東海大學統計學系碩士班

碩士論文

指導教授：林雅俐　博士

Generalized Estimating Equations for Analyzing Correlated Binary Responses on Online Advertising

研究生：陳雅雯

中華民國九十四年七月

Abstract

In this paper, we propose a quantitative approach to modeling consumer response to medium rectangle ad ($300\times$ 250 IMU or 75000 pixels) at a news context Web site. The influences of ad positions, animation length, and exposure times on the click-through rate (CTR) were designed as a factorial experiment with repeated measures. Binary responses in serially correlated click data were collected. The use of generalized estimating equations (GEEs) approach would be introduced to fit logistic regression models with correlated binary data. A goodness-of-fit statistic, quasilikelihood under the independent model information criterion (QIC) for correlated models will be used for evaluating GEE-constructed models. The results showed that a logistic regression model with order effect, two-factor interaction effect of ad type and ad position, as well as ad position and animation length fitted relatively well. In addition, GEE model with AR(1) working correlation also has the smallest QIC, when comparing with other types of correlation structures. Moreover, the graphical method is used to diagnose. Given promotion-type ad, the combination of middle position and animation length of 7.5 seconds could provide the highest ECTR.

Keywords: Advertisement Effectiveness, Browsing Performance, Click-through Rate, Exposure Duration, Online Advertising Recycle Frequency

Contents

List of Tables

List of Figures

1 INTRODUCTION

To attract online users' attention, various kinds of animation are widely used on websites, including animated online advertising. A widely used measurer for evaluating the effectiveness of online advertising is the click-through rate (CTR)-that is, the proportion of viewers who click on an online ad to visit the advertiser's websites (Baltas, 2003; Hanson, 2000). It is noteworthy that the pricing of internet advertising is often based on CTR because advertisers demand results-oriented pricing and question the pricing model of traditional media, which is based on mere impressions (Hofacker & Murphy, 1988; Cho et al., 2001). Lin and Lin (2003) proposed a multiple logistic regression model to fit the CTR data and the experimental results showed the two-factor interactions of Ad size and Ad contents as well as the interaction of gender and Ad position were verified to be significantly affecting CTR. Their results also showed that larger Ad size would have better advertising click-through rate. Using commodities information and sale promotion activity as Ad contents, subjects would have much more click through on medium rectangle size (300×250 IMU or 75,000 pixels) than full banner size (28,080 pixels)and button 2 size (7,200 pixels), as well as have high click through rate on full banner than button 2. "Interactive Advertising" has clearly become a mainstream medium and one that can no longer be ignored as a critical piece of any marketing mix (IAB, 2004). The revenue results reported for 2004 confirm a very healthy environment for online advertising, for both direct marketers seeking immediate performance results, as well as brand advertisers looking to create or enhance an image, product or service.

Continuously, in this study we will investigate the influence of flash animation advertising on online advertising effectiveness. To clarify and quantify the effects of design parameters, an experiments and after-experiment questionnaire survey will be used to collect the related data. With this study, the relationship between animation length and advertising effectiveness as well as the relationship between exposure times and advertising effectiveness on the internet will be discussed.

When we concerned with the analysis of repeated measures of binary response data and tried to find the relationship between CTR and a set of covariates, the response variables will usually be correlated because repeated observations are made on the same individual. However, the conventional generalized linear models (GLMs) proposed by McCullagh and Nelder (1989) require that the observations in the data set are independent. The generalized estimation equations (GEEs) methodology, introduced by Liang and Zeger (1986), provides a method of analyzing correlated data that otherwise could be modeled as a generalized linear model. The GEEs have solutions which are consistent and asymptotically Gaussian even when time dependence is misspecified as we often expect. A consistent variance estimate is also presented by Zeger and Liang (1986).

In this study, the use of the GEE approach with correlated binary data from an experiment would be illustrated. This paper presents the logistic regression models to the analysis of CTR data when independence and autocorrelated structures from each subject are considered. The objectives of this paper are not only to give the ergonomic guidelines on the design of animated online advertising and the introduction of GEEs to the correlated responses, it also offers the advertisers a fair way to evaluate online advertising effectiveness. Moreover, our intension is to motivate what we consider to be a widely applicable methodology for repeated observations.

The rest of this article is organized as follows. Section 2 describes the literatures of GLMs and GEEs methodology. In section 3, research method and model proposed in this study including the design of a factorial with complete randomization in repeated measures will be described. After fitting some feasible models the quasilikelihood under the independent model information criterion (QIC) for correlated models will be used for evaluating GEE-constructed models. Moreover, model diagnostics including graphical assessment of residual analysis will be used to be the final check that the selected model adequately fits the data. The results will show in section 4. The conclusion will be summaried in section 5 of this paper.

2 BRIEF REVIEW OF GLM's

Generalized Linear Models (GLMs) was initially constructed by Wedderburn and Neter in the mid-1970s. The GLM has been used widely for exploring the relationship between response variable and prognostic factors. However, the conventional GLM proposed by McCullagh and Neter (1989) require that the observations in the data set are independent. Liang & Zeger (1986) and Zeger & Liang (1986) proposed the GEE method which is an extension of GLM to the analysis of longitudinal data when the data structure is correlated. In this study, we focus on data sets comprising a short binary time series and a set of time independent covariates for each subject. Given a single binary response for each subject, logistic regression could be used to fit the relationship of response variables and the covariates. With time series, however, methods that account for time autocorrelation are necessary. We begin with a review of likelihood-based logistic regression model for independent binary data. After illustrating the standard techniques for building estimating equations of likelihood-based models, we review the generalized estimating equation of generalized linear models which apply in the field of correlated binary data.

2.1 Independent Binary Data

The standard procedures with independent binary data in the ordinary logistic regression (OLR) model are described as follows:

1. Choose a distribution for the outcome variable.

2. Write the joint distribution for the data set.

3. Convert the joint distribution to a likelihood function, in general, log-likelihood function for the exponential family.

4. Generalized the likelihood via introduction of a linear combination of covariates and associated coefficients.

5. Parameterize the linear combination of covariates to enforce range restrictions on the mean and variance implied by the distribution.

6. Write the estimating equations for obtaining the solutions of unknown parameters.

In this paper, we propose a logistic model with GEE method to assess the relationship between click-through rates and a set of covariates in which repeated measurements on each subject was involved. The following sections are a short review of dependent binary data.

2.2 The ML Estimating Equations for Binary Logistic Regression

Let $(y_1, \ldots, y_n)$ denote values of the outcomes from n individuals. Assume the outcome of interest is dichotomous. Without loss of generality, a successful outcome is coded as 1 and a failure outcome is coded as 0. Let π is the probability of success and $\pi \in [0, 1]$. The probability mass function of a Bernoulli distribution with mean π, and variance, $\pi(1-\pi)$ is

$$f(y_i|\pi) \quad = \quad \pi^{y_i}(1-\pi)^{1-y_i}, \quad y_i = 0, 1, \quad i = 1, 2, \ldots, n \tag{1}$$

Now we have Bernoulli random variables with mean π and variance $\pi(1-\pi)$. The likelihood function of n independent binary outcomes $(y_1,\ldots,y_n)$ is

$$\begin{aligned} L(\pi|y_1,\ldots,y_n) &= \prod_{i=1}^{n} \pi^{y_i}(1-\pi)^{1-y_i} \\ &= \prod_{i=1}^{n} \exp\left\{ y_i \ln\left(\frac{\pi}{1-\pi}\right) + \ln(1-\pi) \right\}. \end{aligned} \tag{2}$$

Replace the expected value π of the Bernoulli with μ and introduce the covariates into the logistic regression model. Then we rewrite equation (2) into:

$$L(\boldsymbol{\mu}|y_1,\ldots,y_n) = \prod_{i=1}^{n} \exp\left\{ y_i \ln\left(\frac{\mu_i}{1-\mu_i}\right) + \ln(1-\mu_i) \right\}. \tag{3}$$

The log likelihood function is

$$\begin{aligned} \ell(\boldsymbol{\mu}|y_1,\ldots,y_n) &= \sum_{i=1}^{n} \left[y_i \ln\left(\frac{\mu_i}{1-\mu_i}\right) + \ln(1-\mu_i) \right] \\ &= \sum_{i=1}^{n} \left[y_i\theta_i - \ln\{1+\exp(\theta_i)\} \right], \end{aligned} \tag{4}$$

where $\theta_i = \ln \frac{\mu_i}{1-\mu_i}$, $i = 1, 2, \ldots, n$.

We introduce the relationship between the expected value and covariates using linear predictor. Let $(x_{i1},\ldots,x_{ip})$ denote values of explanatory variables for subject i. Then, we defined the linear predictor as follows

$$\eta_i = \mathbf{x}_i\boldsymbol{\beta}, \quad i = 1,\ldots,n, \tag{5}$$

where $\mathbf{x}_i = (x_{i1},\ldots,x_{ip})$ and $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)'$.

The range of linear predictor is restricted by its distribution and variance. In this case, the variance of the outcome is given by

$$V(y_i) = \mu_i(1-\mu_i), \quad 0 < \mu_i < 1. \tag{6}$$

We have to parameterize using logit link to transform the linear predictor to enforce a range (0, 1). The form is

$$g(\mu_i) \quad = \quad \ln\frac{\mu_i}{1-\mu_i} = \theta_i = \mathbf{x}_i\boldsymbol{\beta}. \tag{7}$$

Use the chain rule to obtain the likelihood equations $\boldsymbol{\Psi}(\boldsymbol{\beta}) = \mathbf{0}$ that derived as the derivatives of the log likelihood as follows

$$\begin{aligned}
\frac{\partial \ell(\boldsymbol{\beta})}{\partial \beta_j} &= \left[\left(\frac{\partial \ell(\boldsymbol{\mu})}{\partial \theta_i}\right)\left(\frac{\partial \theta_i}{\partial \mu_i}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\left(\frac{\partial \eta_i}{\beta_j}\right)\right] \\
&= \sum_{i=1}^{n}\left(\frac{y_i-\mu_i}{\mu_i(1-\mu_i)}\right)\left(\frac{\partial \mu_i}{\partial \eta_i}\right)\mathbf{x}_i \qquad (8) \\
&= \mathbf{0}, \quad j = 1, \ldots, p. \qquad (9)
\end{aligned}$$

2.3 Parameter Estimations

The solutions of the likelihood equations are obtained using optimization techniques. These techniques iterate toward a solution by updating a set of current estimates to a new set of estimates. Let the k-step set of estimates denote $\boldsymbol{\beta}^{(k)}$. First, we give a initial set of estimates $\boldsymbol{\beta}^{(0)}$. Then the common approach employs a Taylor series expansion of estimating equations given by $\boldsymbol{\Psi}(\boldsymbol{\beta}) = \mathbf{0}$, such that

$$\mathbf{0} \quad = \quad \boldsymbol{\Psi}\left(\boldsymbol{\beta}^{(0)}\right) + \left(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right)\boldsymbol{\Psi}'\left(\boldsymbol{\beta}^{(0)}\right) + \frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right)^2\boldsymbol{\Psi}''\left(\boldsymbol{\beta}^{(0)}\right) + \text{(ommitted terms)}$$

Keeping only the first two terms, we have the linear approximation

$$\begin{aligned}
\mathbf{0} &\approx \boldsymbol{\Psi}\left(\boldsymbol{\beta}^{(0)}\right) + \left(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)}\right)\boldsymbol{\Psi}'\left(\boldsymbol{\beta}^{(0)}\right), \\
\boldsymbol{\beta} &\approx \boldsymbol{\beta}^{(0)} - \frac{\boldsymbol{\Psi}\left(\boldsymbol{\beta}^{(0)}\right)}{\boldsymbol{\Psi}'\left(\boldsymbol{\beta}^{(0)}\right)}. \qquad (10)
\end{aligned}$$

Rewriting this relationship of (10) in matrix notation, then we iterate to obtain a solution using the relationship

$$\begin{aligned}\boldsymbol{\beta}^{(k)} &= \boldsymbol{\beta}^{(k-1)} - \left[\frac{\partial}{\partial\boldsymbol{\beta}}\boldsymbol{\Psi}\left(\boldsymbol{\beta}^{(k-1)}\right)\right]^{-1}\boldsymbol{\Psi}\left(\boldsymbol{\beta}^{(k-1)}\right)\\ &= \boldsymbol{\beta}^{(k-1)} - \mathbf{H}^{-1}\mathbf{s},\end{aligned}\tag{11}$$

where $\mathbf{H}$ is the Hessian (second derivative) matrix, and $\mathbf{s}$ is the gradient (first derivative) vector of estimating equation, both evaluated at the current value of the parameter vector. Thus, given a set of initial estimates $\boldsymbol{\beta}^{(0)}$ as the starting step, we update our set of estimates using the iterated relationship in equation (11).

