INFERENCES IN NON-STATIONARY LONGITUDINAL BINARY MODELS

TRACEY WOODEN

Inferences in Non-stationary Longitudinal Binary Models

by

© Tracey Wooden

A thesis submitted to the

School of Graduate Studies

in partial fulfillment of the

requirements for the degree of

Master of Science in Statistics

Department of Mathematics and Statistics

Memorial University of Newfoundland

May 2013

St John's Newfoundland & Labrador

Abstract

Longitudinal binary data has been analyzed over the last three decades either by using odds ratio or 'working' correlations as a measure of association between the repeated binary responses. Recently, this type of data has been analyzed by modeling the correlations parametrically and estimating the parameters by a generalized quasi-likelihood (GQL) approach. In this thesis, we consider a specific correlation model, namely, the binary autoregressive order 1 (AR(1)) model to generate the data, and study the relative performance of the odds ratio and equi-correlations based estimation approaches with the GQL approach. This comparison is mainly done by simulations under both stationary and non-stationary AR(1) correlation models. A real life data set containing repeated asthma status of a group of children is also analyzed.

Acknowledgments

First and foremost, I would like to thank God for providing me with this opportunity, and granting me the capability to write this thesis. He has blessed my life considerably, and without his strength and direction, I would not be who I am today.

I want to express my deepest appreciation and gratitude to my supervisor, Professor B.C. Sutradhar. Without his guidance, patience and persistent help this thesis would not have been possible. It has been a great privilege to work under his supervision. I also want to thank my co-supervisor, Dr. J.C. Loredo-Osti for his support and help throughout my masters program.

I would like to acknowledge the financial, academic and technical support of Memorial University of Newfoundland and its staff. In particular, I want to acknowledge the financial support provided by the Department of Mathematics and Statistics, Prof. B.C. Sutradhar and Dr. J.C. Loredo-Osti.

To my family and friends, I would like to say thank you for years of support and

encouragement. I am most grateful to my parents and sister, they have given me their unequivocal support throughout, for which my mere expression of thanks does not suffice.

I would like to thank my dear friend Kendel Boehler, who over the years has become like a sister to me, for her kindness, friendship and support.

Last, but by no means least, I thank the love of my life, my husband Leo for his personal support and patience, and always reminding me of the importance of taking a moment to enjoy the small things life has to offer.

Table of Contents

List of Tables

Chapter 1

Introduction

1.1 Background of the Problem

Longitudinal binary data analysis is an important research topic. In this setup, binary responses are repeatedly collected over a small period of time from a large number of independent individuals. For example, consider a binary longitudinal study referenced in Sutradhar (2003) [see also Zeger, Liang, and Albert (1988)]. A data set for 537 children from Ohio state was examined annually from the ages 7-10. Whether the children had respiratory infection in the previous year was reported by the mother of the child. That is, the repeated response is the wheezing status (1=yes, 0=no) of the child. The initial smoking status of the mothers (1=regular smoker, 0= not) were

also collected, and considered as a covariate. The interest in such a study would be to investigate the effect that smoking by a mother on the wheeze status of her child. We also refer to another longitudinal binary data study reported by Fitzmaurice and Lipsitz (1995). In this case, data from 51 subjects from an arthritis clinical trial were collected. In this study, patients complete a self-assessment measurement of arthritis (0 = 'poor', 1 = 'good'). Each patient had a base-line self-assessment measurement at week 0, and then follow-up measurements at weeks $1, 5, 9$ and 13. Patients were then randomly assigned to one of the two treatments. Overall, four covariates including the treatment were considered, the three other covariates being gender, age, and time factor. The interest in this particular study is whether the treatment increases the possibility of a positive self-assessment. In addition, the secondary interest of such a study would be to investigate whether the response differs by age and gender.

Note that the repeated binary responses, such as wheezing status of a child over 4 years in the aforementioned first problem, are likely to be correlated. Similarly, in the second problem, the binary responses, namely the arthritis status, taken repeatedly at four time points are also likely to be correlated. This type of longitudinal correlations must be accounted for while estimating the effects of the covariates (regression effects) on the binary responses. However, the modeling of binary correlations is often diffi-

cult. Consequently, many authors have used a 'working' correlations approach. For example, in relation to the second problem, Fitzmaurice and Lipsitz (1995) [see also Lipsitz *et al.* (1991)] have computed the associated covariate matrix of an individual by using a 'working' odds ratio approach and then using such covariance matrices to construct the so-called generalized estimating equation (GEE) [Liang and Zeger, 1986]. Some authors have used **equi-correlations** instead of odds ratios to obtain correlation structure based regression estimates. See, for example, Prentice (1988) in the context of a dietary screening problem. However, this selection of equi-correlations was not adequately justified. Moreover, when covariates are time dependent, the correlations for binary responses no longer follow equi-correlations (Sutradhar, 2010). In addition, one may argue that a correlation structure exhibiting decay in correlations as time lag increases, would be the better choice. This decaying pattern can be modeled by using a Gaussian type AR(1) (autoregressive) relationship. By this token, it would be more appropriate to use a possible class of correlation structure (Sutradhar 2011, Chapter 7) that may accommodate AR(1) and equi-correlations, and so on, as specialized structures.

Note that the odds ratios are computed from a bivariate binary distribution for a pair of binary responses collected from two transiting times. One may illustrate this

odds ratio computation as follows. Let y_{it} be the binary response, i.e $y_{it} = 1$ or 0, for the i^{th} individual $(i = 1, \ldots, n)$ at time t $(t = 1, \ldots, T)$. Also, let $\pi_{it} = \mathrm{P}(y_{it} = 1)$ be the marginal probabilities at time t. Suppose that $\mathrm{P}(y_{iu} = 1, y_{it} = 1)$ is denoted by $\pi_{i,11}^{(t|u)}$. Then for all possible transitions from time u to t, the joint and marginal probabilities may be written as in Table 1.1.

Table 1.1: Joint probabilities in contingency table form.

		Time t		
Time u		1 (y=1)	2 (y=0)	Total
	1 (y=1)	$\pi_{i,11}^{(t\|u)}$	$\pi_{i,12}^{(t\|u)}$	π_{iu}
	2 (y=0)	$\pi_{i,21}^{(t\|u)}$	$\pi_{i,22}^{(t\|u)}$	$1 - \pi_{iu}$
Total		π_{it}	$1 - \pi_{it}$	1

We may then write the log odds ratios for the i^{th} individual as

$$\psi_{iut} = \log \frac{\pi_{i,11}^{(t|u)} \pi_{i,22}^{(t|u)}}{\pi_{i,12}^{(t|u)} \pi_{i,21}^{(t|u)}} = \log \tau_{iut}, (\text{say}). \tag{1.1.1}$$

Suppose that $\sum_{i=1}^{n} y_{it} = n_t$ and $\sum_{i=1}^{n} y_{iu} y_{it} = n_{11}^{(t|u)}$. When the probabilities in Table 1.1 are the same for all individuals $i = 1, \ldots, n$, we may consequently construct a contingency table using the notation $\sum_{i=1}^{n} y_{it} = n_t$ and $\sum_{i=1}^{n} y_{iu} y_{it} = n_{11}^{(t|u)}$, for example. The 2×2 contingency table may then be formed as

Table 1.2: Contingency table for general time points.

		Time t		
Time u		1 (y=1)	2 (y=0)	Total
	1 (y=1)	$n_{11}^{(t\|u)}$	$n_{12}^{(t\|u)}$	n_u
	2 (y=0)	$n_{21}^{(t\|u)}$	$n_{22}^{(t\|u)}$	$n - n_u$
Total		n_t	$n - n_t$	n

Thus, in this situation, one can estimate the common odds ratio τ_{ut} ($\tau_{ut} \equiv \tau_{iut}$) by

$$\hat{\tau}_{ut} = \frac{n_{11}^{(t|u)} n_{22}^{(t|u)}}{n_{12}^{(t|u)} n_{21}^{(t|u)}}. \tag{1.1.2}$$

Next, using further assumption that the odds ratio are the same for any two time points, i.e, using $\tau_{ut} = \tau$, some authors such as Lipsitz *et al.* (1991) [see also Fitzmaurice and Lipsitz (1995)] computed the joint probabilities in order to construct the 'working' equal odds ratio based GEE. It is, however, clear that common τ based GEE approach is not appropriate when odds ratio vary from individual to individual. Also, they may not be the same for all time points. But the effect of using such τ based GEE when τ_{iut} is appropriate is not adequately discussed in the literature. In this thesis, we revisit this inference issue and examine the performance of the GEE approach by generating data under a non-stationary correlation model.

As opposed to technically using equi-correlations or odds ratio for binary responses

over time, there exist some parametric modeling to understand the correlations. For convenience, we explain below some of these correlation models for a specialized stationary case, where covariates x_{it} corresponding to y_{it} are assumed to be the same for all $t = 1, \ldots, T$. Let $x_{i.} = x_{it}$ to represent this situation. Note that using this notation, we write

$$\mathrm{P}(y_{it} = 1) = \frac{e^{x_{i.}'\beta}}{1 + e^{x_{i.}'\beta}} = \pi_{i.}, \quad t = 1, \ldots, T, \tag{1.1.3}$$

which will be used as the marginal probabilities under all models explained below.

1.1.1 Stationary Correlated Binary Models

1.1.1.1 Bahadur Model

For $T \geq 2$, Bahadur (1961) [see also Sutradhar 2011, Chapter 7] introduced a joint probability function-based correlation model given by

$$f(y_{i1}, \ldots, y_{iT}) = \prod_{t=1}^{T} \pi_{i.}^{y_{it}} (1 - \pi_{i.})^{1 - y_{it}} \times \left[1 + \sum_{t<u}^{T} \rho_{i,ut}^{*} \left(\frac{y_{iu} - \pi_{i.}}{[\pi_{i.}(1-\pi_{i.})]^{\frac{1}{2}}} \right) \left(\frac{y_{it} - \pi_{i.}}{[\pi_{i.}(1-\pi_{i.})]^{\frac{1}{2}}} \right) \right] \tag{1.1.4}$$

where $\rho_{i,ut}^{*}$ denotes the correlation between two binary responses y_{it} and y_{iu}. From (1.1.4), one can then show that

$$\begin{aligned} \mathrm{E}(Y_{it}) &= \pi_{i.} \\ \mathrm{Var}(Y_{it}) &= \pi_{i.}(1-\pi_{i.}) \\ \mathrm{Corr}(Y_{it}, Y_{iu}) &= \rho^*_{i,ut}. \end{aligned} \tag{1.1.5}$$

Alternatively, (1.1.4) can also be expressed as

$$f(y_{i1}, \ldots, y_{iT}) = 1 + \left(\frac{\sum_{t<u}^{T} \rho^*_{i,ut} (-1)^{y_{it}+y_{iu}} \pi_{i.}^{2-y_{it}-y_{iu}} (1-\pi_{i.})^{y_{it}+y_{iu}}}{\pi_{i.}(1-\pi_{i.})} \right), \tag{1.1.6}$$

where $y_{it} = 0, 1$ for any i and all $t = 1, \ldots, T$ (Sutradhar 2011, Chapter 7, eq. 7.12).

This model may however, encounter range restriction problems for correlations $\rho^*_{i,ut}$. To be specific, the range for $\rho^*_{i,ut}$ may be much narrower than -1 to 1.

1.1.1.2 Kanter Model

Suppose that $y_{i1} \sim b(\pi_{i.})$, and let s_{it} and d_{it} denote binary random variables with

$$\begin{aligned} \Pr(s_{it} = 1) &= \gamma_1, \quad \text{with } 0 < \gamma_1 < 1, \\ \Pr(d_{it} = 1) &= \xi^*_i = \frac{\pi_{i.}(1-\gamma_1)}{1-2\gamma_1\pi_{i.}}, \end{aligned} \tag{1.1.7}$$

for $t = 2, \ldots, T$. Kanter (1975) proposed that one may generate the AR(1)-type

correlated responses $y_{i1}, \ldots, y_{iT}$ by using the following model

$$y_{it} = s_{it}\,(y_{i,t-1} \oplus d_{it}) + (1 - s_{it})d_{it} \qquad \text{for } t = 2, \ldots, T, \tag{1.1.8}$$

were $\oplus$ denotes addition mod 2. Since $y_{it} \sim b(\pi_{i.})$ for $t = 1, \ldots, T$, using (1.1.8) one may then show that

$$\begin{aligned} \mathrm{E}(Y_{it}) &= \pi_{i.} \\ \mathrm{Var}(Y_{it}) &= \pi_{i.}(1 - \pi_{i.}). \end{aligned} \tag{1.1.9}$$

In addition, the correlation between y_{it} and y_{iu} can then be found as

$$\mathrm{Corr}(Y_{it}, Y_{iu}) = \rho^{*}_{i,ut} = \rho_i^{|t-u|}, \qquad \text{for } t \neq u, \tag{1.1.10}$$

where

$$\rho_i = \frac{\gamma_1(1 - 2\pi_{i.})}{(1 - 2\gamma_1\pi_{i.})}. \tag{1.1.11}$$

Note that this model also encounters range restrictions for the correlations.

1.1.1.3 Binary Dynamic Logit (BDL) Model

There exists a binary model where correlations are not restricted. For example, consider the following non-linear binary dynamic model:

$$\begin{aligned} p_{i1} &= Pr[y_{i1} = 1] = \pi_{i.} = exp(x'_{i.}\beta)/[1 + exp(x'_{i.}\beta)] \\ p_{it|t-1} &= Pr[y_{it} = 1|y_{i,t-1}] = \frac{exp(x'_{i.}\beta + \theta y_{i,t-1})}{1 + exp(x'_{i.}\beta + \theta y_{i,t-1})}, \end{aligned} \tag{1.1.12}$$

for $t = 2, \ldots, T$, where θ denotes the dynamic dependence parameter.

