ON NONLINEAR DYNAMIC BINARY TIME SERIES

VICKNESWARY TAGORE

On Nonlinear Dynamic Binary Time Series

by

©Vickneswary Tagore

A thesis submitted to the School of Graduate Studies
in partial fulfillment of the requirement for the Degree of
Master of Science in Statistics

Department of Mathematics and Statistics
Memorial University of Newfoundland

St. John's Newfoundland, Canada August 2006

Your file Votre référence
ISBN: 978-0-494-30510-2
Our file Notre référence
ISBN: 978-0-494-30510-2

Abstract

There are many practical situations where one may encounter binary data over a long period of time. For example, in clinical studies, one may be interested in examining the effects of certain time dependent covariates on the binary asthma status (yes or no) of an individual recorded daily over a few months. The analysis of this type of binary time series data is, however, not adequately addressed in the literature. In the thesis, we review three widely used binary time series models and discuss their advantages and draw-backs mainly with regard to their correlation structures. We then provide inferences for a non-linear conditional dynamic binary model which appears to accommodate correlations with full ranges.

With regard to the estimation of the regression and a dynamic dependence parameters we use the well-known maximum likelihood (ML) and various versions of the generalized quasilikelihod (GQL) approaches. The relative performances of these approaches are examined through a simulation study. A conditional GQL (CGQL) approach appears to be quite simple and at the same time it produces the same estimates as that of the ML approach. A lag 1 forecasting for a future binary probability is also studied mainly through simulations.

Acknowledgements

I wish to express my great appreciation to my supervisor Professor B.C. Sutradhar for his encouragement, frequent advice, unbelievable patience and helpful discussions and suggestions throughout my program. It was indeed a great privilege to work on this important problem in the area of discrete time series analysis, which was suggested by Professor Sutradhar.

I extend my thanks to my co supervisor Dr. Gary Sneddon for his support throughout my program.

I sincerely acknowledge the School of Graduate Studies and Department of Mathematics and Statistics for their financial supports. I would like to say thank you to all staff of the Department of Mathematics and Statistics for their kindness and great help.

I also thank the examiners of thesis, Dr. Patrick J.Farrell of Carleton University, and Dr. Alwell Oyet of Memorial University of Newfoundland for their constructive comments and suggestions on the original version of the thesis.

I am also grateful to my husband and my parents for their eternal love, emotional support and encouragement during this program.

It is my great pleasure to thank my friends and well - wishers who directly or indirectly encouraged and helped me in the M.Sc. program and contributed to this dissertation.

Contents

List of Tables

Chapter 1

Introduction

1.1 Background of the Problem

There is a vast literature in time series analysis for continuous data. For example, one may refer to Box and Jenkins (1994) for analysis of (i) two hourly chemical process concentration, (ii) daily IBM common stock closing prices and (iii) monthly international airline passengers in log scale. In practice, we may also encounter discrete data such as binary and count time series. When dichotomous responses are collected from an individual (such as individual person or individual firm or business) over a long period of time, their responses form a binary time series. For example, there are many clinical studies when asthma status of an individual is recorded daily, for a few months. In this type of study, multidimensional covariates are also recorded along with the binary response at a given time. We refer to Zhang, Triche and Leaderer (2000) for such a binary time series analysis. These authors were interested in finding the effects of the associated covariates on the 'yes ' or 'no ' respiratory symptom status of a runny or stuffy nose for mothers followed in southwestern Virginia for the summer period from June 10 to August 31, 1995. In an ecological and environmental study, Guttorp (1986) modelled the daily rainfall data (zero for rainfree day and one for rainy day) observed at Sea-Tac airport in Washington for the month of January 1965 through 1982.

For the analysis of time series for count data we refer to Zeger (1988), where the monthly number of poliomyelitis cases were observed from 1970 to 1983 by U.S Centers for Disease Control.

Note that unlike the modelling of continuous time series, the modelling of discrete time series in particular the binary time series, is however not easy. This happens mainly due to the fact that there is no unique way to model the correlation structure of the repeated binary data. Kanter (1975) introduced a dynamic observation driven AR(1) type correlated binary model for stationary data, where the present observation is related to the past observations through a specified function. But this model does not cover the full range for the correlation parameters. Keenan (1982) has dealt with a latent variable approach. To be specific, under the assumption that a latent process follows a multivariate normal distribution, Keenan (1982) developed a time series model for the observed binary responses. In this approach even though the latent process is assumed to have a simple correlation structure, the resulting correlation structure for the responses, however become extremely complicated. Recently, Qaqish (2003) has introduced a conditional linear dynamic binary model, where the conditional probability of a binary response is assumed to be a linear function of past responses. Similar to the models considered by Kanter (1975) and Keenan (1982), this model of Quaish (2003) also suffers from the range problems for the correlation parameters.

1.2 Objective of The Thesis

In this thesis, we consider a non linear dynamic binary time series model that accommodates binary responses with full ranges from -1 to 1 for the correlation parameters. It is further assumed, under this model that the binary responses are influenced by certain time dependent covariates. The main objective of the thesis is to provide inferences for the effects of the covariates by taking the correlation structure of the time series into account. The specific plan of the thesis is as follows.

In Chapter 2, we review two similar but different binary time series models. One of them [Keenan (1982)] is constructed based on a correlated latent process, whereas the second model discussed in Qaqish (2003) is developed based on a linear dynamic relationship among the responses. The advantages and disadvantages of these two models are discussed. In the same chapter, we introduce the non-linear dynamic model for the binary time series and discuss the basic properties of responses such as the mean, variance and correlation structures. The relative merits of this non-linear dynamic model as compared to the models due to Keenan (1982) and Qaqish (2003) are discussed.

The inference for the proposed non linear dynamic binary time series model is provided in Chapter 3. Note that in addition to the regression effects, it is also of main interest to examine the dynamic dependence among the repeated binary responses. For this purpose, we discuss estimation techniques for the estimation of both regression and the dynamic dependence parameters. To be specific, we examine the performance of four competitive estimation approaches through an extensive simulation study. These approaches are: (i) the generalized quasilikelihood (GQL) approach, (ii) a semi-quasi likelihood (SGQL) approach , (iii) the maximum likelihood (ML) approach and (iv) a conditional quasilikelihood (CGQL) approach.

In Chapter 4, we examine the performance of the ML and CGQL approaches in estimating the probability for a future binary response. This is done mainly by comparing the predicted probabilities and the true probabilities based on a simulation study.

In Chapter 5, we discuss the estimation technique for the estimation of regression and dependence parameters for a lag 2 model. Note that, we examine the performance of the estimation approach through a simulation study.

We conclude the thesis in Chapter 6 with some remarks about the usefulness of the non - linear dynamic binary time series model and the parameter estimation techniques that we constructed in Chapter 3. In the same chapter, we have also noted the possibilities of some future research in this area.

Chapter 2

Non - Stationary Models

Let y_t denote the binary response recorded at time point t from an individual or individual firm. Also let $x_t = (x_{t1}, \ldots, x_{tj}, \ldots, x_{tp})'$ be the p dimensional vector of covariates which explains y_t, and $\beta = (\beta_1, \ldots, \beta_j, \ldots, \beta_p)'$ denote this effect of x_t on y_t. In the independence set up, where $y_1, \ldots, y_t, \ldots, y_T$ are treated as independent binary responses, it is common to use the binary logistic form

$$p(y_t = 1|x_t) = \frac{e^{x_t'\beta}}{[1 + e^{x_t'\beta}]} \quad , \tag{2.1}$$

as a marginal probability model to relate y_t with x_t at time point t. This leads to the likelihood function given by

$$L(\beta) = \prod_{t=1}^{T} p(y_t = 1|x_t) \tag{2.2}$$

which can be maximized for consistent estimation of the β parameter.

When $y_1, \ldots, y_t, \ldots, y_T$ are repeated binary responses from the same individual, it is likely that the responses will be correlated. In this case, we say that the binary responses form a time series. Note however that the construction of the likelihood function in the correlated set up may be complicated. Consequently, the estimation of the regression effects β also becomes complicated. Over the past few decades, some authors have attempted to develop certain correlation structure based likelihood functions for the binary data. For example, we refer to Keenan (1982) for the likelihood construction of a correlated latent process based binary time series, and Qaqish (2003) for a conditional dynamic relationship based correlated binary model. In the following two subsections we review these models and discuss their advantages and limitations. Note that the above two models are recently discussed in detail by Sutradhar and Rao (2006) in the context of longitudinal data analysis.

2.1 Correlated Latent Process Based Dynamic Model: An Overview

2.1.1 Likelihood Computation and Complexity

Let $\{y_t^*, t = 1, \ldots, T\}$ be a sequence of a latent variable, that follows a Gaussian AR(1) process. Suppose that $y^* = (y_1^*, \ldots, y_t^*, \ldots, y_T^*)'$. We then write

$$y^* \sim N(\theta^*, \Sigma^*) \tag{2.3}$$

where $\theta^* = (\theta_1^*, \ldots, \theta_t^*, \ldots, \theta_T^*)$ with $\theta_t^* = x_t'\beta$ and Σ^* is an AR(1) process based suitable covariance matrix. Next suppose that the relationship between the observed binary responses y_t and the latent quantity y_t^* is given by

$$y_t = \begin{cases} 1 & \text{if } y_t^* > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{2.4}$$

The repeated binary responses $(y_1, \ldots, y_t, \ldots, y_T)$ from the same individual are correlated. It then follows that the likelihood function of the binary responses is given by

$$L(\beta,\ \theta^*\ \Sigma^*) = \int_0^\infty \ldots \int_0^\infty g_N(y^*|\theta^*\ \Sigma^*)\ dy_1^* \ldots dy_T^*\ , \tag{2.5}$$

where $g_N(y_1^*, \ldots, y_t^*, \ldots, y_T^*)$ is the T - dimensional normal density function given by

$$g_N(y^*|\theta^*\ \Sigma^*) \ = \ (2\pi)^{\frac{-T}{2}} |\Sigma^*|^{-\frac{1}{2}}\ e^{\frac{-1}{2}[(y^*-\theta^*)'\Sigma^{*-1}(y^*-\theta^*)]}.$$

By shifting the latent process as $z_t^* = [y_t^* - \theta_t^*]$ where $\theta_t^* = x_t^{'}\beta$, one may re- write the equation (2.5) in the cumulative form as

$$L(\beta,\ \theta^*\ \Sigma^*) = \int_{-\infty}^{x_1^{'}\beta} \ldots \int_{-\infty}^{x_T^{'}\beta} g_N(z^*|0\ \Sigma^*)\ dz_1^* \ldots dz_T^* \qquad (2.6)$$

with $z^* = (z_1^*,\ \ldots\ , z_t^*,\ \ldots\ , z_T^*)^{'}$. But the likelihood function (2.5) or (2.6) is extremely complicated for the purpose of the estimation of the main parameter β.

2.1.2 Basic Properties of the Model: Mean, Variance and Correlation

Note that one may attempt to use an alternative approach such as the quasilikelihood approach (Zeger (1988) , Sutradhar (2003)) to estimate this β parameter. This technique requires the formulas for the mean vector and the covariance matrix of the observed response vector $y = (y_1,\ \ldots\ , y_T)^{'}$. Let θ and Σ denote the mean vector and the covariance matrix of y respectively. The computations for the θ vector and the Σ matrix under the present latent process based model, however, also appear to be complicated. To have a feel about the form of the θ vector and the Σ matrix, we now consider the simplest case with T = 2, and demonstrate that even though y_1^* and y_2^* may have a simple correlation structure, the correlation structure of y_1 and y_2 would be complicated.

To compute θ and Σ for T = 2, we first express θ^* and Σ^* of the corresponding latent process as

$$\theta^* = \{\theta_1^*, \theta_2^*\}^{'} \quad \text{and} \quad \Sigma^* = \begin{pmatrix} \sigma_{11}^* & \sigma_{12}^* \\ & \sigma_{22}^* \end{pmatrix} = \begin{pmatrix} \sigma_{11}^* & \rho^*\sqrt{\sigma_{11}^*\ \sigma_{22}^*} \\ & \sigma_{22}^* \end{pmatrix}, \qquad (2.7)$$

where $\theta_t^* = x_t^{'}\beta,\ t = 1, 2$ and ρ^* is the correlation between y_1^* and y_2^*. As the latent variable y_t^* is assumed to follow a normal distribution with mean θ_t^* and variance covariance matrix (σ_{ut}^*), then the unconditional mean of the binary response y_t may be computed as

$$E(Y_t) = P(y_t = 1) \;=\; \int_0^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma^*_{tt}} e^{\frac{-1}{2}\left(\frac{y^*_t - \theta^*_t}{\sqrt{\sigma^*_{tt}}}\right)^2} dy^*_t. \qquad (2.8)$$

By using $z^*_t = [y^*_t - \theta^*_t] \sim N(0, \sigma^*_{tt})$, this expectation may be re - expressed as

$$\begin{aligned} E(Y_t) &= \int_{-\theta^*_t}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma^*_{tt}} e^{\frac{-1}{2\,\sigma^*_{tt}} z^{*2}_t} dz^*_t \\ &= \int_{-\infty}^{\theta^*_t} \frac{1}{\sqrt{2\pi}\,\sigma^*_{tt}} e^{\frac{-1}{2\,\sigma^*_{tt}} z^{*2}_t} dz^*_t \\ &= F_N(\theta^*_t), \; say, \qquad t = 1, 2 \qquad (2.9) \end{aligned}$$

where $F_N(\theta^*_t)$ is the cumulative normal density function for $\theta^*_t = x'_t\beta$.
To compute the covariance structure (σ_{ut}) for the binary responses y_1 and y_2, we first write the the unconditional variance for y_t as

$$var(Y_t) = F_N(\theta^*_t)(1 - F_N(\theta^*_t)) \qquad t = 1, 2, \qquad (2.10)$$

where $F_N(\theta^*_t) = F_N(x'_t\beta)$ is the expectation of y_t as in (2.9). Next, the covariance between y_1 and y_2 is written as

$$\begin{aligned} cov(Y_1, Y_2) &= E(Y_1\,Y_2) - E(Y_1)\,E(Y_2) \\ &= P(Y_1 = 1 \;\; Y_2 = 1) - E(Y_1)\,E(Y_2) \\ &= \int_0^{\infty}\int_0^{\infty} (2\pi)^{\frac{-p}{2}} \mid \Sigma^* \mid^{\frac{-1}{2}} e^{\frac{-1}{2}[(y^* - \theta^*)' \Sigma^{*-1} (y^* - \theta^*)]} dy^*_1 dy^*_2 - F_N(\theta^*_1) F_N(\theta^*_2). \end{aligned} \qquad (2.11)$$

By using $z^*_t = [y^*_t - \theta^*_t]$, (2.11) reduces to

$$\begin{aligned} cov(Y_1, Y_2) &= \int_{-\theta^*_1}^{\infty}\int_{-\theta^*_2}^{\infty} (2\pi)^{\frac{-2}{2}} \mid \Sigma^* \mid^{\frac{-1}{2}} e^{\frac{-1}{2}[z^{*'} \Sigma^{*-1} z^*]} dz^*_1 dz^*_2 - F_N(\theta^*_1) F_N(\theta^*_2) \\ &= \int_{-\infty}^{\theta^*_1}\int_{-\infty}^{\theta^*_2} (2\pi)^{-1} \mid \Sigma^* \mid^{\frac{-1}{2}} e^{\frac{-1}{2}[z^{*'} \Sigma^{*-1} z^*]} dz^*_1 dz^*_2 - F_N(\theta^*_1) F_N(\theta^*_2), \end{aligned} \qquad (2.12)$$

where $z^* = (y^* - \theta^*) \sim N_2(0, \Sigma^*)$, and Σ^* is 2 x 2 variance - covariance matrix of the latent variable. One may then compute the correlation between the binary variables y_1 and y_2 by using

$$corr(Y_1\,, Y_2) = \frac{cov(Y_1\,,\,Y_2)}{\sqrt{var(Y_1)\;\;var(Y_2)}}, \tag{2.13}$$

where $cov(Y_1\,,\,Y_2)$ is given by (2.12) and $var(Y_t)$ for t = 1,2 is given by (2.10). Note that the computation of the covariance by (2.12) is quite complicated even though σ^*_{12} in (2.7) has a simple form, that is, $\sigma^*_{12} = \rho^* \sqrt{\sigma^*_{11}\;\;\sigma^*_{22}}$ with ρ^* as the correlation between y^*_1 and y^*_2. Then, the correlation structure computed by (2.13) would be complicated to use for the computation of the quasilikelihood estimation equation for β. Also the correlation between Y_1 and Y_2 does not satisfy the range from -1 to 1, which makes the model quite limited for any practical application.