2.4 Generalized Estimating Equations (GEEs)

Liang and Zeger (1986) proposed GEEs methodology to extend the OLR model for the analysis of correlated binary data. The estimating equations of this method are related to quasilikelihood methods in that there are no parametric assumptions. The utility of this method is extended outside of the implied log-likelihood due to the work of Wedderburn (1974). We started the introduction of quasilikelihood method and working correlation structure.

2.4.1 Estimating Equation

Wedderburn extended the result by assuming that the variance function is a known function of mean. Hardin & Hilbe (2003) were therefore free to choose any parameterization of the mean and variance function and apply them in the derived estimating equation. That implied that we choose functions are not from an exponential family member. Resulting coefficient estimates are properly called maximum quasilikelihood estimates. It is a generalization of the likelihood. We can use the quasilikelihood estimating equation with no restriction on the choice of the mean and variance functions. The estimation equation were defined as follows.

Let Y_{ij} denote the outcome of jth measurement on ith subject for $i = 1, \ldots, n$ and $j = 1, \ldots, t$. Then, we let $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{it}]'$ represent the vector of measurements on the ith subject and $\boldsymbol{\mu}_i = [\mu_{i1}, \ldots, \mu_{it}]'$ be the corresponding vector of means. In addition, let $\mathbf{V}_i$ be the covariance matrix of $\mathbf{Y}_i$. The estimator vector of the parameters, $\boldsymbol{\beta}$, is found by solving the estimating equation.

$$\Psi(\boldsymbol{\beta}) = \sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0, \quad i = 1, \ldots, n, \tag{12}$$

where $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}'$ and the covariance matrix for ith subject is given by

$$\mathbf{V}_i \;=\; \mathbf{A}_i^{1/2} \mathbf{R}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{1/2} \tag{13}$$

where $\mathbf{A}_i$ is a $n \times n$ diagonal matrix with $v(\mu_{ij})$ as the jth diagonal element and $\mathbf{R}_i(\boldsymbol{\alpha})$ is the "working" correlation matrix. It is estimated in iterative fitting process using the current value of the parameter vector $\boldsymbol{\beta}$ to compute appropriate functions of the Pearson residual.

$$e_{ij} \;=\; \frac{y_{ij} - \mu_{ij}}{\sqrt{v(\mu_{ij})}} \tag{14}$$

2.4.2 Working correlation structure

We defined the dispersion parameter first.

$$\hat{\phi} \;=\; \frac{1}{N-p} \sum_{i=1}^{n} \sum_{j=1}^{t} e_{ij}^2, \quad i = 1, \ldots, n \quad \text{and} \quad j = 1, \ldots, t, \tag{15}$$

where $N = n \times t$ is the total number of measurements and p is the number of parameters in model. The working correlation structures are categorized as follows

1. Independent structure

   The correlation matrix is build from the independent structure. The correlation structure is assumed to be

$$\mathrm{Corr}(Y_{ij}, Y_{ij'}) = \begin{cases} 1, & j = j'; \\ 0, & j \neq j', \end{cases} \tag{16}$$

   The structure is as follow.

$$\mathbf{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}. \tag{17}$$

2. AR(1) structure

   It may be more reasonable to assume a time dependence for the association if the repeated observations within the subjects have natural order. The correlation structure is assumed to be $\mathrm{corr}(y_{ij}, y_{ij'}) = \alpha^{|j-j'|}$. In this case, the correlation matrix is build from the AR(1) structure. The structure is as follow.

$$\mathbf{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha^1 & \dots & \alpha^{t-1} \\ \alpha^1 & 1 & \dots & \alpha^{t-2} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha^{t-1} & \alpha^{t-2} & \dots & 1 \end{bmatrix}, \tag{18}$$

   where

$$\hat{\alpha} = \frac{1}{(N^\star - p)\phi} \sum_{i=1}^{n} \sum_{j \leq t-1} e_{ij} e_{i,j+1}, \quad N^\star = n(t-1). \tag{19}$$

3. m-dependent structure

In a special case, $m = 2$, the working correlation structure is as follows

$$\mathbf{R}(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 & 0 & \dots & 0 \\ \alpha_1 & 1 & \alpha_1 & \alpha_2 & \dots & 0 \\ \alpha_2 & \alpha_1 & 1 & \alpha_1 & \dots & 0 \\ 0 & \alpha_2 & \alpha_1 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}, \tag{20}$$

which we can write the ijth cell of the matrix as follows

$$\text{Corr}(Y_{ij}, Y_{ij'}) = \begin{cases} 1, & |j - j'| = 0; \\ \alpha_{|j-j'|}, & |j - j'| = 1, 2, \dots, m; \\ 0, & |j - j'| > m, \end{cases} \tag{21}$$

where

$$\hat{\alpha}_k = \frac{1}{(N_k - p)\phi} \sum_{i=1}^{n} \sum_{j \le t-k} e_{ij} e_{i,j+k}, \quad N_k = n(n - k). \tag{22}$$

2.4.3 Estimating algorithm

Liang and Zeger (1986) computed a set of estimates $\hat{\boldsymbol{\beta}}$ and based on iterating between a modified Fisher scoring for $\hat{\boldsymbol{\beta}}$ and moment estimation $\hat{\boldsymbol{\alpha}}$ and $\hat{\phi}$. The modified iterative procedure for $\hat{\boldsymbol{\beta}}$ is as follows.

1. Compute initial estimates of $\boldsymbol{\beta}$ with ordinary generalized linear model assuming independence.

2. Compute the working correlations $\mathbf{R}(\boldsymbol{\alpha})$ based on the standardized residuals, the current $\boldsymbol{\beta}$, and the a pre assumed working correlation structure $\mathbf{R}(\boldsymbol{\alpha})$, for example, "independent", "AR(1)", or "m-dependent".

3. Compute an estimate of the covariance:

$$\hat{\mathbf{V}}_i \quad = \quad \phi \mathbf{A}_i^{\frac{1}{2}} \hat{\mathbf{R}}_i(\boldsymbol{\alpha}) \mathbf{A}_i^{\frac{1}{2}} \tag{23}$$

4. Update $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}^{(k+1)} \quad = \quad \hat{\boldsymbol{\beta}}^{(k)} - [\sum_{i=1}^{n} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} \hat{\mathbf{D}}_i]^{-1} [\sum_{i=1}^{n} \hat{\mathbf{D}}_i' \hat{\mathbf{V}}_i^{-1} (\mathbf{Y}_i - \hat{\boldsymbol{\mu}}_i)] \tag{24}$$

5. Iterate steps 2-4 until convergence.

2.4.4 Empirical variance estimators

An empirical variance estimator can be used to estimate covariance matrix of $\hat{\boldsymbol{\beta}}$. This variance estimator is also referred to as a "sandwich" or "robust" estimator. The empirical estimator covariance matrix of $\hat{\boldsymbol{\beta}}$ is given by

$$\hat{\mathbf{V}}_{\mathbf{R}} \quad = \quad \mathbf{I}_0^{-1} \mathbf{I}_1 \mathbf{I}_0^{-1}, \tag{25}$$

where

$$\mathbf{I}_0 \quad = \quad \sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}} \tag{26}$$

$$\mathbf{I}_1 \quad = \quad \sum_{i=1}^{n} \mathbf{D}_i' \mathbf{V}_i^{-1} \mathrm{Cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \frac{\partial \boldsymbol{\mu}_i}{\partial \boldsymbol{\beta}}. \tag{27}$$

It has the property of consistent estimator of the covariance matrix of $\hat{\boldsymbol{\beta}}$, even if the working correlation matrix is misspecified, that is, if $Cov(\mathbf{Y}_i) \neq \mathbf{V}_i$ then is replaced by an estimate, such as

$$(Y_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}}))'(Y_i - \boldsymbol{\mu}_i(\hat{\boldsymbol{\beta}})) \tag{28}$$

2.5 Diagnostics and Testing

Pan (2001) introduced a useful criterion measure that is qusilikeoihood information criterion (QIC). The measure is called QIC and we can use it to choose between several competing correlation structures. It is defined as follows

$$\text{QIC} = -2Q\left(g^{-1}(\mathbf{x}\hat{\boldsymbol{\beta}})\right) + 2\text{trace}(\hat{\mathbf{A}}_{\mathbf{I}}^{-1}\hat{\mathbf{V}}_{\mathbf{R}}) \tag{29}$$

where $Q(y; g^{-1}(\mathbf{x}\hat{\boldsymbol{\beta}}))$ is the value of the quasilikelihood computed using the coefficients from the model with the given correlation structure. We can calculate $Q(y; g^{-1}(\mathbf{x}\hat{\boldsymbol{\beta}}))$ as follows

$$Q(y; g^{-1}(\mathbf{x}\hat{\boldsymbol{\beta}})) = \sum_{i=1}^{n}\sum_{j=1}^{t}\{y_{ij}\log(\hat{\mu}_{ij}/(1-\hat{\mu}_{ij})) + \log(1-\hat{\mu}_{ij})\}. \tag{30}$$

In addition, $\hat{\mathbf{A}}_{\mathbf{I}}$ is the covariance matrix obtained by fitting an independence model. It is defined by

$$\hat{\mathbf{A}}_{\mathbf{I}} = -\mathbf{H}^{-1}, \tag{31}$$

where $\mathbf{H}$ is the Hessian matrix evaluated using the parameter estimates on the last iteration. $\hat{\mathbf{V}}_{\mathbf{R}}$ is based on equation (25) that means the modified sandwich estimate of variance from the model with the given correlation structure.

3 RESEARCH METHODS

3.1 Participants

Fifty-four undergraduate and graduate students from Tunghai University voluntarily participated in this study. The participant pool consisted of 45 females and 9 males, ranging in age of surfing internet from 0.5 to 14.5 years.

3.2 Apparatus and Materials

This study used a Pentium IV desktop computer with Microsoft's Internet Explore 6.0. The animated online ads used Microsoft MSpaint, Microsoft Frontpage 2000, Macromedia Dreamweaver 4.0 (2000), Ulead PhotoImpact 6.0 (2000), and Ulead Graphic Interchange Format (GIF) Animator 4.0. Information gathered from each participant included a background questionnaire and a browsing task with the records of click-through. Each participant continuously viewed 15 different ads imposed on a single Web page with different e-news. A digital video camera recorder (SONY DCR-PC330) was used to record every movement of a mouse that happened in the process of experiment. Ads were positioned in the "High", "Middle", and "Low" pages, approximately (see Figure 1, 2, 3, respectively). When designing web pages,"High" is defined as absolute position, and it's coordinate is "left=694, and top=0." "Middle" coordinate is "left=694, and top=268." "Low" coordinate is "left=694, and top=518." Sizes of advertisements were standard medium rectangle Interactive Marketing Unit (300×250 IMU) in accordance with the Internet Advertising Bureau (IAB) size guidelines for advertisements on the Web (http://www.iab.net/standards/adunits.asp/). Each advertisement was manipulated to be 3 different animation lengths (7.5, 15, and 30 seconds) and 3 different exposure times (1, 3, and more than 3 cyclical times) with company logos and products constantly displayed. Animation on the online advertisement consisted of flashing of words and graphs which were shown as 3 different exposure times 3 different animation lengths, and 3 different positions. Based on the factorial design with complete randomization in repeat, each subject was randomly assigned one of twenty-seven treatment combinations.

