variance 1, and $\beta = (0.3, 0.2)^T$. For a given $y_{i,t-1}$, $\rho * y_{i,t-1}$ denotes the commonly called binomial thinning operation discussed by McKenzie (1988). That is, $\rho * y_{i,t-1} = \sum_{j=1}^{y_{i,t-1}} b_j(\rho)$ with $\Pr[b_j(\rho) = 1] = \rho$, $\Pr[b_j(\rho) = 0] = 1 - \rho$. For the simulation's purpose, we consider the number of time points $m = 5,\ 10$ and number of subjects $k = 100$. We simulated 1000 data sets with $\rho = 0.49, 0.70$ from the above AR(1) stationary dynamic model to generate the data. Under the working exchangeable and MA(1) correlation structure, the correlation parameter $\hat{\alpha}$ can be estimated by using Eqs. (2.5) and (2.7) from Sutradhar and Das (1999). The mean of the estimated values of the regression coefficients and the

**Table 2** Coverage probabilities of regression estimates for a true AR(1) correlation model data under different "working" correlation models (m = 5)

| True model | Method | Parameters | Estimates | Coverage probabilities | |
|---|---|---|---|---|---|
| | | | | 95 % level | 99 % level |
| AR(1) $\rho = 0.70$ | GEEs (AR(1)) | $\beta_1$ | 0.3000 (0.070) | 0.952 (0.279) | 0.987 (0.367) |
| | | $\beta_2$ | 0.2009 (0.073) | 0.950 (0.286) | 0.988 (0.375) |
| | GEEs (EQC) | $\beta_1$ | 0.2997 (0.073) | 0.911 (0.247) | 0.973 (0.325) |
| | | $\beta_2$ | 0.1956 (0.076) | 0.902 (0.252) | 0.963 (0.332) |
| | GQL | $\beta_1$ | 0.3003 (0.070) | 0.952 (0.278) | 0.986 (0.366) |
| | | $\beta_2$ | 0.2007 (0.073) | 0.950 (0.284) | 0.988 (0.374) |
| AR(1) $\rho = 0.49$ | GEEs (AR(1)) | $\beta_1$ | 0.2989 (0.062) | 0.938 (0.237) | 0.988 (0.319) |
| | | $\beta_2$ | 0.1956 (0.062) | 0.940 (0.243) | 0.981 (0.319) |
| | GEEs (EQC) | $\beta_1$ | 0.2992 (0.061) | 0.899 (0.206) | 0.968 (0.272) |
| | | $\beta_2$ | 0.1986 (0.062) | 0.908 (0.211) | 0.980 (0.278) |
| | GEEs (MA(1)) | $\beta_1$ | 0.2991 (0.062) | 0.897 (0.205) | 0.968 (0.270) |
| | | $\beta_2$ | 0.1985 (0.062) | 0.905 (0.210) | 0.981 (0.276) |
| | GQL | $\beta_1$ | 0.2989 (0.061) | 0.931 (0.235) | 0.990 (0.309) |
| | | $\beta_2$ | 0.1955 (0.061) | 0.936 (0.241) | 0.992 (0.317) |

corresponding simulated standard errors in parentheses are reported in Tables 2 and 3 for different $m = 5, 10$ respectively. We also report the coverage probabilities as well as the width of the confidence interval for $\beta_1$ and $\beta_2$ in parentheses for confidence levels 0.95 and 0.99. In this study, we generated the data using the AR(1) correlation structure even though we used all three working correlation structures: AR(1), EQC, and MA(1) for parameter estimation under GEEs and the results are compared with the GQL approach.

We see from Tables 2 and 3 that when we use the true working correlation structure, coverage probabilities based on the GEEs and GQL approaches are

**Table 3** Coverage probabilities of regression estimates for a true AR(1) correlation model data under different "working" correlation models (m = 10)