2.2 Lagged Response Based Conditional Linear Dynamic Model

In the last subsection, a correlated binary model was developed based on a correlated latent process. The resulting correlation structure of that model was found to be complicated. There exists some alternative conditional linear dynamic models in the literature for correlated binary time series. For example, for $\mu^*_t = \frac{e^{x'_t\beta}}{1+e^{x'_t\beta}}$; $t = 1, \ldots, T$, one may consider a correlated binary model given by

$$E(Y_t|y_{t-1}) = \tilde{\lambda}_{t|t-1} = \mu^*_t + \psi(y_{t-1} - \mu^*_{t-1}), t = 2, \ldots, T\;,\;\text{[Qaqish (2003)]}. \tag{2.14}$$

But it was not clear from Qaqish (2003) how one may obtain this conditional linear form (2.14) for modelling binary time series data. In this subsection, we first show that this model, in fact arises based on a uniform latent process, whereas a normal latent process based model was discussed in the last subsection. We then discuss in brief the likelihood approach for the estimation of the parameters of the model. We also provide the basic properties such as mean, variance and covariance of the correlated responses based on this model. These properties are given as in the previous subsection with an objective that one may be able to use these to construct a second moments based quasilikelihood estimation approach.

Suppose that the latent variable y_t^* (which was Gaussian in the previous model) follows a uniform distribution with the probability density function

$$f(y_t^*) = \frac{1}{2}, \quad -1 < y_t^* < 1. \ t = 1, 2, \ldots, T. \tag{2.15}$$

Similar to the previous subsection, the binary variable y_t is generated through the following relationship with y_t^*. That is,

$$y_t = \begin{cases} 1 & \text{if } y_t^* > [1 - 2\tilde{\lambda}_{t|t-1}] \ \ t = 2, \ldots, T \\ 0 & \text{otherwise,} \end{cases} \tag{2.16}$$

which appears to be quite different than the relationship (2.4) used for the normal latent based process discussed in the previous subsection.

Note that for $t = 1$, we assume that $\tilde{\lambda}_{1|0} \equiv \lambda_1 = \mu_1^*$. It then follows that for $t = 1$ the marginal probability of $y_1 = 1$ is given by

$$\begin{aligned} \lambda_1^* = P(y_1 = 1) &= \int_{1-2\tilde{\lambda}_1}^{1} f(y_1^*) \ dy_1^* \\ &= \int_{1-2\tilde{\lambda}_1}^{1} \frac{1}{2} \ dy_1^* \\ &= \tilde{\lambda}_1 = \mu_1^*, \end{aligned} \tag{2.17}$$

whereas for $t = 2, \ldots, T,$ the probability that y_t takes value $y_t = 1$ conditional on y_{t-1} is written as

$$\begin{aligned} \lambda_{t|t-1}^* = P(y_t = 1|y_{t-1}) &= \int_{1-2\tilde{\lambda}_{t|t-1}}^{1} f(y_t^*) \ dy_t^* \\ &= \int_{1-2\tilde{\lambda}_{t|t-1}}^{1} \frac{1}{2} \ dy_t^* \\ &= \tilde{\lambda}_{t|t-1} = \mu_t^* + \psi(y_{t-1} - \mu_{t-1}^*). \end{aligned} \tag{2.18}$$

Thus, it is clear that, the conditional linear model (2.14) or (2.18) is developed based on the uniform latent process given by (2.15). It is also clear that even though (2.18) looks similar to the Gaussian AR(1) model, the ψ parameter however may not satisfy the full range from -1 to 1. This is because the binary probability $\lambda^*_{t|t-1}$ must lie between 0 and 1.

2.2.1 Likelihood Computation and Complexity

Note that it follows from the model (2.14) or (2.18) that the time series responses $(y_1, \ldots, y_t, \ldots, y_T)$ are correlated. Consequently, their likelihood function may be written as

$$\begin{aligned} L(\beta, \psi) &= P(y_1, y_2, \ldots, y_t, \ldots, y_T) \\ &= P(y_1)\, P(y_2, y_3, \ldots, y_T | y_1) \\ &= P(y_1)\, P(y_2|y_1)\, P(y_3, y_4, \ldots, y_T | y_1, y_2) \\ &= P(y_1)\, P(y_2|y_1)\, P(y_3|y_2) \ldots P(y_t|y_{t-1}) \ldots P(y_T|y_{T-1}). \end{aligned} \qquad (2.19)$$

By using (2.17) and (2.18), this likelihood function can be expressed as

$$\begin{aligned} L(\beta, \psi) &= \lambda_1^{*y_1} (1 - \lambda_1^*)^{(1-y_1)}\, \lambda_{2|1}^{*y_2} (1 - \lambda^*_{2|1})^{(1-y_2)} \ldots \lambda_{T|T-1}^{*y_T} (1 - \lambda^*_{T|T-1})^{(1-y_T)} \\ &= \lambda_1^{*y_1} (1 - \lambda_1^*)^{(1-y_1)} \prod_{t=2}^{T} \lambda_{t|t-1}^{*y_t} (1 - \lambda^*_{t|t-1})^{(1-y_t)} \end{aligned} \qquad (2.20)$$

with $\lambda_1^* = \dfrac{e^{x_1'\beta}}{1 + e^{x_1'\beta}} = \mu_1^*$ and $\lambda^*_{t|t-1} = \mu_t^* + \psi(y_{t-1} - \mu^*_{t-1})$ for $t = 2, \ldots, T$ where $\mu_t^* = \dfrac{e^{x_t'\beta}}{1 + e^{x_t'\beta}}$.

Note that maximization of the likelihood function (2.20) with respect to β and ψ simultaneously is complicated. This is because ψ is restricted to a narrower unknown range than -1 to 1, which may cause non - convergence if this range restriction is not taken care of during the estimation.

2.2.2 Basic Properties of the model: Mean, Variance and Correlation

In the previous subsection we have discussed the limitations of the likelihood estimation approach for estimating the regression and the correlation parameters. To avoid this complexity, as mentioned in section 2.1.2, one may attempt to estimate the β parameter by using a quasilikelihood approach, whereas the ψ parameter may be consistently estimated by the well-known method of moments. The construction of the quasilikelihood estimating equation however requires the computation of the moments for the responses up to order 2. Note that the range restriction problem for the ψ parameter can cause problems also in the quasilikelihood estimation which is not due to the fault of the estimation technique but rather a modelling problem that arises because of the use of conditional linear function to relate the responses.

We now turn back to the computation for the moments up to order 2. These moments are provided in the following two lemmas.

Lemma 2.1: Under the conditional linear dynamic model (2.14), the unconditional mean and variance for the binary responses are given by

$$E(Y_t) = \mu_t^* \text{ and } var(Y_t) = \mu_t^*(1-\mu_t^*) \text{ for all } t = 1, \ldots, T.$$

Proof: By using the conditional mean $E(Y_t|y_{t-1}) = P(Y_t = 1|y_{t-1})$ from (2.18), we compute the unconditional expectation of y_t as

$$\begin{aligned} E(Y_t) = P(y_t = 1) &= \sum_{j=0}^{1} P(Y_t = 1|y_{t-1} = j)\, P(Y_{t-1} = j) \\ &= (\mu_t^* - \psi\mu_{t-1}^*)\,(1-\mu_{t-1}^*) + [\mu_t^* + \psi(1-\mu_{t-1}^*)]\,\mu_{t-1}^* \\ &= \mu_t^*. \end{aligned} \tag{2.21}$$

Since y_t is a binary response, it now follows from (2.21) that

$$var(Y_t) = \mu_t^*(1-\mu_t^*). \tag{2.22}$$

Next, we proceed to compute the correlations of the binary responses $(y_1, \ldots, y_t, \ldots, y_T)$ under the model (2.14). For convenience we provide the formulas for these correlations in lemma 2.2.

Lemma 2.2: For $u < t$, auto-correlations of lag $(t-u)$ for the repeated binary responses $y_1, \ldots, y_t, \ldots, y_T$ under the model (2.14) are given by

$$corr(Y_u, Y_t) = \psi^{t-u}\sqrt{\frac{\mu_u^*(1-\mu_u^*)}{\mu_t^*(1-\mu_t^*)}}. \quad (2.23)$$

Proof: Note that by Lemma 2.1, the variance of y_t is given as $var(Y_t) = \mu_t^*(1-\mu_t^*)$ for all $t = 1, \ldots, T$. Now to verify the correlation formula (2.23), we first compute the $E(Y_uY_t)$ as

$$\begin{aligned}
E(Y_uY_t) &= E_{Y_u}E_{Y_u|Y_{u+1}} \cdots E_{Y_{t-1}|Y_{t-2}}E(Y_u Y_t|Y_{t-1}) \\
&= E_{Y_u}E_{Y_u|Y_{u+1}} \cdots E_{Y_{t-1}|Y_{t-2}}[Y_u(\mu_t^* + \psi(Y_{t-1} - \mu_{t-1}^*))] \\
&= E(Y_u)\ \mu_t^* + \psi[E_{Y_u}E_{Y_u|Y_{u+1}} \cdots E_{Y_{t-1}|Y_{t-2}}(Y_uY_{t-1})] - \psi\mu_{t-1}^*E(Y_u) \\
&= \mu_u^*\ \mu_t^* + \psi E[_{y_u}E_{Y_u|Y_{u+1}} \cdots E(Y_uY_{t-1}|y_{t-2})] - \psi\mu_{t-1}^*\mu_u^* \\
&= \mu_u^*\ \mu_t^* + \psi[E_{y_u}E_{Y_u|Y_{u+1}} \cdots E(Y_u(\mu_{t-1}^* + \psi(Y_{t-2} - \mu_{t-2}^*))] - \psi\mu_{t-1}^*\mu_u^* \\
&= \mu_u^*\ \mu_t^* + \psi\mu_{t-1}^*E(Y_u) + \psi^2[E_{Y_u}E_{Y_u|Y_{u+1}} \cdots E(Y_uY_{t-2})] \\
&\quad -\psi^2\mu_{t-2}^*E(Y_u) - \psi\mu_{t-1}^*\mu_u^* \\
&= \mu_u^*\ \mu_t^* + \psi\mu_{t-1}^*\mu_u^* + \psi^2[E_{Y_u}E_{Y_u|Y_{u+1}} \cdots E(Y_uY_{t-2})] - \psi^2\mu_{t-2}^*\mu_u^* - \psi\mu_{t-1}^*\mu_u^* \\
&= \mu_u^*\ \mu_t^* + \psi^2[E_{Y_u}E_{Y_u|Y_{u+1}} \cdots E(Y_uY_{t-2})] - \psi^2\mu_{t-2}^*\mu_u^*. \quad (2.24)
\end{aligned}$$

By doing some algebra in the manner similar to that of (2.24), we finally obtain

$$\begin{aligned}
E(Y_uY_t) &= \mu_u^*\ \mu_t^* + \psi^{t-u}E_{Y_u}(Y_u^2) - \psi^{t-u}\mu_u^2 \\
&= \mu_u^*\ \mu_t^* + \psi^{t-u}[\mu_u^*(1-\mu_u^*)]\ , \quad (2.25)
\end{aligned}$$

which yields the lag $(t-u)$ auto-covariance as

$$\begin{aligned}
cov(Y_u, Y_t) &= E(Y_uY_t) - E(Y_u)E(Y_t) \\
&= \psi^{t-u}\mu_u^*(1-\mu_u^*). \quad (2.26)
\end{aligned}$$

Thus, the lag $(t-u)$ auto correlation between the binary responses y_u and y_t has the formula

$$corr(Y_u, Y_t) = \psi^{t-u}\sqrt{\frac{\mu_u^*(1-\mu_u^*)}{\mu_t^*(1-\mu_t^*)}}\ , \quad (2.27)$$

as in Lemma 2.2.

Note that unlike the correlated latent process based model discussed in Section 2.1, this conditional linear dynamic model (2.14) provides a simpler correlation structure (2.27). Nevertheless, the correlations may not satisfy the full range from -1 to 1. This is mainly because the ψ parameter in (2.14) is restricted to a narrower range (than -1 to 1) in order to have the probability in (2.18) between 0 and 1. Consequently, if the range restrictions for correlations are not taken into account properly, then this would naturally cause certain convergence problems in the estimation of β regardless of whether the likelihood or the quasilikelihood approach is used. In the following section, we propose an alternative model which does not suffer from this type of range restrictions. Consequently, it is expected that the regression effects will be easily computed consistently and efficiently.

2.3 A Non Linear Dynamic Binary Time Series Model

In Section 2.1, we have shown how to construct a correlated binary model by using a multivariate correlated latent variable which follows a multivariate normal distribution. The limitations of this model were also discussed. In the previous section, we have discussed the conditional linear dynamic model and demonstrated how this linear form can arise by using a uniform distribution based latent process. Note that the inferences for these two models whether one uses the likelihood or quasilikelihood approach are complicated because of the range restrictions on the correlations under these models. These disadvantages compel us to seek a proper correlation structure based alternative model to analyze the correlated binary data. We now discuss a logistic latent process based non-linear dynamic model that accommodates a wider correlation structure as compared to the above two models. This model has been used by many researchers in the econometric literature. For example, see Amemiya(1985, p 422). We however first show how this binary logistic model arises from the logistic distribution based latent process.

