Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No. 181
DOI: 10.11159/icsta25.181

Sequential Change-point Detection for Binomial Time Series with Exogenous Variables

Yajun Liu[1], Beth Andrews[2]
[1]Amazon.com
410 Terry Avenue North, Seattle, WA
lyajun@amazon.com
[2]Department of Statistics and Data Science, Northwestern University
2006 Sheridan Road, Evanston, IL
bandrews@northwestern.edu

Abstract **-** A binomial time series describes binary behaviors of individuals within a group, which depend on group behaviors in the past. Binomial time series data is widely applied in fields such as infection tracking and behavior analysis. In this paper, we introduce a generalized Binomial AR(p) model with exogenous variables based on Generalized Linear Model (GLM), prove the statistical properties of the model when $p = 1$, and provide a parameter estimation method. Then, we propose a sequential change-point detection method for the generalized Binomial AR(1) model, facilitate real-time data monitoring and triggering alarms when a change point is detected. We apply the generalized Binomial AR(1) model to weekly pneumonia & influenza mortality data and successfully identify change points related to the COVID-19 outbreak using the proposed method.

***Keywords*:** time-series, binomial data, sequential change-point detection, generalized linear model, COVID-19 analysis

1. Introduction

In time series analysis, the accuracy of forecasts relies on the validity of the fitted model as new data is observed. Therefore, it is crucial to verify if the model is still appropriate for new data points. We call t a change point if any attributes of the model change at time t. Sequential change-point detection is used when monitoring a time series in real-time to identify potential change points as they happen. It is important in practice, such as in data monitoring and risk control, as the earlier a change point is detected, the faster we can take action to manage the associated risks and adjust the model accordingly. The paper written by Chu et al., [1], is the origin of sequential change detection. Since then, sequential detection for various data structures follows, as seen in subsequent studies such as [2], [3], [4], and [5].

Binomial time series models individual behavior within a group, reflected through binary actions (e.g., support/not support, increase/not increase) that depend on past observations. This type of time series has broad applications in fields such as customer engagement analysis, epidemiology, and clinical trials, but it is underdeveloped. A notable early contribution is the binomial thinning process, introduced in [6], which defines a Binomial AR(1) model. [7] extends the Binomial AR(1) model to Binomial AR(p). [8] makes a bivariate extension of the Binomial AR(1) model. However, binomial time series models based on generalized linear model (GLM) and change-point detection for the models haven't been developed yet. In this paper, we develop a binomial model based on generalized linear model (GLM) that incorporates exogenous information and past observations, offering greater flexibility and capability for capturing dynamic dependencies. We then develop the sequential change-point detection for the model, which is based on the monitoring scheme proposed in [9].

In next section, we will introduce the generalized Binomial AR(p) model, prove the Markov chain properties of time series following the model when $p = 1$. We include a parameter estimation method in this section, along with proofs for the consistency and asymptotic normality of the estimated parameters. We then demonstrate the sequential change-point detection method for the generalized Binomial AR(1) model. In the subsequent section, we apply the model to weekly pneumonia and influenza mortality data, successfully identifying change points associated with the COVID-19 outbreak using the proposed approach.

2. Binomial autoregressive model based on GLM

We define the binomial autoregressive model based on GLM as follows

$$X_t | X_{t-1}, X_{t-2}, \ldots, \mathbb{Z}_{t-1}, \mathbb{Z}_{t-2}, \ldots \sim Bin(n, \pi_t(\beta)), \text{ with } g(n\pi_t(\beta)) := g(\mu_t) = \beta^T \mathbb{Z}_{t-1}, \quad (1)$$

in which n is the group size, $g(x) = \log\left(\frac{x}{n-x}\right)$, and $\mathbb{Z}_{t-1}$ is a regressor including autoregressive variables, e.g. X_{t-1}, X_{t-2} or other forms of past information, exogenous variables that affect the time-dependent probability $\pi_t(\beta)$, or a mixture of both. We will refer to the model as the generalized Binomial AR model afterwards.

