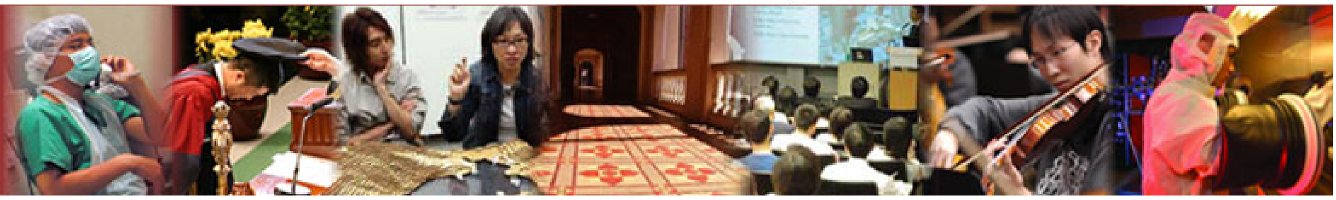

Comput Stat (2010) 25:505–536
DOI 10.1007/s00180-010-0190-8

ORIGINAL PAPER

Binary geometric process model for the modeling of longitudinal binary data with trend

Jennifer S. K. Chan · Doris Y. P. Leung

Received: 14 December 2007 / Accepted: 5 March 2010 / Published online: 30 April 2010

Abstract We propose the Binary Geometric Process (BGP) model for longitudinal binary data with trends. The Geometric Process (GP) model contains two components to capture the dynamics on a trend: the mean of an underlying renewal process and the ratio which measures the direction and strength of the trend. The GP model is extended to binary data using a latent GP. The statistical inference for the BGP models is conducted using the least-square, maximum likelihood (ML) and Bayesian methods. The model is demonstrated through simulation studies and real data analyzes. Results reveal that all estimators perform satisfactorily and that the ML estimator performs the best. Moreover the BGP model is better than the ordinary logistic regression model.

Keywords Geometric process · Longitudinal binary data · Trend data · Threshold model

1 Introduction

Trend data are common, say in clinical trials and business, when longitudinal measurements are made. Lam (1988b) first proposed modeling directly the monotone trend by a monotone process called the Geometric Process (GP). Let $X_1, X_2, \ldots$ be a set of positive random variables. If there exists a positive real number a such that $\{Y_i = a^{i-1}X_i,\ i = 1, 2, \ldots\}$ forms a renewal process (RP) (Feller 1949), then $\{X_i,\ i = 1, 2, \ldots\}$ is called a Geometric Process (GP), and the real number a is called the ratio of the GP. If we define

J. S. K. Chan (✉)
The University of Sydney, Sydney, NSW, Australia
e-mail: jchan@maths.usyd.edu.au

D. Y. P. Leung
The University of Hong Kong, Hong Kong, China

$$E(Y_i) = \mu \quad \text{and} \quad Var(Y_i) = \sigma^2,$$

the mean and variance of a GP are given by

$$E(X_i) = \mu/a^{i-1} \quad \text{and} \quad Var(X_i) = \sigma^2/a^{2(i-1)}, \tag{1}$$

respectively. The trend is increasing if $a < 1$, decreasing if $a > 1$ and stationary if $a = 1$.

The GP model has distinct features over the common one-component generalized linear model (GLM). Firstly it separates exogenous effects onto two components: the underlying initial mean μ and the ratio a which measures the direction and strength of the trend movement. Secondly, the ratio reveals trend movement, ignoring the random noises. It makes forcast simple and straightforward by the inherent geometric structure and the ratio function formulation. Thirdly, Eq. (1) shows that the ratio affects both the mean and variance of a GP and hence allows heteroskedasticity. Essentially, the GP model adopts a geometric relationship to describe a time series. Common daily life examples include the population size within a confined community in demography and the number of infected patients in epidemiology.

Assuming a constant initial level μ and ratio a, original GP model focuses on modeling X_i as the inter-arrival times of a series of events. This model has been applied predominately to the modeling of deteriorating systems and to reliability and maintenance problem in the study of optimal replacement or repairable models (Lam 1988a,b, 1992a; Lam and Zhang 1996; Lam et al. 2002). In recent years, new methods of inference for the GP model have been derived and successfully applied to different areas (Wan and Chan 2009; Chan et al. 2010a,b). In statistical inference, Lam (1992b) proposed some nonparametric (NP) methods of inference including the least square of errors (LSE). Lam and Chan (1998) and Chan et al. (2004) considered parametric maximum likelihood (ML) method adopting, respectively, the Lognormal and Gamma distributions to the RP. They derived large sample distributions for the parameter estimates and showed that ML method is more efficient than NP methods. Lam et al. (2004) showed that GP model out-performed Cox-Lewis model (Cox and Lewis 1966), Weibull process model (Ascher and Feingold 1981) and homogeneous Poisson process model and was easier to implement using NP methods. See Lam (2007) for a brief review and further reference.

Recently, statistical models for recurrent events, including trend-RP of Lindqvist et al. (2003), extended modulated RP of Berman and Turner (1992) and Lawless and Thiagarajah (1996), inhomogeneous gamma process of Berman (1981) and modulated power law process of Lakey and Rigdon (1992), have been proposed. Emphases of these models, like the original GP model, were placed on the modeling of inter-arrival times of a series of events. In fact, a wider class of data including binary, multi-nominal and Poisson count data should be considered. However, none of these models, like GP model, can be easily extended to other type of data. The objective of this paper is to extend GP model to BGP model for longitudinal binary data and derive new modeling methodologies, methods of inference and fields of application for the BGP model.

The paper is presented as follows. Section 2 introduces the basic BGP model and its extensions. Section 3 describes three methods of inference. Performance of the three methods of inference is assessed through simulation experiments in Sect. 4. Then the models are applied to some real data in Sect. 5 and they are compared to the logistic regression model. Hypothesis test on $a = 1$ and likelihood ratio test (LRT) on hierarchical BGP models are performed. Finally, a conclusion is given in Sect. 6.

2 The binary GP model

2.1 The basic model

To extend GP model to binary data, we assume that there is an underlying unobserved GP $\{X_i\}$ and the observed data W_i is the indicator of whether X_i is greater than certain level b which is set $b = 1$, that is,

$$W_i = I(X_i > 1) = I(Y_i > a^{i-1})$$

where $I(E)$ is an indicator function for the event E and $\{Y_i = a^{i-1} X_i\}$ is an underlying RP. Then

$$P_i = P(W_i = 1) = P(X_i > 1) = P(Y_i > a^{i-1}) = 1 - P(Y_i < a^{i-1}) = 1 - F(a^{i-1}),$$

where $F(.)$ is the cumulative distribution function (cdf) of Y_i. The extended model is called the binary GP (BGP) model. The likelihood and log-likelihood functions for $\{W_i\}$ are

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} [1 - F(a^{i-1})]^{w_i} [F(a^{i-1})]^{1-w_i} \tag{2}$$

and

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \{w_i \ln[1 - F(a^{i-1})] + (1 - w_i) \ln[F(a^{i-1})]\} \tag{3}$$

respectively, where $\boldsymbol{\theta}$ is a vector of model parameters. Moreover $\boldsymbol{\theta}$ can also be estimated via the nonparametric LSE approach.

2.2 Extension to lifetime distribution

By assigning some life-time distributions to the latent RP, GP models can be implemented using a parametric approach including ML and Bayesian via Markov chain Monte Carlo (MCMC) approaches. A common lifetime distribution for the RP $\{Y_i\}$ is the two-parameter Weibull distribution with cdf

$$F(y_i) = 1 - \exp[-(\lambda y_i)^{\alpha}]$$

and mean $E(Y_i) = \mu = \Gamma(1+\alpha)/\lambda$ where $\lambda \geq 0$ and $\alpha > 0$ are the scale and shape parameters of the distribution, respectively, and $\Gamma(\cdot)$ is a gamma function. We have

$$P_i = \Pr(W_i = 1) = 1 - F(a^{i-1}) = \exp[-(\lambda a^{i-1})^{\alpha}]. \tag{4}$$

However, if we define a new rate, $\tilde{\lambda} = \lambda^{\alpha}$, and a new ratio, $\tilde{a} = a^{\alpha}$, then (4) becomes

$$p_i = \exp(-\lambda a^{i-1}) \tag{5}$$

which is equivalent to the probability of a BGP model with exponential distribution, rate $\tilde{\lambda}$, ratio $\tilde{a}$ and $\alpha = 1$. As the BGP model with Weibull distribution is over-parameterized, we set $\alpha = 1$ so that $E(Y_i) = \mu = 1/\lambda$ in subsequent analyses. Using (2) and (3), the likelihood and log-likelihood functions for the observed data $\{W_i\}$ are

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} L_i(w_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} [\exp(-\lambda_i a_i^{i-1})]^{w_i} \cdot [1 - \exp(-\lambda_i a_i^{i-1})]^{1-w_i} \tag{6}$$

and

$$\ell(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}) = -\sum_{i=1}^{n} w_i(\lambda_i a_i^{i-1}) + \sum_{i=1}^{n} (1 - w_i) \ln[1 - \exp(-\lambda_i a_i^{i-1})] \tag{7}$$

respectively, where the mean $\lambda_i^{-1} = \lambda^{-1}$ and ratio $a_i = a$ are constant for the BGP model and $\boldsymbol{\theta} = (a, \lambda)$ is a vector of model parameters.

Other lifetime distributions, such as gamma and lognormal, are not considered because their cdfs involve integration. This complicates the Newton Raphson (NR) procedures as numerical method is required to approximate the integrals. Since this paper focuses on establishing the BGP modelling framework, extension of GP model to other lifetime distributions will be investigated in future research.

2.3 Extension to threshold model

Original GP model (Lam 1988a,b) with a constant initial mean μ and ratio a is limited to data with a single trend. However, data may exhibit multiple trends showing different stages of development for a certain event: growing ($a_i < 1$), stabilizing ($a_i = 1$) and declining ($a_i > 1$) stages. Examples include the number of daily-infected cases during the outbreak of the severe acute respiratory syndrome (SARS) epidemic in 2003 reported in Chan et al. (2006). They assigned a separate GP to each stage of development allowing the mean μ_i and ratio a_i to change across stages. Suppose that there are M underlying RPs,

$$RP_m = \{a_{0m}^{i-T_m} X_i;\ T_m \leq i < T_{m+1}\},$$

where the turning points $T_m,\ m = 1, \ldots, M$ mark the time when the trends change their means $\mu_i = 1/\lambda_i$ and/or their ratios a_i. Then

$$\lambda_i^{-1} = \sum_{m=1}^{M} \lambda_{0m}^{-1}\, I(T_m \le i < T_{m+1}), \tag{8}$$

$$a_i = \sum_{m=1}^{M} a_{0m}\, I(T_m \le i < T_{m+1}) \tag{9}$$

where $\lambda_{0m}^{-1} = \exp(\beta_{\mu 0m}), a_{0m} = \exp(\beta_{a0m}), T_1 = 1, T_{M+1} = n+1$ and $m = 1, 2, \ldots, M$ and

$$p_i = \sum_{m=1}^{M} \exp(-\lambda_{0m} a_{0m}^{i-T_m})\, I(T_m \le i < T_{m+1}). \tag{10}$$

This extended model is called threshold BGP (TBGP) model. When $M = 2$, parameters λ_i and a_i can be alternatively written as

$$\lambda_i^{-1} = \lambda_{01}^{-1} + (\lambda_{02}^{-1} - \lambda_{01}^{-1})\, I(i \ge T_2),$$
$$a_i = a_{01} + (a_{02} - a_{01})\, I(i \ge T_2),$$

indicating that TBGP model is nested within BGP model. Hence LRT can be applied to test the significance of hierarchical TBGP model when $M = 2$ against BGP model when $M = 1$.

