Parameter estimation for binary time series using partial likelihood

I Nadiya, Y Widyaningsih, and D Sarwinda
Department of Mathematics, Faculty of Mathematics and Natural Sciences (FMIPA), Universitas Indonesia, Depok 16424, Indonesia
Corresponding author’s email: yekti@sci.ui.ac.id

Abstract. A time series with binary response variable is called a binary time series. Binary time series can be modelled using the Autoregressive general model and nonlinear regression approach. Kedem & Fokianos introduced a binary time series model through the Autoregressive and logistic regression approach. The parameters of binary time series are estimated using the Partial Likelihood method. The Partial Likelihood method is performed by determining the Partial Likelihood function derived from the marginal probability density function (pdf) of Bernoulli distribution. However, in the process of parameter estimation using this method, the form of final function to obtain parameters is not in the closed form equation. To face this problem, Fisher scoring iterations are performed. The application of parameter estimation of the model uses the data about boat racing competition between the University of Cambridge and Oxford University from 1946 to 2011. Based on the data application, parameter estimation of the binary time series model using partial likelihood with different amounts of data resulting in a relatively same or no significant parameter estimator.

Keywords: Binary time series, partial likelihood

1. Introduction
Time series is the set of observations of an event taken from time to time and recorded sequentially on time. The purpose of time series analysis is to understand and model the stochastic mechanism of a time series and predict future time values based on the value of the previous time series [1]. A time series that implements the Generalized Linear Model (GLM) framework as a model of response variables. GLM is a generalization of linear regression with response variables which are distributed from exponential family [2]. GLM is a linear regression generalization by connecting response variables and predictor variables through a link function. Special case of GLM that has a response variable with binary number using logistic regression. Binary logistic regression is a special case of linear regression generalized using GLM, where the dependent variable (response variable) has only two dichotomous (binary) values, 0 and 1, while the independent variables can be either binary or scaled intervals or ratios. Logistic regression model or commonly called logit model has a link function which is the inverse of cumulative distribution function (cdf). Time series that has observations with numbers 0 and 1 is called a binary time series. A binary time series is a time series that has a response variable of binary or Bernoulli-distributed types, in which Bernoulli distribution is included in the exponential family distribution. Therefore, it can be concluded that the time series model which has binary response variable or binary time series can use logistic regression model or
can be called binary logistic regression model which is a special type of Generalized Linear Model (GLM). Furthermore, parameter estimation is required for the binary time series model. Cox et al. introduced the parameter estimation method of Partial Likelihood method [3]. The Partial Likelihood method is also one of the methods in parameter estimation. The advantage of the Partial Likelihood method is that parameter estimators can be obtained using marginal pdf and the estimation of Partial Likelihood is more stable in numerically calculated. This paper is presented in the following sections. Section 2 deals with Autoregressive. Section 3 presents Generalized Linear Model. Section 4 is about logistic regression and continued to section 5 about binary time series. Section 6 presents the result of model’s parameters estimation, section 7 is a case study and section 8 is a summary.

2. Autoregressive

The time series is part of the stochastic process. According to [4], a stochastic process is a set of indexed random variables \( \{ Y_t; t \in T \} \) that meets the laws of opportunity. Time series follows the law of opportunity, where each value in the data depends on time but may change indefinitely (in uncertainty). Each stochastic process contains state space (S) and parameter space (T). The state space (S) is the set of all possible values of \( Y_t \) and the parameter space (T) is the set of indices of \( Y_t \). S and T can be discrete or continuous. Let \( y_t \) value of \( Y_t \) then the sequence value \( \{ y_t; t \in T \} \) is called the realization of \( \{ Y_t; t \in T \} \). Thus, a stochastic process \( Y_t \) is said to be a time series of \( Y_t \) if T is a set of time points with a parameter space (T) of discrete value and sequentially recorded in order of occurrence periodically. One of time series model is the Autoregressive model. The Autoregressive model is a model that illustrates that the dependent variable is influenced by the dependent variable itself in the previous period. The general form of the Autoregressive model (AR) is as follows:

\[
Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \epsilon_t
\]

3. Generalized Linear Model (GLM)

Generalized linear models (GLM) originally appeared as a generalization of classical linear regression models [5]. The differences from the classical linear regression model are as follows: the response variable distribution belongs to the exponential family of distributions (it is not necessarily normal), while a link function gives the relationship between the dependent variable mean and a linear form built on the explanatory variables. The exponential family of distributions includes discrete distributions such as Bernoulli, Poisson, Binomial, Negative Binomial, and also continuous distributions such as Normal, Inverse-Gaussian, or Gamma.