In this paper, we focus on the properties and sequential change-point detection for the generalized Binomial AR(1) model, incorporating the exogenous variable at t. That is, $\mathbb{Z}_{t-1} = (1, X_{t-1}, \boldsymbol{W}_t)^T$, where the exogenous variable $\{\boldsymbol{W}_t\}$ is i.i.d. with a known PDF, and $\boldsymbol{W}_t \in [a,b]^l$ with known a and b.

2.1. Properties of the generalized Binomial AR(1) model and the estimated parameter

In the following of this section, we first prove the statistical properties of the generalized Binomial AR(1) model. Under certain regularity conditions, we provide a parameter estimation method, and show the consistency and the asymptotic normality of the estimated parameter, β.

Lemma 2.1. The binomial time series $\{X_t\}$ defined as (1) is a ψ-irreducible and aperiodic Markov chain.

In [10], the authors prove that the Binary AR(1) model is ψ-irreducible. We can prove the ψ-irreducibility of the generalized Binomial AR(1) following a similar approach. The process is aperiodic since all one-step transition probabilities are strictly positive.

Theorem 2.1. $\{X_t\}$ is a strictly stationary and geometrically ergodic process.

Define the drift operator as $\Delta V(x) = \int P(x, dy)V(y) - V(x)$, where P is the transition probability, we first show that the drift condition

$$\Delta V(x) \leq -\beta V(x) + b \mathbb{I}_\Omega(x), x \in \Omega, \quad (2)$$

holds when $V(x) = x + 1$, $b = n + 1$, and $\beta = 1$. According to [11], if a Markov chain is ψ-irreducible, aperiodic and satisfies the drift condition (2), there exists a unique invariant measure μ. The strict stationarity of $\{X_t\}$ can be proven directly from the definition of $\{X_t\}$. Furthermore, given that the chain is ψ-irreducible and aperiodic, the geometric ergodicity of the chain follows immediately after the drift condition (2) is established.

For time-dependent cases, parameter estimation using maximum likelihood estimation (MLE) can be complicated. Thanks to the statistical properties of the model, instead of using the likelihood function, we can take advantage of partial likelihood, [12], for parameter estimation.

Definition 1. Denote the conditional distribution function following the generalized Binomial AR(1) model as $f_\beta(X_t | X_{t-1}, W_t)$. The partial likelihood is defined as

$$PL(\beta) = \prod_{t=1}^{m} f_\beta(X_t | X_{t-1}, \boldsymbol{W}_t) \propto \prod_{t=1}^{m} \pi_t(\beta)^{X_t} (1 - \pi_t(\beta))^{n - X_t}.$$

The partial score vector is defined as

$$PSV_m(\beta) = \nabla \log PL(\beta) = \sum_{t=1}^{m} \mathbb{Z}_{t-1}(X_t - n\pi_t(\beta)) \triangleq \sum_{t=1}^{m} G(X_t, \beta). \quad (3)$$

From [13], we can conclude that $G(X_t, \beta)$ is stationary and ergodic. Based on the strict stationarity of $\{X_t\}$, it is easy to prove that $G(X_t, \beta)$ is strictly stationary.

The maximum partial likelihood estimation (MPLE) β can be obtained by solving the equation

$$PSV_m(\hat{\beta}) = 0. \quad (4)$$

Before proving the properties of $\hat{\beta}$, we first propose the regularity condition for β.

Assumption 1. The parameter space of β, denoted as $\mathcal{B}$, is convex and compact, and the true parameter β_0 is in the interior of $\mathcal{B}$.

Theorem 2.2. Under Assumption 1,

(a) The MPLE $\hat{\beta}$ is almost surely unique for all sufficiently large m.

(b) $\hat{\beta} \xrightarrow{P} \beta_0$ as $m \to \infty$.

(c) $\sqrt{m}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \left(-\mathbb{E}\nabla G(X_1, \beta_0)\right)^{-1})$ as $m \to \infty$.