| True model | Method | Parameters | Estimates | Coverage probabilities | |
|---|---|---|---|---|---|
| | | | | 95 % level | 99 % level |
| AR(1) $\rho = 0.70$ | GEEs (AR(1)) | $\beta_1$ | 0.2961 (0.057) | 0.956 (0.228) | 0.992 (0.299) |
| | | $\beta_2$ | 0.1978 (0.059) | 0.951 (0.232) | 0.988 (0.306) |
| | GEEs (EQC) | $\beta_1$ | 0.3020 (0.059) | 0.811 (0.159) | 0.922 (0.209) |
| | | $\beta_2$ | 0.1993 (0.062) | 0.802 (0.163) | 0.919 (0.214) |
| | GQL | $\beta_1$ | 0.2960 (0.056) | 0.955 (0.226) | 0.991 (0.231) |
| | | $\beta_2$ | 0.1977 (0.057) | 0.944 (0.231) | 0.986 (0.303) |
| AR(1) $\rho = 0.49$ | GEEs (AR(1)) | $\beta_1$ | 0.2977 (0.047) | 0.945 (0.181) | 0.990 (0.238) |
| | | $\beta_2$ | 0.1994 (0.048) | 0.937 (0.185) | 0.987 (0.243) |
| | GEEs (EQC) | $\beta_1$ | 0.2985 (0.046) | 0.848 (0.135) | 0.952 (0.178) |
| | | $\beta_2$ | 0.1975 (0.050) | 0.826 (0.138) | 0.930 (0.181) |
| | GEEs (MA(1)) | $\beta_1$ | 0.3011 (0.047) | 0.842 (0.134) | 0.945 (0.177) |
| | | $\beta_2$ | 0.1985 (0.048) | 0.847 (0.137) | 0.943 (0.180) |
| | GQL | $\beta_1$ | 0.2981 (0.044) | 0.954 (0.178) | 0.990 (0.233) |
| | | $\beta_2$ | 0.2000 (0.046) | 0.949 (0.182) | 0.988 (0.239) |

almost the same. However, under an arbitrary working correlation structure, the GQL approach performs better than the GEEs approach. This result shows a loss of efficiency of the GEEs estimators due to mis-specification of the correlation structures. We also notice that when $m$ increases GQL has better coverage compared to GEEs based confidence interval. Rather than using any arbitrary "working correlation", it seems much better to define a lag-correlation structure for the longitudinal responses to estimate the parameters. The correlation structure (5) is quite robust, and it accommodates all three correlation structures: AR(1), EQC, and MA(1). Note, however, that the correlation structure is unknown in practice and it makes more sense to use a stationary lag-correlation structure to represent all the three correlation structures in a unique way. We did not consider other cases, for instance in a true EQC and MA(1) correlation models, since under different working correlation structure the correlation parameter $\hat{\alpha}$ does not exist.

## 3 Penalized Generalized Quasi-Likelihood (PGQL)

We use the GQL approach discussed in Sect. 2 and the SCAD penalty function (Fan and Li 2001), to develop the penalized generalized quasi-likelihood for variable selection in the context of a longitudinal data analysis. The regression parameters are estimated by solving the penalized generalized estimating functions

$$\mathscr{U}(\beta) = g(\beta, \hat{\rho}(\beta)) - kp'_{\delta}(|\beta|)\text{sign}(\beta) \quad (7)$$

where $g(\beta, \hat{\rho}(\beta)) = \sum_{i=1}^{k} X_i^T A_i^{1/2} C_i^{*-1}(\rho) A_i^{-1/2}(y_i - \mu_i) = 0$ be the GQL estimating equation given in (4), $p'_{\delta}(*)$ is the first derivative of penalty function, $\text{sign}(\beta) = (\text{sign}(\beta_1), \ldots, \text{sign}(\beta_p))^T$ with $\text{sign}(t) = \text{I}(t > 0) - \text{I}(t < 0)$, and $\delta$ is the tuning parameter. Different penalty functions can be potentially adopted in this modeling. The HARD thresholding penalty proposed by Fan (1997) and Antoniadis (1997) is defined as $p_{\delta}(|\theta|) = \delta^2 - (|\theta| - \delta)^2 I(|\theta| < \delta)$. For a large value of $|\theta|$, the HARD thresholding penalty does not overpenalize. The LASSO penalty function is the $L_1$-penalty, $p_{\delta}(|\theta|) = \delta|\theta|$, proposed by Donoho and Johnstone (1994) in the wavelet setting and extended by Tibshirani (1996) to general likelihood settings. The penalty function used in ridge regression is the $L_2$ penalty, $p_{\delta}(|\theta|) = \delta|\theta|^2$. According to Fan and Li (2001), a good penalty function should result in an estimator with the following three oracle properties:

1. Unbiasedness: To avoid unnecessary modeling bias, the estimator is nearly unbiased when the true unknown parameter is large.
2. Sparsity: This is a thresholding rule that automatically sets small estimated coefficients to zero to reduce the model complexity.
3. Continuity: This property eliminates unnecessary variation in the model prediction.