When $t = 1$, y_{i1} is assumed to be binary with mean $\mu_{i1} = \pi_{i.}$. Let

$$p_{it|t-1(y_{i,t-1}=1)} = \frac{exp(x'_{i.}\beta + \theta)}{1 + exp(x'_{i.}\beta + \theta)} = \tilde{p}_i \quad \text{for all } t = 2, \ldots, T. \tag{1.1.13}$$

The unconditional mean of y_{it} is found using (1.1.13) as,

$$\mu_{it} = \mathrm{E}(Y_{it}) = \Pr(y_{it} = 1) = \pi_{i.} + \mu_{i,t-1}(\tilde{p}_{i.} - \pi_{i.}), \qquad \text{for } t = 2, \ldots, T, \tag{1.1.14}$$

along with variance of y_{it} as

$$\sigma_{i,tt} = \mathrm{var}(Y_{it}) = \mu_{it}(1 - \mu_{it}). \tag{1.1.15}$$

Furthermore, it can be shown that the lag $(t-u)$ autocorrelation between y_{iu} and y_{it} is found as

$$\text{Corr}(Y_{iu}, Y_{it}) = \sqrt{\frac{\mu_{iu}(1-\mu_{iu})}{\mu_{it}(1-\mu_{it})}}\,[\tilde{p}_i - \pi_i]^{t-u} \tag{1.1.16}$$

[Sutradhar (2011, Eqn. (7.150))].

Note that even though this model produces correlations with full range, the mean under various times maintains a recursive relationship, as opposed to Bahadur (1961) and Kanter (1975) models.

1.1.1.4 Linear Dynamic Conditional Probability (LDCP) Model

Some authors have modeled correlated binary data by way of the AR(1) type linear dynamic conditional probability (LDCP) model given by

$$Y_{i1} \sim b(\pi_{i.}) \tag{1.1.17}$$

$$Pr[Y_{it} = 1 | Y_{i,t-1} = y_{i,t-1}] = \pi_{i.} + \rho(y_{i,t-1} - \pi_{i.}), \quad \text{for } t = 2, \ldots, T \tag{1.1.18}$$

[Zeger, Liang, and Self (1985), Qaqish (2003)]. This model will then yield the following marginal mean and variance of y_{it} for all $t = 1, \ldots, T$, as

$$\begin{aligned} \mathrm{E}(Y_{it}) &= \pi_{i.} \\ \mathrm{var}(Y_{it}) &= \pi_{i.}(1-\pi_{i.}), \end{aligned} \tag{1.1.19}$$

for $u < t$. The lag $t-u$ autocorrelation is given as

$$\mathrm{Corr}[Y_{iu}, Y_{it}] = \rho^{*}_{i,ut} = \rho^{t-u}. \tag{1.1.20}$$

However, the lag 1 correlation must satisfy the range restriction

$$\max_i \left[-\frac{\pi_{i.}}{1-\pi_{i.}}, -\frac{1-\pi_{i.}}{\pi_{i.}} \right] \leq \rho \leq 1, \tag{1.1.21}$$

but as shown by Farrell and Sutradhar (2006), the ranges for correlation under this LDCP model are wider than that of Bahadur (1961) and Kanter (1975) models. Note that the comparison for ranges of correlation structures studied by Farrell and Sutradhar (2006) also include other correlation structures such as moving average of order (1) (MA(1)) and equi-correlation (EQC) structures.

1.2 Existing Estimation Approaches

Among all of the models that were discussed in Section 1.1, the LDCP approach appears to have more advantages. This is due to the fact that, except for non-linear models, as pointed out by Farrell and Sutradhar (2006), this linear model allows wider ranges for correlations than other models. For this reason, in this thesis, we concentrate on this model and discuss various inference issues. As far as the existing estimation approaches are concerned, the parameters of this type of model (assuming correlation structure is unknown) have been estimated by using certain 'working' correlation approaches. Quite often, equi-correlations and odds ratio approaches are used. Our intent is to see whether these two working approaches are sufficient in estimating the parameters of this LDCP model ((1.1.17)-(1.1.18)). To check this, we will compare their performance with an existing GQL (generalized quasi-likelihood) approach. This GQL approach (suggested by Sutradhar (2003)) assumes a Gaussian autocorrelations class which accommodates the above LDCP model. This GQL approach minimizes a generalized quadratic distance (GQD) of observations for their means, where the GQD is constructed by exploiting the correlation class.

1.3 Objective of the Thesis

Our main objective is to compare the performance of the existing 'working' odds ratio (WOR) and 'working' equi-correlation (WEQC) approaches with a recently proposed generalized quasi-likelihood (GQL) approach. More specifically, in Chapter 2, we first provide an AR(1) binary model and develop the GQL estimating equation for a general non-stationary data setup. We then described the existing estimation formulas for the WOR and WEQC approaches for the estimation of the regression effects. The formulas for the estimation of correlations and odds ratio are clearly provided. In Chapter 3, we carry out an extensive simulation study, both for stationary and non-stationary data. The relative bias (RB) and mean squared error (MSE) performances for all three approaches are presented in tables and the results discussed. In Chapter 4, for the illustration of the estimation methods, we have applied each method to reanalyze longitudinal asthma data. The results are interpreted along the lines of the simulation study. The thesis concludes in Chapter 5.

Chapter 2

Estimation for Longitudinal Binary Models

As mentioned earlier, the main objective of the thesis is to examine the relative performance of the existing 'working' equi-correlations and odds ratio based estimation approaches as compared to the GQL approach constructed based on known correlation class or structures. As far as the correlation structure is concerned, we assume that the binary data follow a non-stationary AR(1) correlation model, which would be a generalization of the stationary AR(1) model (1.1.17)-(1.1.18), see Sutradhar (2010) and Sutradhar (2011, Chapter 7) for details on such non-stationary AR(1) structures. Note that the Bahadur model given in Section 1.1.1.1 may accommodate AR(1) and

EQC structures; however the other models, namely Kanter's model in Section 1.1.1.2, BDL model in Section 1.1.1.3, and LDCP model in Section 1.1.1.4, were given for AR(1) type relationships only. One of the main reasons to consider the AR(1) structure in this thesis is that the correlation under such a model decreases exponentially as time lag increases, which is considered to be realistic for many practical data.

2.1 Non-Stationary AR(1) Correlation Models and GQL Estimation Approach

To consider a non-stationary AR(1) binary correlation model, let y_{it} be the binary response at a given time point t, with

$$\Pr(y_{it} = 1) = \pi_{it} = \frac{e^{x'_{it}\beta}}{1 + e^{x'_{it}\beta}} \qquad t = 1, \ldots, T. \tag{2.1.1}$$

Note that this marginal probability is a function of the time dependent covariate x_{it}. Suppose that the repeated binary responses $y_{i1}, \ldots, y_{it}, \ldots, y_{iT}$ are generated using the following relationships

$$\begin{aligned} \Pr(Y_{i1} = 1) &= \pi_{i1} \\ \Pr(Y_{it} = 1 | y_{i,t-1}) &= \pi_{it} + \rho(y_{i,t-1} - \pi_{i,t-1}), \qquad \text{for } t = 2, \ldots, T, \end{aligned} \tag{2.1.2}$$

where π_{it} is given in (2.1.1), for all $t = 1, \ldots, T$. Using this model, it can then be shown that the mean and variance are given as

$$\begin{aligned} \mathrm{E}(Y_{it}) &= \mu_i = \pi_{it} \\ \mathrm{var}(Y_{it}) &= \sigma_{i,tt} = \pi_{it}(1 - \pi_{it}), \end{aligned} \tag{2.1.3}$$

for $t = 1, \ldots, T$.

For $u < t$, the covariance between y_{iu} and y_{it} can be found using the model given in (2.1.1) as

$$\begin{aligned} \mathrm{cov}(Y_{iu}, Y_{it}) &= \mathrm{E}(Y_{iu} - \pi_{iu})(Y_{it} - \pi_{it}) \\ &= \mathrm{E}(Y_{iu}Y_{it}) - \mathrm{E}(Y_{iu})\mathrm{E}(Y_{it}) \\ &= \mathrm{E}_{y_{iu}} Y_{iu} \mathrm{E}_{y_{i,t-(t-u-1)}} \left[\ldots \left[\mathrm{E}_{y_{i,t-2}} \left[\mathrm{E}_{y_{i,t-1}} \left[Y_{it} | y_{i,t-1} \right] | y_{i,t-2} \right] \ldots \right] | y_{i,t-(t-u-1)} \right] \\ &\quad - \pi_{iu}\pi_{it} \\ &= \mathrm{E}_{y_{iu}} Y_{iu} \left[\pi_{it} + \sum_{j=1}^{t-u-1} \rho^j \pi_{i,t-j} + \rho^{t-u}(Y_{iu} - \pi_{iu}) - \sum_{j=1}^{t-u-1} \rho^j \pi_{i,t-j} \right] - \pi_{iu}\pi_{it} \end{aligned}$$

$$= \rho^{t-u} \mathrm{E}_{y_{iu}} [Y_{iu}(Y_{iu} - \pi_{iu})]$$

$$= \rho^{t-u} \pi_{iu}(1 - \pi_{iu})$$

$$= \rho^{t-u} \sigma_{i,uu} \tag{2.1.4}$$

[Sutradhar (2011, Eqn. (7.72))]. The non-stationary correlation matrix can then be given as

$$\mathrm{corr}(Y_{iu}, Y_{it}) = \begin{cases} \rho^{t-u} \left[\frac{\sigma_{i,uu}}{\sigma_{i,tt}}\right]^{\frac{1}{2}} & \text{for } u < t \\ \rho^{t-u} \left[\frac{\sigma_{i,tt}}{\sigma_{i,uu}}\right]^{\frac{1}{2}} & \text{for } u > t. \end{cases} \tag{2.1.5}$$

However, the parameter ρ in (2.1.5) must satisfy the following range restriction

$$\max\left[-\frac{\pi_{it}}{1-\pi_{i,t-1}}, -\frac{1-\pi_{it}}{\pi_{i,t-1}}\right] \leq \rho \leq \min\left[-\frac{1-\pi_{it}}{1-\pi_{i,t-1}}, -\frac{\pi_{it}}{\pi_{i,t-1}}\right]. \tag{2.1.6}$$

Note that because $\sigma_{i,tt}$, for example, depends on the time dependent covariates x_{it} through the relationships (2.1.3) and (2.1.1), the correlations defined in (2.1.5) are functions of time dependent covariates, and hence these correlations are non-stationary.

2.1.1 GQL Approach for the Estimation of Regression Effects (β)

It is well known that by treating binary data as independent, one may exploit the mean and variance functions and develop a quasi-likelihood (QL) estimating equation for the regression effect β given by

$$\sum_{i=1}^{n}\sum_{t=1}^{T}\frac{\partial\mu_i'}{\partial\beta}(\sigma_{i,tt})^{-1}(y_{it}-\mu_{it})=0 \tag{2.1.7}$$

[Wedderburn (1974), Sutradhar (2011, Eqn. (7.5))] where $\mu_{it}=\pi_{it}$ is the mean, and $\sigma_{i,tt}$ is the variance of y_{it} as in (2.1.3). However, because the binary data under the AR(1) model in (2.1.2) are correlated following (2.1.5), the solution of the QL estimating equation (2.1.7) for β would produce an inefficient estimate. As a remedy to this inefficiency problem, Sutradhar (2010) suggested a generalization of the QL equation (2.1.7) to accommodate the correlations in estimating β. The generalized quasi-likelihood (GQL) estimating equation is given by

$$\sum_{i=1}^{n}\frac{\partial\mu_i'}{\partial\beta}\Sigma_i^{(ns)^{-1}}(\beta,\rho)(y_i-\mu_i)=0, \tag{2.1.8}$$

where

$$\begin{aligned} y_i &= (y_{i1}, \ldots, y_{iT})' \\ \mu_i &= \pi_i = (\pi_{i1}, \ldots, \pi_{iT})', \end{aligned} \tag{2.1.9}$$

and the covariance matrix of y_i is denoted by $\Sigma_i^{(ns)}(\beta, \rho)$, such that

$$\Sigma_i^{(ns)}(\beta, \rho) = \mathrm{A}_i^{\frac{1}{2}} \mathrm{C}_i^{(ns)}(\rho) \mathrm{A}_i^{\frac{1}{2}}, \tag{2.1.10}$$

where $\mathbf{A}_i$ is defined as

$$\begin{aligned} \mathrm{A}_i &= \mathrm{diag}[\mathrm{var}(y_{i1}), \ldots, \mathrm{var}(y_{it}), \ldots, \mathrm{var}(y_{iT})] \\ &= \mathrm{diag}[\sigma_{i,11}, \ldots, \sigma_{i,tt}, \ldots, \sigma_{i,TT}] \\ &= \mathrm{diag}[\pi_{i1}(1-\pi_{i1}), \ldots, \pi_{it}(1-\pi_{it}), \ldots, \pi_{iT}(1-\pi_{iT})], \end{aligned} \tag{2.1.11}$$

and $\mathrm{C}_i^{(ns)}(\rho)$ as

$$\mathrm{C}_i^{(ns)}(\rho) = \left(\mathrm{c}_{i,u,t}^{(ns)}(x_{iu}, x_{it}, \rho) \right), \tag{2.1.12}$$

for $u, t = 1, \ldots, T$, where by (2.1.8), the $(u,t)^{\text{th}}$ element is given by

$$\begin{aligned} c_{i,ut}^{(ns)}(x_{iu}, x_{it}, \rho) &= \text{corr}(Y_{iu}, Y_{it}) \\ &= \begin{cases} \rho^{t-u}\left[\frac{\sigma_{i,uu}}{\sigma_{i,tt}}\right]^{\frac{1}{2}} & \text{for } u < t \\ \rho^{u-t}\left[\frac{\sigma_{i,tt}}{\sigma_{i,uu}}\right]^{\frac{1}{2}} & \text{for } u > t. \end{cases} \end{aligned} \tag{2.1.13}$$

Note that even though $\Sigma_i^{(ns)}(\beta, \rho)$ in (2.1.8) is a function of β, the construction of the estimating equation (2.1.8) assumes that β in $\Sigma_i^{(ns)}(\beta, \rho)$ is known. This is because when π_{it} in μ_i is known, $\sigma_{i,tt} = \pi_{it}(1 - \pi_{it})$ becomes known as well to construct the A_i matrix contained in $\Sigma_i^{(ns)}(\beta, \rho)$. Also, the correlations given in (2.1.13) depend on β only through $\sigma_{i,tt}$ or π_{it}. Thus, it is sufficient to estimate β involved in the mean function.