In [14], the authors establish a set of regularity conditions under which the existence and uniqueness, and the consistency of the PMLE $\hat{\beta}$ hold and $\sqrt{m}(\hat{\beta} - \beta_0) \xrightarrow{D} N(0, \left(-\mathbb{E}\nabla G(X_1, \beta_0)\right)^{-1})$ holds, as $m \to \infty$. Under Definition 1 and Assumption 1, it can be easily verified that the generalized Binomial AR(1) satisfies the regularity conditions, and the properties follow.

In practice, due to the ergodicity of $\{X_t\}$ and the consistency of $\hat{\beta}$, we are able to estimate the variance-covariance matrix $\left(-\mathbb{E}\nabla G(X_1, \beta_0)\right)^{-1}$ by $[-\sum_{t=1}^{m} \nabla G(X_t, \hat{\beta}) / m]^{-1}$ when m is large enough. Based on (3), $\nabla G(X_t, \beta) = -n \mathbb{Z}_{t-1}\mathbb{Z}_{t-1}^T \frac{\exp\left(-\beta^T\mathbb{Z}_{t-1}\right)}{\left(1 + \exp\left(-\beta^T\mathbb{Z}_{t-1}\right)\right)^2}$.

3. Sequential change-point detection for the generalized Binomial AR(1) model

In this section, we will introduce the close-end monitoring scheme for the generalized Binomial AR(1) model. Given that there are m initial observations and monitoring starts from $m + 1$, close-end monitoring means that the monitoring process will be terminated if there is no change point detected within the Nm new observations after the initial observations, where N is a fixed and predetermined constant. Denote the true parameter of the model for the first m observations is β_0, the hypotheses are

$H_0: \beta_t = \beta_0,$ for all $t \in \{m+1, m+2, ..., m+Nm\}$,

H_a: $\exists$ k^* such that $\beta_t = \beta_0$ for $m+1 \le m+k^* \le m+Nm$, and $\beta_t \ne \beta_0$ for $m+k^* \le t \le m+Nm$.

To ensure the validity of $\hat{\beta}$ based on the first m observations, the time series should satisfy the non-contamination assumptions.

Assumption 2. (a) There is no change point in the exogenous variable $\boldsymbol{W}_t$.

(b) There is no change point in the first m observations X_t.

Under Assumption 2 and (4), the MPLE $\hat{\beta}$ satisfies $\sum_{t=1}^{m} G(X_t, \hat{\beta}) = 0$. Denote $\sum_{t=m+1}^{m+k} G(X_t, \hat{\beta})$ as $S(m, k)$. Under H_0, due to the consistency of $\hat{\beta}$, $S(m, k)$ is expected to be reasonably close to 0; otherwise, suppose a change point appears at k^* and the process never comes back to the original one, $S(m,k)$, would be gradually far from 0 as more and more data arise after X_{m+k^*}. Following this idea, in this paper, we extend the monitoring scheme proposed in [9] to generalized Binomial AR(1) case that H_0 is rejected with a significance level α when

$$\omega^2(m,k) S(m,k)^T A S(m,k) \ge c(\alpha), \quad (5)$$

where $\omega(m,k) = m^{-\frac{1}{2}}\rho\left(\frac{k}{m}\right)$ is the weight function adjusting the sensitivity of the method, and A is a customized symmetric positive definite matrix adjusting the magnitude of the statistic. The regularity conditions for $\omega(m,k)$ are included in [9]. $c(\alpha)$ can be estimated based on the asymptotic distribution of the statistics in (5). To find the asymptotic distribution of (5), we first prove the following lemma and proposition.