However, the penalty functions $L_1, L_2$, and HARD do not satisfy all three oracle properties. A simple penalty function satisfying all three is the SCAD penalty proposed by Fan (1997) where the first derivative is

$$p'_{\delta}(\theta) = \delta\left\{\text{I}(\theta \le \delta) + \frac{(a\delta - \theta)_+}{(a-1)\delta}\text{I}(\theta > \delta)\right\} \quad \text{for some } a > 2 \text{ and } \theta > 0. \quad (8)$$

Necessary conditions for the unbiasedness, sparsity, and continuity of the SCAD penalty have been given by Antoniadis and Fan (2001). This penalty function involves two unknown parameters, $a$ and $\delta$. As shown in Fig. 1, all the penalty functions are singular at the origin, satisfying $p_{\delta}(0+) > 0$. This is the necessary

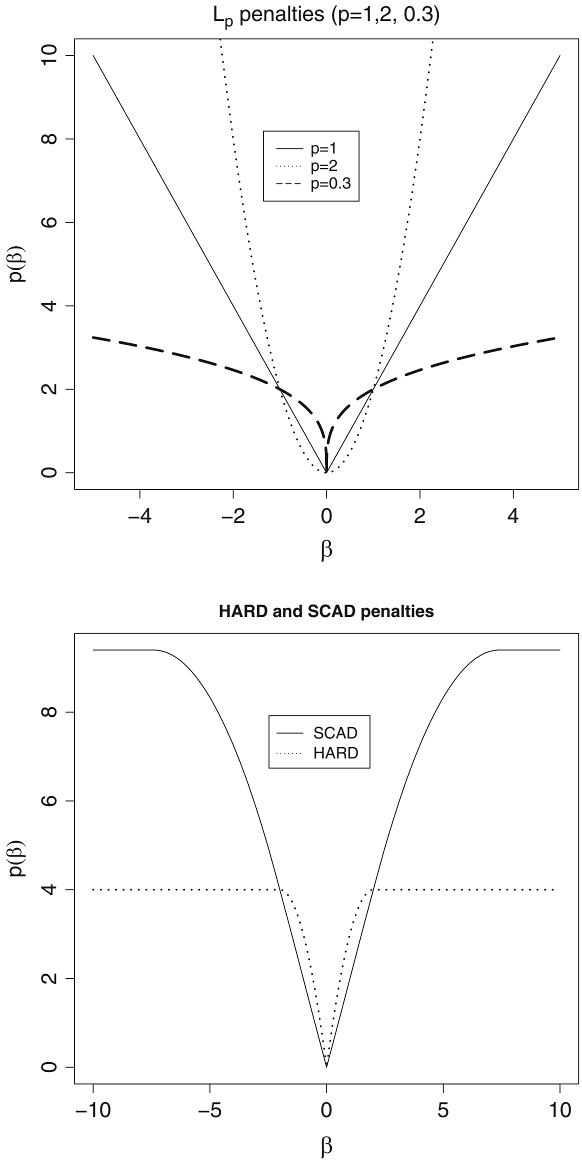



**Fig. 1** $L_p$, SCAD, and HARD penalty functions and their quadratic approximation

condition for sparsity in variable selection. As shown in Fig. 1, the HARD and SCAD penalties are constant when $\beta$ is large, indicating that there is no excessive penalization for large regression coefficients. However, SCAD is smoother than HARD and hence yields a continuous estimator.

By following Fan and Li (2001), we solve (7) to carry out estimation and variable selection simultaneously and arrive the penalized GQL estimates of the regression parameters, $\hat{\beta}_{PGQL}$, which has properties similar to $\hat{\beta}_{PGEEs}$.

### 3.1 Numerical Algorithm

To implement our method, we need an efficient numerical algorithm. The SCAD penalty function involves two unknown parameters, $\delta$ and $a$. From a Bayesian point of view, Fan and Li (2001) suggested setting $a = 3.7$ and using generalized cross-validation (GCV; Craven and Wahba 1979) to select the best value of $\delta$, which is implemented in our simulation studies. We maximize the PGQL with respect to $\beta$ given in (7). We used the modified Newton-Raphson algorithm proposed by Fan and Li (2001), which is numerically stable. At each iteration, we compute the stationary lag-correlation structure $C^*(\rho)$ given in (5) for an updated value of $\beta$. A step-by-step numerical algorithm for estimating $\hat{\beta}_{PGQL}$ for a given value of the tuning parameter $\delta$ is given below.

(a) Set $\beta = \beta^0$, and $\epsilon = 1e - 08$.
(b) Let $C^*(\rho)$ be the estimated value of $C^*(\rho)$.
(c) The parameter $\hat{\beta}_{PGQL}$ is computed iteratively and the solution at the $(h + 1)$th iteration is given by

$$\hat{\beta}^{(h+1)} = \hat{\beta}^{(h)} + \left\{H(\hat{\beta}^h) + kS_\delta(\hat{\beta}^h)\right\}^{-1} \left\{U(\hat{\beta}^h) - kS_\delta(\hat{\beta}^h)\hat{\beta}^h\right\} \tag{9}$$

where

$$H(\hat{\beta}^h) = \sum_{i=1}^{k} X_i^T A_i \Sigma_i^{-1}(\hat{\rho}) A_i X_i, \;\; \Sigma_i \text{ is defined in (4)},$$
$$S_\delta(\hat{\beta}^h) = \text{diag}\left[\frac{p'_\delta(|\hat{\beta}_1^h|)}{|\hat{\beta}_1^h|}, \ldots, \frac{p'_\delta(|\hat{\beta}_p^h|)}{|\hat{\beta}_p^h|}\right].$$