2.4 Extension to mean and ratio functions

TBGP model implicitly assumes that the initial mean μ_i and ratio a_i take different values, λ_{0m}^{-1} and a_{0m} respectively, across the M stages but homogenous within each stage. While this model allows different stages of development, it fails to account for other covariate effects which are time-evolving. For example, the probability of a positive heroin test for a methadone patient may depend on treatment factors like the methadone dose and duration in treatment which vary over time. A natural way to accommodate these time-evolving effects is to adopt a linear function of covariates log-linked to the initial mean function μ_i

$$\mu_i = \lambda_i^{-1} = \exp(\eta_{\mu i}) = \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu 1 i} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu i}) \tag{11}$$

where $z_{\mu k i},\ k = 1, \ldots, q_\mu$ are covariates. Moreover, the ratio a_i may adopt some continuous functions to allow a gradual transition between stages, say from a growing stage to a declining stage when a_i changes slowly from $a_i < 1$ to $a_i > 1$. To allow the transition as well as other covariate effects on the trend movement, the ratio a_i is also log-linked to a linear function of covariates:

$$a_i = \exp(\eta_{ai}) = \exp(\beta_{a0} + \beta_{a1} z_{a1i} + \cdots + \beta_{aq_a} z_{aq_a i}) \tag{12}$$

where $z_{aki},\ k = 1, \ldots, q_a$ are covariates. Different ratio functions a_i describe different trend movements for the probability p_i

$$p_i = \exp(-\lambda_i a_i^{i-1}) \tag{13}$$

where $\mu_i = \lambda_i^{-1}$ and a_i are given by (11) and (12), respectively. This extended model with an adaptive ratio function is called adaptive BGP (ABGP) model as it allows a broad range of applications for the BGP model.

Note that the underlying RP $\{Y_i\}$ under ABGP model is no longer independently and identically distributed (IID) and is a stochastic process (SP) in general and that λ_i in (8) for the TBGP model can be written as (11) when $\beta_{\mu 0} = 0$ and $z_{\mu mi} = I(T_m \leq i < T_{m+1}),\ m = 1, \ldots, M$.

ABGP model describes trend movement on p_i with different stages of development. To investigate the characteristic of these trends, we identify six possible cases. For cases 1–2, we set $\mu_i = 1$ and let a to vary. For cases 3–6, we set $a_i = \exp(\beta_{a0} + \beta_{a1} \ln i)$, $\mu_i = \exp(\beta_{\mu 0})$ and let β_{a0} or β_{a1} to vary. These six cases which describe most trend movements for p_i and a_i in any trend data are summarized below:

Case 1. Fixed $a_i = a \leq 1 : p_i$ is increasing at a decreasing rate (Fig. 1a). When $a = 1$, p_i is a constant.
Case 2. Fixed $a_i = a > 1 : p_i$ is decreasing at a decreasing rate (Fig. 1b).
Case 3. Increasing $a_i < 1 : p_i$ is increasing at an increasing rate and then a decreasing rate (Fig. 2a).
Case 4. Decreasing $a_i > 1 : p_i$ is decreasing at an increasing rate and then a decreasing rate (Fig. 2b).
Case 5. Increasing a_i from $a_i < 1$ to $a_i > 1 : p_i$ is increasing according to case 3 and then decreasing according to case 4 (Fig. 3a).
Case 6. Decreasing a_i from $a_i > 1$ to $a_i < 1 : p_i$ is decreasing according to case 4 and then increasing according to case 3 (Fig. 3b).

Note that Fig. 3c illustrates the trends of p_i and a_i when $\beta_{\mu 0}$ is varied.

3 Methodology

3.1 Least-square-error method

Least-square-error (LSE) method is perhaps the simplest method of inference. Lam (1992b) considered log-LSE method and Chan et al. (2006) considered both log-LSE and LSE methods for positive continuous data. In these methods, parameters are estimated by minimizing the sum of squared-error (SSE) on $\ln X_i$ and X_i, respectively. In this paper, we consider LSE method for W_i and the SSE is given by

$$SSE = \sum_{i=1}^{n} (W_i - p_i)^2 \tag{14}$$

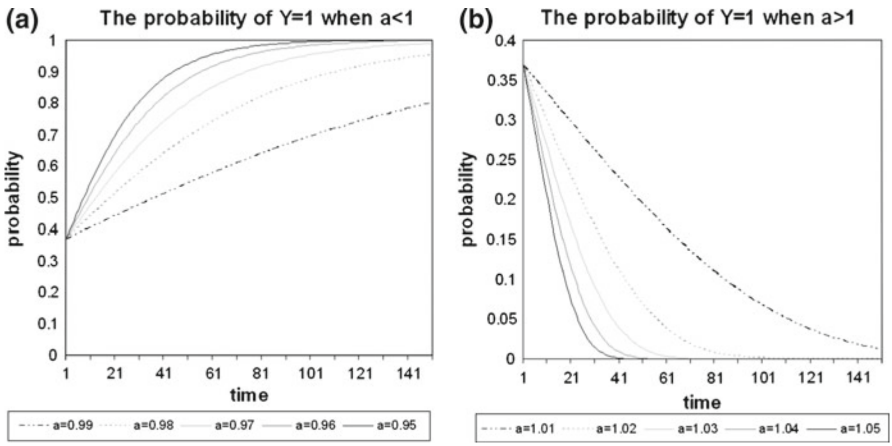

Fig. 1 1 Probability p_i for various ratio parameters a when $\mu_i = 1$ and **a** $a_i = a < 1$; **b** $a_i = a > 1$

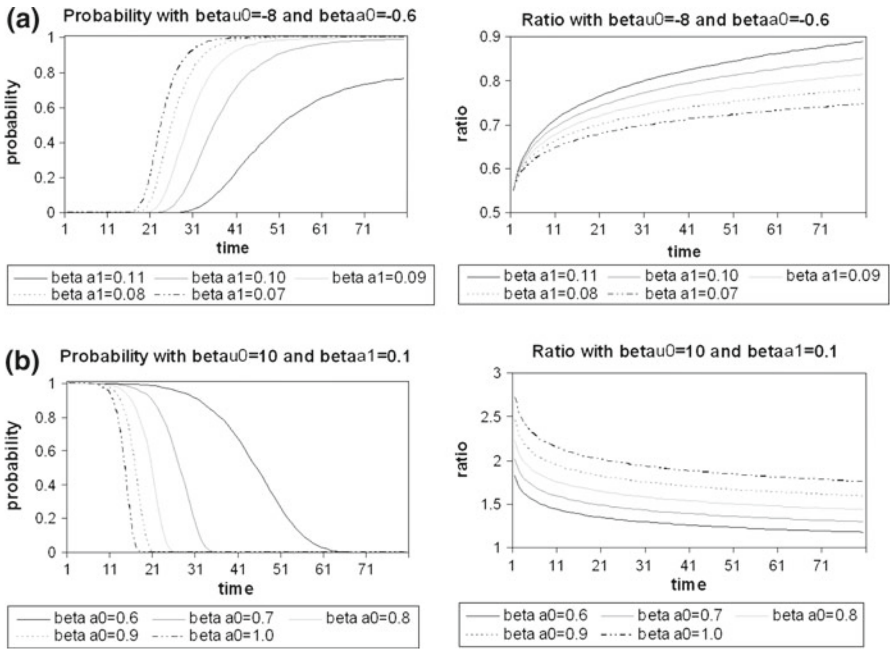

Fig. 2 **a** Probability p_i and ratio a_i for various β_{a1} when $\mu_i = \exp(-8)$, $a_i = \exp(-0.6 + \beta_{a1} \ln i) < 1$ & increasing. **b** Probability p_i and ratio a_i for various β_{a0} when $\mu_i = \exp(10)$, $a_i = \exp(\beta_{a0} + 0.1 \ln i) > 1$ & decreasing

where $E(W_i) = p_i$ is a function of a_i and μ_i given by (5), (10) and (13) for the BGP, TBGP and ABGP models, respectively. Newton Raphson (NR) iterative procedure

$$\boldsymbol{\theta}^{(v+1)} = \boldsymbol{\theta}^{(v)} - [SSE''(\boldsymbol{\theta}^{(v)})]^{-1} SSE'(\boldsymbol{\theta}^{(v)}), \tag{15}$$

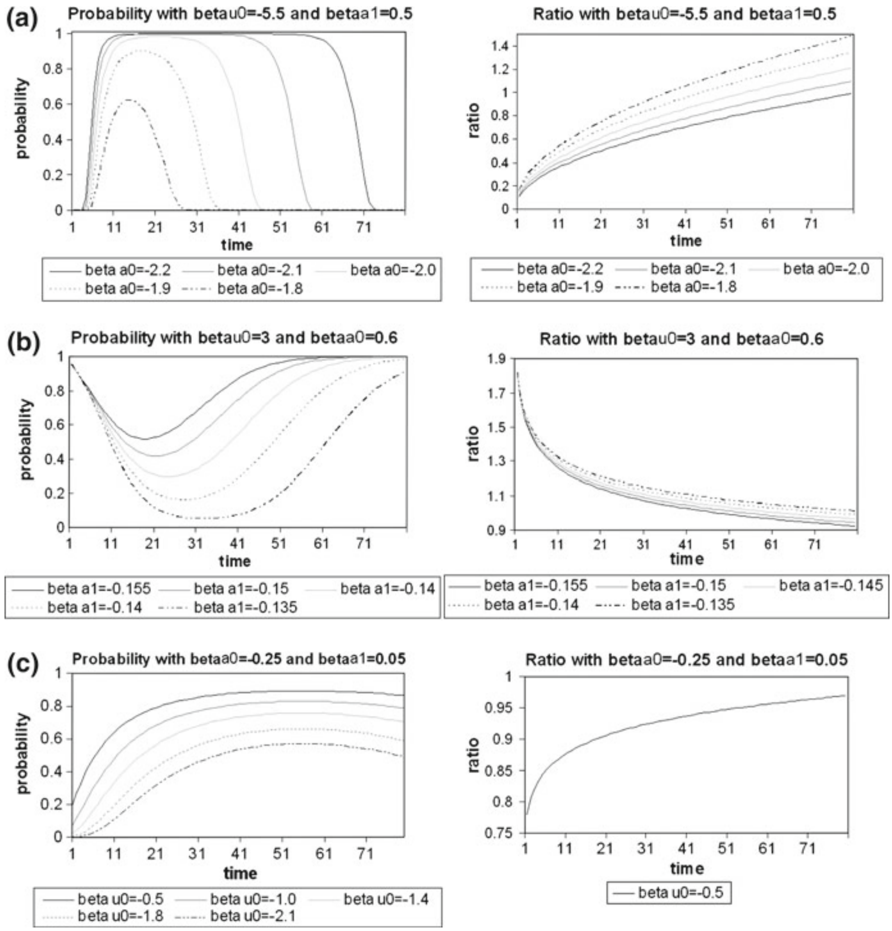

Fig. 3 **a** Probability p_i and ratio a_i for various β_{a0} when $\mu_i = \exp(-5.5), a_i = \exp(\beta_{a0} + 0.5 \ln i)$ increasing from <1 to >1. **b** Probability p_i and ratio a_i for various β_{a1} when $\mu_i = \exp(3), a_i = \exp(0.6 + \beta_{a1} \ln i)$ decreases from >1 to <1. **c** Probability p_i and ratio a_i for various $\mu_i = \exp(\beta_{\mu 0})$ when $a_i = \exp(-0.25 + 0.05 \ln i)$

where $\boldsymbol{\theta}^{(v)}$ denotes the vector of parameter estimates in the v-th iteration, is run until $\| \boldsymbol{\theta}^{(v+1)} - \boldsymbol{\theta}^{(v)} \|$ is sufficiently small. Then the LSE estimates are given by $\widehat{\boldsymbol{\theta}}_{LSE} = \boldsymbol{\theta}^{(v+1)}$ and their covariance matrix is given by the inverse of information matrix, i.e. $[SSE''(\widehat{\boldsymbol{\theta}}_{LSE})]^{-1}$ in (22) and (23). We suggest to use Bayesian estimates as initial values. The first and second order derivatives, $SSE'(\boldsymbol{\theta})$ and $SSE''(\boldsymbol{\theta})$, respectively, as required in the NR procedure are given in "Appendix A".

