Binary Time Series Modeling with Application to Adhesion Frequency Experiments

Ying Hung[†], Veronika Zarnitsyna[‡], Yan Zhang[‡], Cheng Zhu[‡],

C. F. Jeff Wu[†*]

[†]H. Milton Stewart School of Industrial and Systems Engineering,

[‡]The Wallace H. Coulter Department of Biomedical Engineering,

Georgia Institute of Technology, Atlanta, GA

[*]Corresponding author

Abstract

Repeated adhesion frequency assay is the only published method for measuring the kinetic rates of cell adhesion. Cell adhesion plays an important role in many physiological and pathological processes. Traditional analysis of adhesion frequency experiments assumes that the adhesion test cycles are independent Bernoulli trials. This assumption can often be violated in practice. Motivated by the analysis of repeated adhesion tests, a binary time series model incorporating random effects is developed in this paper. A goodness-of-fit statistic is introduced to assess the adequacy of distribution assumptions on the dependent binary data with random effects. The asymptotic distribution of the goodness-of-fit statistic is derived and its finite-sample performance is examined via a simulation study. Application of the proposed methodology to real data from a T-cell experiment reveals some interesting information, including the dependency between repeated adhesion tests.

KEY WORDS: Goodness-of-fit test, Random effects, Cell adhesion, Micropipette experiments.

1. Introduction

This research is motivated by the statistical analysis of time series data from biomechanical experiments that study protein, DNA, and RNA at the level of single molecules (Mehta et al., 1999). Single molecule mechanics experiments employ ultrasensitive force techniques to characterize mechanically a single pair of molecules that physically links the force sensor to a sample surface. Figure 1 illustrates a simple experiment - the micropipette adhesion frequency assay (Chesla et al., 1998). Here, a human red blood cell (Figure 1, left) pressurized by micropipette suction is used as a force transducer to test interactions between molecules presented on the red cell membrane and the counter molecules on the surface of another cell (Figure 1, right, only partly shown). The two cells are put together for a pre-determined duration (Figure 1B), then retracted away. The simplest measurement is whether a controlled contact results in adhesion. If adhesion is resulted, retraction will stretch the red cell (Figure 1C). If no adhesion is resulted, the red cell will not be stretched (Figure 1A).

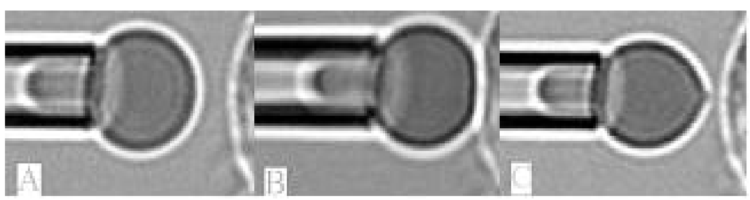

Figure 1: Photomicrographs of micropipette adhesion frequency assay

To ensure adhesion to be mediated by a single molecular bond, the experimental condition is designed such that adhesion is infrequent (Zhu et al., 2002). As such, in any particular test both positive (i.e., adhesion, scored 1) as well as negative (i.e., no adhesion, scored 0) outcomes are possible and random. Due to the inherent stochastic nature of single molecular interactions, such analysis would require a large number of measurements to obtain their statistical properties. For example, the probability of adhesion can be estimated from the frequency of occurrence of adhesion in a large number of contacts (Chesla et al., 1998). The probability distribution of single bond lifetimes can be estimated from the histogram of a large number of lifetime measurements (Marshall et al., 2003). Experimentally, these are obtained by sequentially repeating the measurements many times.

A crucial assumption that allows measurements from repeated tests to be used for probability calculation is that all measurements are identical yet independent from each other, i.e., the test sequence consists of independent and identically distributed random variables. However, this may or may not be valid depending on the particular biological system in question. Recently, Zarnitsyna et al. (2007) demonstrated that this assumption is not valid in some biological systems. Specifically, it is shown that the occurrence of adhesion in the immediate past test can either increase or decrease the likelihood for the next test to result in an adhesion. A simple analysis has been developed to determine whether the independent assumption is valid, and if not, to measure the amount of change in the probability of adhesion in the next test due to the occurrence of adhesion in the immediate past test.

In this article, we extend the simply analysis to a more sophisticated binary time series model. Numerous methods for binary time series analysis are available in the literature (Zeger and Qaqish, 1988; Li, 1994; Slud and Kedem, 1994; Benjamin et al., 2003). Most of these methods are developed for a single series of observations. Extensions to multiple binary time series modeling and related inferences have not been systematically studied. Both Li (1994) and Kedem and Fokianos (2002, p. 84) pointed out the importance of extensions to cases where a series is collected for each individual. This is different from classical time series analysis in that the binary time series are observed on different replicates of the experimental units. Correlation among the repeated observations may arise not only from memory effects but also from shared unobserved variables. Therefore, more general models are required to incorporate the correlations among repeated observations. Another important issue is model diagnostic. In distinction to Pearson's χ^2 test which works under the independence assumption, new test statistics and their theoretical properties need to be developed.

Table 1: Example of adhesion frequency experiment data

Average Number of Bonds (ANB)	50 Repeated Adhesion Tests
0.085	01010011011101010000 $\cdots$
0.085	00010000100010100110 $\cdots$
$\vdots$	$\vdots$
1.360	00111010000001000011 $\cdots$
1.360	11110000111000000011 $\cdots$

The remainder of this article is organized as follows. Some preliminary analysis results for an adhesion frequency experiment are presented in Section 2. In Section 3, a class of multiple binary time series models is proposed. A goodness-of-fit test for model assumptions and their asymptotic properties are derived in Section 4 and its finite-sample performance is examined via a simulation study. In Section 5, the proposed model and inferences are applied to the same experiment and the results are compared with those in Section 2. Summary and concluding remarks are given in Section 6.

2. Preliminary analysis of an adhesion frequency Experiment

In the micropipette adhesion frequency assay, adhesion between the two cells are staged by placing them onto controlled contact with given contact time and area via a computer-driven micromanipulation to ensure each contact was as close to identical to any other contacts as possible (Figure 1). Average number of bonds (ANB) is a transformation of the contact time (Chesla et al.,1998). For each ANB, several replicates of cell pairs are tested in the experiment. For each pair of cells, adhesion test cycle (i.e. contact and retraction) was repeated 50 times. Test scores (denoted by y) are recorded in binary form (i.e., $y = 0$ or 1), which results in multiple binary time series of the type exemplified in Table 1.

Under independent Bernoulli trial assumption (Chesla et al., 1998), the average adhesion probability (P_{ANB}) can simply be estimated by the adhesion frequency calculated as

$$P_{ANB} = \frac{\text{number of adhesions}}{\text{number of test cycles}}. \tag{1}$$

Figure 2 shows an example of the relationship between adhesion probability and ANB (unpublished data, courtesy of Y. Zhang and C. Zhu). In this micropipette experiment, the adhesion test was conducted with seven different ANBs (0.085, 0.17, 0.255, 0.34, 0.51, 0.68, and 1.36). For the first two ANBs, each has six pairs of replicates; for the remaining, there are five pairs each. Each point in Figure 2 represents the P_{ANB} value for one pair of cells, and is calculated from equation (1). The solid line represents the average over all the replicates under the same ANB.

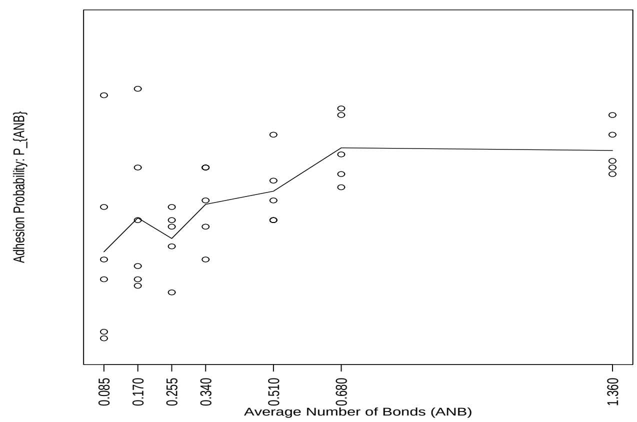

Figure 2: Adhesion probability (P_{ANB}) varies with the average number of bonds (ANB)

To understand the relationship between P_{ANB} and ANB, the existing method (Chesla et al., 1998) is based on the assumption that the binary time series data (e.g., Table 1) form Bernoulli sequences. However, for each pair of cells, the adhesion test cycles are observed repeatedly. The independence assumption may not hold as recently demonstrated (Zarnitsyna et al., 2007). Therefore, it is necessary to check the adequacy of the distributional assumption before applying the method. One graphical technique to assess this assumption is the probability plot. If the data are collected from independent Bernoulli trials, the number of trials needed to get one success will follow a geometric distribution with probability p, where p=Prob(y=1). For each ANB, the numbers of tests needed to get one success are calculated over all replicates. Then, its empirical cumulative distribution can be plotted against the geometric distribution, where the parameter p is estimated by (1) at each ANB. In the probability plots (not shown here to save space), significant deviations from the straight line would indicate violation of the independent Bernoulli assumption. Similar conclusions were first observed by using a different analysis in Zarnitsyna et al. (2007) which motivated our present work.

To gain further insight on the violation of the independent Bernoulli assumption, additional graphical plots are used here to better understand the dependence among repeated binary observations. The idea is to compare the conditional adhesion probability given the previous test results. Define P(1|1) to be the conditional adhesion probability given adhesion in the previous test, and P(1|0) the conditional probability given no adhesion in the previous test. If the test results are independent, P(1|1) should be equal to P(1|0) and both can be estimated by P_{ANB} in (1). In Figure 3, for each ANB, the green points represent the conditional probability P(1|1) calculated for each replicate. The green line stands for P(1|1) calculated over all replicates under the same ANB. Similarly, the red points and red line are those for the conditional probability P(1|0). For comparison, the black lines shows the adhesion probability P_{ANB} calculated by (1) at each ANB. As the green line and points are much higher than the red ones, the adhesion probability is higher if adhesion occurs in the previous test. This lends strong evidence for *memory effect* on repeated tests. A more in-depth biological discussion can be found in Zarnitsyna et al. (2007), where the memory effect was first observed by using a different analysis. From Figure 3, one can visually infer about the existence of serial correlation and interactions. The heterogeneity among subjects is also transparent from Figure 3. To describe and quantify significant effects on the adhesion probability, one should consider the use of a new binary time series model which incorporates the various effects suggested by the plots.