(d) If $\min\left|\hat{\beta}^{(h+1)} - \hat{\beta}^{(h)}\right| < \epsilon$ stop the algorithm and report $\hat{\beta}^{(h+1)}$; otherwise, $h = h + 1$ and go to Step 3.

## 4 Performance Analysis

### 4.1 Stationary Count Data

A class of stationary correlation models for longitudinal count data are considered for our simulation studies, which are given in Table 1. For the purpose of the simulation, we consider five covariates $\tilde{X}_i = (\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_5})$, assumed to have distributions as $\tilde{x}_{i1} \sim$ Bernoulli(0.5), $\tilde{x}_{i2}$ to $\tilde{x}_{i5}$ from the standard normal with $\text{cor}(x_i, x_j) = 0.5^{|i-j|}$, and $\beta = (0.5, 0.5, 0.6, 0, 0)^T$. The performance analysis is carried out based on two measures, viz (a) the median of the relative model error (MRME) and (b) the average number of correct zero and non-zero coefficients. The estimated values of the nonzero coefficients and the corresponding simulated standard errors are also reported. The model error (ME) is defined as $\text{ME}(\hat{\beta}) = E_x\left\{\mu(X\beta) - \mu(X\hat{\beta})\right\}^2$, where $\mu(X\beta) = E(y|X)$ and computed the relative model error as $\text{RME} = \text{ME}/\text{ME}_{\text{full}}$ where $\text{ME}_{\text{full}}$ is the model error calculated by fitting the data with the full model and ME is the model error of the selected model. For the purpose of this simulation, we consider the number of time points $m = 5, 10$ and number of subjects $k = 100$. The entire simulation study was repeated 1000 times for all the three true models and a summary of the performance measures is given in Tables 4 and 5. From Tables 4 and 5 we see that, when we use the true working correlation structure, the MRME of PGEEs is very close to the MRME arrived based on PGQL. The average number of zero coefficients for PGEEs and PGQL are closer to the target of two and the nonzero regression parameter estimates are close to the true values in all cases. This shows, the bias of the non-zero parameter estimates are approximately zero. However, the proposed PGQL approach has smaller MRME compared to PGEEs under an arbitrary working correlation structure. Also, we noticed the average number of zero coefficients for PGQL is closer to the target of two in all cases but not for PGEEs with a wrong correlation structure. We repeated the simulation with smaller sample sizes, and the overall conclusions were similar, so they have not been provided here. These simulations studies clearly shows that PGQL is performing better compared to PGEEs with mis-specified working approaches. Note that, in practice, it is very difficult to know the true correlation structure, so the PGQL approach is preferred for the variable selection as well as estimation of regression parameters.

## 4.2 Over-Dispersed Stationary Count Data

In this section, we consider the performance of our method when the model is misspecified in the context of stationary count data. We generate over-dispersed stationary count data $y_{it}$ using $\tilde{\mu}_i = u_i \exp(\tilde{x}_i\beta)$ on the models which are discussed in Table 1 with $u_i$ a random sample such that $\text{E}(u_i) = 1$ and $\text{Var}(u_i) = \omega$. Marginally, we have $\text{E}(y_{it}) = \tilde{\mu}_i$ and $\text{Var}(y_{it}) = \tilde{\mu}_i(1 + \tilde{\mu}_i\omega)$. The distribution of $u$ is chosen to be gamma with parameters $(\omega, 1/\omega)$ with $\omega$ being the over-dispersion parameter. For the purpose of the simulation, we consider five covariates $\tilde{X}_i = (\tilde{x}_{i_1}, \ldots, \tilde{x}_{i_5})$, assumed to have distributions as $\tilde{x}_{i1} \sim$ Bernoulli(0.5), $\tilde{x}_{i2}$ to $\tilde{x}_{i5}$ are generated from a multivariate normal distribution with a mean of zero, and the correlation between $x_i$ and $x_j$ is $0.5^{|i-j|}$, $\omega = 0.25$, and $\beta = (0.5, 0.5, 0.6, 0, 0)^T$. In each simulation, we generated $m = 5$ repeated over-dispersed count data for a sample of