Suppose that the latent variable y_1^* follows a logistic distribution with mean $g_1^* = x_1^{'}\beta$ and variance $\frac{\pi^2}{3}$ (Johnson and Kotz (1970)), whereas for $t = 2, \ldots, T$, y_t^* follows the same logistic distribution but with mean $g_t^* = x_t^{'}\beta + \gamma_1 y_{t-1} + \ldots + \gamma_{t-1} y_1$ and variance $\frac{\pi^2}{3}$. To be specific the probability density function of y_1^* and y_t^* $(t = 2, \ldots, T)$ are written as

$$f_L(y_1^*) = \frac{exp[-\{y_1^* - x_1^{'}\beta\}]}{\{1 + exp[-\{y_1^* - x_1^{'}\beta\}]\}^2} \ , \tag{2.28}$$

and

$$f_L(y_t^*) = \frac{exp[-\{y_t^* - g_t^*\}]}{\{1 + exp[-\{y_t^* - g_t^*\}]\}^2} \ , \tag{2.29}$$

respectively.

Now by using the relationship (2.4) between the latent variable y_t^* and the binary variable y_t, one can obtain the marginal probability function for y_1 as,

$$\begin{aligned} P(y_1 = 1) &= \int_{-\infty}^{x_1^{'}\beta} f_L(y_1^*) dy_1^* \\ &= F_L(x_1^{'}\beta) = \frac{exp(x_1^{'}\beta)}{1 + exp(x_1^{'}\beta)} \\ &= p_1 = \lambda_1^* \ , say \ , \end{aligned} \tag{2.30}$$

and for $t = 2, \ldots, T$ the conditional probability for $y_t = 1$ given $y_{t-1}, \ldots, y_1$ as

$$\begin{aligned} P(y_t = 1 | y_{t-1}, \ldots, y_1) &= \int_{-\infty}^{g_t^*} f_L(y_1^*) dy_1^* \\ &= F_L(g_t^*) = \frac{exp(g_t^*)}{[1 + exp(g_t^*)]} \\ &= p_{ty_{t-1}, y_{t-2}, \ldots, y_1} = \tilde{\lambda}_t^* \ , \end{aligned} \tag{2.31}$$

where $g_t^* = x_t^{'}\beta + \gamma_1 y_{t-1} + \ldots + \gamma_{t-1} y_1$.

It is clear from (2.30) and (2.31) that the λ_t^* for all $t = 1, \ldots, T$ ranges between 0 and 1 for any values of $-\infty < \gamma_1, \gamma_2, \ldots, \gamma_{T-1} < \infty$. This implies that there is no range restriction for the dynamic dependence parameters $\gamma_1, \gamma_2, \ldots, \gamma_{T-1}$ under the present non-linear model, whereas similar dependence parameters under the models discussed in Sections 2.1 and 2.2 were restricted to certain narrower ranges. In Subsection 2.3.2 we demonstrate that unlike the correlated latent process based dynamic model (Section 2.1), and the conditional linear dynamic model (Section 2.2), the present model accommodates full ranges for the correlations.

Note that it is in general complicated to deal with the inference for the general full-lagged model (2.30)-(2.31). In many practical situations it may be however sufficient to fit smaller lagged dependent models such as AR(1), AR(2) types. In the following subsection, we provide the likelihood function and basic properties of a lag 1 dependent model.

2.3.1 Likelihood Function

For a lag1 dependent model, that is, where g_t^* in (2.31) depends only on x_t and y_{t-1}, g_t^* is written as

$$g_t^* = x_t'\beta + \gamma_1 y_{t-1} \ , \tag{2.32}$$

leading to the conditional probability function given by,

$$\begin{aligned} P(y_t = 1|y_{t-1}) &= \frac{exp(x_t'\beta + \gamma_1 y_{t-1})}{[1 + exp(x_t'\beta + \gamma_1 y_{t-1})]} \\ &= p_{ty_{t-1}} = \lambda_t^*. \end{aligned} \tag{2.33}$$

Note that it follows from the model (2.33) that the time series responses $(y_1, \ldots, y_t, \ldots, y_T)$ are correlated. Consequently, their likelihood function may be written as

$$\begin{aligned} L(\beta\ , \gamma_1) &= P(y_1, y_2, \ \ldots\ , \ y_t, \ \ldots\ , y_T) \\ &= P(y_1)P(y_2, y_3, \ \ldots\ , y_T|y_1) \\ &= P(y_1)P(y_2|y_1)P(y_3, y_4, \ \ldots\ , y_T|y_1, y_2) \\ &= P(y_1)P(y_2|y_1)P(y_3|y_2) \ldots P(y_t|y_{t-1}) \ldots P(y_T|y_{T-1}), \end{aligned} \tag{2.34}$$

where by (2.33), the conditional probability $P(y_t|y_{t-1})$ is given by

$$P(y_t|y_{t-1}) = p_{ty_{t-1}}^{y_t}(1 - p_{ty_{t-1}})^{1-y_t}, \tag{2.35}$$

where $p_{ty_{t-1}}$ is as in (2.33). Thus, we re-write the likelihood function (2.34) as

$$L(\beta\,,\gamma_1) \;=\; {p_1}^{y_1}(1-p_1)^{(1-y_1)}\prod_{t=2}^{T} p_{t\,y_{t-1}}^{y_t}(1-p_{t\,y_{t-1}})^{(1-y_t)} \qquad (2.36)$$

with $p_1 = exp(x_1^{'}\beta)/[1+exp(x_1^{'}\beta)]$ and $p_{ty_{t-1}} = exp(x_t^{'}\beta+\gamma_1 y_{t-1})/[1+exp(x_t^{'}\beta+\gamma_1 y_{t-1})]$ for $t\ = 2,\ \ldots\ ,T$.

Note that the conditional probability function $p_{ty_{t-1}}$ contains a lag1 dependence parameter γ_1. It is clear from their relationship that $p_{ty_{t-1}}$ satisfies the 0 to 1 range for any value of the dependence/correlation parameter $-\infty < \gamma_1 < \infty$. This is in contrast to the other two models discussed in Section 2.1 and 2.2, a big advantage of this model over the other two models from a data analysis point of view.

As far as the likelihood inference for the data is concerned in Chapter 3, we discuss the maximization of the likelihood function (ML) (2.36) with respect to β and γ_1. We will also consider an alternative GQL estimation approach and compare the relative performances of the GQL and the ML approaches.

2.3.2 Basic Properties of the Non-Linear Dynamic Model: Mean, Variance and Correlation

In Section 2.1 and 2.2 we have provided the likelihood functions of the correlated binary data under two specific models. Note however that to understand the basic nature of the data, it is useful to know the mean, variance and correlation structure under a specific model. This type of lower order moments may also be useful to develop simpler estimation technique, such as the GQL (mentioned in section 2.1.2 and 2.2.2) estimation approach, as an alternative to the ML approach. For the purpose, we now provide these moments (see also Farrell and Sutradhar (2006)) under the model (2.33) in Lemmas 2.3 and 2.4 below. These moments will be used in chapter 3 for the construction of the GQL estimating equation.

Lemma 2.3: Under the model (2.33), y_1 has the mean $\mu_1 = E(Y_1) = exp(x_1^{'}\beta)/[1+exp(x_1^{'}\beta)]$

and the mean of y_t for $t = 2, \ldots, T$, has a recurrence relationship given by

$$\mu_t = E(Y_t) \;=\; p_{t0} + \mu_{t-1}(p_{t1} - p_{t0}), \tag{2.37}$$

where $p_{t1} = exp(x_t^{'}\beta + \gamma_1)/[1 + exp(x_t^{'}\beta + \gamma_1)]$ and $p_{t0} = exp(x_t^{'}\beta)/[1 + exp(x_t^{'}\beta)]$. Furthermore, the variance of y_t *for all* $t = 1, \ldots, T$ is given by

$$var(Y_t) = \sigma_{tt} = \mu_t\,(1 - \mu_t). \tag{2.38}$$

Proof: It is obvious from (2.31) that $\mu_1 = P(y_1 = 1) = exp(x_1^{'}\beta)/[1 + exp(x_1^{'}\beta)]$. Next, by (2.33) we obtain

$$\begin{aligned} \mu_2 = P(y_2 = 1) &= \sum_{j=0}^{1} P(Y_2 = 1 | y_1 = j)\; P(Y_1 = j) \\ &= p_{20}\,(1 - \mu_1) + p_{21}\,\mu_1 \\ &= p_{20} + \mu_1(p_{21} - p_{20}) \end{aligned}$$

with $p_{21} = exp(x_2^{'}\beta + \gamma_1)/[1 + exp(x_2^{'}\beta + \gamma_1)]$ and $p_{20} = exp(x_2^{'}\beta)/[1 + exp(x_2^{'}\beta)]$. Similarly, we obtain

$$\begin{aligned} \mu_3 = P(y_3 = 1) &= \sum_{j=0}^{1} P(Y_3 = 1 | y_2 = j)\; P(Y_2 = j) \\ &= p_{30}\,(1 - \mu_2) + p_{31}\,\mu_2 \\ &= p_{30} + \mu_2(p_{31} - p_{30}) \end{aligned}$$

with $p_{31} = exp(x_3^{'}\beta + \gamma_1)/[1 + exp(x_3^{'}\beta + \gamma_1)]$ and $p_{30} = exp(x_2^{'}\beta)/[1 + exp(x_2^{'}\beta)]$. By following this pattern, we write

$$\begin{aligned} \mu_t = P(y_t = 1) &= \sum_{j=0}^{1} P(Y_t = 1 | y_{t-1} = j)\; P(Y_{t-1} = j) \\ &= p_{t0}(1 - \mu_{t-1}) + p_{t1}\mu_{t-1} \\ &= p_{t0} + \mu_{t-1}(p_{t1} - p_{t0}) \end{aligned} \tag{2.39}$$

with $p_{t1} = exp(x_t^{'}\beta + \gamma_1)/[1 + exp(x_t^{'}\beta + \gamma_1)]$ and $p_{t0} = exp(x_t^{'}\beta)/[1 + exp(x_t^{'}\beta)]$. This provides the proof for the first part of the lemma.

Next, as y_t is a binary response with $P(y_t = 1) = \mu_t$ (2.39), it then follows that

$$var(Y_t) \;=\; \mu_t\,(1 - \mu_t) \tag{2.40}$$

which proves the second part of the lemma. This completes the proof for Lemma 2.3.

Note that there is a big difference in the formulas for the means provided by this non-linear dynamic model and the other two models presented in the previous subsections. To be specific the present model has a dynamic relationship for the means of the data, whereas the other two models provide the marginal means. The variances of these three models may also be interpreted similarly. To examine the difference in correlation forms, we now compute the correlations for the present non-linear dynamic model as in the following lemma.

Lemma 2.4 For $u < t$, the auto-correlations of lag $t - u$ for the repeated binary responses $y_1, \ldots, y_t, \ldots, y_T$ under the non linear dynamic model (2.31) are given by

$$corr(Y_u, Y_t) = \sqrt{\frac{\mu_u(1-\mu_u)}{\mu_t(1-\mu_t)}} \prod_{j=u+1}^{t} (p_{j1} - p_{j0}), \tag{2.41}$$

where $p_{j1} = exp(x_j'\beta + \gamma_1)/[1 + exp(x_j'\beta + \gamma_1)]$ and $p_{j0} = exp(x_j'\beta)/[1 + exp(x_j'\beta)]$ by (2.33).

Proof: For $u < t$, we first derive the lag $t - u$ auto-covariance as

$$cov(Y_u, Y_t) = \mu_u(1-\mu_u) \prod_{j=u+1}^{t} (p_{j1} - p_{j0}). \tag{2.42}$$

This result may be verified by induction. To begin with, we compute the two lag-1 auto-covariances, $cov(Y_1, Y_2)$ and $cov(Y_2, Y_3)$ as follows. To be specific, for the computation of $cov(Y_1, Y_2)$ we use

$$cov(Y_1, Y_2) = E(Y_1Y_2) - E(Y_1)E(Y_2)$$

where

$$\begin{aligned} E(Y_1Y_2) &= P(y_2 = 1, y_1 = 1) \\ &= P(y_2 = 1 | y_1 = 1)P(y_1 = 1) \\ &= p_{21}\mu_1, \end{aligned}$$

This leads to the covariance of Y_1 and Y_2 as

$$\begin{aligned} cov(Y_1, Y_2) &= p_{21}\mu_1 - \mu_1[p_{20} + \mu_1(p_{21} - p_{20})] \\ &= \mu_1(1-\mu_1)(p_{21} - p_{20}) \\ &= \mu_1(1-\mu_1)\prod_{j=2}^{2}(p_{j1} - p_{j0}). \end{aligned} \tag{2.43}$$

Next, to compute $cov(Y_2, Y_3)$, we first compute

$$\begin{aligned} E(Y_2Y_3) &= P(y_3 = 1, y_2 = 1) \\ &= P(y_3 = 1|y_2 = 1)P(y_2 = 1) \\ &= p_{31}\mu_2, \end{aligned}$$

which yields the covariance of Y_2 and Y_3 as

$$\begin{aligned} cov(Y_2, Y_3) &= p_{31}\mu_2 - \mu_2[p_{30} + \mu_2(p_{31} - p_{30})] \\ &= \mu_2(1-\mu_2)(p_{31} - p_{30}) \\ &= \mu_2(1-\mu_2)\prod_{j=3}^{3}(p_{j1} - p_{j0}). \end{aligned} \tag{2.44}$$

Similarly, to compute the lag-2 auto-covariances, $cov(Y_1, Y_3)$ and $cov(Y_2, Y_4)$, we first compute

$$\begin{aligned} E(Y_1Y_3) &= P(y_3 = 1, y_1 = 1) \\ &= P(y_3 = 1|y_2 = 0)P(y_2 = 0|y_1 = 1)P(y_1 = 1) + \\ &\quad P(y_3 = 1|y_2 = 1)P(y_2 = 1|y_1 = 1)P(y_1 = 1) \\ &= p_{30}(1 - p_{21})\mu_1 + p_{31}p_{21}\mu_1, \end{aligned}$$

and

$$\begin{aligned} E(Y_2Y_4) &= P(y_2 = 1, y_4 = 1) \\ &= P(y_4 = 1|y_3 = 0)P(y_3 = 0|y_2 = 1)P(y_2 = 1) + \\ &\quad P(y_4 = 1|y_3 = 1)P(y_3 = 1|y_2 = 1)P(y_2 = 1) \\ &= p_{40}(1 - p_{31})\mu_2 + p_{41}p_{31}\mu_2, \end{aligned}$$

yielding

$$cov(Y_1, Y_3) = p_{30}(1 - p_{21})\mu_1 + p_{31}p_{21}\mu_1 - \mu_1[p_{30} + \mu_2(p_{31} - p_{30})], \quad (2.45)$$

and

$$cov(Y_2, Y_4) = p_{40}(1 - p_{31})\mu_2 + p_{41}p_{31}\mu_2 - \mu_2[p_{40} + \mu_3(p_{41} - p_{40})], \quad (2.46)$$

respectively.

Next by using the recurrence relationship for μ_2 and μ_3 given in Lemma 2.3, the above two equations (2.45) and (2.46) reduce to the formulas given by

$$cov(Y_1, Y_3) = \mu_1(1 - \mu_1)\prod_{j=2}^{3}(p_{j1} - p_{j0}), \quad (2.47)$$

and

$$cov(Y_2, Y_4) = \mu_2(1 - \mu_2)\prod_{j=3}^{4}(p_{j1} - p_{j0}), \quad (2.48)$$

respectively.