3. Modeling and Estimation

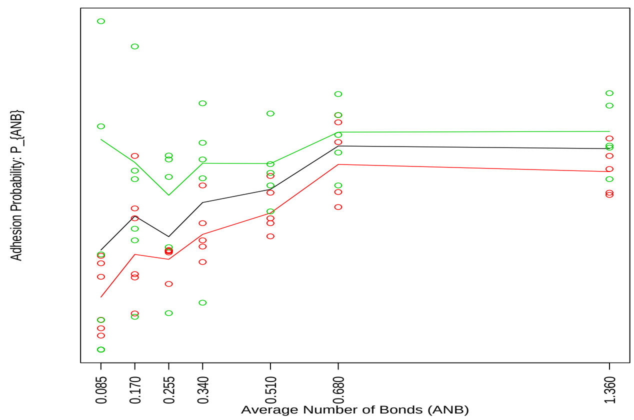

Figure 3: Memory effects in micropipette experiments

3.1 Modeling

In this section, a new binary time series model will be proposed. First, we need to review some existing models.

3.1.1 Random effects models

Random effects models are most useful in longitudinal data analysis when correlation arises from some unobservable variables shared among repeated observations. Consider a binary realization $\{y_{ij}\}$ taking values 0 or 1 for subject i at jth observation. For given subject-specific coefficients β_i, assuming the repeated observations for each individual are independent, the random effects model takes the form

$$log\frac{Pr(y_{ij}=1 \mid \beta_i)}{1-Pr(y_{ij}=1 \mid \beta_i)} = \beta_0 + \beta_i + x'_{ij}\alpha, \tag{2}$$

where the vector x_{ij} denotes the covariates associated with the fixed effects α, and the random effects β_i's are mutually independent with a common underlying multivariate distribution. This model is used to represent the natural heterogeneity across individuals in the regression coefficient. More discussion about this model can be found in Diggle et al. (2002).

3.1.2 Binary time series models

Non-Gaussian time series modeling techniques has been extensively discussed in the literature. Benjamin et al. (2003) proposed a generalized autoregressive moving average (GARMA) model. Applying this GARMA model with logistic link, a binary time series $\{y_t\}$ can be fitted as

$$logit(\mu_t) = x_t'\alpha + \sum_{r=1}^{R} \varphi_r \mathcal{A}(y_{t-r}) + \sum_{q=1}^{Q} \zeta_q \mathcal{M}(y_{t-q}, \mu_{t-q}), \tag{3}$$

where x_t are covariates at time t, $\mu_t = E(y_t \mid H_t)$ is the conditional mean given the previous information $H_t = \{x_t, \cdots, x_1, y_{t-1}, \cdots, y_1, \mu_{t-1}, \cdots, \mu_1\}$. $\mathcal{A}$ and $\mathcal{M}$ are functions representing the autoregressive (AR) and moving average (MA) terms with corresponding order R and Q. These two functions together are denoted by ARMA(R,Q). φ_r's and ζ_q's are the AR and MA parameters. For binary time series, a reasonable choice for $\mathcal{A}$ and $\mathcal{M}$ can be respectively y_t and residuals such as $y_t - \mu_t$.

Model (3) includes many well-known models as special cases. One important submodel is the Zeger-Qaqish model with logistic link

$$logit(\mu_t) = x_t'\alpha + \sum_{r=1}^{R} \varphi_r y_{t-r}, \tag{4}$$

and the moving average form for this model (Li, 1994)

$$logit(\mu_t) = x_t'\alpha + \sum_{q=1}^{Q} \zeta_q \Big(y_{t-q} - \mu_{t-q} \Big). \tag{5}$$

Asymptotic properties for the autoregressive logistic regression models are discussed via conditional likelihood (Kaufmann, 1987) and partial likelihood (Kedem and Fokianos, 2002).

We propose a *binary time series mixed model* (BTSM). It is a multiple logistic time series model with random effects that takes into account the heterogeneity among experimental units. Consider a binary time series realization $\{y_{it}\}$ taking values 0 or 1 for subject i at time t, where $i = 1, \cdots, m$, $t = 1, \cdots, n$, and $mn = N$. Suppose the experimental units are sampled from a population. It is reasonable to assume that the random effects β_i's are independent from a normal distribution with mean b and variance σ_b^2. For the vector $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m)'$, its distribution can be written as $\mathcal{N}(\mathbf{b}, \boldsymbol{\Sigma})$, where $\mathbf{b}$ is a column of b's having length m, $\boldsymbol{\Sigma} = \sigma_b^2 \mathbf{I}_m$, and $\mathbf{I}_m$ is the $m \times m$ identity matrix. The vector $x_{it} = \{x_{it,1}, \cdots, x_{it,p}\}'$ denotes the covariates associated with the p-dimensional fixed effects $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_p)'$, and $z_{it} = \{z_{it,1}, \cdots, z_{it,m}\}'$ the design matrix for the random effects $\boldsymbol{\beta}$ such that $z'_{it}\boldsymbol{\beta} = \beta_i$, that is $z_{it,i} = 1$ and $z_{it,j} = 0$ for all $j \neq i$.

Denote the conditional mean $\mu_{it} = \mathrm{E}(y_{it} \mid H_{it})$. Given the previous information $H_{it} = \{x_{it}, x_{it-1}, x_{it-2}, \cdots, y_{it-1}, y_{it-2}, \cdots, \mu_{it-1}, \mu_{it-2}, \cdots\}$ and random effects, y_{it} are conditionally independent with mean $\mathrm{E}(y_{it}|\boldsymbol{\beta}, H_{it}) = \mu^{\boldsymbol{\beta}}_{it}$. By logistic link function, the conditional mean $\mu^{\boldsymbol{\beta}}_{it}$ is related to the linear predictor $\eta^{\boldsymbol{\beta}}_{it}$ by

$$logit(\mu^{\boldsymbol{\beta}}_{it}) = \eta^{\boldsymbol{\beta}}_{it} = z'_{it}\boldsymbol{\beta} + x'_{it}\boldsymbol{\alpha} + \sum_{l=1}^{L} \gamma_i x_{it-l} y_{it-l} + \sum_{r=1}^{R} \varphi_r y_{it-r} + \sum_{q=1}^{Q} \zeta_q (y_{it-q} - \mu_{it-q}). \qquad (6)$$

This model is called a BTSM model. The random effects $\boldsymbol{\beta}$ are used to represent a variety of situations, including subject heterogeneity, unobserved covariates, and other forms of overdispersion. Here the heterogeneity is modelled directly through subject-specific parameter. If random intercept alone may not sufficiently capture the variation exhibited in the data, this model can be easily extended to a general form by incorporating more complicated random effects. Given β_i, the y_{it}'s are correlated because y_{it-l} explicitly influence y_{it}. This correlation can be explain by the AR and MA components in (6). The MA process which involves $\mu_{i,t-q}$ makes the model more complicated. In this formulation, the interaction terms $(x_{it-1}y_{it-1}, \cdots, x_{it-L}y_{it-L})$ between covariates and past outcomes provide flexibility in adjusting the time series structure with respect to different covariates settings.

The proposed BTSM model is general and includes the models discussed heretofore. The development here is based on the logistic link because it is popular and easy to interpret. It can, however, be easily extended to other link functions. The random effects model in (2) is a submodel of BTSM under the assumption that the repeated measurements for each unit are independent, and the correlation among repeated observations arises only from some unobserved variables. With the logistic link function, the GARMA model in (3) is a special case of BTSM if no random effect is included. That is, based on the population average, it models the time series structure without considering the heterogeneities among the units.

More than a simple extension of existing models, the BTSM model poses some challenging tasks. By considering the hidden variables shared among units, it incorporates random effects in logistic time series regression. This makes the estimation and inference more complicated and different from standard binary time series analysis. Another important issue is the goodness-of-fit test for model diagnostic. There are related works for linear mixed models in the literature (Jiang, 2001a,b). There is, however, no existing method for testing the distributional assumption in binary time series models with random effects. Furthermore, the asymptotic χ^2 distribution cannot be applied to the new test statistics because of its independence assumption. Instead, a martingale central limit theorem will be used in the next section to derive the asymptotic properties.

3.2 Estimation by partial likelihood

Model fitting procedure herein is based on partial likelihood (PL). PL was introduced by Cox (1975). More formal definition and theoretical justification can be found in Wong (1986). Fokianos and Kedem (2004) have discussions on using PL in time series which follow generalized linear models. For the BTSM model, the presence of random effects causes some integration difficulty, which makes the estimation different from standard methods in time series analysis. In this section, an approximation procedure will be proposed to tackle this problem.

Denote the observation vector by $\mathbf{y} = (y_1, \cdots, y_m)'$, where the observations for subject i are $y_i = (y_{i1}, \cdots, y_{in})'$, and $\boldsymbol{\gamma} = (\gamma_1, \cdots, \gamma_O)'$, $\boldsymbol{\varphi} = (\varphi_1, \cdots, \varphi_R)'$, $\boldsymbol{\zeta} = (\zeta_1, \cdots, \zeta_Q)'$. Assume

$$\omega = (\boldsymbol{\alpha}', \boldsymbol{\gamma}', \boldsymbol{\varphi}', \boldsymbol{\zeta}')'$$

are s-dimensional fixed effects, and $\mathbf{X}$ the corresponding matrix with rows

$$X_{it}' = (x_{it}', x_{i\,t-1}y_{i\,t-1}, \cdots, x_{i\,t-O}y_{i\,t-O}, y_{i\,t-1}, \cdots, y_{i\,t-R}, (y_{i,\,t-1} - \mu_{it-1}), \cdots, (y_{i,\,t-Q} - \mu_{it-Q})).$$

Similarly, with rows z_{it}', the design matrices for the random effects are denoted by $\mathbf{Z}$. Given the previous information H_{it}, the corresponding partial likelihood for fixed effects is

$$PL(\omega|\boldsymbol{\beta}) = \prod_{i=1}^{m}\prod_{t=1}^{n} pl_{\omega}(y_{it}|\boldsymbol{\beta}, H_{it}) = \prod_{i=1}^{m}\prod_{t=1}^{n}[\pi_{it}(\omega|\boldsymbol{\beta})]^{y_{it}}[1-\pi_{it}(\omega|\boldsymbol{\beta})]^{1-y_{it}},$$

where $\pi_{it}(\omega|\boldsymbol{\beta}) = P_{\omega|\boldsymbol{\beta}}(y_{it} = 1 \mid H_{it}) = \mu_{it}^{\boldsymbol{\beta}}$.