For data with multiple trends, there exists a problem of detecting the turning point(s) T_m in the TBGP model. Since the LSE method cannot be applied to estimate T_m, Chan et al. (2006) proposed a moving window technique to estimate T_m. A direct way to estimate the turning points is to take condition on M. We first set $M = 2$ and search T_2 over certain interval not too close to the end points 1 and n. Condition on each T_2, two GP models are fitted to data with time $i < T_2$ and $i \geq T_2$, respectively. Optimal T_2 is chosen to minimize the SSE in (14). Then M is set to 3 and the search for T_3 given T_2 is similarly repeated from the remaining time points not too close to T_2, 1 and n. This method is essentially a partial LSE method where $\boldsymbol{\theta}$ is obtained by LSE method but $\{T_m\}$ by searching.

3.2 Maximum likelihood method

In parametric inference, the log-likelihood function $\ell(\boldsymbol{\theta})$ to be maximized is given by (7) where a_i and λ_i^{-1} are given by (8) and (9) for TBGP model and (11) and (12) for ABGP model, respectively. Again the NR procedure is used to solve for the ML estimates in $\ell'(\boldsymbol{\theta}) = 0$. The turning points for the TBGP model are estimated via a partial ML method similar to that of the partial LSE method. The first and second order derivatives, $\ell'(\boldsymbol{\theta})$ and $\ell''(\boldsymbol{\theta})$, respectively, as required in the NR procedure are given in "Appendix B". Then the ML estimates are given by $\widehat{\boldsymbol{\theta}}_{ML} = \boldsymbol{\theta}^{(v+1)}$ when v is sufficiently large so that $\boldsymbol{\theta}^{(v+1)}$ converges and their covariance matrix is given by the negative inverse of information matrix, i.e. $-[\ell''(\widehat{\boldsymbol{\theta}}_{ML})]^{-1}$ given in (25) and (26). Large sample properties for the ML estimates $\widehat{\boldsymbol{\theta}}_{ML}$ are given in the following theorem:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, n^{-1}\boldsymbol{\Sigma}) \tag{16}$$

where $n^{-2}\boldsymbol{\Sigma}$ is the large sample covariance matrix for $\widehat{\boldsymbol{\theta}}_{ML}$ and $\boldsymbol{\Sigma}$ is given in (27) in "Appendix C". With these asymptotic distributions, we can construct confidence intervals and perform hypothesis testing on $\boldsymbol{\theta}$.

3.3 Bayesian method

Comparing to the classical likelihood approach, Bayesian approach using MCMC techniques converts an optimization problem into a sampling problem, thus avoiding the numerical difficulties associated with the maximization of complicated high-dimensional likelihood functions. This is done by iterative simulation of model parameters from the posterior distributions. In case of nonstandard posterior distributions, MCMC method (Smith and Roberts 1993; Gilks et al. 1996) with Gibbs sampling and Metropolis Hastings algorithm (Hastings 1970 and Metropolis et al. 1953) produce samples from the intractable posterior distributions of all unknown parameters. Moreover, WinBUGS, a Bayesian software using MCMC techniques, enables easy model implementation (Spiegelhalter et al. 2004).

To implement the MCMC techniques, we derive the conditional posterior densities from which parameters are sampled. The conditional posterior densities are proportional to the joint density $f(\boldsymbol{w}, \boldsymbol{\theta})$ of data $f(\boldsymbol{w}|\,\boldsymbol{\theta})$ and priors $f(\boldsymbol{\theta})$. Parameter estimates $\widehat{\boldsymbol{\theta}}_B$ and their standard error (SE) are given by the posterior means (or median) and standard derivations of the posterior samples. Command codes for the three proposed models are given in "Appendix D".

Unlike the LSE and ML methods, the turning points T_m are set as model parameters in Bayesian method and are estimated from simulation. In this way, the uncertainty due to the estimation of T_m is measured simultaneously with the uncertainty due to the remaining model parameters. This offers an advantage over LSE and ML methods because the SE of T_m can be obtained and the SE of other parameters will not be underestimated. Without specific information on the parameters, we assign non-informative priors with large variances to all parameters, for example, we fix M and assign uniform priors to T_m. Then the number of turning points M can be chosen by some model selection criteria.

The Bayesian hierarchies for the three models are

$$\begin{aligned}
\boldsymbol{w} &\sim \text{bernoulli}(\boldsymbol{p}),\\
\boldsymbol{p} &= h(\boldsymbol{\mu}, \boldsymbol{a}),\\
\boldsymbol{\beta}_j &\sim MVN_{q_j}(\mathbf{0}, \boldsymbol{\Sigma}_j),\ \ j = a, \mu,\\
T_m &\sim U(T_{m-1} + b, n - b),\ \ m = 1, \ldots, M, \qquad \text{for TBGP model,}
\end{aligned}$$

where $\boldsymbol{w}$ and $\boldsymbol{p}$ are $n \times 1$ vectors of w_i and p_i, respectively, p_i is a function of a_i and μ_i given by (5), (10) and (13), respectively for BGP, TBGP and ABGP models, priors for the $q_j \times 1$ vectors of parameters $\boldsymbol{\beta}_j$, $j = a, \mu$ are multivariate normal with means $\mathbf{0}$ (a vector of zeros) and covariance matrices $\boldsymbol{\Sigma}_j$, $\boldsymbol{\Sigma}_j = \text{diag}(\tau_{jq'_j})$, $q'_j = 1, \ldots, q_j$, $j = a, \mu$ are diagonal matrices with large diagonal values $\tau_{jq'_j}$ and b is the minimum width for any trend. Parameters for BGP and ABGP models are drawn from Gibbs sampler from the posterior conditional distributions $[\boldsymbol{\beta}_\mu | \boldsymbol{\beta}_a, \boldsymbol{w}]$ and $[\boldsymbol{\beta}_a | \boldsymbol{\beta}_\mu, \boldsymbol{w}]$ whereas they are drawn from $[\boldsymbol{\beta}_\mu | \boldsymbol{\beta}_a, T_m, \boldsymbol{w}]$, $[\boldsymbol{\beta}_a | T_m, \boldsymbol{\beta}_\mu, \boldsymbol{w}]$ and $[T_m | \boldsymbol{\beta}_\mu, \boldsymbol{\beta}_a, \boldsymbol{w}]$ for TBGP model.

For each model, 55000 iterations are run and the first 5000 iterations are set as the burn-in period to ensure that convergence has reached. Thereafter parameters are sub-sampled from every 50th iteration ($h = 50$) to reduce the auto-correlation in the sample. Simulated values from the Gibbs sampler after the burn-in period are taken to mimic a random sample of size 1000 from the joint posterior distribution for posterior inference. History and ACF plots show that the posterior samples have converged and are close to independence.

4 Simulation experiment

In this simulation experiment, we consider three models:

$$\begin{aligned}
\text{BGP model: } (a_i, 1/\lambda_i) &= (a_0, 1/\lambda_0)\\
\text{TBGP model: } (a_i, 1/\lambda_i) &= \begin{cases} (\exp(\beta_{a01}), \exp(\beta_{\mu 01})) & \text{if } i < T;\\ (\exp(\beta_{a02}), \exp(\beta_{\mu 02})) & \text{if } i \geq T \end{cases} \qquad (17)\\
\text{ABGP model: } (a_i, 1/\lambda_i) &= (\exp(\beta_{a0} + \beta_{a1} \ln i), \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu i}))
\end{aligned}$$

and p_i are given by (5), (10) and (13), respectively. We set the number of realizations $N = 200$ and the sample size in each realization $n = 100$ unless otherwise stated. For each realization, we first simulate u_i, $i = 1, \ldots, 100$ from a uniform distribution $U(0, 1)$ and set $w_i = I(u_i < \exp[-\lambda_i a_i^{i-1}])$. For each parameter, the reported $\widehat{\theta}$ is the mean of 200 estimates θ_j where j indicates the realization and θ denotes the true parameter value. We compare the performance of different parameter estimates by four goodness-of-fit (GOF) measures:

$$PE = \frac{\widehat{\theta} - \theta}{|\theta|},$$

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(\theta_j - \theta)^2,$$

$$R = \frac{SD}{SE} = \left[\frac{1}{N-1}\sum_{j=1}^{N}(\theta_j - \widehat{\theta})^2\right]^{\frac{1}{2}}\left[\frac{1}{N}\sum_{j=1}^{N}SE(\theta_j)\right]^{-1},$$

$$PC = 1 - \frac{1}{N}\sum_{j=1}^{N} I[\theta \in (\theta_j - 1.96\, SE(\theta_j),\ \theta_j + 1.96\, SE(\theta_j))],$$

where $SE(\theta_j)$ are square roots of diagonal elements in the large sample covariance matrix $n^{-2}\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}$ is given in (16) and (27). The first two measures reveal the accuracy and precision of $\widehat{\theta}$ while the last two statistics measure the reliability of $SE(\theta_j)$. The ratio R compares the mean of standard deviations $SD(\theta_j)$ to the mean of the asymptotic standard errors $SE(\theta_j)$. If $SE(\theta_j)$ are good estimates of $SD(\theta_j)$, R should be close to 1 and the proportion of confidence intervals PC which do not contain θ should be close to 0.05.

We conduct the simulation experiment using all three methods of inference, namely ML, LSE and Bayesian. We report all results using ML method in Tables 1, 2, and 3 but a subset of results using LSE and Bayesian methods for comparison in Table 4 due to limited space. We do not report R and PC in Table 4 because large sample covariance matrix cannot be obtained for LSE and Bayesian methods. In general, the performance of ML method is better than those of LSE and Bayesian methods but the size of effects for each parameter are similar. Hence we focus on the results using ML method in the following sections.

4.1 The BGP model

We set the mean $\mu_i = \frac{1}{\lambda_0}$ and ratio $a_i = a_0$ to be constant across time i. We consider the following combination of parameters: $\lambda_0 = 0.2,\ 0.5,\ 1.0, a_0 = 0.96, 0.98, 1.0, 1.02, 1.04$ and $n = 50, 100$. Results are reported in Table 1 and 4.

Table 1 shows that $\hat{a}_0$ is good under all conditions, especially, when a_0 is close to 1 which indicates a more gentle trend. On the other hand, performance of $\widehat{\lambda_0}$ depends on values of a_0: it is better when λ_0 is large but a_0 is small and when λ_0 is small but a_0 is large. Both cases indicate a moderate level of $p_i = \exp(-\lambda_0 a_0^{i-1})$. In summary, ML estimators perform satisfactorily for BGP models and that longer trend (larger n) gives better estimates.

Table 1 Simulation studies for the BGP models using ML method

a_0	$n=50$ $\lambda_0=0.2$		$n=50$ $\lambda_0=0.5$		$n=50$ $\lambda_0=1.0$		$n=100$ $\lambda_0=0.2$		$n=100$ $\lambda_0=0.5$		$n=100$ $\lambda_0=1.0$	
	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$
0.96	0.9496	0.2976	0.9526	0.6301	0.9573	1.1232	0.9503	0.2754	0.9560	0.5733	0.9574	1.0904
PE	*− 0.0108*	*0.4878*†	*−0.0077*	*0.2601*†	*−0.0028*	*0.1232*	*−0.0101*	*0.3769*†	*−0.0041*	*0.1466*	*−0.0027*	*0.0904*
MSE	*0.0015*	*0.0501*	*0.0010*	*0.1583*	*0.0004*	*0.2448*	*0.0007*	*0.0290*	*0.0002*	*0.0648*	*0.0001*	*0.1805*
R	*1.1253*	*1.1890*†	*0.8702*	*1.0568*	*1.0015*	*1.0568*	*1.2059*‡	*1.2587*‡	*1.0455*	*1.1217*	*0.9888*	*0.9917*
PC	*0.0150**	*0.0300*	*0.0400*	*0.0400*	*0.0250*	*0.0200*	*0.0100**	*0.0150**	*0.0200*	*0.0500*	*0.0400*	*0.0250*
0.98	0.9671	0.3015	0.9781	0.5724	0.9796	1.0864	0.9751	0.2641	0.9778	0.5618	0.9796	1.0542
PE	*−0.0132*	*0.5075*†	*−0.0019*	*0.1447*	*−0.0004*	*0.0864*	*−0.0050*	*0.3203*†	*−0.0022*	*0.1236*	*−0.0005*	*0.0542*
MSE	*0.0012*	*0.0533*	*0.0004*	*0.0924*	*0.0002*	*0.2017*	*0.0002*	*0.0218*	*0.0001*	*0.0507*	*0.0000*	*0.1066*
R	*1.0242*	*1.0801*	*0.9976*	*0.9973*	*0.8038*‡	*1.0182*	*1.0496*	*1.1252*	*1.0697*	*1.0173*	*1.0164*	*1.0164*
PC	*0.0300*	*0.0300*	*0.0400*	*0.0500*	*0.0300*	*0.0600*	*0.0200*	*0.0200*	*0.0350*	*0.0150**	*0.0400*	*0.0450*
1.00	0.9951	0.2536	0.9996	0.5572	1.0001	1.0854	0.9987	0.2308	0.9997	0.5327	0.9999	1.0465
PE	*−0.0049*	*0.2682*†	*−0.0004*	*0.1144*	*0.0001*	*0.0854*	*−0.0013*	*0.1538*†	*−0.0003*	*0.0655*	*−0.0001*	*0.0465*
MSE	*0.0006*	*0.0263*	*0.0002*	*0.0648*	*0.0001*	*0.1819*	*0.0001*	*0.0098*	*0.0000*	*0.0295*	*0.0000*	*0.0911*
R	*1.0348*	*1.1008*	*1.0914*	*1.0037*	*1.0868*	*1.0034*	*1.1235*	*1.1317*	*1.0583*	*1.0200*	*0.9298*	*0.9397*
PC	*0.0200*	*0.0450*	*0.0200*	*0.0350*	*0.0250*	*0.0400*	*0.0200*	*0.0350*	*0.0300*	*0.0500*	*0.0550*	*0.0650*