**Table 4** Performance measures for count data with stationary covariates (m = 5)

| True model | Method | MRME% | Avg. no. of zero coefficients | | Estimates of nonzero coefficients | | |
|---|---|---|---|---|---|---|---|
| | | | Correct | Incorrect | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ |
| AR(1) $\rho = 0.70$ | PGEEs(IND) | 86.86 | 1.25 | 0.0 | 0.5002 (0.068) | 0.5023 (0.075) | 0.5909 (0.076) |
| | PGEEs(AR(1)) | 65.60 | 1.85 | 0.0 | 0.5029 (0.066) | 0.5058 (0.073) | 0.5930 (0.071) |
| | PGEEs(EQC) | 69.76 | 1.84 | 0.0 | 0.5030 (0.066) | 0.5047 (0.073) | 0.5935 (0.072) |
| | PGQL | 66.90 | 1.85 | 0.0 | 0.5034 (0.066) | 0.5052 (0.073) | 0.5930 (0.071) |
| AR(1) $\rho = 0.49$ | PGEEs(AR(1)) | 63.60 | 1.80 | 0.0 | 0.5003 (0.056) | 0.5025 (0.056) | 0.5968 (0.057) |
| | PGEEs(MA(1)) | 70.57 | 1.65 | 0.0 | 0.5011 (0.060) | 0.5021 (0.061) | 0.5986 (0.062) |
| | PGQL | 67.64 | 1.79 | 0.0 | 0.5014 (0.059) | 0.5029 (0.060) | 0.5973 (0.061) |
| EQC $\rho = 0.70$ | PGEEs(IND) | 77.95 | 1.23 | 0.0 | 0.5059 (0.074) | 0.5053 (0.076) | 0.5913 (0.080) |
| | PGEEs(EQC) | 61.43 | 1.87 | 0.0 | 0.5022 (0.073) | 0.5066 (0.076) | 0.5921 (0.074) |
| | PGEEs(AR(1)) | 63.37 | 1.70 | 0.0 | 0.5025 (0.074) | 0.5066 (0.076) | 0.5914 (0.076) |
| | PGQL | 62.61 | 1.87 | 0.0 | 0.5026 (0.073) | 0.5069 (0.076) | 0.5916 (0.073) |
| EQC $\rho = 0.49$ | PGEEs(EQC) | 65.39 | 1.82 | 0.0 | 0.5023 (0.062) | 0.5057 (0.065) | 0.5922 (0.064) |
| | PGEEs(MA(1)) | 75.50 | 1.59 | 0.0 | 0.4996 (0.065) | 0.5022 (0.068) | 0.5980 (0.073) |
| | PGQL | 66.40 | 1.82 | 0.0 | 0.5017 (0.064) | 0.5046 (0.068) | 0.5938 (0.069) |
| MA(1) $\rho = 0.67$ | PGEEs(IND) | 70.29 | 1.54 | 0.0 | 0.5002 (0.052) | 0.5006 (0.054) | 0.5994 (0.054) |
| | PGEEs(MA(1)) | 63.56 | 1.72 | 0.0 | 0.5004 (0.052) | 0.4993 (0.053) | 0.5981 (0.052) |
| | PGEEs(AR(1)) | 69.39 | 1.78 | 0.0 | 0.5018 (0.052) | 0.5013 (0.056) | 0.5963 (0.054) |
| | PGEEs(EQC) | 71.37 | 1.71 | 0.0 | 0.5006 (0.052) | 0.5008 (0.056) | 0.5974 (0.058) |
| | PGQL | 65.20 | 1.75 | 0.0 | 0.5005 (0.051) | 0.5004 (0.053) | 0.5970 (0.053) |