There are three components of forming in GLM [6], namely:

1. The model’s random component is the dependent random variables \( Y_1, Y_2, \ldots, Y_n \), which are not identically distributed, although their distributions have the same form in the exponential family. In addition, the mean of dependent variables is denoted by \( \mu_i \).
2. The model’s systematic component or the “linear predictor” uses \( p + 1 \) parameters \( \beta = (\beta_0, \beta_1, \ldots, \beta_p)^T \) and \( p \) explanatory variables \( \eta_i = \beta + \sum_{k=1}^{p} x_{ik} \beta_k, i = 1, 2, \ldots, n \).
3. The “link function” between the random and systematic components, namely a monotonic and differentiable function \( g \), such that \( \eta_i = g(\mu_i) = \beta + \sum_{k=1}^{p} x_{ik} \beta_k, \quad i = 1, 2, \ldots, n \).

This research used Logistic regression model. The Logistic regression model is written as follows:

\[
Y_i = \mu_i + \epsilon_i, \quad i = 1, 2, \ldots, n
\]
with \( Y_i \) is dependent variable that state the number events in the form of binary data based on Bernoulli’s distribution, \( \mu_i \) is mean of dependent variable, and \( \varepsilon_i \) is error for the \( i \)-th observation.

Note that in the Bernoulli regression model, the conditional variable response of the predictor variable is assumed to be Bernoulli distributed. Since \( Y_i \) is Bernoulli distributed, mean \( \mu_i \) must exist at interval \((0, 1)\), while the value of \( \beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + \cdots + \beta_pX_{pi} \) lies at interval \((-\infty, \infty)\). Therefore, the link function \( g \) is needed to solve the problem, so that the values of the left and right sides of the equation at the same interval, ie \((-\infty, \infty)\). The appropriate link function \( g \) is a logarithm, so that it is obtained:

\[
g(\mu_i) = \ln\left( \frac{\mu_i}{1 - \mu_i} \right) = \beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + \cdots + \beta_pX_{pi}.
\]

\[
\mu_i = \exp(\beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + \cdots + \beta_pX_{pi}).
\]

The maximum likelihood method is used to estimate the parameters.

4. Logistic regression

Logistic regression is a modelling method in a condition where the dependent variable only has two values (dichotomy). The model established to analyze the relationship between response variable and explanatory variables can be written as follows: [7]

\[
y_i = x_i \mathbf{\beta} + \varepsilon_i
\]

\[
= [1 \ x_{i1} \ \cdots \ x_{pi}] \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \varepsilon_i
\]

By \( y_i \) is the response variable for the \( i \)-th observation, and \( y_i \) is 0 or 1. Assumed \( y_i \) is Bernoulli distributed, with the probability of \( y_i = 1 \) is denoted as \( \pi(x_i) \) and the probability of \( y_i = 0 \) is denoted as \( 1 - \pi(x_i) \), or it can be written as \( \Pr(y_i = 1) = \pi(x_i) \) and \( \Pr(y_i = 0) = 1 - \pi(x_i) \). The expectation of the error is assumed to be zero, so the expected value of \( y_i \) is:

\[
E(y_i) = y_i \Pr(y_i = 1) + y_i \Pr(y_i = 0) = 1(\pi(x_i)) + 0(1 - \pi(x_i)) = \pi(x_i)
\]

Based on the above equation, \( E(y_i) = \pi(x_i) = x^\mathbf{\beta} \). Equation 1 can therefore be written as follows:

\[
y_i = \pi(x_i) + \varepsilon_i
\]