The integrated quasi-partial likelihood function used to estimate (ω, σ_b^2) is defined by

$$\mid \boldsymbol{\Sigma} \mid^{-1/2} \int exp\left[logPL(\omega|\boldsymbol{\beta}) - \frac{1}{2}\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}\right] d\boldsymbol{\beta}.$$

Because of the difficulty in implementing the full partial likelihood, we use penalized quasi-partial likelihood (PQPL) as an approximation. The integrated quasi-partial log-likelihood can be approximated by Laplace's method (Barndorff-Nielsen and Cox, 1989; Breslow and Clayton, 1993)

$$-\frac{1}{2}log|\mathbf{I}_m + \mathbf{Z}^t\mathbf{W}\mathbf{Z}\boldsymbol{\Sigma}| + \sum_{i=1}^{m}\sum_{t=1}^{n}(y_{it}log\frac{\pi_{it}(\omega|\tilde{\boldsymbol{\beta}})}{1-\pi_{it}(\omega|\tilde{\boldsymbol{\beta}})} + log(1-\pi_t(\omega|\tilde{\boldsymbol{\beta}}))) - \frac{1}{2}\tilde{\boldsymbol{\beta}}'\boldsymbol{\Sigma}^{-1}\tilde{\boldsymbol{\beta}}, \quad (7)$$

where $\mathbf{W}$ is the $N \times N$ diagonal matrix with diagonal terms $w_{it} = \pi_{it}(\omega|\tilde{\boldsymbol{\beta}})(1-\pi_{it}(\omega|\tilde{\boldsymbol{\beta}}))$ and $\tilde{\boldsymbol{\beta}} = \tilde{\boldsymbol{\beta}}(\omega, \sigma_b)$ is the solution of $\sum_{i=1}^{m}\sum_{t=1}^{n}(y_{it} - \pi_{it}(\omega|\boldsymbol{\beta}))z_{it} - \frac{\boldsymbol{\beta}}{\sigma_b^2} = 0$, which maximizes the sum of the last two terms in (7). Using derivations similar to Breslow and Clayton (1993), the penalized quasi-partial score equations for fixed effects ω and random effects $\boldsymbol{\beta}$ are:

$$\sum_{i=1}^{m}\sum_{t=1}^{n} X_{it}(y_{it} - \pi_{it}(\omega, \sigma_b)) = 0, \quad (8)$$

$$\sum_{i=1}^{m}\sum_{t=1}^{n} z_{it}(y_{it} - \pi_{it}(\omega, \sigma_b)) = \boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}, \quad (9)$$

where $\pi_{it}(\omega, \sigma_b) = P_{\omega,\sigma_b}(y_{it} = 1|H_{it})$. Given σ_b, the maximum quasi-partial likelihood estimator(MQPLE) of $(\hat{\omega}, \hat{\boldsymbol{\beta}})$ can be obtained by solving these two score equations. An important role in partial likelihood inference is played by the score process (8) and (9), which is a vector of martingales with respect to H_{it}. Hence, in Section 3.4 the study of asymptotic behavior of the MQPLE $\hat{\omega}$ will be based on central limit theorems for martingales.

Questions regarding existence and uniqueness of the MQPLE are important. Similar questions for the traditional maximum likelihood estimators (MLE) have been addressed by a number of authors (Silvapulle, 1981; Albert and Anderson, 1984; Wedderburn, 1976; Kaufmann, 1987). These results can also be applied to MQPLE and provide the essential conditions needed for existence and uniqueness of MQPLE.

The restricted maximum likelihood (REML) (Patterson and Thompson, 1971) version of the approximated profile quasi-likelihood function for the variance components can be written as

$$ql(\hat{\omega}(\sigma_b), \sigma_b) \approx -\frac{1}{2} log|\mathbf{V}| - \frac{1}{2} log|\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}| - \frac{1}{2}(\mathbf{Y} - \mathbf{X}\hat{\omega})'\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\hat{\omega}). \tag{10}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the design matrices, $\mathbf{V} = \mathbf{W}^{-1} + \mathbf{Z}\boldsymbol{\Sigma}\mathbf{Z}'$, and $\mathbf{Y}$ is a vector whose components are $Y_{it} = \eta_{it}^{\boldsymbol{\beta}} + \frac{(y_{it} - \mu_{it}^{\boldsymbol{\beta}})}{\mu_{it}^{\boldsymbol{\beta}}(1-\mu_{it}^{\boldsymbol{\beta}})}$. Differentiating (10) with respect to σ_b^2 gives the estimating equation for the variance components (Harville, 1977; Searle et al., 1992):

$$-\frac{1}{2}\left[(\mathbf{Y} - \mathbf{X}\omega)'\mathbf{V}^{-1}\frac{\partial \mathbf{V}}{\partial \sigma_b^2}\mathbf{V}^{-1}(\mathbf{Y} - \mathbf{X}\omega) - tr\left(P\frac{\partial \mathbf{V}}{\partial \sigma_b^2}\right)\right] = 0, \tag{11}$$

where $P = \mathbf{V}^{-1} - \mathbf{V}^{-1}\mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$.

Estimation of the fixed effects and variance components can be obtained by iteratively solving (8), (9) and (11). This estimation procedure is different from standard GLMM since the new model involves time series structure, i.e., the μ_{it-q} term depends on all the previous observations throughout the iterations. We may compute μ_{it-q} by setting initial μ_{it-q}'s to zero or to the sample mean of y_{it}. This should have negligible effect for a long enough iteration. Estimation can be carried out by simple modification in standard statistical software for GLMM such as SAS GLIMMIX package. Details about GLMM can be found in Breslow and Clayton (1993). Questions regarding robust estimation and efficient algorithm have been addressed by a number of authors (McCulloch, 1994; Lin and Breslow, 1996).

3.3 Asymptotic properties

Large sample properties for fixed effects and variance components in BTSM model are studied in this section. Considering a model which includes time-dependent covatiates, Fokianos and Kedem (2004) studied the asymptotic behavior of fixed effects ω in generalized linear time series models using partial likelihood inference. Theorem 1 is an extension of Fokianos and Kedem (2004) to multiple binary time series models with random effects. Based on the quasi-partial likelihood, Theorem 1 gives the consistency and asymptotic normality for the fixed effects estimators $\hat{\omega}$. With the help of the working dependent variables Y defined in Section 3, Theorem 2 gives the asymptotic properties for the REML estimator of σ_b^2 based on some asymptotic properties for linear mixed models with GLM iterative weights (Jiang, 1996). Assumptions and proofs are given in the appendix.

THEOREM 1. *Under assumptions A1 and A2, the maximum quasi-partial likelihood estimator (MQPLE) for the fixed effects $\hat{\omega}$ are consistent and asymptotically normal as $N \to \infty$:*

$$\sqrt{N}(\hat{\omega} - \omega) = \Lambda_N^{-1} \frac{1}{\sqrt{N}} S_n(\omega, \sigma_b) + o_p(1), \tag{12}$$

$$\sqrt{N}\Lambda_N^{1/2}(\hat{\omega} - \omega) \xrightarrow{d} \mathcal{N}(\mathbf{0}, I_s), \tag{13}$$

where the sample information matrix $\Lambda_N = \frac{1}{N}\sum_{t=1}^{n}\sum_{i=1}^{m} X_{it}X_{it}'\pi_{it}(\omega, \sigma_b)(1 - \pi_{it}(\omega, \sigma_b))$, *and* $S_n(\omega, \sigma_b) = \sum_{t=1}^{n}\sum_{i=1}^{m} X_{it}(y_{it} - \pi_{it}(\omega, \sigma_b)) = \mathbf{X}'(y - \pi(\omega, \sigma_b))$.

Based on Breslow and Clayton (1993), inference on variance component in model (6) can be formulated as an iterative procedure to estimate linear mixed model with the GLM iterative weight $\mathbf{W}^{-1}$ as

$$\mathbf{Y} = \mathbf{X}\omega + \mathbf{Z}\boldsymbol{\beta} + \epsilon, \tag{14}$$

where $\boldsymbol{\beta}$ comes from $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, $\epsilon = (\epsilon_1, \cdots, \epsilon_N)$ follows $\mathcal{N}(\mathbf{0}, \mathbf{W}^{-1})$, and the corresponding w_{it}'s are rewritten as $(w_1, \cdots, w_N)$. Recall that $\mathbf{\Sigma} = \sigma_b^2 \mathbf{I}_m$. Jiang (1996) developed rigorous asymptotic properties for REML estimates of variance components in linear mixed model (LMM) without the normality assumption on random effects and errors. Hence, Theorem 2 is a special case of Jiang (1996) for LMM with known unequal weights. Here we borrow some notation from Jiang (1996). Define

$$\mathbf{V}^* = A(A^t\mathbf{W}^{1/2}\mathbf{V}\mathbf{W}^{1/2}A)^{-1}A^t,$$

where A is any $N \times (N-s)$ matrix such that $\text{rank}(A) = N - s$ and $A^t \mathbf{W}^{1/2} \mathbf{X} = 0$,