3.3 Experimental Procedure

Participants were shown a single Web page and asked to complete a browsing task one page at a time. 15 different online ads will accompany with e-news shown on the Web page. Before beginning, participants were given a few minutes to familiarize themselves with the content of the page. Participants were informed that all information needed to complete the browsing tasks was present within the web pages, including free clicking through the ad shown in the same page. Each subject was asked for viewing 15 different online ads and then the data with or without click-through would be taken by a digital video camera recorder (SONY DCR-PC330). After the task was completed, participants were asked to respond to the reasons of clicking through ads or not in terms of after-experiment questionnaire.

3.4 Experimental Design

Three-factor factorial with repeated measures was used in this study to collect the click-through data of repeated measures. Doyle et al. proposed a theory as follows: click through will be greater if the banner ad is placed approximately 1/3 of the way down

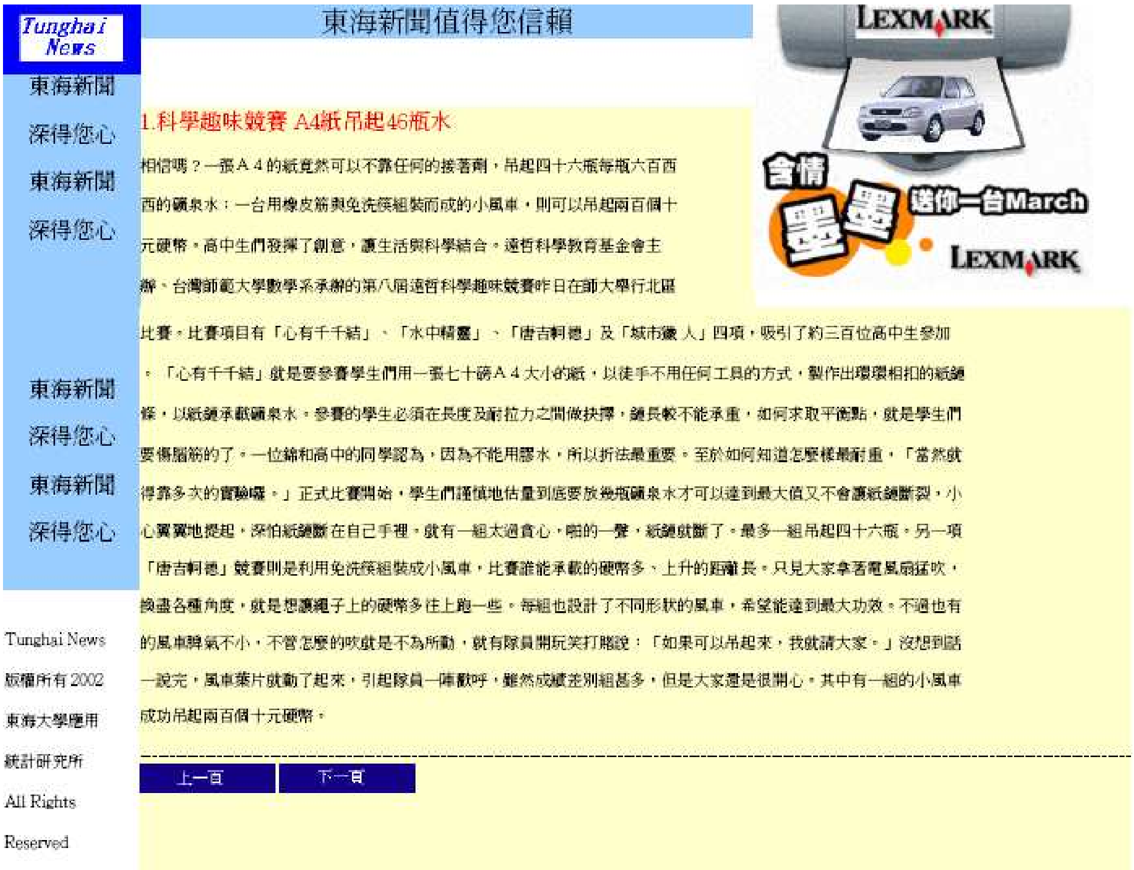

Figure 1: Advertising medium rectangle put on the high position of Web page in this study.

the page. Based on their results, the placement of the ad down the screen would increase click-through 77% for the Photodisc ad, however, the Webreference ad showed the same trend, but the result was not statistically significant. In this study, the medium rectangle ads were placed on the three different positions (high, middle, and low placements). The other factors are design factors of animation ads, animation length and exposure times of advertising. The three levels of animation length and exposure times are 7.5, 15, and 30 seconds as well as 1, 3, and more than 3 times respectively. There are totally twenty-seven treatment combinations in this experiment. Two subjects as the experimental units would be randomly assigned to one of twenty-seven treatment combinations as the replicates.

Figure 2: Advertising medium rectangle put on the middle position of Web page in this study.

Fifteen different advertisings would be exposed in 3 different ad contents, commodities information, sale promotion, and entertainments which are arranged randomly and each consisting of 5 different ads for each subject and the corresponding click through data of repeated measures would be collected from this experiment.

3.5 Proposed Models

In this study, each subject will be assigned 15 different advertising and their click-through data will also be recorded. Using the notation $(\mathbf{x}_i,\ y_{ij})$, $i = 1, \ldots, 54$, $j =$

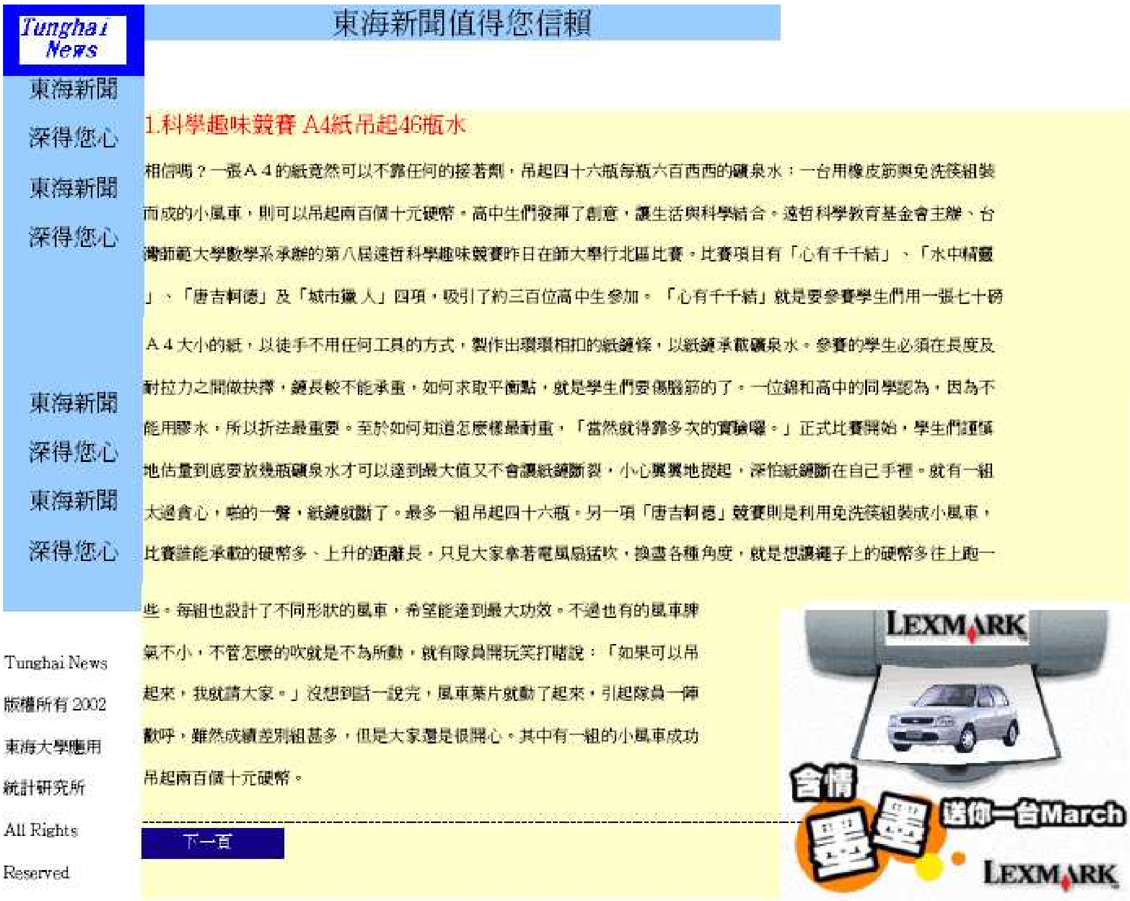

Figure 3: Advertising medium rectangle put on the low position of Web page in this study.

$1, \ldots, 15$, to denote the covariates vector of dimensions p corresponding to the ith subject and the jth binary outcome of clicking through for the ith subject. We start by supposing that the distribution of Y_{ij} is Bernoulli, given an covariate vector $\mathbf{x}_i$ as follows.

$$
\begin{aligned}
f(y_{ij}|\mathbf{x}_i) &= \pi_{ij}^{y_{ij}}(1-\pi_{ij})^{1-y_{ij}} \\
&= \exp\left[y_{ij}\log\frac{\pi_{ij}}{1-\pi_{ij}} + \log(1-\pi_{ij})\right] \\
&= \exp\left[y_{ij}\theta_{ij} - \log\{1+\exp(\theta_{ij})\}\right]. \qquad (32)
\end{aligned}
$$

We denote by π_{ij} be the probability of success at time j for the ith subject. It implies that $\pi_{ij} = P(Y_{ij} = 1)$. Then, we assume multiple logistic model with GEEs approach as follows

$$\begin{aligned}\text{logit}(\pi_{ij}) &= \theta_{ij} \\ &= \log\left[\pi_{ij}/(1-\pi_{ij})\right] \\ &= \mathbf{x}_i'\boldsymbol{\beta}, \qquad (33)\end{aligned}$$

where $\boldsymbol{\beta}$ is a vector of model parameters. The covariates vector consists of the order of repeated measures, gender, position, animation length , exposure times, the two-factor interaction with position and animation length, the two-factor interaction with position and exposure times, and the two-factor interaction with animation length and exposure times. Table 1 showed the coding sheet describing the variables of the CTR data set. In this study, we use GEE approach to fit the logistic regression models and apply to CTR data.

Base on the equation (33), we rewrite the GEE model with p category predictors as follows,

$$\text{logit } P(Y_{ij}=1|x_{i1},\ldots,x_{ip}) = \beta_0+\beta_1 x_{i1}+\ldots+\beta_p x_{ip}, \qquad (34)$$

such that the odds ratio given by

$$\text{OR} = \frac{P(Y_{ij}=1|x_{i1},\ldots,x_{i,p-1},x_{ip}=1)/P(Y_{ij}=0|x_{i1},\ldots,x_{i,p-1},x_{ip}=1)}{P(Y_{ij}=1|x_{i1},\ldots,x_{i,p-1},x_{ip}=0)/P(Y_{ij}=0|x_{i1},\ldots,x_{i,p-1},x_{ip}=0)} = \exp(\beta_p).$$

It represents the ratio of the odds of a given predictors containing pth variable if the outcome of response equals 1 compared to the odds of a given predictors excluding pth variable if the outcome of response equals 0. Next, we define dummy variable to all category predictors.