In logistic regression, the probability of success denoted as \( \pi(x_i) = \Pr(Y_i = 1) \) has a value restriction that lies at intervals 0 and 1 or can be written as \( 0 \leq \pi(x_i) \leq 1 \). Therefore, to limit the value of \( \pi(x_i) \) to lie at interval \([0,1]\) a function that transforms \( \beta_0 + \beta_1X_{1i} \) from interval \([-\infty, \infty]\) to interval \([0,1]\).
A suitable function for this purpose is the logistic distribution function $F$ [8]. The logistics distribution function can be written as follows:

$$F(z) = \frac{1}{1 + e^{-z}}$$

Based on the logistics distribution function, then in transform $\beta_0 + \beta_1X_i$ to interval $[0,1]$ can be written equation as follows:

$$\pi(x_i) = F(\beta_0 + \beta_1X_i)$$

$$\pi(x_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1x_i)}}$$

It can also be written by the following equation:

$$\pi(x_i) = \frac{e^{(\beta_0 + \beta_1x_i)}}{1 + e^{(\beta_0 + \beta_1x_i)}}$$

From equation 2 obtained:

$$\frac{1}{\pi(x_i)} = 1 + e^{-(\beta_0 + \beta_1x_i)}$$

$$\frac{1}{\pi(x_i)} - 1 = e^{-(\beta_0 + \beta_1x_i)}$$

$$\frac{1 - \pi(x_i)}{\pi(x_i)} = e^{-(\beta_0 + \beta_1x_i)}$$

$$\frac{\pi(x_i)}{1 - \pi(x_i)} = e^{(\beta_0 + \beta_1x_i)}$$

$$\ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right) = \beta_0 + \beta_1x_i$$

Then obtained logit model as in equation 5 or can be written as follows:

$$\text{logit}(\pi(x_i)) = \ln\left(\frac{\pi(x_i)}{1 - \pi(x_i)}\right)$$

5. Binary time series

Based on the Autoregressive model, substituting $\beta'Y_t$, then the binary time series model constructed based on logistic regression model and Autoregressive model can be written as follows:
Then the binary time series model can be written as follows:

\[
\beta' Y_* = \log \left( \frac{e^{\beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i}}}{1 + e^{\beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i}} + \sum_{i=1}^{p} \beta_i Y_{t-i}} \right)
\]

6. Results

**Definition 1.** Let \( Y_1, Y_2, ..., Y_n \) be the random variables in the probability. The marginal pdf of \( Y_t \) known \( Y_{(t-1)} \) is \( f_t(y_t; \theta) \) where \( \theta \in \mathbb{R}^p \) is a parameters vector. Then the partial likelihood function is

\[
PL(\theta; y_1, y_2, ..., y_n) = \prod_{t=1}^{n} f_t(y_t; \theta)
\]

with \( PL(\theta; y_1, y_2, ..., y_n) \) is partial likelihood function and \( f_t(y_t; \theta) \) is marginal pdf of \( Y_t \).

The parameters that estimated in the binary time series model are \( \beta \)'s. In estimating the parameters, the most commonly used method is the maximum likelihood estimator (MLE), but in maximum likelihood estimator (MLE), the probability density function (pdf) is \( f(y_1, y_2, ..., y_n; \beta) \), where \( Y_1, Y_2, Y_3, ..., Y_n \) is a random variable and \( \beta \) is an unknown parameter vector. However, since the binary time series model has a random variable \( Y_1, Y_2, Y_3, ..., Y_n \) which is not independent then it is difficult to obtain the joint probability density function (pdf) \( Y_1, Y_2, Y_3, ..., Y_n \). Therefore, in estimating the \( \beta \) parameter we will use
the maximum partial likelihood method, where the maximum partial likelihood estimation does not work using a joint probability density function (pdf) of $Y_1, Y_2, \ldots, Y_n$, but using the marginal probability density function (pdf) $f_t(Y_t; \beta)$. The following partial likelihood function: [9]