Lemma 3.1. Under Assumption 1-2 and under H_0,

$$sup_{k\geq 1}\omega(m,k)//\sum_{t=m+1}^{m+k} G(X_t,\beta) - (\sum_{t=m+1}^{m+k} G(X_t,\beta_0) - \frac{k}{m}\sum_{t=1}^{m} G(X_t,\beta_0)//= o_p(1).$$

Proof. According to the regularity conditions for $\omega(m,k)$in [9], $\lim_{t\to 0} t^{\gamma}\rho(t) < \infty$ implies $\rho(t) = \rho(\frac{k}{m}) = O_p((\frac{k}{m})^{-\gamma})$ as $t\to 0$. Then, $\omega(m,k) = m^{-\frac{1}{2}}\rho(\frac{k}{m}) = O_p(k^{-\gamma}m^{\gamma-\frac{1}{2}})$ as $\frac{k}{m}\to 0$. Similarly, $\omega(m,k) = O_p(m^{\frac{1}{2}}k^{-1})$ as $\frac{k}{m}\to\infty$. There always exist a constant $C>0$ so that

$$\omega(m,k) \leq \begin{cases} Ck^{-\gamma}m^{\gamma-\frac{1}{2}}, & \text{when } k \leq m, \\ Cm^{\frac{1}{2}}k^{-1}, & \text{when } k > m. \end{cases}$$

From (4), we know that $\sum_{t=1}^{m} G(X_t,\beta) = 0$, then the norm component of the lemma can be written as

$$(\sum_{t=m+1}^{m+k} G(X_t,\beta) - G(X_t,\beta_0)) - \frac{k}{m}(\sum_{t=1}^{m} (G(X_t,\beta) - G(X_t,\beta_0)) := D_1(m,k) - D_2(m,k).$$

For the j^{th} element in $G(X_t,\cdot)$, the following equation holds

$$\sum_{t=m+1}^{m+k} G_j(X_t,\beta) - G_j(X_t,\beta_0)) = \sum_{t=m+1}^{m+k} (\nabla G_j(X_t,\beta_0)^T(\beta-\beta_0) + \frac{1}{2}(\beta-\beta_0)^T\nabla^2 G_j(X_t,\beta^*)(\beta-\beta_0)),$$

where β^* is between β_0 and β. Based on Definition 1, we can conclude that $\nabla^2 G_j(X_t,\beta^*)$ is bounded. The following equation holds in a convex neighborhood of β_0, $U_{\beta_0} \in \mathcal{B}$,

$$sup_{k\geq 1} sup_{\beta^*\in U_{\beta_0}} \frac{1}{k}\sum_{t=m+1}^{m+k} \| \nabla^2 G_j(X_t,\beta^*) \|_\infty = O_p(1). \quad (6)$$

Furthermore, by Birkhoff's Ergodic Theorem, the following equation holds

$$\frac{1}{k}\sum_{t=m+1}^{m+k} \nabla G(X_t,\beta_0) - \mathbb{E}\nabla G(X_1,\beta_0) = o_p(1). \quad (7)$$

Combining (6), (7) and Theorem 2.2, we can show that

$$\begin{aligned} & sup_{k\geq 1}\omega(m,k)\left| D_1(m,k) - k\mathbb{E}\nabla G(X_1,\beta_0)^T(\beta-\beta_0)\right| \\ &= sup_{k\geq 1}\omega(m,k)/\sum_{t=m+1}^{m+k}(\nabla G_j(X_t,\beta_0) - \mathbb{E}\nabla G(X_1,\beta_0))^T(\beta-\beta_0) + \frac{1}{2}(\beta-\beta_0)^T\nabla^2 G_j(X_t,\beta^*)(\beta-\beta_0)/ \\ &= sup_{k\geq 1}\omega(m,k)\left(o_p(\tfrac{k}{\sqrt{m}}) + O_p(\tfrac{k}{m})\right) \\ &= sup_{k\leq\sqrt{m}} O_p(k^{1-\gamma}m^{\gamma-\frac{3}{2}}) + sup_{\sqrt{m}<k\leq m} o_p(\tfrac{k}{m})^{1-\gamma} + sup_{k>m} O_p(\tfrac{1}{\sqrt{m}}) = o_p(1). \end{aligned} \quad (8)$$