2.1.1.1 Computation of $\frac{\partial \mu_i'}{\partial \beta}$

The derivative $\frac{\partial \mu_i'}{\partial \beta}$ can be computed by using the formula for $\frac{\partial \pi_{it}}{\partial \beta}$, given by

$$\frac{\partial \pi_{it}}{\partial \beta} = \pi_{it}(1 - \pi_{it})x_{it} \quad : \quad p \times 1, \tag{2.1.14}$$

for all $t = 1, \ldots, T$. Thus,

$$\begin{aligned}\frac{\partial \mu_i'}{\partial \beta} &= \left[\frac{\partial \pi_{i1}}{\partial \beta}, \ldots, \frac{\partial \pi_{it}}{\partial \beta}, \ldots, \frac{\partial \pi_{iT}}{\partial \beta}\right] \\ &= [\pi_{i1}(1-\pi_{i1})x_{i1}, \ldots, \pi_{it}(1-\pi_{it})x_{it}, \ldots, \pi_{iT}(1-\pi_{iT})x_{iT}] \quad : \quad p \times T \\ &= \mathbf{X}_i' \mathbf{A}_i, \end{aligned} \tag{2.1.15}$$

where

$$\mathbf{X}_i' = [x_{i1}, \ldots, x_{it}, \ldots, x_{iT}] \quad : \quad p \times T, \tag{2.1.16}$$

and A_i is the $T \times T$ diagonal matrix as in (2.1.11).

2.1.1.2 Iterative Formula for GQL Estimator

Now for known ρ, we can solve the GQL estimating equation given in (2.1.8) for β iteratively by using the so-called Newton-Raphson formula given by

$$\begin{aligned}\hat{\beta}_{r+1} = \hat{\beta}_r + &\left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i^{(ns)^{-1}}(\beta, \rho) \mathbf{A}_i \mathbf{X}_i\right]_r^{-1} \\ &\times \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i^{(ns)^{-1}}(\beta, \rho)(y_i - \mu_i)\right]_r \end{aligned} \tag{2.1.17}$$

where $[\quad]_r$ denotes that the expression contained within the brackets is evaluated at $\beta = \hat{\beta}(r)$; that is, the r^{th} iterative value for β. Let $\hat{\beta}_{GQL}$ be the final estimate

obtained from (2.1.17).

2.1.1.3 Estimation of ρ

Note that in solving for $\hat{\beta}$ by (2.1.17), we have assumed that the correlation index parameter ρ is known. In practice, this is not usually the case, and the $AR(1)$ index parameter is unknown. To estimate this parameter consistently, we first observe that ρ is in fact a lag 1 correlation index parameter. That is, if ρ is known, then ρ^{t-u} is also known.

Consequently, to estimate the lag 1 index parameter, we construct a moment estimating equation by equating the sample lag 1 auto-covariance to its expected value. In other words, for $\tilde{y}_{it} = |y_{it} - \pi_{it}|/\sqrt{\sigma_{i,tt}}$ with $\sigma_{i,tt} = \pi_{it}(1 - \pi_{it})$,

$$\begin{aligned} \mathrm{E}\left[\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}\tilde{y}_{i,t-1}\right] &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\frac{\mathrm{E}(y_{it}-\pi_{it})(y_{i,t-1}-\pi_{i,t-1})}{\sqrt{(\sigma_{i,tt})(\sigma_{i,t-1,t-1})}} \\ &= \frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=2}^{T}\frac{\rho\sigma_{i,t-1,t-1}}{\sqrt{(\sigma_{i,tt})(\sigma_{i,t-1,t-1})}} \end{aligned} \tag{2.1.18}$$

by (2.1.4). Thus,

$$\mathrm{E}\left[\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}\tilde{y}_{i,t-1}\right] = \rho\sum_{i=1}^{n}\sum_{t=2}^{T}\sqrt{\frac{\sigma_{i,t-1,t-1}}{\sigma_{i,tt}}}. \tag{2.1.19}$$

Next, it is clear that

$$\begin{aligned} \mathrm{E}\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}^2 &= \sum_{i=1}^{n}\sum_{t=2}^{T}\frac{\mathrm{E}(y_{it}-\pi_{it})^2}{\sigma_{i,tt}} \\ &= nT. \end{aligned} \tag{2.1.20}$$

Now by combining (2.1.19) and (2.1.20), we write

$$\mathrm{E}\left[\frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}^2}\right] \simeq \frac{\mathrm{E}\left[\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}\tilde{y}_{i,t-1}\right]}{\mathrm{E}\left[\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}^2\right]} \tag{2.1.21}$$

$$= \frac{\rho\sum_{i=1}^{n}\sum_{t=2}^{T}\sqrt{\sigma_{i,t-1,t-1}/\sigma_{i,tt}}}{nT}, \tag{2.1.22}$$

yielding the moment estimate for ρ as

$$\hat{\rho} = \frac{\sum_{i=1}^{n}\sum_{t=2}^{T}\tilde{y}_{it}\tilde{y}_{i,t-1}}{\sum_{i=1}^{n}\sum_{t=1}^{T}\tilde{y}_{it}^2}\frac{nT}{\sum_{i=1}^{n}\sum_{t=2}^{T}\left[\sigma_{i,t-1,t-1}/\sigma_{i,tt}\right]^{\frac{1}{2}}}. \tag{2.1.23}$$

This estimate $\hat{\rho}$ will be used in (2.1.17) for ρ.

Note that the purpose of the thesis is to compare the finite sample performances of (2.1.17) with other possible competitive estimators, such as $\hat{\beta}_{GEE}$ based on the equi-correlation assumption [Prentice (1988, Section 4)], and $\hat{\beta}_{GEE}$ based on an odds

ratio approach. The latter two methods are discussed in detail in the next two sections.

Further note that $\hat{\beta}_{GQL}$ obtained from (2.1.3) is consistent for β. This is because

$$\mathrm{E}(\hat{\beta}_{GQL}) = \beta, \tag{2.1.24}$$

using the fact that $\mathrm{E}(Y_i) = \mu_i$ under the present $AR(1)$ model (2.1.12), yielding

$$\mathrm{E}\left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i^{(ns)^{-1}}(\beta, \rho)(y_i - \mu_i)\right] = 0. \tag{2.1.25}$$

Furthermore, it can be shown that (Sutradhar (2011, Eqn. (7.87))) $\hat{\beta}_{GQL}$ has asymptotic covariance given by

$$\mathrm{cov}(\hat{\beta}_{GQL}) = \lim_{n\to\infty} \left\{\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i^{\frac{1}{2}} \mathbf{C}_i^{(ns)}(\rho) \mathbf{A}_i^{\frac{1}{2}} \mathbf{X}_i\right\}^{-1}, \tag{2.1.26}$$

which provides bounded variances when fixed design matrices X_i are chosen properly.

Also, $\hat{\beta}_{GQL}$ has demonstrated to be more efficient than $\hat{\beta}_{QL}$, for example, where

$\hat{\beta}_{QL}$ is obtained by solving

$$\sum_{i=1}^{n} \mathbf{X}_i'(y_i - \mu_i) = 0, \tag{2.1.27}$$

[see (2.1.7)]. This can be examined by comparing the asymptotic variance of $\hat{\beta}_{QL}$, given by

$$\mathrm{V}(\hat{\beta}_{QL}) = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{A}_i\mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i'\Sigma_i^{(ns)^{-1}}\mathbf{X}_i \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{A}_i\mathbf{X}_i\right)^{-1}, \tag{2.1.28}$$

with the asymptotic variance of $\hat{\beta}_{GQL}$ given in (2.1.26).

2.1.2 Joint GQL Estimation for β and ρ

Since β is the regression effect and ρ is the correlation index parameter, for their joint estimation, following Sutradhar (2004), we exploit both first and second order responses. For this purpose, we define $u_i = (y_i', g_i')'$ where

$$\begin{aligned} y_i &= (y_{i1}, \ldots, y_{iT})' \\ g_i &= (y_{i1}y_{i2}, \ldots, y_{i,t-1}y_{it}, \ldots, y_{i,T-1}y_{iT})'. \end{aligned}$$

Note that we are considering only the lag 1 pairwise responses, which are appropriate

for any lag 1 correlation model. It is known from (2.1.9) that

$$\mathrm{E}(Y_i) \;=\; \mu_i,$$

and by denoting $\mathrm{E}(g_i)$ as

$$\mathrm{E}(g_i) \;=\; \eta_i = (\eta_{i,12}, \ldots, \eta_{i,t-1,t}, \ldots, \eta_{i,T-1,T})',$$

one can compute the general element $\eta_{i,t-1,t}$ as

$$\begin{aligned}
\eta_{i,t-1,t} &= \mathrm{E}(Y_{it}Y_{i,t-1}) \\
&= \mathrm{cov}(Y_{it}, Y_{i,t-1}) + \mathrm{E}(Y_{it})\mathrm{E}(Y_{i,t-1}) \\
&= \sigma_{i,t-1,t} + \pi_{i,t-1}\pi_{it} \\
&= \rho\Big(\sqrt{\frac{\sigma_{i,t-1,t-1}}{\sigma_{i,tt}}}\Big)\sqrt{\sigma_{i,tl}}\sqrt{\sigma_{i,t-1,t-1}} + \pi_{i,t-1}\pi_{it}, \qquad (2.1.29)
\end{aligned}$$

by (2.1.13), yielding

$$\eta_{i,t-1,t} = \rho\sigma_{i,t-1,t-1} + \pi_{i,t-1}\pi_{it}. \qquad (2.1.30)$$

Note that in general, for lag $|t-u|$ apart responses, one may write

$$\begin{aligned} \eta_{i,ut} &= \mathrm{E}(Y_{iu}Y_{it}) = \pi_{i,11}^{(t|u)} \\ &= \rho_{|u-t|}\sqrt{\sigma_{i,uu}\sigma_{i,tt}} + \pi_{iu}\pi_{it}. \end{aligned} \tag{2.1.31}$$

for time points u and t. The non-stationary AR(1) model, (2.1.31), reduces to

$$\begin{aligned} \eta_{i,ut} &= \rho^{t-u}\left(\sqrt{\frac{\sigma_{i,uu}}{\sigma_{i,tt}}}\right)\sqrt{\sigma_{i,uu}}\sqrt{\sigma_{i,tt}} + \pi_{iu}\pi_{it} \\ &= \rho_{t-u}\sqrt{\sigma_{i,uu}\sigma_{i,tt}} + \pi_{iu}\pi_{it}, \qquad \text{for } u < t, \end{aligned} \tag{2.1.32}$$

by way of (2.1.13).

Let

$$\begin{aligned} \mathrm{E}(U_i) &= [\mathrm{E}(Y_i'), \mathrm{E}(g_i')]' \\ &= (\mu_i', \eta_i')' \\ &\equiv \lambda_i, \end{aligned} \tag{2.1.33}$$

and

$$\mathrm{cov}(U_i) = \Omega_i, \tag{2.1.34}$$

where Ω_i is the $(2T-1)\times(2T-1)$ covariance matrix. Then, for $\theta=(\beta',\rho)'$, following Sutradhar (2004), we write the GQL estimating equation as

$$\sum_{i=1}^{n}\frac{\partial\lambda_i'}{\partial\theta}\Omega_i^{-1}(u_i-\lambda_i)=0 \quad :(p+1)\times 1. \tag{2.1.35}$$

Computation of Ω_i (Normality Based):

$$\begin{aligned}\Omega_i &= \text{cov}(u_i)=\text{cov}\begin{pmatrix}Y_i\\ g_i\end{pmatrix}\\ &= \begin{bmatrix}\text{cov}(Y_i) & \text{cov}(Y_i,g_i')\\ & \text{cov}(g_i)\end{bmatrix}\\ &= \begin{bmatrix}\Sigma_i & \mathrm{B}_i\\ & \Lambda_i\end{bmatrix}.\end{aligned} \tag{2.1.36}$$

To compute the covariance matrix Ω_i, we construct each component matrix in (2.1.36) as follows.

Computation of Σ_i:

Note that Σ_i is used for $\Sigma_i^{(ns)}(\beta,\rho)$ defined in (2.1.10). Thus, Σ_i in (2.1.36) is computed by (2.1.10).

