



Fixed versus Mixed Effects Based Marginal Models for Clustered Correlated Binary Data: an Overview on Advances and Challenges

Brajendra C. Sutradhar
Memorial University St. John's Canada

Abstract

In a cross-sectional cluster setup, the binary responses from the individuals in a cluster become correlated as they share a common cluster effect, whereas longitudinal responses from an individual those form a cluster become correlated as the present and past responses are likely to maintain a suitable dynamic relationship. In both cluster and longitudinal setups, the marginal means may or may not be specified as the function of regression effects/parameters only. In a cluster setup, this depends on the distributional assumption of the random cluster effects and in a longitudinal setup this depends on the form such as linear or non-linear dynamic relationships used to construct a conditional model. However, over the last four decades, many studies arbitrarily pre-specified the marginal means as the function of regression effects only under both cluster and longitudinal setups and accommodated correlations also using arbitrarily selected 'working' correlation structures. This paper makes a thorough in-depth review of these decades long binary correlation models for consistent and efficient estimation of the regression effects. Both progress and drawbacks of these works are presented clearly showing how the inconsistency can arise if the pre-specified marginal fixed model is used when in fact such a marginal fixed effects model does not exist. This is because, some of the conditional random effects models in a cluster setup produce mixed effect models for the marginal means, and conditional non-linear dynamic models in a longitudinal setup produce history based marginal recursive/dynamic models. As the practitioners in both cluster and longitudinal setups deal with large data sets, it is demonstrated for their benefits how one can use the GQL (generalized quasi-likelihood) estimation approach both in cluster and longitudinal setups. Furthermore, there exist many studies using the Bayesian approach where unlike the aforementioned parametric correlation structure based inferences, the marginal mixed effects models have been used for inferences for correlated binary data without specifying their correlation structures, under both cluster and longitudinal setup. We also provide a brief review on this alternative approach.

AMS (2000) subject classification. Primary 62F10, 62H20; Secondary 62F12.
Keywords and phrases. Asymptotic properties such as consistency and normality, Correlated binary data, Cluster variance, Cross-sectional cluster setup,

Dynamic dependence, Dynamic relationship causing correlations in a longitudinal setup, Generalized quasi-likelihood and likelihood estimation, Random cluster effects causing correlations in a cluster setup, Regression effects

1 Introduction

In a cross-sectional cluster setup, the responses from the individuals in a given cluster are correlated as these responses share a common random cluster effect, whereas in a longitudinal setup the repeated responses collected from an individual form a cluster and these clustered responses from the same individual become correlated as they are likely to follow a dynamic relationship. Thus the correlation structure under cross-sectional clusters and longitudinal clusters are supposed to be different. In both setups, it is of primary objective to examine the regression effects (of the associated covariates) after accommodating the respective correlation structure.

To facilitate the discussion on the cluster regression models in a cross-sectional setup, suppose that there are I independent clusters and for a cluster i ($i = 1, \dots, I$), n_i denotes its size. Let y_{ij} denotes a binary response from the j -th ($i = 1, \dots, n_i$) individual of the i -th cluster. Further, let \mathbf{x}_{ij} be a p -dimensional fixed covariate vector, and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_u, \dots, \beta_p)'$ be the regression effect of \mathbf{x}_{ij} on y_{ij} , for all $i = 1, \dots, I$; $j = 1, \dots, n_i$. Notice that in this cluster setup, there is likely to be a cluster effect on the responses belonging to the same cluster. Let γ_i denote the random cluster effect of the i -th cluster which is shared by the responses belonging to this cluster. Thus, on top of $\boldsymbol{\beta}$, there is an influence of γ_i on the responses ($\{y_{ij}, j = 1, \dots, n_i\}$) belonging to the i -th cluster. This additional influence, along with the influence of \mathbf{x}_{ij} , is reflected on the binary response y_{ij} through a cluster-specific conditional mean model given by

$$\begin{aligned} E[Y_{ij}|\gamma_i] &= Pr[Y_{ij} = 1|\mathbf{x}_{ij}, \gamma_i] \\ &= p_{ij}^*(\boldsymbol{\beta}, \gamma_i) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)/[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)], \quad (1.1) \end{aligned}$$

where it is customarily assumed that $\gamma_i \stackrel{iid}{\sim} (0, \sigma_\gamma^2)$. Notice that Eq. 1.1 is a marginal model for y_{ij} conditional on γ_i . Hence this model may be referred to as the marginal-conditional (MC) binary logistic model. This model in Eq. 1.1 is also known as the so-called random effects model where σ_γ^2 plays multiple roles. More specifically, depending on the distributional assumption of γ_i or $p_{ij}^*(\gamma_i)$, (1) the unconditional mean, that is, $E[Y_{ij}] = Pr[Y_{ij} = 1|\mathbf{x}_{ij}]$,

may or may not be a function of σ_γ^2 ; (2) similarly $\text{var}[Y_{ij}]$ may or may not exhibit overdispersion (McCullagh and Nelder, 1989); but (3) because γ_i is the common clustered effect shared by all responses $\{y_{ij}$ and y_{ik} , for $k \neq j$; $j, k = 1, \dots, n_i\}$ y_{ij} and y_{ik} are correlated, and this within cluster correlation must be a function of σ_γ^2 . For the third reason (3), σ_γ^2 may preferably be referred to as the cluster correlation parameter.

As far as the applications of the model (1.1) is concerned, there are many practical situations, where one needs to analyze cluster specific binary data following (1.1). For example, in a chronic obstructive pulmonary disease (COPD) study (Liang et al., 1992; Ekholm et al., 1995; Sutradhar and Mukerjee, 2005), y_{ij} denotes the impaired pulmonary function (IPF) status (yes or no), and \mathbf{x}_{ij} is the vector of covariates such as gender, race, age, and smoking status, for the j -th sibling of the i -th COPD patient. In this problem it is likely that the IPF status for n_i siblings may be influenced by an unobservable random effect (γ_i) due to the i -th COPD patient. This common random effect makes the binary responses of any two siblings of the same patient correlated. It is of scientific interest to find the effects of the covariates on the binary responses (i.e., β in the mean function, $E[Y_{ij}] = Pr[Y_{ij} = 1|\mathbf{x}_{ij}]$), after taking the within cluster correlations into account. Thus, it is desired to derive the formula for $P[Y_{ij} = 1|\mathbf{x}_{ij}] = E[Y_{ij}|\mathbf{x}_{ij}]$ from Eq. 1.1 under a suitable distribution for γ_i , or using a nonparametric density. In this paper, we will confine our discussion to a parametric setup.

With regard to constructing a marginal (fixed or mixed) model for $E[Y_{ij}] = Pr[Y_{ij} = 1|\mathbf{x}_{ij}]$ we remark that because γ_i in Eq. 1.1 may be treated as an additive random covariate in the linear predictor $\mathbf{x}'_{ij} + \gamma_i$, it would be highly reasonable to assume that γ_i follows a normal (N) distribution, specifically $\gamma_i \stackrel{iid}{\sim} N(0, \sigma_\gamma^2)$ (Breslow and Clayton, 1993; Lee and Nelder, 1996; Sutradhar, 2004). This normality assumption produces a marginal mixed effects (MME) model. For convenience of further discussion in the next section and so on, we name this model as cluster model A (CM-A). Some studies such as Wang and Louis (2003) assume a so-called “bridge” distribution for γ_i , which provides a marginal fixed effects (MFE) model. We name this model as CM-B-1. Some other studies such as Prentice (1986) (see also Haseman and Kuper, 1979) assumed a beta-binary distribution which also produces a MFE model. We name this as CM-B-2.

There exists another group of studies (Zeger et al. (1988, Section 3.1), Neuhaus et al. (1991, Eqn. (4)), and Chen et al. (2011, Sections 2.1, 3.1)) where without any justification how the cluster effect may contribute to the modeling for mean, variance and correlations, they assumed a subject specific

(SS) arbitrary marginal fixed effects (AMFE) model for this mean function, and further assume that a user's choice 'working' correlation structure can be used for the estimation of the marginal fixed effects parameter. Thus, in this approach both means and correlations have 'working' models, which is a naive approach, and is bound to produce invalid such as inconsistent regression and correlation estimates in many practical situations where true means and correlations may be generated based on normal random cluster effects, γ_i . We name this naive/working model as CM-C. A brief review is given in Section 2 on the advantages and drawbacks of all these for cluster models (CM-A, CM-B-1, CM-B-2, CM-C) and their respective inferences.

We now consider the clustered binary data in a longitudinal setup, where a cluster is formed with repeated responses from the same individual. For convenience, we consider I independent individuals, whereas in the cross-sectional cluster setup, the same I was used to represent total number of independent clusters. However to form the i -th ($i = 1, \dots, I$) cluster for individual i , with repeated responses, we assume that these responses are recorded over a small period of time T , such as $T = 4$ or 5 weeks/months/years. Hence we denote the binary response recorded at time t ($t = 1, \dots, T$) from the i -th individual by y_{it} , whereas in cross-sectional cluster setup y_{ij} , $j = 1, \dots, n_i$, was used to represent the binary response from the j -th member of the i -th clusters. Next we denote by \mathbf{x}_{it} , a time dependent covariate vector corresponding to y_{it} . Here it is natural to expect that these repeated responses $\{y_{it}, t = 1, \dots, T, \}$ will be correlated most possibly through a dynamic relationship similar to time series data.

Similar to the cross-sectional clusters setup. it is of primary interest in this longitudinal setup, to find out the effect of \mathbf{x}_{it} on the binary response y_{it} . This is equivalent to compute the effect of \mathbf{x}_{it} on $E[Y_{it}|\mathbf{x}_{it}] = Pr[Y_{it} = 1|\mathbf{x}_{it}]$. Note however that unlike in the cross-sectional setup, in some situations it may be of interest to find the effects of the past history on the current response y_{it} . This is equivalent to compute the effect of the covariate history

$$H_{i,t}(\cdot) \equiv [\mathbf{x}_{i1}, \dots, \mathbf{x}_{i,t-1}, \mathbf{x}_{i,t}]$$

on y_{it} , i.e., to compute $E[Y_{it}|H_{it}] = Pr[Y_{it} = 1|H_{it}]$. Suppose that the effect of \mathbf{x}_{it} on y_{it} is measured by β which is similar but different than in the cross-sectional case where it represents the effect of \mathbf{x}_{ij} on y_{ij} , j being the j -th individual in the cluster. On top of this difference, a major difference between the models in both setups (cross-sectional and longitudinal clusters) arises because of the different nature of the binary responses under their respective clusters. More specifically, in the longitudinal setup, the correlation between y_{it} and $y_{i,t-1}$, for $t = 2, \dots, T$ arises because these responses

are likely to be directly related through a dynamic dependence relationship, whereas in cross-sectional setup, y_{ij} and y_{ik} for $j \neq k; j, k = 1, \dots, n_i$, are correlated as they share a common random cluster effect γ_i . For this reason, a big volume of existing studies (Laird and Ware, 1982; Stiratelli et al., 1984; Neuhaus, 2002; Parzen and et al. 2011) where longitudinal binary responses are analyzed using random/mixed effects model, fail to accommodate longitudinal correlations adequately. In particular these random effects based models have limited or no values to address the dynamic dependence among repeated responses.

As far as the marginal models for time specific binary means are concerned, there exists many situations where a marginal fixed effects (MFE) model involving only regression parameters (β) can be used for the marginal means. This is similar to the cross-sectional setup but correlation models are quite different under both cross-sectional and longitudinal setups. For MFE based longitudinal models, one may refer to specific dynamic models suggested, for example, by Bahadur (1961) (see also Cox, 1972) multivariate binary density (MBD) based model, Kanter (1975) for an observation driven dynamic (ODD) model, and Zeger et al. (1985) for a linear dynamic conditional probability (LDCP) model (see Sutradhar (2011, Section 7.2) for details). All these models yield the marginal mean function, i.e., the formula for the unconditional means ($E[Y_{it}|\mathbf{x}_{it}] = Pr[Y_{it} = 1|\mathbf{x}_{it}]$) for all $t = 1, \dots, T$, in terms of β parameter only. For our discussion involving a MFE model, we consider, for example the AR(1) (auto-regressive order 1) type linear dynamic model from Zeger et al. (1985), given by

$$\begin{aligned} Pr[Y_{i1} = 1|\mathbf{x}_{i1}] &= \tilde{p}_{i1}(\beta) \\ Pr[Y_{it} = 1|y_{i,t-1}, \mathbf{x}_{it}, \mathbf{x}_{i,t-1}] &= \tilde{p}_{it}(\beta) + \rho(y_{i,t-1} - \tilde{p}_{i,t-1}), \quad t = 2, \dots, T, \end{aligned} \quad (1.2)$$

with $\tilde{p}_{it}(\beta) = \exp(\mathbf{x}'_{it}\beta)/[1 + \exp(\mathbf{x}'_{it}\beta)]$, yielding the marginal means and variances as the function of β only, that is they are free of the dynamic dependence parameter ρ . We refer to this MFE model as longitudinal model 1 (LM(1)), and express it as

$$\begin{aligned} \text{LM(1): A marginal fixed effects (MFE) model} \\ E[Y_{it}] &= Pr[Y_{it} = 1] = \tilde{p}_{it}(\beta) = \exp(\mathbf{x}'_{it}\beta)/[1 + \exp(\mathbf{x}'_{it}\beta)] \quad (1.3) \\ \text{var}[Y_{it}] &= \tilde{p}_{it}(\beta)(1 - \tilde{p}_{it}(\beta)). \end{aligned}$$

There are, however, many other situations where the MFE models are not appropriate for the marginal means of the longitudinal binary data. This mostly happens when y_{it} depends on the history H_{it} , rather than on \mathbf{x}_{it} . In

this case, marginal means will be the function of both β and dynamic dependence parameter ρ . In cross-sectional cluster setup, a MME (marginal mixed effects model (CM(1)), means involving β and σ_γ^2) was used to represent the marginal means, but in the present longitudinal setup it is more appropriate to refer the model as the MD (marginal dynamic) or MR (marginal recursive) model. More specifically this MD/MR model for marginal means is derived from a non-linear conditional dynamic logit model Sutradhar and Farrell (2007) given by

$$\begin{aligned} Pr[Y_{i1} = 1 | \mathbf{x}_{i1}] &= \tilde{p}_{i1}(\beta) = \exp(\mathbf{x}'_{i1}\beta) / [1 + \exp(\mathbf{x}'_{i1}\beta)] \\ Pr[Y_{it} = 1 | y_{i,t-1}, \mathbf{x}_{it}] &= \frac{\exp(\mathbf{x}'_{it}\beta + \rho y_{i,t-1})}{1 + \exp(\mathbf{x}'_{it}\beta + \rho y_{i,t-1})}, \quad t = 2, \dots, T, \end{aligned} \quad (1.4)$$

(see also Loredo-Osti and Sutradhar, 2012; Fokianos and Kedem, 2003 in a time series setup) yielding the marginal dynamic/recursive (MD/MR) model:

$$\begin{aligned} \text{LM(2) : marginal dynamic/recursive (MD/MR) model} \\ \mu_{i1}(\beta) &= E[Y_{i1} | \mathbf{x}_{i1}] = Pr[Y_{i1} = 1 | \mathbf{x}_{i1}] = \tilde{p}_{i1}(\beta) \\ \mu_{it}(\beta, \rho) &= E[Y_{it} | H_{it}] = \tilde{p}_{it}(\beta) + \mu_{i,t-1}(\cdot)(\tilde{p}_{it}(\beta, \rho) - \tilde{p}_{it}(\beta)), \\ & \quad t = 2, \dots, T, \end{aligned} \quad (1.5)$$

[Sutradhar and Farrell (2007), Sutradhar (2011, Section 7.7.2)] where $\tilde{p}_{it}(\beta, \rho) = \exp(\mathbf{x}'_{it}\beta + \rho) / [1 + \exp(\mathbf{x}'_{it}\beta + \rho)]$.

For the sake of completeness, we also include another MFE model (on top of LM(1)) where, similar to CM-C in cross-section cluster setup, an arbitrary MFE (AMFE) model is used for the marginal means in terms of β , and longitudinal correlations are not modeled but substituted by certain ‘working’ correlations (Liang and Zeger, 1986) for the inference about β . We name this AMFE based model as LM(3). we briefly review these models LM(1), LM(2), and LM(3) in Section 3, along with available approaches for their parameters estimation. The advantages and drawbacks of these models and estimation approaches are also discussed.