Table 1: Variables in the CTR Data Set

Name(Abbreviation)	Description	Codes/Values
Order(O)	the order of the repeated measurements	$1, 2, \ldots, 15$
Gender(G)	the gender of the participants	1 = Female
		2 = Male
Type	the type of advertising	1 = Commodities Information
		2 = Sale Promotion
		3 = Entertainments
Position(P)	the layout position of a middle	1 = High
	rectangular advertising	2 = Middle
		3 = Low
Animation Length(L)	the animation length of ad	7.5, 15, and 30 seconds
Exposure Times(T)	the exposure times of advertising	1, 3, and more than 3 times

3.5.1 GEE model with categorical predictors

A GEE model was fitted containing variables for order, gender, two dummy variables for type, two dummy variables for position , two dummy variables for animation length,and two dummy variables for exposure times. Define the notation of the covariates in the logistic regression model with categorical factors as follows,

$$
\begin{aligned}
G &= \begin{cases} 1, & \text{if gender is female;} \\ 0, & \text{else.} \end{cases} \\
\mathrm{Type}_1 &= \begin{cases} 1, & \text{if ad type is commodities information;} \\ 0, & \text{else.} \end{cases} \\
\mathrm{Type}_2 &= \begin{cases} 1, & \text{if ad type is sale promotion;} \\ 0, & \text{else.} \end{cases}
\end{aligned}
$$

$$P_1 = \begin{cases} 1, & \text{if ad position is high;} \\ 0, & \text{else.} \end{cases}$$

$$P_2 = \begin{cases} 1, & \text{if ad position is middle;} \\ 0, & \text{else.} \end{cases}$$

$$L_1 = \begin{cases} 1, & \text{if animation length is 7.5 seconds;} \\ 0, & \text{else.} \end{cases}$$

$$L_2 = \begin{cases} 1, & \text{if animation length is 15 seconds;} \\ 0, & \text{else.} \end{cases}$$

$$T_1 = \begin{cases} 1, & \text{if exposure times is 1 time;} \\ 0, & \text{else.} \end{cases}$$

$$T_2 = \begin{cases} 1, & \text{if exposure times is 3 times;} \\ 0, & \text{else.} \end{cases}$$

3.5.2 Parameter estimation

In this study, SAS/STAT Release 9.1 was used to fit the proposed model. "PROC GENMOD" provides a support of GEE in the logistic regression model. "REPEATED" statement followed by the "MODEL" statement in GENMOD procedure was used to specify the working correlation structure including "Independent", "AR(1)" or "m-dependent". The parameters estimates of OLR model are used to be the initial values of iteratively fitting algorithm in the GEE model.

3.5.3 Strategies of model selection

Traditionally, comparison of observed to predicted values in logistic regression is based on the log likelihood function. Akaike's information criterion (AIC) is one useful measure to assess the goodness-of-fit of fitted model. QIC is introduced to apply in assessing quasilikelihood based model by Pan (2001). In this study, we used QIC measures to assess goodness-of-fit of fitted model.

First, we began the complex model to fit model with three types working correlation structures, respectively. Second, remove nonsignificant parameters in the fitted model. Repeated the second step until all parameter of fitted model is significant. Then, we calculated the QIC of final fitted model with three different working correlation structures. Final, we chose the final model with the smallest QIC measure.

After choosing the final model, we draw the scatter plot of raw residual versus fitted value. Use the plot to check the fitted model.

4 RESULTS

According to the results of the experiment we've collected 15 repeated binary observations from the 54 subjects. We began a complex model of multiple logistic regression by GEEs approach with "Independent", "AR(1)" and "m-dependent ($m = 2$)" working correlation structures, respectively. The model is

$$\begin{aligned}
\text{logit}(\pi_{ij}) \quad = \quad & \beta_0 + \beta_1 O + \beta_2 G + \beta_3 \text{Type}_1 + \beta_4 \text{Type}_2 + \beta_5 P_1 + \beta_6 P_2 + \beta_7 \text{Type}_1 \times P_1 \\
& + \beta_8 \text{Type}_1 \times P_2 + \beta_9 \text{Type}_2 \times P_1 + \beta_{10} \text{Type}_2 \times P_2 + \beta_{11} L_1 + \beta_{12} L_2 \\
& + \beta_{13} \text{Type}_1 \times L_1 + \beta_{14} \text{Type}_1 \times L_2 + \beta_{15} \text{Type}_2 \times L_1 + \beta_{16} \text{Type}_2 \times L_2 \\
& + \beta_{17} T_1 + \beta_{18} T_2 + \beta_{19} \text{Type}_1 \times T_1 + \beta_{20} \text{Type}_1 \times T_2 + \beta_{21} \text{Type}_2 \times T_1 \\
& + \beta_{22} \text{Type}_2 \times T_2 + \beta_{23} P_1 \times L_1 + \beta_{24} P_1 \times L_2 + \beta_{25} P_2 \times L_1 + \beta_{26} P_2 \times L_2 \\
& + \beta_{27} P_1 \times T_1 + \beta_{28} P_1 \times T_2 + \beta_{29} P_2 \times T_1 + \beta_{30} P_2 \times T_2 + \beta_{31} L_1 \times T_1 \\
& + \beta_{32} L_1 \times T_2 + \beta_{33} L_2 \times T_1 + \beta_{34} L_2 \times T_2. \qquad (35)
\end{aligned}$$

The parameters estimates ,the corresponding standard errors estimates , z-value, and p-value in model (35) are showed in Table 2. Remove all nonsignificant parameters in Table 2 and refit a model as follows.

$$\begin{aligned}
\text{logit}(\pi_{ij}) \quad = \quad & \beta_0 + \beta_1 O + \beta_2 G + \beta_3 \text{Type}_1 + \beta_4 \text{Type}_2 + \beta_5 P_1 + \beta_6 P_2 + \beta_7 \text{Type}_1 \times P_1 \\
& + \beta_8 \text{Type}_1 \times P_2 + \beta_9 \text{Type}_2 \times P_1 + \beta_{10} \text{Type}_2 \times P_2 + \beta_{11} L_1 + \beta_{12} L_2 \\
& + \beta_{13} P_1 \times L_1 + \beta_{14} P_1 \times L_2 + \beta_{15} P_2 \times L_1 + \beta_{16} P_2 \times L_2. \qquad (36)
\end{aligned}$$

The parameters estimates, their standard error estimates, z-value, and p-value in model (36) are showed in Table 3. Remove gender and refit a model as follows.

Table 2: GEE parameter estimates, empirical standard error estimates, z-value, and p-value for equation (35), including three types of working correlation structures.

	Independent				m-dependent ($m = 2$)				AR(1)			
Parameters	Estimate	SE	Z	Pr > $\|Z\|$	Estimate	SE	Z	Pr > $\|Z\|$	Estimate	SE	Z	Pr > $\|Z\|$
Intercept	-2.9326	1.1523	-2.54	0.0109*	-2.8306	1.1473	-2.47	0.0136*	-2.8392	1.1123	-2.55	0.0107*
O	0.1113	0.0200	5.56	<.0001*	0.1253	0.0220	5.70	<.0001*	0.1181	0.0202	5.86	<.0001*
G	1.3046	0.4897	2.66	0.0077*	1.3475	0.5243	2.57	0.0102*	1.2990	0.4975	2.61	0.0090*
Type_1	0.6078	0.6934	0.88	0.3808	-0.1937	0.6839	-0.28	0.7770	0.2671	0.6254	0.43	0.6693
Type_2	-0.2016	0.5636	-0.36	0.7206	-0.5014	0.7565	-0.66	0.5075	-0.3991	0.5548	-0.72	0.4720
P_1	-0.6866	0.9997	-0.69	0.4922	-0.7818	0.9862	-0.79	0.4279	-0.6023	0.9869	-0.61	0.5417
P_2	-1.8531	1.1164	-1.66	0.0969	-1.8346	1.1028	-1.66	0.0962	-1.6828	1.1159	-1.51	0.1316
$\text{Type}_1 \times P_1$	-0.1620	0.4818	-0.34	0.7366	0.2693	0.5214	0.52	0.6055	-0.1899	0.4899	-0.39	0.6982
$\text{Type}_1 \times P_2$	0.0515	0.6112	0.08	0.9329	0.3794	0.6210	0.61	0.5413	-0.0301	0.5837	-0.05	0.9589
$\text{Type}_2 \times P_1$	0.2446	0.4961	0.49	0.6220	0.7483	0.5679	1.32	0.1877	0.3617	0.5122	0.71	0.4800
$\text{Type}_2 \times P_2$	1.3255	0.4605	2.88	0.0040*	1.5788	0.6117	2.58	0.0099*	1.2502	0.5286	2.36	0.0180*
L_1	-0.3525	1.0807	-0.33	0.7443	-0.4408	1.0793	-0.41	0.6830	-0.4645	1.0551	-0.44	0.6598
L_2	0.5441	1.3148	0.41	0.6790	0.3792	1.2710	0.30	0.7654	0.4092	1.2918	0.32	0.7514
$\text{Type}_1 \times L_1$	-0.1019	0.6655	-0.15	0.8784	0.1021	0.6879	0.15	0.8821	0.0727	0.6537	0.11	0.9115
$\text{Type}_1 \times L_2$	0.1046	0.5090	0.21	0.8372	0.2472	0.4883	0.51	0.6127	0.2543	0.4624	0.55	0.5823
$\text{Type}_2 \times L_1$	-0.4436	0.4859	-0.91	0.3612	-0.1477	0.6686	-0.22	0.8252	-0.2511	0.5486	-0.46	0.6471
$\text{Type}_2 \times L_2$	-0.4295	0.4260	-1.01	0.3133	-0.0463	0.4786	-0.10	0.9229	-0.2296	0.4287	-0.54	0.5923
T_1	0.7636	1.0501	0.73	0.4671	0.7095	1.0221	0.69	0.4876	0.6688	1.0155	0.66	0.5101
T_2	0.6586	1.1499	0.57	0.5668	0.4957	1.1812	0.42	0.6747	0.4805	1.1320	0.42	0.6712
$\text{Type}_1 \times T_1$	-0.7407	0.5825	-1.27	0.2036	-0.4253	0.5960	-0.71	0.4755	-0.6453	0.5466	-1.18	0.2378
$\text{Type}_1 \times T_2$	-0.6698	0.5046	-1.33	0.1844	-0.5670	0.5397	-1.05	0.2934	-0.3942	0.4940	-0.80	0.4248
$\text{Type}_2 \times T_1$	-0.8607	0.4558	-1.89	0.0590	-0.7287	0.5677	-1.28	0.1993	-0.7547	0.4643	-1.63	0.1041
$\text{Type}_2 \times T_2$	-0.2469	0.4888	-0.51	0.6135	0.0729	0.5568	0.13	0.8958	-0.1732	0.4911	-0.35	0.7243
$P_1 \times L_1$	1.4069	0.9832	1.43	0.1525	1.2616	0.9861	1.28	0.2008	1.2944	0.9817	1.32	0.1873
$P_1 \times L_2$	1.1534	1.0561	1.09	0.2748	1.1482	1.0365	1.11	0.2680	1.1308	1.0530	1.07	0.2829
$P_2 \times L_1$	2.6556	0.9256	2.87	0.0041*	2.5042	0.9401	2.66	0.0077*	2.5252	0.9307	2.71	0.0067*
$P_2 \times L_2$	1.1163	1.2802	0.87	0.3832	0.9665	1.3020	0.74	0.4579	1.0471	1.2876	0.81	0.4161
$P_1 \times T_1$	0.7934	1.0510	0.75	0.4503	0.5427	1.0337	0.53	0.5996	0.7010	1.0441	0.67	0.5019
$P_1 \times T_2$	-0.8903	1.1646	-0.76	0.4446	-0.9114	1.1840	-0.77	0.4414	-0.8667	1.1724	-0.74	0.4598
$P_2 \times T_1$	0.2899	1.3487	0.21	0.8298	0.1800	1.3602	0.13	0.8947	0.2744	1.3367	0.21	0.8373
$P_2 \times T_2$	1.0494	1.1034	0.95	0.3416	0.8874	1.1293	0.79	0.4320	0.9754	1.1184	0.87	0.3831
$L_1 \times T_1$	-1.3509	1.0446	-1.29	0.1960	-1.3214	1.0735	-1.23	0.2183	-1.2126	1.0423	-1.16	0.2447
$L_1 \times T_2$	-1.0141	1.0723	-0.95	0.3443	-1.1025	1.1260	-0.98	0.3275	-0.9902	1.0783	-0.92	0.3584
$L_2 \times T_1$	-1.3255	1.2160	-1.09	0.2757	-1.3176	1.2074	-1.09	0.2751	-1.2743	1.2109	-1.05	0.2926
$L_2 \times T_2$	-1.4277	1.1457	-1.25	0.2127	-1.2893	1.1536	-1.12	0.2637	-1.3440	1.1421	-1.18	0.2393

Note: " * " denotes statistically significance at $\alpha = 0.05$

Table 3: GEE parameter estimates, empirical standard error estimates, z-value, and p-value for equation (36), including three types of working correlation structures.