Computation of B_i and Λ_i:

To compute B_i and Λ_i, we use a normal approximation. To do so, pretend that $y_i = (y_{i1}, \ldots, y_{iT})'$ follows a normal distribution but with correct mean μ_i and covariance matrix Σ_i as in (2.1.10), for the binary responses. It then follows that

$$\begin{aligned} \mathrm{E}(Y_{iu} - \pi_{iu})(Y_{it} - \pi_{it})(Y_{il} - \pi_{il}) &= 0 \qquad (2.1.37) \\ \mathrm{E}(Y_{iu} - \pi_{iu})(Y_{it} - \pi_{it})(Y_{il} - \pi_{il})(Y_{im} - \pi_{im}) &= \sigma_{i,ut}\sigma_{i,lm} + \sigma_{i,ul}\sigma_{i,tm} + \sigma_{i,um}\sigma_{i,tl}, \end{aligned}$$

where $\sigma_{i,ut}$, for example, is given by

$$\sigma_{i,ut} = \rho^{t-u}\sigma_{i,uu},$$

for $u < t$, by (2.1.4). After some algebra, one may then write the raw third order moments from (2.1.37) as,

$$\begin{aligned} \mathrm{E}(y_{iu}y_{it}y_{il}) &= \pi_{il}\mathrm{E}(y_{iu}y_{it}) + \pi_{it}\mathrm{E}(y_{iu}y_{il}) + \pi_{iu}\mathrm{E}(y_{it}y_{il}) - 2\pi_{it}\pi_{il}\pi_{iu} \\ &= \pi_{il}\eta_{i,ut} + \pi_{it}\eta_{i,ul} + \pi_{iu}\eta_{i,tl} - 2\pi_{it}\pi_{il}\pi_{iu} \\ &= \gamma_{i,utl}, \qquad (2.1.38) \end{aligned}$$

where, for example,

$$\begin{aligned} \eta_{i,ut} &= \mathrm{E}(Y_{iu}Y_{i,t}) \\ &= \sigma_{i,ut} + \pi_{i,t-1}\pi_{it} \\ &= \rho^{t-u}\sigma_{i,uu} + \pi_{i,t-1}\pi_{it}, \qquad \text{for } u < t, \end{aligned} \tag{2.1.39}$$

by (2.1.13). Now we compute,

$$\begin{aligned} B_i &= \mathrm{cov}(y_i, g_i') \\ &= \mathrm{cov}\left(\begin{pmatrix} y_{i1} \\ \vdots \\ y_{it} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad (y_{i1}y_{i2}, \ldots, y_{i,t-1}y_{it}, \ldots, y_{i,T-1}y_{iT}) \right), \end{aligned} \tag{2.1.40}$$

where in general

$$\begin{aligned} \mathrm{cov}(Y_{iu}, Y_{it}Y_{il}) &= \mathrm{E}(Y_{iu}Y_{i,t}Y_{il}) - \mathrm{E}(Y_{iu})\mathrm{E}(Y_{it}Y_{il}) \\ &= \gamma_{i,utl} - \pi_{iu}\eta_{i,tl}, \end{aligned} \tag{2.1.41}$$

where $\gamma_{i,utl}$ are computed by (2.1.38) for selected values of u, t, and l. Thus, B_i has been computed.

Likewise, one may also write the raw fourth order moments from (2.1.37) as,

$$\begin{aligned}
\mathrm{E}(y_{iu}y_{it}y_{il}y_{im}) &= \sigma_{i,ut}\sigma_{i,lm} + \sigma_{i,ul}\sigma_{i,tm} + \sigma_{i,um}\sigma_{i,tl} + 3\pi_{iu}\pi_{it}\pi_{il}\pi_{im} \\
&+ \pi_{im}\gamma_{i,utl} + \pi_{il}\gamma_{i,utm} + \pi_{it}\gamma_{i,ulm} + \pi_{iu}\gamma_{i,tlm} \\
&- (\pi_{it}\pi_{il}\eta_{i,um} + \pi_{iu}\pi_{im}\eta_{i,tl} + \pi_{iu}\pi_{im}\eta_{i,tl} + \pi_{iu}\pi_{it}\eta_{i,lm} \\
&+ \pi_{il}\pi_{im}\eta_{i,ut} + \pi_{it}\pi_{im}\eta_{i,ul}) \\
&= \phi_{i,utlm}. \qquad (2.1.42)
\end{aligned}$$

The formulas from (2.1.38) and (2.1.42) may be used to compute Λ_i. To be specific, for the computation of Λ_i, we first compute

$$\Lambda_i = \mathrm{cov}(g_i, g_i') = \mathrm{cov}\left(\begin{pmatrix} y_{i12} \\ \vdots \\ y_{i,u-1,u} \\ \vdots \\ y_{i,T-1,T} \end{pmatrix}, \quad (y_{i12}, \ldots, y_{i,l-1}y_{il}, \ldots, y_{i,T-1}y_{iT}) \right), \qquad (2.1.43)$$

where in general

$$\begin{aligned} \text{cov}(Y_{iu}Y_{i,t}, Y_{il}y_{im}) &= \text{E}(Y_{iu}Y_{it}Y_{il}Y_{im}) - \text{E}(Y_{iu}Y_{it})\text{E}(Y_{il}Y_{im}) \\ &= \phi_{i,ullm} - \eta_{i,ut}\eta_{i,lm}, \end{aligned} \tag{2.1.44}$$

where $\phi_{i,utlm}$ is computed by (2.1.42) for selected values of u, t, l, and m, and $\eta_{i,ut}$ is given by (2.1.39). Thus, Λ_i has been computed.

Computation for $\frac{\partial \lambda_i'}{\partial \theta}$:

To compute $\frac{\partial \lambda_i'}{\partial \theta}$, first recall that $\theta = (\beta', \rho)'$ and $\lambda_i = (\mu_i', \eta_i')'$. This implies that $\lambda_i = (\mu_i', \eta_i')'$. One may then write,

$$\frac{\partial \lambda_i'}{\partial \theta} = \left(\frac{\partial \mu_i'}{\partial \theta}, \frac{\partial \eta_i'}{\partial \theta} \right) \quad : (p+1) \times (2T-1), \tag{2.1.45}$$

such that

$$\frac{\partial \mu_i'}{\partial \theta} = \begin{pmatrix} \frac{\partial \mu_i'}{\partial \beta} \\ \\ \frac{\partial \mu_i'}{\partial \rho} \end{pmatrix} \quad : (p+1) \times T \tag{2.1.46}$$

and

$$\frac{\partial \eta_i'}{\partial \theta} = \begin{pmatrix} \frac{\partial \eta_i'}{\partial \beta} \\ \\ \frac{\partial \eta_i'}{\partial \rho} \end{pmatrix} \quad : (p+1) \times (T-1), \tag{2.1.47}$$

where

$$\begin{aligned} \frac{\partial \mu_i'}{\partial \beta} &= \mathbf{X}_i' \mathbf{A}_i \quad \text{as in (2.1.15).} \\ \frac{\partial \mu_i'}{\partial \rho} &= \mathbf{0} 1_T', \end{aligned}$$

and

$$\frac{\partial \eta_i'}{\partial \beta} = \left(\frac{\partial \eta_{i12}'}{\partial \beta}, \ldots, \frac{\partial \eta_{i,t-1,t}'}{\partial \beta}, \ldots, \frac{\partial \eta_{i,T-1,T}'}{\partial \beta} \right) \quad : p \times (T-1) \tag{2.1.48}$$

$$\frac{\partial \eta_i'}{\partial \rho_1} = (\sigma_{i,11}, \ldots, \sigma_{i,t-1,t-1}, \ldots, \sigma_{i,T-1,T-1}) \quad : 1 \times (T-1). \tag{2.1.49}$$

In (2.1.48),

$$\begin{aligned} \frac{\partial \eta_i'}{\partial \beta} &= \rho \frac{\partial \sigma_{i,t-1,t-1}}{\partial \beta} + \pi_{it} \frac{\partial \pi_{i,t-1}}{\partial \beta} + \pi_{i,t-1} \frac{\partial \pi_{it}}{\partial \beta} \\ &= \rho \Big[- \pi_{i,t-1} \frac{\partial \pi_{i,t-1}}{\partial \beta} + (1 - \pi_{i,t-1}) \frac{\partial \pi_{i,t-1}}{\partial \beta} \Big] \\ &+ \pi_{it} \frac{\partial \pi_{i,t-1}}{\partial \beta} + \pi_{i,t-1} \frac{\partial \pi_{i,t}}{\partial \beta} \\ &= \pi_{i,t-1} \frac{\partial \pi_{i,t}}{\partial \beta} + \frac{\partial \pi_{i,t-1}}{\partial \beta} \Big[\pi_{it} + \rho(1 - 2\pi_{i,t-1}) \Big], \end{aligned} \tag{2.1.50}$$

with

$$\frac{\partial \pi_{i,t}}{\partial \beta} = \pi_{it}(1-\pi_{it})x_{it}.$$

Thus, $\frac{\partial \lambda_i'}{\partial \theta}$ has been computed.

2.2 'Working' Correlations Based GEE Approaches

When the correlation structure for the repeated binary data is unknown, Liang and Zeger (1986) have used a 'working' correlations based approach for the estimation of the regression parameter β. Because there is no guidance for the selection of correlation structure, some authors, such as Prentice (1988, Section 4) and Lipsitz *et al.* (1991, Table 1, p. 158) used an equi-correlations (EQC) structure which may not be appropriate in many situations. This raises concerns about the correlation model misspecification effects. For this reason, we include this structure in our empirical study in Chapter 3 to examine its performance when repeated binary data are in fact generated following the most practical $AR(1)$ model defined in (2.1.2). In the next section, we provide the EQC based GEE, including the estimation of the equi-correlation parameter.

2.2.1 'Working' Equi-Correlations (WEQC) Based Approach for Regression Effects

Similar to Section 2.1.1, we may write the equi-correlations based estimating equation as

$$\sum_{i=1}^{n} \frac{\partial \mu_i'}{\partial \beta} \left[\mathbf{A}_i^{\frac{1}{2}} R_i(\alpha) \mathbf{A}_i^{\frac{1}{2}} \right]^{-1} (y_i - \mu_i) = 0, \tag{2.2.1}$$

where

$$\begin{aligned} y_i &= (y_{i1}, \ldots, y_{iT})' \\ \mu_i &= \pi_i = (\pi_{i1}, \ldots, \pi_{iT})', \end{aligned} \tag{2.2.2}$$

and

$$\mathbf{A}_i = \text{diag}[\pi_{i1}(1-\pi_{i1}), \ldots, \pi_{it}(1-\pi_{it}), \ldots, \pi_{iT}(1-\pi_{iT})], \tag{2.2.3}$$

are as in (2.1.8), but $R_i(\alpha)$ is a 'working' EQC matrix given by

$$\begin{aligned} R_i(\alpha) &= \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ & 1 & \cdots & \alpha \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix} \quad : \quad T \times T \\ &= R(\alpha) \qquad \text{say.} \end{aligned} \tag{2.2.4}$$

Note that we have used ρ in (2.1.8) as a correlation index parameter, assuming that the data follows an $AR(1)$ structure as in (2.1.12), but have used α in (2.2.4) to indicate that the 'working' correlation parameters are different than the true index parameters. The computation of $\frac{\partial \mu_i'}{\partial \beta}$ is found in the same way as in Section 2.1.1.1.

For known α, we can solve the estimating equation given in (2.2.1) for β iteratively by using the so-called Newton-Raphson formula given by

$$\hat{\beta}_{r+1} = \hat{\beta}_r + \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i R(\alpha)^{-1} \mathbf{A}_i \mathbf{X}_i\right]_r^{-1} \times \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i R(\alpha)^{-1} (y_i - \mu_i)\right]_r \qquad (2.2.5)$$

As in the GQL approach, $[\ \ ]_r$ denotes that the expression contained within the brackets is evaluated at $\beta = \hat{\beta}(r)$; that is, the r^{th} iterative value for β.

2.2.1.1 Estimation of α

Note that by using Table 1.1 (for joint probabilities) we can write the general correlation between y_{iu} and y_{it} as

$$\begin{aligned}
\text{corr}(Y_{iu}, Y_{it}) &= \frac{\text{cov}(Y_{iu}, Y_{it})}{\sqrt{\text{var}(Y_{iu})\text{var}(Y_{it})}} \\
&= \frac{E(Y_{iu}Y_{it}) - E(Y_{iu})E(Y_{iu})}{\sqrt{[EY_{iu}^2 - (EY_{iu})^2][EY_{it}^2 - (EY_{it})^2]}} \\
&= \frac{\Pr(Y_{iu} = 1, Y_{it} = 1) - \pi_{iu}\pi_{it}}{\sqrt{\pi_{iu}(1-\pi_{iu})\pi_{it}(1-\pi_{it})}} \\
&= \frac{\pi_{i,11}^{(t|u)} - \pi_{iu}\pi_{it}}{\sqrt{\pi_{iu}(1-\pi_{iu})\pi_{it}(1-\pi_{it})}} \\
&= \alpha_{i,ut}. \qquad (2.2.6)
\end{aligned}$$

As pointed out earlier, some authors such as Lipsitz et al. (1991, Section 4) and Prentice (1988, Section 4.1), did not model the correlations among the repeated binary responses. Instead, they have assumed constant correlation over time. Based on this assumption first for any $u < t$, by using $\alpha_{i,ut} = \alpha_{ut}$ for all i, one may estimate α_{ut} as

$$\begin{aligned}\hat{\alpha}_{ut} &= \frac{\left[\sum_{i=1}^{n} y_{iu}y_{it} - \frac{\sum_{i=1}^{n} y_{iu}\sum_{i=1}^{n} y_{it}}{n}\right]}{\sqrt{\left[\sum_{i=1}^{n} y_{iu}^2 - \frac{(\sum_{i=1}^{n} y_{iu})^2}{n}\right]\left[\sum_{i=1}^{n} y_{it}^2 - \frac{(\sum_{i=1}^{n} y_{it})^2}{n}\right]}} \\ &= \frac{\frac{n_{11}^{(t|u)}}{n} - \frac{n_u n_t}{n^2}}{\sqrt{\left(\frac{n_u}{n}\left(1-\frac{n_u}{n}\right)\right)\left(\frac{n_t}{n}\left(1-\frac{n_t}{n}\right)\right)}} \\ &= \frac{p_{11}^{(t|u)} - p_u p_t}{p_u(1-p_u)p_t(1-p_t)}. \end{aligned} \tag{2.2.7}$$

Furthermore, when it is assumed that y_{iu} and y_{it} have equi-correlation over time, one may exploit $\hat{\alpha}_{ut}$ from (2.2.7) and estimate the equi-correlation α, say, by

$$\hat{\alpha} = \frac{2}{T(T-1)} \sum_{t=u+1}^{T} \sum_{u=1}^{T-1} \hat{\alpha}_{ut}. \tag{2.2.8}$$

Note that (2.2.7) and (2.2.8) are applicable only when $R_i(\alpha) \equiv R(\alpha)$ for all $i = 1, \ldots, n$. In such cases, $\hat{\beta}$ is obtained by

$$\hat{\beta}_{r+1} = \hat{\beta}_r + \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i R^{-1}(\alpha) \mathbf{A}_i \mathbf{X}_i\right]_r^{-1} \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i R^{-1}(\alpha)(y_i - \mu_i)\right]_r \tag{2.2.9}$$

Furthermore, note that in the non-stationary case, covariates are time dependent. In addition, the covariates may change from individual to individual. This raises questions with regard to the summation over i in (2.2.7), and over time t in (2.2.8), to obtain an estimate of correlation α.