Furthermore, because the CM-A as opposed to CM-B-1, CM-B-2 and CM-C in the cross-sectional cluster setup shows under the normality assumption of the random cluster effect (γ_i) that the marginal means contain both β and σ_γ^2 parameters, we consider this general model further in Section 4 and demonstrate how to construct a computationally simpler GQL (generalized quasi-likelihood) approach than maximum likelihood (ML) approach for the estimation of β , for known σ_γ^2 . When σ_γ^2 is unknown we provide a

consistent MM (method of moments) approach for its estimation. Asymptotic properties such as consistency of these GQL and MM estimators are also given in the same section. As far as the estimation of β and ρ under the general longitudinal model LM(2) is concerned, one may refer to Sutradhar and Farrell (2007) for their GQL and MM based estimation. In Section 5, we provide a brief review on the use of the GLMM (generalized linear mixed model) in a Bayesian frame work for inferences for correlated binary data in both cluster and longitudinal setups. Apart from computational complexity, it is outlined that because the random effects based models, in general, do not produce the time lag dependent correlations, choosing necessary proper prior distributions under longitudinal setup may be problematic. To tackle this problem to some extent, there appears a few studies using dynamic models for longitudinal binary data in a Bayesian frame work. This approach is discussed in brief as well. The paper concludes in Section 6.

2 Existing Marginal Models and Estimation for Cross-Sectional Clustered Binary Data

2.1. *CM-A: Population Average (PA) Based Marginal Mixed Effects (MME) Models* Refer to the marginal conditional model (1.1) which is written by adding the random cluster effect γ_i to the linear predictor $\mathbf{x}'_{ij}\beta$ used in the binary logistic probability function. This model is also as random effects model for binary data. Under the assumption that γ_i has a suitable probability distribution with probability density $g_D(\gamma_i|0, \sigma_\gamma^2)$, one may then write the likelihood function as

$$\begin{aligned} L(\beta, \sigma_\gamma^2) &= \prod_{i=1}^I Pr((y_{i1}, \dots, y_{ij}, \dots, y_{in_i})|\beta, \sigma_\gamma^2) \\ &= \prod_{i=1}^I \int_{\gamma_i} \prod_{j=1}^{n_i} Pr(y_{ij}|\gamma_i)g_D(\gamma_i)d\gamma_i = \prod_{i=1}^I \int_{\gamma_i} \prod_{j=1}^{n_i} [p_{ij}^*(\beta, \gamma_i)]^{y_{ij}} [1 - p_{ij}^*(\beta, \gamma_i)]^{1-y_{ij}} g_D(\gamma_i)d\gamma_i \\ &= \prod_{i=1}^I \int \frac{\exp\{\sum_{j=1}^{n_i} y_{ij}(\mathbf{x}'_{ij}\beta + \gamma_i)\}}{\prod_{j=1}^{n_i} \{1 + \exp(\mathbf{x}'_{ij}\beta + \gamma_i)\}} g_D(\gamma_i)d\gamma_i, \end{aligned} \tag{2.1}$$

which, and some of its modification such as penalized quasi-likelihood, hierarchical likelihood, were exploited by many researchers over the last four decades under varieties form of $g_D(\cdot)$, mainly for the estimation of β and σ_γ^2 . Among varieties form for $g_D(\cdot)$, normality assumption based $g_N(\gamma_i)$ is widely used. See for example, Stiratelli et al. (1984, Eqn. (3.1)), Breslow and Clayton (1993), Lee and Nelder (1996), Sutradhar and Mukerjee (2005). Some authors have used a specialized “bridge” (b) distribution with density, say

$g_b(\cdot)$ (Wang and Louis (2003, Eqns. (4.1)-(4.2)), Parzen and et al. (2011)), which, unlike the normal distribution ($g_N(\cdot)$), yields a marginal fixed effects model for the marginal means. But this “bridge” distribution appears to be restrictive and too technical for practical use. When $g_D(\cdot) \equiv g_N(\cdot)$, one obtains a MME (marginal mixed effects) based mean model given by

$$\begin{aligned}
 E[Y_{ij}] &= Pr[Y_{ij} = 1 | \mathbf{x}_{ij}] \quad j = 1, \dots, n_i \\
 &= \int_{\gamma_i} p_{ij}^*(\boldsymbol{\beta}, \gamma_i) g_N(\gamma_i) d\gamma_i = \int \left[\frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)}{[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)]} \right] dG_N(\gamma_i, \sigma_\gamma^2) \\
 &= \mu_{ij}(\boldsymbol{\beta}, \sigma_\gamma^2), \text{ (say), for all } j = 1, \dots, n_i \tag{2.2}
 \end{aligned}$$

$$\neq \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})]} = p_{ij}(\boldsymbol{\beta}). \tag{2.3}$$

Notice that the $g_N(\cdot)$ based likelihood estimates for $\boldsymbol{\beta}$ and σ_γ^2 (Sutradhar and Mukerjee, 2005), obtained by maximizing the likelihood function (2.1), can be used in the marginal mean $\mu_{ij}(\boldsymbol{\beta}, \sigma_\gamma^2)$ to interpret the effects of \mathbf{x}_{ij} on the binary mean response $E[Y_{ij}] = Pr[Y_{ij} = 1 | \mathbf{x}_{ij}]$. However, some studies attempt to estimate $\boldsymbol{\beta}$ in $p_{ij}(\boldsymbol{\beta})$ and interpret the effects of \mathbf{x}_{ij} on the mean response. But clearly it would be an incorrect or inconsistent estimate under normal random cluster effects, as $\boldsymbol{\beta}$ in Eq. 2.2 can not be estimated without estimating σ_γ^2 at least consistently. This is also evident from Zeger et al. (1988) that under normality, the population average (PA) based $\boldsymbol{\beta}$, that is, $\boldsymbol{\beta}^{PA}$ in $\mu_{ij}(\boldsymbol{\beta}^{PA}, \sigma_\gamma^2)$ in Eq. 2.3 has an approximate relationship with the subject specific (SS) $\boldsymbol{\beta}$, i.e., $\boldsymbol{\beta}^{SS}$ in $p_{ij}(\boldsymbol{\beta}^{SS})$, as

$$\boldsymbol{\beta}^{PA} \approx \boldsymbol{\beta}^{SS} / \left[\sqrt{1 + \left(\frac{16}{15}\right)^2 \frac{3}{\pi^2} \sigma_\gamma^2} \right]. \tag{2.4}$$

Thus, the desired $\boldsymbol{\beta}^{PA}$ can not be estimated without estimating σ_γ^2 under the present cluster setup.

However, there remains two relatively complex issues in this $\boldsymbol{\beta} = \boldsymbol{\beta}^{PA}$ estimation. First, $\mu_{ij}(\boldsymbol{\beta}, \sigma_\gamma^2)$ is an implicit function, hence it is not easy to interpret the role of $\boldsymbol{\beta}$ on this mean function in the presence of an estimate of σ_γ^2 . Second, the likelihood estimation for $\boldsymbol{\beta}$ and σ_γ^2 is complex. As a remedy, following a binomial approximation (BA) to the normal distribution of γ_i (Sutradhar (2011, Chapter 5, Eqn. (4.24))), one may compute this mean function $\mu_{ij}(\cdot)$ as follows and interpret it as the function of $\boldsymbol{\beta}$ for given σ_γ^2 . Use $\gamma_i^* = \gamma_i / \sigma_\gamma$ in Eq. 2.2 and express $p_{ij}^*(\boldsymbol{\beta}, \gamma_i)$ as $p_{ij}^*(\mathbf{x}_j; \boldsymbol{\beta}, \sigma_\gamma \gamma_i^*)$.

Further consider v_i as a binomial variable with parameters V and $1/2$, i.e., $v_i \sim \text{binomial}(V, 1/2)$. Next using

$$\gamma_i^* = \frac{v_i - V(1/2)}{\sqrt{V(1/2)(1/2)}} \equiv h(v_i) \quad (2.5)$$

we may express the MME based mean function in Eq. 2.2 as

$$\mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = \sum_{v_i=0}^V p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}, \quad (2.6)$$

where V is assumed to be relatively large such as $V = 10$. Note that this formula in Eq. 2.6 for the computation of the BA-based individual specific mean function is explicit, whereas the mean function was implicitly defined in Eq. 2.2. Thus one may use the likelihood estimates of $\boldsymbol{\beta}$ and σ_γ^2 obtained by exploiting the $g_N(\gamma_i)$ -based likelihood function (2.1) (Sutradhar and Mukerjee, 2005) into this mean formula in Eq. 2.6, and easily examine/interpret the effects of individual covariate \mathbf{x}_{ij} on the binary mean function $Pr[Y_{ij} = 1 | \mathbf{x}_{ij}] = E[Y_{ij} | \mathbf{x}_{ij}]$. We remark however that as the likelihood estimation is relatively complex, in Section 4 we demonstrate how one can develop a GQL approach (which produces consistent and highly efficient estimate, the ML estimate being optimal) for the estimation of the main regression parameters, and a MM approach for consistent estimation of σ_γ^2 . These GQL and MM approach exploit moments of the clustered binary data up to order 2 containing all squared and pairwise (from 2 individuals in the cluster) products of the binary responses. The following section provides a brief discussion on some other (than ML and GQL) existing estimated approaches along with their limitations.

2.1.1. Some Highly Competing Estimation Approaches in the Cross-Sectional Cluster Setup and their Drawbacks.

A BLUP (Best Linear Unbiased Prediction) Approach Under normality, i.e., when $g_D(\gamma_i) \equiv g_N(\gamma_i)$ in Eq. 2.1, many authors such as Stiratelli et al. (1984, Eqn. (3.1)), Schall (1991), Karim and Zeger (1992), Breslow and Clayton (1993), McGilchrist (1994), Kuk (1995), Lin and Breslow (1996), and Lee and Nelder (1996) have used a BLUP analogue estimation approach, where cluster/familial random effects are treated to be the fixed effects [Henderson (1963)] and the regression and variance components of the mixed model (2.1) are estimated based on the so-called estimates of the random effects. Because γ_i has to be estimated using the data from the i -th

cluster only, in general, this BLUP procedure may yield biased estimate for γ_i specially when i th cluster size is small, which may subsequently produce biased regression and variance estimates, variance estimates being more adversely affected than regression estimates. In order to remove biases in the estimates, Kuk (1995) and Lin and Breslow (1996), among others, provided certain asymptotic bias corrections both for the regression and the variance component estimates. But, as Breslow and Lin (1995, p. 90) have shown that the bias corrections appear to improve the asymptotic performance of the uncorrected quantities only when the true variance component is small, more specifically, less than or equal to 0.25. But in practice, variance component can be much larger. We further remark that the above BLUP analogue approaches are essentially using a likelihood technique for the present non-linear binary regression analysis. For example, Breslow and Clayton (1993) specifically use a PQL (penalized quasi-likelihood) approach, similarly Lee and Nelder (1996) use a HL (hierarchical likelihood) approach. These two approaches are similar, because in the first step, both PQL and HL approaches estimate the regression parameters and the random effects. The difference between the two approaches is that the PQL approach estimates them by maximizing a penalized quasi-likelihood function, whereas the HL approach maximizes a hierarchical likelihood function. In the second step, in estimating the variance of the random effects, the PQL approach maximizes a profile quasi-likelihood function, whereas the HL approach maximizes an adjusted profile hierarchical likelihood function. Thus both approaches encounter biases in the estimates in a similar way.

Another major drawback of the above mentioned BLUP oriented likelihood approaches is that no attempt is made to compute the marginal means from the respective likelihood function, whereas this computation of the marginal means is essential to interpret the effects of the covariates \mathbf{x}_{ij} on the marginal means $Pr[Y_{ij} = 1|\mathbf{x}_{ij}] = E[Y_{ij}|\mathbf{x}_{ij}] = \mu_{ij}(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma^2)$.

2.2. *CM-B-1: Subject Specific (SS) Marginal Fixed Effects (MFE) Model Based on “bridge” Random Cluster Effects* In some situations depending on the assumption about the distribution of the random effects (γ_i), the PA-based mixed model may yield a fixed effects model for the marginal means. More specifically, by using a slightly different (than $p_{ij}^*(\boldsymbol{\beta}, \gamma_i)$ in Eq. 2.2) marginal-conditional probability given by

$$\begin{aligned} Pr[Y_{ij} = 1|\mathbf{x}_{ij}, \gamma_i] &= p_{ij}^{**}(\boldsymbol{\beta}, \phi(\sigma_\gamma^2), \gamma_i) \\ &= \exp(\{\phi(\sigma_\gamma^2)\}^{-1} \mathbf{x}'_{ij} \boldsymbol{\beta} + \gamma_i) [1 + \exp(\{\phi(\sigma_\gamma^2)\}^{-1} \mathbf{x}'_{ij} \boldsymbol{\beta} + \gamma_i)], \\ 0 &< \phi(\sigma_\gamma^2) < 1, \end{aligned} \tag{2.7}$$

Wang and Louis (2003, Eqn. (4.2)) have shown that

$$\begin{aligned} Pr[Y_{ij} = 1 | \mathbf{x}_{ij}] &= \int_{-\infty}^{\infty} p_{ij}^{**}(\boldsymbol{\beta}, \gamma_i) g_D(\gamma_i) d\gamma_i \\ &= p_{ij}(\boldsymbol{\beta}) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta}) / [1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta})], \end{aligned} \tag{2.8}$$

a MFE (marginal fixed effects) based model involving only $\boldsymbol{\beta}$ parameters, when

$$g_D(\gamma_i) \Rightarrow g_b(\gamma_i),$$

$g_b(\gamma_i)$ being the so-called “bridge” density of the form

$$g_b(\gamma_i) = \frac{1}{2\pi} \frac{\sin(\phi\pi)}{\cosh(\phi\gamma_i) + \cos(\phi\pi)}; \quad 0 < \phi(\sigma_\gamma^2) < 1, \quad -\infty < \gamma_i < \infty. \tag{2.9}$$

where ϕ is related to σ_γ^2 through the relationship, $\sigma_\gamma^2 = \pi^2(\phi^{-2} - 1)/3$.

We remark that the MFE model for the marginal means given in Eq. 2.8 is simpler than the MME model (2.2) to interpret the effects of the covariates \mathbf{x}_{ij} on the SS binary means $E[Y_{ij} | \mathbf{x}_{ij}]$ as it is σ_γ^2 free. Also the likelihood estimate of $\boldsymbol{\beta}$ obtained by exploiting the likelihood function

$$L(\boldsymbol{\beta}, \phi/\sigma_\gamma^2) = \prod_{i=1}^I \int \frac{\exp\{\sum_{j=1}^{n_i} y_{ij}(\phi^{-1}\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)\}}{\prod_{j=1}^{n_i} \{1 + \exp(\phi^{-1}\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)\}} g_b(\gamma_i) d\gamma_i, \tag{2.10}$$

can be used for $\boldsymbol{\beta}$ in the MFE model (2.8). This is because the MFE model (2.8) can be obtained from the joint probability function used in the likelihood function (2.10). That is $\boldsymbol{\beta}^{PA} \equiv \boldsymbol{\beta}^{SS}$.

However, some of the major drawbacks of this “bridge” random effects based fixed model are:

(i) Notice that γ_i involved in the linear mixed predictor in the conditional probability function in Eq. 2.7 may be treated as a random covariate, whereas \mathbf{x}_{ij} 's are known to be fixed covariates. As far as its distributional properties are concerned, even though the bridge distribution (2.9) (which has a complex trigonometrical ratio form) technically yields the marginal fixed effects model, the suitability of this distributional assumption, as opposed to the normality assumption (e.g., Breslow and Clayton, 1993; Lee and Nelder, 1996 in GLMM setup) in practical contexts, is not discussed adequately in the literature.

(ii) Even though the MFE model (2.8) does not contain ϕ/σ_γ^2 , this parameter has to be estimated anyway as it contains in the likelihood function or any possible correlation structure. Moreover, the likelihood estimation using the likelihood function (2.10) would be much more complex than using the normal clusters based likelihood function.

(iii) β parameter in the MFE model (2.8) could be estimated using a QL (quasi-likelihood) approach rather than using the complicated likelihood approach, provided one could compute the pair-wise correlations among the clustered binary responses. The computation of these correlations appear to be complicated under this “bridge” cluster effects assumption.