	Independent				m-dependent ($m = 2$)				AR(1)			
Parameters	Estimate	SE	Z	Pr > $\|Z\|$	Estimate	SE	Z	Pr > $\|Z\|$	Estimate	SE	Z	Pr > $\|Z\|$
Intercept	-2.1705	0.8075	2.69	0.0072*	-2.1482	0.8457	-2.54	0.0111*	-2.2025	0.7968	-2.76	0.0057*
O	0.1179	0.0216	5.45	<.0001*	0.1297	0.0225	5.75	<.0001*	0.1211	0.0211	5.74	<.0001*
G	0.9368	0.5901	1.59	0.1124	0.9820	0.6111	1.61	0.1080	0.9447	0.5812	1.63	0.1041
Type_1	0.1696	0.3884	0.44	0.6624	-0.4967	0.4294	-1.16	0.2474	0.0285	0.4019	0.07	0.9434
Type_2	-0.7052	0.3715	1.90	0.0577	-0.7148	0.4939	-1.45	0.1478	-0.7680	0.4137	-1.86	0.0634
P_1	-0.4334	0.7463	0.58	0.5614	-0.7431	0.7742	-0.96	0.3372	-0.4213	0.7393	-0.57	0.5687
P_2	-1.0465	0.7132	1.47	0.1423	-1.3742	0.7097	-1.94	0.0528	-1.0088	0.6843	-1.47	0.1404
$\text{Type}_1 \times P_1$	-0.3288	0.4556	0.72	0.4705	0.2783	0.4984	0.56	0.5766	-0.3058	0.4653	-0.66	0.5110
$\text{Type}_1 \times P_2$	-0.1289	0.5576	0.23	0.8172	0.4388	0.5713	0.77	0.4424	-0.1244	0.5381	-0.23	0.8172
$\text{Type}_2 \times P_1$	0.0033	0.4783	0.01	0.9946	0.6709	0.5617	1.19	0.2324	0.1915	0.4919	0.39	0.6971
$\text{Type}_2 \times P_2$	1.0305	0.4419	2.33	0.0197*	1.5615	0.5787	2.70	0.0070*	1.0520	0.4968	2.12	0.0342*
L_1	-1.2541	0.6536	1.92	0.0550	-1.2761	0.6806	-1.87	0.0608	-1.2198	0.6568	-1.86	0.0633
L_2	-0.3622	0.9353	0.39	0.6986	-0.3303	0.9207	-0.36	0.7198	-0.3211	0.9323	-0.34	0.7305
$P_1 \times L_1$	1.0325	1.1017	0.94	0.3487	0.9587	1.1257	0.85	0.3944	0.9654	1.0996	0.88	0.3799
$P_1 \times L_2$	0.9347	1.1191	0.84	0.4036	0.9222	1.1196	0.82	0.4101	0.9087	1.1124	0.82	0.4140
$P_2 \times L_1$	2.3719	0.9256	2.56	0.0104*	2.3560	0.9445	2.49	0.0126*	2.3085	0.9131	2.53	0.0115*
$P_2 \times L_2$	0.8858	1.1890	0.74	0.4563	0.8212	1.1913	0.69	0.4906	0.8357	1.1894	0.70	0.4823

Note: " * " denotes statistically significance at $\alpha = 0.05$

Table 4: GEE parameter estimates, empirical standard error estimates, z-value, and p-value of equation (37), including three types of working correlation structures.

	Independent				m-dependent ($m=2$)				AR(1)			
Parameters	Estimate	SE	Z	Pr $>\|Z\|$	Estimate	SE	Z	Pr $>\|Z\|$	Estimate	SE	Z	Pr $>\|Z\|$
Intercept	-1.3651	0.6580	-2.07	0.0380*	-1.2815	0.6728	-1.90	0.0568	-1.3902	0.6458	-2.15	0.0313*
O	0.1156	0.0207	5.58	<.0001*	0.1283	0.0218	5.88	<.0001*	0.1189	0.0202	5.90	<.0001*
Type_1	0.1915	0.3826	0.50	0.6168	-0.5456	0.4485	-1.22	0.2238	0.0395	0.4002	0.10	0.9213
Type_2	-0.7019	0.3749	-1.87	0.0611	-0.6867	0.5131	-1.34	0.1808	-0.7630	0.4237	-1.80	0.0718
P_1	-0.3136	0.7959	-0.39	0.6936	-0.6461	0.8371	-0.77	0.4402	-0.2961	0.7907	-0.37	0.7081
P_2	-1.0152	0.7887	-1.29	0.1981	-1.3848	0.7927	-1.75	0.0807	-0.9738	0.7615	-1.28	0.2009
$\text{Type}_1 \times P_1$	-0.2969	0.4492	-0.66	0.5086	0.3422	0.5228	0.65	0.5128	-0.2811	0.4635	-0.61	0.5442
$\text{Type}_1 \times P_2$	-0.1571	0.5569	-0.28	0.7778	0.4951	0.5884	0.84	0.4001	-0.1494	0.5389	-0.28	0.7816
$\text{Type}_2 \times P_1$	0.0405	0.4804	0.08	0.9328	0.7280	0.5817	1.25	0.2107	0.2352	0.4974	0.47	0.6363
$\text{Type}_2 \times P_2$	1.0214	0.4459	2.29	0.0220*	1.5658	0.6043	2.59	0.0096*	1.0365	0.5088	2.04	0.0416*
L_1	-1.2261	0.6678	-1.84	0.0663	-1.2669	0.7033	-1.80	0.0716	-1.1959	0.6683	-1.79	0.0735
L_2	-0.2197	0.9560	-0.23	0.8182	-0.1964	0.9464	-0.21	0.8356	-0.1779	0.9527	-0.19	0.8518
$P_1 \times L_1$	0.6006	1.0718	0.56	0.5752	0.5331	1.1107	0.48	0.6312	0.5356	1.0696	0.50	0.6166
$P_1 \times L_2$	0.6525	1.1313	0.58	0.5641	0.6439	1.1335	0.57	0.5700	0.6246	1.1223	0.56	0.5779
$P_2 \times L_1$	2.1750	0.9534	2.28	0.0225*	2.1467	0.9903	2.17	0.0302*	2.1160	0.9419	2.25	0.0247*
$P_2 \times L_2$	0.8679	1.2238	0.71	0.4782	0.8025	1.2304	0.65	0.5142	0.8146	1.2234	0.67	0.5055

Note: " * " denotes statistically significance at $\alpha = 0.05$

$$\begin{aligned}
\text{logit}(\pi_{ij}) \quad = \quad & \beta_0 + \beta_1 O + \beta_2 \text{Type}_1 + \beta_3 \text{Type}_2 + \beta_4 P_1 + \beta_5 P_2 + \beta_6 \text{Type}_1 \times P_1 \\
& + \beta_7 \text{Type}_1 \times P_2 + \beta_8 \text{Type}_2 \times P_1 + \beta_9 \text{Type}_2 \times P_2 + \beta_{10} L_1 \\
& + \beta_{11} L_2 + \beta_{12} P_1 \times L_1 + \beta_{13} P_1 \times L_2 + \beta_{14} P_2 \times L_1 + \beta_{15} P_2 \times L_2. \quad (37)
\end{aligned}$$

The parameters estimates, their standard error estimates, z-value, and p-value in model (37) are showed in Table 4. It showed that the two-factor interaction effect of ad type and position (Type $\times$ P), the interaction effect of position and animation length ($P \times L$), as well as the order effect ("O") were statistically significant at $\alpha = 0.05$.

Furthermore, we calculated the QIC for the fitted model (37) with three different working correlation structures. The results of QIC values based on three working correlation were showed in Table 5. GEE model with AR(1) working correlation has the smallest value of QIC. According to Pan's (2001) suggestion, model (38) with AR(1) working correlation structure was selected to fit CTR data well.

Table 5: The QIC measures of fitted equation (37) for CTR data.

Working correlation	**DF**	**QIC**	**Quasilikelihood**	α_1	α_2
Autoregressive	794	1001.205	-448.1486	0.3523	0.1241
m-dependent ($m = 2$)	794	1023.534	-455.9307	0.3618	0.3912
Independent	794	1001.214	-447.5849	0	0

The fitted equation is

$$\begin{aligned}
\text{logit}(\hat{\pi}_{ij}) &= -1.3902 + 0.1189O + 0.0395\text{Type}_1 - 0.7630\text{Type}_2 \\
&\quad -0.2961P_1 - 0.9738P_2 - 0.2811\text{Type}_1 \times P_1 \\
&\quad -0.1494\text{Type}_1 \times P_2 + 0.2352\text{Type}_2 \times P_1 + 1.0365\text{Type}_2 \times P_2 \\
&\quad -1.1959L_1 - 0.1779L_2 + 0.5356P_1 \times L_1 + 0.6246P_1 \times L_2 \\
&\quad +2.116P_2 \times L_1 + 0.8146P_2 \times L_2. \qquad (38)
\end{aligned}$$

We also tried out some other possible candidates for the "working correlation", for example , m-dependent with $m = 1$ or 2; the results of model selection were almost unchanged. One possible reason for this is that the estimate of the covariance matrix of the logistic regression coefficients is robust. That is, the GEEs method has the property of being a consistent estimator of the covariance matrix of the estimators of logistic regression coefficients even if the working correlation matrix is mispecified (Zeger and Liang, 1986).

For identifying the systematic departure, we checked the scatter plot of residuals against the fitted values in Figure 4. The scatter plot of residuals looks like to be lopsided across the fitted values but the spread of the residuals was not too large. The reason we guess was partly due to the highly negative correlations between the residuals and the fitted values for the correlated binary data. The scatter plot of residuals against follow-up time orders (denoted by "O") by the model (38), in Figure 5, also showed that

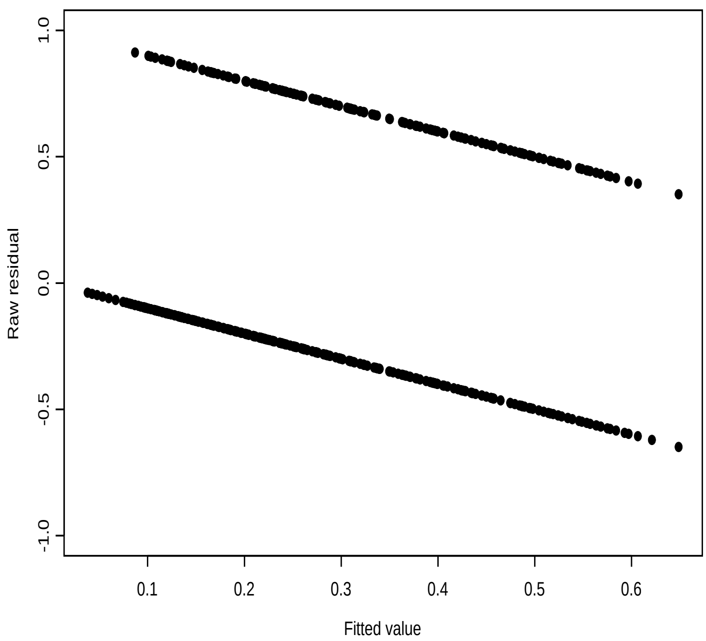

Figure 4: The plot of the raw residuals against the fitted value, $\hat{Y}$'s (predicted values by the model (38)).

the order effect on the CTR data is important. We tried to add other curvilinear effects into the model, there was no further improvement in the diagnostic scatter plots. Both quantitative results (the p-value in Table 4), QIC in Table 5, and graphical diagnosis (Figures 4 and 5) showed that the current model (38) fitted the data acceptably.

Moreover, using the fitted model (38) to calculate the odds ratio (OR) and to estimate the click-through-ratio for all treatment combinations. We transformed the equation (38) to the estimated odds of click-through as follows.