Table 1 continued

a_0	$n = 50$, $\lambda_0 = 0.2$		$n = 50$, $\lambda_0 = 0.5$		$n = 50$, $\lambda_0 = 1.0$		$n = 100$, $\lambda_0 = 0.2$		$n = 100$, $\lambda_0 = 0.5$		$n = 100$, $\lambda_0 = 1.0$	
	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$	$\hat{a}_0$	$\hat{\lambda}_0$
1.02	1.0181	0.2372	1.0210	0.5418	1.0245	1.0340	1.0194	0.2189	1.0201	0.5243	1.0232	1.0112
PE	*−0.0019*	*0.1859*†	*0.0010*	*0.0837*	*0.0044*	*0.0340*	*−0.0006*	*0.0947*	*0.0001*	*0.0485*	*0.0031*	*0.0112*
MSE	*0.0004*	*0.0189*	*0.0002*	*0.0564*	*0.0003*	*0.1957*	*0.0000*	*0.0059*	*0.0000*	*0.0261*	*0.0001*	*0.1278*
R	*1.0693*	*1.0652*	*1.0587*	*0.9925*	*0.9452*	*0.9245*	*1.0916*	*1.1148*	*0.9245*	*0.9622*	*0.9164*	*0.8645*
PC	*0.0450*	*0.0500*	*0.0300*	*0.0550*	*0.0400*	*0.0850**	*0.0500*	*0.0400*	*0.0600*	*0.0750*	*0.0400*	*0.0750*
1.04	1.0393	0.2356	1.0454	0.5221	1.0516	0.9936	1.0408	0.2154	1.0476	0.4942	1.0608	0.9319
PE	*−0.0006*	*0.1780*†	*0.0052*	*0.0442*	*0.0112*	*−0.0064*	*0.0008*	*0.0771*	*0.0073*	*−0.0115*	*0.0200*	*−0.0681*
MSE	*0.0003*	*0.0140*	*0.0003*	*0.0556*	*0.0007*	*0.2931*	*0.0001*	*0.0067*	*0.0004*	*0.0449*	*0.0021*	*0.1905*
R	*1.0729*	*1.1128*	*0.9646*	*0.9571*	*1.1486*	*0.8897*	*0.9266*	*1.0407*	*0.7336*‡	*0.9155*	*0.5260*‡	*1.0033*
PC	*0.0350*	*0.0550*	*0.0300*	*0.0550*	*0.0250*	*0.1350**	*0.0650*	*0.0450*	*0.0250*	*0.0600*	*0.0350*	*0.0650*

$^{\dagger}|PE| > 15\%$. $^{\ddagger}R < 0.85$ or $R > 1.15$. $^{\star}PC < 0.02$ or $PC > 0.08$

Table 2 Simulation studies for the TBGP models using ML method

	$\beta_{\mu 01}$	$\beta_{\mu 02}$	β_{a01}	β_{a02}	T		$\beta_{\mu 01}$	$\beta_{\mu 02}$	β_{a01}	β_{a02}	T
Set 1						Set 2					
True	−1.0000	1.0000	0.0100	−0.0100	101.00	True	−1.0000	1.0000	0.0200	−0.0300	101.00
Est.	−1.0130	1.0378	0.0111	−0.0100	102.08	Est.	−1.0264	1.0530	0.0220	−0.0315	101.80
PE	*−0.0130*	*0.0378*	*0.1100*	*0.0000*	*0.0107*	PE	*−0.0264*	*0.0530*	*0.1000*	*−0.0500*	*0.0079*
MSE	*0.2841*	*0.1352*	*0.0046*	*0.0052*	*3.226*	MSE	*0.3251*	*0.1770*	*0.0067*	*0.0082*	*1.8360*
R	*0.9616*	*1.0288*	*1.0876*	*0.9762*		R	*0.9337*	*0.9808*	*0.8772*	*0.9362*	
PC	*0.0450*	*0.0300*	*0.0150*⋆	*0.0400*		PC	*0.0450*	*0.0250*	*0.0750*	*0.0600*	
Set 3						Set 4					
True	−1.0000	1.0000	0.0100	−0.0200	101.00	True	−1.0000	1.0000	0.0300	−0.0300	101.00
Est.	−1.0042	1.0426	0.0113	−0.0204	102.31	Est.	−1.0100	1.0511	0.0326	−0.0313	101.85
PE	*−0.0042*	*0.0426*	*0.1300*	*0.0200*	*0.0130*	PE	*−0.0100*	*0.0511*	*0.0867*	*−0.0433*	*0.0084*
MSE	*0.3173*	*0.1410*	*0.0050*	*0.0067*	*4.3280*	MSE	*0.3306*	*0.1898*	*0.0084*	*0.0081*	*1.4390*
R	*0.8890*	*1.0592*	*0.9758*	*0.9123*		R	*0.9599*	*0.9847*	*0.8560*	*0.9611*	
PC	*0.0550*	*0.0200*	*0.0450*	*0.0600*		PC	*0.0400*	*0.0200*	*0.0450*	*0.0550*	
Set 5						Set 6					
True	−1.0000	1.0000	0.0200	−0.0100	101.00	True	−1.0000	1.0000	−0.0100	0.0300	101.00
Est.	−1.0137	1.0315	0.0214	−0.0100	101.66	Est.	−1.0567	1.0522	−0.0104	0.0316	99.47
PE	*−0.0137*	*0.0315*	*0.0700*	*0.0000*	*0.0065*	PE	*−0.0567*	*0.0522*	*−0.0400*	*0.0533*	*−0.0151*
MSE	*0.2871*	*0.1486*	*0.0057*	*0.0051*	*1.6610*	MSE	*0.3102*	*0.1531*	*0.0055*	*0.0070*	*4.6950*
R	*0.9747*	*0.9946*	*0.9420*	*0.9823*		R	*0.9412*	*1.0384*	*0.9418*	*0.9130*	
PC	*0.0550*	*0.0250*	*0.0600*	*0.0400*		PC	*0.0500*	*0.0300*	*0.0550*	*0.0500*	

Table 2 continued

	$\beta_{\mu 01}$	$\beta_{\mu 02}$	β_{a01}	β_{a02}	T		$\beta_{\mu 01}$	$\beta_{\mu 02}$	β_{a01}	β_{a02}	T
Set 7						Set 8					
True	−1.0000	1.0000	0.0200	−0.0200	101.00	True	−1.0000	1.0000	−0.0300	0.0100	101.00
Est.	−1.0040	1.0346	0.0217	−0.0203	101.87	Est.	−1.0743	1.0453	−0.0317	0.0103	100.43
PE	*−0.0040*	*0.0346*	*0.0850*	*−0.0150*	*0.0086*	PE	*−0.0743*	*0.0453*	*−0.0567*	*0.0300*	*−0.0056*
MSE	*0.2952*	*0.1581*	*0.0061*	*0.0063*	*1.9740*	MSE	*0.3065*	*0.1495*	*0.0072*	*0.0043*	*1.4150*
R	*0.9799*	*1.0125*	*0.9172*	*0.9590*		R	*0.9477*	*1.0542*	*1.0324*	*1.1474*	
PC	*0.0300*	*0.0400*	*0.0650*	*0.0550*		PC	*0.0450*	*0.0350*	*0.0250*	*0.0300*	
Set 9						Set 10					
True	1.0000	−0.5000	0.0200	−0.0200	101.00	True	−1.0000	0.5000	0.0200	−0.0200	101.00
Est.	1.0626	−0.5278	0.0216	−0.0205	101.00	Est.	−0.9884	0.5169	0.0226	−0.0204	102.62
PE	*0.0626*	*−0.0556*	*0.0800*	*−0.0250*	*0.0000*	PE	*0.0116*	*0.0338*	*0.1300*	*−0.0200*	*0.0160*
MSE	*0.2893*	*0.1101*	*0.0051*	*0.0060*	*1.9070*	MSE	*0.2783*	*0.1221*	*0.0079*	*0.0058*	2.5320
R	*1.0919*	*1.1121*	*1.0388*	*0.9621*		R	*0.9907*	*1.0231*	*0.8462*[‡]	*0.9007*	
PC	*0.0400*	*0.0100*[⋆]	*0.0350*	*0.0600*		PC	*0.0300*	*0.0400*	*0.0300*	*0.0700*	
Set 11						Set 12					
True	−1.0000	2.0000	0.0200	−0.0200	101.00	True	−0.2000	0.0000	0.0200	−0.0200	101.00
Est.	−1.0809	2.1178	0.0209	−0.0218	101.67	Est.	−0.1295	−0.0011	0.0253	−0.0198	102.11
PE	*−0.0809*	*0.0589*	*0.0450*	*−0.0900*	*0.0066*	PE	*0.3525*[†]	*−0.0011*	*0.2650*[†]	*0.0100*	*0.0110*
MSE	*0.4100*	*0.3528*	*0.0077*	*0.0093*	3.5840	MSE	*0.2732*	*0.1159*	*0.0094*	*0.0051*	2.2660
R	*0.9073*	*0.9099*	*0.9191*	*0.8511*		R	*0.9837*	*0.9452*	*0.9101*	*1.0374*	
PC	*0.0350*	*0.0500*	*0.0550*	*0.0700*		PC	*0.0500*	*0.0600*	*0.0150*[⋆]	*0.0250*	

$PE = \widehat{\theta} - \theta$ when $\theta = 0$. [†] $|PE| > 15\%$
[‡] $R < 0.85$ or $R > 1.15$. [⋆] $PC < 0.02$ or $PC > 0.08$

Table 3 Simulation studies for the ABGP models using ML method

	$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}		$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}
Set 1				Set 2			
True	0.00000	1.00000	0.01000	True	0.00000	1.00000	0.02000
Est.	−0.01560	1.06093	0.01084	Est.	−0.03686	1.08373	0.02177
PE	*−0.01560*	*0.06093*	*0.08370*	PE	*−0.03686*	*0.08373*	*0.08849*
MSE	*0.16684*	*0.04529*	*0.00005*	MSE	*0.14173*	*0.05248*	*0.00006*
R	*0.93889*	*1.00061*	*0.94233*	R	*0.93889*	*1.00061*	*0.94233*
PC	*0.06500*	*0.04000*	*0.04500*	PC	*0.06500*	*0.04000*	*0.36500*
Set 3				Set 4			
True	1.00000	1.00000	0.01000	True	0.00000	1.00000	−0.01000
Est.	0.99695	1.07041	0.01010	Est.	−0.06451	1.13895	−0.01118
PE	*−0.00305*	*0.07041*	*0.01018*	PE	*−0.06451*	*0.13895*	*−0.11792*
MSE	*0.24632*	*0.09573*	*0.00006*	MSE	*0.26054*	*0.10593*	*0.00008*
R	*1.07420*	*0.99996*	*1.03983*	R	*0.95773*	*0.98321*	*0.92777*
PC	*0.03500*	*0.02000*	*0.02500*	PC	*0.07000*	*0.03500*	*0.06500*
Set 5				Set 6			
True	−1.00000	1.00000	0.01000	True	0.00000	1.00000	−0.02000
Est.	−1.10790	1.09168	0.01096	Est.	−0.08004	1.00048	−0.02225
PE	*−0.10790*	*0.09168*	*0.09644*	PE	*−0.08004*	*0.00048*	*−0.11228*
MSE	*0.20326*	*0.05613*	*0.00004*	MSE	*0.62895*	*0.17203*	*0.00012*
R	*0.91226*	*0.95387*	*1.00651*	R	*0.99892*	*0.99634*	*1.00064*
PC	*0.04000*	*0.03500*	*0.03500*	PC	*0.04000*	*0.01500**	*0.03000*