2.3. *CM-B-2: SS Marginal Fixed Effects (MFE) Model Based on Beta-Binary Random Clustered Probability Function* Similar to the CM-B-1 model (Wang and Louis, 2003, 2004), there exists some early studies (Prentice, 1986; Haseman and Kuper, 1979) in the context of clustered binary regression analysis in a longitudinal setup, where, in order to obtain a MFE model for binary means, an extended assumption about the distribution of a function of γ_i , specifically for $p_{ij}^*(\beta, \gamma_i)$, was used. For a bounded scale parameter τ , as a function of σ_γ^2 , say $\tau(\sigma_\gamma^2)$ satisfying the range $0 < \tau(\sigma_\gamma^2) < 1$, suppose that $p_{ij}^*(\beta, \gamma)$ in Eq. 1.1 (see also Eq. 2.2) follows a beta-distribution (\tilde{g}_B) of first kind with parameters $(\{\tau(\sigma_\gamma^2)\}^{-1} - 1)p_{ij}(\beta)$ and $(\{\tau(\sigma_\gamma^2)\}^{-1} - 1)q_{ij}(\beta)$ with $q_{ij}(\beta) = 1 - p_{ij}(\beta)$. More specifically,

Distributional assumption for the random logistic function $p_{ij}^*(\beta, \gamma_i)$:

A beta-distribution of first kind

$$\tilde{g}_B(p_{ij}^*; \tau, p_{ij}(\beta)) = \frac{p_{ij}^{*(\tau^{-1}-1)p_{ij}-1} (1 - p_{ij}^*)^{(\tau^{-1}-1)q_{ij}-1}}{B((\tau^{-1} - 1)p_{ij}, (\tau^{-1} - 1)q_{ij})}; \quad 0 \leq p_{ij}^* \leq 1, \quad (2.11)$$

which yields the marginal probability

$$\begin{aligned} Pr[Y_{ij} = 1 | \mathbf{x}_{ij}] &= \int_0^1 p_{ij}^*(\beta, \gamma_i) \tilde{g}_B(p_{ij}^*) dp_{ij}^* \\ &= p_{ij}(\beta) = \exp(\mathbf{x}'_{ij}\beta) / [1 + \exp(\mathbf{x}'_{ij}\beta)], \end{aligned} \quad (2.12)$$

as in Eq. 2.8 under CM-B-1 model, which is the same as the marginal probability in Eq. 2.3.

Consequently, under this mixed model approach, one may examine the effects of \mathbf{x}_{ij} on the marginal response means $E[Y_{ij} | \mathbf{x}_{ij}] = Pr[Y_{ij} = 1 | \mathbf{x}_{ij}]$ by computing β parameter involved in the simpler MFE model Eq. 2.12, i.e., in $p_{ij}(\beta) = \exp(\mathbf{x}'_{ij}\beta) / [1 + \exp(\mathbf{x}'_{ij}\beta)]$. This estimation can be achieved by maximizing the likelihood function

$$\begin{aligned} L(\beta, \tau) &= \prod_{i=1}^I \prod_{j=1}^{n_i} \int_0^1 \frac{p_{ij}^{*\{\tau^{-1}-1\}p_{ij}+y_{ij}-1} (1 - p_{ij}^*)^{\{\tau^{-1}-1\}q_{ij}+y_{ij}+1}-1}}{B((\tau^{-1} - 1)p_{ij}, (\tau^{-1} - 1)q_{ij})} dp_{ij}^* \\ &= \prod_{i=1}^I \prod_{j=1}^{n_i} \frac{\Gamma(\tau^{-1} - 1)p_{ij} + y_{ij} \Gamma(\tau^{-1} - 1)q_{ij} + y_{ij} + 1}{\Gamma(\tau^{-1} - 1) + 2y_{ij} + 1 B((\tau^{-1} - 1)p_{ij}, (\tau^{-1} - 1)q_{ij})}, \end{aligned} \quad (2.13)$$

with respect to β and τ .

However, some of the major drawbacks with this beta-binary approach based marginal fixed model, are:

(i) The likelihood estimation by exploiting the likelihood function (2.13) is complex. See, for example, Sutradhar and Das (1997), for an approximate QL approach estimation in a similar setup.

(ii) The assumption that the whole conditional probability $p_{ij}^*(\beta, \gamma_i) = \exp(\mathbf{x}'_{ij}\beta + \gamma_i)/[1 + \exp(\mathbf{x}'_{ij}\beta + \gamma_i)]$ in Eq. 1.1 follows a beta distribution, rather than assuming a distribution for γ_i such as normality, appears to be too restrictive and hence it may be impractical, in order to obtain a marginal fixed model.

(iii) The pair-wise correlations among the clustered responses are not understood as they may not be easy to compute. This is because, such a computation will require first the computation of the correlations between $p_{ij}^*(\beta, \gamma_i) = \exp(\mathbf{x}'_{ij}\beta + \gamma_i)/[1 + \exp(\mathbf{x}'_{ij}\beta + \gamma_i)]$ and $p_{ik}^*(\beta, \gamma_i) = \exp(\mathbf{x}'_{ik}\beta + \gamma_i)/[1 + \exp(\mathbf{x}'_{ik}\beta + \gamma_i)]$ for $j \neq k; j, k = 1, \dots, n_i$ which is not possible without making further joint, say bivariate, distributional assumptions for p_{ij}^* and p_{ik}^* .

(iv) Even though ML approach may give an estimate for τ , estimating σ_γ^2 , the cluster variance, is, however, not possible without knowing the specific relationship between τ and σ_γ^2 , $\tau(\sigma_\gamma^2)$ being currently an implicit function only.

2.4. *CM-C: A SS Arbitrary Marginal Fixed Effects (AMFE) Model* Some authors, for example in an early study, Zeger et al. (1988, Section 3.1), considered a clustered binary data analysis and suggested to use the MFE model for the marginal means, given by

$$E[Y_{ij}] = Pr[Y_{ij} = 1 | \mathbf{x}_{ij}] = p_{ij}(\beta) = \exp(\mathbf{x}'_{ij}\beta) / [1 + \exp(\mathbf{x}'_{ij}\beta)], \quad (2.14)$$

for all $j = 1, \dots, n_i$. Because the clustered responses are correlated, for inferences about β , these authors have suggested the use of a ‘working’ correlation structure based GEE (generalized estimating equations) approach discussed by Liang and Zeger (1986). It is clear that neither the means nor the correlations were modeled under this approach. Hence, the MFE model in Eq. 2.14 is purely an arbitrary fixed effects model. Notice that even though this model in Eq. 2.14 appears to be the same as the marginal models in Eqs. 2.8 and 2.12, it is however an assumed model without showing its connection with the marginal-conditional model (1.1), whereas the models in Eqs. 2.8 and 2.12 were derived from Eq. 1.1 under certain distributional (“bridge” and beta-binary) assumptions for the cluster effects γ_i . Furthermore, as shown by Eq. 2.2 (see also Eq. 2.6), this marginal model (2.14) can not be derived from

Eq. 1.1 under normality assumption for γ_i , and in such normal based cases, the MFE model (2.14) would produce biased and hence inconsistent regression estimate due to ignoring σ_γ^2 from the marginal mean function. In this token we remark that the marginal model (2.14) suggested by Zeger et al. (1988, Section 3.1), therefore, gives a wrong impression that it can be used for any clustered correlated binary data. This impression is further noticed in a recent paper by Chen et al. (2011, Section 2.1), where this marginal model (2.14) was used under a clustered correlated ‘response process’ without misclassification, and was generalized for a possible ‘misclassification process’. The difference between their models is that Zeger et al. (1988, Section 3.1) suggested a ‘working’ correlation structure to construct their GEE, whereas Chen et al. (2011, Section 2.1) suggested a ‘working’ odds ratio based ‘working’ covariance (or bivariate probability) structure to develop the GEE for β estimation. However, in a longitudinal setup for binary data, it is well known that these GEE approaches may produce inefficient estimates as compared to the so-called independence assumption based simpler MM and QL approaches (Sutradhar and Das (1999), Sutradhar (2011, Section 7.3.6), Sutradhar and Zheng (2018), Sutradhar (2014, Section 4.2)), which is a serious inference drawback.

To have a feel about the possible adverse performance of the odds ratio based GEE approach in the present cross-sectional cluster setup, we consider the most likely practical case with normal random cluster effects ($\gamma_i \sim N(0, \sigma_\gamma^2)$) as discussed in Section 2.1, and compute the odds ratio as follows to examine whether one can express log of this odds ratio in a linear form as suggested in Chen et al. (2011, Section 2.1). Because, the odds ratio for y_{ij} and y_{ik} ($j \neq k; j, k = 1, \dots, n_i$) has the formula

$$\psi_{ijk} = \frac{Pr(Y_{ij} = 1, Y_{ik} = 1)Pr(Y_{ij} = 0, Y_{ik} = 0)}{Pr(Y_{ij} = 1, Y_{ik} = 0)Pr(Y_{ij} = 0, Y_{ik} = 1)}, \tag{2.15}$$

we compute these joint probabilities involved in Eq. 2.15, by exploiting the independence of y_{ij} and y_{ik} conditional on γ_i in Eq. 1.1 and then taking population average over the normal distribution of γ_i . More specifically, for $\gamma_i^* = \frac{v_i - V(1/2)}{\sqrt{V(1/2)(1/2)}} \equiv h(v_i)$ as in Eq. 2.5, following Eq. 2.6, we write

$$\begin{aligned} \lambda_{ijk}^{(1,1)}(\beta, \sigma_\gamma^2) &= Pr(Y_{ij} = 1, Y_{ik} = 1) \tag{2.16} \\ &= \sum_{v_i=0}^V \frac{\exp((\mathbf{x}'_{ij} + \mathbf{x}'_{ik})\beta + 2\sigma_\gamma h(v_i))}{[1 + \exp(\mathbf{x}'_{ij}\beta + \sigma_\gamma h(v_i))][1 + \exp(\mathbf{x}'_{ik}\beta + \sigma_\gamma h(v_i))]} \binom{V}{v_i} \\ &\quad (1/2)^{v_i} (1/2)^{V-v_i} \end{aligned}$$

$$\lambda_{ijk}^{(0,0)}(\boldsymbol{\beta}, \sigma_\gamma^2) = Pr(Y_{ij} = 0, Y_{ik} = 0) \tag{2.17}$$

$$= \sum_{v_i=0}^V \frac{1}{[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \sigma_\gamma h(v_i))][1 + \exp(\mathbf{x}'_{ik}\boldsymbol{\beta} + \sigma_\gamma h(v_i))]} \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}$$

$$\lambda_{ijk}^{(1,0)}(\boldsymbol{\beta}, \sigma_\gamma^2) = Pr(Y_{ij} = 1, Y_{ik} = 0) \tag{2.18}$$

$$= \sum_{v_i=0}^V \frac{\exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \sigma_\gamma h(v_i))}{[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \sigma_\gamma h(v_i))][1 + \exp(\mathbf{x}'_{ik}\boldsymbol{\beta} + \sigma_\gamma h(v_i))]} \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}$$

$$\lambda_{ijk}^{(0,1)}(\boldsymbol{\beta}, \sigma_\gamma^2) = Pr(Y_{ij} = 0, Y_{ik} = 1) \tag{2.19}$$

$$= \sum_{v_i=0}^V \frac{\exp(\mathbf{x}'_{ik}\boldsymbol{\beta} + \sigma_\gamma h(v_i))}{[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \sigma_\gamma h(v_i))][1 + \exp(\mathbf{x}'_{ik}\boldsymbol{\beta} + \sigma_\gamma h(v_i))]} \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i},$$

yielding the odds ratio as

$$\psi_{ijk} = \frac{\lambda_{ijk}^{(1,1)}(\boldsymbol{\beta}, \sigma_\gamma^2)\lambda_{ijk}^{(0,0)}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\lambda_{ijk}^{(1,0)}(\boldsymbol{\beta}, \sigma_\gamma^2)\lambda_{ijk}^{(0,1)}(\boldsymbol{\beta}, \sigma_\gamma^2)}. \tag{2.20}$$

In Chen et al. (2011), these joint probabilities are unknown, and $\lambda_{ijk}^{(1,1)}(\boldsymbol{\beta}, \sigma_\gamma^2)$ is expressed as a function of $\psi_{ijk}, p_{ij}(\boldsymbol{\beta}), p_{ik}(\boldsymbol{\beta})$, and then estimate this joint probability by using an estimate of ψ_{ijk} . For the estimation, in this approach they use an ‘working’ log linear model, namely

$$\psi_{ijk} = \exp(\mathbf{u}'_{ijk}\boldsymbol{\alpha}),$$

where \mathbf{u}_{ijk} is a set of suitable covariates and $\boldsymbol{\alpha}$ is set of new regression parameters. Notice however that the odds ratio, which is known, and computed by Eq. 2.20, is far different than what is modeled using a log linear relationship. Thus, this aforementioned example demonstrates that the ‘working’ odds ratio approach by fitting a log linear model for odds ratio estimation may yield inconsistent estimate for the joint probability, restricting its use for GEE construction.

3 Existing Marginal Models and Estimation for Longitudinal Clustered Binary Data

As opposed to cross-sectional clustered data collection, in a longitudinal setup a cluster is formed with repeated responses over a period of time T ,

from an individual i , for all $i = 1, \dots, I$. As explained in Section 1, specifically in Eqs. 1.2 and 1.4, the correlations among repeated responses arise through certain dynamic relationships between past and present responses of the same individual. We refer to Sutradhar (2010, Section 2.2) and Sutradhar and Zheng (2018), for example, for some low order non-stationary (time dependent covariates based) correlation such as AR(1) (auto-regressive order 1), MA(1) (moving average order 1), and exchangeable/equi-correlation structures for repeated binary data. Similar but ‘working’ correlation structures for stationary/non-stationary repeated binary data are also found in Liang and Zeger (1986), Zeger et al. (1985), and Lin and Carroll (2001), for example.

As far as the marginal models for the binary means at a given time t are concerned, similar to the cross-sectional clustered binary models, these models can be (1) MFE model (LM(1)) such as in Eq. 1.4 obtained from a linear dynamic conditional model, or (2) MD/MR model (LM(2)) such as in Eq. 1.5 obtained from a non-linear dynamic conditional logits model, or (3) AMFE (arbitrary marginal fixed effects) model (LM(3)), where correlations are thought not to play any roles in mean specification. For convenience, we refer to Zeger et al. (1985), Sutradhar (2010, 2011), and Sutradhar and Zheng (2018), for LM(1) type MFE model; Fokianos and Kedem (2003), Sutradhar and Farrell (2007), and Sutradhar (2011, Section 7.7.2), for LM(2) type MD/MR (marginal dynamic/recursive) model; and Liang and Zeger (1986), and Lin and Carroll (2001), for the AMFE model.

We further remark that some studies (e.g., Laird and Ware, 1982; Stiratelli et al., 1984; Parzen and et al., 2011) have used random effects models those are similar to the cross-sectional clustered models discussed in Section 2. However, these models can accommodate only EQC/exchangeable type correlations, and hence they have limited or no values for longitudinal data where one encounters correlations through time series type dynamic models. As discussed in Section 2, these models also have limitations for specification of marginal means for the cross-sectional cluster binary data. Thus, we do not include these models any further in our discussion.

3.1. LM(1): Time Specific (TS) Marginal Fixed Effects Model Recall from Section 1 that the linear dynamic conditional probability ($Pr[Y_{it} = 1|y_{i,t-1}]$) model (1.2) relates $y_{i,t-1}$ to y_{it} , for $t = 2, \dots, T$, through an AR(1) type relationship. This model produces a MFE model for the binary means at time t as in Eq. 1.4, specifically it yields

$$E[Y_{it}] = Pr[Y_{it} = 1] = \tilde{p}_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta})/[1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})],$$

where β explains the effects of the fixed covariates \mathbf{x}_{it} on y_{it} , specifically on $E[Y_{it}]$. Notice that the conditional model (1.2) also produces the lag $(t - u)$ correlations between two responses y_{iu} and y_{it} , for $u < t$, say, as

$$\text{corr}(Y_{iu}, Y_{it}) = \rho^{t-u} \left[\frac{\sigma_{iuu}}{\sigma_{itt}} \right]^{\frac{1}{2}}, \tag{3.1}$$

[Sutradhar (2011), Eqn. (7.73)] where σ_{itt} is the variance of y_{it} , for all $t = 1, \dots, T$, and is given in Eq. 1.4 as $\sigma_{itt} = \tilde{p}_{it}(1 - \tilde{p}_{it})$.