$$g(\sigma_b) = (I_N, \sqrt{\sigma_b^2} \mathbf{W}^{1/2} \mathbf{Z})',$$

$$\varpi_l = \begin{cases} \epsilon_l \sqrt{w_l} \ , & 1 \le l \le N, \\ \frac{\beta_{l-N}}{\sqrt{\sigma_b^2}} \ , & N+1 \le l \le N+m, \end{cases}$$

$$\mathbf{V}_1 = (A'\mathbf{W}^{1/2}\mathbf{V}\mathbf{W}^{1/2}A)^{-1/2} A'\mathbf{W}^{1/2}\mathbf{Z}\mathbf{Z}'\mathbf{W}^{1/2}A(A'\mathbf{W}^{1/2}V\mathbf{W}^{1/2}A)^{-1/2},$$

$$V_1(\sigma_b) = g(\sigma_b)\mathbf{V}^* \mathbf{W}^{1/2}\mathbf{Z}\mathbf{Z}'\mathbf{W}^{1/2}\mathbf{V}^* g(\sigma_b)',$$

$$I^N = \frac{\text{tr}(\mathbf{V}_1 \mathbf{V}_1)}{m}, \ K = \frac{1}{m} \sum_{l=1}^{N+m} (E\varpi_l^4 - 3) V_1(\sigma_b)_{ll}^2 \,,$$

$$J = 2I^N + K.$$

Theorem 2. *Under assumptions A3 and A4, as $N \to \infty$ and $m \to \infty$, the REML estimate for variance components is consistent and asymptotically normal with*

$$J^{-1/2} I^N \sqrt{m} (\hat{\sigma}_b^2 - \sigma_b^2) \ \xrightarrow{d} \ \mathcal{N}(0,1). \tag{15}$$

4. Goodness-of-fit for model diagnostics

4.1 Goodness-of-fit test

Pearson's χ^2 test is generally used to test if data follow some specific distribution. An important assumption for this test is the independence of the observations. How to perform testing for model assumptions when the data come from a binary time series model? One approach is to classify the responses according to mutually exclusive events in terms of the previous output and then check the difference between observed and theoretical frequencies in each category. This can be written as follows.

Assume the binary data y_{it} comes from a binomial distribution with probability p depending on H_{it}. H_{it}, defined in Section 3.1, can be decomposed into several mutually exclusive events. Recall that $H_{it} = \{x_{it}, x_{it-1}, x_{it-2}, \cdots, y_{it-1}, y_{it-2}, \cdots, \mu_{it-1}, \mu_{it-2}, \cdots\}$. Suppose the decomposition is decided by n_1 covariates $(x_{it}, \cdots, x_{i\,t-n_1})$ decomposed into c_1 exclusive subsets, n_2 autoregression effects $(y_{i\,t-1}, \cdots, y_{i\,t-n_2})$ decomposed into c_2 exclusive subsets, and n_3 moving average effects $(\mu_{i\,t-1}, \cdots, \mu_{i\,t-n_3})$ decomposed into c_3 exclusive subsets, where $c_i \geq 1$, for $i = 1, 2, 3$. Therefore, there are K exclusive events denoted by $E_1, \ldots, E_K$ with

$$K = c_1 c_2 c_3, \tag{16}$$

and define

$$M_k \equiv \sum_{i=1}^{m} \sum_{t=1}^{n} y_{it} \mathbf{1}_{(H_{it} \in E_k)},$$

$$e_k(\omega, \sigma_b) \equiv \sum_{i=1}^{m} \sum_{t=1}^{n_i} \mathbf{1}_{(H_{it} \in E_k)} P_{\omega, \sigma_b}(y_{it} = 1 \mid H_{it}).$$

Similar to Pearson's χ^2-test, a test statistic can be defined by

$$\chi^* \equiv \sum_{k=1}^{K} \frac{(M_k - e_k(\omega, \sigma_b))^2}{\mathrm{E}(M_k)}. \tag{17}$$

If the parameters ω and σ are known, the asymptotic distribution for this test statistic is χ^2 with K degrees of freedom.

For the BTSM model, new construction and asymptotic properties of the goodness-of-fit test need to be rigorously established for two reasons. First, the probability $P_{\omega,\sigma_b}(y_{it} = 1 \mid H_{it})$ is not completely specified under the null hypothesis because it involves the unknown parameters (ω, σ_b). After replacing the unknown parameters in $e_j(\omega, \sigma_b)$ by the estimated values $(\hat{\omega}, \hat{\sigma_b})$, the χ^2 approximation may not be valid (Chernoff and Lehmann, 1954; Jiang, 2001b). Second, because of the random effects and time series structures in the BTSM model, the observations are correlated. Accordingly, the asymptotic χ^2 result may not follow from the classic central limit theorem.

Jiang (2001b) derived the asymptotic distribution for goodness-of-fit test in linear mixed models (LMM) with continuous response to assess the adequacy of distributional assumptions. A new test statistic is constructed here based on binary observations and the corresponding time series model. Furthermore, the BTSM model includes time-dependent covariates. For this general formulation, inference is made based on the partial likelihood function. Asymptotic distribution of the new test statistic is derived by exploiting the martingale properties of the quasi-partial score process, which is different from Jiang (2001b).

Define a new goodness-of-fit test statistics for distributional assumptions in the BTSM model as

$$\hat{\chi}^2 = \frac{1}{N}\sum_{k=1}^{K}(M_k - e_k(\hat{\omega}, \hat{\sigma}_b))^2. \tag{18}$$

Unlike the Pearson's χ^2 test, the asymptotic distribution for this new statistic may not be χ^2. Hence, there is no need to have a normalizing constant in the test statistic to achieve χ^2 distribution. Instead, for simplicity, we choose a unified N as suggested in Jiang (2001b).

The asymptotic properties for the test statistic (18) are given in Theorem 3. Proofs are given in the appendix. The following notation is used in the theorem. Define

$$\theta = (\omega', \sigma_b),\ D = \left[\frac{1}{N}\sum_i\sum_t \mathbf{1}_{(H_{it}\in E_k)}\frac{\partial}{\partial\omega'}\pi_{it}(\theta)\right]_{1\le k\le K}\Lambda_N^{-1},$$

$$G_{it} = [\mathbf{1}_{(H_{it}\in E_k)} - DX_{it}]_{1\le k\le K},$$

$$C = \frac{1}{\sqrt{m}}\mathbf{W}^{1/2}\mathbf{V}^*\mathbf{W}^{1/2}\mathbf{Z}\mathbf{Z}'\mathbf{W}^{1/2}\mathbf{V}^*\mathbf{W}^{1/2},$$

$$\Phi = \frac{(I^N)^{-1}}{\sqrt{m}}\left[\sum_i\sum_t(\mathbf{1}_{(H_{it}\in E_k)}\frac{\partial}{\partial\sigma_b^2}\pi_{it}(\theta))\right]_{1\le k\le K},$$

$$h_{it} = G_{it}(y_{it} - \pi_{it}(\theta)) - \Phi C_{it}(Y_{it} - X'_{it}\omega)^2,$$

$Vh = \sum_{i=1}^m\sum_{t=1}^n \mathrm{Var}(h_{it})$, $R = \mathrm{tr}((C\mathbf{V})^2) - \sum_i\sum_t(C\mathbf{V})_{it}^2$ and

$$\Psi_N = (N)^{-1}[Vh + 2\Phi R\Phi']. \tag{19}$$

Note that, for $N\times N$ matrixes C and $C\mathbf{V}$, C_{it} and $(C\mathbf{V})_{it}$ indicate the $((i-1)n+t)$-th diagonal elements respectively.

THEOREM 3. *Suppose Ψ_N in (19) converges to a limiting value Ψ. Under the assumptions A1-A8, as $N\to\infty$, the asymptotic distribution of the goodness-of-fit statistics (18) is*

$$\hat{\chi}^2 \xrightarrow{d} \textstyle\sum_{j=1}^K \lambda_i\mathbb{Z}_i^2, \tag{20}$$

where $\Gamma = diag(\lambda_1, \cdots, \lambda_K)$, *and* λ_i *are the eigenvalues of* Ψ *and* $\mathbb{Z}_1, \cdots, \mathbb{Z}_K$ *are i.i.d.* $\mathcal{N}(0,1)$.

Let $\hat{\Psi} = N^{-1}[\widehat{Vh} + 2\hat{\Phi}\hat{R}\hat{\Phi}']$ denote the estimate of (19). Computation of $\hat{\Psi}$ is essential to obtaining the critical values in the goodness-of-fit test. In practice, it is often straightforward to evaluate $\hat{\Psi}$ by Monte-Carlo method as follows:

$$\begin{aligned}
\hat{\Psi} &= N^{-1}[\textstyle\sum_i \sum_t \widehat{\mathrm{Var}(h_{it})} + 2\hat{\Phi}\hat{R}\hat{\Phi}'] \\
&\approx N^{-1}\left[\textstyle\sum_i \sum_t \frac{1}{U}\sum_{u=1}^{U}(\hat{h}_{it,(u)} - \overline{h_{it}})(\hat{h}_{it,(u)} - \overline{h_{it}})' + 2\frac{1}{U}\sum_{u=1}^{U}[\hat{\Phi}_{(u)}\hat{R}_{(u)}\hat{\Phi}'_{(u)}]\right] \\
&= \frac{1}{U}\left[N^{-1}\textstyle\sum_i \sum_t (\hat{h}_{it,(u)} - \overline{h_{it}})(\hat{h}_{it,(u)} - \overline{h_{it}})' + 2\frac{1}{N}\sum_{u=1}^{U}\hat{\Phi}_{(u)}\hat{R}_{(u)}\hat{\Phi}'_{(u)}\right],
\end{aligned}$$

where U are the number of Monte-Carlo simulations, $\hat{h}_{it,(l)}$, $\hat{\Phi}_{(l)}$, $\hat{I}^N_{(l)}$, $\hat{R}_{(l)}$ are estimates with θ replaced by $\hat{\theta}$, $\overline{h_{it}} = \frac{1}{U}\sum_{u=1}^{U}\hat{h}_{it,(u)}$, y_{it} are a sample from Bernoulli trials with probability following the fitted BTSM model and β_i are i.i.d. variables generated from $\mathcal{N}(\hat{b}, \hat{\sigma}_b)$. As mentioned in Section 3.2, Laplace's method can be applied to approximate integration in $\frac{\partial \pi_{it}(\theta)}{\partial \omega'}$ and $\frac{\partial \pi_{it}(\theta)}{\partial \sigma_b^2}$.

