Abstract of thesis entitled

ANALYSIS OF POISSON COUNT DATA USING GEOMETRIC PROCESS MODEL

Submitted by

WAN WAI YIN

for the degree of Master of Philosophy

at The University of Hong Kong

in November 2006

A Poisson Geometric Process (PGP) model is proposed to study Poisson count data with trends. The model is extended from the Geometric Process (GP) model first introduced by Lam (1988) in the modeling of monotone trend for positive continuous data. A stochastic process (SP) $\{X_t, t = 1, 2, \dots\}$ is a GP if there exists a positive real number $a > 0$ such that $\{Y_t = a^{t-1}X_t\}$ generates a renewal process (RP) with a mean μ and a ratio a.

In the PGP model, it was assumed that the observed count data $\{W_t\}$ follow a Poisson distribution with a mean $\lambda_t = X_t$ where $\{X_t\}$ is a latent GP and the corresponding RP $\{Y_t\}$ follows exponential distribution with a mean $E(Y_t) = \mu_t$. Alternatively the PGP model was simplified by setting $\lambda_t = \mu_t/a_t^{t-1}$. The two versions of the PGP model with different shapes of distribution are suitable for modeling different types of data.

The PGP model was also extended to allow for covariate effects in the data, to fit data with multiple trends, to incorporate mixture effects to account for the population heterogeneity and clustering, and also to solve the problem of "excess zeros" in the data.

These extended models were implemented using different estimation methods and were demonstrated and assessed through simulation studies and real life applications in diverse areas. As a whole, the PGP models and its extended models take into account a rather pronounced dependence structure and the overdispersion in the time series of count data. It surely would become a viable alternative to the traditional time series model for count data in the future.

ANALYSIS OF POISSON COUNT DATA USING GEOMETRIC PROCESS MODEL

by

WAN WAI YIN

A thesis submitted in partial fulfillment of the requirements for

the Degree of Master of Philosophy

at The University of Hong Kong.

November 2006

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.

Signed _____

WAN Wai Yin

Acknowledgements

I would like to express my deep and sincere gratitude to my supervisor, Dr. Jennifer S. K. Chan, at the Department of Statistics and Actuarial Science, The University of Hong Kong. Her understanding, encouraging, inspirational and personal guidance have provided a good basis for the present thesis. She has been a role model in her diligence in research, enthusiasm in teaching and ingenuity of ideas, to a budding beginner who sometimes gets lost in midst of difficulties. I would also like to thank all the friends and staff in the Department for their support. Without anyone of them, I could not have such wonderful years.

Most of all, I owe my family for always believing in me and encouraging me to achieve my goals. Their endless love and support have been invaluable in helping me to focus on my academic pursuits.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main focus of this thesis is the extension of Geometric Process (GP) model to count data. In this chapter, we first explain why count data is our focus. Then, the definitions, properties, and advantages of the GP models would be discussed. Lastly, other related models would be studied, in particular, those models for the count data.

1.1 Background

A time series of count data which exhibits monotone increasing or decreasing trend patterns are common. In epidemiology, we observe the daily number of deaths in an epidemic outbreak increases as the epidemic spreads; in engineering, we find that the number of failures during a fixed period of an operating system demonstrates a monotone rising trend; in business, we discover the sales of a new electronic product increase at the beginning of the promotion period, but finally drop when other new products enter into the market. The list of areas in which time series are observed and analyzed is endless.

The simplest way to model time series count data is the use of the Poisson dis-

tribution, in which the mean, conditional on past observations, is autoregressive. However, the model is equidispersed because the variance is equal to the mean. Hence, it may not be appropriate to model count data with overdispersion. A successful model should take into account a rather pronounced dependence structure and the extra variation relative to the mean of the time series. The generalized linear autoregressive moving average (GLARMA) model and the autoregressive conditional Poisson (ACP) model are some successful models which address the autocorrelation of the time series and the problem of overdispersion of count data. It is modified from the traditional autoregressive (AR) model and moving average (MA) model which are commonly used in time series analysis.

Perhaps, there would be another more straightforward approach to fit count data with trend. We proposed in this research the Poisson Geometric Process (PGP) model which is originated from the Geometric Process (GP) model. Lam ((28), (29)) first proposed modeling directly the time series with monotone trend by a monotone process called the Geometric Process (GP). The data $\{X_t\}$ follows a GP if $\{Y_t = a^{t-1}X_t, t = 1, \ldots, n\}$ forms a renewal process (RP) which is a sequence of identically and independently distributed (IID) non-negative random variables. The RP $\{Y_t\}$ can be assigned different lifetime distributions.