Next, we also check the formulas for two lag-3 auto-covariances, namely $cov(Y_2, Y_5)$ and $cov(Y_3, Y_6)$. To compute the covariance of Y_2 and Y_5, we first compute

$$\begin{aligned} E(Y_2Y_5) &= P(y_2 = 1, y_5 = 1) \\ &= P(y_5 = 1|y_4 = 0)P(y_4 = 0|y_3 = 1)P(y_3 = 1|y_2 = 1)P(y_2 = 1) + \\ &\quad P(y_5 = 1|y_4 = 0)P(y_4 = 0|y_3 = 0)P(y_3 = 0|y_2 = 1)P(y_2 = 1) + \\ &\quad P(y_5 = 1|y_4 = 1)P(y_4 = 1|y_3 = 1)P(y_3 = 1|y_2 = 1)P(y_2 = 1) + \\ &\quad P(y_5 = 1|y_4 = 1)P(y_4 = 1|y_3 = 0)P(y_3 = 0|y_2 = 1)P(y_2 = 1) \\ &= p_{50}(1 - p_{41})p_{31}\mu_2 + p_{50}(1 - p_{40})(1 - p_{31})\mu_2 + \\ &\quad p_{51}p_{41}p_{31}\mu_2 + p_{51}p_{40}(1 - p_{31})\mu_2 \\ &= \mu_2[p_{50}(1 - p_{40})(1 - p_{31}) + p_{51}p_{40}(1 - p_{31}) + p_{50}(1 - p_{41})p_{31} + p_{51}p_{41}p_{31}], \end{aligned}$$

which produces the covariance

$$\begin{aligned} cov(Y_2, Y_5) &= \mu_2[p_{50}(1 - p_{40})(1 - p_{31}) + p_{51}p_{40}(1 - p_{31}) + p_{50}(1 - p_{41})p_{31} + p_{51}p_{41}p_{31}] - \\ &\quad \mu_2\mu_5. \end{aligned} \quad (2.49)$$

Now, using the recurrence relationship for μ_5 from lemma 2.3 in equation (2.49) and the recurrence relationship for μ_4 and μ_3 from the same lemma 2.3, we get the following form

$$\begin{aligned} cov(Y_2, Y_5) &= \mu_2(1-\mu_2)(p_{31}-p_{30})(p_{41}-p_{40})(p_{51}-p_{50}) \\ &= \mu_2(1-\mu_2)\prod_{j=3}^{5}(p_{j1}-p_{j0}). \end{aligned} \tag{2.50}$$

In the similar fashion, we compute the formula for $cov(Y_3, Y_6)$. For this,

$$\begin{aligned} E(Y_3Y_6) &= P(y_3=1, y_6=1) \\ &= P(y_6=1|y_5=0)P(y_5=0|y_4=1)P(y_4=1|y_3=1)P(y_3=1)+ \\ &\quad P(y_6=1|y_5=0)P(y_5=0|y_4=0)P(y_4=0|y_3=1)P(y_3=1)+ \\ &\quad P(y_6=1|y_5=1)P(y_5=1|y_4=1)P(y_4=1|y_3=1)P(y_3=1)+ \\ &\quad P(y_6=1|y_5=1)P(y_5=1|y_4=0)P(y_4=0|y_3=1)P(y_3=1) \\ &= p_{60}(1-p_{51})p_{41}\mu_3 + p_{60}(1-p_{50})(1-p_{41})\mu_3 + \\ &\quad p_{61}p_{51}p_{41}\mu_3 + p_{61}p_{50}(1-p_{41})\mu_3 \\ &= \mu_3[p_{60}(1-p_{50})(1-p_{41}) + p_{61}p_{50}(1-p_{41}) + p_{60}(1-p_{51})p_{41} + p_{61}p_{51}p_{41}], \end{aligned}$$

which provides the covariance

$$\begin{aligned} cov(Y_3, Y_6) &= \mu_3[p_{60}(1-p_{50})(1-p_{41}) + p_{61}p_{50}(1-p_{41}) + p_{60}(1-p_{51})p_{41} + p_{61}p_{51}p_{41}] \\ &\quad -\mu_3\mu_6. \end{aligned} \tag{2.51}$$

Now using the recurrence relationship for μ_6 from lemma 2.3 in equation (2.51) and the recurrence relationship for μ_5 and μ_4 from the same lemma 2.3, we obtain the formula for the covariance

$$\begin{aligned} cov(Y_3, Y_6) &= \mu_3(1-\mu_3)(p_{41}-p_{40})(p_{51}-p_{50})(p_{61}-p_{60}) \\ &= \mu_3(1-\mu_3)\prod_{j=4}^{6}(p_{j1}-p_{j0}). \end{aligned} \tag{2.52}$$

It is clear from (2.43), (2.44); (2.47), (2.48); and (2.50), (2.52) that for $u < t$ we may write the formula for the lag(t - u) auto-covariance between Y_u and Y_t as

$$cov(Y_u, Y_t) = \mu_u(1 - \mu_u) \prod_{j=u+1}^{t} (p_{j1} - p_{j0}), \tag{2.53}$$

where, similar to (2.39), p_{j1} and p_{j0} are defined as $p_{j1} = exp(x_j^{'}\beta + \gamma_1)/[1 + exp(x_j^{'}\beta + \gamma_1)]$ and $p_{j0} = exp(x_j^{'}\beta)/[1 + exp(x_t^{'}\beta)]$, respectively. It then follows that for $u < t$, the correlation between Y_u and Y_t is given by

$$\begin{aligned} corr(Y_u, Y_t) &= \frac{cov(Y_u, Y_t)}{\sqrt{var(Y_u)\ var(Y_t)}} \\ &= \frac{\mu_u(1 - \mu_u)\ \prod_{j=u+1}^{t}(p_{j1} - p_{j0})}{\sqrt{\mu_u(1 - \mu_u)\ \mu_t(1 - \mu_t)}} \\ &= \sqrt{\frac{\mu_u(1 - \mu_u)}{\mu_t(1 - \mu_t)}} \prod_{j=u+1}^{t} (p_{j1} - p_{j0}). \end{aligned} \tag{2.54}$$

This completes the proof of Lemma 2.4.

Note that in this chapter we have discussed three binary time series models. Among these models, the correlated latent process based binary time series model discussed in Section 2.1 provides a complicated correlation structure (2.13). The conditional linear dynamic based model discussed in Section 2.2 gives a simpler correlation structure (2.27) but the correlations are found to satisfy narrower range as compared to the full range -1 to 1. The non linear dynamic model discussed in this section appears to be simple to interpret and allows for correlations covering the full range unlike the other two models. For a detailed numerical comparison on the ranges of the correlations under different models including the present non-linear dynamic model, we refer to Farrell and Sutradhar (2006). Note that there is, however, no adequate discussion in the literature on the statistical inferences based on this simpler and attractive time series model. This motivated us to seek estimation approaches for the parameters of this model which we present in the next chapter. In Chapter 4, we discuss forecasting aspects of this model.

Chapter 3

Estimation of the Model Parameters of Lag 1 Model

Recall that the non-linear dynamic model was discussed in Section 2.3. For convenience, we re-write this model here as follows.

$$\begin{aligned} P(y_1 = 1) &= \frac{exp(x_1'\beta)}{[1 + exp(x_1'\beta)]} \\ &= \lambda_1^*. \end{aligned} \tag{3.1}$$

and

$$\begin{aligned} P(y_t = 1|y_{t-1}) &= \frac{exp(x_t'\beta + \gamma_1 y_{t-1})}{[1 + exp(x_t'\beta + \gamma_1 y_{t-1})]} \\ &= p_{ty_{t-1}} = \lambda_t^*. \end{aligned} \tag{3.2}$$

It is of interest to estimate the regression effect β and the lag 1 dependence parameterγ_1. For this purpose, we choose to explore the performance of the traditional maximum likelihood (ML) approach and a less familiar generalized quasi likelihood (GQL) approach for the estimation of these parameters.

In Section 3.1, we conduct an initial study to examine the performances of these two approaches in estimating β only by using a "working" independence approach which treats the data as independent even though data are truly correlated. As we explain below, it follows that when data are generated independently, these "working" independence assumption based ML and GQL approaches work quite well in estimating the regression parameters, whereas the performance of these approaches is extremely poor for the estimation of the same parameters where the data are truly correlated.

3.1 A "working " Independence Approach

Let $y_1, \ldots, y_t, \ldots, y_T$ be a time series of binary observations following the model (3.1- 3.2) with $x_t = (x_{t1}, \ldots, x_{tj}, \ldots, x_{tp})'$ as a vector of p-dimensional covariates associated with y_t. To estimate the regression parameter β, in this subsection we treat the observations as independent and develop the likelihood as well as the GQL estimating equations, even though data are generated with certain correlations represented by non-zero values of γ_1. That is, data are generated with $\gamma_1 \neq 0$ but estimation will be done by treating $\gamma_1 = 0$.

3.1.1 Likelihood Estimation Equation

As we are treating the observations as independent, we may write the "working" likelihood function as,

$$L_I(\beta|y) = \prod_{t=1}^{T} f(y_t) = \prod_{t=1}^{T} p_t(I)^{y_t}(1 - p_t(I))^{(1-y_t)}, \quad (3.3)$$

with $\mu_t(I) = p_t(I) = exp(x_t' \beta)/[1 + exp(x_t' \beta)]$. Here $L_I(\beta|y)$ represents an independence(I) assumption based likelihood function. The log-likelihood function is then written as

$$l_I(\beta|y) = logL_I(\beta|y) = \sum_{t=1}^{T} (y_t \log p_t(I) + (1 - y_t) \log(1 - p_t(I))),$$

yielding the derivative function as

$$\frac{\partial l_I(\beta|y)}{\partial \beta} = \sum_{t=1}^{T} \left(y_t \frac{p_t(I)(1 - p_t(I))}{p_t(I)} x_t + (1 - y_t) \frac{(1 - p_t(I))}{(1 - p_t(I))} p_t(I)(-x_t) \right)$$

$$= \sum_{i=1}^{T}(y_t x_t(1-p_t(I)) - (1-y_t)p_t(I)x_t)$$
$$= \sum_{t=1}^{T}(y_t - p_t(I))x_t. \quad (3.4)$$

The likelihood estimating equation is,

$$\frac{\partial l_I(\beta|y)}{\partial \beta} = 0. \quad (3.5)$$

By using a Taylor's series expansion, it follows from (3.5) that the β parameter may be estimated by using the iterative equation ,

$$\hat{\beta}_{WML}(r+1) = \hat{\beta}_{WML}(r) + \left[[\frac{\partial^2 l_I(\beta|y)}{\partial\beta\partial\beta'}]^{-1} \frac{\partial l_I(\beta|y)}{\partial\beta} \right]_{\hat{\beta}_{WML}(r)}, \quad (3.6)$$

where $\hat{\beta}_{WML}(r)$, the "working " maximum likelihood estimate of β is a solution of (3.4) at the r-th interation , and $[.]_{\hat{\beta}_{WML}(r)}$is the value of the expression in the square bracket evaluated at $\beta = \hat{\beta}_{WML}(r)$. Note that in (3.6), the first derivative $\frac{\partial l_I(\beta|y)}{\partial\beta}$ has the formula given by (3.4). Similarly one may obtain the second derivative as $\frac{\partial^2 l_I(\beta|y)}{\partial\beta\partial\beta'} = -\sum_{t=1}^{T} x_t p_t(I)(1-p_t(I))x_t'$.

3.1.2 "Working "Generalized Quasilikelihood Estimation for the Regression Parameter β

Let $\mu(I) = (\mu_1(I), \ldots, \mu_T(I))$ be the T- dimensional mean vector of $y = (y_1, \ldots, y_t, \ldots, y_T)$ and $\Sigma(I) = (\sigma_{tt'}(I))$ be the covariance matrix of y. Here $\sigma_{tt}(I) = var(Y_t|\gamma_1 = 0)$ and $\sigma_{tt'}(I) = cov(Y_t, Y_{t'}|\gamma_1 = 0)$. Note that covariance of Y_t and $Y_{t'}$ is zero, because the binary observations are assumed to be independent. Thus, $\Sigma(I)$ has only the diagonal elements.

One may write the generalized quasilikelihood(GQL) estimating equation for β as

$$\frac{\partial\mu'}{\partial\beta}(I)\Sigma^{-1}(I)(y - \mu(I)) = 0. \quad (3.7)$$

Let $\hat{\beta}_{WGQL}$ be the GQL estimator of β obtained from (3.7). Note that the computation of $\hat{\beta}_{WGQL}$ is usually done by an iterative method. More specifically, $\hat{\beta}_{WGQL}$ is obtained by using the Newton-Raphson iterative equations

$$\hat{\beta}_{WGQL}(r+1) = \hat{\beta}_{WGQL}(r) + \left[\left(\frac{\partial \mu'(I)}{\partial \beta}\Sigma^{-1}(I)\frac{\partial \mu(I)}{\partial \beta}\right)^{-1} \frac{\partial \mu(I)}{\partial \beta}\Sigma^{-1}(I)(y-\mu(I))\right]_{\hat{\beta}_{WGQL}(r)}. \tag{3.8}$$

Note that in (3.8) to compute $\frac{\partial \mu'(I)}{\partial \beta}$, we first compute $\frac{\partial \mu^t(I)}{\partial \beta} = (x_{t1}\mu_t(I)(1-\mu_t)(I), \ldots, x_{tp}\mu_t(I)(1-\mu_t(I)))'$, which yields $\frac{\partial \mu'(I)}{\partial \beta} = X'A(I)$ where $X = (x_1, x_2, \ldots, x_t, \ldots x_T)'$ with $x_t = (x_{t1}, \ldots, x_{tp})'$, and $A(I) = diag[\mu_1(I)(1-\mu_1(I)), \ldots, \mu_T(I)(1-\mu_T(I))]$.

It then follows that for computational convenience the WGQL estimating equation (3.8) may be re-expressed as

$$\hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) + \left[\left(X'(I)A(I)\Sigma^{-1}(I)A(I)X(I)\right)^{-1} X'(I)A(I)\Sigma^{-1}(I)(y-\mu(I))\right]_{\hat{\beta}_{GQL}(r)}. \tag{3.9}$$

3.1.3 A Simulation Study

In this subsection we conduct a simulation study to examine the performance of $\hat{\beta}_{WML}$ and $\hat{\beta}_{WGQL}$ to estimate the β parameter of the true model. Recall that under the true model the responses are correlated, whereas $\hat{\beta}_{WML}$ and $\hat{\beta}_{WGQL}$ are computed by treating the responses as independent.

Under the true time series model (3.1) - (3.2), we generated $y_1, \ldots, y_t, \ldots, y_T$ for T = 100, 200 and 300 with p = 2 covariates such that $\beta_1 = \beta_2 = 0.5$, and with dependence parameter value as $\gamma_1 = -1, 0, 1$ for a given design matrix X. As far as the choice of this X matrix is concerned we consider the following four designs. Two covariates were chosen under each of the four designs.