4.2 Finite-sample performance and empirical application

To examine the finite-sample performance of the proposed tests, we carry out some simulations under nulls and alternatives. Each result is calculated based on 5000 simulations with 5% significant level. Two sample sizes N=480 (m=25, n=20) and N=160 (m=16, n=10) and four different partitions (K=2, 4, 6, 8) are studied. For simplicity, we only focus on equal cell partitions in this simulation study. As mentioned in Section 4.1, when unknown parameters are involved, there is no existing test which has valid asymptotic distribution. Thus, we compare our method with a naive test, namely, the Pearson χ^2-test in (17) but with parameters estimated. Since parameters are not assumed to be known, a naive way to apply the Pearson χ^2-test is to modify the asymptotic χ^2 distribution with $K-1-a$ degrees of freedom, where a is the number of parameters being estimated. Here, the comparison is conducted only for K=8. For example, for the second model (BTSM-AR(2)) in Table 2, there are five parameters being estimated (three fixed effects, one random effect, and one corresponding variance). Thus, the naive Pearson χ^2-test under comparison has an asymptotic χ^2 distribution with

Table 2: BTSM models with four different time series structures

BTSM–	Model	β_i
AR(1)	$logit(\mu_{it}) = \beta_i + 1.3y_{it-1} + 0.3x_{it}$, $x_{it} = 0.2, 0.4, 0.6, 0.8$	$\mathcal{N}(-0.3, 0.5)$
MA(1)	$logit(\mu_{it}) = \beta_i + 1.3(y_{it-1} - \mu_{it-1}) + 0.3x_{it}$, $x_{it} \in (0, 1)$	$\mathcal{N}(-0.3, 0.5)$
AR(2)	$logit(\mu_{it}) = \beta_i + y_{it-1} + 0.5y_{it-2} + 0.3x_{it}$, $x_{it} \in (0, 1)$	$\mathcal{N}(-1, 0.5)$
ARMA(1,1)	$logit(\mu_{it}) = \beta + 1.5y_{it-1} + 0.5(y_{it-1} - \mu_{it-1}) - 0.5x_{it}$, $x_{it} \in (0, 1)$	$\mathcal{N}(-0.3, 0.5)$

two degrees of freedom ($2 = 8 - 1 - 5$).

Binary data are generated by using the BTSM models listed in Table 2 with four different time series structures. Table 3 reports the empirical rejection probabilities associated with these four models to examine the empirical level of the test. In general, when the sample size increases, the empirical level of the proposed test becomes more stable with respect to the number of partitions K. Compared with the naive Pearson χ^2-test, the proposed method performs better in the following two respects. First, when the number of estimated parameters involved in the model increases, the proposed method provides a more stable empirical level. For example, the empirical level of the naive test almost doubles and far exceeds the nominal 5% level when the number of estimated parameters increases from four (AR(1) or MA(1)) to five (AR(2) or ARMA(1,1)). This is because the critical value of the naive test decreases rapidly when the number of estimated parameters increases. The other advantage of the proposed method is the performance robustness to sample size. For the naive test, the empirical level increases dramatically when the sample size decreases, while for the proposed method, the increase is slight to modest.

In terms of power, we choose two types of alternatives to assess the distributional assumptions involved in the fitted model (at 5% level), including the Bernoulli assumption for the binary data and the normal assumption for the random effects. The first alternative assumes that the random effects are normally distributed and the binary data follow a *beta-binomial* distribution. That is, y_{it} are generated from Bernoulli($\mathcal{P}_{it}$) distribution, and $\mathcal{P}_{it}$ is a random

Table 3: Empirical level of the goodness-of-fit test at 5 %

Model	(m,n)	$K=2$	$K=4$	$K=6$	$K=8$	χ^2
BTSM–AR(1)	$(24,20)$	0.042	0.046	0.046	0.044	0.094
	$(16,10)$	0.047	0.049	0.054	0.055	0.119
BTSM–MA(1)	$(24,20)$	0.046	0.052	0.054	0.062	0.084
	$(16,10)$	0.063	0.056	0.065	0.074	0.172
BTSM–AR(2)	$(24,20)$	0.056	0.042	0.048	0.06	0.151
	$(16,10)$	0.06	0.048	0.054	0.062	0.277
BTSM–ARMA(1,1)	$(24,20)$	0.044	0.046	0.042	0.046	0.294
	$(16,10)$	0.055	0.058	0.051	0.056	0.346

Table 4: Power of testing Bernoulli assumption under beta-binomial distribution

Model	(m,n)	$K=2$	$K=4$	$K=6$	$K=8$	χ^2
BTSM–AR(1)	$(24,20)$	0.651	0.659	0.718	0.772	0.574
	$(16,10)$	0.361	0.403	0.608	0.548	0.296
BTSM–MA(1)	$(24,20)$	0.792	0.806	0.811	0.912	0.778
	$(16,10)$	0.528	0.729	0.579	0.634	0.428
BTSM–AR(2)	$(24,20)$	0.688	0.858	0.762	0.756	0.728
	$(16,10)$	0.603	0.596	0.586	0.594	0.578
BTSM–ARMA(1,1)	$(24,20)$	0.402	0.818	0.754	0.746	0.648
	$(16,10)$	0.216	0.286	0.428	0.502	0.486

variable with a Beta(μ_{it}, $1-\mu_{it}$) distribution. The other alternative assumes a departure from the normal assumption for random effects. Let y_{it} follow Bernoulli(μ_{it}) distribution, and the random effect $\boldsymbol{\beta}$ follows a mixture of two normal distributions $\mathcal{N}(\mathbf{b}_1, 1)$ and $\mathcal{N}(\mathbf{b}_2, 1)$ with probability *prob* and $1-prob$, denoted by $MIXN(\mathbf{b}_1, \mathbf{b}_2, prob)$. In the simulation, the random effect is assume to be $MIXN(-\mathbf{0.5}, \mathbf{0.5}, 0.3)$. For each alternative, μ_{it} are obtained from the values specified in the four models given in Table 2. Based on the generated data, models are fitted by the procedure described in Section 3.

Tables 4 and 5 report the empirical rejection probability for both alternatives associated with four BTSM models to examine the empirical power. Clearly, for the first two models, the

Table 5: Power of testing normal random effect under mixed normal distribution

Model	(m, n)	$K = 2$	$K = 4$	$K = 6$	$K = 8$	χ^2
BTST–AR(1)	$(24, 20)$	0.319	0.427	0.486	0.674	0.354
	$(16, 10)$	0.205	0.314	0.458	0.327	0.135
BTSM–MA(1)	$(24, 20)$	0.994	1	1	0.998	0.993
	$(16, 10)$	0.924	0.983	0.988	0.97	0.826
BTSM–AR(2)	$(24, 20)$	0.596	0.607	0.686	0.702	0.618
	$(16, 10)$	0.336	0.375	0.395	0.448	0.432
BTSM–ARMA(1,1)	$(24, 20)$	0.546	0.658	0.586	0.616	0.518
	$(16, 10)$	0.323	0.348	0.356	0.434	0.346

proposed test is more powerful than the naive test for both alternatives (with the exception of BTSM–AR(1), $m = 24$, $n = 20$, $K = 2$ in Table 5). In some cases of the last two models, when K is small (mostly for $K = 2$, and some for $K = 4$), the naive method has more power than the proposed method, but this is due to the higher empirical levels of the former in Table 3. Another issue is the dependence of performance on K, the number of cells. It is well known that the power of this type of goodness-of-fit test can vary greatly with K. This is observed in the simulation results, especially when the sample size is smaller. Therefore, proper choice of partitions is important. This leads to the following guidelines for choosing the optimal number of partitions.

Although the construction of the goodness-of-fit test allows arbitrary partition of the cells, its performance depends on a proper choice of the number of exclusive subsets K in (16). How to choose the optimal number of partitions? First, to ensure enough power, K should not be too small, because the fewer cells the more difficult to distinguish between two distributions. On the other hand, if there are too many cells, the size of the test may become a problem. This is because the asymptotic distribution of the test is based on a K-dimensional central limit theorem. A necessary condition to maintain this asymptotic property is that $K/N^{1/5} \to 0$ (Senatov, 1980; Jiang, 2001b). Therefore, the proper number of partitions should be chosen from 1 to $\lceil N^{1/5} \rceil$. Within this range, conducting a simulation with comparable sample size will be helpful in determining the optimal number of partitions. R code for the simulations are available on http://www2.isye.gatech.edu/∼jeffwu/publications/, which can be easily implemented.

5. Application in adhesion frequency experiment

In this section, we revisit the adhesion frequency experiment data and apply the proposed model to predict the adhesion probability. As in Section 2, there are 37 pairs of cells used in this experiment. Adhesion test cycles for each pair are repeated 50 times. To study the time series behavior, for every subject, the first five observations are treated as additional predictor variables. Therefore, in this example, m=37, n=45. The covariate here is the average number of bonds denoted by ANB_i for the i-th pair of cells. For each pair of cells, the ANB is fixed. Therefore, there is no time-dependent covariate in this example and the one-dimensional (p=1) covariates in model (6) can be simplified by assuming $x_{it} = x_{it,1} = ANB_i$, for all t.