Analogously,

$$sup_{k\geq 1}\omega(m,k)\left| D_2(m,k) - k\mathbb{E}\nabla G(X_1,\beta_0)^T(\beta-\beta_0)\right| = o_p(1). \quad (9)$$

Combining (8) and (9) , $sup_{k\geq 1}\omega(m,k)//\sum_{t=m+1}^{m+k} G(X_t,\beta) - (\sum_{t=m+1}^{m+k} G(X_t,\beta_0) - \frac{k}{m}\sum_{t=1}^{m} G(X_t,\beta_0)//= o_p(1).$

Proposition 3.1. For any fixed $N > 0$, under Assumption 1-2 and under H_0 , as $m \to \infty$, the sequence of functions $\frac{1}{\sqrt{m}} \sum_{t=1}^{\lfloor ms \rfloor} G(X_t, \beta_0)$, for $0 < s \le N$, converges weakly to a Wiener process $W(s)$ with covariance matrix $\Sigma = \mathbb{E}[G(X_1, \beta_0)G(X_1, \beta_0)^T] < \infty$.

Proof. Based on the functional central limit theorem in [15], if the process $G(X_t, \beta_0)$ is
(1). bounded,
(2). strictly stationary,
(3). $\mathbb{E}G(X_t, \beta_0) = 0$,
(4). a strong mixing process,
and satisfies the two inequalities $\sum_{n=1}^{\infty} \alpha_{G(n)} < \infty$, and $\alpha_{G(n)} \le \frac{M_G}{n \log n}$, where $\alpha_G(n)$ is the strong mixing coefficient for $\{G(X_t, \beta_0)\}$, then $\{G(X_t, \beta_0)\}$ converges weakly to a Wiener process with covariance matrix

$$\Sigma = \mathbb{E}\left[G(X_1, \beta_0)^T G(X_1, \beta_0)\right] + 2\sum_{t=2}^{\infty} \mathbb{E}\left[G(X_1, \beta_0)^T G(X_t, \beta_0)\right] < \infty .$$

Conditions (1)-(3) can be easily verified by the definition of $G(X_t, \beta_0)$ ((3)) and Theorem 2.1. Then we show the process $G(X_t, \beta_0)$ is strong mixing. In Theorem 3.1, we have proved that $\{X_t\}$ is strictly stationary and geometrically ergodic. Given that $\{W_t\}$ is an i.i.d. sequence, it is naturally strictly stationary and geometrically ergodic. So $\mathbb{Z}_{t-1} = (1, X_{t-1}, W_t)$ is strictly stationary and geometrically ergodic. By [16], $\mathbb{Z}_{t-1}$ has $\beta(n) = O_p(e^{-\theta n}), \theta > 0$. The definition of the measure β can be found in [16]. What's worth pointing out is that according to [16],

$$2\alpha(n) \le \beta(n),$$

where $\alpha(n)$ is the strong mixing coefficient for $\{\mathbb{Z}_t\}$. That is, for $\{\mathbb{Z}_t\}$, there exists a $\theta > 0$ so that the strong mixing coefficient

$$\alpha(n) = sup_{j \in \mathbb{N}^+} \alpha(\mathcal{F}_0^j, \mathcal{F}_{j+n}^{\infty}) = sup_{j \in \mathbb{N}^+} |P(A \cap B) - P(A)P(B)| = O_p(e^{-\theta n}), \text{ for } \forall A \in \mathcal{F}_0^j, B \in \mathcal{F}_{j+n}^{\infty},$$

where $\mathcal{F}_J^L = \sigma(\mathbb{Z}_t, J \le t \le L)$. Then it is easy to conclude that $\alpha_G(n) = O_p(e^{-\theta n})$. As $n \to \infty$, $\alpha_G(n) \to 0$, which implies that $G(X_t, \beta_0)$ is a strong mixing process. Additionally, we can find a constant $M > 0$ such that

$$\alpha_G(n) \le Me^{-\theta n}, \forall n \in \mathbb{N}^+ ,$$

then the inequality

$$\sum_{n=1}^{\infty} \alpha_G(n) \le \sum_{n=1}^{\infty} Me^{-\theta n} < \infty \qquad (10)$$

holds.