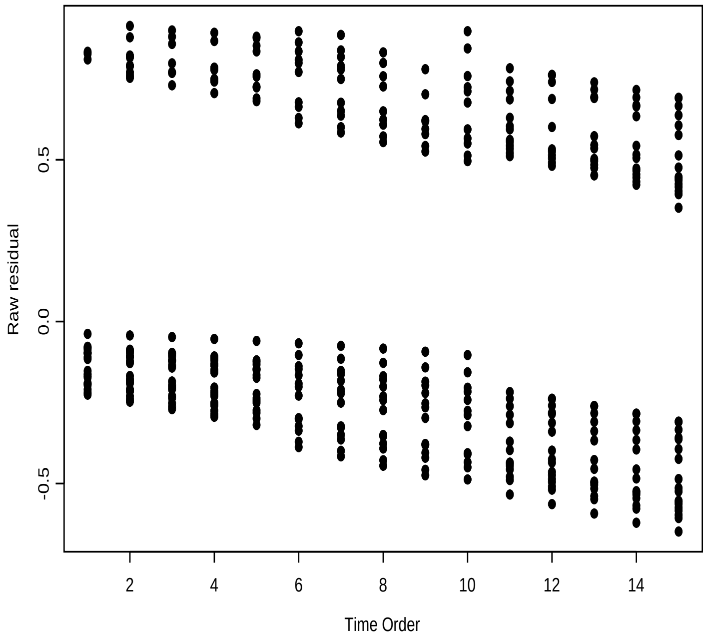

Figure 5: The plot of the raw residuals against time order.

$$\frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}} = \exp\left(\begin{array}{l} -1.3902 + 0.1189O + 0.0395\text{Type}_1 - 0.7630\text{Type}_2 \\ -0.2961P_1 - 0.9738P_2 - 0.2811\text{Type}_1 \times P_1 \\ -0.1494\text{Type}_1 \times P_2 + 0.2352\text{Type}_2 \times P_1 \\ +1.0365\text{Type}_2 \times P_2 - 1.1959L_1 - 0.1779L_2 \\ +0.5356P_1 \times L_1 + 0.6246P_1 \times L_2 + 2.116P_2 \times L_1 \\ +0.8146P_2 \times L_2. \end{array}\right) \quad (39)$$

According to model (39), the estimated odds ratios for all treatment combinations were calculated and shown in Tables 6, 7, 8 and 9 respectively. First, we considered the odds ratios between the interaction effect of ad type and ad position (Type$\times L$). Under the condition of first ad and animation length of 7.5 seconds, we compared the odds of two different ad position with the same ad type.

For instance, under the first ad and animation length of 7.5 seconds, $O = 1$, $L_1 =$

1, $L_2 = 0$, given ad is promotion-type, $T_1 = 0$, $T_2 = 1$, given ad position is middle, $P_1 = 0$, $P_2 = 1$, substitute the preceding values into equation (38), $\text{logit}(\hat{\pi}) = -1.0515$; by contrast, for ad position is low, $\text{logit}(\hat{\pi}) = -3.2302$. Changing ad position from low to middle, a multiplicative effect of $\exp(2.1787) = 8.8348$ on the odds that subject would click through this ad ($Y = 1$). At the same condition except ad position, the odds equals 0.0636 for an high position, 0.3494 for a middle position, for which odds ratio is 0.1820 or $1/0.1820 = 5.49$ for the middle against high positions. As for the odds of high against low positions we have odds equal 0.063577 for an high position, and 0.03955 for a low position, for which odds ratio is 1.6075. In other words, the odds of click through ad of middle position is slightly greater than low and high ones, given the ad type is sale promotion. Similarly, we obtained other odds ratios under different conditions of ad types (Table 6). When the ad type is commodities-type, the odds of click through ad of middle position is greater than low and high ones. Given the ad is entertainment-type , the result is also similar to the ones of commodities-type ads.

In addition, we also compared the odds of two different ad types under the same ad position (Table 7). Given the positions of ad are high and middle, the odds of click through ad of any two kinds of ad types are not significantly different. However, given the position of ad is low, the odds of click through ads of sale promotion type is smaller than commodities-type and entertainment-type ads.

Meanwhile, we also considered the odds ratios between the interaction effect of position and animation length ($P \times L$) under the condition of first ad and commodities-type ad (Table 8). Given the position of ad is high, the odds of click through ad of animation length of 15 seconds is slightly greater than 7.5 seconds. Given the position is middle, the odds of click through ad animation length of 7.5 seconds is slightly greater than 30

Table 6: The estimated odds ratio among ad type and ad position under the condition of first ad and animation length of 7.5 seconds.

Compare		Base		
Type	Position	Type	Position	Odds Ratio
Commodities	High	Commodities	Low	0.9593
Commodities	Middle	Commodities	Low	2.6988
Commodities	High	Commodities	Middle	0.3554
Promotion	High	Promotion	Low	1.6075
Promotion	Middle	Promotion	Low	8.8348
Promotion	High	Promotion	Middle	0.1820
Entertainment	High	Entertainment	Low	1.2706
Entertainment	Middle	Entertainment	Low	3.1337
Entertainment	High	Entertainment	Middle	0.4055

Table 7: The estimated odds ratio among ad type and ad position under first ad and animation length of 7.5 seconds.

Compare		Base		
Position	Type	Position	Type	Odds Ratio
High	Commodities	High	Entertainment	0.7854
High	Promotion	High	Entertainment	0.5899
High	Commodities	High	Promotion	1.3314
Middle	Commodities	Middle	Entertainment	0.8959
Middle	Promotion	Middle	Entertainment	1.3146
Middle	Commodities	Middle	Promotion	0.6815
Low	Commodities	Low	Entertainment	1.0403
Low	Promotion	Low	Entertainment	0.4663
Low	Commodities	Low	Promotion	2.2311

Table 8: The estimated odds ratio among position and animation length under fixed first ad and commodities-type ad.

Compare		Base		
Position	Animation Length	Position	Animation Length	Odds Ratio
High	7.5	High	30	0.5167
High	15	High	30	1.5631
High	7.5	High	15	0.3305
Middle	7.5	Middle	30	2.5095
Middle	15	Middle	30	1.8902
Middle	7.5	Middle	15	1.3276
Low	7.5	Low	30	0.3024
Low	15	Low	30	0.8370
Low	7.5	Low	15	0.3613

seconds. Given the position is low, the odds of click through ad animation length of 15 and 30 seconds are both greater than 7.5 seconds.

Next, we also considered the odds ratios of two different ad positions that are conditional on the first ad, commodities-type ad, and the animation length (Table 9). Given animation length is 7.5 seconds, the odds of click through ad of middle position is greater than high and low positions. It's partly similar to the results in Table 6 (for commodities-type ads). Given animation length is 15 seconds, the odds ratios among any two kinds of positions are not significantly different. Given animation length is 30 seconds, the odds of click through ad in low position is greater than middle one.

After discussing the odds ratios between the interaction effects of Type$\times P$ and $P \times L$, we would calculate the estimated click-through-ratio (ECTR) as follows. Using the equation (37) with AR(1) structure in GEEs, the estimated click through rate would be

Table 9: The estimated odds ratio among position and animation length under fixed first ad and commodities-type ad.

Compare		Base		
Animation Length	Position	Animation Length	Position	Odds Ratio
7.5	High	7.5	Low	0.9593
7.5	Middle	7.5	Low	2.6988
7.5	High	7.5	Middle	0.3554
15	High	15	Low	1.0485
15	Middle	15	Low	0.7345
15	High	15	Middle	1.4276
30	High	30	Low	0.5615
30	Middle	30	Low	0.3252
30	High	30	Middle	1.7263

calculated.

$$\hat{\pi}_{it} = \frac{\exp\begin{pmatrix} -1.3902 + 0.1189O + 0.0395\text{Type}_1 - 0.7630\text{Type}_2 \\ -0.2961P_1 - 0.9738P_2 - 0.2811\text{Type}_1 \times P_1 \\ -0.1494\text{Type}_1 \times P_2 + 0.2352\text{Type}_2 \times P_1 \\ +1.0365\text{Type}_2 \times P_2 - 1.1959L_1 - 0.1779L_2 \\ +0.5356P_1 \times L_1 + 0.6246P_1 \times L_2 + 2.116P_2 \times L_1 \\ +0.8146P_2 \times L_2 \end{pmatrix}}{1 + \exp\begin{pmatrix} -1.3902 + 0.1189O + 0.0395\text{Type}_1 - 0.7630\text{Type}_2 \\ -0.2961P_1 - 0.9738P_2 - 0.2811\text{Type}_1 \times P_1 \\ -0.1494\text{Type}_1 \times P_2 + 0.2352\text{Type}_2 \times P_1 \\ +1.0365\text{Type}_2 \times P_2 - 1.1959L_1 - 0.1779L_2 \\ +0.5356P_1 \times L_1 + 0.6246P_1 \times L_2 + 2.116P_2 \times L_1 \\ +0.8146P_2 \times L_2 \end{pmatrix}}. \quad (40)$$

ECTR and 95% confidence interval of CTR among ad position and ad type that given conditions on first ad and the 3 different animation lengths (see Tables 10-12). ECTR in Table 10, for example, under first ad and animation length fixed at 7.5 seconds, $O = 1,\ L_1 = 1,\ L_2 = 0$, given middle position and promotion-type ad , $P_1 = 0,\ P_2 = 1,\ T_1 =$

Table 10: ECTR and 95% confidence interval of CTR among the nine combinations of ad positions and types under animation length fixed at 7.5 seconds.

Animation Length (7.5 seconds)			95% C.I.	
Ad Position	Ad Type	ECTR	Lower	Upper
High	Commodities	0.0780	0.0208	0.2520
High	Promotion	0.0598	0.0139	0.2234
High	Entertainment	0.0973	0.0261	0.3028
Middle	Commodities	0.1923	0.0733	0.4175
Middle	Promotion	0.2589	0.1067	0.5056
Middle	Entertainment	0.2100	0.0870	0.4257
Low	Commodities	0.0811	0.0395	0.1591
Low	Promotion	0.0380	0.0169	0.0834
Low	Entertainment	0.0782	0.0327	0.1755

0, $T_2 = 1$, and $\text{logit}(\hat{\pi}) = -1.0515$. Using equation (40), we have $\exp(-1.0515)/[1 + \exp(-1.0515)]$, this is, ECTR equal 0.2589. From Table 10, we drew ECTR plot as Figure 6. We could find that the greatest ECTR (0.2589) is the combination of middle position and promotion-type ad, and the smallest ECTR (0.0380) is the combination of low position and promotion-type ad, under animation length fixed at 7.5 seconds and the first ad ($O = 1$). Similarly, from Table 11, we could draw ECTR plot in Figure 7. The greatest ECTR (0.2459) is the combination of high position and entertainment-type ad under animation length fixed at 15 seconds and first ad ($O = 1$). From Table 12, Figure 8 showed the plot of ECTR. From the plot, we found that the greatest ECTR (0.2259) is the combination of low position and commodities type ad and the smallest ECTR (0.0867) is the combination of middle position and commodities type ad under animation length.