With fixed effects

$$\omega = (\alpha_1, \gamma_1, \varphi_1, \zeta_1),$$

and the corresponding $X'_{it} = (ANB_i, ANB_i \times y_{it-1}, y_{i\,t-1}, (y_{i\,t-1} - \mu_{i\,t-1}))$, the fitted BTSM model for adhesion probability is given below:

$$logit(\mu_{it}) = \beta_i + \alpha_1 ANB_i + \gamma_1 ANB_i \times y_{it-1} + \varphi_1 y_{it-1} + \zeta_1 (y_{i\,t-1} - \mu_{i\,t-1}), \qquad (21)$$

where $\beta_i \sim \mathcal{N}(-1.33, 0.44)$. The value of the MQPLE is

$$\hat{\omega} = (0.97, -0.62, 1.76, -0.86)$$

with the corresponding p-values 0.004, 0.031, $<$ 0.001, and 0.006. The estimated variance component $\hat{\sigma}_b$=0.4 (with standard deviation 0.14) provides clear evidence on the substantial heterogeneity among subjects. In model (21), ANB_i and $y_{i\,t-1}$ have significant effects on the cell adhesion probability at time t. The positive α_1 value of 0.97 indicates that the cell adhesion probability increases with respect to the ANB. The adhesion memory can be described by a first-order autoregressive and moving average process. The positive φ_1 value of 1.76 indicates that the adhesion probability is higher if adhesion occurs in the previous test. The significant interaction ($ANB_i \times y_{it-1}$) plays an important role in the model interpretation. Based on the fitted model (21), the coefficient of the ANB_i, $0.97 - 0.62y_{i\,t-1}$, shows that the effect of ANB is smaller if an adhesion occurs in the previous test ($y_{i\,t-1}$=1). On the other hand, based on the coefficient of $y_{i\,t-1}$, i.e., $1.76-0.86-0.62ANB_i$=$0.9-0.62ANB_i$, the effect of $y_{i\,t-1}$ is reduced as the average number of bonds increases. Furthermore, the memory effect is close to 0 if the ANB is around 1.45 (=0.9/0.62). It implies that, if two cells with ANB more than 4.3 seconds, the repeated adhesion tests become nearly independent. This model gives so much new information on the adhesion frequency analysis, because it provides not only a flexible model for considering the memory effect but also the conditions under which the independence assumption may hold.

The distributional assumptions here are the normally distributed random effects and the dependent Bernoulli distributed responses. To assess their adequacy, the proposed goodness-of-fit test (18) is applied in this example. Based on some simulation studies that we suggested in section 4.2, the optimal number of partition in this example is $K = 4$. Therefore, we first partition the the previous information space H_{it} into four disjoint events as follows:

$$E_1 = (y_{i\,t-1} = 0, \mu_{i\,t-1} > 0.5), \quad E_2 = (y_{i\,t-1} = 0, \mu_{i\,t-1} \leq 0.5),$$
$$E_3 = (y_{i\,t-1} = 1, \mu_{i\,t-1} > 0.5), \quad E_4 = (y_{i\,t-1} = 1, \mu_{i\,t-1} \leq 0.5).$$

That is, c_1=1, c_2=2, c_3=2 in (16). 5000 Monte-Carlo simulations are conducted to evaluate $\hat{\Psi}$. The corresponding eigenvalues for $\hat{\Psi}$ are {0.3100, 0.1428, 0.0350, 0.0252}. By Theorem 3, the critical values of the proposed goodness-of-fit test at α=0.01, 0.05, and 0.1 are 2.3819, 1.6271, and 1.2556 respectively. The test statistic under model (21) is $\hat{\chi}^2$=0.9392, which is much smaller than the critical values. There is thus no evidence to reject the hypothesis that the binary responses in adhesion tests follow a dependent Bernoulli distribution with probability given by model (21). Similar to the study in Section 4.2, we compare the proposed test with the naive χ^2 test in (17) with one degree of freedom. The nave test statistic has the value 6.8838 with the corresponding p-value of 0.0087. This would lead to the rejection of the hypothesis of dependency and the model in (21). In view of the simulation results in Section 4.2 that the naive test can have an exceedingly large test statistic value, such a conclusion cannot be taken seriously.

Recall that the preliminary analysis in Section 2 shows some memory effects in the repeated observations. By applying the BTSM model, the cell adhesion memory can be described by an ARMA(1,1) process. Besides, model (21) can quantify the effect of ANB and identify a significant interaction between ANB and the previous test result. This is a great advantage, because in practice it is difficult to assess the moving average and interaction effects by graphical analysis. As shown in this example, by including the interaction term, the BTSM model provides flexibility in capturing different time series structures with respect to different covariates. Given the fitted models, goodness-of-fit tests are conducted to check the distributional assumptions. The test result provides statistical evidence on the adequacy of the distributional assumption and supports model-based predictions. Another advantage of the BTSM model is that it incorporates the random effects. Thus inference and predictions can be made beyond the particular subjects used in the experiment.

6. Summary and concluding remarks

Despite the prevalence of multiple binary time series data in many applications, their modeling and inference have not been systematically studied in the literature. We propose a binary time series mixed model (BTSM) to analyze data when a repeated binary time series is observed for each subject. It handles multiple time series by incorporating random effects to borrow strength across different subjects. Thus, inference and predictions can be made beyond the specific units in the study. The BTSM model includes numerous known models as special cases. Moreover, it may have applications in longitudinal analysis.

Estimators for the fixed effects and variance components are shown to be consistent and asymptotically normally distributed. To assess the adequacy of the distributional assumptions in the BTSM model, we propose a new goodness-of-fit test. Because there are some unknown parameters and the data are dependent, the asymptotic distribution for the test statistic is derived by using a martingale central limit theorem. Not surprisingly, the results are different from the classical Pearson's χ^2 test. The proposed test outperforms the naive Pearson's χ^2 test in an simulation study. Some guidelines are given on the choice of the optimal number K of partitions.

As an application, the BTSM model is applied to fit some multiple binary time series observed on T-cell adhesion frequency experiment. This study demonstrates how the BTSM model can help in quantitatively describing the effects of significant factors. Furthermore, the fitted model provides valuable information on moving average and interaction effects, which cannot be obtained from graphical analysis. This example shows that the first-order autocorrelation effect can be observed from graphical analysis, but not when higher order autocorrelations are present. The goodness-of-fit test is also demonstrated in this example. Although the covariates in this example are independent of time, the proposed model and inference are generally applicable to problems with time-dependent covariates.

Acknowledgments

The research of CFJW is supported by NSF grant DMS 0305996. The research of CZ is supported by NIH grants AI38282 and AI44902. We are grateful to the Editor, an AE and two referees for their helpful comments and suggestions.

Appendix A: Assumptions

A1. The parameter ω belongs to an open set $B \subseteq R^s$.

A2. The covariate matrix X_{it} lies almost surely in a nonrandom compact subset of R^s such that $P[\sum_i \sum_t X_{it}' X_{it} > 0] = 1$.

A3. $\sigma_b \geq 0$ *and* $Var(\beta_1^2) > 0$.

A4. *As* $N \to \infty$, *lim inf* $\lambda_{min} Cor(I_{N-s}, \mathbf{V}_1) > 0$ *and lim* $tr(\mathbf{V}_1'\mathbf{V}_1)^{1/2} = \infty$, *where for matrices* $A_1, A_2, Cor(A_1, A_2) = tr(A_1'A_2)/[tr^{1/2}(A_1'A_1)tr^{1/2}(A_2'A_2)]$.

A5. $\|N^{-1/2}a_T'D\|$ *and* $\|N^{-1/2}a_T'\Phi J^{1/2}\|$ *are bounded, where* $\|\kappa\|=(\kappa'\kappa)^{1/2}$ *for any vector* κ.

A6. Define $\Delta' = (\mathbf{Y} - \mathbf{X}\omega)'\mathbf{V}^{-1/2}$ *as a vector with elements* Δ_{it}, *and* $\sigma_N^2 = \sum_i \sum_t Var\bigg((C^*\mathbf{V})_{it}\Delta_{it}^2 + G_{it}^*(y_{it} - \pi_{it}(\theta))\bigg) + 2\sum_{(it)' \neq it}(C^*\mathbf{V})^2_{it,(it)'}$. *There exists numbers* L_{it}, *such that as* $N \to \infty$, *the following quantities converge to 0:* $\sigma_N^{-2}\{\sum_{it} E[(C^*\mathbf{V})_{it}(\Delta_{it}^2 - 1)]^2 \mathbf{1}_{(|\Delta_{it}| > L_{it})} + \frac{1}{2}\sum_{it \neq (it)'}((C^*\mathbf{V})_{it,(it)'})^2[\delta_{it} + \delta_{(it)'}]\}$, *and* $\sigma_N^{-2}\sum_{it}(G_{it}^*)^2 E(y_{it} - \pi_{it}(\theta))^2 \mathbf{1}_{(|y_{it} - \pi_{it}(\theta)| > L_{it})}$, *where* $\delta_{it} = E\Delta_{it}^2 \mathbf{1}_{(|\Delta_{it}| > L_{it})}$.

A7. There exists numbers L_{it}, *such that as* $N \to \infty$, *the following quantities converge to 0:* $\sigma_N^{-4}\{\sum_{it} E[(C^*\mathbf{V})_{it}(\Delta_{it}^2 - 1)]^4 \mathbf{1}_{(|\Delta_{it}| \leq L_{it})} + \sum_{it \neq (it)'}((C^*\mathbf{V})_{it,(it)'})^4 \delta_{it}^* \delta_{(it)'}^* + \sum_{it}[\sum_{(it)' \neq it}((C^*\mathbf{V})_{it,(it)'})^2]^2 \delta_{it}^*\}$, *and* $\sigma_N^{-4}\sum_{it}(G_{it}^*)^4 E(y_{it} - \pi_{it}(\theta))^4 \mathbf{1}_{(|y_{it} - \pi_{it}(\theta)| \leq L_{it})}$, *where* $\delta_{it}^* = E\Delta_{it}^4 \mathbf{1}_{(|\Delta_{it}| \leq L_{it})}$.

A8. As $N \to \infty$, $\lambda_{max}\xi'\xi/\sigma_N^2 \to 0$, *where* $\xi = (C^*\mathbf{V}) - diag(C^*\mathbf{V})$.

Assumptions A1 and A2 are required for the asymptotic properties for fixed effects estimated from partial likelihood. Lindeberg's condition holds under assumption A2 (Fokianos and Kedem, 1998), which leads to the proof of Theorem 1. Assumptions A3 and A4 are the same as in Jiang (1996).

Only a sketch of the proofs is given in the appendix. Details can be found on http://www.amstat.org/publications/jasa/supplemental_materials.