Table 3 continued

	$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{a 0}$		$\beta_{\mu 0}$	$\beta_{\mu 1}$	$\beta_{a 0}$	$\beta_{a 1}$
Set 7				Set 8				
True	0.00000	−0.10000	0.01000	True	0.00000	1.00000	0.01000	0.00000
Est.	0.00128	−0.10082	0.01063	Est.	0.00077	1.00048	0.01001	0.00001
PE	0.00128	−0.00816	0.06297	PE	0.00077	0.00048	0.00125	0.00001
MSE	0.13219	0.01727	0.00003	MSE	0.62895	0.05199	0.00990	0.00043
R	0.87391	0.92406	0.84904‡	R	0.91332	0.96609	0.89088	0.89118
PC	0.09000*	0.06000	0.07000	PC	0.06000	0.05000	0.06500	0.07000
Set 9				Set 10				
True	0.00000	−0.05000	0.01000	True	0.00000	1.00000	0.01000	−0.00050
Est.	−0.00644	−0.04536	0.01053	Est.	−0.02879	1.07080	0.00964	−0.00034
PE	−0.00644	0.09285	0.05272	PE	−0.02879	0.07080	−0.03630	0.31177†
MSE	0.11756	0.01536	0.00003	MSE	0.63080	0.05300	0.00944	0.00042
R	0.91001	0.95671	0.88175	R	0.92210	0.96712	0.92802	0.92117
PC	0.06500	0.06000	0.06000	PC	0.06000	0.04000	0.05500	0.05500
Set 11				Set 12				
True	0.00000	0.05000	0.01000	True	0.00000	1.00000	0.01000	0.00050
Est.	−0.02170	0.06540	0.01026	Est.	−0.06503	1.07034	0.00327	0.00219
PE	−0.02170	0.30797†	0.02556	PE	−0.06503	0.07034	−0.67317†	3.38064†
MSE	0.11091	0.01619	0.00002	MSE	0.60170	0.04791	0.00829	0.00037
R	0.92249	0.92412	0.94926	R	0.93258	0.98320	0.96776	0.96444
PC	0.07500	0.07000	0.05000	PC	0.03500	0.04000	0.04500	0.04500

Table 3 continued

	$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}		$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}	β_{a1}
Set 13				Set 14				
True	0.00000	−0.01000	0.01000	True	0.00000	1.00000	0.01000	−0.00100
Est.	−0.01224	−0.00296	0.01041	Est.	−0.01171	1.06854	0.01226	−0.00151
PE	*−0.01224*	*0.70368*†	*0.04105*	PE	*−0.01171*	*0.06854*	*0.22647*†	*−0.51078*†
MSE	*0.11167*	*0.01496*	*0.00003*	MSE	*0.60815*	*0.05565*	*0.00890*	*0.00040*
R	*0.92451*	*0.95938*	*0.93179*	R	*0.94561*	*0.96807*	*0.97644*	*0.97228*
PC	*0.06500*	*0.05000*	*0.05000*	PC	*0.04500*	*0.02000*	*0.04000*	*0.06000*
Set 15				Set 16				
True	0.00000	0.01000	0.01000	True	0.00000	1.00000	0.01000	0.00100
Est.	−0.00917	0.01714	0.01049	Est.	−0.06907	1.07059	0.00329	0.00276
PE	*−0.00917*	*0.71445*†	*0.04881*	PE	*−0.06907*	*0.07059*	*−0.67116*†	*1.75814*†
MSE	*0.10973*	*0.01486*	*0.00003*	MSE	*0.57545*	*0.04808*	*0.00839*	*0.00038*
R	*0.93033*	*0.96091*	*0.92939*	R	*0.94743*	*0.97578*	*0.95336*	*0.93931*
PC	*0.06500*	*0.06000*	*0.04500*	PC	*0.04500*	*0.03500*	*0.05000*	*0.05500*

$PE = \widehat{\theta} - \theta$ when $\theta = 0$; $^{\dagger}|PE| > 15\%$; $^{\star}PC > 0.02$ or $PC < 0.08$

Table 4 Simulation studies for the BGP, TBGP and ABGP models using LSE and Bayesian methods

	LSE					Bayesian				
BGP ($n = 100$)	λ		a			λ		a		
True	0.2000		0.9800			0.2000		0.9800		
Est.	0.2822		0.9575			0.2278		0.9794		
PE	*0.4109*†		*−0.0230*			*0.1390*		*−0.0006*		
MSE	*0.1065*		*0.0083*			*0.0143*		*0.0001*		
TBGP (Set 1)	$\beta_{\mu01}$	$\beta_{\mu02}$	β_{a01}	β_{a02}	T	$\beta_{\mu01}$	$\beta_{\mu02}$	β_{a01}	β_{a02}	T
True	−1.0000	1.0000	0.0100	−0.0100	101.00	−1.0000	1.0000	0.0100	−0.0100	101.00
Est.	−1.3007	1.0175	0.3928	−0.0103	100.92	−0.6215	1.0113	0.0766	−0.0105	100.11
PE	*0.3007*†	*0.0175*	*38.281*†	*0.0338*	*−0.0008*	*−0.3785*†	*0.0113*	*6.6588*†	*0.0467*	*−0.0088*
MSE	*3.1351*	*0.1564*	*0.6491*	*0.0001*	*0.9182*	*0.5060*	*0.2073*	*0.0396*	*0.0001*	*8.4370*
ABGP (Set 1)	$\beta_{\mu0}$	$\beta_{\mu1}$	β_{a0}			$\beta_{\mu0}$	$\beta_{\mu1}$	β_{a0}		
True	0.0000	1.0000	0.0100			0.0000	1.0000	0.0100		
Est.	−0.0187	1.1219	0.0117			0.0241	1.1006	0.0114		
PE	*−0.0187*	*0.1219*	*0.1673*†			*0.0241*	*0.1006*	*0.1423*		
MSE	*0.1924*	*0.0913*	*0.0001*			*0.1734*	*0.0574*	*0.0001*		
ABGP (Set 10)	$\beta_{\mu0}$	$\beta_{\mu1}$	β_{a0}	β_{a1}		$\beta_{\mu0}$	$\beta_{\mu1}$	β_{a0}	β_{a1}	
True	0.0000	1.0000	0.0100	−0.0005		0.0000	1.0000	0.0100	−0.0005	
Est.	−0.0812	1.1693	0.0047	0.0010		−0.2444	1.1202	−0.0264	0.0074	
PE	*−0.0812*	*0.1693*†	*−0.5319*†	*−2.9036*†		*−0.2444*†	*0.1202*	*−3.6404*†	*−15.708*†	
MSE	*0.8547*	*0.1840*	*0.0135*	*0.0006*		*0.6734*	*0.0691*	*0.0101*	*0.0004*	

$PE = \widehat{\theta} - \theta$ when $\theta = 0$. $^{\dagger}|PE| > 15\%$

4.2 The TBGP model

For simplicity, we set the number of trends $M = 2, n = 200$ and $T = T_2 = 101$. We vary the pair of parameters $(\beta_{\mu 01}, \beta_{\mu 02})$ for the mean function and $(\beta_{a01}, \beta_{a02})$ for the ratio function before and after the turning point T. Results from Table 2 show that the performance of ML estimators for the TBGP model as revealed by all the four GOF measures is satisfactory under all conditions.

4.3 The ABGP model

We set the covariate $z_{\mu i}$ to be cyclic with values 0, 1, 2, 3 in each cycle. The set of parameters chosen for the initial mean and ratio functions are $\beta_{\mu 0} = 0.00, \beta_{\mu 1} = 1.00, \beta_{a0} = 0.01, \beta_{a1} = 0.00$. We vary the value of one parameter in the set at a time to detect changes in the four GOF measures.

Again, Table 3 shows that the performance of ML estimators for the ABGP model is satisfactory. The only exception is on the parameters β_{a0} and β_{a1} when β_{a1} varies. It is clear that p_i in (13) for the ABGP model is more sensitive to changes in β_{a0} and β_{a1} as the ratio $a_i = \exp(\beta_{a0} + \beta_{a1} \ln i)$ in (13) is raised to the power of $i - 1$. This sensitivity of p_i to changes in β_{a0} and β_{a1} becomes more pronounced when the trend is long, for example, $n = 100$ in the simulation experiment. Hence the performance of ML estimators for β_{a0} and β_{a1} is better for smaller n.

5 Example

The three proposed BGP models are applied to two data sets and each model is estimated by three methods, namely the LSE, ML and Bayesian. We compare BGP models with the logistic model given by

$$p_i = \frac{\exp(\eta_i)}{1 + \exp(\eta_i)}$$

where η_i is a linear function of covariates. SE of parameters using the LSE and ML methods are given by the square root of the diagonal elements in the information matrix $-[SSE''(\widehat{\boldsymbol{\theta}}_{LSE})]^{-1}$ and $-[\ell''(\widehat{\boldsymbol{\theta}}_{ML})]^{-1}$, respectively and for Bayesian method, the posterior standard deviations.

To choose between models, we consider two goodness-of-fit (GOF) measures:

$$\text{Mean squares of error: } MSE = \frac{1}{n} \sum_{i=1}^{n} (w_i - p_i)^2, \tag{18}$$

$$\text{Posterior expected utility: } U = \frac{1}{n} \sum_{i=1}^{n} \ln[L_i(w_i | \widehat{\boldsymbol{\theta}})] \tag{19}$$

where the likelihood function $L_i(w_i | \widehat{\boldsymbol{\theta}})$, evaluated at $\widehat{\boldsymbol{\theta}}$, is given by (6), $\lambda_i = 1/\mu_i$, μ_i is given by (11) and a_i is given by (12). The measure U is simplified from U' (Walker and Gutiérrez-Peña 1999) which is defined as

$$U' = \frac{1}{Rn}\sum_{r=1}^{R}\sum_{i=1}^{n}\ln[L_{ir}(w_i\,|\,\boldsymbol{\theta}_r)]$$

where the likelihood function $L_{ir}(w_i\,|\,\boldsymbol{\theta}_r)$ is evaluated at the r-th set of posterior parameter estimates $\boldsymbol{\theta}_r$ and R is the number of posterior samples. Obviously, a model with the largest U and/or the least MSE is chosen to be the best model as it indicates the highest likelihood and/or the least errors of the data given the model. For each type of BGP model, Tables 5 and 6 show that ML estimates often give the largest U while LSE estimates often provide the least MSE because the former maximize the sum of $\ln[L_i(w_i\,|\,\hat{\boldsymbol{\theta}})]$ while the latter minimize the sum of $(w_i - p_i)^2$.
Apart from the two GOF measures, models can also be compared using the plots of observed W_i and predicted $P_i = E(W_i)$ in Figs. 4a, 5a, 6a and 7a and their sums, $SW_i = \sum_{i'=1}^{i} W_{i'}$ and $SP_i = \sum_{i'=1}^{i} P_{i'}$ in Figs 4b, 5b, 6b and 7b. If the data points are densely (sparsely) located along the line $y = 1$, P_i will be large (small) and SW_i and SP_i will rise up rapidly (slowly). In particular, if the points are located with equal density along $y = 1$, P_i will stay at a constant level and SW_i and SP_i will rise up linearly showing a very gentle trend. On the other hand, if the points are becoming more densely (sparsely) located along $y = 1$, P_i will rise up (drop down) and SW_i and SP_i will rise up along a concave (convex) curve.