In addition, we also calculated ECTR and 95% confidence interval of CTR among ad position and animation length that given conditions on first ad and 3 different ad

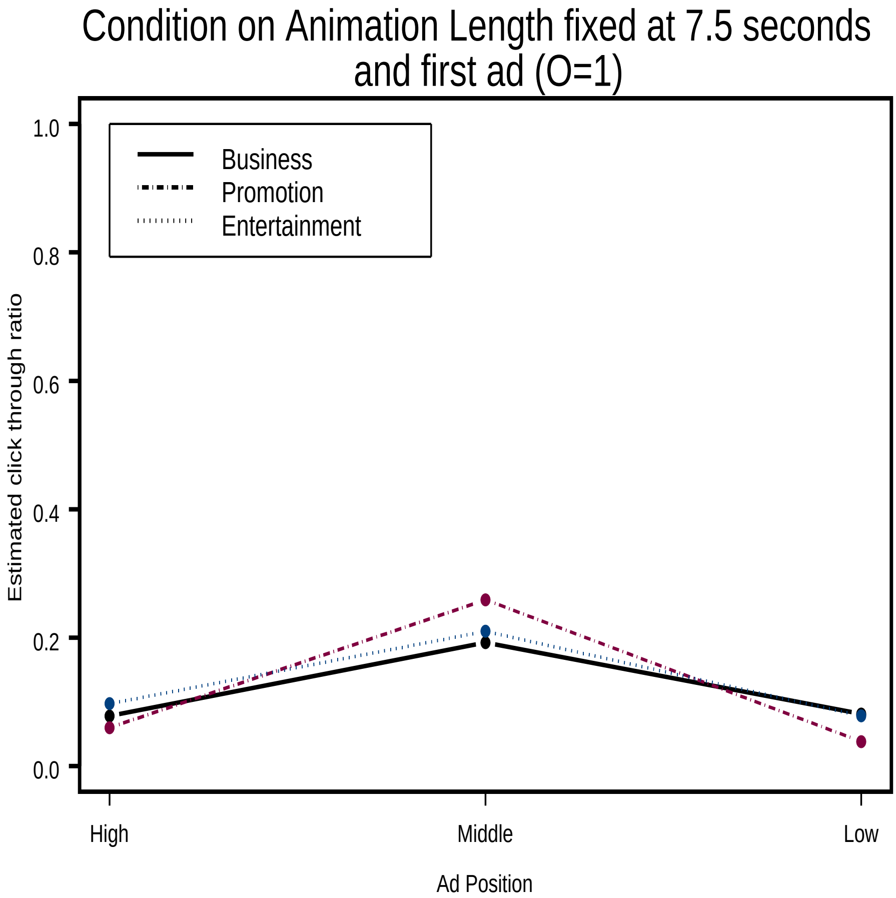

Figure 6: The plot of ECTR at ad positions and ad types under fixed animation length fixed at 7.5 seconds.

Table 11: ECTR and 95% confidence interval of CTR among the nine combinations of ad positions and types under animation length fixed at 15 secons.

Animation Length (15 seconds)			95% C.I.	
Ad Position	Ad Type	ECTR	Lower	Upper
High	Commodities	0.2039	0.1040	0.3609
High	Promotion	0.1613	0.0818	0.2934
High	Entertainment	0.2459	0.1292	0.4173
Middle	Commodities	0.1521	0.0438	0.4128
Middle	Promotion	0.2084	0.0637	0.5046
Middle	Entertainment	0.1668	0.0427	0.4732
Low	Commodities	0.1963	0.0514	0.5241
Low	Promotion	0.0987	0.0194	0.3772
Low	Entertainment	0.1901	0.0424	0.5544

Table 12: ECTR and 95% confidence interval of CTR among the nine combinations of ad positions and types under animation length fixed at 30 seconds.

Animation Length(30 seconds)			95% C.I.	
Ad Position	Ad Type	ECTR	Lower	Upper
High	Commodities	0.1408	0.0530	0.3243
High	Promotion	0.1096	0.0401	0.2659
High	Entertainment	0.1726	0.0716	0.3606
Middle	Commodities	0.0867	0.0368	0.1908
Middle	Promotion	0.1222	0.0523	0.2599
Middle	Entertainment	0.0958	0.0420	0.2038
Low	Commodities	0.2259	0.0765	0.5067
Low	Promotion	0.1156	0.0374	0.3056
Low	Entertainment	0.2190	0.0739	0.4963

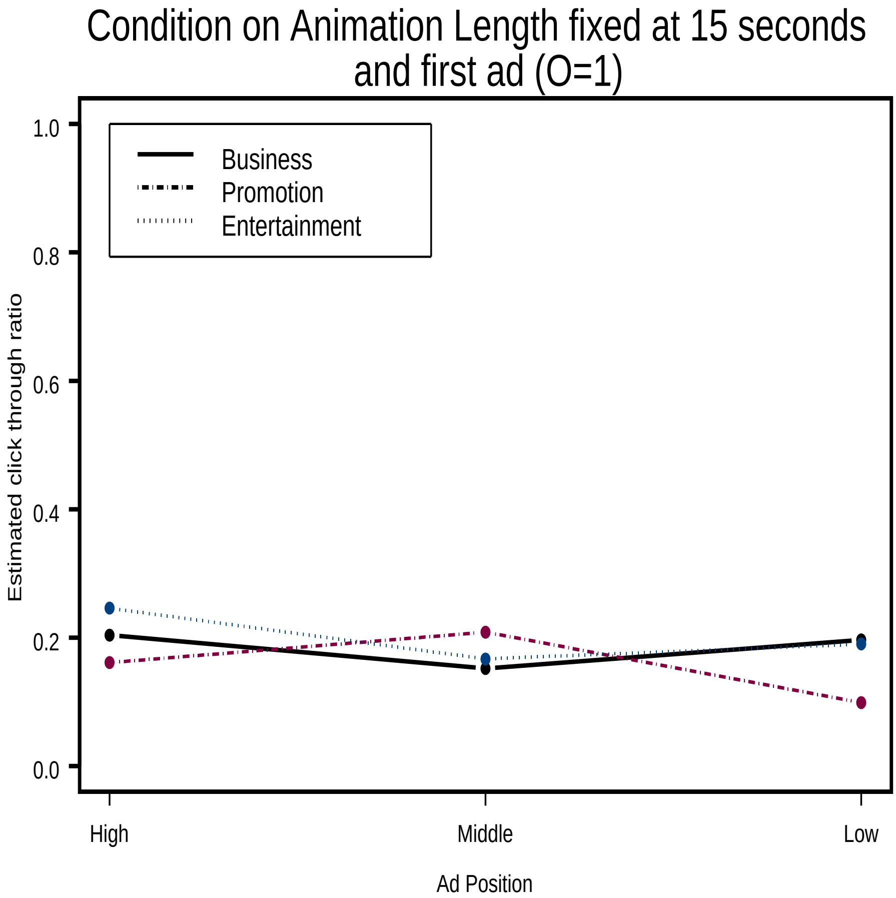

Figure 7: The plot of ECTR at ad position and ad type under animation length fixed at 15 seconds.

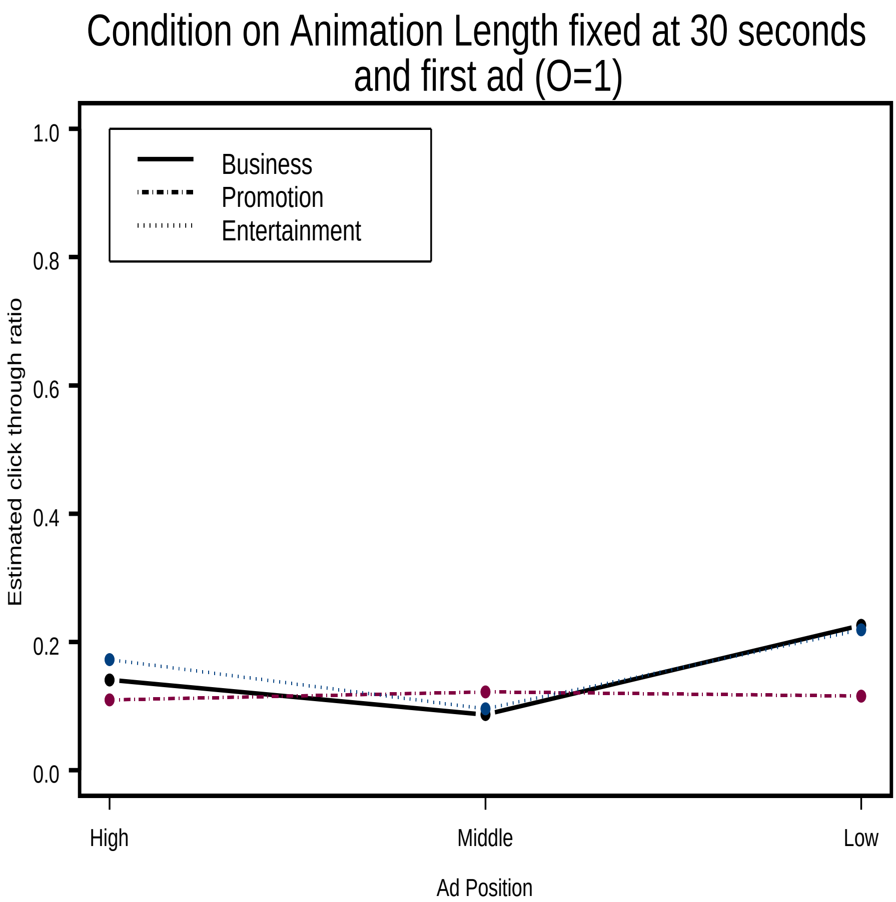

Figure 8: The plot of ECTR at ad position and ad type under animation length fixed at 30 seconds.

Table 13: ECTR and 95% confidence interval of CTR among the nine combinations of ad positions and animation lengths under commodities-type advertising.

Commodities-type advertising			95% C.I.	
Ad Position	Animation length	ECTR	Lower	Upper
High	7.5	0.0780	0.0208	0.2520
High	15	0.2039	0.1040	0.3609
High	30	0.1408	0.0530	0.3243
Middle	7.5	0.1923	0.0733	0.4175
Middle	15	0.1521	0.0438	0.4128
Middle	30	0.0867	0.0368	0.1908
Low	7.5	0.0811	0.0395	0.1591
Low	15	0.1963	0.0514	0.5241
Low	30	0.2259	0.0765	0.5067

types (Tables 13-15). From Table 13, given ad type is commodities-type advertising, the greatest ECTR (0.2259) is the combination of low position and animation length held at 30 seconds and the smallest ECTR (0.0780) is the combination of high position and animation length held at 7.5 seconds. The ECTR plot is shown in Figure 9. From Table 14, given the condition of promotion-type ad, the ECTR (0.2589) of middle position and animation length of 7.5 seconds is the greatest, and low position and animation length of 7.5 seconds is the smallest (0.0380). The ECTR plot is shown in Figure 10. In addition the greatest ECTR is the treatment combination of high position and animation length of 15 seconds; the smallest ECTR is low position and animation length of 7.5 seconds under fixed entertainment type ad.

In summary, the best and worst combinations of ECTR would give the ad designers and advertisers an more objective guideline to determine the golden layout for different ad types.

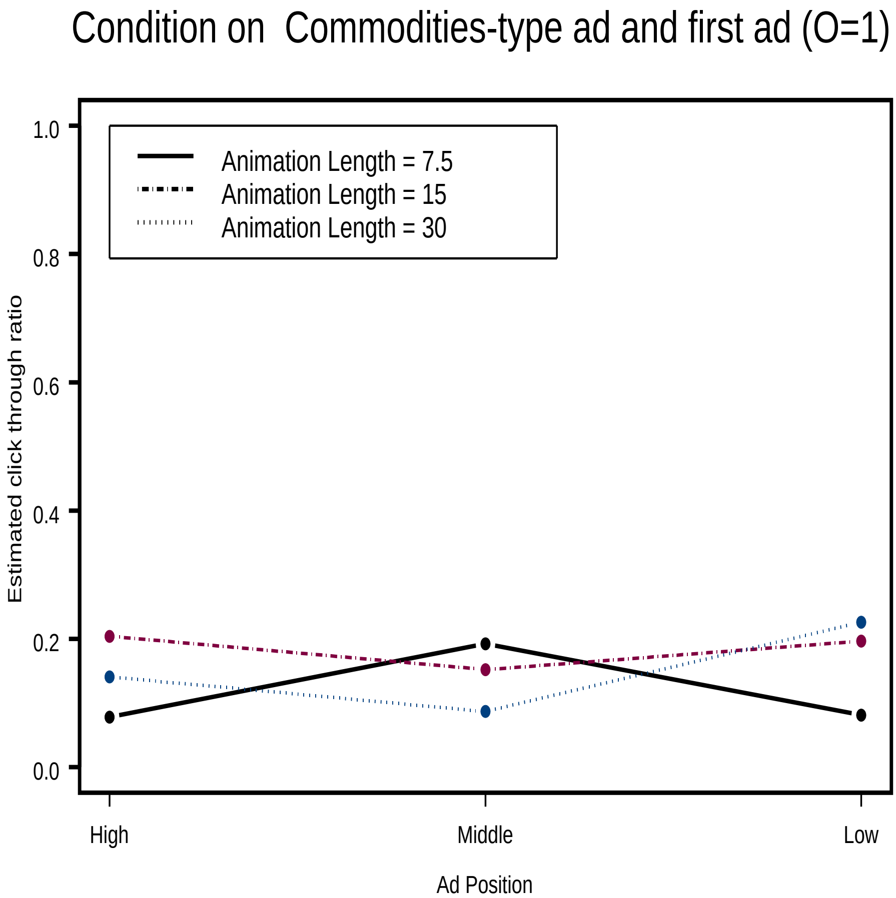

Figure 9: The plot of ECTR at ad position and ad type under fixed commodities-type advertising.