D_1 (Design 1) :

$$x_{i1} = 1 \text{ for } i = 1.\ldots.T$$

and

$$x_{i2} = \begin{cases} -1 & \text{for } 0 < i < T/4 \\ 0 & \text{for } T/4 < i < 3T/4 \\ 1 & \text{otherwise.} \end{cases}$$

D_2 (Design 2) :

$$x_{i1} = 1 \text{ for } i = 1.\ldots.T.$$

and

$$x_{i2} = i/T \text{ for } i = 1.\ldots.T.$$

D_3 (Design 3) :

$$x_{i1} = 1 \text{ for } i = 1.\ldots.T.$$

and

$$x_{i2} = \begin{cases} 0.01 & \text{for i} = 1 \\ x(i-1,2) + 0.01 & \text{for } 1 < i < T/4 \\ x(i-1,2) + 0.05 & \text{for } T/4 < i < 3T/4 \\ x(i-1,2) + 0.10 & \text{for } 3T/4 < i < T. \end{cases}$$

D_4 (Design 4) :

$$x_{i1} = \begin{cases} 0 & \text{for } 0 < i < T/2 \\ 1 & \text{otherwise.} \end{cases}$$

and

$$x_{i2} = \begin{cases} -1 & \text{for } 0 < i < T/4 \\ 0 & \text{for } T/4 < i < 3T/4 \\ 1 & \text{otherwise.} \end{cases}$$

Now by using the generated $y = (y_1, \ldots, y_t, \ldots, y_T)$ and the design matrix constructed under the chosen design, we apply (3.6) and (3.9) to obtain the WML and WGQL estimates of β_1 and β_2. All together we consider 1000 simulations. The average values for each of the components of $\hat{\beta}_{WML} = (\hat{\beta}_{WML,1}, \hat{\beta}_{WML,2})$ and $\hat{\beta}_{WGQL} = (\hat{\beta}_{WGQL,1}, \hat{\beta}_{WGQL,2})$ based on 1000 simulations are reported in Tables 3.1, 3.2 and 3.3 for γ_1= -1, 0 and 1 respectively. The standard errors and mean squared errors of these estimates are also reported in the same tables.

Table 3.1: Simulated mean (SM), simulated standard error (SSE) and simulated mean squared error (SMSE) of the WML (Independence based) and WGQL(Independence based) estimates for regression coefficients with T = 100, 200, 300; $\beta_1 = \beta_2 = 0.5$; based on 1000 simulations, for the case when γ_1= -1, 0, 1

					Estimates	
γ_1	Size T	Design	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$
-1	100	D_1	WML/WGQL	SM	0.006	0.394
				SSE	0.159	0.230
				SMSE	0.269	0.064
		D_2	WML/WGQL	SM	0.002	0.388
				SSE	0.310	0.534
				SMSE	0.344	0.298
		D_3	WML/WGQL	SM	-0.024	0.427
				SSE	0.248	0.140
				SMSE	0.336	0.025
		D_4	WML/WGQL	SM	-0.115	0.650
				SSE	0.274	0.266
				SMSE	0.453	0.093
	200	D_1	WML/WGQL	SM	-0.001	0.391
				SSE	0.112	0.166
				SMSE	0.264	0.039
		D_2	WML/WGQL	SM	-0.001	0.379
				SSE	0.221	0.394
				SMSE	0.300	0.170
		D_3	WML/WGQL	SM	-0.045	0.440
				SSE	0.198	0.084
				SMSE	0.336	0.011
		D_4	WML/WGQL	SM	-0.123	0.646
				SSE	0.179	0.199
				SMSE	0.420	0.061

(Table 3.1 Contd....)

γ_1	Size T	Design	Method	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$
	300	D_1	WML/WGQL	SM	0.001	0.382
				SSE	0.096	0.133
				SMSE	0.258	0.032
		D_2	WML/WGQL	SM	0.001	0.386
				SSE	0.182	0.324
				SMSE	0.282	0.118
		D_3	WML/WGQL	SM	-0.053	0.445
				SSE	0.177	0.068
				SMSE	0.337	0.008
		D_4	WML/WGQL	SM	-0.129	0.644
				SSE	0.153	0.159
				SMSE	0.419	0.046
0	100	D_1	WML/WGQL	SM	0.524	0.537
				SSE	0.222	0.327
				SMSE	0.050	0.108
		D_2	WML/WGQL	SM	0.510	0.554
				SSE	0.448	0.783
				SMSE	0.201	0.616
		D_3	WML/WGQL	SM	0.500	0.546
				SSE	0.374	0.225
				SMSE	0.140	0.053
		D_4	WML/WGQL	SM	0.531	0.534
				SSE	0.380	0.369
				SMSE	0.145	0.137

(Table 3.1 Contd....)

γ_1	Size T	Design	Method	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$
	200	D_1	WML/WGQL	SM	0.513	0.524
				SSE	0.148	0.232
				SMSE	0.022	0.054
		D_2	WML/WGQL	SM	0.516	0.515
				SSE	0.299	0.524
				SMSE	0.089	0.275
		D_3	WML/WGQL	SM	0.500	0.529
				SSE	0.287	0.138
				SMSE	0.082	0.020
		D_4	WML/WGQL	SM	0.514	0.521
				SSE	0.246	0.250
				SMSE	0.061	0.063
	300	D_1	WML/WGQL	SM	0.510	0.513
				SSE	0.123	0.177
				SMSE	0.015	0.031
		D_2	WML/WGQL	SM	0.513	0.510
				SSE	0.254	0.446
				SMSE	0.064	0.199
		D_3	WML/WGQL	SM	0.499	0.524
				SSE	0.254	0.118
				SMSE	0.064	0.015
		D_4	WML/WGQL	SM	0.509	0.513
				SSE	0.201	0.203
				SMSE	0.040	0.041

(Table 3.1 Contd....)

γ_1	Size T	Design	Method	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$
1	100	D_1	WML/WGQL	SM	1.280	0.642
				SSE	0.325	0.455
				SMSE	0.714	0.227
		D_2	WML/WGQL	SM	1.244	0.709
				SSE	0.644	1.151
				SMSE	0.968	1.368
		D_3	WML/WGQL	SM	1.230	0.728
				SSE	0.543	0.858
				SMSE	0.828	0.788
		D_4	WML/WGQL	SM	1.510	0.180
				SSE	0.553	0.476
				SMSE	1.326	0.329
	200	D_1	WML/WGQL	SM	1.250	0.622
				SSE	0.223	0.304
				SMSE	0.612	0.107
		D_2	WML/WGQL	SM	1.235	0.643
				SSE	0.430	0.765
				SMSE	0.725	0.606
		D_3	WML/WGQL	SM	1.231	0.646
				SSE	0.425	0.371
				SMSE	0.715	0.159
		D_4	WML/WGQL	SM	1.454	0.161
				SSE	0.365	0.314
				SMSE	1.043	0.214

(Table 3.1 Contd....)

γ_1	Size T	Design	Method	Quantity	Estimates $\hat{\beta}_1$	$\hat{\beta}_2$
	300	D_1	WML/WGQL	SM	1.240	0.613
				SSE	0.172	0.240
				SMSE	0.577	0.070
		D_2	WML/WGQL	SM	1.230	0.630
				SSE	0.358	0.633
				SMSE	0.661	0.418
		D_3	WML/WGQL	SM	1.230	0.622
				SSE	0.386	0.263
				SMSE	0.682	0.084
		D_4	WML/WGQL	SM	1.448	0.147
				SSE	0.290	0.262
				SMSE	0.983	0.193

The results of Tables 3.1 show that when data are generated with certain correlations ($\gamma_1 = -1, \gamma_1 = 1$)the independent assumption based WML and WGQL approaches perform very poorly in estimating both β_1 and β_2. To be specific, when data are generated with γ_1= -1, the results of the table show that the independence assumption based approaches grossly underestimate β_1 and β_2 for all designs except D_4. For the cases with D_4, WML and WGQL approaches overestimate β_2 and highly underestimate β_1. For example, when $\gamma_1 = -1$, WML and WGQL produce the same mean estimates of β_1 and β_2 as 0.001 and 0.382 under D_1 with T = 300, whereas β_1 and β_2 have the true values as $\beta_1 = \beta_2 = 0.5$. When T = 300, under D_4, the mean estimates of β_1 and β_2 produced by WML and WGQL are found to be -0.129 and 0.644 for the same true values of β_1 and β_2, that is for $\beta_1 = \beta_2 = 0.5$. It is therefore clear that when data are correlated (with $\gamma_1 = -1$) but it is attempted to obtain the estimates by assuming independence, the estimates are bound to be poor. Similar results hold for the case with $\gamma_1 = 1$. In this case both WML and WGQL approaches generally overestimate both β_1 and β_2 for D_1, D_2 and D_3, whereas for D_4, β_1 is overestimated and β_2 is highly underestimated.

When the data are generated under the independence condition ($\gamma_1 = 0$) and the estimation approaches are also based on the independence assumption, the WML and WGQL approaches perform extremely well in estimating β_1 and β_2. For example, when T = 300, under D_1, the mean estimates of β_1 and β_2 are found to be 0.510 and 0.513 with standard errors 0.123 and 0.177 respectively, whereas $\beta_1 = \beta_2 = 0.5$ truly. These results therefore indicate that the ML and GQL approaches may perform well in the dependence case provided the correlation parameter γ_1 is accounted for while estimating β_1 and β_2.The results for this case with $\gamma_1 = 0$ also show that as T increases, the WML and WGQL approaches perform better in estimating the parameters.

In summary, when the data are truly independent and parameters are estimated by the "working "ML and GQL approaches under an independence assumption, these approaches work very well, whereas when data are generated with correlation, the "working "independence assumption based WML and WGQL approaches provide poor estimates. This indicates that it is important to estimate β_1 and β_2 after taking the correlations (γ_1) into account. The purpose of the next section is to estimate β_1 and β_2 as well as γ_1 such that the estimates of β_1 and β_2 are obtained by taking the correlations into account.

3.2 True Correlation Structure Based Approach

The model (3.2) involves two unknown parameters: (i) β, the p-dimensional vector of regression parameters and (ii)γ_1, the correlation or dependence parameter. In the last section, we have estimated β by treating $\gamma_1 = 0$, even though the data may be correlated. It was found that when data were generated with $\gamma_1 \neq 0$, the independence assumption based ML (i.e WML) and GQL (i.e WGQL) approaches yield highly biased, i.e, inconsistent estimates. We also emphasize that under the present non-linear dynamic model, γ_1 along with β is also an important parameter. This is because all of these parameters are involved in the expressions for the mean and variance of the responses. The other higher order moments also will be functions of these parameters. Consequently, any estimation approaches constructed by ignoring γ_1 will not produce good estimators for β. This motivates us to estimate all parameters β and γ_1 simultaneously. To be specific, we consider the well-known ML and different versions of the GQL approach to estimate these parameters. We also refer to Sutradhar and Farrell (2006) where some of these approaches were used in the context of longitudinal data analysis, whereas we consider the estimation problem for binary time series.

3.2.1 Likelihood Estimating Equations for β and the Dependence Parameter γ_1

Note that the model (3.1) - (3.2) generates y_t as a function of x_t and y_{t-1}. Recall that the likelihood function of β and γ_1 for this model was written in (2.36). Here we maximize this likelihood function (2.36) with respect to β and γ_1.
For the purpose, we first write the log-likelihood function as,

$$\begin{aligned}\log L(\beta,\gamma_1) &= l(\beta,\gamma_1) = y_1 \log p_1 + (1-y_1)\log(1-p_1) \\ &\quad + \sum_{t=2}^{T}[y_t \log p_{ty_{t-1}} + (1-y_t)\log(1-p_{ty_{t-1}})], \end{aligned} \qquad (3.10)$$

with $p_1 = exp(x_1'\beta)/[1+exp(x_1'\beta)]$ and $p_{ty_{t-1}} = exp(x_t'\beta + \gamma_1 y_{t-1})/[1+exp(x_t'\beta+\gamma_1 y_{t-1})]$ for $t = 2, \ldots, T$. Now the log-likelihood function with respect to β is given by

$$\frac{\partial \log L(\beta,\gamma_1)}{\partial \beta} = (y_1 - p_1)x_1 + \sum_{t=2}^{T}\left[(y_t - p_{ty_{t-1}})x_t\right]. \qquad (3.11)$$

Similarly the first derivative of the log-likelihood function with respect to γ_1 is given by

$$\frac{\partial \log L(\beta,\gamma_1)}{\partial \gamma_1} = \sum_{t=2}^{T}\left[(y_t - p_{ty_{t-1}})y_{t-1}\right]. \qquad (3.12)$$

We now solve the likelihood equations,

$$\frac{\partial l(\theta|y)}{\partial \theta} = \begin{pmatrix} \frac{\partial \log L(\beta,\gamma_1)}{\partial \beta} \\ \frac{\partial \log L(\beta,\gamma_1)}{\partial \gamma_1} \end{pmatrix} = 0 \tag{3.13}$$

for $\theta = (\beta, \gamma_1)'$.

By using a Taylor's series expansion, it follows from (3.13) that the θ parameter may be estimated by using the iterative equation

$$\hat{\theta}_{ML}(r+1) = \hat{\theta}_{ML}(r) + \left[[\frac{\partial^2 l(\theta|y)}{\partial \theta \partial \theta'}]^{-1} \frac{\partial l(\theta|y)}{\partial \theta} \right]_{\hat{\theta}_{ML}(r)} . \tag{3.14}$$

where $\hat{\theta}_{ML}(r)$ denotes the maximum likelihood estimate of θ as a solution of (3.13) at the r-th interation , and $[.]_{\hat{\theta}_{ML}(r)}$is the value of the expression in the square bracket evaluated at $\theta = \hat{\theta}_{ML}(r)$. Note that the equation in (3.14) requires the computation of the second derivative matrix $[\frac{\partial^2 l(\theta|y)}{\partial \theta \partial \theta'}]$, which is of order $(p+1) \times (p+1)$. The components of this matrix may be computed by using the following formulas:

$$\frac{\partial^2 L(\beta,\gamma_1)}{\partial \beta \partial \beta'} = -\left[x_1 x_1' p_1 (1-p_1) + \sum_{t=2}^{T} x_t x_t' p_{ty_{t-1}} (1-p_{ty_{t-1}}) \right] . \tag{3.15}$$

$$\frac{\partial^2 L(\beta,\gamma_1)}{\partial \beta \partial \gamma_1} = -\left[\sum_{t=2}^{T} p_{ty_{t-1}} (1-p_{ty_{t-1}}) x_t y_{t-1} \right] . \tag{3.16}$$

$$\frac{\partial^2 L(\beta,\gamma_1)}{\partial \gamma_1^2} = -\left[\sum_{t=2}^{T} y_{t-1}^2 p_{ty_{t-1}} (1-p_{ty_{t-1}}) \right] . \tag{3.17}$$

3.2.2 Generalized Quasilikelihood Estimating Equation: An Unconditional Approach

In this approach we consider three different cases: (i) Estimation of the regression parameters β with known γ_1 ; (ii) Estimation of the regression parameters and the correlation parameter simultaneously by using the GQL approach; (iii) Estimation of the regression parameters by using the GQL approach and the correlation parameter by using the method of moments.