As far as the estimation of β is concerned, if one is willing to ignore the correlation structure (3.1) (which is equivalent to use $\rho = 0$ in Eq. 3.1), then (1) one may solve the MM (method of moments) estimating equation

$$\sum_{i=1}^I \sum_{t=1}^T \frac{\partial \tilde{p}_{it}}{\partial \beta} (y_{it} - \tilde{p}_{it}) = 0, \tag{3.2}$$

or (2) a QL (quasi-likelihood) estimating equation

$$\sum_{i=1}^I \sum_{t=1}^T \frac{\partial \tilde{p}_{it}}{\partial \beta} \sigma_{itt}^{-1} (y_{it} - \tilde{p}_{it}) = 0, \tag{3.3}$$

[Wedderburn (1974)] to obtain MM or QL estimate for β . Note that both MM (3.2) and QL (3.3) estimating equations are unbiased as $E[Y_{it}] = \tilde{p}_{it}(\beta)$ yielding $E[Y_{it} - \tilde{p}_{it}(\beta)] = 0$. Consequently, as $I \rightarrow \infty$, MM and QL estimators will be consistent under some mild regularity conditions. But these estimators will be inefficient as compared to other moments based estimators obtained by accommodating the underlying correlation structure (3.1).

Let $\Sigma_i(\beta, \rho) = (\sigma_{iut}(\cdot))$ denote the $T \times T$ covariance matrix constructed based on the correlation structure from Eq. 3.1. One may then obtain a highly efficient estimate of β by solving the GQL estimating equation

$$\sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \Sigma_i^{-1}(\beta, \rho) (\mathbf{y}_i - \tilde{\mathbf{p}}_i(\beta)) = 0, \tag{3.4}$$

[Sutradhar (2003, Section 3)] where

$$\mathbf{y}_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})', \quad \tilde{\mathbf{p}}_i(\beta) = (\tilde{p}_{i1}, \dots, \tilde{p}_{it}, \dots, \tilde{p}_{iT})'$$

Alternatively, one may obtain an optimal estimate of β by solving a likelihood estimating equation for $\theta = (\beta', \rho)'$ given by

$$\frac{\partial \log L(\beta, \rho)}{\partial \theta} = 0, \tag{3.5}$$

where

$$L(\boldsymbol{\beta}, \rho) = \prod_{i=1}^I [f(y_{i1}) \prod_{t=2}^T f(y_{it}|y_{i,t-1})], \tag{3.6}$$

with

$$f(y_{i1}) = \tilde{p}_{i1}^{y_{i1}} [1 - \tilde{p}_{i1}]^{1-y_{i1}}$$

as the binary density at $t = 1$, and conditional density of the form

$$f(y_{it}|y_{i,t-1}) = [\lambda_{it}^*(\boldsymbol{\beta}, \rho|y_{i,t-1})]^{y_{it}} [1 - \lambda_{it}^*(\boldsymbol{\beta}, \rho|y_{i,t-1})]^{1-y_{it}}, \tag{3.7}$$

for $t = 2, \dots, T$, with $\lambda_{it}^*(\boldsymbol{\beta}, \rho|y_{i,t-1}) = P[y_{it} = 1|y_{i,t-1}]$ as the conditional probability as given in Eq. 1.2.

Note that as the likelihood estimation is more complex than the GQL estimation approach, GQL approach becomes practically useful as it also provides more efficient estimates than the MM and QL approaches. We further remark that under this MFE model (LM(1)), where correlations are specified by Eq. 3.1, the so-called GEE approach (Liang and Zeger (1986)) becomes redundant because no ‘working’ correlation structure is needed when true correlation structure is known.

3.2. *LM(2): Time Specific (TS) Marginal Dynamic/Recursive (MD/MR) Model* Many existing studies such as Liang and Zeger (1986), Zeger et al (1988, Section 3.1), Lipsitz et al. (1991), and Yi and Cook (2002), among others, have specified the marginal binary means as a function of regression parameters only, specifically as

$$E[Y_{it}] = Pr[Y_{it} = 1] = \tilde{p}_{it} = \exp(\mathbf{x}'_{it}\boldsymbol{\beta})/[1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})], \tag{3.8}$$

which is similar to Eq. 2.14 in a cross-sectional cluster setup, and estimated $\boldsymbol{\beta}$ using ‘working’ correlations based so-called GEE approach. Thus, it is clear that these and other follow up works neither did model the marginal means nor the correlation structure for the underlying longitudinal binary responses. Between these two specifications, i.e., specifying the marginal means by Eq. 3.8, and specifying a ‘working’ correlations matrix, for repeated binary data, the former specification can seriously effect the validation of the regression estimates when the marginal means for correlated binary data can not be specified as a function of regression parameters only. One such important situation is indicated by Eq. 1.5 under Section 1, where marginal means for the longitudinal binary data appear to involve both regression ($\boldsymbol{\beta}$)

and correlation (ρ) parameters. More specifically, the dynamic logit model (1.4) yields

$$\begin{aligned} \mu_{i1}(\boldsymbol{\beta}) &= E[Y_{i1}|\mathbf{x}_{i1}] = \tilde{p}_{i1}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'_{i1}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{i1}\boldsymbol{\beta})]} \\ \mu_{i2}(\boldsymbol{\beta}, \rho) &= E[Y_{i2}|H_{i2}] = \frac{\exp(\mathbf{x}'_{i2}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{i2}\boldsymbol{\beta})]} \\ &\quad + \mu_{i1}(\boldsymbol{\beta}) \left(\frac{\exp(\mathbf{x}'_{i2}\boldsymbol{\beta} + \rho)}{[1 + \exp(\mathbf{x}'_{i2}\boldsymbol{\beta} + \rho)]} - \frac{\exp(\mathbf{x}'_{i2}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{i2}\boldsymbol{\beta})]} \right) \\ \mu_{i3}(\boldsymbol{\beta}, \rho) &= E[Y_{i3}|H_{i3}] = \frac{\exp(\mathbf{x}'_{i3}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{i3}\boldsymbol{\beta})]} \\ &\quad + \mu_{i2}(\boldsymbol{\beta}, \rho) \left(\frac{\exp(\mathbf{x}'_{i3}\boldsymbol{\beta} + \rho)}{[1 + \exp(\mathbf{x}'_{i3}\boldsymbol{\beta} + \rho)]} - \frac{\exp(\mathbf{x}'_{i3}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{i3}\boldsymbol{\beta})]} \right), \end{aligned} \quad (3.9)$$

and so on. Clearly, these means show a recursive relationship. These marginal means take the so-called fixed effects model form, i.e., $\mu_{it}(\cdot) = \tilde{p}_{it}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})]}$, for all $t = 1, \dots, T$, only when $\rho = 0$. Otherwise, $\mu_{i2}(\cdot)$ (marginal mean at time point $t = 2$) is $\tilde{p}_{i2}(\boldsymbol{\beta})$ plus an increment or decrement due to ρ weighted by previous mean μ_{i1} , and so on. It is then clear that one can no longer estimate the regression effects $\boldsymbol{\beta}$ by using the so-called GEE approach (Liang and Zeger, 1986). This is because the binary means (1.5) under this BDL model are not free of correlation parameter, whereas GEE approach is developed for the estimation of fixed effects based marginal means all containing only regression parameters $\boldsymbol{\beta}$, correlations are being nuisance. In summary, any $\boldsymbol{\beta}$ estimates for the mean model (3.8) when in fact the mean model by Eq. 1.5 is true, would produce inconsistent regression estimates, which is a serious inference issue.

For the regression analysis of the BDL (binary dynamic logit) model (1.4) which produces the marginal recursive (MR) means as in Eq. 1.5 involving both $\boldsymbol{\beta}$ and ρ , (Sutradhar and Farrell, 2007) (see also Amemiya (1985, p. 422) in a time series setup) have develop a GQL (generalized quasi-likelihood) estimation approach which exploits the true correlation structure of the data. For $u < t$, the formula for the lag $(t-u)$ auto-correlation between y_{iu} and y_{it} , is given by

$$\begin{aligned} \text{corr}(Y_{iu}, Y_{it}) &= \tilde{\rho}_{t-u}(\boldsymbol{\beta}, \rho) \\ &= \sqrt{\frac{\mu_{iu}(\cdot)(1 - \mu_{iu}(\cdot))}{\mu_{it}(\cdot)(1 - \mu_{it}(\cdot))}} \Pi_{v=u+1}^t (\tilde{p}_{iv}(\boldsymbol{\beta}, \rho) - \tilde{p}_{iv}(\boldsymbol{\beta})), \end{aligned} \quad (3.10)$$

[Sutradhar and Farrell (2007)] where

$$\tilde{p}_{it}(\boldsymbol{\beta}) = \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta})}{[1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})]}, \text{ for all } t = 1, \dots, T,$$

$$\begin{aligned} \tilde{p}_{it}(\boldsymbol{\beta}, \rho) &= \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \rho) / [1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + \rho)], \text{ for all } t = 2, \dots, T, \\ \mu_{i1}(\boldsymbol{\beta}) &= \tilde{p}_{i1}(\boldsymbol{\beta}), \end{aligned}$$

and $\mu_{it}(\boldsymbol{\beta}, \rho)$ for $t = 2, \dots, T$, have the recursive/dynamic formula as in Eq. 1.5. Subsequently, one may construct the $T \times T$ true covariance matrix of the response vector $\mathbf{y}_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$, of the i -th individual, as

$$\tilde{\Sigma}_i(\boldsymbol{\beta}, \rho) = \text{cov}[\mathbf{Y}_i] = \mathbf{A}_i^{\frac{1}{2}}(\boldsymbol{\beta}, \rho) \tilde{\rho}_M(\boldsymbol{\beta}, \rho) \mathbf{A}_i^{\frac{1}{2}}(\boldsymbol{\beta}, \rho), \quad (3.11)$$

where

$$\begin{aligned} \mathbf{A}_i(\boldsymbol{\beta}, \rho) &= \text{diag}[\mu_{i1}(\cdot)(1 - \mu_{i1}(\cdot)), \dots, \mu_{iT}(\cdot)(1 - \mu_{iT}(\cdot))] \\ \tilde{\rho}_M(\boldsymbol{\beta}, \rho) &= \begin{pmatrix} 1 & \tilde{\rho}_1 & \tilde{\rho}_2 & \dots & \tilde{\rho}_\ell & \dots & \tilde{\rho}_{T-2} & \tilde{\rho}_{T-1} \\ \cdot & 1 & \tilde{\rho}_1 & \dots & \tilde{\rho}_{\ell-1} & \dots & \tilde{\rho}_{T-3} & \tilde{\rho}_{T-2} \\ \vdots & \vdots & \vdots & \dots & \vdots & \dots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \dots & \cdot & \dots & 1 & \tilde{\rho}_1 \\ \cdot & \cdot & \cdot & \dots & \cdot & \dots & \cdot & 1 \end{pmatrix}. \end{aligned} \quad (3.12)$$

One may then exploit the mean vector $\boldsymbol{\mu}_i(\boldsymbol{\beta}, \rho) = E[\mathbf{Y}_i] = (\mu_{i1}(\boldsymbol{\beta}), \mu_{i2}(\boldsymbol{\beta}, \rho), \dots, \mu_{iT}(\boldsymbol{\beta}, \rho))'$ and the above covariance matrix $\tilde{\Sigma}_i(\boldsymbol{\beta}, \rho)$ from Eq. 3.11, for the GQL estimation of the main regression parameter $\boldsymbol{\beta}$. The dynamic dependence or correlation index parameter ρ can be estimated by using the method of moments (MM). See Section 5, for specific GQL and MM estimating equations for these parameters. In the same section, it is shown that as $I \rightarrow \infty$, the GQL estimator of $\boldsymbol{\beta}$ and the MM estimator of ρ are consistent under some mild regularity conditions. The asymptotic normality of the GQL estimator of the main parameter $\boldsymbol{\beta}$ is also given for convenience of the construction of confidence intervals, when needed.

3.3. *LM(3): A TS (Time Specific) Arbitrary Marginal Fixed Effects (AMFE) Model for Longitudinal Binary Data* This model is similar to the AMFE model (2.14) under the cross-sectional cluster setup. More specifically, the AMFE model under the longitudinal setup is written as in Eq. 3.8, i.e.,

$$E[Y_{it}] = Pr[Y_{it} = 1] = \tilde{p}_{it}(\boldsymbol{\beta}) = \exp(\mathbf{x}'_{it}\boldsymbol{\beta}) / [1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta})],$$

without writing any correlation structures for its derivation, as the underlying structure is assumed to be unknown. Some possible correlation models those might yield the above marginal mean model (as in Eq. 3.8) are: (a) the AR(1) type model given in Eq. 1.2. (b) MA(1) (moving average order 1), and (c) EQC (equi-correlations)/Exchange, models. We may refer to Sutradhar (2011, Sections 7.4.1 to 7.4.3) for a detailed discussion about these

three basic longitudinal models, all yielding the same marginal fixed effects based mean model.

Notice that to obtain a consistent estimate of β involved in the above marginal mean $\tilde{p}_{it}(\beta)$, one may solve the MM or QL estimating equations shown in Eqs. 3.1 and 3.2, as they are unbiased estimating equations. These MM and QL estimating equations are free of correlations and hence the regression estimates obtained from them are bound to be less efficient than any moments based equations involving correlations. However, for the cases where true correlation models are unknown, Liang and Zeger (1986) proposed a ‘working’ correlation matrix based GEE (generalized estimating equations) approach for efficient estimation of β . More specifically, for efficient β estimation, they define a ‘working’ correlation matrix as $R_i(\alpha)$, α being a set of working correlation index parameters, and solve the GEE given by

$$\sum_{i=1}^I \frac{\partial \tilde{p}'_i(\beta)}{\partial \beta} \mathbf{V}_i^{-1}(\beta, \alpha)(\mathbf{y}_i - \tilde{\mathbf{p}}_i(\beta)) = 0, \tag{3.13}$$

where $\tilde{\mathbf{p}}_i(\beta) = (\tilde{p}_{i1}(\beta), \dots, \tilde{p}_{it}(\beta), \dots, \tilde{p}_{iT}(\beta))'$, and $\mathbf{V}_i(\beta, \alpha) = \tilde{\mathbf{A}}_i^{\frac{1}{2}}(\beta) \mathbf{R}_i(\beta, \alpha) \tilde{\mathbf{A}}_i^{\frac{1}{2}}(\beta)$, with $\tilde{\mathbf{A}}_i(\beta) = \text{diag}[\tilde{p}_{i1}, \dots, \tilde{p}_{it}, \dots, \tilde{p}_{iT}]$. Because this GEE approach was ambitiously aimed to deal with any types of correlated binary data, it was used by hundreds and hundreds researchers over two decades or so until it was discovered that this approach may in fact yield less efficient estimates than an independence assumption-based estimating equation approach (Sutradhar and Das (1999), Sutradhar (2011, Section 7.3.6; see also Sutradhar and Zheng (2018) under a semi-parametric setup)) such as QL approach in Eq. 3.2 (also may be referred to as independence based GEE (GEE(I)).

Further note that as pointed out in the last section, one can not at all use the marginal fixed effects (MFE) based GEE approach for certain longitudinal binary data where correlation parameters enter to the formulas for the binary marginal means. More specifically, if GEE is used in such cases, it will produce inconsistent regression estimates. For example, suppose that the longitudinal responses follow the BDL (binary dynamic logit) model (1.4) yielding the marginal mean models as in Eq. 1.5. Under this BDL model, the response vector $\mathbf{y}_i = (y_{i1}, \dots, y_{it}, \dots, y_{iT})'$ has the mean $\boldsymbol{\mu}_i(\beta, \rho) = E[\mathbf{Y}_i] = (\mu_{i1}(\beta), \mu_{i2}(\beta, \rho), \dots, \mu_{iT}(\beta, \rho))'$ (see Eq. 1.5) and the covariance matrix $\boldsymbol{\Sigma}_i(\beta, \rho)$ as in Eq. 3.11, i.e.,

$$\mathbf{Y}_i \sim (\boldsymbol{\mu}_i(\beta, \rho), \boldsymbol{\Sigma}_i(\beta, \rho)), \tag{3.14}$$

where marginal means are function of both β and ρ , whereas the GEE approach will specify the mean vector as $\tilde{\mathbf{p}}_i(\beta) = (\tilde{p}_{i1}(\beta), \dots, \tilde{p}_{it}(\beta), \dots, \tilde{p}_{iT}(\beta))'$, with $\tilde{p}_{it}(\beta) = \exp(\mathbf{x}'_{it}\beta)/[1 + \exp(\mathbf{x}'_{it}\beta)]$.