Table 14: ECTR and 95% confidence interval of CTR among the nine combinations of ad positions and animation lengths under fixed promotion type advertising.

Promotion-type Advertising			95% C.I.	
Ad Position	Animation length	ECTR	Lower	Upper
High	7.5	0.0598	0.0139	0.2234
High	15	0.1613	0.0818	0.2934
High	30	0.1096	0.0401	0.2659
Middle	7.5	0.2589	0.1067	0.5056
Middle	15	0.2084	0.0637	0.5046
Middle	30	0.1222	0.0523	0.2599
Low	7.5	0.0380	0.0169	0.0834
Low	15	0.0987	0.0194	0.3772
Low	30	0.1156	0.0374	0.3056

Table 15: ECTR and 95% confidence interval of CTR among the nine combinations of ad positions and animation lengths under entertainment type advertising.

Entertainment-type Advertising			95% C.I.	
Ad Position	Animation length	ECTR	Lower	Upper
High	7.5	0.0973	0.0261	0.3028
High	15	0.2459	0.1292	0.4173
High	30	0.1726	0.0716	0.3606
Middle	7.5	0.2100	0.0870	0.4257
Middle	15	0.1668	0.0427	0.4732
Middle	30	0.0958	0.0420	0.2038
Low	7.5	0.0782	0.0327	0.1755
Low	15	0.1901	0.0424	0.5544
Low	30	0.2190	0.0739	0.4963

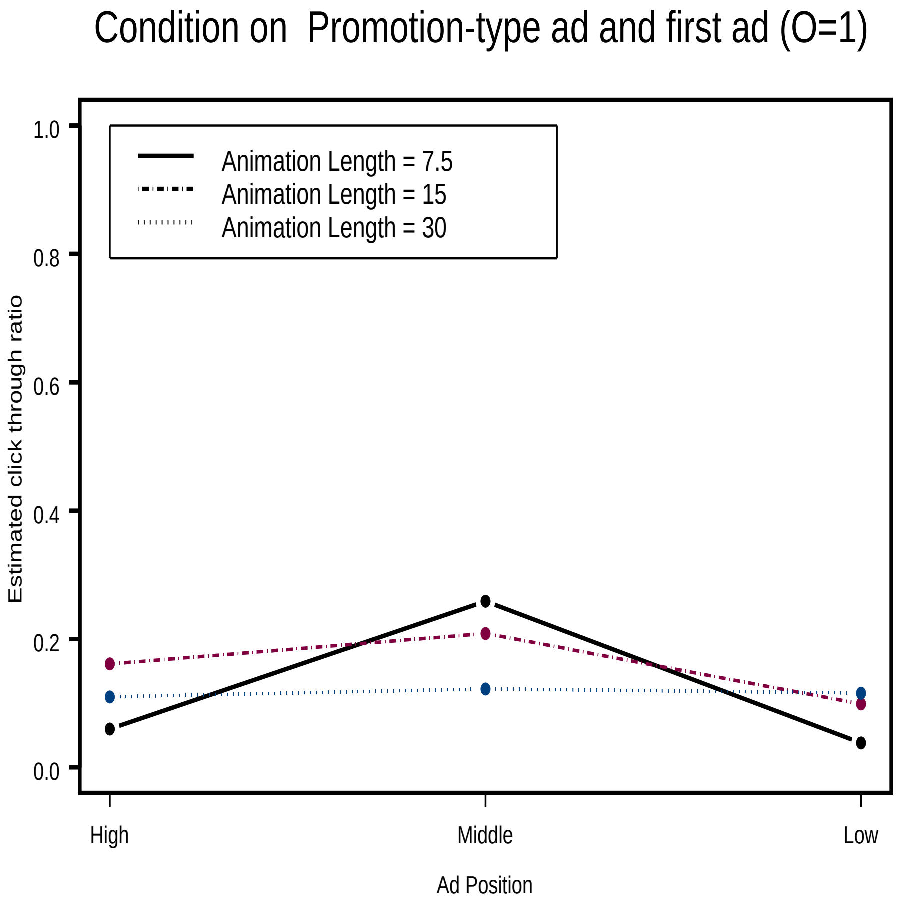

Figure 10: The plot of ECTR at ad position and ad type under fixed promotion-type advertising.

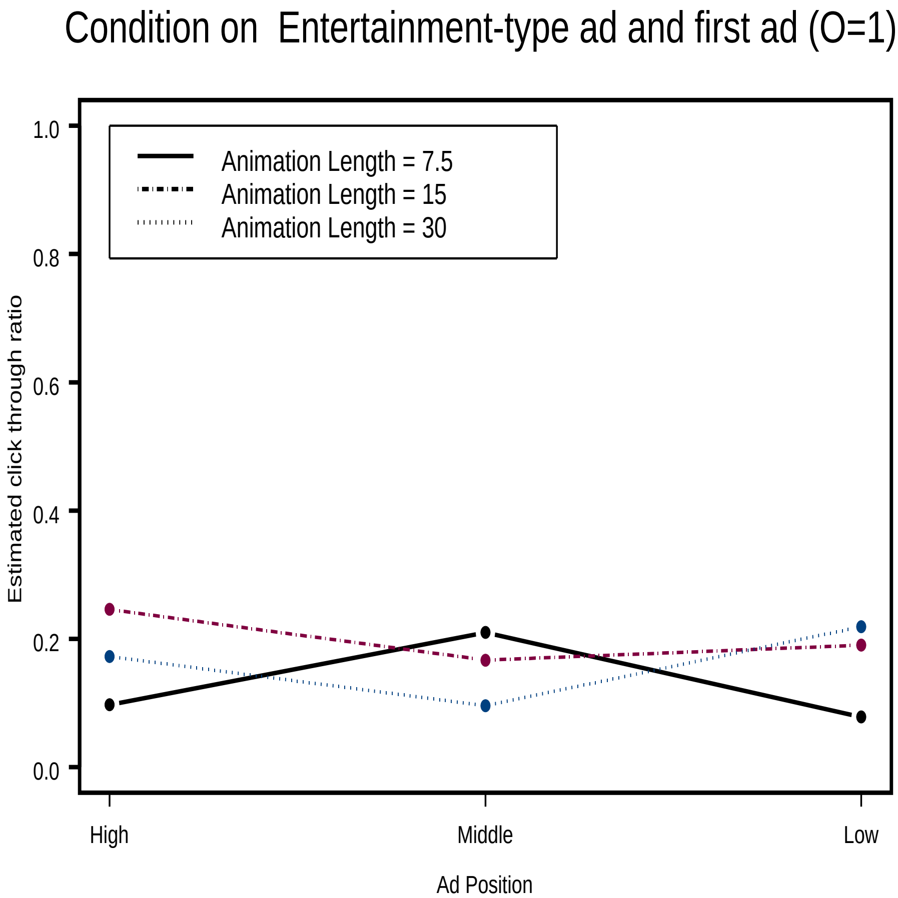

Figure 11: The plot of ECTR at ad position and ad type under fixed entertainment-type advertising.

5 CONCLUSIONS

In this study, the multiple logistic regression models with GEEs approach under the consideration of correlation structures were used to fit a CTR data. The empirical results have shown that the GEEs model is more appropriate than considering the outcomes as independent. We summarized the main results as follows.

1. To assess the relationship between responses of click-through of consumers and design factors of medium rectangle advertising, a multiple logistic regression model using GEEs methods with AR(1) correlation structure were verified to fit the data acceptably.

2. Both quantitative results, z-test based on the p-value and goodness-of-fit test statistic QIC, and graphical diagnosis of residual plots showed that the fitted equation (38) fits adequately. The results indicated that time order, two-factor interaction of ad type and ad position, as well as the interaction of ad position and animation length are statistically significant.

3. Given promotion-type ad, the combination of middle position and animation length of 7.5 seconds could provide the highest ECTR. (Hofacker & Murphy, 1988; Cho et al, 2001) proposed that the pricing of internet advertising is often based on CTR. The results would provide to ad designers and advertisers an to determine the golden layout or the pricing of internet advertising.

4. However, given commodities-type ad, the combination of position and animation length could be different from the ones of promotion-type ad. We've found the combination of lower position and animation length of 30 seconds would provide

higher ECTR (0.2259) than others. In addition, we've also found the combination of upper position and animation length of 15 seconds would result in higher ECTR (0.2459) than other combinations given that entertainment-type ad.

References

[1] Agresti, A. (1996), *An Introduction to Categorical Data Analysis*, John Wiley & Sons, INC.

[2] Agresti, A. (2002), *Categorical Data Analysis*, Second Edition, John Wiley & Sons, INC.

[3] Baltas, G. (2003), Determinants of internet advertising effectiveness: an empirical styd, *International Journal of Market Research* **45**.

[4] Chang, Y.-C. (2000), Residuals analysis of the generalized linear models for longitudinal data, *Statistics in Medicine* **19**, pp.1277-1293.

[5] Cho, C. H., Lee, J. G. & Taharp, M. (2001), Different force-exposure levels to banner advertisement, *Journal of Advertising Research* **41**, 1, pp.35-42.

[6] Doyle, K., Minor, A. & Weyrich, C. (1997), Banner ad placement study . Avilable at URL:http://webreference.com./dev/banners/onethird.html.

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

[8] Hanson, W. (2000), *Principles of Internet Marketing*, Ohio: South-Western College Publishing.

[9] Hardin, J. W. and Hilbe, J. M. (2003), Review of generalized estimating equations by Hardin and Hilbe, *The Stata Journal* **3**, pp.208-210.

[10] Hardin, J. W. and Hilbe, J. M. (2003), *Generalized Estimating Equations*, CHAPMAN & HALL/CRC.

[11] Hofacker, C. F. & Murphy, J. (1988), World wide web banner advertisement copy testing, *European Journal of Marketing* **32**, 7/8, pp.703-712.

[12] Hosmer, D. W. and Lemeshow, S. (2000), *Applied Logistic Regression*, Second Edition, John Wiley & Sons, INC.

[13] Interactive Advertising Bureau (2002), http://www.iab.net/.

[14] Johnston G. Repeated measures analysis with discrete data using the SAS system, SAS institute Inc., Cary, NC

[15] Liang, K. Y. & Zeger, S. L. (1986), Longitudinal data analysis using generalized linear models, *Biometrika* **11**, pp.59-67.

[16] Lin, Ya-Li and Lin, Kai-Chiang (2003), Influence Study of Key Factors of Dynamic Ad Design on the View Clicking Rate of Web Advertising, *Conference on Chinese Instituteof Industrial Engineers*, Taiwan.

[17] Mccullagh, P., and Nelder, J. A. (1983), Quasi-likelihood functions, *Annals of Statistics* **73(1)**, pp.13-22.

[18] Mccullagh, P., and Nelder, J. A. (1989), *Generalized Linear Models*, Second Edition, Chapman & Hall, London.

[19] Ochi Y. and Prentice R. L. (1984), Likelihood inference in a correlated probit regression-model, *Biometrika* **71**, pp.531-543.

[20] Pan, W. (2001), Akaike's information criterion in generalized estimating equations, *Biometrics* **57**, pp.120-125.

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

[22] Chang, Y. C. (2000), Residuals analysis of the generalized linear models for longitudinal data, *Statistics in medicine* **19**, pp.1277-1293.

[23] Zeger, S. L., Liang, K. Y., and Self, S. G. (1985), The analysis of binary longitudinal data with time-independent covariates, *Biometrika* **72(1)**, pp.31-38.

[24] Zeger, S. L. & Liang, K. Y. (1986), Longitudinal data analysis for discrete and continuous outcomes, *Biometrics* **42**, pp.121-130.