**Table 5** Performance measures for count data with stationary covariates (m = 10)

| True model | Method | MRME% | Avg. no. of zero coefficients | | Estimates of nonzero coefficients | | |
|---|---|---|---|---|---|---|---|
| | | | Correct | Incorrect | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ |
| AR(1) $\rho = 0.70$ | PGEEs (IND) | 84.21 | 0.93 | 0.0 | 0.5007 (0.052) | 0.5017 (0.056) | 0.5963 (0.063) |
| | PGEEs (AR(1)) | 60.96 | 1.76 | 0.0 | 0.5000 (0.051) | 0.5022 (0.053) | 0.5967 (0.054) |
| | PGEEs (EQC) | 60.56 | 1.79 | 0.0 | 0.5000 (0.052) | 0.5020 (0.055) | 0.5981 (0.055) |
| | PGQL | 60.32 | 1.76 | 0.0 | 0.5003 (0.051) | 0.5005 (0.053) | 0.5988 (0.053) |
| AR(1) $\rho = 0.49$ | PGEEs (AR(1)) | 71.54 | 1.69 | 0.0 | 0.5010 (0.042) | 0.5002 (0.043) | 0.5996 (0.043) |
| | PGEEs (MA(1)) | 81.85 | 1.27 | 0.0 | 0.5012 (0.042) | 0.4989 (0.044) | 0.6004 (0.046) |
| | PGQL | 77.21 | 1.67 | 0.0 | 0.5010 (0.041) | 0.5013 (0.043) | 0.5980 (0.044) |
| EQC $\rho = 0.70$ | PGEEs (IND) | 99.04 | 0.743 | 0.0 | 0.5009 (0.074) | 0.5041 (0.078) | 0.5935 (0.086) |
| | PGEEs (EQC) | 63.46 | 1.87 | 0.0 | 0.5014 (0.072) | 0.5058 (0.073) | 0.5928 (0.075) |
| | PGEEs (AR(1)) | 69.42 | 1.67 | 0.0 | 0.5026 (0.072) | 0.5032 (0.076) | 0.5943 (0.076) |
| | PGQL | 64.15 | 1.86 | 0.0 | 0.5010 (0.071) | 0.5050 (0.074) | 0.5945 (0.075) |
| MA(1) $\rho = 0.67$ | PGEEs(IND) | 83.44 | 1.34 | 0.0 | 0.5015 (0.034) | 0.4991 (0.037) | 0.6005 (0.038) |
| | PGEEs (EQC) | 71.85 | 1.56 | 0.0 | 0.5008 (0.033) | 0.5010 (0.036) | 0.5976 (0.038) |
| | PGEEs (AR(1)) | 72.39 | 1.62 | 0.0 | 0.4998 (0.033) | 0.5005 (0.036) | 0.5974 (0.038) |
| | PGQL | 71.12 | 1.59 | 0.0 | 0.5001 (0.032) | 0.5000 (0.035) | 0.5983 (0.037) |

size $k = 100$ individuals with three different correlation structures and arrived the parameter estimates for each method. The MRME, the average number of zero and nonzero coefficients, the estimated values of the nonzero regression coefficients and the corresponding standard errors over 1000 simulated data sets are summarized in Table 6. When there is an over-dispersion, as we noticed from Table 6, the proposed PGQL approach has smaller MRME compared to PGEEs but the average number of zero coefficients for PGEEs and PGQL are close to each other and the nonzero regression parameter estimates are close to the true values in all cases. This shows,

**Table 6** Performance measures for over-dispersion count data with stationary covariates (m = 5)

| True model | Method | MRME% | Avg. no. of zero coefficients | | Estimates of nonzero coefficients | | |
|---|---|---|---|---|---|---|---|
| | | | Correct | Incorrect | $\hat{\beta}_1$ | $\hat{\beta}_2$ | $\hat{\beta}_3$ |
| AR(1) $\rho = 0.50$ | PGEEs(IND) | 95.62 | 0.89 | 0.0 | 0.5009 (0.089) | 0.5077 (0.106) | 0.5897 (0.118) |
| | PGEEs(AR(1)) | 73.07 | 1.47 | 0.0 | 0.5018 (0.088) | 0.5121 (0.102) | 0.5864 (0.108) |
| | PGQL | 69.04 | 1.52 | 0.0 | 0.5037 (0.086) | 0.5110 (0.102) | 0.5874 (0.108) |
| EQC $\rho = 0.50$ | PGEEs(IND) | 95.91 | 0.86 | 0.0 | 0.4991 (0.096) | 0.5059 (0.109) | 0.5910 (0.120) |
| | PGEEs(EQC) | 82.80 | 1.53 | 0.0 | 0.5117 (0.092) | 0.5075 (0.106) | 0.5832 (0.116) |
| | PGQL | 81.79 | 1.51 | 0.0 | 0.5104 (0.091) | 0.5082 (0.106) | 0.5850 (0.116) |
| MA(1) $\rho = 0.50$ | PGEEs(IND) | 85.43 | 0.95 | 0.0 | 0.5004 (0.086) | 0.5029 (0.098) | 0.5921 (0.114) |
| | PGEEs(MA(1)) | 73.78 | 1.31 | 0.0 | 0.5059 (0.086) | 0.5071 (0.099) | 0.5892 (0.111) |
| | PGQL | 68.37 | 1.46 | 0.0 | 0.5054 (0.086) | 0.5058 (0.097) | 0.5892 (0.106) |

the bias of the non-zero parameter estimates are approximately zero. Again, these simulation studies clearly show that PGQL is performing better compared to PGEEs with over-dispersion data. Note that, in practice, it is very difficult to know the true correlation structure; however, it is reasonable to assume that the correlation structure remains the same for all individuals.