5.1 Coal mining disasters data

5.1.1 The data

The Coal mining disasters data, reported in Andrews and Herzberg (1985) , contain 190 interarrival times between successive disasters in Great Britain. It was originally reported in Maguire et al. (1952) and was studied by Cox and Lewis (1966) in the analysis of trend. Afterward, the data were extended so that it covers the period from 15th May 1851 to 22nd March 1962, a total of 40,550 days or 112 years. Lam (1992b), Lam and Chan (1998), Lam et al. (2004) and Chan et al. (2004) analyzed the data using GP models and showed that the interarrival times between successive disasters follow an increasing trend.
In this study, we define the outcomes W_i to be the indicator of whether a disaster occurs during the i-th quarter of 112 years. We have totally 448 observations. In reality, binary outcomes of hazards or failures for a monitored system often arise. If the likelihood of a system failure is high during certain period regardless of the intensity of failures, certain control policies should be implemented. Hence models for binary trend data are useful for the studies of reliability and system maintenance.

5.1.2 Result and comment

Since the data contain no information of covariate effects, we adopt the mean function $\mu_i = \exp(\beta_{\mu 0})$ for all three BGP models in (17). Plot of SW_i in Fig. 4b shows the presence of two trends: SW_i rises nearly linearly before $i = 147$ but it rises at a lower

Table 5 Parameter estimates and standard errors in italic for the coal mining disaster data

Method		LSE			ML			Bay		
Model	Par.	Est.	SE	GOF	Est.	SE	GOF	Est.	SE	GOF
BGP	$\beta_{\mu 0}$	0.5537	*0.2015*	0.1923	0.5292	*0.0871*	0.1923	0.5274	*0.1292*	0.1926
	a_0	1.0033	*0.00089*		1.0032	*0.0004*	*−0.5682*	1.0030	*0.00049*	*−0.5689*
TBGP	$\beta_{\mu 01}$	0.3139	*0.2703*	0.1870	0.3132*	*0.2376*	0.1900	0.3931	*0.2440*	0.1947
	$\beta_{\mu 02}$	−0.2666	*0.6105*		−0.2539*	*0.1454*	*−0.5568**	−0.2452	*0.1652*	*−0.5606*
	a_{01}	0.9979	*0.0029*		0.9979*	*0.0029*		0.9992	*0.0030*	
	a_{02}	1.0014	*0.0033*		1.0015*	*0.0008*		1.0020	*0.00093*	
	T	147			147*			151.50	*10.76*	
ABGP	$\beta_{\mu 0}$	0.6817	*0.0186*	0.1922	0.6860	*0.0105*	0.1922	0.7216	*0.2406*	0.1922
	β_{a0}	0.0088	*0.00068*		0.0094	*0.0004*	*−0.5678*	0.0104	*0.0078*	*−0.5678*
	β_{a1}	−0.0009	*0.00039*		−0.0010	*0.0002*		−0.0011	*0.0012*	
Logistic	$\beta_{\mu 0}$				2.3881	*0.6218*	0.1971			
	$\beta_{\mu t}$				−0.6282	*0.1169*	*−0.5793*			
	$\beta_{\mu p}$				−0.0872	*0.2374*				

The chosen model is indicated by ‘*’
The two GOF measures are MSE and U (in italic), respectively

Table 6 Parameter estimates and standard errors in italic for the methadone clinic data

Patient		119			509			513		
Model	Par.	Est.	SE	GOF	Est.	SE	GOF	Est.	SE	GOF
BGP	$\beta_{\mu 0}$	3.9054	*3.4678*	0.1839	−0.2126	*0.9532*	0.2193	0.2010	*0.4637*	0.1614
LSE	$\beta_{\mu d}$	−0.0358	*0.0534*		−0.0132	*0.0271*		−0.0005	*0.0075*	
	a_0	1.0517	*0.0376*		0.9768	*0.0155*		1.0042	*0.0018*	
BGP	$\beta_{\mu 0}$	2.8574	*1.7092*	0.1850	−0.0286	*0.6382*	0.2196	0.2127	*0.2854*	0.1617
ML	$\beta_{\mu d}$	−0.0211	*0.0304*	*−0.5599*	−0.0155	*0.0191*	*−0.6324*	0.0004	*0.0043*	*−0.4902*
	a_0	1.0494	*0.0218*		0.9773	*0.0104*		1.0048	*0.0010*	
BGP	$\beta_{\mu 0}$	3.6450	*1.6660*	0.1869	−0.0251	*0.6154*	0.2198	0.2269	*0.2889*	0.1619
Bay	$\beta_{\mu d}$	−0.0399	*0.0273*	*−0.5668*	−0.0151	*0.0181*	*−0.6325*	0.0005	*0.0044*	*−0.4903*
	a_0	1.0330	*0.0132*		0.9774	*0.0096*		1.0050	*0.0011*	
TBGP	$\beta_{\mu 0}$	5.1244	*5.4491*	0.1631	0.7958	*1.3040*	0.2003	1.6594	*0.7396*	0.1321
LSE	$\beta_{\mu d}$	−0.0525	*0.0848*		−0.0147	*0.0322*		−0.0132	*0.0097*	
	a_{01}	1.0562	*0.0703*		1.0235	*0.0235*		1.0775	*0.0442*	
	a_{02}	2.5479	*3.2227*		0.9666	*0.0341*		1.0160	*0.0050*	
	T	35			46			69		
TBGP	$\beta_{\mu 0}$	0.8360*	*1.6993*	0.1720	0.9574*	*0.5422*	0.2004	0.9847*	*0.2138*	0.1354
ML	$\beta_{\mu d}$	−0.0027*	*0.0263*	*−0.4973**	−0.0185*	*0.0102*	*−0.5823**	−0.0081*	*0.0043*	*−0.4336**
	a_{01}	0.5256*	*0.2842*		1.0237*	*0.0078*		1.0318*	*0.0044*	
	a_{02}	1.0409*	*0.0106*		0.9627*	*0.0081*		1.0105*	*0.0008*	
	T	13*			46*			69*		
TBGP	$\beta_{\mu 0}$	3.7010	*1.9470*	0.1865	0.5311	*0.7151*	0.2096	0.9589	*0.3507*	0.1373
Bay	$\beta_{\mu d}$	−0.0510	*0.0338*	*−0.5609*	−0.0148	*0.0178*	*−0.6085*	−0.0076	*0.0050*	*−0.4373*
	a_{01}	0.9856	*0.0263*		1.0100	*0.0217*		1.0310	*0.0061*	
	a_{02}	1.0290	*0.0190*		0.9686	*0.0142*		1.0100	*0.0017*	
	T	21.57	*12.06*		36.99	*12.00*		67.75	1.7080	
ABGP	$\beta_{\mu 0}$	5.4410	*5.4005*	0.1706	1.2674	*1.3920*	0.2054	0.1592	*0.5086*	0.1572
LSE	$\beta_{\mu d}$	−0.1208	*0.1423*		−0.0329	*0.0346*		−0.0117	*0.0107*	
	β_{a0}	−0.2985	*0.3903*		0.0233	*0.0332*		−0.0119	*0.0109*	
	β_{a1}	0.0062	*0.0072*		−0.00060	*0.00040*		0.00007	*0.00005*	
ABGP	$\beta_{\mu 0}$	4.5101	*2.2267*	0.1729	1.2927	*0.8963*	0.2070	NA	NA	
ML	$\beta_{\mu d}$	−0.0864	*0.0550*	*−0.5307*	−0.0300	*0.0215*	*−0.5990*			
	β_{a0}	−0.1674	*0.1365*		0.0355	*0.0209*				
	β_{a1}	0.0036	*0.0023*		−0.00071	*0.00025*				
ABGP	$\beta_{\mu 0}$	5.6120	*2.4100*	0.1742	1.7640	*0.9852*	0.2063	0.0749	*0.2939*	0.1587
Bay	$\beta_{\mu d}$	−0.1078	*0.0594*	*−0.5341*	−0.0451	*0.0245*	*−0.5960*	−0.0034	*0.0049*	*−0.4846*
	β_{a0}	−0.1976	*0.1464*		0.0259	*0.0226*		−0.0025	*0.0041*	
	β_{a1}	0.0041	*0.0024*		−0.00071	*0.00027*		0.00003	*0.00001*	
Logistic	$\beta_{\mu 0}$	3.8942	*0.9481*	0.2048	−2.0754	*0.1798*	0.2146	−0.0916	*0.0991*	0.1470
ML	$\beta_{\mu d}$	−0.0694	*0.0075*	*−0.5975*	0.0178	*0.0057*	*−0.6230*	0.0061	*0.0018*	*−0.4616*
	$\beta_{\mu t}$	−0.0539	*0.1101*		0.1064	*0.0451*		−0.4222	*0.0243*	
	$\beta_{\mu p}$	0.8384	*0.3958*		1.1343	*0.3490*		1.6067	*0.1801*	

The chosen model is indicated by ‘*’

The two GOF measures are MSE and U (in italic), respectively

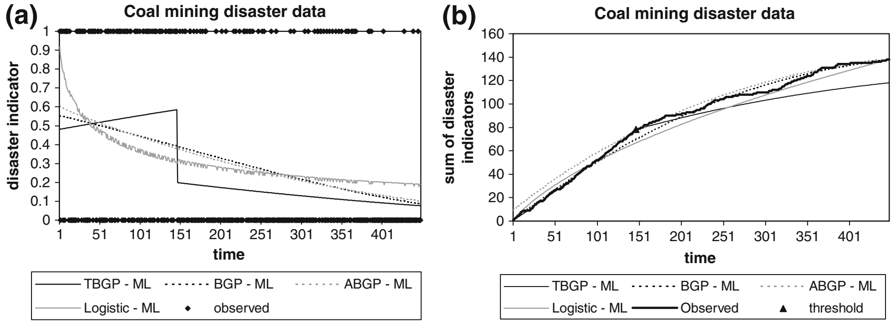

Fig. 4 **a** Observed and predicted disaster indicator and **b** their sum in the coal mining diseaster data

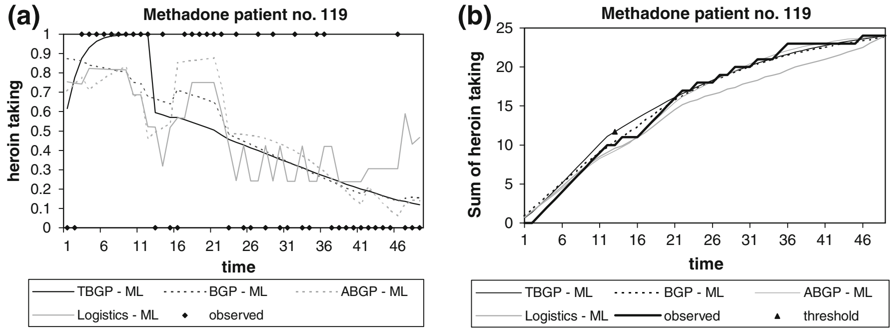

Fig. 5 **a** Observed and predicted heroin use and **b** their sum for patient 119 in the methadone clinic data

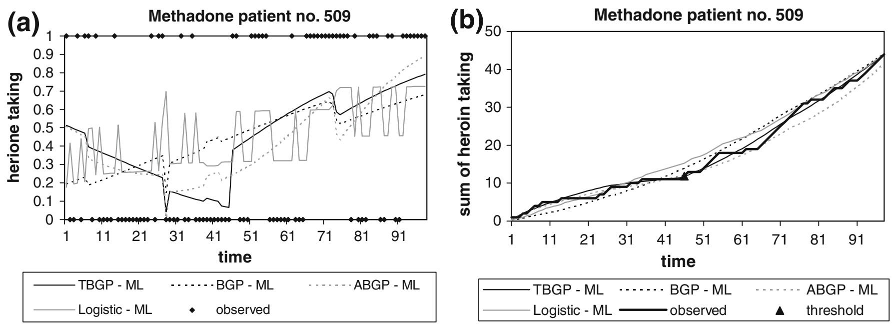

Fig. 6 **a** Observed and predicted heroin use and **b** their sum for patient 509 in the methadone clinic data

rate and subject to higher variability from $i = 147$. Hence we set $M = 2$ for the TBGP model. For model comparison, logistic model with a lag-1 autoregressive (AR1) term:

$$\text{logit}(p_i) = \beta_{\mu 0} + \beta_{\mu t} \ln i + \beta_{\mu p} w_{i-1}$$

to allow autocorrelation is also fitted to the data. Table 5 shows that the three estimation methods give similar estimates for each type of BGP model. Across the three BGP models, TBGP model using ML and LSE methods give the best fit according to U