$$l(\beta) = \prod_{t=1}^{n} [f_t(Y_t; \beta)]$$

$$= \prod_{t=1}^{n} [\pi(Y_t)]^{y_t} [1 - \pi(Y_t)]^{1 - y_t}$$

$$= \prod_{t=1}^{n} [F(\beta' Y_s)]^{y_t} [1 - F(\beta' Y_s)]^{1 - y_t}$$

where $l(\beta)$ is a partial likelihood function, $Y_t$ is a random variable at a time $t$, $f_t(Y_t; \beta)$ is the marginal probability density function (pdf), $\pi(Y_t)$ or $F(\beta' Y_s)$ is the probability of $Y_t = 1$ with known $Y_{t-i}; i = 1, 2, \ldots, p$. Then will be determined log-partial function likelihood with the formula as follows:

$$L(\beta) = \log l(\beta)$$

$$= \log \left( \prod_{t=1}^{n} [\pi(Y_t)]^{y_t} [1 - \pi(Y_t)]^{1 - y_t} \right)$$

$$= \sum_{t=1}^{n} y_t \log(\pi(Y_t)) + (1 - y_t) \log(1 - \pi(Y_t))$$

$$= \sum_{t=1}^{n} y_t \log(F(\beta' Y_s)) + (1 - y_t) \log(1 - F(\beta' Y_s))$$

(9)

assume equation 9 can be derived, then:

$$\nabla \log l(\beta) = 0,$$

with

$$\nabla = \left( \frac{\partial}{\partial \beta_0}, \frac{\partial}{\partial \beta_1}, \frac{\partial}{\partial \beta_2}, \ldots, \frac{\partial}{\partial \beta_p} \right)'$$

let $\nabla \log l(\beta) = S_N(\beta)$, then:

$$S_N(\beta) = \frac{\partial}{\partial \beta} \left( \sum_{t=1}^{n} y_t \log(F(\beta' Y_s)) + (1 - y_t) \log(1 - F(\beta' Y_s)) \right)$$
with \( \log F(x) = \frac{1}{F(x)} F'(x) \) and \( F'(x) = f(x) \), then:

\[
S_{N}(\beta) = \sum_{t=1}^{n} \left[ \frac{y_t}{F(\beta'Y_t)} f(\beta'Y_t)Y_* - (1 - y_t) \frac{1}{1 - F(\beta'Y_t)} f(\beta'Y_t)Y_* \right]
= \sum_{t=1}^{n} \left[ f(\beta'Y_t)Y_* \left( \frac{y_t}{F(\beta'Y_t)} - \frac{(1 - y_t)}{1 - F(\beta'Y_t)} \right) \right]
= \sum_{t=1}^{n} \left[ f(\beta'Y_t)Y_* \left( \frac{y_t(1 - F(\beta'Y_t)) - (1 - y_t)F(\beta'Y_t)}{F(\beta'Y_t)(1 - F(\beta'Y_t))} \right) \right]
= \sum_{t=1}^{n} \left[ f(\beta'Y_t)Y_* \left( \frac{y_t - y_tF(\beta'Y_t) - F(\beta'Y_t) + y_tF(\beta'Y_t)}{F(\beta'Y_t)(1 - F(\beta'Y_t))} \right) \right]
= \sum_{t=1}^{n} \left[ f(\beta'Y_t)Y_* \left( \frac{y_t - F(\beta'Y_t)}{F(\beta'Y_t)(1 - F(\beta'Y_t))} \right) \right]
= \sum_{t=1}^{n} Y_* \left( \frac{f(\beta'Y_t)}{F(\beta'Y_t)(1 - F(\beta'Y_t))} \right) (y_t - F(\beta'Y_t))
= \sum_{t=1}^{n} Y_* D(\beta'Y_*)(y_t - \pi(Y_t)),
\]

then next step is look for a solution \( S_{N}(\beta) = 0 \), but since the equation is not linear and not in the closed form, it is difficult to find the solution directly then the solution can be obtained by using Fisher Scoring iteration.