Alternative Estimation for α

An alternative approach may be considered to estimate the equi-correlation α as

$$\hat{\alpha}_{ut} = \frac{\sum_{i=1}^{n} \tilde{y}_{iu}\tilde{y}_{it}/n}{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}_{it}^2/nT}, \tag{2.2.10}$$

for $u < t$ where

$$\tilde{y}_{it} = \frac{y_{it} - \pi_{it}}{\sqrt{\pi_{it}(1-\pi_{it})}}, \tag{2.2.11}$$

and then calculate $\hat{\alpha}$ by using (2.2.8). The difference between the formulas denoted in (2.2.7) and (2.2.10) is that (2.2.7) exploits the sample mean, whereas in (2.2.10) we exploit the estimated population means.

2.2.2 'Working' Odds Ratios (WOR) Based GEE Approach for Regression Effects

When the covariance matrix Σ_i is computed based on odds ratios, we write the GEE for β as

$$\sum_{i=1}^{n} \frac{\partial \mu_i'}{\partial \beta} \Sigma_i^{-1}(\tau)(y_i - \mu_i) = 0, \tag{2.2.12}$$

where $y_i = (y_{i1}, \ldots, y_{iT})'$ and $\mu_i = \pi_i = (\pi_{i1}, \ldots, \pi_{iT})'$, as before, and $\Sigma_i(\tau)$ is the $T \times T$ covariance matrix with its elements based on the odds ratios.

2.2.2.1 Construction of $\Sigma_i(\tau)$:

To compute the elements of this matrix, we first write the general covariance as

$$\text{cov}(y_{iu}, y_{it}) = \pi_{i,11}^{(t|u)}(\tau) - \pi_{iu}\pi_{it}, \tag{2.2.13}$$

where

$$\pi_{i,11}^{(t|u)}(\tau) = \text{E}(Y_{iu}Y_{it}), \tag{2.2.14}$$

to be computed by taking advantage of the odds ratios. For this purpose we refer to the joint probabilities in Table 1.1 in Chapter 1, and write the odds ratio for the i^{th} individual at time t conditional on the individual's response at time u, as

$$\begin{aligned} \tau_{iut} &= \frac{\pi_{i,11}^{(t|u)}\pi_{i,22}^{(t|u)}}{\pi_{i,12}^{(t|u)}\pi_{i,21}^{(t|u)}} \\ &= \frac{\pi_{i,11}^{(t|u)}(1-\pi_{it}-\pi_{i,12}^{(t|u)})}{\pi_{i,12}^{(t|u)}\pi_{i,21}^{(t|u)}} \\ &= \frac{\pi_{i,11}^{(t|u)}(1-\pi_{it}-\pi_{iu}+\pi_{i,11}^{(t|u)})}{(\pi_{iu}-\pi_{i,11}^{(t|u)})(\pi_{it}-\pi_{i,11}^{(t|u)})}. \end{aligned} \tag{2.2.15}$$

The above equation can be rewritten as

$$\begin{aligned} \pi_{i,11}^{(t|u)}\left(1-\pi_{it}-\pi_{iu}+\pi_{i,11}^{(t|u)}\right) &= \tau_{i,ut}\left[(\pi_{iu}-\pi_{i,11}^{(t|u)})(\pi_{it}-\pi_{i,11}^{(t|u)})\right] \\ &= \tau_{i,ut}\left[\pi_{i,11}^{2(t|u)} - (\pi_{iu}+\pi_{it})\pi_{i,11}^{(t|u)} + \pi_{iu}\pi_{it}\right], \end{aligned} \tag{2.2.16}$$

yielding

$$\pi_{i,11}^{2(t|u)} + \pi_{i,11}^{(t|u)}(1 - \pi_{it} - \pi_{iu}) = \tau_{i,ut}\left[\pi_{i,11}^{2(t|u)} - (\pi_{iu} + \pi_{it})\pi_{i,11}^{(t|u)} + \pi_{iu}\pi_{it}\right]. \quad (2.2.17)$$

This in turn can be re-expressed as

$$(1 - \tau_{i,ut})\pi_{i,11}^{2(t|u)} + \left(1 - \pi_{it} - \pi_{iu} + (\pi_{iu} + \pi_{it})\tau_{i,ut}\right)\pi_{i,11}^{(t|u)} - \pi_{iu}\pi_{it}\tau_{i,ut} = 0. \quad (2.2.18)$$

This further reduces to a general quadratic equation that can be written in the form

$$a\pi_{i,11}^{2(t|u)} + b\pi_{i,11}^{(t|u)} + c = 0, \quad (2.2.19)$$

where

$$\begin{aligned} a &= 1 - \tau_{i,ut} \\ b &= 1 - \pi_{it} - \pi_{iu} + (\pi_{iu} + \pi_{it})\tau_{i,ut} \\ c &= -\pi_{iu}\pi_{it}\tau_{i,ut}. \end{aligned} \quad (2.2.20)$$

The solution of (2.2.19) for $\pi_{i,11}^{(t|u)}$ has the form

$$\pi_{i,11}^{(t|u)} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}. \quad (2.2.21)$$

The solution to (2.2.21), after some algebra, is

$$\pi_{i,11}^{(t|u)}(\tau) = \begin{cases} \frac{[1-(1-\tau_{i,ut})(\pi_{iu}+\pi_{it})]-\sqrt{[1-(1-\tau_{i,ut})(\pi_{iu}+\pi_{it})]^2-4\tau_{i,ut}(\tau_{i,ut}-1)\pi_{iu}\pi_{it}}}{2(\tau_{i,ut}-1)} & \text{if } \tau_{i,ut} \neq 1 \\ \pi_{iu}\pi_{it} & \text{if } \tau_{i,ut} = 1 \end{cases} \tag{2.2.22}$$

[Lipsitz et al., (1991, Eqn. (6), p.155), Yi and Cook(2001, Eqn. (3), p.1072)], which will always lie in $[0, 1]$. Finally,

$$\begin{aligned} \text{cov}(y_{iu}, y_{it}) &= \sigma_{i,ut}(\tau) \\ &= \pi_{i,11}^{(t|u)} - \pi_{iu}\pi_{it} \qquad u, t = 1, \ldots, T. \\ &= \begin{cases} \pi_{it}(1-\pi_{it}) & \text{for } t = u, \\ \pi_{i,11}^{(t|u)} - \pi_{iu}\pi_{it} & \text{for } u < t, \end{cases} \end{aligned} \tag{2.2.23}$$

where $\pi_{i,11}^{(t|u)}$ is a function of τ, as given in (2.2.22). One may then construct the covariance matrix as

$$\Sigma_i(\tau) = \Big(\sigma_{i,ut}(\tau)\Big). \tag{2.2.24}$$

For the computation of $\frac{\partial \mu_i'}{\partial \beta}$, refer to (2.1.15). For convenience, we rewrite

$$\frac{\partial \mu_i'}{\partial \beta} = \mathbf{X}_i' \mathbf{A}_i, \tag{2.2.25}$$

where

$$\mathbf{X}_i' = [x_{i1}, \ldots, x_{it}, \ldots, x_{iT}] \quad : \quad p \times T, \tag{2.2.26}$$

and A_i is the $T \times T$ diagonal matrix as in (2.1.11).

Solving the GEE (2.2.12) for β:

For known τ, we may then solve the estimating equation in (2.2.12) for β iteratively. Using (2.2.26) and (2.2.24), one may use the Newton-Raphson formula to find an estimate of $\hat{\beta}$ as

$$\hat{\beta}_{r+1} = \hat{\beta}_r + \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i(\tau)^{-1} \mathbf{A}_i \mathbf{X}_i\right]_r^{-1} \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i(\tau)^{-1} (y_i - \mu_i)\right]_r \tag{2.2.27}$$

2.2.2.2 Estimation of τ

Note that for the construction of the covariance elements $\sigma_{i,ut}(\tau)$, we need to first compute the joint probabilities $\pi_{i,11}^{(t|u)}(\tau)$ by (2.2.22), which in turn requires the estimates of $\tau_{i,ut}$ for all i, u and t. In general, for all i, the computation of this odds ratio $\tau_{i,ut}$ is not possible unless one can use a model for all individuals. For this reason, some authors, such as Williamson et al. (1995, Eqn. (3)) [see also Cook and Yi (2002, Eqn. (1), p. 1072)] have used linear regression modeling as

$$\log \tau_{i,ut} = \Delta + \Delta_u + \Delta_t + \Delta_{ut} + \xi x'_{ic}, \tag{2.2.28}$$

where Δ is an intercept parameter, Δ_u and Δ_t are the marginal effects at time u and t, and Δ_{ut} is the interaction. Further, x'_{ic} is a suitable vector of covariates responsible for the correlation of y_{iu} and y_{it}, and ξ is the effect of x_{ic}. This extra regression model for association parameters, however, appears to be arbitrary.

Constant odds ratio (τ_{ut}) estimation:

Note that all individuals having constant odds ratio is equivalent to considering that the joint probabilities in Table 1.1 are free from i ($i = 1, \ldots, n$). Some authors, such as Lipsitz *et al.* (1991) have used a constant odds ratio assumption and estimated this parameter as follows. Suppose that for given u and t, all individuals ($i = 1, \ldots, n$) have the constant odds ratio τ_{ut}.

Recall Table 1.1 from Chapter 1, which contains the joint and marginal probabilities for all possible transitions from time u to t. Thus, for τ_{ut} estimation, one uses the probabilities as in Table 2.1 below.

Table 2.1: Joint probabilities in contingency table form.

		Time t		
Time u		1 (y=1)	2 (y=0)	Total
	1 (y=1)	$\pi_{11}^{(t\|u)}$	$\pi_{12}^{(t\|u)}$	$\pi_{\cdot u}$
	2 (y=0)	$\pi_{21}^{(t\|u)}$	$\pi_{22}^{(t\|u)}$	$1-\pi_{\cdot u}$
Total		$\pi_{\cdot t}$	$1-\pi_{\cdot t}$	1

When individuals are grouped by way of Table 2.1, one obtains the contingency table in Table 1.2, which we reiterate in Table 2.2, for convenience.

Table 2.2: Contingency table for general time points.

		Time t		
Time u		1 (y=1)	2 (y=0)	Total
	1 (y=1)	$n_{11}^{(t\|u)}$	$n_{12}^{(t\|u)}$	n_u
	2 (y=0)	$n_{21}^{(t\|u)}$	$n_{22}^{(t\|u)}$	$n-n_u$
Total		n_t	$n-n_t$	n

Using Tables 2.1 and 2.2, one then obtains the moment estimates for the joint probabilities as

$$\hat{\pi}_{11}^{(t|u)} = \frac{n_{11}^{(t|u)}}{n} \qquad \hat{\pi}_{21}^{(t|u)} = \frac{n_{21}^{(t|u)}}{n} \tag{2.2.29}$$

$$\hat{\pi}_{12}^{(t|u)} = \frac{n_{12}^{(t|u)}}{n} \qquad \hat{\pi}_{22}^{(t|u)} = \frac{n_{22}^{(t|u)}}{n}. \tag{2.2.30}$$

By using (2.2.29)-(2.2.30), it follows from (2.2.15) that the constant odds ratio may be estimated as

$$\begin{aligned}\hat{\tau}_{ut} &= \frac{\hat{\pi}_{11}^{(t|u)}\hat{\pi}_{22}^{(t|u)}}{\hat{\pi}_{12}^{(t|u)}\hat{\pi}_{21}^{(t|u)}} \\ &= \frac{n_{11}^{(t|u)}n_{22}^{(t|u)}}{n_{12}^{(t|u)}n_{21}^{(t|u)}}.\end{aligned} \tag{2.2.31}$$

Note that when the longitudinal binary data $\{y_{it}\}$ is available, the counts in Table 2.2 may be expressed as follows:

$$\begin{aligned}n_{11}^{(t|u)} &= \sum_{i=1}^{n} y_{iu}y_{it} \\ n_{12}^{(t|u)} &= \sum_{i=1}^{n} y_{iu}(1-y_{it}) \\ n_{21}^{(t|u)} &= \sum_{i=1}^{n} y_{it}(1-y_{iu}) \\ n_{22}^{(t|u)} &= \sum_{i=1}^{n}(1-y_{iu})(1-y_{it}),\end{aligned} \tag{2.2.32}$$

where

$$\begin{aligned}n_u &= \sum_{i=1}^{n} y_{iu} \\ n_t &= \sum_{i=1}^{n} y_{it},\end{aligned} \tag{2.2.33}$$

and

$$\begin{aligned} n_{12}^{(t|u)} &= n_u - n_{11}^{(t|u)} \\ &= \sum_{i=1}^{n} y_{iu} - \sum_{i=1}^{n} y_{iu}y_{it} \\ &= \sum_{i=1}^{n} y_{iu}(1 - y_{it}) \end{aligned} \tag{2.2.34}$$

$$\begin{aligned} n_{21}^{(t|u)} &= n_t - n_{11}^{(t|u)} \\ &= \sum_{i=1}^{n} y_{it} - \sum_{i=1}^{n} y_{iu}y_{it} \\ &= \sum_{i=1}^{n} y_{it}(1 - y_{iu}) \end{aligned} \tag{2.2.35}$$

$$\begin{aligned} n_{22}^{(t|u)} &= n - n_t - n_{12}^{(t|u)} \\ &= n - \sum_{i=1}^{n} y_{it} - \Big(\sum_{i=1}^{n} y_{iu} - \sum_{i=1}^{n} y_{iu}y_{it}\Big) \\ &= n - \sum_{i=1}^{n} y_{it} - \sum_{i=1}^{n} y_{iu} + \sum_{i=1}^{n} y_{iu}y_{it} \\ &= \sum_{i=1}^{n} \Big(1 - y_{it} - y_{iu} + y_{it}y_{iu}\Big) \\ &= \sum_{i=1}^{n} (1 - y_{iu})(1 - y_{it}). \end{aligned} \tag{2.2.36}$$