## 5 Health Care Utilization Data Study

We applied our proposed methodology to a real life data-set on the health care utilization problem, studied by Sutradhar (2003). This data-set was collected by the General Hospital of St. Johns, Newfoundland, Canada. These longitudinal count data contain the complete records for $k = 144$ individuals for 4 years ($m = 4$) from 1985 to 1988. The number of visits to a physician by each individual during a given year was recorded as the response, and this was repeated for 4 years. Also, the information on four covariates: gender, number of chronic conditions, education level, and age were also recorded for each individual. We are also interested in examining whether there are any interaction effects between the parametric covariates, so we included some of these interactions in our model. In

**Table 7** Estimates of the regression parameters under PGQL and PGEEs approaches in fitting health care utilization count data

| | Penalized estimates | | | |
|---|---|---|---|---|
| Variable | PGEEs (AR(1)) | PGEEs (EQC) | PGEEs (MA(1)) | PGQL |
| GENDER | 0.000 | 0.000 | 0.000 | 0.000 |
| CHRONIC | 0.105 | 0.103 | 0.104 | 0.104 |
| EDUCATION | −0.492 | −0.432 | −0.489 | −0.443 |
| AGE | 0.033 | 0.032 | 0.033 | 0.033 |
| GENDER*CHRONIC | 0.143 | 0.144 | 0.144 | 0.143 |
| GENDER*EDUCATION | 0.053 | 0.000 | 0.050 | 0.000 |
| GENDER*AGE | −0.009 | −0.009 | −0.009 | −0.009 |

view of the background information, it is appropriate to assume that the response variable, marginally, follows the Poisson distribution, and the repeated counts recorded for 4 years will be longitudinally correlated. We are interested in taking the longitudinal correlations into account and examining the effects of the above covariates and their interaction effects on the physician visits. We used PGEEs under different correlation structure for variable selection and compared the results with our proposed method, PGQL. A summary of the results is given in Table 7. From Table 7, we see that all methods identified CHRONIC, EDUCATION, AGE, GENDER*CHRONIC, and GENDER*AGE are the significant variables and the covariate GENDER as unimportant. Under PGEEs method with AR(1) & MA(1) working correlation structure, results indicates that GENDER*EDUCATION also significant where as it is not significant for other two methods. Parameter estimates and identified variables are almost similar for PGEEs (EQC) and PGQL. This clearly indicate that PGEEs based variable selection procedure is sensitive to the choice of covariance structure, leading to different results for different covariance structures. Since in practical situations the true correlation structure is often unknown, the PGQL approach is more appropriate since it can accommodate all three correlation structures in a unique way.

## 6 Concluding Remarks

We propose a penalized GQL approach for variable selection in longitudinal data analysis where both estimation and variable selection are carried out simultaneously. We used SCAD penalty to achieve oracle properties. Our performance analysis shows that the PGQL approach produces consistent as well as more efficient regression estimates as compared to the independence assumption-based PGEEs approach. The proposed PGQL approach assumes a known longitudinal lag-correlation structure with unknown correlation parameters. When the correlation structure is known the PGQL method has similar performance compared to the

PGEEs approach. However, when the model is mis-specified such as variance function and correlation structure our proposed PGQL approach outperforms the PGEEs method. The main advantage of this PGQL approach is that there is unique way to specify the correlation structure compared to the PGEEs method.

**Acknowledgements** The authors' are grateful for the opportunity to present their work at the 2015 International Symposium in Statistics (ISS) on Advances in Parametric and Semiparametric Analysis of Multivariate, Time Series, Spatial-temporal, and Familial-longitudinal Data. Special thanks go to Professor Brajendra Sutradhar for organizing the conference, to members of the symposium audience for insightful discussion of our presentation, and to two anonymous referees for their thoughtful comments on our manuscript. The authors' research was partially supported by grants from Natural Sciences & Engineering Research Council of Canada and Canadian Institute of Health Research.

## References

Akaike, H: Information theory as a extension of maximum likelihood principle. In: Petrove, B.N., Csaki, F. (eds.) Second Symposium of Information Theory, pp. 267–282. Akademiai Kiado, Budapest (1973)

Akaike, H: A new look at the statistical model identification. IEEE Trans. Autom. Control **19**, 716–723 (1974)

Antoniadis, A.: Wavelets in statistics: a review (with discussion). J. Italian Stat. Assoc. **6**, 97–144 (1997)

Antoniadis, A., Fan, J.: Regularization of wavelets approximations. J. Am. Stat. Assoc. **96**, 939–967 (2001)

Cantoni, E., Flemming, J.M., Ronchetti, E.: Variable selection for marginal longitudinal generalized linear models. Biometrika **61**, 507–514 (2005)

Craven, P., Wahba, G.: Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math. **31**, 377–403 (1979)