Estimation of the regression parameters with known γ_1

Recall that in section 3.1.2, we have used a 'working'GQL approach to estimate the regression parameters. To be specific, the 'working 'approach was developed by treating the data as independent, i.e, by using $\gamma_1 = 0$, even though the data were correlated. In this subsection we discuss the GQL estimation for the regression parameters β under the assumption that γ_1 is known, even though in practice γ_1 is unknown. The estimation of β for unknown γ_1 is discussed in the next subsection.

Consider notation that we have used in section 3.1.2 and define $\Sigma = (\sigma_{tt'})$ as the covariance matrix of y, where $\gamma_1 \neq 0$. To be specific, $\sigma_{tt} = var(Y_t)$ and $\sigma_{tt'} = cov(Y_t, Y_{t'})$. The formulas for these variances (σ_{tt}) and covariances$(\sigma_{tt'})$are given by (2.40) and (2.53). By using the Σ matrix, we now write the generalized quasilikelihood(GQL) estimating equation for β as

$$\frac{\partial \mu'}{\partial \beta}\Sigma^{-1}(y-\mu) = 0, \tag{3.18}$$

which can be solved iteratively as in (3.8). Let $\hat{\beta}_{GQL}$ be the solution of (3.18). The iterative equation for $\hat{\beta}_{GQL}$ is given by

$$\hat{\beta}_{GQL}(r+1) = \hat{\beta}_{GQL}(r) + \left[\left(\frac{\partial \mu'}{\partial \beta}\Sigma^{-1}\frac{\partial \mu}{\partial \beta}\right)^{-1}\frac{\partial \mu}{\partial \beta}\Sigma^{-1}(y-\mu)\right]_{\hat{\beta}_{GQL}(r)}. \tag{3.19}$$

Since by (2.39) $\mu_t = p_{t0} + \mu_{t-1}(p_{t1} - p_{t0})$ with $p_{t0} = exp(x_t'\beta)/[1+exp(x_t'\beta)]$, and $p_{t1} = exp(x_t'\beta + \gamma_1)/[1+exp(x_t'\beta+\gamma_1)]$, and $x_t = (x_{t1}, \ldots, x_{tj}, \ldots, x_{tp})'$, to compute the derivative matrix $\frac{\partial \mu'}{\partial \beta}$ in (3.19) it is sufficient to compute the derivative vector $\frac{\partial \mu_t}{\partial \beta}$ for all $t = 1, \ldots, T$. This derivative vector for $t = 2, \ldots, T$ has the formula given by

$$\frac{\partial \mu_t}{\partial \beta} = [p_{t0}(1-p_{t0})(1-\mu_{t-1}) + p_{t1}(1-p_{t1})\mu_{t-1}]x_t + (p_{t1}-p_{t0})\frac{\partial \mu_{t-1}}{\partial \beta}.$$

whereas $\frac{\partial \mu_1}{\partial \beta}$ has the formula

$$\frac{\partial \mu_1}{\partial \beta} = p_1(1-p_1)x_1,$$

where $p_1 = exp(x_1'\beta)/[1+exp(x_1'\beta)]$.

Estimation of Regression and Correlation parameters by using GQL approach

As γ_1 is unknown in practice, in this subsection we estimate all parameters (β, γ_1) by using the GQL approach.
By writing $\theta = (\beta, \gamma_1)'$ and following (3.18) the GQL estimating equation for θ may be written as

$$\frac{\partial \mu'}{\partial \theta} \Sigma^{-1}(y - \mu) = 0. \tag{3.20}$$

Let $\hat{\theta}_{GQL}$ be the GQL estimator of θ obtained from (3.20). Similar to (3.9) and (3.19), $\hat{\theta}_{GQL}$ is obtained by using the iterative equation

$$\hat{\theta}_{GQL}(r+1) = \hat{\theta}_{GQL}(r) + \left[\left(\frac{\partial \mu'}{\partial \theta} \Sigma^{-1} \frac{\partial \mu}{\partial \theta} \right)^{-1} \frac{\partial \mu}{\partial \theta} \Sigma^{-1}(y - \mu) \right]_{\hat{\theta}_{GQL}(r)} . \tag{3.21}$$

Now to compute $\frac{\partial \mu'}{\partial \theta}$, we use the formula for $\frac{\partial \mu'}{\partial \beta}$ from the previous subsection, whereas $\frac{\partial \mu'}{\partial \gamma_1}$ may be computed by computing $\frac{\partial \mu_t}{\partial \gamma_1}$ for all $t = 1, \ldots, T$. Note that $\frac{\partial \mu_1}{\partial \gamma_1} = 0$ whereas for $t = 2, \ldots, T$, $\frac{\partial \mu_t}{\partial \gamma_1} = p_{t1}(1 - p_{t1})\mu_{t-1} + (p_{t1} - p_{t0})\frac{\partial \mu_{t-1}}{\partial \gamma_1}$.

Estimation of β by GQL and γ_1 by Method of Moments

Here, the estimation of β and γ_1 will be done in cycles of iterations. For a given γ_1, we first estimate β parameter by using (3.18) - (3.19). Once we get this estimate, we use it in a moment estimating equation for γ_1. The moment estimating equation for γ_1 is derived by equating $\sum_{t=2}^{T} \frac{y_t y_{t-1}}{T-1}$ with its expected value. Now to compute the expectation, that is, $E(Y_t Y_{t-1})$, we use the formula

$$E[Y_t Y_{t-1}] = Cov(Y_t, Y_{t-1}) + \mu_{t-1}\mu_t, \tag{3.22}$$

where by (2.53), the $Cov(Y_t, Y_{t-1})$ is given by

$$Cov(Y_t, Y_{t-1}) = \mu_{t-1}(1 - \mu_{t-1})(p_{t1} - p_{t0}).$$

It then follows that the moment estimating equation for γ_1 has the formula given by,

$$\begin{aligned} E\sum_{t=2}^{T}(Y_tY_{t-1}) - \sum_{t=2}^{T} y_ty_{t-1} &= 0 \\ \sum_{t=2}^{T}[\mu_{t-1}(1-\mu_{t-1})(p_{t1}-p_{t0}) + \mu_{t-1}\mu_t] - \sum_{t=2}^{T} y_ty_{t-1} &= 0 \\ g^*(\gamma_1) &= 0, \qquad (3.23) \end{aligned}$$

where $g^*(\gamma_1) = g(\gamma_1) - \sum_{t=2}^{T} y_ty_{t-1}$ and $g(\gamma_1) = \sum_{t=2}^{T}[\mu_{t-1}(1-\mu_{t-1})(p_{t1}-p_{t0}) + \mu_{t-1}\mu_t]$.

By using a Taylor's series expansion, it follows from (3.23) that the γ_1 parameter may be estimated by using the iterative equation

$$\gamma_1(\hat{r+1}) = \gamma_1\hat{(r)}_{old} - \left[\left(\frac{\partial g^*(\gamma_1)}{\partial\gamma_1}\right)^{-1} g^*(\gamma_1)\right]_{\hat{\gamma}_1(r)}, \qquad (3.24)$$

where $\hat{\gamma}_1(r)$ denotes the moment estimate of γ_1 as a solution of (3.23) at the r-th interation , and $[.]_{\hat{\gamma}_1(r)}$is the value of the expression in the square bracket evaluated at $\gamma_1 = \hat{\gamma}_1(r)$. Note that the equation in (3.24) requires the computation of the derivative $\frac{\partial g^*(\gamma_1)}{\partial\gamma_1}$, which has the formula given by

$$\begin{aligned} \frac{\partial g^*(\gamma_1)}{\partial\gamma_1} &= \sum_{t=2}^{T}[\mu_{t-1}p_{t1}(1-p_{t1}) + p_{t1}\frac{\partial\mu_{t-1}}{\partial\gamma_1} - \mu_{t-1}^2 p_{t1}(1-p_{t1}) - p_{t1}.2.\mu_{t-1}\frac{\partial\mu_{t-1}}{\partial\gamma_1} - \\ &\quad \mu_{t-1}.0 - p_{t0}\frac{\partial\mu_{t-1}}{\partial\gamma_1} + \mu_{t-1}^2.0 + p_{t0}.2.\mu_{t-1}\frac{\partial\mu_{t-1}}{\partial\gamma_1} + \mu_{t-1}\frac{\partial\mu_t}{\partial\gamma_1} + \mu_t\frac{\partial\mu_{t-1}}{\partial\gamma_1}] \\ &= \sum_{t=2}^{T}[\mu_{t-1}p_{t1}(1-p_{t1})(1-\mu_{t-1}) + (p_{t1}-p_{t0})\frac{\partial\mu_{t-1}}{\partial\gamma_1} - \\ &\quad 2(p_{t1}-p_{t0})\mu_{t-1}\frac{\partial\mu_{t-1}}{\partial\gamma_1} + \frac{\partial\mu_t}{\partial\gamma_1}\mu_{t-1} + \mu_t\frac{\partial\mu_{t-1}}{\partial\gamma_1}] \\ &= \sum_{t=2}^{T}[\mu_{t-1}p_{t1}(1-p_{t1})(1-\mu_{t-1}) + \frac{\partial\mu_{t-1}}{\partial\gamma_1}\{(p_{t1}-p_{t0}) - \\ &\quad 2.(p_{t1}-p_{t0})\mu_{t-1} + \mu_t\} + \frac{\partial\mu_t}{\partial\gamma_1}\mu_{t-1}]. \end{aligned}$$

As mentioned before, the estimate of β obtained from (3.19) is used in (3.24) to obtain an improved estimate of γ_1. The improved estimate of γ_1 is then used in

(3.19) to obtain an improved estimate of β. This cycle of iteration continues until convergence. We refer to this GQL and moments based combined approach as the Semi-GQL(SGQL)approach. In notation, the estimators for β and γ_1 based on this SGQL approach will be denoted by $\hat{\theta}_{SGQL} = (\hat{\beta}_{SGQL}, \hat{\gamma}_{SGQL,1})'$.

3.2.3 Generalized Quasilikelihood Estimating Equation: A Conditional Approach

In the last section we have discussed an unconditional quasilikelihood estimating equation approach, where it was necessary to compute the unconditional mean, variance and all lag covariances. Note that the computations for the unconditional covariance structure was somewhat involved. To reduce the computational burden in this subsection we now use a simpler conditional mean and covariance structure based conditional GQL approach. For this purpose the conditional first and second order moments readily follow from the model (3.1) - (3.2). To be specific, we provide these conditional moments as follows.

Conditional Mean:

It follows from (3.1) that for t=1, the expected value of y_1 is given by

$$E(Y_1) = p_{10} = \frac{exp(x_1'\,\beta)}{[1 + exp(x_1'\,\beta)]} = \mu_1\ , \tag{3.25}$$

For convenience we write $\mu_1^* = \mu_1$ under the conditional setup.

Next, it is clear from the model (3.2) that for $t = 2, \ldots, T$, the conditional mean of Y_t given y_{t-1} is given by

$$\begin{aligned} E(Y_t|y_{t-1}) &= P(Y_t = 1|y_{t-1}) = p_{ty_{t-1}} \\ &= \frac{exp(x_t'\,\beta + \gamma_1\ y_{t-1})}{[1 + exp(x_t'\,\beta + \gamma_1\ y_{t-1})]} = \mu_t^*,\ (say). \end{aligned} \tag{3.26}$$

Conditional Variance:

As the time series observations are binary, the conditional variance is given by

$$var(Y_t|y_{t-1}) = \mu_t^*(1 - \mu_t^*),$$

where μ_t^* are defined in (3.25) and (3.26).

Conditional Covariance:

Since we are considering a Lag1 model, it follow that covariance of $U = Y_{t-1}|Y_{t-2}$ and $V = Y_t|Y_{t-1}$ is zero. This is because,

$$\begin{aligned} cov(U,V) &= E(U\,V) - \mu_t^*\mu_{t-1}^* \\ &= E(Y_{t-1}Y_t|Y_{t-1}Y_{t-2}) - \mu_t^*\mu_{t-1}^* \\ &= E(Y_{t-1}|Y_{t-2})E(Y_t|Y_{t-1},Y_{t-2}) - \mu_t^*\mu_{t-1}^*. \end{aligned} \tag{3.27}$$

Note that by model (3.2), Y_t depends on y_{t-1} but not on y_{t-2}. Consequently, (3.27) may be written as

$$\begin{aligned} cov(U,V) &= E(Y_{t-1}|Y_{t-2})E(Y_t|Y_{t-1}) - \mu_t^*\mu_{t-1}^* \\ &= \mu_{t-1}^*\mu_t^* - \mu_{t-1}^*\mu_t^* \\ &= 0. \end{aligned} \tag{3.28}$$

Now writing $\mu^* = (\mu_1^*, \ldots, \mu_t^*, \ldots, \mu_T^*)$ and $\Sigma^* = diag[\mu_1^*(1-\mu_1^*), \ldots, \mu_t^*(1-\mu_t^*), \ldots, \mu_T^*(1-\mu_T^*)]$ we follow (3.20) and write the conditional GQL(CGQL) estimating equation for $\theta = (\beta, \gamma_1)'$ as

$$\frac{\partial \mu^{*'}}{\partial \theta}\Sigma^{*-1}(y - \mu^*) = 0, \tag{3.29}$$

where the derivative matrix $\frac{\partial \mu^{*'}}{\partial \theta}$ may be computed by exploiting the derivatives of μ_1^*, $\mu_t^*(t = 2, \ldots, T)$ with respect to β and γ_1. For the sake of computation, we provide the formulas for these derivatives as:

$$\begin{aligned} \frac{\partial \mu_1^*}{\partial \beta} &= p_1(1-p_1)x_1, \qquad \text{and} \quad \frac{\partial \mu_1^*}{\partial \gamma_1} = 0. \\ \frac{\partial \mu_t^*}{\partial \beta} &= p_{ty_{t-1}}(1-p_{ty_{t-1}})x_t, \quad \text{and} \quad \frac{\partial \mu_t^*}{\partial \gamma_1} = p_{ty_{t-1}}(1-p_{ty_{t-1}})y_{t-1} \quad t = 2, \ldots, T, \end{aligned}$$

where $p_{ty_{t-1}} = exp(x_t'\,\beta + \gamma_1\,y_{t-1})/[1 + exp(x_t'\,\beta + \gamma_1\,y_{t-1})]$.

Let $\hat{\theta}_{CGQL}$ be the CGQL estimator of θ obtained from (3.29). Similar to the unconditional GQL approach, $\hat{\theta}_{CGQL}$ may now be obtained by using the iterative equation

$$\hat{\theta}_{CGQL}(r+1) = \hat{\theta}_{CGQL}(r) + \left[\left(\frac{\partial \mu^{*'}}{\partial \theta}\Sigma^{*-1}\frac{\partial \mu^{*}}{\partial \theta}\right)^{-1}\frac{\partial \mu^{*}}{\partial \theta}\Sigma^{*-1}(y-\mu^{*})\right]_{\hat{\theta}_{CGQL}(r)}. \quad (3.30)$$

where $\hat{\theta}_{CGQL}(r)$ denotes the CGQL estimate of θ as a solution of (3.30) at the r-th interation , and $[.]_{\hat{\theta}_{CGQL}(r)}$is the value of the expression in the square bracket evaluated at $\theta = \hat{\theta}_{CGQL}(r)$.