Constant odds ratio estimation for any two time points:

By assuming equal odds ratios at any two time points (Lipsitz *et al.*(1991), Table 1, p. 158), i.e., $\tau_{ut} = \tau$ for all $u < t$, one can estimate τ by

$$\hat{\tau} = \frac{2}{T(T-1)} \sum_{t=u+1}^{T} \sum_{u=1}^{T-1} \hat{\tau}_{ut}. \tag{2.2.37}$$

Chapter 3

Relative Performance of Estimation Approaches: A Simulation Study

3.1 A Simulation Study for Stationary Correlation Model

In this section, a special case of the non-stationary correlation model, discussed in Section 2.1, is explored. The purpose of this section is to generate correlated data

adhering to the following AR(1) model

$$Pr[y_{i1} = 1] = \frac{e^{x_{i\cdot}'\beta}}{1 + e^{x_{i\cdot}'\beta}} = \pi_{i\cdot}$$

$$Pr[Y_{it}|y_{i,t-1}] = \pi_{i\cdot} + \rho(y_{i,t-1} - \pi_{i\cdot}), \qquad \text{for } t = 2, \ldots, T, \tag{3.1.1}$$

and estimate β by the three different approaches, namely, GQL discussed in Section 2.1, WEQC (working equi-correlation) based GEE which is discussed in Section 2.2.1, and WOR (working odds ratio) based GEE discussed in Section 2.2.2.

Note that in practice, the correlation model is unknown. Because the GQL approach assumes a large class of correlations, it is practical to assume that an unknown correlation structure may well be taken care of by using such a class. However, some 'working' approaches do not even care for such a class and furthermore, without any diagnostics, this approach uses a certain 'working' correlation model to develop estimating equations. Since these are arbitrary, in this thesis, we examine their performance by generating data from an AR(1) model, for the purpose of checking the correlation mis-specification effect on estimation.

3.1.1 Simulation Design and Data Generation

For this simulation study, we consider the parameters as follows:

$$\begin{aligned} n &= 100 \\ T &= 4 \\ p &= 1 \\ \rho &= 0.0, 0.3, 0.5, 0.7. \end{aligned} \tag{3.1.2}$$

For the scalar ($p = 1$) covariate, we choose the stationary design (D_s) covariates as

$$x_{it} = x_{i\cdot} = \begin{cases} -1 & \text{for } i = 1, \ldots, n/4 \\ 0 & \text{for } i = (n/4) + 1, \ldots, 3n/4 \\ 1 & \text{for } i = (3n/4) + 1, \ldots, n, \end{cases} \tag{3.1.3}$$

for all $t = 1, \ldots, 4$. Note that x_{it} is stationary, and does not depend on time t. For the covariate effect, we choose

$$\beta_1 \equiv \beta = 0.0, 0.5, \text{and } 1.0. \tag{3.1.4}$$

Next we generate the repeated binary responses. In a given simulation we generate

the first response y_{i1} for all $i = 1, \ldots, n$, following

$$y_{i1} \sim Bin(\pi_{i\cdot}), \tag{3.1.5}$$

with

$$\pi_{i\cdot} = \frac{e^{x_{i\cdot}\beta}}{1 + e^{x_{i\cdot}\beta}}.$$

For a given value of ρ, we then use (3.1.1) to generate y_{i2} using

$$y_{i2} \sim Bin(\lambda_{i2} = \pi_{i\cdot} + \rho(y_{i1} - \pi_{i\cdot})). \tag{3.1.6}$$

Once y_{i2} is generated, we use it to generate y_{i3} as

$$y_{i3} \sim Bin(\lambda_{i3} = \pi_{i\cdot} + \rho(y_{i2} - \pi_{i\cdot})). \tag{3.1.7}$$

We follow this above procedure and proceed to generate $y_{i1}, \ldots, y_{iT}$.

3.1.2 GQL Versus WOR and WEQC Approaches for β Estimation

In a given simulation, using the responses $\{y_{it}, t = 1, \ldots, 4; i = 1, \ldots, 100\}$ generated as in Section 3.1.1, and $x_{i\cdot}$ as in (3.1.3), we compute the GQL, WOR, and WEQC estimates of β by solving the respective iterative equations as follows:

GQL Estimation:

To obtain the GQL estimate, we use the stationary version of the iterative equation given in (2.1.17), namely,

$$\hat{\beta}_{r+1} = \hat{\beta}_r + \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i^{-1}(\beta, \rho) \mathbf{A}_i \mathbf{X}_i\right]_r^{-1} \times \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i^{-1}(\beta, \rho)(y_i - \mu_i)\right]_r \quad (3.1.8)$$

where in this stationary case,

$$\begin{aligned} X_i' &= (x_{i\cdot}, \ldots, x_{i\cdot}, \ldots, x_{i\cdot}) \qquad : p \times T \\ &= x_{i\cdot} \otimes \mathbf{1}_T', \end{aligned} \quad (3.1.9)$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{1}_T'$ is the $1 \times T$ unit vector. In addition,

$$\begin{aligned} y_i &= \big(y_{i1}, \ldots, y_{iT}\big)' \\ A_i &= \pi_{i\cdot}\big(1 - \pi_{i\cdot}\big)\mathbf{I}_T \\ \Sigma_i^{(s)} &= A_i^{\frac{1}{2}} C_i^{(s)} A_i^{\frac{1}{2}}, \end{aligned} \tag{3.1.10}$$

with

$$C_i^{(s)} = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{T-1} \\ & 1 & \cdots & \rho_{T-2} \\ & & \ddots & 1 \end{pmatrix}. \tag{3.1.11}$$

Note that in order to use (3.1.1), we must first compute the lag correlations ρ_1, ρ_2 and ρ_3 by using the formula for $\hat{\rho}_l$ $(l = 1, 2, 3)$ obtained from (2.1.23) under the present stationary case. More specifically, we compute

$$\hat{\rho}_l = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T-l} \tilde{y}_{it} \tilde{y}_{i,t+l}}{\sum_{i=1}^{n} \sum_{t=1}^{T-l} \tilde{y}_{it}^2}, \tag{3.1.12}$$

where

$$\tilde{y}_{it} = \frac{(y_{it} - \pi_{i\cdot})}{\sqrt{\pi_{i\cdot}(1 - \pi_{i\cdot})}}. \tag{3.1.13}$$

Note that $v(\hat{\beta})$ has the form

$$v(\hat{\beta}) = \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma \Sigma \Sigma_{\Sigma_i^{-1} \mathbf{A}_i \mathbf{X}_i^{-1}}, \right. \quad (3.1.14)$$

which is a stationary form of (2.1.26). **WOR Estimation:**

Here we use the stationary version of the WOR based iterative equation given in (2.2.27). More specifically, the X_i' and A_i matrices will now have the forms:

$$\begin{aligned} X_i' &= \big(x_{i\cdot}, \ldots, x_{i\cdot}, \ldots, x_{i\cdot}\big) \qquad : p \times T \\ &= x_{i\cdot} \otimes \mathbf{1}_T', \end{aligned} \quad (3.1.15)$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{1}_T'$ is the $1 \times T$ unit vector, and

$$A_i = \pi_{i\cdot}\big(1 - \pi_{i\cdot}\big)\mathbf{I}_T,$$

as in the stationary GQL case. In this stationary case, to compute the elements of

$$\Sigma_i(\tau) = \Big(\sigma_{i,ut}(\tau)\Big),$$

we use the stationary version of (2.2.24). That is,

$$\sigma_{i,ut}(\tau) = \begin{cases} \pi_{i\cdot}(1-\pi_{i\cdot}) & \text{for } t = u, \\ \pi_{i,11}^{(t|u)} - \pi_{i\cdot}^2 & \text{for } u < t, \end{cases}$$

where $\pi_{i,11}^{(t|u)}$ now has the form

$$\pi_{i,11}^{(t|u)}(\tau) = \begin{cases} \frac{[1-2(1-\tau)\pi_{i\cdot}] - \sqrt{[1-2(1-\tau)\pi_{i\cdot}]^2 - 4\tau(\tau-1)\pi_{i\cdot}^2}}{2(\tau-1)} & \text{if } \tau \neq 1 \\ \pi_{i\cdot}^2 & \text{if } \tau = 1. \end{cases} \tag{3.1.16}$$

WEQC Estimation:

The stationary version of the WEQC based estimates are obtained by using the iterative equation (2.2.9), with

$$\begin{aligned} X_i' &= (x_{i\cdot}, \ldots, x_{i\cdot}, \ldots, x_{i\cdot}) \qquad : p \times T \\ &= x_{i\cdot} \otimes \mathbf{1}_T', \end{aligned}$$

where $\otimes$ denotes the Kronecker product, and $\mathbf{1}_T'$ is the $1 \times T$ unit vector, and

$$A_i = \pi_{i\cdot}(1-\pi_{i\cdot})\mathbf{I}_T.$$

The above computations for β based on all three approaches, namely GQL, WOR and WEQC, are repeated 500 times. The average of these 500 values under a given method is computed to obtain the final estimate. These estimates are denoted by $\hat{\beta}_{GQL}$, $\hat{\beta}_{WOR}$, and $\hat{\beta}_{WEQC}$, respectively. We also compute the standard error of these 500 values. The estimates and standard errors are referred to as SM (simulated mean) and SSE (simulated standard error). In addition, we also compute the percentage RB (relative bias) and the MSE (mean squared error) of these estimates. For example, for the GQL estimates, the percentage RB and MSE have the formulas

$$\mathrm{RB}(\hat{\beta}_{GQL}) = \frac{|\hat{\beta}_{GQL} - \beta|}{\mathrm{SSE}(\hat{\beta}_{GQL})} \times 100, \tag{3.1.17}$$

and

$$\mathrm{MSE}(\hat{\beta}_{GQL}) = \left(\hat{\beta}_{GQL} - \beta\right)^2 + \mathrm{SSE}^2, \tag{3.1.18}$$

respectively. The simulation results for the stationary GQL, WOR, and WEQC estimates are given in Tables 3.1, 3.2, and 3.3, respectively.

For the WEQC approach, we also use an alternative estimation method for the 'working' equi-correlation parameter α. In this alternative approach, we use the

Table 3.1: GQL estimate ($\hat{\beta}_{GQL}$) for the regression effect along with its standard error ($\sigma_{\hat{\beta}_{GQL}}$), MSE, and RB, for stationary design D_S with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{GQL}$	$\sigma_{\hat{\beta}_{GQL}}$	RB(%)	$\hat{\rho}_1$	$\hat{\rho}_2$	$\hat{\rho}_3$	MSE
0.0	0.0	-0.021	0.336	6.128	0.012	0.008	0.005	0.113
	0.3	-0.009	0.438	2.078	0.315	0.111	0.040	0.192
	0.5	-0.021	0.680	3.506	0.524	0.283	0.156	0.463
	0.7	-0.084	1.333	6.302	0.732	0.537	0.395	1.784
0.5	0.0	0.514	0.354	3.880	0.011	0.009	0.009	0.126
	0.3	0.522	0.480	4.588	0.322	0.117	0.042	0.231
	0.5	0.531	0.575	5.367	0.523	0.279	0.148	0.332
	0.7	0.537	1.504	2.451	0.730	0.533	0.390	2.263
1.0	0.0	1.033	0.394	8.319	0.013	0.009	0.004	0.156
	0.3	1.043	0.521	8.170	0.318	0.109	0.035	0.273
	0.5	1.043	0.659	6.618	0.521	0.278	0.150	0.436
	0.7	1.230	1.935	12.114	0.726	0.528	0.386	3.797

Table 3.2: WOR based GEE estimate ($\hat{\beta}_{WOR}$) for regression effect along with its standard error ($\sigma_{\hat{\beta}_{WOR}}$), MSE, and RB, for stationary design D_S with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{WOR}$	$\sigma_{\hat{\beta}_{WOR}}$	RB (%)	$\hat{\tau}$	$\sigma_{\hat{\tau}}$	MSE
0.0	0.0	-0.014	0.298	4.500	1.094	0.191	0.089
	0.3	-0.012	0.541	2.300	2.790	0.747	0.293
	0.5	-0.022	0.832	2.700	7.151	2.585	0.693
	0.7	-0.032	1.293	2.500	28.418	15.430	1.673
0.5	0.0	0.509	0.354	2.600	1.247	0.241	0.125
	0.3	0.509	0.607	1.600	3.097	0.859	0.369
	0.5	0.497	0.880	0.300	7.789	2.739	0.774
	0.7	0.382	1.319	8.900	30.946	15.119	1.754
1.0	0.0	1.034	0.474	7.300	1.034	0.474	0.226
	0.3	1.024	0.726	3.400	4.026	1.155	0.528
	0.5	0.954	1.057	4.400	10.084	4.043	1.119
	0.7	0.787	1.455	14.600	39.436	21.795	2.162

Table 3.3: WEQC based GEE estimate ($\hat{\beta}_{WEQC}$) for regression effect along with its standard error ($\sigma_{\hat{\beta}_{WOR}}$), MSE, and RB, for stationary design D_S with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{WEQC}$	$\sigma_{\hat{\beta}_{WEQC}}$	RB(%)	$\hat{\alpha}$	MSE
0.0	0.0	0.107	0.251	42.700	0.001	0.074
	0.3	0.113	0.294	38.500	0.192	0.099
	0.5	0.105	0.318	33.000	0.363	0.112
	0.7	0.079	0.335	23.500	0.575	0.118
0.5	0.0	0.632	0.283	46.500	0.033	0.098
	0.3	0.637	0.331	41.312	0.217	0.128
	0.5	0.630	0.352	36.900	0.381	0.141
	0.7	0.602	0.364	28.000	0.589	0.143
1.0	0.0	1.171	0.364	46.900	0.111	0.162
	0.3	1.181	0.427	42.300	0.278	0.215
	0.5	1.168	0.445	37.900	0.432	0.226
	0.7	1.137	0.439	31.200	0.624	0.211

stationary version of the estimating formula (2.2.10). That is,

$$\hat{\alpha} = \frac{1}{T(T-1)} \sum \hat{\alpha}^*_{ut}, \tag{3.1.19}$$

where

$$\hat{\alpha}^*_{ut} = \frac{\sum_{i=1}^{n} \tilde{y}_{iu}\tilde{y}_{it}/n}{\sum_{i=1}^{n}\sum_{t=1}^{T} \tilde{y}^2_{it}/nT}, \tag{3.1.20}$$

for $u < t$, where

$$\tilde{y}_{it} = \frac{y_{it} - \pi_{i\cdot}}{\sqrt{\pi_{i\cdot}(1-\pi_{i\cdot})}}, \tag{3.1.21}$$

with

$$\pi_{i\cdot} = \frac{e^{x_{i\cdot}\beta}}{1 + e^{x_{i\cdot}\beta}},$$

whereas in the former WEQC approach, the estimation of α by (2.2.7) was free from a β estimate. The simulation results for this alternative WEQC (AWEQC) estimates are given in Table 3.4.