Appendix B: Proof of Theorem 1

Based on the partial likelihood, the partial score process for ω can be written as

$$S_n(\omega, \sigma_b) = \sum_{t=1}^{n}\sum_{i=1}^{m} X_{it}(y_{it} - \pi_{it}(\omega, \sigma_b)).$$

Assume a σ-field is generated from the past data and covariates $\mathcal{F}_{n-1} = \sigma(H_{1n}, H_{2n}, \cdots, H_{mn})$. It is clear that $\text{E}[S_n(\omega, \sigma_b) \mid \mathcal{F}_{n-1}] = S_{n-1}(\omega, \sigma_b)$, and $\text{E}[S_n(\omega, \sigma_b)]$=0. Base on this fact and

A1 and A2, it is easy to see that the partial score process $S_n(\omega, \sigma_b)$ is the sum of zero-mean martingale differences with respect to $\mathcal{F}_{n-1}$. The asymptotic normality follows from the martingale central limit theorem. Detail of the proof is analogue to Slud and Kedem (1994).

Appendix C: Proof of Theorem 2

The inference for the variance component can be formulated as a linear mixed model with variances of error terms following the GLM iterative weights in (14). Define $\mathbf{Y}^*= \mathbf{W}^{1/2}\mathbf{Y}$, $\mathbf{X}^*=\mathbf{W}^{1/2}\mathbf{X}$, $\mathbf{Z}^*=\mathbf{W}^{1/2}\mathbf{Z}$, $\epsilon^*=\mathbf{W}^{1/2}\epsilon$. Replacing them in (14), the results directly follow as a special case of Theorem 4.1 of Jiang (1996).

Appendix D: Proof of Theorem 3

The proof is along the lines of Jiang (2001b). It consists of several lemmas that culminate in the final proof.

Lemma D.1. *Under the same assumptions in Theorem 3, define*

$$\Psi_n^{(1)} = \Psi_n^{(1)}(\theta) = n^{-1} \sum_{i=1}^{n} Var(h_{n,i}^{(1)}),$$

where

$$h_i^{(1)} = \left[\mathbf{1}_{(H_i \in E_k)} - \left(\frac{1}{n} \sum_{j=1}^{n} (\mathbf{1}_{(H_j \in E_k)} X_j'(1-\pi_j(\omega))\pi_j(\omega)) \right) \Lambda_n^{-1}(\omega) X_i \right]_{1\le k \le K} (y_i - \pi_i(\omega)).$$

Suppose $\Psi_n^{(1)}$ converges to a limiting value $\Psi^{(1)}$. If there is no random effect in model (6)(i.e. $m = 1$ in Theorem 3), the asymptotic distribution of the test statistic (18) is

$$\hat{\chi}^2 = \tfrac{1}{n} \textstyle\sum_{j=1}^{K} (M_j - e_j(\hat{\omega}))^2 \xrightarrow{d} \textstyle\sum_{k=1}^{K} \lambda_k \mathbb{Z}_k^2, \tag{22}$$

where $\lambda_1, \cdots, \lambda_K$ are the eigenvalues of $\Psi^{(1)}$.

The proof is omitted because it is similar to that in Jiang (2001a). The only difference is that the asymptotics here will be proved by using a martingale central limit theorem. Because of the use of partial likelihood, this result is more general than Jiang (2001a).

Lemma D.2. *Using the notation in Section 3.3, for any $\mu \in R \setminus \{0\}$,*

$$\mu J^{-1/2} I^N \sqrt{m}(\hat{\sigma}_b^2 - \sigma_b^2) = [(\mathbf{Y} - \mathbf{X}\omega)' B_N^* (\mathbf{Y} - \mathbf{X}\omega) - E((\mathbf{Y} - \mathbf{X}\omega)' B_N^* (\mathbf{Y} - \mathbf{X}\omega))], \quad (23)$$

where $B_N^* = J^{-1/2} \mu \mathbf{W}^{1/2} V^* \mathbf{W}^{1/2} Z Z' \mathbf{W}^{1/2} V^* \mathbf{W}^{1/2} / \sqrt{m}$.

Proof: Following the same argument as in Theorem 2, consider the LMM with GLM weights, we can obtain this result by modifying the first formula on page 276 of Jiang (1996) into

$$\mu J^{-1/2} I^N \sqrt{m}(\hat{\sigma}_b^2 - \sigma_b^2) = \varpi' B_N \varpi - \mathrm{E}(\varpi' B_N \varpi),$$

where $B_N = J^{-1/2} \mu V_1(\sigma_b)/\sqrt{m}$, and $V_1(\sigma_b)$ is defined in Theorem 2. Lemma D.2 follows because $\varpi' g(\sigma_b) \mathbf{W}^{-1/2} = (\mathbf{Y} - \mathbf{X}\omega)'$.

Lemma D.3. *Denote $\theta = (\omega', \sigma_b)$, $\xi_k = M_k - e_k(\hat{\theta})$, and $\xi = (\xi_k)_{1 \le k \le K}$. Let T be an orthogonal matrix such that $T' \Psi_N T = diag(\lambda_{N,1}, \cdots, \lambda_{N,K})$, where $\lambda_{N,1}, \cdots, \lambda_{N,K}$ are the eigenvalues of Ψ_N. For any $a \in R^K$,*

$$a'((N)^{-1/2} T' \xi) = \sum_{i=1}^{m} \sum_{t=1}^{n} \Upsilon_{it} + o_p(1), \quad (24)$$

where $Ta = a_T = (a_{T,1}, \cdots, a_{T,K})'$, $G_{it}^* = (N)^{-1/2} a_T' G_{it}$, $\Delta' = (\Delta_{it}) = (\mathbf{Y} - \mathbf{X}\omega)' \mathbf{V}^{-1/2}$, $Var(\Delta_{it}) = 1$, $C^* = (N)^{-1/2} a_T' \left[\sum_i \sum_t (\mathbf{1}_{(H_{it} \in E_k)} \frac{\partial}{\partial \sigma_b^2} \pi_{it}(\theta)) \right]_{1 \le k \le K} \frac{(I^N)^{-1}}{\sqrt{m}} C = (N)^{-1/2} a_T' \Phi C$. and

$$\Upsilon_{it} = G_{it}^* (y_{it} - \pi_{it}(\theta)) - (C^* \mathbf{V})_{it} \Delta_{it}^2 - (\sum_{(it)' \neq it} (C^* \mathbf{V})_{it,(it)'} \Delta_{(it)'}) \Delta_{it} + (C^* \mathbf{V})_{it}.$$

(Notice that C^* is a $N \times N$ matrix with $C^*_{i't',it}$ indicating the element in C^* with $[(i'-1)n + t']$-th column and $(i-1)n + t$-th row, and by the definition before Theorem 3, $C^*_{it,it} = C^*_{it}$.)

Proof: For $1 \le k \le K$, $\xi_k = M_k - e_k(\theta) - (e_k(\hat{\theta}) - e_k(\theta))$. By definition,

$$a'((N)^{-1/2} T' \xi) = (N)^{-1/2} \textstyle\sum_{k=1}^{K} a_{T,k} (M_k - e_k(\theta)) - (N)^{-1/2} \sum_{k=1}^{K} a_{T,k} (e_k(\hat{\theta}) - e_k(\theta)).$$

By Taylor expansion, the second term on the right hand side can be approximated by

$$(N)^{-1/2} \sum_{k=1}^{K} a_{T,k} \left[\left(\sum_{i=1}^{m} \sum_{t=1}^{n} \mathbf{1}_{(H_{it} \in E_k)} \frac{\partial}{\partial \omega'} \pi_{it}(\theta) \right) (\hat{\omega} - \omega) + \left(\sum_{i=1}^{m} \sum_{t=1}^{n} \mathbf{1}_{(H_{it} \in E_k)} \frac{\partial}{\partial \sigma_b^2} \pi_{it}(\theta) \right) (\hat{\sigma}_b^2 - \sigma_b^2) \right].$$

The result follows by Theorem 1, Assumption A5, and Lemma D.2.

Lemma D.4. *Under A6-A8, as $N \to \infty$,*

$$\textstyle\sum_{i=1}^m \sum_{t=1}^n \Upsilon_{it} \xrightarrow{d} \mathcal{N}(0, a'\Gamma a). \tag{25}$$

Proof: First, derive the asymptotic distribution of $\sum_{i=1}^m \sum_{t=1}^n \Upsilon_{it}/\sigma_N$, where σ_N is defined in A6. Decompose $\Upsilon_{it}/\sigma_N = \Upsilon_{it}^{(1)} + \Upsilon_{it}^{(2)}$, where

$$\Upsilon_{it}^{(1)} = \frac{1}{\sigma_N}\Bigg(G_{it}^* u_{it}^* + (C^*\mathbf{V})_{it} U_{it} + \sum_{(it)' \neq it} \Big((C^*\mathbf{V})_{it,(it)'} u_{(it)'}\Big) u_{it}\Bigg),$$

$$\Upsilon_{it}^{(2)} = \frac{1}{\sigma_N}\Bigg(G_{it}^* v_{it}^* + (C^*\mathbf{V})_{it} V_{it} + \Big(\sum_{(it)' \neq it} (C^*\mathbf{V})_{it,(it)'} v_{(it)'}\Big) u_{it} + \Big(\sum_{(it)' \neq it} \Delta_{(it)'} (C^*\mathbf{V})_{it,(it)'}\Big) v_{it}\Bigg).$$

Define

$$\begin{aligned}
U_{it} &= (\Delta_{it}^2 - 1)\mathbf{1}_{(|\Delta_{it}| < L_{it})} - \mathrm{E}(\Delta_{it}^2 - 1)\mathbf{1}_{(|\Delta_{it}| < L_{it})}, \\
V_{it} &= (\Delta_{it}^2 - 1) - U_{it}, \\
u_{it} &= \Delta_{it}\mathbf{1}_{(|\Delta_{it}| < L_{it})} - \mathrm{E}\Delta_{it}\mathbf{1}_{(|\Delta_{it}| < L_{it})}, \\
v_{it} &= \Delta_{it} - u_{it}, \\
u_{it}^* &= (y_{it} - \pi_{it}(\theta))\mathbf{1}_{(|y_{it} - \pi_{it}(\theta)| < L_{it})} - \mathrm{E}(y_{it} - \pi_{it}(\theta))\mathbf{1}_{(|y_{it} - \pi_{it}(\theta)| < L_{it})}, \\
v_{it}^* &= (y_{it} - \pi_{it}(\theta)) - u_{it}^*.
\end{aligned}$$

By assumption A6, we can easily show that $\sum_{i=1}^m \sum_{t=1}^n \Upsilon_{it}^{(2)}$ converges to 0 in L_2. Next, consider $\Upsilon_{it}^{(1)}$ which is an array of martingale differences by following the same argument in Theorem 5.2 of Jiang (1996). Based on assumption A7 and Rosenthal's inequality (Hall and Heyde, 1980), $\max_{it} \mid \Upsilon_{it}^{(1)} \mid$ is bounded in L_2 and converges to 0 in probability.