The inequality

$$\alpha_{G(n)} \le \frac{M_G}{n \log n} \quad (11)$$

can be proven by defining $M_G = \max_{n \in \mathbb{N}^+} M e^{-\theta n} n \log n$, which is guaranteed to exist. According to Theorem 1 in [15], (10) and (11) imply that

$$\left\{\frac{1}{\sqrt{m}}\sum_{t=1}^{\lfloor ms\rfloor} G\left(X_t,\beta_0\right)\right\} \xrightarrow{\mathcal{D}} W(s), \text{ as } m\to\infty,$$

where $W(s)$ is a Wiener process with covariance

$\Sigma = \mathbb{E}\left[G\left(X_1, \beta_0\right)^T G\left(X_1, \beta_0\right)\right] + 2\sum_{t=2}^{\infty}\mathbb{E}\left[G\left(X_1, \beta_0\right)^T G\left(X_t, \beta_0\right)\right] < \infty$.

For any $t > 1$,

$$\begin{aligned}\mathbb{E}\left(G\left(X_1, \beta_0\right)^T G\left(X_t,\beta_0\right)\right) &= \mathbb{E}(\mathbb{E}(G\left(X_1, \beta_0\right)^T G\left(X_t,\beta_0\right)/\mathcal{C}_t))\\ &= \mathbb{E}(G\left(X_1,\beta_0\right)\mathbb{Z}_{t-1}\mathbb{E}((X_t - n\pi_t(\beta_0))/\mathcal{C}_t))\\ &= 0,\end{aligned}$$

where $\mathcal{C}_t = \sigma(\boldsymbol{W}_t, X_{t-1}, \boldsymbol{W}_{t-1}, X_{t-2},\ldots,\boldsymbol{W}_0, X_1)$.

Therefore, the covariance matrix can be simplified as

$$\Sigma = \mathbb{E}[G\left(X_1, \beta_0\right)G\left(X_1, \beta_0\right)^T] < \infty.$$

Theorem 3.1. Under H_0, for any fixed symmetric positive definite matrix A,

$$\begin{aligned}&\lim_{m\to\infty} P\left(sup_{1\le k\le Nm}\,\omega\left(m,k\right)^2 S\left(m,k\right)^T A S\left(m,k\right) \le c\right)\\ &= P(sup_{0<s\le N}\rho^2\left(s\right)\left(W_1\left(s\right) - sW_2\left(1\right)\right)^T A(W_1\left(s\right) - sW_2\left(1\right)) \le c,\end{aligned}$$

where $W_1(.)$ and $W_2(.)$ are independent Wiener processes with the same covariance matrix $\Sigma = \mathbb{E}[G\left(X_1, \beta_0\right)G\left(X_1, \beta_0\right)^T]$.

Proof. Denote $Z^T A Z$ as $//Z//_A^2$. By using Lemma 3.1, the test statistic

$$\begin{aligned}&sup_{1\le k\le Nm}\,\omega\left(m,k\right)^2 //S\left(m,k\right)//_A^2\\ &= sup_{1\le k\le Nm}\left(\frac{1}{\sqrt{m}}\rho\left(\frac{k}{m}\right)\right)^2 //\sum_{t=m+1}^{m+k} G(X_t, \beta_0) - \frac{k}{m}\sum_{t=1}^{m} G(X_t, \beta_0)//_A^2 + o_p(1)\\ &= sup_{0<s\le N}\rho\left(s\right)^2 //\frac{1}{\sqrt{m}}\sum_{t=m+1}^{m+k} G(X_t, \beta_0) - \frac{1}{\sqrt{m}}\frac{k}{m}\sum_{t=1}^{m} G(X_t, \beta_0)//_A^2 + o_p(1), \quad (12)\end{aligned}$$

where $s \triangleq \frac{k}{m}$.