Table 3.4: AWEQC based GEE estimate ($\hat{\beta}_{AWEQC}$) for regression effect along with its standard error ($\sigma_{\hat{\beta}_{AWEQC}}$), MSE, and RB, for stationary design D_S with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{AWEQC}$	$\sigma_{\hat{\beta}_{AWEQC}}$	RB(%)	$\hat{R}$	MSE
0.0	0.0	0.026	0.192	13.300	-0.009	0.038
	0.3	0.040	0.204	19.600	-0.022	0.043
	0.5	0.053	0.219	24.300	-0.017	0.051
	0.7	0.073	0.232	31.400	-0.009	0.059
0.5	0.0	0.537	0.217	17.200	-0.012	0.048
	0.3	0.546	0.209	21.900	-0.007	0.046
	0.5	0.561	0.215	28.200	-0.005	0.050
	0.7	0.580	0.229	34.800	0.004	0.059
1.0	0.0	1.037	0.227	16.100	-0.014	0.053
	0.3	1.047	0.223	21.000	-0.025	0.052
	0.5	1.058	0.223	26.200	-0.017	0.053
	0.7	1.076	0.257	29.700	-0.002	0.072

3.1.3 Simulation Results: A Comparison

The results from Table 3.3 show that the WEQC approach produces a biased regression estimate for β, for all values of β and ρ. For example, when the true $\beta = 0.5$ and

$\rho = 0.5$, the WEQC estimate is found to be $\hat{\beta}_{WEQC} = 0.630$. However, the results from both Tables 3.1 and 3.2 show that the GQL and WOR approaches produce a nearly unbiased estimate for β, with the exception of the GQL and WOR estimates for large ρ. Consequently, to compare the performances of the WEQC approach with the GQL and WOR approaches, we examine the relative biases (RB) of the estimates. It is clear from all three tables that the RBs are much large in Table 3.3 as compared to those of Tables 3.1 and 3.2. Note that this occurred because the WEQC approach produces biased estimates with smaller standard errors. Therefore, the WEQC estimates do not appear converge to the true value.

Now to compare the performances of the GQL and WOR approaches, we use the mean squared error (MSE). This is due to the fact that, in general,both approaches produce unbiased estimates. The results show that for large values of ρ, such as $\rho = 0.7$, the MSEs of the GQL estimates are, in general, larger than those of the WOR approach. For example, when $\rho = 0.7$, and $\beta = 1.0$, the GQL approach has an MSE of 3.797, while the WOR approach has an MSE of 2.162. However, for smaller correlation values such as $\rho = 0.3$ and 0.5, the MSEs of $\hat{\beta}_{GQL}$ are 0.272 and 0.436, whereas the MSEs of $\hat{\beta}_{WOR}$ are 0.528 and 1.119, which are much larger. Thus, the GQL approach appears to perform better than the WOR approach except for large

values of ρ. Note that this relative performance between the GQL and WOR approaches for large ρ appear to reverse in the non-stationary case, which we discuss in the next section. The better performance of the WOR approach in the stationary case for large ρ, however, indicates that the WOR approach is a reasonable one for the stationary data.

When Table 3.4 is compared with Tables 3.1 and 3.2, we obtain similar results to those mentioned above. That is, this alternative WEQC approach also produces worse estimates as compared to the GQL and WOR approaches.

3.2 A Simulation Study for Non-Stationary Correlation Model

The purpose of this section is to generate correlated data adhering to the following AR(1) model

$$Pr[y_{i1} = 1] = \frac{e^{x'_{i1}\beta}}{1 + e^{x'_{i1}\beta}} = \pi_{i1}$$

$$Pr[Y_{it}|y_{i,t-1}] = \pi_{it} + \rho(y_{i,t-1} - \pi_{it}), \qquad \text{for } t = 2, \ldots, T, \tag{3.2.1}$$

with time dependent covariates and estimate β by the three different approaches, namely, GQL discussed in Section 2.1, WEQC (working equi-correlation) based GEE which is discussed in Section 2.2.1, and WOR (working odds ratio) based GEE discussed in Section 2.2.2.

3.2.1 Simulation Design and Data Generation

Similar to the simulation study for the stationary case, we use the design parameters as in 3.1.1. That is, $n = 100$, $T = 4$, $p = 1$, and

$$\rho = 0.0, 0.3, 0.5, 0.7.$$

However, to mark the difference between the stationary and non-stationary models, unlike in the stationary case (3.1.3), we now choose the non-stationary design (D_{NS}) covariates as

$$x_{it} = \begin{cases} \frac{1}{2} & \text{for } t = 1, 2;\ i = 1, \ldots, n/4 \\ 1 & \text{for } t = 3, 4;\ i = 1, \ldots, n/4 \\ -\frac{1}{2} & \text{for } t = 1;\ i = (n/4) + 1, \ldots, 3n/4 \\ 0 & \text{for } t = 2, 3;\ i = (n/4) + 1, \ldots, 3n/4 \\ \frac{1}{2} & \text{for } t = 4;\ i = (n/4) + 1, \ldots, 3n/4 \\ \frac{t}{8} & \text{for } t = 1, \ldots, T;\ i = (3n/4) + 1, \ldots, n. \end{cases} \tag{3.2.2}$$

It is clear that, unlike the stationary model discussed in Section 3.1, the non-stationary model uses covariates that are dependent on t. For the covariate effect, we choose

$$\beta_1 \equiv \beta = 0.0, 0.5, \text{and } 1.0, \tag{3.2.3}$$

the same as in the stationary case.

Next we generate the repeated binary responses as follows. Recall that

$$\pi_{it} = \frac{e^{x_{it}\beta}}{1 + e^{x_{it}\beta}}, \tag{3.2.4}$$

for all $t = 1, \ldots, 4$. Thus, to generate the first response, we use

$$y_{i1} \sim Bin(\pi_{i1}). \tag{3.2.5}$$

Next, for a given value of ρ, we use (3.2.1) to generate y_{i2} using

$$y_{i2} \sim Bin(\lambda_{i2} = \pi_{i2} + \rho(y_{i1} - \pi_{i1})). \tag{3.2.6}$$

Once y_{i2} is generated, we use it to generate y_{i3} as

$$y_{i3} \sim Bin(\lambda_{i3} = \pi_{i3} + \rho(y_{i2} - \pi_{i2})). \tag{3.2.7}$$

We follow this procedure and proceed to generate all responses, that is, $y_{i1}, \ldots, y_{iT}$.

3.2.2 GQL Versus WOR and WEQC Approaches for β Estimation

In a given simulation, using the responses $\{y_{it}, t = 1, \ldots, 4; i = 1, \ldots, 100\}$ generated in Section 3.2.1, and x_{it} as in (3.2.2), we compute the GQL, WOR, and WEQC estimates of β by solving the appropriate estimating equations from Chapter 2.

GQL Estimation:

To be specific, to compute $\hat{\beta}_{GQL}$ for β, we use the iterative equation given in (2.1.17), where X_i' is given as in (2.1.16), A_i is given as in (2.1.11), and the $C_i^{(ns)}$ matrix in $\Sigma_i^{(ns)} = A_i^{\frac{1}{2}} C_i^{(ns)} A_i^{\frac{1}{2}}$ is given as in (2.1.12) and (2.1.13). In order to use the iterative equation (2.1.17), we need to compute the non-stationary correlation matrix $C_i^{(ns)} = \left(c_{i,ut}^{(ns)}\right)$ by (2.1.13). For this, we need to estimate ρ, which we do by (2.1.23).

WOR Estimation:

To compute $\hat{\beta}_{WOR}$ for β, we use the iterative equation given in (2.2.27), where X_i' is given in (2.1.16) and A_i is the $T \times T$ diagonal matrix as in (2.1.11). To use (2.2.27), in this approach, we need to compute the covariance matrix $\Sigma_i(\tau)$ by using (2.2.23), where $\pi_{i,11}^{(t|u)}(\tau)$ is given in (2.2.22). The constant odds ratio τ is estimated using (2.2.37). We then use (2.2.27) to obtain $\hat{\beta}_{WOR}$.

WEQC Estimation:

To compute $\hat{\beta}_{WEQC}$ for β, we use the iterative equation given in (2.2.5), where X_i' is given in (2.1.16) and A_i is the $T \times T$ diagonal matrix as in (2.2.3). To use (2.2.5), we need to compute the $R(\alpha)$ matrix, where α is the equi-correlation parameter. This we have done, similar to the stationary case, by using (2.2.7) and (2.2.8).

As in the stationary case, the above computations for β estimation based on all three approaches, namely GQL, WOR and WEQC, are now repeated 500 times. The average of these 500 values under a given method is computed to obtain the final estimate. These estimates are denoted by $\hat{\beta}_{GQL}$, $\hat{\beta}_{WOR}$, and $\hat{\beta}_{WEQC}$, respectively. In addition, the standard error of these 500 values (SSE), the RB, and MSE are also computed. The percentage RB and MSE have the formulas, for example for the GQL, as given in (3.1.17) and (3.1.18), respectively. The simulation results for the non-stationary GQL, WOR, and WEQC approaches are given in Tables 3.5, 3.6, and 3.4, respectively. Note that for $\beta = 1.0$ and $\rho = 0.7$, the WOR approach had convergence problems and hence results cannot be shown.

Table 3.5: GQL estimate ($\hat{\beta}_{GQL}$) for regression effect along with its standard error ($\sigma_{\hat{\beta}_{GQL}}$), MSE, and RB, for non-stationary design D_{NS} with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{GQL}$	$\sigma_{\hat{\beta}_{GQL}}$	RB(%)	$\hat{\rho}$	MSE
0.0	0.0	-0.010	0.213	4.700	0.000	0.046
	0.3	0.005	0.224	2.000	0.305	0.050
	0 5	0.004	0.221	1.600	0.506	0.049
	0.7	0.000	0.216	0.056	0.703	0.046
0.5	0.0	0.524	0.209	11.500	-0.002	0.044
	0.3	0.493	0.229	3.200	0.307	0.052
	0.5	0.494	0.229	2.500	0.506	0.052
	0.7	0.492	0.208	3.600	0.704	0.043
1.0	0.0	1.032	0.225	14.100	0.000	0.052
	0.3	1.014	0.234	6.000	0.306	0.055
	0.5	0.996	0.238	1.600	0.505	0.057
	0.7	0.991	0.211	4.000	0.704	0.045

As in the stationary case, we also use an alternative estimation method for the 'working' equi-correlation parameter α. In this alternative approach, we use the formula for α as given in (2.2.10), where $\tilde{y}_{it}$ is given in (2.2.11). The simulation results for this alternative WEQC (AWEQC) estimates are given in Table 3.8.