Now to examine the convergence of $\hat{\beta}_{GEE}$ obtained from Eq. 3.13 when it is known that \mathbf{y}_i has the true mean vector and covariance matrix as in Eq. 3.14, we first write the iterative equation to obtain $\hat{\beta}_{GEE}$, as follows:

$$\begin{aligned} \hat{\beta}_{GEE}(r+1) &= \hat{\beta}_{GEE}(r) \\ &+ \left[\left\{ \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i(\beta)}{\partial \beta'} \mathbf{V}_i^{-1}(\beta, \hat{\alpha}) \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \right\} \right. \\ &\quad \left. \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \mathbf{V}_i^{-1}(\beta, \hat{\alpha})(\mathbf{y}_i - \tilde{\mathbf{p}}_i(\beta)) \right]_{\beta = \hat{\beta}_{GEE}(r)}. \end{aligned} \tag{3.15}$$

Notice that because the mean vector and covariance matrix of \mathbf{y}_i are the function of both β and ρ , and because $\hat{\alpha}$ is usually a moment estimator, it then follows that $\hat{\alpha}$ will converge to a quantity, say α_0 , which must be a function of ρ . That is,

$$\hat{\alpha} \rightarrow \alpha_0(\rho) \tag{3.16}$$

[Crowder (1995), Sutradhar and Das (1999)]. Thus, one may approximate the limiting (as $I \rightarrow \infty$) difference between $\hat{\beta}_{GEE}$ and true parameter β , as

$$\begin{aligned} &\lim_{I \rightarrow \infty} [\hat{\beta}_{GEE} - \beta] \\ &\approx \lim_{I \rightarrow \infty} \left[\left\{ \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i(\beta)}{\partial \beta'} \mathbf{V}_i^{-1}(\beta, \alpha_0(\rho)) \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \right\} \right. \\ &\quad \left. \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \mathbf{V}_i^{-1}(\beta, \alpha_0(\rho))(\mathbf{y}_i - \tilde{\mathbf{p}}_i(\beta)) \right] \\ &\rightarrow E_{\mathbf{y}} \left[\left\{ \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i(\beta)}{\partial \beta'} \mathbf{V}_i^{-1}(\beta, \alpha_0(\rho)) \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \right\} \right. \\ &\quad \left. \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \mathbf{V}_i^{-1}(\beta, \alpha_0(\rho))(\mathbf{y}_i - \tilde{\mathbf{p}}_i(\beta)) \right] \\ &= \left[\left\{ \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i(\beta)}{\partial \beta'} \mathbf{V}_i^{-1}(\beta, \alpha_0(\rho)) \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \right\} \right. \\ &\quad \left. \sum_{i=1}^I \frac{\partial \tilde{\mathbf{p}}_i'(\beta)}{\partial \beta} \mathbf{V}_i^{-1}(\beta, \alpha_0(\rho))(\boldsymbol{\mu}_i(\beta, \rho) - \tilde{\mathbf{p}}_i(\beta)) \right] \neq 0, \end{aligned} \tag{3.17}$$

because of the fact that under the BDL model (1.4), $E[\mathbf{Y}_i] = \boldsymbol{\mu}_i(\boldsymbol{\beta}, \rho)$ as in Eq. 3.9 (see also Eq. 3.14), which is quite different than $\tilde{\boldsymbol{\mu}}_i(\boldsymbol{\beta})$. Thus $\hat{\boldsymbol{\beta}}_{GEE}$ obtained from Eq. 3.13 is asymptotically biased and can not converge to $\boldsymbol{\beta}$ unless $\rho = 0$, which is unlikely to happen in the longitudinal setup.

4 Further Estimation and Asymptotic Properties in Cross-sectional Cluster Setup

4.1. GQL and MM Estimation Recall from Section 2 that except the MME based general cluster model A (CM-A), the remaining MFE based cluster models were developed either under restrictive assumptions about the distribution of the random effects such as “bridge” distribution leading to the fixed effects model CM-B-1, and beta-binary distribution leading to the fixed effects model CM-B-2, or using ‘working’ specification both for means and correlations leading to the AMFE model (CM-C). As discussed in details in the same section, these later three MFE based models have limited practical use, in particular the AMFE model (CM-C) can not be trusted at all as it does not justify how a fixed effects based marginal mean model can be derived from the conditional random effects model (1.1). For these reasons, we concentrate in this section only on the estimation of the parameters of the MME based CM-A model.

More specifically we turn back to the CM-A model described in Section 2.1. The model parameters $\boldsymbol{\beta}$ and σ_γ^2 are involved in the marginal mean function $\mu_{ij}(\boldsymbol{\beta}, \sigma_\gamma^2)$ in Eq. 2.2 which has its BA (binomial approximation) based computational version given by Eq. 2.6. For the estimation of these parameters, as discussed in Section 2.1.1 that there exist several likelihood based approaches (exact likelihood, PQL (penalized quasi-likelihood), HL (hierarchical likelihood)), but they are either computationally involved or they produce biased and hence mean squared error inconsistent estimates specially for large values of σ_γ^2 . In this section, following the GQL approach of Sutradhar (2004) developed under the GLMM (generalized linear mixed model) setup, we simplify the binomial approximation (to standard normal random effects) based GQL estimating equation for $\boldsymbol{\beta}$, and MM estimating equation for σ_γ^2 . Furthermore, as in practice one (specially the statistical agencies) deals with large number of clusters each containing large number of individuals so that $\sum_{i=1}^I n_i \rightarrow \infty$, in the next section we make sure for the benefit to these practitioners that the GQL estimator of $\boldsymbol{\beta}$ and the MM estimator of σ_γ^2 , are consistent. In Section 4.2.2, we show that the GQL estimator of the main regression parameters $\boldsymbol{\beta}$ has asymptotically normal

distribution providing an opportunity for confidence interval construction, when needed.

4.1.1. *GQL Estimation of β .* Once the mean function is specified, one requires the true covariance/correlation structure to construct the desired GQL estimating equation (Sutradhar (2003, Section 3.1)) for the parameter of interest. Under the conditional cluster model (1.1) with normal random cluster effects (γ_i), the mean function of the i -th cluster response vector $\mathbf{y}_i = (y_{i1}, \dots, y_{ij}, \dots, y_{in_i})'$ is computed as in Eq. 3.14. More specifically, by using the BA (binomial approximation) to the standard normal cluster effect γ_i^* as in Eq. 2.5, we write the BA based mean function as

$$\begin{aligned} E[\mathbf{Y}_i] &= \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) \\ &= (\mu_{i1}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2), \dots, \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2), \dots, \mu_{in_i}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2))', \end{aligned} \tag{4.1}$$

where by Eq. 2.6,

$$\mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = \sum_{v_i=0}^V p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i},$$

for all $j = 1, \dots, n_i$, V being a large number such as $V = 10$, or 15. It immediately follows that

$$\text{var}(Y_{ij}) = \sigma_{i,jj}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)(1 - \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)). \tag{4.2}$$

We now turn to the computation of the $n_i \times n_i$ covariance matrix of \mathbf{y}_i . For two responses y_{ij} and y_{ik} , $j \neq k$; $j, k = 1, \dots, n_i$, by Eq. 1.1, we first write

$$\begin{aligned} \lambda_{i,jk}(\boldsymbol{\beta}, \sigma_\gamma^2) &= E[Y_{ij}Y_{ik}] = E_{\gamma_i} E[Y_{ij}Y_{ik}|\gamma_i] \\ &= E_{\gamma_i} [E(Y_{ij}|\gamma_i)E(Y_{ik}|\gamma_i)] = \int p_{ij}^*(\boldsymbol{\beta}, \gamma_i)p_{ik}^*(\boldsymbol{\beta}, \gamma_i)g_N(\gamma_i)d\gamma_i, \end{aligned} \tag{4.3}$$

where $p_{ij}^*(\boldsymbol{\beta}, \gamma_i) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)/[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)]$, and $g_N(\gamma_i) \equiv [\gamma_i \sim N(0, \sigma_\gamma^2)]$. Notice that the normal integration (4.3) of a complex function in γ_i , can be computed as in Eq. 2.6 using the BA. More specifically, for $\gamma_i^* = \gamma_i/\sigma_\gamma \equiv h(v_i)$ as in Eq. 2.5, the integration in Eq. 4.3 is approximated as

$$\begin{aligned} &\lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) \\ &= \sum_{v_i=0}^V p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i))p_{ik}^*(\mathbf{x}_{ik}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}, \end{aligned} \tag{4.4}$$

yielding the covariance between y_{ij} and y_{ik} as

$$\begin{aligned} \text{cov}[Y_{ij}, Y_{ik}] &= E[Y_{ij}Y_{ik}] - E[Y_{ij}]E[Y_{ik}] \\ &= \lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) - \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)\mu_{ik}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = \sigma_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2), \end{aligned} \quad (4.5)$$

where the formula for $\mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$, for example, is given by Eq. 4.1 (see also Eq. 2.6). Subsequently, combining Eqs. 4.2 and 4.5 we obtain the $n_i \times n_i$ covariance matrix of \mathbf{y}_i , as

$$\text{cov}[\mathbf{Y}_i] = \boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = (\sigma_{i,jk}^{BA}) : n_i \times n_i. \quad (4.6)$$

Because $E[\mathbf{Y}_i] = \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$ and $\text{cov}[\mathbf{Y}_i] = \boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$ can be computed by Eqs. 4.1 and 4.6, respectively, following Sutradhar (2003, Section 3.1), for given σ_γ^2 , one may then construct the desired GQL estimating equation for $\boldsymbol{\beta}$ as

$$\sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)) = 0. \quad (4.7)$$

Let the solution of this GQL estimating Eq. 4.7 be denoted by $\hat{\boldsymbol{\beta}}_{GQL}$. For practical benefit, the asymptotic properties of this estimator must be studied. The consistency of this estimator is examined in Section 4.2.1, along with its asymptotic normality property in Section 4.2.2. To solve the estimating Eq. 4.7, it remains to compute the matrix derivative involved in the equation, which we derive as follows.

Computation of the Derivative $\frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]'}{\partial \boldsymbol{\beta}}$ For this matrix computation it is sufficient to compute the derivative vector, $\frac{\partial \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \boldsymbol{\beta}}$, which, following Eq. 4.1, can be derived as

$$\begin{aligned} \frac{\partial \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \boldsymbol{\beta}} &= \sum_{v_i=0}^V \frac{\partial p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i))}{\partial \boldsymbol{\beta}} \begin{pmatrix} V \\ v_i \end{pmatrix} (1/2)^{v_i} (1/2)^{V-v_i} \quad (4.8) \\ &= \sum_{v_i=0}^V \mathbf{x}_{ij} p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) q_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \begin{pmatrix} V \\ v_i \end{pmatrix} \\ &\quad (1/2)^{v_i} (1/2)^{V-v_i} : p \times 1, \end{aligned}$$

where $q_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) = 1 - p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) = [1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \sigma_\gamma h(v_i))]^{-1}$.

4.1.2. *MM Estimation of σ_γ^2 .* Notice that the GQL estimating Eq. 4.24 for β was developed for known σ_γ^2 , which is, however, unknown in practice. Similar to Sutradhar (2004) (see also Jiang, 1998), in this section we estimate this parameter by exploiting second order binary responses, whereas β was estimated using the first order responses. However, because σ_γ^2 is a parameter of secondary interest, as opposed to the GQL approach for σ_γ^2 estimation by Sutradhar (2004), for simplicity, we use the well known method of moments (MM). It is shown in Section 4.2.3, this simpler MM estimation produces consistent σ_γ^2 estimate, similar to the consistency property of the GQL regression estimator $\hat{\beta}_{GQL}$. As pointed out above, this MM estimator of σ_γ^2 is expected to be less efficient than its GQL estimator, this efficiency is not being a concerning issue as σ_γ^2 is a parameter of secondary interest.

Let the second order response vectors under the present clustered binary setup, be denoted by

$$\begin{aligned} \mathbf{g}_i &= (y_{i1}^2, \dots, y_{ij}^2, \dots, y_{in_i}^2)' \equiv (y_{i1}, \dots, y_{ij}, \dots, y_{in_i})' = \mathbf{y}_i : n_i \times 1 \\ \mathbf{q}_i &= (y_{i1}y_{i2}, \dots, y_{ij}y_{ik}, \dots, y_{i(n_i-1)}y_{in_i})' : j < k : \frac{n_i(n_i - 1)}{2} \times 1 \\ &\equiv (q_{i,12}, \dots, q_{i,jk}, \dots, q_{i,(n_i-1)n_i})', \end{aligned} \tag{4.9}$$

containing all possible squared and pair-wise responses. Because, $\mathbf{g}_i = \mathbf{y}_i$, clearly $E[\mathbf{G}_i] = \boldsymbol{\mu}_i^{BA}(\beta, \sigma_\gamma^2)$ as in Eq. 4.1. Next, by Eqs. 4.3 and 4.4, we write

$$\begin{aligned} E[\mathbf{Q}_i] &= (\lambda_{i,12}^{BA}(\beta, \sigma_\gamma^2), \dots, \lambda_{i,jk}^{BA}(\beta, \sigma_\gamma^2), \dots, \lambda_{i,(n_i-1)n_i}^{BA}(\beta, \sigma_\gamma^2))' \\ &= \boldsymbol{\lambda}_i^{BA}(\beta, \sigma_\gamma^2), \text{ (say)}, \end{aligned} \tag{4.10}$$

where the BA based formula for $E[Y_{ij}Y_{ik}] = \lambda_{i,jk}(\beta, \sigma_\gamma^2)$ is given in Eq. 4.4. Because both $\boldsymbol{\mu}_i^{BA}(\cdot)$ and $\boldsymbol{\lambda}_i^{BA}(\cdot)$ contain σ_γ^2 on top of β , we may construct a MM estimating equation for σ_γ^2 , as

$$\begin{aligned} &\sum_{i=1}^I \left[\frac{\partial[\boldsymbol{\mu}_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial \sigma_\gamma^2} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\beta, \sigma_\gamma^2)) \right. \\ &\left. + \frac{\partial[\boldsymbol{\lambda}_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial \sigma_\gamma^2} (\mathbf{q}_i - \boldsymbol{\lambda}_i^{BA}(\beta, \sigma_\gamma^2)) \right] = 0, \end{aligned} \tag{4.11}$$

where

$$\frac{\partial \boldsymbol{\mu}_i^{BA}(\beta, \sigma_\gamma^2)}{\partial \sigma_\gamma^2} = \left(\frac{\partial \mu_{i1}^{BA}(\beta, \sigma_\gamma^2)}{\partial \sigma_\gamma^2}, \dots, \frac{\partial \mu_{ij}^{BA}(\beta, \sigma_\gamma^2)}{\partial \sigma_\gamma^2}, \dots, \frac{\partial \mu_{in_i}^{BA}(\beta, \sigma_\gamma^2)}{\partial \sigma_\gamma^2} \right)', \tag{4.12}$$

with

$$\frac{\partial \mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \sigma_\gamma^2} \tag{4.13}$$

$$= \sum_{v_i=0}^V \frac{1}{2\sigma_\gamma} h(v_i) p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) q_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i},$$

and

$$\begin{aligned} \frac{\partial \lambda_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \sigma_\gamma^2} &= \left(\frac{\partial \lambda_{i,12}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \sigma_\gamma^2}, \dots, \frac{\partial \lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \sigma_\gamma^2}, \dots, \right. \\ &\quad \left. \frac{\partial \lambda_{i,(n_1-1)n_i}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \sigma_\gamma^2} \right)', \end{aligned} \tag{4.14}$$

where, it follows from Eq. 4.4 that

$$\begin{aligned} \frac{\partial \lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)}{\partial \sigma_\gamma^2} &= \sum_{v_i=0}^V \frac{1}{2\sigma_\gamma} h(v_i) p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) p_{ik}^*(\mathbf{x}_{ik}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \\ &\times [q_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) + q_{ik}^*(\mathbf{x}_{ik}; \boldsymbol{\beta}, \sigma_\gamma h(v_i))] \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}, \end{aligned} \tag{4.15}$$

for all $j < k; j, k = 1, \dots, n_i$.