According to Proposition 3.1, as $m\to\infty$,

$$\frac{1}{\sqrt{m}}\sum_{t=m+1}^{m+k} G(X_t, \beta_0) = \frac{1}{\sqrt{m}}\sum_{t=1}^{m+k} G\left(X_t, \beta_0\right) - \sum_{t=1}^{m} G\left(X_t, \beta_0\right) \xrightarrow{\mathcal{D}} W_1\left(\tfrac{k}{m}\right) := W_1\left(s\right), \quad (13)$$

and

$$\frac{1}{\sqrt{m}}\frac{k}{m}\sum_{t=1}^{m} G\left(X_t, \beta_0\right) \xrightarrow{\mathcal{D}} sW_2(1), \quad (14)$$

of which $W_1(.)$ and $W_2(.)$are Wiener processes defined in Proposition 3.1. The continuity of $\rho^2(s)\|W_1(s) - sW_2(1)\|_A^2$ on $s \in (0, N]$, together with (12), (13) and (14) yield that

$$\lim_{m \to \infty} P\left(\sup_{1 \le k \le Nm} \omega(m,k)^2 S(m,k)^T A S(m,k) \le c\right)$$
$$= P(\sup_{0 < s \le N} \rho^2(s)\left(W_1(s) - sW_2(1)\right)^T A (W_1(s) - sW_2(1)) \le c,$$

for a given threshold c.

The covariance matrix Σ is unknown. Due to the ergodicity of $\{X_t\}$ and the consistency of β, we can use $\left[\sum_{t=1}^{m} G(X_t, \beta) G^T(X_t, \beta)\right] / m$ as the estimation of the covariance matrix.

4. Application to Weekly Pneumonia & Influenza Mortality Rate

In this section, we apply the monitoring procedure to a binomial time series recording the weekly changes in the Pneumonia & Influenza (P&I) mortality rate from the 1st week of 2017 to the 5th week of 2021. The method successfully identifies several change points associated with the COVID-19 outbreak.

4.1. Data

We collect the statewide weekly P&I mortality rate of from the 40th week of 2013 (CDC treats the 40th week of each year as the start of the corresponding flu season) to the 5th week of 2021[1]. The fluctuation of P&I rates reflects the extent of the spread of pneumonia and influenza. For each state, we first calculate the weekly average P&I rate from 2013 to 2016, the years without severe pandemics. Treating the weekly average rates from 2013 to 2016 as benchmarks, we create a binary time series for each state from the 1st week of 2017 to the 5th week of 2021, indicating whether the P&I rate exceeds the corresponding benchmark value. We choose the binary time series of the six states: California, Colorado, Illinois, Florida, New York and Washington, and the sum of the six binary time series is a Binomial time series with total 6. Geographically, the six states can represent the nationwide flu activity trend. Meanwhile, the six states are far enough to be considered mutually independent at each time.

4.2. Detection Results

As a preliminary analysis of the data set, we calculate the AIC of the fitted generalized Binomial AR(1) model with $\mathbb{Z}_{t-1} = (1, X_{t-1})$, and the AIC of a simple Binomial model with a constant π that all the observations are considered as i.i.d random variables following $Bin(6, \pi)$. The AIC of the simple Binomial model is 2228.31 and the AIC of the generalized Binomial AR(1) model is 1209.48. Compared to a simple Binomial model, the generalized Binomial AR(1) model can capture the dependence between observations and estimate the time-varying π_t.