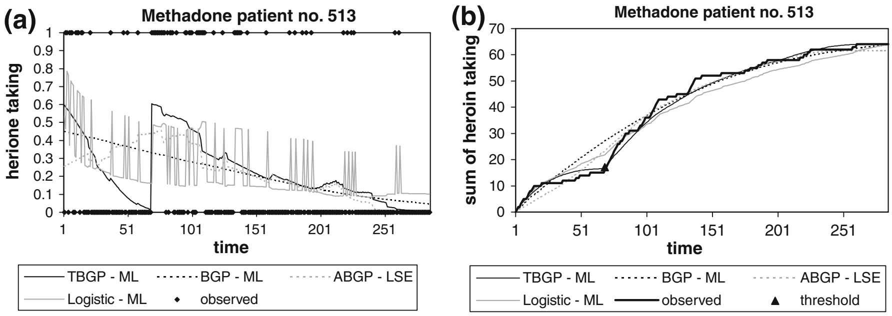

Fig. 7 **a** Observed and predicted heroin use and **b** their sum for patient 513 in the methadone clinic data

and MSE, respectively. TBGP model using the ML method is the chosen model as its MSE is the second best. For this model, threshold occurs at time $\widehat{T}_{ML} = 147$ and the two ratio parameters are $\widehat{a}_{ML,01} = 0.998$ and $\widehat{a}_{ML,02} = 1.002$. They are close to 1 and statistically significant showing a gentle increasing trend of P_i from $p_1 = 0.481$ and then a gentle decreasing trend of P_i from a lower level of $p_{147} = 0.276$ after $i = 147$. Thresholds using other methods are close to $\widehat{T}_{ML}$ ($\widehat{T}_{LSE} = 147$; $\widehat{T}_B = 151.5$).

The trends of the observed W_i and its predicted P_i for various models are shown in Fig. 4a. All P_i are smooth and drop continuously except the logistic and TBGP models. The former drops rapidly at the beginning according to the logit function and it frustrates 'up and down' due to the AR1 term in the model. The latter, the chosen TBGP model, is set to have two different ratios a_i giving an increasing and then a decreasing trends of P_i. It provides nearly perfect fit of SW_i during the increasing trend. The P_i for both BGP and ABGP models are very similar: a_i in the ABGP model decreases slowly from $a_1 = 1.009$ to $a_{448} = 1.003$ whereas $a_i = 1.003$ in the BGP model when ML method is used. This is probably because the increasing trend before T is not strong. Hence the ABGP model detects an overall decreasing trend rather than an increasing and then a decreasing trend. The logistic model gives the worst fit according to both MSE and U. This shows that the one component logistic model fails to account for the trend effect in the data even though an AR1 term is included to account for autocorrelation.

5.1.3 The test

To test for the existence of trends in the data, we test whether $a = 1$ in the BGP model using the ML estimator. Using the limiting distribution of (16), the asymptotic standard error $ASE(\widehat{a}_{ML,0})$ using (27) is found to be 0.00054. The test statistic is

$$Z = \frac{\widehat{a}_{ML,0} - 1}{ASE(\widehat{a}_{ML,0})} = 5.88$$

which is significant. Hence we conclude that there is a trend in the data.

To test for the significance of the TBGP and ABGP models against the BGP model, the LRT is used. The test statistics are

$$\chi_{TBGP} = 2(448)(U_{TBGP} - U_{BGP}) = 10.21 \quad \text{and}$$
$$\chi_{ABGP} = 2(448)(U_{ABGP} - U_{BGP}) = 0.40$$

and their p-values are 0.006 and 0.819, respectively, showing that the TBGP model should be chosen.

5.2 Methadone clinic data

5.2.1 The data

The data contain results of weekly urine drug screens, positive or negative for heroin use, from 136 patients in a methadone clinic in Western Sydney in 1986. Previous analyzes were performed using a restricted data set of all 136 patients including only results of urine screens collected in the first 26 weeks of treatment to avoid the drop-out effect of patients (Chan et al. 1997, 1998). The aim of this analysis is to study the trend of heroin use over time and we use the whole series of observations from three selected patients (patient number 119, 509 and 513) who exhibit special trend patterns as shown in the plots of W_i in Figs. 5a, 6a and 7a.

In Fig. 5a, patient 119 ($n = 49$) shows a high density of data points along $y = 1$ before $i = 13$ and a decreasing density of points along $y = 1$ after $i = 13$. Hence SW_i increases linearly and then along a convex curve after $i = 13$. In Fig. 6a, patient 509 ($n = 98$) shows a sparse density of points along $y = 1$ at the beginning. Then the density decreases further and then increases again after $i = 45$. Accordingly SW_i increases along a convex curve and then along a concave curve after $i = 45$. Lastly, for patient 513 ($n = 285$), Fig. 7a shows that the density of points decreases along $y = 1$ from $i = 1$ and $i = 69$ but it decreases at a slower rate from $i = 69$. Consequently, SW_i increases along two convex curves. We demonstrate that BGP models can model these trends well.

5.2.2 Result and comment

Previous studies showed that methadone dose d_i in mg and the log of treatment time, $\ln i$, in weeks are significant treatment factors. Interaction effect between dose and time was found to be insignificant and hence was dropped subsequently. As a result, we set the initial mean function to be

$$\mu_i = \exp(\beta_{\mu 0} + \beta_{\mu d}\, d_i) \tag{20}$$

to account for the dose effect ($\beta_{\mu d}$) and let the ratio a_i to account for the time effect for all BGP models. With the incorporation of dose effects, μ_i is no longer a constant over time. As Figs. 5, 6 and 7 show the presence of 2 trends for the three patients, we set $M = 2$ for the TBGP models and let the ratio parameter a_{0m} to change after the threshold T but the initial mean function μ_i remains unchanged in (20). For the ABGP models, we set $a_i = \exp(\beta_{a0} + \beta_{a1} \ln i)$. The logistic model

$$\mathrm{logit}(p_i) = \beta_{\mu 0} + \beta_{\mu d}\, d_i + \beta_{\mu t} \ln i + \beta_{\mu p}\, w_{i-1}$$

as in Chan et al. (1997, 1998) is also adopted for comparison.

Previous studies on all patients showed that the dose effect was always significant and negative. In the current studies in which analyzes are done separately for each patient, the dose effect is insignificant for most BGP models. One possible reason is that part of the dose effect is accounted for by the trend effect in the ratio a_i. Results from Table 6 show that across all BGP and logistic models, TBGP model using ML method is always the best model according to U for all the three patients. The thresholds T are 13, 46 and 69, respectively, for the three patients.

For patient 119, the two ratio parameters for the chosen TBGP model are $\widehat{a}_{ML,01} = 0.526$ and $\widehat{a}_{ML,02} = 1.041$, respectively, indicating an increasing and then a decreasing trends of P_i, the heroin use, before and after $i = \widehat{T}_{ML} = 13$. Thresholds using other methods are $\widehat{T}_{LSE} = 35$ and $\widehat{T}_B = 21.6$, respectively. In fact $i = 35$ is the second choice of T according to U for the ML method. BGP and ABGP models give similar results: both show a decreasing trend of P_i in general. The ratio $a_i = 1.049$ for the BGP model whereas a_i increases from $a_1 = 0.849$ to $a_{49} = 1.009$ in the ABGP model. According to a_i in the ABGP model, P_i should be increasing at a decreasing rate. However, P_i shows a general decreasing trend. This is possibly because the trend effect is out-weighted by the dose effect ($\beta_{\mu d}$) which acts adversely to the increasing trend. For the chosen model, the dose effect is insignificant and hence the trends of P_i are clear.

For patient 509, the chosen TBGP model shows a decreasing and then an increasing trends of P_i after $i = \widehat{T}_{ML} = 46$ since the two ratio parameters are $\widehat{a}_{ML,01} = 1.024$ and $\widehat{a}_{ML,01} = 0.963$, respectively. Thresholds using other methods are $\widehat{T}_{LSE} = 46$ and $\widehat{T}_B = 37$, respectively. Again $i = 37$ is the second choice of T according to U for the ML method. The P_i in BGP model is increasing in general as $a_i = 0.977$. On the other hand, a_i decreases from $a_1 = 1.035$ to $a_{98} = 0.967$ in ABGP model. As a_i passes through 1, P_i in ABGP model follows a U-shape, in general. The dose effect is not negligible for all BGP models. Hence P_i curves often show a 'kink' whenever the dosage changes levels.

Lastly, for patient 513, the two ratio parameters for the chosen TBGP model are $\widehat{a}_{ML,01} = 1.032$ and $\widehat{a}_{ML,02} = 1.011$, respectively showing two decreasing trends throughout the period. However, P_i drops to zero at a faster rate before $i = 69$. Thresholds using other methods are $\widehat{T}_{LSE} = 69$ and $\widehat{T}_B = 67.8$ which are close. Note that parameters for the ABGP model using ML method do not converge and hence P_i and SP_i in Fig. 7 are estimated from LSE method. Experience shows that parameter β_{a1} which accounts for different non-monotone trend patterns will be unstable when the ratio function a_i does not apply to the whole time series, particularly when the time series is long ($n = 285$). For BGP model, the ratio $a_i = 1.005$ indicates a general downwards trend of P_i whereas a_i in ABGP model using LSE method changes slightly from $a_1 = 0.988$ to $a_{285} = 1.008$ indicating a gentle inverted U-shape trend for P_i. The dose effect for the BGP model is very weak and hence P_i drops nearly continuously. Logistic model gives nearly the worst fit to all three data.

6 Conclusion

Original GP model was limited to the modeling of inter-arrival times of a series of events where the underlying RP has a constant mean μ and a constant ratio a. Studies on this model were focused on inference and application to reliability and maintenance problems. This paper generalizes the GP model to a broader class of models with different modeling methodologies, methods of inference and fields of application.

Firstly, we consider a wider class of data and binary data is the focus in this paper. Following the methodology, extension to other type of data, say the Poisson counts, within the framework of GP model, is straightforward (Wan and Chan 2009, 2010; Chan et al. 2010a,b). Secondly, three methods of inference, namely the LSE, ML and Bayesian methods, are derived for the basic and extended BGP models which incorporate threshold and covariate effects to the initial level μ_i and/or the ratio a_i. Lastly, BGP models are applied to two different types of data, the coal mining disasters data for industrial application and the methadone clinic data for medical application.

Results from simulation studies show that all three methods of inference provide good parameter estimates. Performance of BGP models is also demonstrated through real data analyses. Results show that BGP models perform better than the logistic model in all analyzes and they also identify trend patterns clearly through the ratio parameters. Among the three BGP models, TBGP model is often chosen. Regarding statistical inference, ML method provides the best model fit according to U and LSE method according to MSE. Bayesian method also provides similar model fit. Model selection can be performed using the likelihood ratio test on hierarchical BGP models. Significance of trend patterns and covariate effects can be tested using the limiting distribution of the corresponding ML estimates.
In summary, proposed BGP models offer a good modeling alternative to longitudinal binary data and hence widen the scope of application of GP models considerably.

Acknowledgments We gratefully acknowledge Dr. James Bell of the Drug and Alcohol Unit, Prince of Wales Hospital, Sydney, Australia, for kindly providing the data set on Methadone Maintenance Treatment Programme analyzed in this paper.