Let $\hat{\sigma}_{\gamma,MM}^2$ denote the solution of the moment (4.11). In Section 4.2.3 we show that this MM estimator for σ_γ^2 is a consistent estimator under some mild regularity conditions.

4.2. Consistency and Asymptotic Normality

4.2.1. Consistency of $\hat{\boldsymbol{\beta}}_{GQL}$ Obtained from Eq. 4.7. We first apply a first order Taylor series expansion to the GQL estimating function in the left hand side of the estimating in Eq. 4.7 and obtain

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GQL} - \boldsymbol{\beta} &\simeq - \left[\sum_{i=1}^I \frac{\partial [\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial [\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]}{\partial \boldsymbol{\beta}'} \right]^{-1} \\ &\times \left[\sum_{i=1}^I \frac{\partial [\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)) \right] \\ &+ o_p(1/\sqrt{N}), \end{aligned} \tag{4.16}$$

where $N = \sum_{i=1}^I n_i$. Let G_N be a N -dependent finite and bounded quantity, and it increases as N gets larger. Notice that the $p \times p$ matrix in the first

term in the right hand side of Eq. 4.16 is free from responses $\{y\}$. Suppose that this $p \times p$ matrix satisfies the regularity condition

$$\frac{1}{\sum_{i=1}^I n_i} \sum_{i=1}^I \left| \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]}{\partial \boldsymbol{\beta}'} \right| \leq G_N, \quad (4.17)$$

implying that

$$\sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]}{\partial \boldsymbol{\beta}'} \equiv O(NG_N). \quad (4.18)$$

Next the second term in the right hand side of Eq. 4.16 converges to zero as

$$\begin{aligned} & \sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)) \\ \rightarrow E \left[\sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)) \right] &= 0, \quad (4.19) \end{aligned}$$

in the order of

$$\begin{aligned} & \left[\text{cov} \left\{ \sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)) \right\} \right]^{\frac{1}{2}} \quad (4.20) \\ &= \left| \left[\sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]}{\partial \boldsymbol{\beta}'} \right]^{\frac{1}{2}} \right| \\ &\simeq O_p(\sqrt{NG_N}), \text{ by (4.35).} \end{aligned}$$

Hence applying Eqs. 4.18 and 4.20 into 4.16, we obtain

$$\begin{aligned} [\hat{\boldsymbol{\beta}}_{GQL} - \boldsymbol{\beta}] &= O(N^{-1}G_N^{-1})O_p(\sqrt{NG_N}) + o_p(1/\sqrt{N}) \\ &= O_p((1/\sqrt{N})G_N^{-\frac{1}{2}}) + o_p(1/\sqrt{N}) \equiv o_p(1/\sqrt{N}), \quad (4.21) \end{aligned}$$

because G_N is a finite and bounded quantity. It then follows that

$$\lim_{N \rightarrow \infty} [\hat{\boldsymbol{\beta}}_{GQL} - \boldsymbol{\beta}] \rightarrow 0. \quad (4.22)$$

Thus, $\hat{\boldsymbol{\beta}}_{GQL}$ obtained from Eq. 4.7 is consistent for $\boldsymbol{\beta}$.

Note that as $\hat{\beta}_{GQL}$ is asymptotically unbiased for β , it follows from Eq. 4.16 that its asymptotic covariance matrix is given by

$$\begin{aligned} & \lim_{I \rightarrow \infty} \text{cov}(\hat{\beta}_{GQL}) \\ &= \left[\sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial \boldsymbol{\beta}'} \right]^{-1}, \end{aligned} \tag{4.23}$$

which can be estimated by replacing β and σ_γ^2 , with $\hat{\beta}_{GQL}$ and $\hat{\sigma}_{\gamma,MM}^2$, respectively, provided $\hat{\sigma}_{\gamma,MM}^2$ is a consistent estimator of σ_γ^2 . This later consistency property is examined in Section 4.2.3. Further note that the aforementioned estimate for $\text{cov}(\hat{\beta}_{GQL})$ becomes more useful when confidence interval construction for the β parameter is needed. However, for such a confidence interval construction one needs to examine the asymptotic distribution of $\hat{\beta}_{GQL}$, which we do in the following section.

4.2.2. *Asymptotic Normality of $\hat{\beta}_{GQL}$.* We outline the derivation of the asymptotic distribution as follows. Notice from Eq. 4.16 that for large I , β estimator satisfy the approximation

$$\begin{aligned} & \hat{\beta}_{GQL} - \beta \simeq - \left[\frac{1}{I} \sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial \boldsymbol{\beta}'} \right]^{-1} \\ & \times \left[\frac{1}{I} \sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]'}{\partial \boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)) \right], \end{aligned} \tag{4.24}$$

which we re-express as

$$\hat{\beta}_{GQL} - \beta \simeq - \left[\sum_{i=1}^I \frac{\partial \mathbf{f}_i(\boldsymbol{\beta} | \sigma_\gamma^2, \mathbf{y}_i)}{\partial \boldsymbol{\beta}'} \right]^{-1} \left[\sum_{i=1}^I \mathbf{f}_i(\boldsymbol{\beta} | \sigma_\gamma^2, \mathbf{y}_i) \right]. \tag{4.25}$$

Let

$$\bar{\mathbf{f}}_I(\boldsymbol{\beta} | \sigma_\gamma^2) = \frac{1}{I} \sum_{i=1}^I \mathbf{f}_i(\boldsymbol{\beta} | \sigma_\gamma^2, \mathbf{y}_i), \tag{4.26}$$

where \mathbf{f}_i 's are clearly independent because $\mathbf{y}_1, \dots, \mathbf{y}_i, \dots, \mathbf{y}_I$ are independent vectors from I independent clusters. But, they are not identically distributed because of the fact that

$$\{\mathbf{Y}_i : n_i \times 1\} \sim (\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2), \boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)), \tag{4.27}$$

by Eqs. 4.1 and 4.6, i.e., the means, variances and covariances are cluster dependent, i.e., they vary from cluster to cluster. Notice from Eqs. 4.24–4.26 that $\bar{\mathbf{f}}_I(\boldsymbol{\beta}|\sigma_\gamma^2, \mathbf{y}_i)$ in Eq. 4.26 has the mean vector and covariance matrix as given by

$$\begin{aligned} E[\bar{\mathbf{f}}_I(\boldsymbol{\beta})] &= 0, \text{ and} \\ \text{cov}[\bar{\mathbf{f}}_I(\boldsymbol{\beta})] &= \frac{1}{I^2} \sum_{i=1}^I \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]'}{\partial\boldsymbol{\beta}} [\boldsymbol{\Sigma}_i^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} \frac{\partial[\boldsymbol{\mu}_i^{BA}(\boldsymbol{\beta}, \sigma^2)]}{\partial\boldsymbol{\beta}'} \\ &= \frac{1}{I^2} \mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2), \text{ (say)}. \end{aligned} \tag{4.28}$$

Next we assume that the multivariate version of Lindeberg’s condition holds, that is,

$$\lim_{I \rightarrow \infty} \mathbf{V}_I^{*-1} \sum_{i=1}^I \sum_{\{\mathbf{f}_i' \mathbf{V}_I^{*-1} \mathbf{f}_i\} > \epsilon} \mathbf{f}_i \mathbf{f}_i' p^\dagger(\mathbf{f}_i) = 0 \tag{4.29}$$

holds, for all $\epsilon > 0$, $p^\dagger(\cdot)$ being the probability distribution of \mathbf{f}_i . Then the Lindeberg-Feller central limit theorem [Amemiya (1985). Theorem 3.3.6], McDonald (2005, Theorem 2.2)] implies the following convergence in distribution (\rightarrow_d):

$$\mathbf{Z}_I = I[\mathbf{V}_I^*]^{-\frac{1}{2}} \bar{\mathbf{f}}_I(\boldsymbol{\beta}) \rightarrow_d N_p(0, I_p). \tag{4.30}$$

I_p being the $p \times p$ identity matrix.

By using the notations from Eq. 4.26 it follows from Eqs. 4.25 and 4.30 that

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{GQL} - \boldsymbol{\beta} &\simeq - \left[\sum_{i=1}^I \frac{\partial \mathbf{f}_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \right]^{-1} \left[\sum_{i=1}^I \mathbf{f}_i \right] \\ &= [\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1} [\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{\frac{1}{2}} [\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-\frac{1}{2}} I \bar{\mathbf{f}}_I(\boldsymbol{\beta}) \\ &= [\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-\frac{1}{2}} \mathbf{Z}_I. \end{aligned} \tag{4.31}$$

Clearly, by Eq. 4.30, the quantity in Eq. 4.31 converges in distribution, as

$$[\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-\frac{1}{2}} \mathbf{Z}_I \rightarrow_d N_p(0, [\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1}). \tag{4.32}$$

Notice that this normal covariance matrix $[\mathbf{V}_I^*(\boldsymbol{\beta}, \sigma_\gamma^2)]^{-1}$ is the same as the limiting covariance matrix in Eq. 4.23, as expected.

4.2.3. *Consistency of $\hat{\sigma}_{\gamma,MM}^2$ Obtained from Eq. 4.11.* Consistency property of $\hat{\sigma}_{\gamma,MM}^2$ can be established in a similar way as that of $\hat{\beta}_{GQL}$ discussed in Section 4.2.1. For convenience we, however, highlight the main steps below. Because $\hat{\sigma}_{\gamma,MM}^2$ is the solution of the MM estimating Eq. 4.11, a first order Taylor series expansion of the estimating function in the left hand side of Eq. 4.11 about σ_γ^2 provides

$$\begin{aligned}
 & [\hat{\sigma}_{\gamma,MM}^2 - \sigma_\gamma^2] \\
 \simeq & - \left[\sum_{i=1}^I \left\{ \frac{\partial[\mu_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial\sigma_\gamma^2} \frac{\partial[\mu_i^{BA}(\beta, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} + \frac{\partial[\lambda_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial\sigma_\gamma^2} \frac{\partial[\lambda_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial\sigma_\gamma^2} \right\} \right]^{-1} \\
 \times & \left[\sum_{i=1}^I \left\{ \frac{\partial[\mu_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial\sigma_\gamma^2} (\mathbf{y}_i - \mu_i^{BA}(\beta, \sigma_\gamma^2)) + \frac{\partial[\lambda_i^{BA}(\beta, \sigma_\gamma^2)]'}{\partial\sigma_\gamma^2} (\mathbf{q}_i - \lambda_i^{BA}(\beta, \sigma_\gamma^2)) \right\} \right] \\
 + & o_p(1/\sqrt{N}), \tag{4.33}
 \end{aligned}$$

where $N = \sum_{i=1}^I n_i$. For convenience of further calculations, we re-express the equation in (4.33) as

$$\hat{\sigma}_{\gamma,MM}^2 - \sigma_\gamma^2 \simeq -S_1^{-1} S_{2,y} + o_p(1/\sqrt{N}). \tag{4.34}$$

Suppose that H_N is a N -dependent increasing but finite and bounded quantity, and S_1 in Eq. 4.34 satisfies the following regularity condition:

$$\frac{1}{\sum_{i=1}^I n_i} S_1 \leq H_N, \tag{4.35}$$

implying that

$$S_1 \approx O(NH_N). \tag{4.36}$$

Notice that $E_Y[S_{2,y}] = 0$. It then follows that $S_{2,y} \rightarrow_p E_Y[S_{2,y}] = 0$, but in order of $[\text{var}(S_{2,y})]^{1/2}$. To compute this variance formula, it is convenient to re-express $S_{2,y}$, by using Eq. 4.33, as

$$\begin{aligned}
 S_{2,y} = & \sum_{i=1}^I \left\{ \frac{\sum_{j=1}^{n_i} \partial[\mu_{ij}^{BA}(\beta, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} (y_{ij} - \mu_{ij}^{BA}(\beta, \sigma_\gamma^2)) \right. \\
 + & \left. \sum_{j < k}^{n_i} \frac{\partial[\lambda_{i,jk}^{BA}(\beta, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} (y_{ij}y_{ik} - \lambda_{i,jk}^{BA}(\beta, \sigma_\gamma^2)) \right\}, \tag{4.37}
 \end{aligned}$$

and obtain its variance as

$$\text{var}(S_{2,y})$$

$$\begin{aligned}
&= \sum_{i=1}^I \left[\left\{ \sum_{j=1}^{n_i} \left(\frac{\partial[\mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \right)^2 \text{var}(Y_{ij}) \right. \right. \\
&\quad \left. \left. + 2 \sum_{j < k}^{n_i} \left(\frac{\partial[\mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \frac{\partial[\mu_{ik}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \right) \right. \right. \\
&\quad \times \text{cov}(Y_{ij}, Y_{ik}) \left. \left. + \left\{ \sum_{j < k}^{n_i} \left(\frac{\partial[\lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \right)^2 \text{var}(Y_{ij}Y_{ik}) \right. \right. \right. \\
&\quad \left. \left. + \sum_{j < k}^{n_i} \sum_{\ell < m}^{n_i} \left(\frac{\partial[\lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \frac{\partial[\lambda_{i,\ell m}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \right) \text{cov}(Y_{ij}Y_{ik}, Y_{i\ell}Y_{im}) \right\} \right. \\
&\quad \left. \left. + \left\{ \sum_{j=1}^{n_i} \sum_{k < \ell}^{n_i} \left(\frac{\partial[\mu_{ij}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \frac{\partial[\lambda_{i,k\ell}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)]}{\partial\sigma_\gamma^2} \right) \text{cov}(Y_{ij}, Y_{ik}Y_{i\ell}) \right\} \right] \\
&= \sum_{i=1}^I \Omega_i(\boldsymbol{\beta}, \sigma_\gamma^2), \text{ (say)}, \tag{4.38}
\end{aligned}$$

where $\text{var}(Y_{ij}) = \sigma_{i,jj}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$, and $\text{cov}(Y_{ij}, Y_{ik}) = \sigma_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$, are given by Eqs. 4.2 and. 4.5, respectively. The computational formulas for the remaining third and fourth order moments, i.e., for $\text{var}(Y_{ij}Y_{ik}) = \omega_{i,jjkk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$; $\text{cov}(Y_{ij}Y_{ik}, Y_{i\ell}Y_{im}) = \omega_{i,jk\ell m}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$; $\text{cov}(Y_{ij}, Y_{ik}Y_{i\ell}) = \phi_{i,jk\ell}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)$, are relatively lengthy and given in Appendix A, for convenience.

Suppose that for a N -dependent finite and bounded quantity K_N , $\text{var}(S_{2,y})$ satisfies the regularity condition

$$\frac{1}{\sum_{i=1}^I n_i} \sum_{i=1}^I \Omega_i(\boldsymbol{\beta}, \sigma_\gamma^2) \leq K_N, \tag{4.39}$$

implying that

$$[\text{var}(S_{2,y})] = O_p(\sqrt{NK_N}). \tag{4.40}$$

Now by applying Eqs. 4.35 and 4.40 to 4.34, one obtains

$$\begin{aligned}
\hat{\sigma}_{\gamma,MM}^2 - \sigma_\gamma^2 &\simeq O(N^{-1}H_N^{-1})O_p(\sqrt{NK_N}) + o_p(1/\sqrt{N}) \\
&= O_p(N^{-\frac{1}{2}}\frac{\sqrt{K_N}}{H_N}) + o_p(1/\sqrt{N}) = o_p(1/\sqrt{N}), \tag{4.41}
\end{aligned}$$

because both H_N and K_N are finite and bounded. Hence,

$$\lim_{N \rightarrow \infty} [\hat{\sigma}_{\gamma,MM}^2 - \sigma_\gamma^2] \rightarrow_p 0, \tag{4.42}$$

justifying that $\hat{\sigma}_{\gamma,MM}^2$ is consistent for σ_γ^2 , and this can be used in the GQL (4.7) while estimating β .