3.2.4 A Simulation Study

Recall that in Section 3.1 we examined the performance of the independent assumption based ML and GQL approaches through a simulation study. Altogether four different designs were considered under each T = 100, 200 and 300.

Note that as opposed to Section 3.1, in Section 3.2 we have considered ML, GQL, SGQL and CGQL approaches to estimate β and γ_1 parameters. Two versions of GQL approaches were considered. First, under the GQL approach only β was estimated by using known, γ_1. In the second version, we have estimated all three parameters by using the GQL approach. These two versions will be referred to as GQL1 and GQL2 approaches.

In the present subsection, we conduct a simulation study with p =2, to examine the performances of the above five (ML,GQL1,GQL2,SGQL,and CGQL) approaches in estimating with p=2 covariates such that $\beta = (\beta_1, \beta_2)'$ as well as γ_1, wherever applicable. As far as the design is concerned, we, for simplicity use the fourth design (D_4) from section 3.1.3. Furthermore, for the size of the binary time series, we consider T = 200. The simulation results based on 1000 simulations are reported in Table 3.2 for three selected values of $\gamma_1 = -1, 0, 1$.

Table 3.2: Simulated mean (SM), simulated standard error (SSE) and simulated mean squared error (SMSE) of the ML, GQL1, GQL2, SGQL and CGQL estimates for all parameters with T = 200 $\beta_1 = \beta_2 = 0.5$; based on 1000 simulations, for the case when γ_1= -1, 0, 1

				Estimates		
γ_1	Method	Quantity	Simulations	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$
-1	ML	SM	1000	0.552	0.513	-1.089
		SSE		0.305	0.261	0.304
		SMSE		0.096	0.068	0.100
	GQL1	SM	1000	0.497	0.518	-
		SSE		0.233	0.255	-
		SMSE		0.054	0.065	-
	GQL2	SM	700	0.549	0.527	-1.073
		SSE		0.510	0.270	0.689
		SMSE		0.263	0.074	0.481
	SGQL	SM	900	0.521	0.518	-1.063
		SSE		0.290	0.258	0.296
		SMSE		0.084	0.067	0.092
	CGQL	SM	1000	0.552	0.513	-1.089
		SSE		0.305	0.261	0.304
		SMSE		0.096	0.068	0.100
0	ML	SM	1000	0.554	0.514	-0.049
		SE		0.326	0.258	0.272
		MSE		0.109	0.067	0.076
	GQL1	SM	1000	0.514	0.521	-
		SSE		0.246	0.250	-
		SMSE		0.061	0.063	-
	GQL2	SM	800	0.554	0.521	-0.042
		SSE		0.470	0.278	0.506
		SMSE		0.224	0.077	0.258
	SGQL	SM	700	0.547	0.515	-0.046
		SSE		0.338	0.262	0.276
		SMSE		0.116	0.069	0.078
	CGQL	SM	1000	0.554	0.514	-0.049
		SSE		0.326	0.258	0.272
		SMSE		0.109	0.067	0.076

(Table 3.2 Contd....)

γ_1	Method	Quantity	Simulations	Estimates		
				$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$
1	ML	SM	1000	0.550	0.511	0.986
		SSE		0.402	0.281	0.292
		SMSE		0.164	0.079	0.085
	GQL1	SM		0.532	0.516	-
		SSE		0.298	0.258	-
		SMSE		0.090	0.067	-
	GQL2	SM		0.543	0.520	1.004
		SSE		0.547	0.307	0.497
		SMSE		0.301	0.095	0.247
	SGQL	SM		0.579	0.500	0.973
		SSE		0.421	0.284	0.319
		SMSE		0.184	0.081	0.325
	CGQL	SM		0.550	0.511	0.986
		SSE		0.402	0.281	0.292
		SMSE		0.164	0.079	0.085

The results of Table 3.2 show that the CGQL approach gives the same estimates for all three parameters β_1, β_2 and γ_1 as that of the ML approach. For γ_1 =1, none of the approaches encountered any convergence problems. However for γ_1 = -1, the GQL2 approach converged in 700 simulations and the SGQL approach converged in 900 simulations. For γ_1 = 0, the GQL2 encountered convergence difficulties in 200 simulations whereas SGQL converged in 700 simulations. The ML, GQL1(γ_1 known) and CGQL had no convergence problems.

Note that when all three parameters are estimated at simultaneously, the ML and CGQL methods appear to produce their estimates with smaller mean squared errors(MSE) under all three cases with γ_1 = -1, 0, and 1. The GQL1 performs better than these two (ML and CGQL) approaches in estimating β_1 and β_2, but, it is not of much interest. This is because in practice γ_1 has to be estimated too. To be specific about the performances of the ML and CGQL approaches, for γ_1 = 1 case for example, they produce simulated MSE (SMSE) for $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\gamma}_1$ as 0.164, 0.079

and 0.085 respectively, whereas GQL2 produces their estimates with SMSE as 0.301, 0.095 and 0.247, and similarly SGQL approach produces their estimates with SMSE 0.184, 0.081 and 0.325. Thus, the ML and CGQL approaches perform better than other competitive approaches in estimating all three parameters.

Chapter 4

Forecasting Binary Probability

In the previous chapter we have discussed the estimation approaches for the three parameters involved in our proposed the non-linear dynamic model. We have examined the performance of estimation approaches through a simulation study. In order to examine the forecasting performance of the ML and/or CGQL approach, in this chapter we consider forecasting inference for a Lag1 model only. Note that in the binary case, it is appropriate to forecast the probability of the occurrence of the future binary observation. Thus, we want to use the available data up to time t to forecast $\mu_{t+1} = P(y_{t+1} = 1)$.

Recall from Chapter 2 that for $t = 1, \ldots, T$, μ_{t+1} can be obtained by using the recurrence relationship

$$\mu_{t+1} = p_{t+1,0} + \mu_t(p_{t+1,1} - p_{t+1,0}), \tag{4.1}$$

where $\mu_1 = exp(x_1'\beta)/[1 + exp(x_1'\beta)]$, $p_{t+1,1} = exp(x_{t+1}'\beta + \gamma_1)/[1 + exp(x_{t+1}'\beta + \gamma_1)]$ and $p_{t+1,0} = exp(x_{t+1}'\beta)/[1 + exp(x_{t+1}'\beta)]$.

Now suppose that data up to time point T, i.e, $y_1, \ldots, y_t, \ldots, y_T$ as well as $x_1, \ldots, x_t, \ldots, x_T$ are available. Also suppose that the covariate x_{T+1} is available. It then follows from the model (3.1) - (3.2) that one may forecast μ_{t+1} by using $E(Y_{t+1}|y_t)$ which has the formula as

$$E(Y_{t+1}|y_t) = P(Y_{t+1} = 1|y_t)$$

$$= exp(x_{t+1}^{'}\beta + \gamma_1 y_t)/[1 + exp(x_{t+1}^{'}\beta + \gamma_1 y_t)]$$
$$= p_{f,t+1}, say. \quad (4.2)$$

We evaluate the performance of this forecasting function (4.2) through a simulation study. For this purpose we consider, p = 2 with $\beta_1 = \beta_2 = 0.5$ and three values of γ_1 = -1, 0, 1. Further more, we consider T =200 and examine the performance of the forecasting function $\hat{p}_{f,201}(4.2)$ to forecast $\mu_{201}(4.1)$. As far as the design matrix is concerned, we consider the same design D_4 for $t = 1, \ldots, T =$ (200), and we chose $x_{201} = (1,1)^{'}$.

Note that based on $x_{T+1} = (1,1)^{'}$, and the values of β_1, β_2 and γ_1, we first compute μ_{201} which is reported in the second column of Table 4.1. We remark that for given values of β_1, β_2 and γ_1, these values of μ_{201} are fixed as they do not depend on the response variable, rather they depend only on the covariate x_{t+1}. Thus, there is no simulation needed for this.

We carry out a simulation study to compute the simulation average of the forecasting function $\hat{p}_{f,201}$. As far as the values of $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\gamma}_1$ are concerned , we obtain these estimates in a given simulation based on T = 200 observations, by using both ML (3.13) and CGQL (3.29) approach as explained in chapter 3. In each simulations, these estimates are then used to compute $\hat{p}_{f,201}$ by using x_{201}. The simulation average along with their standard errors based on 1000 simulations are reported in the fourth column of Table 4.1.

Table 4.1: The value of μ_{201}(fixed) simulated average value of $\hat{p}_{f,201}$ based on the ML and/or CGQL approach with T = 200; $\beta_1 = \beta_2 = 0.5$; based on 1000 simulations, for the case when γ_1=-1,0,1.

γ_1	μ_{201}	Method	$\hat{p}_{f,201}$(SSE)
-1	0.363	ML/CGQL	0.354(0.157)
0	0.500	ML/CGQL	0.484(0.140)
1	0.699	ML/CGQL	0.670(0.168)

Note that, as the ML and the CGQL approaches provide the same estimates for the parameters (see Table 3.2) of the model, they also provide the same forecasted value. The simulation average appears to be very close to the value of μ_{201} in all three cases with γ_1 = -1, 0, 1, indicating that the forecasting function $p_{f,201}$ forecasts μ_{t+1}very well. For example, where γ_1 = -1, the simulated average forecast to be 0.354, while the true value of the counterpart probability is found to be 0.363.

Chapter 5

Estimation of the Model Parameters of Lag 2 Model: A Generalization of Lag 1 Model

Recall from chapter 3 that the ML and the CGQL approaches were found to be the best as compared to other competitive approaches in estimating the parameters of the lag 1 binary time series model. Note that, even though the lag 1 model is more practical as compared to other higher order lag based models, there may be some situations where lag 2 model may be appropriate.

In this chapter, we examine the performance of the ML and the CGQL approaches in estimating the parameters of a lag 2 based binary time series model. For convenience, we write this model as follows:

$$\begin{aligned} P(y_1 = 1) &= \frac{exp(x_1'\beta)}{[1 + exp(x_1'\beta)]} \\ &= \lambda_1^*. \end{aligned} \tag{5.1}$$

$$\begin{aligned} P(y_2 = 1|y_1) &= \frac{exp(\, x_2'\beta + \gamma_1 y_1)}{[1 + exp(\, x_2'\beta + \gamma_1 y_1)]} \\ &= p_{2\,y_1} = \lambda_2^*. \end{aligned} \tag{5.2}$$

and

$$P(y_t = 1|y_{t-1}y_{t-2}) = \frac{exp(\ x_t'\beta + \gamma_1 y_{t-1} + \gamma_2 y_{t-2})}{[1 + exp(\ x_t'\beta + \gamma_1 y_{t-1} + \gamma_2 y_{t-2})]} = p_{t\, y_{t-1}\, y_{t-2}}, \quad (5.3)$$

for $t = 3, \ldots, T$.

It is of interest to estimate the regression parameter β and the dynamic dependence parameter $\gamma = (\gamma_1, \gamma_2)'$. For this purpose, we discuss the ML and the CGQL approaches for the estimation of all four parameters, as these approaches were found to be best as compared to other competitive approaches explained in chapter 3.

We consider the time series $y_1, \ldots, y_t, \ldots, y_T$, along with $x_t = (x_{t1}, \ldots, x_{tj}, \ldots, x_{tp})'$ as a vector of p-dimensional covariates associated with y_t, as before, but the responses now are generated following the model (5.1) - (5.3). The likelihood as well as the CGQL estimating equations are developed in sections 5.1 and 5.2, respectively.

5.1 Likelihood Estimating Equations for β and γ

Note that for the lag 1 model, the likelihood function of β, and γ_1 was written in (2.36). We now extend the likelihood function (2.36) to the lag 2 model (5.1) - (5.3). To be specific, under (5.1) - (5.3) the likelihood function has the form given by

$$L(\beta, \gamma_1, \gamma_2) = P(y_1, y_2, \ldots, y_t, \ldots, y_T) = P(y_1)P(y_2|y_1)P(y_3|y_2, y_1) \ldots P(y_t|y_{t-1}, y_{t-2}) \ldots P(y_T|y_{T-1}, y_{T-2}). \quad (5.4)$$

By (5.3) the conditional probability $P(y_t|y_{t-1}y_{t-2})$ is given as

$$P(y_t|y_{t-1}y_{t-2}) = p_{ty_{t-1}y_{t-2}}^{y_t}(1 - p_{ty_{t-1}y_{t-2}})^{1-y_t} \quad (5.5)$$

with $p_{ty_{t-1}y_{t-2}}$ as in (5.3). Thus, we can re-write the likelihood function (5.4) as

$$L(\beta, \gamma_1, \gamma_2) = p_1^{y_1}(1 - p_1)^{(1-y_1)} p_{2y_1}^{y_2}(1 - p_{2y_1})^{(1-y_2)} \prod_{t=3}^{T} p_{ty_{t-1}y_{t-2}}^{y_t}(1 - p_{ty_{t-1}y_{t-2}})^{(1-y_t)} \quad (5.6)$$

yielding the log likelihood function

$$\log L(\beta,\gamma_1,\gamma_2) = y_1 \log p_1 + (1-y_1)\log(1-p_1) + y_2 \log p_{2y_1} + (1-y_2)\log(1-p_{2y_1}) + \sum_{t=3}^{T}[y_t \log p_{ty_{t-1}y_{t-2}} + (1-y_t)\log(1-p_{ty_{t-1}y_{t-2}})] \quad (5.7)$$

with $p_1 = exp(x_1'\beta)/[1+exp(x_1'\beta)]$, $p_{2y_1} = exp(x_2'\beta+\gamma_1 y_1)/[1+exp(x_2'\beta+\gamma_1 y_1)]$ and $p_{ty_{t-1}y_{t-2}} = exp(\ x_t'\beta+\gamma_1 y_{t-1}+\gamma_2 y_{t-2})/[1+exp(\ x_t'\beta+\gamma_1 y_{t-1}+\gamma_2 y_{t-2})]$ for $t=3,\ \ldots,T$.

Now by similar calculations as in (3.11) the first derivative of the log likelihood function with respect to β is given by

$$\frac{\partial \log L(\beta,\gamma_1,\gamma_2)}{\partial\beta} = (y_1-p_1)x_1 + (y_2-p_{2y_1})x_2 + \sum_{t=3}^{T}\Big[(y_t-p_{ty_{t-1}y_{t-2}})x_t\Big]. \quad (5.8)$$

Similarly the first derivative with respect to γ_1 is given by

$$\frac{\partial \ln L(\beta,\gamma_1,\gamma_2)}{\partial\gamma_1} = (y_2-p_{2y_1})y_1 + \sum_{t=3}^{T}\Big[(y_t-p_{ty_{t-1}y_{t-2}})y_{t-1}\Big], \quad (5.9)$$

whereas the first derivative with respect to γ_2 has the form

$$\frac{\partial \ln L(\beta,\gamma_1,\gamma_2)}{\partial\gamma_2} = \sum_{t=3}^{T}\Big[(y_t-p_{ty_{t-1}y_{t-2}})y_{t-2}\Big]. \quad (5.10)$$

Next by combining (5.8) to (5.10), we solve the likelihood equation

$$\frac{\partial l(\theta|y)}{\partial\theta} = \begin{pmatrix} \frac{\partial \log L(\beta,\gamma_1,\gamma_2)}{\partial\beta} \\ \frac{\partial \log L(\beta,\gamma_1)}{\partial\gamma_1} \\ \frac{\partial \log L(\beta,\gamma_1)}{\partial\gamma_2} \end{pmatrix} = 0 \quad (5.11)$$

for $\theta = (\beta,\gamma_1,\gamma_2)'$.