7. Case study
In this study, the parameter estimation method is applied to the data of boat racing between the University of Cambridge and the University of Oxford from 1946 to 2011. The sample in this study is the winner data from 1946 to 2011 which can be accessed at http://www.theboatrace.org. The time series data which is analyzed is the sample data set for 66 years, calculated from year 1946 until year 2011. The response variable in this study is the result of a boat racing competition between the University of Cambridge and the University of Oxford. The response variable contains binary data, where the data contains 0 or 1. The value is 0 if the University of Cambridge loses or the University of Oxford wins, and 1 if the University of Cambridge wins or the University of Oxford loses. The dependent variable which is also the response variable in this study is denoted by the vector of size \( n \times 1 \) where \( y \) is the indicator of the result of the boat race and \( t = 1 \) refers to the competition result of 1946 and \( t = n \) refers to the result of competition in 2011. The model is an autoregressive model with the lag equal to one, so equation (7) becomes: \( \text{logit}(\pi(Y_t)) = (\beta_0 + \beta_1Y_{t-1}) + \epsilon_t \). In the program simulation, performed 3 experiments for different N values. The values of N or the number of selected data are 22, 44, and 66. The result of the binary time series estimation are shown in table 1.
### Table 1. The parameter estimation result of binary time series

| N   | $\beta_0$ | S.E  | $\beta_1$ | S.E  |
|-----|-----------|------|-----------|------|
| 22  | -26.57    | 53.13| 1.26e+5   | 1.88e+5|
| 44  | -27.57    | 55.13| 1.22e+5   | 1.55e+5|
| 66  | -27.57    | 55.13| 1.03e+5   | 1.50e+5|

Based on table 1, it can be concluded that the greater the number of observations, the fewer errors will remain. In the previous section (section 6), the result of parameter estimation was shown for $p+1$ parameters. This section presents an application on the data when $p = 1$. The case study could be extended for data with larger values of $p$. The usefulness of the result can be divided into two views: theoretically, the parameter estimation result can be extended for other types of the time series model, while in the application view, the model can be applied for more lag, based on the condition of the data.

### 8. Conclusion

In the partial likelihood method, the probability density function (pdf) is not a joint pdf as in the Maximum Likelihood Estimation (MLE) method, but uses a marginal pdf. Then maximize the partial likelihood function, by finding the log-partial likelihood function and the first derivative of the log-partial likelihood function against its parameters and equated to zero. Based on the data simulation, the results of the observation parameters for observations were 22 observations, 44 observations, and 66 observations using partial likelihood method with Fisher Scoring iteration resulted in parameter estimation which did not differ significantly. However, it was seen from the standard error obtained that if the number of observation is greater than the standard error will be smaller.

### Acknowledgments

This work was financially supported by Universitas Indonesia under research grant PITTA with grant contract number No. 2336/UN2.R3.1/HKP.05.00/2018.

### References

[1] Cryer J D and Chan K 2010 *Time Series Analysis with Applications in R* 2nd edition (New York: Springer Science and Business Media LLC)

[2] Fokianos K and Kedem B 2000 *Regression Models for Binary Time Series* (Cyprus: Department of Mathematics and Statistics Maryland and Cyprus USA)

[3] Cox D R 1975 *Biometrika* 62 269-76

[4] Wei W W S 2006 *Time Series Analysis Univariate and Multivariate Method* 2nd edition (New York: Pearson Education)

[5] Zheng B and Agresti A 2000 *Stat. Med.* 19 1771-81

[6] Covrig M, Mircea L, Zbaganu G, Coser A and Tindeche A 2015 *Romanian Statistical Review* 3 33-45

[7] Montgomery D C, Peck E A and Vinning G G 2012 *Linear Regression Analysis* (New York: John Wiley & Sons)

[8] Lawless J F 2003 *Statistical Modells and Methods for Lifetime Data* (New York: John Willey and Sons)

[9] Wooldridge J M 2010 *Econometric Analysis of Cross Section and Panel Data* (London: MIT Press)