We remark that because by Eq. 4.42, $\hat{\sigma}_{\gamma,MM}^2$ is asymptotically unbiased for σ_γ^2 , one may then compute the asymptotic variance of $\hat{\sigma}_{\gamma,MM}^2$ by exploiting Eqs. 4.33 and 4.34. More specifically,

$$\begin{aligned} \text{var}[\hat{\sigma}_{\gamma,MM}^2] &= S^{-1}(\beta, \sigma_\gamma^2) \text{var}[S_{2,y}] S^{-1}(\beta, \sigma_\gamma^2) \\ &= S^{-1}(\beta, \sigma_\gamma^2) \sum_{i=1}^I \Omega_i(\beta, \sigma_\gamma^2) S^{-1}(\beta, \sigma_\gamma^2), \end{aligned} \tag{4.43}$$

by Eq. 4.38, which can be estimated by replacing β with $\hat{\beta}_{GQL}$, and σ_γ^2 with $\hat{\sigma}_{\gamma,MM}^2$.

5. On the Bayesian Approach for Correlated Binary Data

We continue discussing mixed effects models (1.1) for cross-sectional cluster binary data and a time dynamic fixed effects models (1.4) for longitudinal cluster data. However, as opposed to the parametric correlation structures based regression analysis, here we focus on some of the the existing alternative studies using the Bayesian approach where, without specifying the correlation structures, multilevel conditional models are used to estimate the main parameters such as the individual level covariates effects (β) and cluster specific parameters such as cluster variation σ_γ^2 in Eq. 1.1 under cross-sectional cluster model (e.g., McCulloch, 1997), or dynamic dependence parameter (ρ) in Eq. 1.4 under longitudinal cluster model (e.g Chib and Jeliazkov, 2006). We remark that because the mixed effects models are also used by some authors such as Stiratelli et al. (1984), and Zeger et al. (1988) for binary longitudinal data, we also include these models under the longitudinal setup on top of the dynamic fixed effects models.

5.1. Monte Carlo Based Likelihood Estimation for Cluster Binary Data

We keep focussing on the clustered binary model but instead of common random cluster effect γ_i , consider a more general situation using γ_{ij} as the cluster specific individual random effect for the j -th individual in the i -th cluster. Let z_{ij}^* denote a cluster-specific scalar covariate associated with random cluster effects γ_{ij} . Thus, $z_{ij}^* = 1$, and $\gamma_{ij} = \gamma_i$ for all $j = 1, \dots, n_i$, would refer to the basic cluster-specific mean model (1.1). Similar to Stiratelli et al.

(1984, Eqn. (2.2)), Zeger et al. (1988, Eqn. (2.1), and Daniels and Gatsonis (1999, Eqns. (1)-(2)), we may write the logit link for this general case, as

$$\text{logit}(p_{ij}^*(\boldsymbol{\beta}, \boldsymbol{\gamma}_{ij})) \equiv \ell(p_{ij}^*) = \mathbf{x}'_{ij}\boldsymbol{\beta} + z_{ij}^*\boldsymbol{\gamma}_{ij}, \quad (5.1)$$

for $j = 1, \dots, n_i$; $i = 1, \dots, I$. Notice that Daniels and Gatsonis (1999) have expressed the logit link as $\ell(p_{ij}^*) = \mathbf{x}'_{ij}\boldsymbol{\alpha}_i$. For our discussion it is convenient to use the notation in Eq. 5.1. Write $\boldsymbol{\gamma}_i = (\gamma_{i1}, \dots, \gamma_{ij}, \dots, \gamma_{n_i})'$. Similar to the normality assumption in Eq. 2.2, and also in the aforementioned studies, one may assume that

$$\boldsymbol{\gamma}_i \sim N(0, \mathbf{D}_i), \quad (5.2)$$

where \mathbf{D}_i is the $n_i \times n_i$ covariance matrix.

Recall from Eq. 2.1 that a closed-form likelihood function cannot be obtained due to the problem of integration over the distribution of the random effect $\boldsymbol{\gamma}_i$. To handle such an integration problem, some numerical algorithms are developed where $\boldsymbol{\gamma}_i$ is considered to be a missing data, and it is drawn from a conditional distribution of $\boldsymbol{\gamma}_i|y$ by using the so-called Metropolis algorithm (Gelfand and Carlin, 1993), which does not require specification of the unconditional density of the binary data y . More specifically, the Metropolis algorithm is used to simulate the random effects and the so-called expectation-maximization (EM) or Newton–Ralphson (NR) technique is used to maximize the Monte Carlo (simulated) (MC) based approximate likelihood function for the estimation of the regression effects $\boldsymbol{\beta}$. We may refer to McCulloch (1997), for example, for these MCEM and MCNR approaches. Some authors such as Daniels and Gatsonis (1999), instead of normality in Eq. 5.2, have assumed more general symmetric multivariate t distribution for $\boldsymbol{\gamma}_i$ given by

$$\boldsymbol{\gamma}_i \sim t_\nu(\mathbf{G}_i\boldsymbol{\gamma}, \mathbf{D}_i), \quad (5.3)$$

where \mathbf{G}_i is a cluster level covariates dependent known matrix of dimension $n_i \times q$, and $\boldsymbol{\gamma}$ is a q -dimensional vector of suitable parameters, and ν is the unknown degrees of freedom parameter. Next, using suitable proper prior distributions for $\boldsymbol{\gamma}$ and \mathbf{D}_i , Daniels and Gatsonis (1999, Section 2.2) used the Markov Chain Monte Carlo (MCMC) approach for the desired model fitting.

However, as expected the above monte carlo based likelihood inference approach is computationally expensive. Moreover, the selection of proper prior distributions is a challenge in this approach. For example, while under

normality assumption for γ_i (5.2), it is reasonable to consider \mathbf{D}_i^{-1} has the prior so-called Wishart distribution, but it may not be a proper prior distribution when γ_i follows the multivariate t distribution as in Eq. 5.3. This is because as Sutradhar and Ali (1989), for example, derived a Wishart distribution under the multivariate t -model which is different than the usual normality based Wishart distribution. More specifically, it is also dependent of the degrees of freedom of the t -distribution.

Turning back to the parametric inferences discussed in Section 4, when the normality assumption in Eq. 5.2 holds, using the so-called Binomial approximation (BA), one may easily construct the correlation structure as explained in Section 4.1.1 and apply the generalized quasi-likelihood estimation approach (4.7) to obtain consistent and highly efficient estimate for β , and method of moments (MM) estimation approach (4.11) to obtain consistent estimates for the variance parameters involved in \mathbf{D}_i . Alternatively, as also pointed out by Daniels and Gatsonis (1999, Section 1), one may use the other parametric approaches such as the PQL (penalized quasi-likelihood) approach of Breslow and Clayton (1993) or hierarchical likelihood (HQL) approach of Lee and Nelder (1996), which are simpler than the MCMC based Bayesian approach.

5.2. *Monte Carlo Based Likelihood Estimation for Longitudinal Binary Data* With regard to the analysis of longitudinal binary data in Bayesian setup, the most of the existing studies used the same random effects based logit link model (5.1) with some modifications as follows. Using the notations from Section 3, we re-write the logit link model for longitudinal data as

$$\text{logit}(\tilde{p}_{it}(\beta, \gamma_i)) \equiv \ell(\tilde{p}_{it}) = \mathbf{x}'_{it}\beta + \tilde{z}_{it}\gamma_i, \quad (5.4)$$

for the binary response y_{it} , recorded at time t ($t = 1, \dots, T$), for the i -th individual. Here both \mathbf{x}_{it} and \tilde{z}_{it} are time dependent covariates. Under this model, the binary responses y_{iu} at time u , and y_{it} at time t , become correlated under the assumption that i -th individual's random effect remains the same over time.

We remark that in a longitudinal setup, irrespective of the nature of the responses whether linear, count or binary, it is expected in practice that as time lag increases the correlations between two responses must decrease. Two dynamic models considered in the literature for binary responses, such as the AR(1) type dynamic model (1.2) and binary dynamic logit (BDL) model (1.4) satisfies this lag dependent decaying correlation property. Specifically, the AR(1) model (1.2) produced the correlations $\text{corr}(Y_{iu}, Y_{it}) = \rho^{t-u} \left[\frac{\sigma_{iuu}}{\sigma_{itt}} \right]^{\frac{1}{2}}$ as in Eq. 3.1 which (a) becomes smaller as $|t - u|$

increases, and also (b) it contains time varying covariates involved in σ_{itt} . Similarly, the BDL model (1.4) produced the correlations

$$\text{corr}(Y_{iu}, Y_{it}) = \sqrt{\frac{\mu_{iu}(\cdot)(1 - \mu_{iu}(\cdot))}{\mu_{it}(\cdot)(1 - \mu_{it}(\cdot))}} \prod_{v=u+1}^t (\tilde{p}_{iv}(\boldsymbol{\beta}, \rho) - \tilde{p}_{iv}(\boldsymbol{\beta}))$$

as in Eq. 3.10, which decay as $|t - u|$ increases. This is because

$$0 < (\tilde{p}_{iv}(\boldsymbol{\beta}, \rho) - \tilde{p}_{iv}(\boldsymbol{\beta})) < 1,$$

for all $v = (u + 1), \dots, t$. Also it contains time varying covariates.

Note that the random effects model (5.4), for example, does not satisfy decaying correlation property (a) mentioned above. It, however, satisfies (b) indicating that correlations contain time varying covariates. This is because under the assumption that $\gamma_i \stackrel{iid}{\sim} N(0, \tilde{\sigma}_\gamma^2)$, we can compute

$$\begin{aligned} E[Y_{it}] &= E_{\gamma_i} [E(Y_{it}|\gamma_i)] = \tilde{\mu}_{it}(\mathbf{x}_{it}, z_{it}^*, \boldsymbol{\beta}, \tilde{\sigma}_\gamma^2) \\ &= \int \left[\frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + z_{it}^*\gamma_i)}{[1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + z_{it}^*\gamma_i)]} \right] dG_N(\gamma_i, \tilde{\sigma}_\gamma^2), \end{aligned} \tag{5.5}$$

and

$$\begin{aligned} E[Y_{iu}Y_{it}] &= E_{\gamma_i} [E(Y_{iu}|\gamma_i)E(Y_{it}|\gamma_i)] = \tilde{\lambda}_{iut}(\mathbf{x}_{iu}, \mathbf{x}_{it}, z_{iu}^*, z_{it}^*, \boldsymbol{\beta}, \tilde{\sigma}_\gamma^2) \\ &= \int \left[\left\{ \frac{\exp(\mathbf{x}'_{iu}\boldsymbol{\beta} + z_{iu}^*\gamma_i)}{[1 + \exp(\mathbf{x}'_{iu}\boldsymbol{\beta} + z_{iu}^*\gamma_i)]} \right\} \left\{ \frac{\exp(\mathbf{x}'_{it}\boldsymbol{\beta} + z_{it}^*\gamma_i)}{[1 + \exp(\mathbf{x}'_{it}\boldsymbol{\beta} + z_{it}^*\gamma_i)]} \right\} \right] \\ &\quad dG_N(\gamma_i, \tilde{\sigma}_\gamma^2) \\ &\quad - \tilde{\mu}_{iu}(\mathbf{x}_{iu}, z_{iu}^*, \boldsymbol{\beta}, \tilde{\sigma}_\gamma^2)\tilde{\mu}_{it}(\mathbf{x}_{it}, z_{it}^*, \boldsymbol{\beta}, \tilde{\sigma}_\gamma^2), \end{aligned} \tag{5.6}$$

yielding the $(t - u)$ lag correlation as

$$\begin{aligned} \text{corr}[Y_{iu}, Y_{it}] &= \frac{\tilde{\lambda}_{iut}(\mathbf{x}_{iu}, \mathbf{x}_{it}, z_{iu}^*, z_{it}^*, \boldsymbol{\beta}, \tilde{\sigma}_\gamma^2)}{[\tilde{\mu}_{iu}(\cdot)(1 - \tilde{\mu}_{iu}(\cdot))\tilde{\mu}_{it}(\cdot)(1 - \tilde{\mu}_{it}(\cdot))]^{\frac{1}{2}}} \\ &\quad - \frac{[\tilde{\mu}_{iu}(\cdot)\tilde{\mu}_{it}(\cdot)]^{\frac{1}{2}}}{[(1 - \tilde{\mu}_{iu}(\cdot))(1 - \tilde{\mu}_{it}(\cdot))]^{\frac{1}{2}}}, \end{aligned} \tag{5.7}$$

which contains time varying covariates but provides equi-correlations for any (u, t) when these covariates are same over time. Thus, it does not show any decaying correlations when lag $|t - u|$ increases.

Some authors in the past such as Stiratelli et al. (1984, Eqns. (2.2), (3.1)–(3.2)), recognizing that serial correlations play an important role for

longitudinal binary data, to reflect such correlations, as opposed to Eq. 5.4, they have used a logit link random effects model similar but different than Eq. 5.1. More specifically,

$$\text{logit}(\tilde{p}_{it}(\boldsymbol{\beta}, \gamma_{it})) \equiv \ell(\tilde{p}_{it}) = \mathbf{x}'_{it}(y_{i,t-1}, \dots, y_{i1})\boldsymbol{\beta} + \tilde{z}_{it}\gamma_{it}, \quad (5.8)$$

where the covariates \mathbf{x}_{it} is composed of the past binary responses as given covariates, and $\tilde{\boldsymbol{\gamma}}_i = (\gamma_{i1}, \dots, \gamma_{it}, \dots, \gamma_{iT})'$ denote the variable random effects of the i -th individual over the time period T . As far as the distribution of $\tilde{\boldsymbol{\gamma}}_i$ is concerned, the authors have considered

$$\tilde{\boldsymbol{\gamma}}_i \sim N(0, \tilde{D}), \quad (5.9)$$

similar to Eq. 5.2, and estimated $\boldsymbol{\beta}$ and $\tilde{D} : T \times T$, using the so-called empirical Bayes estimation approach. Notice that the dimension of $\boldsymbol{\beta}$ in Eqs. 5.8 and 5.4 is different. This is because $\boldsymbol{\beta}$ in Eq. 5.8 also contains the regression effects/parameters of the past binary responses.

As compared to the logit link model (5.8), Chib and Jeliazkov (2006, Eqns. (1),(6)) have used a more general semi-parametric dynamic mixed model for longitudinal binary data, constructed based on a latent linear semi-parametric dynamic mixed model. This allows one either to use logit or probit links. More specifically, suppose that y_{it}^* is an unobservable continuous variable satisfying a linear semi-parametric dynamic mixed model, as

$$g_{it}^* = E[Y_{it}^*|\cdot] = \mathbf{x}'_{it}\boldsymbol{\beta} + \tilde{z}_{it}\gamma_{it} + \phi_1 y_{i,t-1} + \dots + \phi_m y_{i,t-m} + g(s_{it}) + \epsilon_{it}, \quad (5.10)$$

where $y_{i,t-j}$ is the binary response occurred in the past at time $(t - j)$ with its regression effect ϕ_{t-j} , $g(s_{it})$ is a smooth non-parametric function in s_{it} covariates, and ϵ_{it} is the model error. Next, suppose that the binary response y_{it} be determined based on the relationship

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (5.11)$$

Note that if y_{it}^* follows a logistic (L) distribution (e.g., Johnson and Kotz (1970)) with mean g_{it}^* as in Eq. 5.10, and variance $\frac{\pi^2}{3}$, then by using the condition in Eq. 5.11, one can compute the binary probability as

$$Pr(Y_{it} = 1|g_{it}^*) = \int_{-\infty}^{g_{it}^*} f_L(y_{it}^*) dy_{it}^* = \frac{\exp(g_{it}^*)}{1 + \exp(g_{it}^*)} = \pi_{it}^{**}(\cdot), \quad (5.12)$$

which has the same form as in Eq. 1.4, with a difference in the formula for g_{it}^* , specifically in Eq. 1.4 g_{it}^* has the dynamic form, whereas g_{it}^* in Eq. 5.10 considered by Chib and Jeliazkov (2006) has the dynamic mixed model form which is a logit link function ($\text{logit}(\pi_{it}^{**}) = g_{it}^*$) for clustered longitudinal data (see Sutradhar (2011, Chapter 11) for similar familial/cluster longitudinal binary data). For linear data, g_{it}^* itself is the linear link function which has been studied by some authors such as Das et al. (2013) in a Bayesian setup. For some more discussions on binary dynamic mixed models, similar to that of Chib and Jeliazkov (2006), one may be referred to Sutradhar et al. (2010), for example, in a parametric setup, and Congdon (2014, Section 7.1.1, p. 287), among others, in a Bayesian frame work.