Similar to section 3.2.1, by using a Taylor's series expansion, it follows from (5.11) that the θ parameter may be estimated by using the iterative equation

$$\hat{\theta}_{ML}(r+1) = \hat{\theta}_{ML}(r) + \left[[\frac{\partial^2 l(\theta|y)}{\partial\theta\partial\theta'}]^{-1}\frac{\partial l(\theta|y)}{\partial\theta}\right]_{\hat{\theta}_{ML}(r)}, \quad (5.12)$$

where $\hat{\theta}_{ML}(r)$ denotes the maximum likelihood estimate of θ as a solution of (5.11) at the r-th iteration , and $[.]_{\hat{\theta}_{ML}(r)}$is the value of the expression in the square bracket evaluated at $\theta = \hat{\theta}_{ML}(r)$. Note that the equation in (5.12) requires the computation of the second derivative matrix $[\frac{\partial^2 l(\theta|y)}{\partial\theta\partial\theta'}]$ which is of order $(p+2) \times (p+2)$. The components of this matrix may be computed by using the following formulas:

$$\begin{aligned}\frac{\partial^2 L(\beta,\gamma_1,\gamma_2)}{\partial\beta\partial\beta'} &= -[p_1(1-p_1)x_1x_1' + p_{2y_1}(1-p_{2y_1})x_2x_2' + \\ &\quad \sum_{t=3}^{T} p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})x_tx_t'], \quad (5.13)\\ \frac{\partial^2 L(\beta,\gamma_1,\gamma_2)}{\partial\beta\partial\gamma_1} &= -\left[p_{2y_1}(1-p_{2y_1})x_2y_1 + \sum_{t=3}^{T} p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})x_ty_{t-1}\right], \quad (5.14)\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L(\beta,\gamma_1,\gamma_2)}{\partial\beta\partial\gamma_2} &= -\left[\sum_{t=3}^{T} p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})x_ty_{t-2}\right], \quad (5.15)\\ \frac{\partial^2 L(\beta,\gamma_1,\gamma_2)}{\partial\gamma_1\partial\gamma_1'} &= -\left[p_{2y_1}(1-p_{2y_1})y_1^2 + \sum_{t=3}^{T} p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})y_{t-1}^2\right], \quad (5.16)\\ \frac{\partial^2 L(\beta,\gamma_1,\gamma_2)}{\partial\gamma_1\partial\gamma_2'} &= -\left[\sum_{t=3}^{T} p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})y_{t-1}y_{t-2}\right], \quad (5.17)\end{aligned}$$

and

$$\frac{\partial^2 L(\beta,\gamma_1,\gamma_2)}{\partial\gamma_2\partial\gamma_2'} = -\left[\sum_{t=3}^{T} p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})y_{t-2}^2\right]. \quad (5.18)$$

5.2 Generalized Quasilikelihood Estimating Equation for β and γ: A Conditional Approach

The conditional GQL (CGQL) approach is much easier than the ML approach and it performs as well as the ML approach in estimating the parameters. In this section, we discuss the lag 2 model parameter estimation using a CGQL approach. Similar to section 3.2.3 we use the conditional mean and conditional covariance structure to construct the CGQL estimating equations. The conditional mean, variance and covariance may be simplified as follows:

Conditional Mean

From model (5.1) for t=1, the expected value of y_1 is given by,

$$E(Y_1) = p_{10} = \frac{exp(x_1'\beta)}{[1+exp(x_1'\beta)]} = \mu_1, \tag{5.19}$$

which for convenience we write as $\mu_1 = \mu_1^*$. Next, the conditional expectation of y_2 given y_1 is written as

$$E(Y_2|y_1) = p_{2y_1} = \frac{exp(x_2'\beta + \gamma_1 y_1)}{[1+exp(x_1'\beta + \gamma_1 y_1)]} = \mu_2 = \mu_2^*. \tag{5.20}$$

Similarly, for $t = 3, \ldots, T$, it follows from the model (5.3) that the conditional mean of y_t given y_{t-1} and y_{t-2} has the form,

$$E(Y_t|y_{t-1}, y_{t-2}) = P(y_t = 1|y_{t-1}, y_{t-2}) = \frac{exp(x_t' \, \beta + \gamma_1 \, y_{t-1} + \gamma_2 y_{t-2})}{1 + exp(x_t' \, \beta + \gamma_1 \, y_{t-1} + \gamma_2 y_{t-2})} = \mu_t^*, (say). \tag{5.21}$$

Conditional Variance

As the time series observations are binary, the conditional variance is given by,

$$var(Y_t|y_{t-1}, y_{t-2}) = \mu_t^*(1 - \mu_t^*), \tag{5.22}$$

where μ_t^* are define in (5.19) - (5.21).

Conditional Covariance

All conditional covariances are zero. This may be shown as follows. Suppose $u = \{y_{t-1}|y_{t-2}, y_{t-3}\}$ and $v = \{y_t|y_{t-1}, y_{t-2}\}$. Then, the covariance between u and v can be written as

$$\begin{aligned} cov(U,V) &= E(UV) - \mu_t^*\mu_{t-1}^* \\ &= E(Y_{t-1}Y_t|Y_{t-1}Y_{t-2}Y_{t-3}) - \mu_t^*\mu_{t-1}^* \\ &= E(Y_{t-1}|y_{t-2}, y_{t-3})E(Y_t|y_{t-1}, y_{t-2}, y_{t-3}) - \mu_t^*\mu_{t-1}^*. \end{aligned} \tag{5.23}$$

Note that it follows from (5.21) that y_t depends on y_{t-1} and y_{t-2}, but not on y_{t-3}. Consequently, (5.23) may be written as,

$$\begin{aligned} cov(U,V) &= E(Y_{t-1}|y_{t-2},y_{t-3})E(Y_t|y_{t-1},y_{t-2}) - \mu_t^*\mu_{t-1}^* \\ &= \mu_t^*\mu_{t-1}^* - \mu_t^*\mu_{t-1}^* \\ &= 0. \end{aligned} \quad (5.24)$$

CGQL Estimating Equation:

Now writing $\mu^* = (\mu_1^*, \ldots, \mu_t^*, \ldots, \mu_T^*)$ and $\Sigma^* = \text{diag}[\mu_1^*(1-\mu_1^*), \ldots, \mu_t^*(1-\mu_t^*), \ldots, \mu_T^*(1-\mu_T^*)]$, we may write the conditional GQL(CGQL) estimating equation for $\theta = (\beta, \gamma_1, \gamma_2)'$ as

$$\frac{\partial \mu^{*'}}{\partial \theta}\Sigma^{*-1}(y-\mu^*) = 0\,, \quad (5.25)$$

where the derivative matrix $\frac{\partial \mu^{*'}}{\partial \theta}$ may be computed by exploiting the derivatives of μ_1^*, μ_2^*, μ_t^* $(t = 3, \ldots, T)$ with respect to β and $\gamma = (\gamma_1, \gamma_2)'$. We provide the formulas for these derivatives as:

$$\begin{aligned} \frac{\partial \mu_1^*}{\partial \beta} &= p_1(1-p_1)x_1, \quad \frac{\partial \mu_1^*}{\partial \gamma_1} = 0 \;\; and \;\; \frac{\partial \mu_1^*}{\partial \gamma_2} = 0. \\ \frac{\partial \mu_2^*}{\partial \beta} &= p_{2y_1}(1-p_{2y_1})x_2, \quad \frac{\partial \mu_2^*}{\partial \gamma_1} = p_{2y_1}(1-p_{2y_1})y_1 \;\; and \;\; \frac{\partial \mu_2^*}{\partial \gamma_2} = 0. \\ \frac{\partial \mu_t^*}{\partial \beta} &= p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})x_t, \\ \frac{\partial \mu_t^*}{\partial \gamma_1} &= p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})y_{t-1}, \\ \frac{\partial \mu_t^*}{\partial \gamma_2} &= p_{ty_{t-1}y_{t-2}}(1-p_{ty_{t-1}y_{t-2}})y_{t-2}, \end{aligned}$$

where $p_{ty_{t-1}y_{t-2}} = exp(x_t'\beta + \gamma_1 y_{t-1} + \gamma_2 y_{t-2})/[1 + exp(x_t'\beta + \gamma_1 y_{t-1} + \gamma_2 y_{t-2})]$.

Let $\hat{\theta}_{CGQL}$ be the CGQL estimator of θ obtained from (5.25). Similar to the CGQL estimating equating written in (3.30), $\hat{\theta}_{CGQL}$ may now be obtained for this larger model by using the iterative equation

$$\hat{\theta}_{CGQL}(r+1) = \hat{\theta}_{CGQL}(r) + \left[\left(\frac{\partial \mu^{*'}}{\partial \theta}\Sigma^{*-1}\frac{\partial \mu^*}{\partial \theta}\right)^{-1}\frac{\partial \mu^*}{\partial \theta}\Sigma^{*-1}(y-\mu^*)\right]_{\hat{\theta}_{CGQL}(r)}, \quad (5.26)$$

where $\hat{\theta}_{CGQL}(r)$ denotes the CGQL estimate of θ as a solution of (5.25) at the r-th interation , and $[.]_{\hat{\theta}_{CGQL}(r)}$is the value of the expression in the square bracket evaluated at $\theta = \hat{\theta}_{CGQL}(r)$. Note that the computation of the CGQL iterative equation (5.26) is simpler than that of the ML equation (5.12). This is because unlike the ML approach, the CGQL estimating equation is only required to construct the conditional mean vector and conditional covariance matrix which is of diagonal form.

5.3 A Simulation Study

In this section we conduct a simulation study to examine the performance of the parameters of the lag 2 model. We use the same design matrix D_4 from section 3.2.4. In the design matrix the number of covariates is two, i.e $\beta = (\beta_1, \beta_2)'$. Further, we consider T=200 for the size of the binary time series. The results based on 1000 simulations are reported in Table 5.1 for three values of $\gamma_1 = \gamma_2 = -1$, $\gamma_1 = \gamma_2 = 0$ and $\gamma_1 = \gamma_2 = 1$.

The results of Table 5.1 show that the CGQL and ML approaches gives the same estimates for all four parameters $\beta_1, \beta_2, \gamma_1$ and γ_2. Note that the approaches produces good estimates for all parameters except β_1. For example, when $\gamma_1 = -1$ and γ_2 = -1 , the simulated mean (SM) of the estimates of $\beta_1, \beta_2, \gamma_1,$ andγ_2 are found to be $\hat{\beta}_1$ =0.572, $\hat{\beta}_2$ = 0.496 $\hat{\gamma}_1$ = -1.091 and $\hat{\gamma}_2$= -1.086, whereas the true β is given as $\beta = (0.5, 0.5)'$. However, the ML and CGQL overestimate β_1 in all three cases. However, the approach overall perform well in estimating the parameters.

Table 5.1: Simulated mean (SM), and simulated standard error (SSE) and simulated mean squared error (SMSE) of the ML and CGQL estimates of the regression and the correlation parameters with T = 200; $\beta_1 = \beta_2 = 0.5$; based on 1000 simulations, for the case when γ_1 = -1, 0, 1 and γ_2 = -1, 0, 1.

				Estimates			
γ_1	γ_2	Method	Quantity	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\gamma}_1$	$\hat{\gamma}_2$
-1	-1	ML/CGQL	SM	0.572	0.500	-1.091	-1.086
			SSE	0.348	0.269	0.319	0.332
			SMSE	0.126	0.0724	0.102	0.110
0	0	ML/CGQL	SM	0.582	0.509	-0.040	-0.043
			SSE	0.366	0.265	0.287	0.281
			SMSE	0.133	0.070	0.084	0.079
1	1	ML/CGQL	SM	0.672	0.453	1.022	0.956
			SSE	0.788	0.373	0.436	0.430
			SMSE	0.621	0.139	0.190	0.186

Chapter 6

Concluding Remarks

As opposed to the Gaussian (Keenan(1982)) and uniform (Qaqish (2003)) latent process based binary time series models, in this thesis we have considered a logistic latent process based non-linear dynamic binary time series model (Amemiya (1985)). This latter model has advantages over the other models because of its correlation structure that allows correlations to have full ranges.

With regard to the estimation of the parameters of the non-linear dynamic binary time series model, we first examined the effect of ignoring correlations in estimating the regression parameters. It was found that ignoring correlations has detrimental effects on the regression parameter estimation. To estimate the regression as well as the dynamic dependence parameters, we have used the ML and the GQL approaches and found that the conditional GQL (CGQL) approach performs as good as the ML approach in estimating these parameters. Also, the CGQL approach appears to be simpler as compared to the other approaches.

We have also considered the forecasting problem. It was found that the non-linear conditional probability function works quite well in forecasting a future response probability.

This work should be useful to the researchers interested in analysing discrete time series data. Note that in some situations the binary time series may be contaminated by certain outliers. One may attempt to use robust estimation approach to tackle such outlier cases, which is however beyond the scope of the present thesis.

Bibliography

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

[2] Box, George E. P and Jenkins, G.M.(1994). Time series analysis : forecasting and control. *Prentice Hall.*

[3] Guttorp,P. (1986). On Binary Time Series Obtained From Continuous Time Point Process Describing Rainfalls. *Water Resources Research.* **22**, 897-904.

[4] Johnson, N.L, and Kotz.S. (1970). Continuous univariate distributions-2. *John Wiley and Sons.*

[5] Kanter,M. (1975). Autoregression for Discrete Processes Mod 2. *Journal of Applied Probability*, **12**, 371-375.

[6] Keenan,D.M. (1982). A Time Series Analysis of Binary Data. *Journal of American Statistical Association*, **77**, 816-821.

[7] 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.

[8] Sutradhar, B.C. (2003). An Overview on Regression Models for Discrete Longitudinal Responses. *Statistical Science*, **18**, 377-393.

[9] Farrell, P.J. and Sutradhar, B.C.(2006). A non-linear conditional probability model for generating correlated binary data *Statistics & Probability Letters*, **76**, 353-361.

[10] Sutradhar, B.C. and Farrell,P.J (2006) Analyzing longitudinal binary data by using nonlinear dynamic modelling. *Unpublished Manuscript*

[11] Sutradhar, B.C. and Rao, R.P (2006) On dynamic mixed models for binary panel data and estimtions of parameters. *Unpublished Manuscript*

[12] Zeger, S.L. (1988). A Regression Model for Time Series of Counts. *Biometrika*, **75**, 621-629.

[13] Zhang, H. Triche,E. and Leaderer,B. (2000). Model for the Analysis of Binary Time Series of Respiratory Symptoms. *American Journal of Epidemiology*, **151**, 1206-1215.