For the monitoring process, we use the weight function $\omega(m,k) = m^{-\frac{1}{2}}\left(1 + \frac{k}{m}\right)^{-1}\left(\frac{k}{m+k}\right)^{-\gamma}$, $0 \le \gamma < \frac{1}{2}$, which has been widely used in the literature, e.g. [2] and [3]. Table 1 shows the change points detected by models built on different past observations. We can see that the sensitivity of the method increases as γ increases. Except the only data point when $m = 100$ and $\gamma = 0.4$, all other change points are in April, 2020. When $m = 150$ and $\gamma = 0.25, 0.4$, the detected change points are 4/4/2020. Notably, on 3/11/2020 (the 166th observation), the WHO officially declared the COVID-19 outbreak a pandemic[2]. That is, the method can successfully detect change points associated with the COVID-19 outbreak.

Table 1: The detected change points relying on models based on different past observations

[1] Data source: https://gis.cdc.gov/grasp/fluview/mortality.html

[2] https://www.cdc.gov/museum/timeline/covid19.html

Number of past observations Detection launch date	m=100 12/08/2018	m=130 07/06/2019	m=150 11/23/2019
$\gamma = 0$	04/18/2020	04/25/2020	04/11/2020
$\gamma = 0.25$	04/11/2020	04/18/2020	04/04/2020
$\gamma = 0.4$	01/19/2019	04/18/2020	04/04/2020

References

[1] Chu, C.-S.J., Stinchcombe, M. and White, H., 1996. Monitoring structural change. *Econometrica: Journal of the Econometric Society*, 64(5), pp.1045–1065.

[2] Horváth, L., Hušková, M., Kokoszka, P. and Steinebach, J., 2004. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126(1), pp.225–251.

[3] Horváth, L., Kokoszka, P. and Steinebach, J., 2007. On sequential detection of parameter changes in linear regression. *Statistics & Probability Letters*, 77(9), pp.885–895.

[4] Gombay, E. and Serban, D., 2009. Monitoring parameter change in AR(p) time series models. *Journal of Multivariate Analysis*, 100(4), pp.715–725.

[5] Berkes, I., Gombay, E., Horváth, L. and Kokoszka, P., 2004. Sequential change-point detection in GARCH(p, q) models. *Econometric Theory*, 20(6), pp.1140–1167.

[6] McKenzie, E., 1985. Some simple models for discrete variate time series 1. *JAWRA Journal of the American Water Resources Association*, 21(4), pp.645–650.

[7] Weiß, C.H., 2009. A new class of autoregressive models for time series of binomial counts. *Communications in Statistics—Theory and Methods*, *38*(4), pp.447-460.

[8] Scotto, M.G., Weiß, C.H., Silva, M.E. and Pereira, I., 2014. Bivariate binomial autoregressive models. *Journal of Multivariate Analysis*, *125*, pp.233-251.

[9] Kirch, C. and Kamgaing, J.T., 2015. On the use of estimating functions in monitoring time series for change points. *Journal of Statistical Planning and Inference*, *161*, pp.25-49.

[10] Wang, C. and Li, W.K., 2011. On the autopersistence functions and the autopersistence graphs of binary autoregressive time series. *Journal of Time Series Analysis*, *32*(6), pp.639-646.

[11] Meyn, S.P. and Tweedie, R.L., 2012. *Markov chains and stochastic stability*. Springer Science & Business Media.

[12] Cox, D.R., 1975. Partial likelihood. *Biometrika*, *62*(2), pp.269-276.

[13] Straumann, D., 2005. *Estimation in conditionally heteroscedastic time series models* (Vol. 181). Springer Science & Business Media.

[14] Fokianos, K. and Kedem, B., 2004. Partial likelihood inference for time series following generalized linear models. *Journal of Time Series Analysis*, *25*(2), pp.173-197.

[15] Oodaira, H. and Yoshihara, K.I., 1972. Functional central limit theorems for strictly stationary processes satisfyinc the strong mixing condition. In *Kodai Mathematical Seminar Reports* (Vol. 24, No. 3, pp. 259-269). Department of Mathematics, Tokyo Institute of Technology.

[16] Bradley, R.C., 2005. Basic properties of strong mixing conditions. A survey and some open questions.