Crowder, M.J.: On use of a working correlation matrix in using generalized linear models for repeated measures. Biometrika **82**, 407–410 (1995)

Donoho, D.L., Johnstone, I.M.: Ideal spatial adaptation by wavelet shrinkage. Biometrika **81**, 425–455 (1994)

Dziak, J.J., Li, R., Qu, A.: An overview on quadratic inference function approaches for longitudinal data. In Frontiers of Statistics, Volume 1: New Developments in Biostatistics and Bioinformatics, J. Fan, J.S. Liu, and X. Lin (eds), Chapter 3, 49–72. 5 Toh Tuch Link, Singapore: World Scientific Publishing, (2009)

Fan, J.: Comments on "Wavelets in Statistics: A Review" by A. Antoniadis. J. Italian Stat. Assoc. **6**, 131–138 (1997)

Fan, J., Li, R.: Variable selection via non concave penalized likelihood and its oracle properties. J. Am. Stat. Assoc. **96**, 1348–1360 (2001)

Fan, J., Li, R.: New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis. J. Am. Stat. Assoc. **99**, 710–723 (2004)

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

Lv, J., Fan, Y.: A unified approach to model selection and sparse recovery using regularized least squares. Ann. Stat. **37**, 3498–3528 (2009)

McKenzie, E.: Some ARMA models for dependent sequences of Poisson counts. Adv. Appl. Probab. **20**, 822–835 (1988)

Nadarajah, T.: Penalized empirical likelihood based variable selection. M.Sc. thesis, Memorial University of Newfoundland, St. John's (2011)
Pan, W.: Akaike's information criterion in generalized estimating equations. Biometrics **57**, 120–125 (2001)
Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)
Sutradhar, B.C.: An overview on regression models for discrete longitudinal responses. Stat. Sci. **18**, 377–393 (2003)
Sutradhar, B. C. Dynamic Mixed Models for Familial Longitudinal Data. New York: Springer (2011)
Sutradhar, B.C., Das, K.: On the efficiency of regression estimators in generalized linear models for longitudinal data. Biometrika **86**, 459–465 (1999)
Sutradhar, B.C., Kovacevic, M.: Analysing ordinal longitudinal survey data: generalized estimating equations approach. Biometrika **87**, 837–848 (2000)
Tibshirani, R.: Regression shrinkage and selection via the lasso. J. R. Stat. Soc. B **58**, 267–288 (1996)
Variyath, A.M.: Variable selection in generalized linear models by empirical likelihood. Ph.D. thesis, University of Waterloo, Waterloo (2006)
Variyath, A.M., Chen, J., Abraham, B.: Empirical likelihood based variable selection. J. Stat. Plan. Infer. **140**, 971–981 (2010)
Wang, H., Leng, C.: Unified LASSO estimation via least squares approximation. J. Am. Stat. Assoc. **102**, 1039–1048 (2007)
Wang, L., Qu, A.: Consistent model selection and data-driven smooth tests for longitudinal data in the estimating equations approach. J. R. Stat. Soc. Ser. B **71**, 177–190 (2009)
Wang, L., Li, H., Huang, J.: Variable selection in nonparametric varying- coefficient models for analysis of repeated measurements. J. Am. Stat. Assoc. **103**, 1556–1569 (2008)
Wang, L., Zhou, J., Qu, A.: Penalized generalized estimating equations for high-dimensional longitudinal data analysis. Biometrics **68**, 353–360 (2012)
Wedderburn, R.W.M.: Quasi-likelihood functions, generalized linear models, and the Gauss–Newton method. Biometrika **61**(3), 439–444 (1974)
Xu, P., Wu, P., Wang, Y., Zhu, L.X.: A GEE based shrinkage estimation for the generalized linear model in longitudinal data analysis. Technical report, Department of Mathematics, Hong Kong Baptist University, Hong Kong (2010)
Xue, L., Qu, A., Zhou, J.: Consistent model selection for marginal generalized additive model for correlated data. J. Am. Stat. Assoc. **105**, 1518–1530 (2010)
Xiao, N., Zhang, D., Zhang, H.H.: Variable selection for semiparametric mixed models in longitudinal studies. Biometrics **66**, 79–88 (2009)
Zhang, H.H., Lu, W.: Adaptive Lasso for Cox's proportional hazards model. Biometrika **94**, 691–703 (2007)
Zou, H.: The adaptive Lasso and its oracle properties. J. Am. Stat. Assoc. **101**(476), 1418–1429 (2006)

# Index





B.C. Sutradhar (ed.), *Advances and Challenges in Parametric and Semi-parametric Analysis for Correlated Data*, Lecture Notes in Statistics 218,
DOI 10.1007/978-3-319-31260-6

**H**



**I**



**J**



**L**



**M**