6 Concluding Remarks

This review paper clarifies at least two main misconceptions around the analysis of correlated binary data collected under a cluster in both cross-sectional cluster and longitudinal cluster data setups. First many authors over the past forty years used random effect models to model the correlations for longitudinal binary data. This approach is either misleading or too restrictive. This is because similar to time series modeling, the longitudinal correlations are best modeled through suitable dynamic models relating repeated responses from the same individual. More clearly a common individual random effect among longitudinal responses is unable to address the time effects on the binary responses, rather it generates equi-correlations type structure among the repeated responses which is too restrictive.

Second, in both cross-sectional and longitudinal cluster setups, many studies have pre-specified the marginal means as the function of regression parameters only which may lead to inconsistent regression estimates when a mixed effects model is the true model for the marginal means. To clarify this issue in the cross-sectional cluster setup, we have considered 3 different important situations where fixed effects based marginal means may or may not appropriate. (A) In the first approach, random cluster effects are assumed to follow a normal distribution, and a likelihood function is constructed averaging (referred to as population average (PA)) the conditional likelihood function (which is a product of independent binary distributions conditional on the cluster effects) over the normal random effects, and then the likelihood estimates of the regression and cluster variance parameters are obtained and interpreted. Under this approach, the binary response means were shown to have a marginal mixed effects (MME) model. Thus any fixed effects based marginal mean specification in such cases is bound

to produce inconsistent regression estimates, which is a serious inference issue. (B) In the second approach, certain suitable distributions for the random cluster effects were technically developed so that it provides a marginal fixed effects (MFE) model for the binary means involving only the regression parameters (referred to as the subject specific (SS) regression effects), which may be estimated and interpreted by using the likelihood estimates computed in the same way as in (A). But, these distributional assumptions such as so-called “bridge” or beta- binary distributions for the random effects or their functions, are too narrow or restrictive for practical use. (C) In the third approach, no assumption is made about the random cluster effects distribution, in stead an arbitrary MFE (AMFE) model was used for the means involving only regression parameters. Also in this approach no attempts were made to develop any correlation structure or likelihood function, in stead a ‘working’ correlation structure based GEE (generalized estimating equations) approach was used for the SS regression parameters estimation. This approach is misleading as under (A) one never gets a fixed mean model, and under (B) only a limited number of assumptions for the distribution of random effects, those too technically restrictive, may lead to a marginal fixed effects based mean model. In summary, because normal random effects based cluster model (A) is quite practical, we have given details for estimation of the mean model (involving both β and σ_γ^2) using the so-called GQL approach. As asymptotic properties of such estimators are not available, they (consistency and normality) are discussed in details.

The paper has also equally studied the longitudinal clustered models for binary data where repeated responses from an individual are collected over a short period. Longitudinal correlations arise due to a dynamic relationship among the present and past binary responses and they are different than clustered correlations. Similar to the cluster setup, it is shown that in many situations MFE model can not be used to study the regression effects. For example, an alternative MD/MR (marginal dynamic/recursive) model does not produce fixed effects based mean model. The existing GEE approach is not useful in such a situation because the recursive means contains both regression and correlation parameters, whereas GEE is based on fixed effects based marginal means.

Furthermore, there also exists some studies dealing with clustered and/or panel/longitudinal binary data in a Bayesian setup. These studies are based on generalized linear mixed models with certain suitable link functions to reflect the correlations of the clustered and/or longitudinal data. But the inferences are not made based on any correlation structures, rather they

exploit conditional likelihood using monte carlo techniques. We have highlighted some of these important studies in this paper.

Acknowledgments. The author would like to thank a referee and the Associate Editor for their valuable comments and suggestions that lead to the improvement of the paper.

References

- AMEMIYA, T. (1985). *Advanced Econometrics*. Harvard University Press, Cambridge.
- BAHADUR, R.R. (1961). *A representation of the joint distribution of responses to n dichotomous items*, 6, SOLOMON, H. (ed.), p. 158–168.
- BRESLOW, N.E. and CLAYTON, D.G. (1993). Approximate inference in generalized linear mixed models. *Journal of American Statistical Association* **88**, 9–25.
- BRESLOW, N.E. and LIN, X. (1995). Bias correction in generalized linear models with a single component of dispersion. *Biometrika* **82**, 81–92.
- CHEN, Z., YI, G.Y. and WU, C. (2011). Marginal methods for correlated binary data with misclassified responses. *Biometrika* **98**, 647–662.
- CHIB, S. and JELIAZKOV, I. (2006). Inference in semiparametric dynamic models for binary longitudinal data. *Journal of American Statistical Association* **101**, 685–700.
- CROWDER, M. (1995). On the use of a working correlation matrix in using generalized linear models for repeated measures. *Biometrika* **82**, 407–410.
- CONGDON, P. (2014). *Applied Bayesian Modelling*. Wiley, New York.
- COX, D.R. (1972). The analysis of multivariate binary data. *Appl. Stat.* **21**, 113–120.
- DANIELS, M.J. and GATSONIS, C. (1999). Hierarchical generalized linear models in the analysis of variations in health care utilization. *Journal of American Statistical Association* **94**, 29–42.
- DAS, K., LI, R., SENGUPTA, S. and WU, R. (2013). A Bayesian semiparametric model for bivariate sparse longitudinal data. *Stat. Med.* **32**, 3899–3910.
- EKHOLM, A., SMITH, P.W.F. and MCDONALD, J.W. (1995). Marginal regression analysis of a multivariate binary response. *Biometrika* **82**, 847–854.
- FOKIANOS, K. and KEDEM, B. (2003). Regression theory for categorical time series. *Stat. Sci.* **18**, 357–376.
- GELFAND, A.E. and CARLIN, B.P. (1993). Maximum likelihood estimation for constrained or missing data problems. *Canadian Journal of Statistics* **21**, 303–311.
- HASEMAN, J.K. and KUPER, J.K. (1979). Analysis of dichotomous response data from certain toxicological experiments. *Biometrics* **35**, 281–294.
- HENDERSON, C.R. (1963). *Selection index and expected genetic advance*. National Academy of Sciences, p. 141–63.
- JIANG, J. (1998). Consistent estimators in generalized linear mixed models. *Journal of American Statistical Association* **93**, 720–729.
- JOHNSON, N. L. and KOTZ, S. (1970). *Continuous Multivariate Distributions-2*. Wiley, New York.
- KANTER, M. (1975). Auto-regression for discrete processes mod 2. *Journal of Applied Probability* **12**, 371–375.
- KARIM, M.R. and ZEGER, S.L. (1992). Generalized linear models with random effects: Salamander mating revisited. *Biometrics* **48**, 631–644.
- KUK, A.Y.C. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *J. R. Statist. Soc. B* **58**, 619–678.

- LAIRD, N.M. and WARE, J.H. (1982). Random effects models for longitudinal data. *Biometrics* **38**, 963–974.
- LEE, Y. and NELDER, J. (1996). Hierarchical generalized linear models. *Journal of Royal Statistical Society, B* **58**, 619–678.
- LIANG, K.Y. and ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.
- LIANG, K.-Y., ZEGER, S.L. and QAQISH, B. (1992). Multivariate regression analysis for categorical data. *J. Roy. Statist. Soc. Ser. B* **54**, 3–40.
- LIN, X. and BRESLOW, N.E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Am. Statist. Assoc.* **91**, 1007–1016.
- LIN, X. and CARROLL, R.J. (2001). Semiparametric regression for cluster data using generalized estimating equations. *J. Am. Statist. Asso.* **96**, 1045–1056.
- LIPSITZ, S.R., LAIRD, N.M. and HARRINGTON, D.P. (1991). Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. *Biometrika* **78**, 153–160.
- LOREDO-OSTI, J.C. and SUTRADHAR, B.C. (2012). Estimation of regression and dynamic dependence parameters for non-stationary multinomial time series. *J. Time Ser. Anal.* **33**, 458–467.
- MCCULLAGH, P. and NELDER, J.A. (1989). *Generalized Linear Models*. Chapman and Hall, London.
- MCCULLOCH, C.E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *Journal of American Statistical Association* **92**, 162–170.
- MCDONALD, D.R. (2005). The local limit theorem: a historical perspective. *Journal of Iranian Statistical Society* **4**, 73–86.
- MCGILCHRIST, C.A. (1994). Estimation in generalised linear mixed models. *J. R. Statist. Soc.* **B56**, 61–69.
- NEUHAUS, J.M. (2002). Analysis of clustered and longitudinal binary data subject to response misclassification. *Biometrics* **58**, 675–683.
- NEUHAUS, J.M., KALBFLEISCH, J.D. and HAUCK, W.W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *Int. Stat. Rev.* **59**, 25–35.
- PARZEN, M. and ET AL. (2011). A generalized linear mixed model for longitudinal binary data with a marginal logit link function. *The Annals of Applied Statistics* **5**, 449–467.
- PRENTICE, R.L. (1986). Binary regression using an extended Beta-binomial distribution. *Journal of American Statistical Association* **81**, 321–327.
- SCHALL, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–727.
- STIRATELLI, R., LAIRD, N. and WARE, J.H. (1984). Random effects model for serial observations with binary response. *Biometrics* **40**, 961–971.
- SUTRADHAR, B.C. (2003). An overview on regression models for discrete longitudinal responses. *Stat. Sci.* **18**, 377–393.
- SUTRADHAR, B.C. (2004). On exact quasi-likelihood inference in generalized linear mixed models. *Sankhya B: The Indian Journal of Statistics* **66**, 261–289.
- SUTRADHAR, B.C. (2010). Inferences in generalized linear longitudinal mixed models, vol. 38.
- SUTRADHAR, B.C. (2011). *Dynamic Mixed Models for Familial Longitudinal Data*. Springer, New York.
- SUTRADHAR, B.C. (2014). *Longitudinal Categorical Data Analysis*. Springer, New York.

- SUTRADHAR, B.C. and ALI, M.M. (1989). A generalization of the Wishart distribution for the elliptical model and its moments for the multivariate t model. *J. Multivar. Anal.* **29**, 155–162.
- SUTRADHAR, B.C. and DAS, K. (1997). Generalized linear models for beta correlated binary longitudinal data. *Communications in Statistics- Theory and Methods* **26**, 617–635.
- SUTRADHAR, B.C. and DAS, K. (1999). On the efficiency of regression estimators in generalized linear models for longitudinal data. *Biometrika* **86**, 459–465.
- SUTRADHAR, B.C., BARI, W. and DAS, K. (2010). On probit versus logit dynamic mixed models for binary panel data. *J. Stat. Comput. Simul.* **80**, 421–441.
- SUTRADHAR, B.C. and FARRELL, P.J. (2007). On optimal lag 1 dependence estimation for dynamic binary models with application to asthma data. *Sankhya B* **69**, 448–467.
- SUTRADHAR, B.C. and MUKERJEE, R. (2005). On likelihood inference in binary mixed model with an application to COPD data. *Computational Statistics and Data Analysis* **48**, 345–361.
- SUTRADHAR, B.C. and ZHENG, N. (2018). Inferences in binary dynamic fixed models in a semiparametric setup. *Sankhya B* **80**, 263–291.
- WANG, Z. and LOUIS, T.A. (2003). Matching conditional and marginal shapes in binary random intercept models using a bridge distribution function. *Biometrika* **90**, 765–775.
- WANG, Z. and LOUIS, T.A. (2004). Marginalized binary mixed-effects models with covariate-dependent random effects and likelihood inference. *Biometrics* **60**, 884–891.
- WEDDERBURN, R. (1974). Quasilielihood functions, generalized linear models and the Gauss-Newton method. *Biometrika* **61**, 439–447.
- YI, G.Y. and COOK, R.J. (2002). Marginal methods for incomplete longitudinal data arising in clusters. *Journal of American Statistical Association* **97**, 1071–1080.
- ZEGER, S.L., LIANG, K.Y. and ALBERT, P.S. (1988). Models for longitudinal data: a generalized estimating equations approach. *Biometrics* **44**, 1049–1060.
- ZEGER, S.L., LIANG, K.Y. and SELF, S.G. (1985). The analysis of binary longitudinal data with time independent covariates. *Biometrika* **72**, 31–38.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Appendix A. Higher Order Moments (up to order 4) for Clustered Binary Responses

To compute $\Omega_i(\boldsymbol{\beta}, \sigma_\gamma^2)$ in Eq. 4.38 on top of $\text{var}(Y_{ij})$ and $\text{cov}(Y_{ij}, Y_{ik})$, we need the formulas for certain specific third and fourth order moments as follows.

Computation of $\text{var}(Y_{ij}Y_{ik})$ This variance is computed as

$$\begin{aligned} \text{var}(Y_{ij}Y_{ik}) &= E[Y_{ij}^2 Y_{ik}^2] - [E(Y_{ij}Y_{ik})]^2 \\ &= E(Y_{ij}Y_{ik}) - [E(Y_{ij}Y_{ik})]^2 = \lambda_{i,jk}^{BA} [1 - \lambda_{i,jk}^{BA}] = \omega_{i,jkk}^{BA}, \end{aligned} \quad (\text{A.1})$$

where $\lambda_{i,jk}^{BA}$ is computed by Eq. 4.4.

Computation of $\text{cov}(Y_{ij}, Y_{ik}Y_{il}) = \phi_{i,jkl}(\boldsymbol{\beta}, \sigma_\gamma^2)$ Because

$$\text{cov}(Y_{ij}, Y_{ik}Y_{il}) = E[Y_{ij}Y_{ik}Y_{il}] - \mu_{ij}^{BA}\lambda_{i,kl}^{BA}, \tag{A.2}$$

we need the formula for the third order moments, namely

$$\begin{aligned} E[Y_{ij}Y_{ik}Y_{il}] &= E_{\gamma_i} E[\{Y_{ij}Y_{ik}Y_{il}\}|\gamma_i] \\ &= E_{\gamma_i} [E(Y_{ij}|\gamma_i)E(Y_{ik}|\gamma_i)E(Y_{il}|\gamma_i)] \\ &= \int p_{ij}^*(\boldsymbol{\beta}, \gamma_i)p_{ik}^*(\boldsymbol{\beta}, \gamma_i)p_{il}^*(\boldsymbol{\beta}, \gamma_i)g_N(\gamma_i)d\gamma_i, \end{aligned} \tag{A.3}$$

where, for example, $p_{ij}^*(\boldsymbol{\beta}, \gamma_i) = \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)/[1 + \exp(\mathbf{x}'_{ij}\boldsymbol{\beta} + \gamma_i)]$, and $g_N(\gamma_i) \equiv [\gamma_i \sim N(0, \sigma_\gamma^2)]$. Similar to Eq. 4.4, this normal integration in Eq. A.3 may be computed approximately by

$$\begin{aligned} &\lambda_{i,jkl}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) \\ &= \sum_{v_i=0}^V p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i))p_{ik}^*(\mathbf{x}_{ik}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \\ &\times p_{il}^*(\mathbf{x}_{il}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}, \end{aligned} \tag{A.4}$$

yielding

$$\phi_{i,jkl}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = \lambda_{i,jkl}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) - \mu_{ij}^{BA}\lambda_{i,kl}^{BA}. \tag{A.5}$$

Computation of $\text{cov}(Y_{ij}Y_{ik}, Y_{il}Y_{im}) = \omega_{i,jklm}(\boldsymbol{\beta}, \sigma_\gamma^2)$ By similar calculations as in Eq. A.4, one obtains

$$\omega_{i,jklm}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) = \lambda_{i,jklm}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) - \lambda_{i,jk}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2)\lambda_{i,lm}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2), \tag{A.6}$$

where

$$\begin{aligned} &\lambda_{i,jklm}^{BA}(\boldsymbol{\beta}, \sigma_\gamma^2) \\ &= \sum_{v_i=0}^V p_{ij}^*(\mathbf{x}_{ij}; \boldsymbol{\beta}, \sigma_\gamma h(v_i))p_{ik}^*(\mathbf{x}_{ik}; \boldsymbol{\beta}, \sigma_\gamma h(v_i))p_{il}^*(\mathbf{x}_{il}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \\ &\times p_{im}^*(\mathbf{x}_{im}; \boldsymbol{\beta}, \sigma_\gamma h(v_i)) \binom{V}{v_i} (1/2)^{v_i} (1/2)^{V-v_i}. \end{aligned} \tag{A.7}$$

BRAJENDRA C. SUTRADHAR
MEMORIAL UNIVERSITY, ST. JOHN'S, NL,
CANADA
E-mail: bsutradh@mun.ca

Paper received: 31 December 2020; accepted 5 June 2021.