Appendix A: First and second order derivatives of the SSE for LSE method

We set $E(W_i) = p_i$ in the three BGP models defined as

$$\text{BGP model}: p_i = \exp[-\lambda_i a_i^{i-1}],\ a_i = a_0,\ \lambda_i^{-1} = \lambda_0^{-1} = \exp(\beta_{\mu 0}),$$

$$\text{TBGP model}: \quad p_i = \exp[-\lambda_i a_i^{i-T_m}],\ a_i = a_{0m},$$

$$\lambda_i^{-1} = \lambda_{0m}^{-1} = \exp(\beta_{\mu 0m}) \quad \text{if } T_m \le i < T_{m+1},$$

$$\text{ABGP model}: \quad p_i = \exp[-\lambda_i a_i^{i-1}],\ a_i = \exp\left(\beta_{a0} + \sum_{k=1}^{q_a} \beta_{ak} z_{aki}\right),$$
$$\lambda_i^{-1} = \exp\left(\beta_{\mu 0} + \sum_{k=1}^{q_\mu} \beta_{\mu k} z_{\mu ki}\right).$$

Alternatively, the initial mean function for TBGP model can be written as $\lambda_i^{-1} = \exp(\beta_{\mu 0} + \sum_{k=1}^{q_\mu} \beta_{\mu k} z_{\mu ki}), \beta_{\mu 0} = 0, \beta_{\mu k} = \beta_{\mu 0k}, q_\mu = M, z_{\mu ki} = I(T_k \le i < T_{k+1}) = I_{ki}$ and $k = m = 1, \ldots, M$.

Then the sum of squared errors $SSE(\boldsymbol{\theta})$ and its first and second order derivatives are

$$SSE(\boldsymbol{\theta}) = \sum_{i=1}^{n} (w_i - p_i)^2,$$
$$\frac{\partial SSE(\boldsymbol{\theta})}{\partial \beta_{jk}} = -2 \sum_{i=1}^{n} (w_i - p_i) z_{jki}^{*}\, p_i \ln p_i, \tag{21}$$
$$\frac{\partial^2 SSE(\boldsymbol{\theta})}{\partial \beta_{j_1 k_1} \partial \beta_{j_2 k_2}} = -2 \sum_{i=1}^{n} z_{j_1 k_1 i}^{*} z_{j_2 k_2 i}^{*} p_i \ln p_i [(w_i - p_i)(1 + \ln p_i) - p_i \ln p_i] \tag{22}$$

where $j, j_1, j_2 = a, \mu;\ k, k_1, k_2 = 0, \ldots, q_j; z_{j0i} = 1, z_{aki}^{*} = (i-1) z_{aki}$ and $z_{\mu ki}^{*} = -z_{\mu ki}$.

When $a_i = a_0$ or a_{0k}, they replace β_{a0} and β_{ak}, respectively, $z_{a0i}^{*} = (i-1)/a_0$ and $z_{aki}^{*} = (i - T_k) I_{ki} / a_{0k}$ in (21). Moreover (21) becomes

$$\frac{\partial^2 SSE(\boldsymbol{\theta})}{\partial a_{0k}^2} = -2 \sum_{i=1}^{n} z_{aki}^{*} p_i \ln p_i \times \left[z_{aki}^{\star} (w_i - p_i) + z_{aki}^{*} (w_i - p_i) \ln p_i - z_{aki}^{*}\, p_i \ln p_i \right] \tag{23}$$

where $z_{aki}^{\star} = (i - T_k - 1) I_{ki} / a_{0k}$. When the parameter is a_0 for BGP model, set $k = 0$ and $T_k = T_1 = 1$ in (23).

Appendix B: First and second order derivatives of the log-likelihood function for ML method

The log-likelihood function $\ell(\boldsymbol{\theta})$ and its first and second order derivatives are

$$\ell(\boldsymbol{\theta}) = -\sum_{i=1}^{n} [w_i \ln p_i + (1 - w_i) \ln(1 - p_i)],$$
$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \beta_{jk}} = \sum_{i=1}^{n} z_{jki}^{*} \ln p_i \left(w_i - \frac{1 - w_i}{1 - p_i} p_i \right), \tag{24}$$

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \beta_{j_1k_1} \partial \beta_{j_2k_2}} = \sum_{i=1}^{n} w_i z^*_{j_1k_1i} z^*_{j_2k_2i} \ln p_i - \sum_{i=1}^{n} \frac{1-w_i}{1-p_i} z^*_{j_1k_1i} z^*_{j_2k_2i} \, p_i \ln p_i \left(\frac{p_i \ln p_i}{1-p_i} + \ln p_i + 1 \right), \tag{25}$$

where $j, j_1, j_2 = a, \mu;\ k, k_1, k_2 = 0, \ldots, q_j$; $z^*_{aki} = (i-1)z_{aki}$, $z^*_{\mu ki} = -z_{\mu ki}$ and $z_{j0i} = 1$. When $a_i = a_0$ or a_{0k} for TBGP model, they replace β_{a0} and β_{ak}, respectively, $z^*_{a0i} = (i-1)/a_0$ and $z^*_{aki} = (i-T_k)I_{ki}/a_{0k}$ When $\lambda_i = \lambda$, λ replaces $\beta_{\mu 0}$, and $z^*_{\mu i} = \frac{1}{\lambda}$. Moreover (25) becomes

$$\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial a_{0k}^2} = \sum_{i=1}^{n} w_i z^*_{aki} z^\star_{aki} \ln p_i - \sum_{i=1}^{n} \frac{1-w_i}{1-p_i} z^*_{aki} z^*_{aki} \, p_i \ln p_i \left(\frac{p_i \ln p_i}{1-p_i} + \ln p_i + 1 - \frac{1}{i-1} \right) \tag{26}$$

where $z^\star_{aki} = (i - T_k - 1) I_{ki}/a_{0k}$. When the parameter is a_0 for BGP model, set $k = 0$ and $T_k = T_1 = 1$ in (25).

Appendix C: Limiting distribution for the ML estimators

Taking $E(w_i) = p_i$ in (25), we have

$$s_{uv} = -E\left[\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_u \partial \theta_v} \right] = \sum_{i=1}^{n} z^*_{ui} z^*_{vi} \frac{p_i (\ln p_i)^2}{1-p_i} \tag{27}$$

where s_{uv} is the element in the u-th row and v-th column of the covariance matrix $\boldsymbol{\Sigma}$. The parameter θ_u and θ_v may represent any of the model parameters β_{jk}, where $j = a, \mu;\ k = 0, \ldots, q_j$ and $k = 0, \ldots, M$ for TBGP model. Accordingly, the variables z^*_{ui} and z^*_{vi} may represent any of the z_{jki}.

Appendix D: WinBUGS code for the BGP models

```
model
{ c <- 10000
for (i in 1:N){
ai1[i] <- pow(a,i-1)      # for BGP
ai1[i] < - step(turn-i)*pow(a01,i-1)+(1-step(turn-i))* pow(a02,i-turn)
# for TBGP
ai1[i] <- pow(exp(betaa0+betaa1*log(time[i])),time[i]-1)    # for ABGP
mu[i] <- 1/lam     # for BGP
```

```
mu[i] <- step(turn-i)/lam1+(1-step(turn-i))/lam2     # for TBGP, M=2
mu[i] <- exp(betam0+betam1*z[i])     # for ABGP
e[i] <- exp(-1*ai1[i]/mu[i])
ones[i] <- 1
ones[i] ~ dbern( p[i] )
p[i] <- pow(e[i],w[i])*pow(1-e[i],1-w[i]) / c
}
lam ~ dgamma(0.001,0.001)    # for BGP
a ~ dunif(0.95,1.05)     # same
lam1 ~ dgamma(0.001,0.001)     # for TBGP, M=2
lam2 ~ dgamma(0.001,0.001)     # same
a01 ~ dunif(0.95,1.05)     # same
a02 ~ dunif(0.95,1.05)     # same
turn ~ dunif(150,200)     # same
betam0 ~ dnorm(0,0.00001)     # for ABGP
betam1 ~ dnorm(0,0.00001)     # same
betaa0 ~ dnorm(0,0.00001)     # same
betaa1 ~ dnorm(0,0.00001)     # same
}
```

Note: when one model is run, comment commands for other models by inserting # in front.

References

Andrews DF, Herzberg AM (1985) Data. Springer, New York

Ascher H, Feingold H (1981) Repairable systems reliability. Marcel Dekker, New York

Berman M (1981) Inhomogeneous and modulated gamma processes. Biometrika 68:143–152

Berman M, Turner TR (1992) Approximate point process likelihoods with GLIM. Appl Stat 41:31–38

Chan JSK, Kuk AYC, Bell J (1997) A Likelihood approach to analysing longitudinal bivariate binary data. Biom J 39(4):409–421

Chan JSK, Kuk AYC, Bell J, McGilchrist C (1998) The analysis of methadone clinic data using marginal and conditional logistic models with mixture or random effects. Aust N Z J Stat 40:1–10

Chan JSK, Lam Y, Leung DYP (2004) Statistical inference for geometric processes with gamma distributions. Comput Stat Data Anal 47:565–581

Chan JSK, Yu PLH, Lam Y, Ho APK (2006) Modeling SARS data using threshold geometric process. Stat Med 25:1826–1839

Chan JSK, Lam CPY, Yu PLH, Choy STB (2010a) Bayesian conditional autoregressive geometric process model for range data. submitted to journal

Chan JSK, Wan WY, Lee CK, Lin CK, Yu PLH (2010b) Predicting drop-out and committed first time blood donors: the use of Poisson GP model. submitted to journal

Cox DR, Lewis PA (1966) The statistical analysis of series of events. Mathuen, London

Feller W (1949) Fluctuation theory of recurrent events. Trans Am Math Soc 67:98–119

Gilks WR, Richardson S, Spiegelhalter DJ (1996) Markov chain Monte Carlo in practice. Chapman and Hall, UK

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109

Lakey MJ, Rigdon SE (1992) The modulated power law process. In: Proceedings of the 45th annual quality congress, pp 559–563

Lam Y (1988a) A note on the optimal replacement problem. Adv Appl Probab 20:479–482

Lam Y (1988b) Geometric process and replacement problem. Acta Math Appl Sin 4:366–377

Lam Y (1992a) Optimal geometric process replacement model. Acta Math Appl Sin 8:73–81
Lam Y (1992b) Nonparametric inference for geometric processes. Commun Stat Theory Methods 21:2083–2105
Lam Y (2007) The geometric process and its application. World Scientific Publishing Co. Pte. Ltd, Singapore
Lam Y, Chan JSK (1998) Statistical inference for geometric processes with lognormal distribution. Comput Stat Data Anal 27:99–112
Lam Y, Zhang YL (1996) Analysis of a two-component series system with a geometric process model. Nav Res Logistics 43:491–502
Lam Y, Zhang YL, Zheng YH (2002) A geometric process equivalent model for a multistate degenerative system. Eur J Oper Res 142:21–29
Lam Y, Zhu LX, Chan JSK, Liu Q (2004) Analysis of data from a series of events by a geometric process model. Acta Math Appl Sin, English Series 20:263–282
Lawless JF, Thiagarajah K (1996) A point-process model incorporating renewals and time trends, with application to repairable systems. Technometric 38:131–138
Lindqvist BH, Elvebakk GE, Heggland K (2003) The trend-renewal process for statistical analysis of repairable systems. Technometric 45:31–44
Maguire BA, Pearson ES, Wynn AHA (1952) The time intervals between industrial accidents. Biometrika 39:168–180
Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH (1953) Equations of state calculations by fast computing machines. J Chem Phys 21:087–1091
Smith AFM, Roberts GO (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. J R Stat Soc Series B 55:3–23
Spiegelhalter D, Thomas A, Best NG, Lunn D (2004) Bayesian inference using Gibbs sampling for Window version (WinBUGS) Version 1.4.1. The University of Cambridge. http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml
Walker SG, Gutiérrez-Peña E (1999) Robustifying Bayesian procedures. Bayesian Stat 6:685–710
Wan WY, Chan JSK (2009) A new approach for handling longitudinal count data with zero inflation and overdispersion: Poisson geometric process model. Biom J 51:556–570
Wan WY, Chan JSK (2010) Bayesian analysis of robust Poisson geometric process model using heavy-tailed distributions. submitted to journal