By Theorem 3.2 of Hall and Heyde (1980), to prove Lemma D.4, one has to show that $\sum_i \sum_t (\Upsilon_{it}^{(1)})^2$ converges to $a'\Gamma a$ in probabilty. First, it can be decomposed as

$$\sum_{i=1}^m \sum_{t=1}^n (\Upsilon_{it}^{(1)})^2 = \sum_{j=1}^3 t_i + \sum_{j=1}^3 s_i,$$

where

$$
\begin{aligned}
t_1 &= \sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n [((C^*\mathbf{V})_{it} U_{it} + G^*_{it} u^*_{it})^2 - \mathrm{E}((C^*\mathbf{V})_{it} U_{it} + G^*_{it} u^*_{it})^2], \\
t_2 &= 2\sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n (\sum_{(it)' \neq it} (C^*\mathbf{V})_{it,(it)'} u_{(it)'}) [(C^*\mathbf{V})_{it} (U_{it} u_{it} - \mathrm{E}(U_{it} u_{it})) \\
&\quad + (G^*_{it} u^*_{it}) u_{it} - \mathrm{E}((G^*_{it} u^*_{it}) u_{it})], \\
t_3 &= 2\sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n \Big((\sum_{(it)' \neq it} (C^*\mathbf{V})_{it,(it)'} u_{(it)'})^2 (u_{it}^2 - \mathrm{E}u_{it}^2) \Big), \\
s_1 &= \sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n \mathrm{E}((C^*\mathbf{V})_{it} U_{it} + G^*_{it} u^*_{it})^2, \\
s_2 &= 2\sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n (\sum_{(it)' \neq it} (C^*\mathbf{V})_{it,(it)'} u_{(it)'}) [\mathrm{E}((C^*\mathbf{V})_{it} U_{it} u_{it}) + \mathrm{E}((G^*_{it} u^*_{it}) u_{it})], \\
s_3 &= 2\sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n \Big((\sum_{(it)' \neq it} (C^*\mathbf{V})_{it,(it)'} u_{(it)'})^2 \mathrm{E}u_{it}^2 \Big).
\end{aligned}
$$

By assumption A7 and Rosenthal's inequality, we can show that $t_i \to 0$ in L_2 for $i = 1, 2, 3$, which is similar to the result in Theorem 5.2 of Jiang (1996). By assumption A7,

$$
s_1 = \sigma_N^{-2} \sum_{i=1}^m \sum_{t=1}^n \mathrm{Var}\Big((C^*\mathbf{V})_{it} \Delta_{it}^2 + G^*_{it}(y_{it} - \pi_{it}(\theta)) \Big) + o_p(1). \tag{26}
$$

Analogous to Theorem 5.2 of Jiang (1996), by assumptions A6-A8, we have

$$
Es_2^2 \leq c \left[\frac{\lambda \max(\xi'\xi)}{\sigma_N^2} \right]^{1/2} \to 0, \tag{27}
$$

where c stands for a constant and

$$
s_3 = 2\sigma_N^{-2} \sum_{(it)' \neq it} (C^*\mathbf{V})^2_{it,(it)'} + o_p(1). \tag{28}
$$

By (26) and (28), $\sum_i \sum_t (\Upsilon^{(1)}_{it})^2 = 1 + o_p(1)$. Because $\sigma_N^2 = a'T'\Psi_N Ta$, it converges to $a'\Gamma a$ in probability. Consequently, (25) follows.

Proof of Theorem 3: From Lemmas D.2 to D.4, we have, for any a,

$$
a'(N^{-1/2} T'\xi) \xrightarrow{d} (a'\Gamma a)^{1/2} \mathbb{Z},
$$

where $\mathbb{Z} \sim \mathcal{N}(0, \mathbf{1})$, form which, $N^{-1/2}T'\xi \xrightarrow{d} \mathcal{N}(0, \Gamma)$ follows.

References

Albert, A. and Anderson, J. A. (1984), "On the Existence of Maximum Likelihood Estimates in Logistic Regression Models," *Biometrika* **71**, 1-10.

Barndorff-Nielsen, O. E. and Cox, D. R. (1989), *Asymptotic Techniques for Use in Statistics,* London: Chapman and Hall.

Benjamin, M. A., Robert, A. R., and Stasinopoulos, D. M. (2003), "Generalized Autoregressive Moving Average Models," *Journal of American Statistical Association* **98**, 214-223.

Breslow, N. E. and Clayton, D. G. (1993), "Approximate Inference to Generalized Linear Mixed Models," *Journal of American Statistical Association* **88**, 9-25.

Chernoff, H. and Lehmann, E. L. (1954), "The Use of Maximum Likelihood Estimations in χ^2 Tests for Goodness of fit," *The Annals of Mathematical Statistics* **25**, 579-586.

Chesla, S. E., Selvaraj, P., and Zhu, C. (1998), "Measuring Two-Dimensional Receptor-Ligand Binding Kinetics by Micropipette," *Biophysical Journal* **75**, 1553-1572.

Cox, D. R. (1975), "Partial Likelihood," *Biometrika* **62**, 69-76.

Diggle, P., Heagerty, P., Liang, K-Y., and Zeger, S. (2002), *Analysis of Longitudinal Data,* 2nd Ed. Oxford: Oxford University Press.

Fokianos, K. and Kedem, B. (1998), "Prediction and Classification of Non-stationary Categorical Time Series," *Journal of Multivariate Analysis* **67**, 277-296.

—— (2004), "Partial Likelihood Inference for Time Series Following Generalized Linear Models," *Journal of Time Series Analysis* **25**, 173-197.

Hall, P. and Heyde, C. C. (1980), *Martingale Limit Theory and its Application,* New York: Academic Press.

Harville, D. A. (1977), “Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems,” *Journal of American Statistical Association* **72**, 320-340.

Jiang, J. (1996), “REML Estimation: Asymptotic Behavior and Related Topics,” *The Annals of Statistics* **24**, 255-286.

—— (2001a), “A Nonstandard Chi-square Test with Application to Generalized Linear Model Diagnostics,” *Statistics and Probability Letters* **53**, 101-109.

—— (2001b), “Goodness-of-fit Tests for Mixed Model Diagnostics,” *The Annals of Statistics* **29**, 1137-1164.

Kaufmann, H. (1987), “Regression Models for Nonstationary Categorical Time Series: Asymptotic Estimation Theory,” *The Annals of Statistics* **15**, 79-98.

Kedem, B. and Fokianos, K. (2002), *Regression Models for Time Series Analysis,* New York: Wiley.

Li, W, K. (1994), “Time Series Models Based on Generalized Linear Models: Some Further Results,” *Biometrics* **50**, 506-511.

Lin, X. and Breslow, N. E. (1996), “Bias Correction in Generalized Linear Mixed Models With Multiple Components of Dispersion,” *Journal of American Statistical Association* **91**, 1007-1016.

Marshall, B. T., Long, M., Piper, J. W., Yago, T., McEver, R. P., and Zhu, C. (2003), “Direct Observation of Catch Bonds Involving Cell-adhesion Molecules,” *Nature* **423** 190-193.

Mehta, A. D., Rief, M., Spudich, J. A., Smith, D. A., and Simmons, R. M. (1999), “Single-molecule Biomechanics with Optical Methods.” *Science* **283**(5408) 1689-1695.

McCulloch, D. E. (1994), "Maximum Likelihood Variance Components Estimation for Binary Data," *Journal of American Statistical Association* **89**, 330-335.

Patterson, H. D. and Thompson, R. (1971), "Recovery of Interblock Information When Block Sizes Are Unequal," *Biometrika* **58**, 545-554.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.

Senatov, V. V. (1980), "Uniform Estimates of the Rate of Convergence in the Multi-dimensional Central Limit Theorem," *Theory of Probability and its Applications* **25**, 745-759.

Silvapulle, M. (1981), " On the Existence of Maximum Likelihood Estimates for the Binomial Response Models," *Journal of the Royal Statistical Society,* Ser. B, **43**, 310-313.

Slud, E. and Kedem, B. (1994), "Partial Likelihood Analysis of Logistic Regression and Autoregression," *Statistica Sinica* **4**, 89-106.

Wedderburn, R. W. M. (1976), "On the Existence and Uniqueness of the Maximum Likelihood Estimates," *Biometrika* **63**, 27-32.

Wong, W. H. (1986), "Theory of Partial Likelihood," *The Annals of Statistics* **14**, 88-123.

Zarnitsyna, V. I., Huang, J., Zhang, F., Chien, Y-H., Leckband, D., and Zhu, C. (2007), "Memory in Receptor-ligand Mediated Cell Adhesion," *Proceedings of the National Academy of Sciences USA.*, to appear.

Zeger, S. L. and Qaqish, B. (1988), "Markov Models for Time Series: A Quasi-Likelihood Approach," *Biometrics* **44**, 1019-1032.

Zhu, C., Long, M., Chesla, S. E., and Bongrand, P. (2002), "Measuring Receptor/Ligand Interaction at the Single-bond Level: Experimental and Interpretative Issues," *Annals of Biomedical Engineering* **30**, 305-314.