3.2.3 Simulation Results: A Comparison

In general, all three approaches in the present non-stationary case produced unbiased regression estimates. For this reason, to examine the relative performances of the GQL and WOR approaches, we compare their respective MSEs. The results show

Table 3.6: Equal odds ratio based GEE estimate ($\hat{\beta}_{WOR}$) for regression effect along with its standard error ($\sigma_{\hat{\beta}_{WOR}}$), MSE, and RB, for non-stationary design D_{NS} with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{WOR}$	$\sigma_{\hat{\beta}_{WOR}}$	RB (%)	MSE
0.0	0.0	0.009	0.462	1.984	0.213
	0.3	0.011	0.588	1.900	0.346
	0.5	-0.012	0.697	1.771	0.485
	0.7	0.002	0.872	1.948	0.760
0.5	0.0	0.505	0.447	1.063	0.200
	0.3	0.491	0.592	1.500	0.351
	0.5	0.504	0.690	0.532	0.461
	0.7	0.541	0.803	5.106	0.646
1.0	0.0	1.013	0.477	2.700	0.228
	0.3	0.992	0.639	1.300	0.409
	0.5	1.001	0.716	0.097	0.513
	0.7	-	-	-	-

Table 3.7: WEQC based GEE estimate ($\hat{\beta}_{WEQC}$) for regression effect along with its standard error ($\sigma_{\hat{\beta}_{WEQC}}$), MSE, and RB, for non-stationary design D_{NS} with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{WEQC}$	$\sigma_{\hat{\beta}_{WEQC}}$	RB (%)	$\hat{\alpha}$	MSE
0.0	0.0	0.072	0.252	28.900	0.001	0.069
	0.3	0.174	0.265	65.700	0.192	0.101
	0.5	0.224	0.268	83.400	0.363	0.122
	0.7	0.234	0.263	89.200	0.575	0.124
0.5	0.0	0.612	0.255	44.100	0.008	0.078
	0.3	0.659	0.264	60.200	0.197	0.095
	0.5	0.712	0.255	82.900	0.367	0.110
	0.7	0.708	0.236	87.900	0.580	0.099
1.0	0.0	1.123	0.274	44.700	0.024	0.090
	0.3	1.171	0.277	61.900	0.208	0.106
	0.5	1.204	0.278	73.500	0.379	0.119
	0.7	1.195	0.251	77.700	0.592	0.101

Table 3.8: AWEQC based GEE estimate ($\hat{\beta}_{AWEQC}$) for regression effect along with its standard errors ($\sigma_{\hat{\beta}_{AWEQC}}$), MSE, and RB, for non-stationary design D_{NS} with $t = 4$, $n = 100$, and selected values of ρ; 500 simulations.

True β	True ρ	$\hat{\beta}_{AWEQC}$	$\sigma_{\hat{\beta}_{AWEQC}}$	RB (%)	$\hat{\alpha}$	MSE
0.0	0.0	0.013	0.596	2.200	-0.007	0.355
	0.3	0.055	0.605	9.200	0.006	0.369
	0.5	0.092	0.609	15.100	0.018	0.379
	0.7	0.121	0.609	19.900	0.033	0.386
0.5	0.0	0.425	0.616	12.200	-0.006	0.385
	0.3	0.500	0.623	0.038	0.007	0.388
	0.5	0.533	0.622	5.400	0.019	0.388
	0.7	0.560	0.617	9.800	0.033	0.384
1.0	0.0	0.850	0.728	20.600	-0.007	0.552
	0.3	0.920	0.732	10.970	0.006	0.542
	0.5	0.950	0.734	6.700	0.018	0.541
	0.7	0.978	0.727	3.100	0.033	0.529

that for all values of β and ρ, the MSEs for the GQL approach, given in Table 3.5, are quite smaller than those for the WOR approach, which are given in Table 3.6. For example, for true $\beta = 0.0$ and $\rho = 0.7$, the MSE of $\hat{\beta}_{GQL}$ is 0.046, while $\hat{\beta}_{WOR}$ has an MSE of 0.760. Unlike that stationary case, these results hold true for both small and large values of ρ.

If we were to also compare the performances of the WEQC approach to the GQL and WOR approaches through their respective MSEs, we would also find similar results to those discussed above. That is to say, the MSEs of the GQL method are much

lower than those of the other two approaches. For example, consider when $\beta = 0.5$ and $\rho = 0.5$. The MSE for the WEQC approach is 0.110, and for the WOR approach the MSE is 0.461. However, for the GQL approach, when true $\beta = 0.5$ and $\rho = 0.5$, the MSE is 0.052.

When Table 3.8 is compared with Tables 3.5 - 3.7, we find that the GQL approach once again produces estimates with lower MSEs. We now compare between the WOR and AWEQC approaches. For small values of ρ, such as 0.0 and 0.3, the MSEs of the AWEQC approach are larger than those of the WOR approach. For example, when $\beta = 1.0$ and $\rho = 0.0$, the WOR approach has an MSE of 0.228, while the AWEQC approach has an MSE of 0.552. In this regard, the WOR approach appears to perform better than the AWEQC approach. When the AWEQC is compared with both the GQL and WEQC, one may note that the AWEQC approach produces larger MSEs for all β and ρ.

In the next chapter, we provide an illustration of the relative performance of the GQL, WOR, and WEQC approaches for an asthma data set.

Chapter 4

Analyzing Asthma Data: An Illustration

In Chapter 3 we discussed the relative performances of the GQL, WEQC, and WOR approaches through a simulation study. The purpose of this chapter is to provide a numerical illustration of the application of these three approaches by analyzing an asthma data set. To be more specific, we want to apply the GQL, WEQC, and WOR approaches to a data set for 537 children from Ohio state, who were examined annually from the ages 7-10. As mentioned in Chapter 1, this data set considers whether the children had respiratory infection in the previous year (as reported by the mother of the child). In other words, the repeated response for this data set is the

wheezing status (1=yes, 0=no) of the child. The initial smoking status of the mothers (1=regular smoker, 0= not) were also collected, and considered as a covariate. Thus, for $t = 1, \ldots, 4$,

$$\begin{aligned} x_{it} &= (x_{it_1}, x_{it_2})' \\ &= \big(1, \text{initial smoking status of mother}\big)'. \end{aligned}$$

showing that the covariates are stationary. Furthermore, $y_{it} = 1$ or 0 for all $i = 1, \ldots, 537$; and $t = 1, \ldots, 4$.

The scientific interest of the study is to examine the effect that smoking by the mother has on the wheezing status of her child. In addition, it is of interest to estimate β once one has taken the longitudinal correlations of the responses into account. This particular data set was earlier analyzed by Zeger, Liang and Albert (1988), Sutradhar (2003), and Sutradhar (2011, Chapter 7), among others.

4.1 Estimation of Smoking Effect

4.1.1 GQL Estimation of β

For the GQL estimation, we use the correlation structure

$$C_i(\rho) = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{T-1} \\ & 1 & \rho_1 & \cdots & \rho_{T-2} \\ & & \ddots & & \vdots \\ & & & 1 & \rho_1 \\ & & & & 1 \end{pmatrix}, \tag{4.1.1}$$

as used in Sutradhar (2003). Using the stationary GQL approach discussed in Section 3.1.2, we obtain the GQL estimate of $\beta = (\beta_1, \beta_2)'$

$$\begin{aligned} \hat{\beta}_1(\text{intercept}) &= -1.826 \\ \hat{\beta}_2(\text{mothers smoking effect}) &= 0.263. \end{aligned}$$

and the standard errors of the regression estimates were found to be

$$\begin{aligned} \text{s.e}(\hat{\beta}_1) &= 0.111 \\ \text{s.e}(\hat{\beta}_2) &= 0.178. \end{aligned}$$

Note that these standard errors of the regression estimates were calculated by exploiting the formula

$$\text{cov}(\hat{\beta}) = \left(\sqrt{\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma(\beta, \rho)_i^{-1} \mathbf{A}_i \mathbf{X}_i} \right)^{-1} \tag{4.1.2}$$

derived from (2.1.26).

The above β estimation was done based on the moment estimates for the lag correlation, which were found to be

$$\begin{aligned} \hat{\rho}_1 &= 0.401 \\ \hat{\rho}_2 &= 0.314 \\ \hat{\rho}_3 &= 0.297, \end{aligned}$$

by (3.1.12).

4.1.2 WEQC Estimation of β

Using the WEQC approach discussed in Section 2.1.1, we obtain the WEQC estimate of β as

$$\hat{\beta}_1(\text{intercept}) = -1.349$$

$$\hat{\beta}_2(\text{mothers smoking effect}) = -0.094,$$

with standard errors of the regression estimates given by

$$\text{s.e}(\hat{\beta}_1) = 0.116$$

$$\text{s.e}(\hat{\beta}_2) = 0.169.$$

Note that these standard errors were computed from the formula

$$\text{cov}(\hat{\beta}_{\text{WEQC}}) = \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i R(\alpha)^{-1} \mathbf{A}_i \mathbf{X}_i\right]^{-1}, \tag{4.1.3}$$

derived from (2.2.5), and using equi-correlation structure for $R(\alpha)$.

The equi-correlation, α, is estimated using (2.2.8) is found to be

$$\hat{\alpha} = 0.357.$$

Note that $\hat{\alpha}$ was computed based on $\hat{\alpha}_{ut}$ as

$$\begin{pmatrix} \hat{\alpha}_{12} & \hat{\alpha}_{13} & \hat{\alpha}_{14} \\ & \hat{\alpha}_{23} & \hat{\alpha}_{24} \\ & & \hat{\alpha}_{34} \end{pmatrix} = \begin{pmatrix} 0.354 & 0.308 & 0.327 \\ & 0.443 & 0.329 \\ & & 0.381 \end{pmatrix}, \tag{4.1.4}$$

using (2.2.7).

4.1.3 WOR Estimation of β

Using the WOR approach discussed in Section 2.2.2, we obtain the WOR estimate of β as

$$\begin{aligned} \hat{\beta}_1(\text{intercept}) &= -1.821 \\ \hat{\beta}_2(\text{mothers smoking effect}) &= 0.272, \end{aligned}$$

along with the standard errors of the regression estimates given by

$$\begin{aligned} \text{s.e}(\hat{\beta}_1) &= 0.111 \\ \text{s.e}(\hat{\beta}_2) &= 0.180. \end{aligned}$$

Note that these standard errors were computed using the formula

$$\text{cov}(\hat{\beta}_{\text{WOR}}) = \left[\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{A}_i \Sigma_i(\tau)^{-1} \mathbf{A}_i \mathbf{X}_i\right]^{-1}, \tag{4.1.5}$$

derived from (2.2.27).

The constant odds ratio, τ_{ut}, for $u < t$, is given as

$$\begin{pmatrix} \hat{\tau}_{12} & \hat{\tau}_{13} & \hat{\tau}_{14} \\ & \hat{\tau}_{23} & \hat{\tau}_{24} \\ & & \hat{\tau}_{34} \end{pmatrix} = \begin{pmatrix} 7.130 & 5.777 & 7.231 \\ & 11.469 & 7.261 \\ & & 9.972 \end{pmatrix}. \tag{4.1.6}$$

We then compute an estimate of τ as

$$\hat{\tau} = 8.099,$$

by way of (2.2.37).

4.2 Discussion

All three approaches, that is to say, GQL, WEQC, and WOR, produced virtually the same estimate for the intercept parameter $\hat{\beta}_1$, the GQL estimate being $\hat{\beta}_{1,GQL} = -1.826$, the WEQC estimate being $\hat{\beta}_{1,WEQC} = -1.349$, and the WOR estimate being $\hat{\beta}_{1,WOR} = -1.821$. However, for the effect of the mother's smoking habit on the asthma status of her child, the GQL and WOR approaches produced almost the same estimate. This result agrees with the simulation pattern for the stationary case that we have discussed in Section 3.1. The WEQC based approach produced a negative estimate, which is counter intuitive.

The standard errors for both $\hat{\beta}_1$ and $\hat{\beta}_2$ in all three approaches, that is GQL, WEQC, and WOR, appear to be quite similar. Hence, the WEQC approach appears to produce a different result only for the estimation of $\hat{\beta}_2$, as compared to the GQL and WOR approaches. As mentioned above, this result is expected, as we saw in Table 3.3 that for the stationary case, the WEQC approach produced more biased estimates for the regression parameters as compared to the other methods.

Chapter 5

Concluding Remarks

In the longitudinal setup, there exists some comparisons between a recently proposed GQL estimation approach and the so-called 'working' correlation based GEE approaches for the estimation of the regression effect. The GQL approach, in general, tends to produce more efficient estimates, both for stationary and non-stationary data. However, there exists an odds ratio approach, where the joint probabilities are computed in terms of odds ratios. Yet, as a result of estimation difficulty for the odds ratio, some authors have used an equal odds ratio assumption for the construction of the joint probability based estimating equations. In this thesis, we have made a comparative study between the 'working' equal odds ratio (WOR) and GQL approaches, and found that the GQL approach performed better than the WOR approach in the

non-stationary case. In general, this was also true for the stationary case. We also included the 'working' EQC (WEQC) approach in the comparison.

In addition to the simulation study, we have also included a real life data analysis for the comparison of the three approaches. The GQL and WOR approaches were found to produce similar estimates.

Bibliography

[1] Bahadur, R.R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction. Stanford Mathematical Studies in the Social Sciences*; Solomon, H., Ed.; Vol. 6, 158-168.

[2] Cook, Richard J., & Yi, Grace Y. (2002). Marginal methods for incomplete longitudinal data arising in clusters. *J. Amer. Statist. Assoc.*, **97**, 1071-1080.

[3] Farrell, P. J. and Sutradhar, B. C. (2006). A non-linear conditional probability model for generating correlated binary data. *Statist. Probab. Lett.*, **76**, 353-361.

[4] Fitzmaurice, G.M. and Laird, N. M. (1993). A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141-151.

[5] Fitzmaurice, G.M., and Lipsitz, S.R. (1995). A model for binary time series data with serial odds ratio patterns. *Applied Statistics*, **44**, 51-61.

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

[7] Lipsitz *et al.* (1991). Generalized estimating equations for correlated binary data: Using the odds ratio as a measure of association. *Biometrika*, **78**, 153-160.

[8] Kanter, M. (1975). Autoregression for discrete processes mod 2. *J. Appl. Probab.*, **12**, 371-375.

[9] Prentice, Ross L. (1988). Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 4, 1033-1048.

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

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

[12] Sutradhar, B.C. (2004). On exact quasi-likelihood inference in generalized linear models. *Sankhya: The Ind. J. of Statist.*, **66**, 263-291.

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

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

[15] Wedderburn, R. W. M. (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, **61**, 439-447.

[16] Williamson, John M., Kim, KyungMann, and Lipsitz, Stuart R. (1995). Analyzing Bivariate Ordinal Data Using a Global Odds Ratio. *J. Amer. Statist. Assoc.* , **90**, 1432-1437.

[17] Zeger, S.L., Liang, K.Y. and Albert, P.S. (1988). Models for longitudinal data: A generalized estimating equations approach. *Biometrics*, **44**, 1049-1060.

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