In most of the previous studies, the GP model was used to model the inter-arrival times of a series of events which are positive continuous (Lam (34)). Re-

cently, the model is also extended to binary data (Chan and Leung (9)). However, it is not yet developed to model time series count data which are indeed very commonly found in our daily lives. In this thesis, we aim to extend the type of data for the GP model to Poisson count time series. The resultant model is called the Poisson Geometric Process (PGP) model. In the PGP model, we assume that the latent GP $\{X_t\}$ is the mean of some Poisson distributions for the observed count data $\{W_t\}$ such that autocorrelation with different dependency structures is introduced via a ratio function a_t and overdispersion is allowed by the source of variability from the latent GP. A simpler version of the PGP model can be defined by assigning directly the mean of the GP $\{X_t\}$ to the mean of the Poisson distributions. The PGP model can be implemented by using different parametric approaches such as the maximum likelihood (ML) or Bayesian methods, and non-parametric approach say the least-square error (LSE) method.

Furthermore, following the framework of the exponential family of sampling distribution within the generalized linear mixed model (GLMM), the GP model can be easily extended to allow for covariates, mixture and random effects by adopting a linear function of these effects linked by a suitable link function to the mean of some lifetime distributions for the unobserved RP $\{Y_t\}$. The inclusion of mixture and random effects into the mean function helps to accommodate variability in the data due to overdispersion, clustering and population heterogeneity.

Such extensions is definitely significant in the development of GP models.

1.2 Basic Geometric Process model

1.2.1 Previous researches

For inference and model development during the past years, increasing attention has been drawn on the GP models. New methods of inference have been derived and have been successfully applied to several areas. In statistical inference, Lam (31) proposed some non-parametric methods of inference. Lam and Chan (35) and Chan *et al.* (8) considered parametric likelihood method adopting respectively the lognormal and gamma distributions to the RP. They derived large sample distributions for the parameter estimators and showed that the parametric methods are more efficient. Lam *et al.* (40) showed that the GP model out-performed the Cox-Lewis model (Cox and Lewis (12)), the Weibull process model (Ascher and Feingold (3)) and the homogeneous Poisson process model and was easier to implement using non-parametric methods.

Later, Ho (25) extended the basic GP model to the generalized Geometric Process (GGP) model by using the idea of the generalized linear model (GLM). The GGP model can account for covariate effects in the data and thus making the model more comprehensive in modeling trend data. Recently, the GP model

has been extended to study multiple trend data. Chan *et al.* (10) applied the threshold GP (TGP) model to study the Severe Acute Respiratory Syndrome (SARS) epidemic in Hong Kong in 2003 because the number of daily infected cases showed a clear growing trend at the beginning of the outbreak and declined afterwards. Besides the extension of GP model to account for multiple trend data, the binary Geometric Process (BGP) model was also introduced by Ho (25) and Chan and Leung (9) to model some binary data. In the study of the methadone clinic data in Western Sydney in 1986 in Chan and Leung (9), it was found that the BGP model could fit the binary trend data even better than the common autoregressive logistic regression model. In the next chapter, the development of the GP models will be reviewed in details.

1.2.2 Basic Geometric Process model

In the analysis of a system with trend, the nonhomogeneous Poisson process (NHPP) has been used for a long time. If the successive inter-arrival times are monotone, the Cox-Lewis model and exponential process model are commonly used. Lam ((28), (29)) however, adopted a more direct approach to model such monotone processes. He first proposed the Geometric Process model (GP) with the following definition:

Definition: Given a set of random variables $\{X_1, X_2, \dots\}$, if for some $a > 0$,

$\{Y_t = a^{t-1}X_t, t = 1, 2, ...\}$ forms a renewal process (RP), then $\{X_t, t = 1, 2, \dots\}$ is called a Geometric Process (GP), and the real number a is called the ratio of the GP. This ratio measures both the direction and strength of the trend. In fact, a GP is a sequence of non-negative random variables which is monotone increasing if $a < 1$, and monotone decreasing if $a > 1$. When $a = 1$, it becomes a stationary renewal process (RP) which is independently and identically distributed (IID) with the same distribution F. The individual terms Y_t are called renewals (Feller (19)). The GP model asserts that if the ratio a discounts the t^{th} outcome X_t geometrically by $t - 1$ times, the resulting process $\{Y_t\}$ becomes stationary and forms a RP, which may follow some parametric distributions such as exponential, gamma, Weibull and lognormal distributions.

For the RP $\{Y_t\}$, the mean and variance are:

$$E(Y_t) = \mu \qquad \text{and} \qquad Var(Y_t) = \sigma^2$$

So, the mean and variance of the GP $\{X_t\}$ would have the form:

$$E(X_t) = \frac{\mu}{a^{t-1}} \qquad \text{and} \qquad Var(X_t) = \frac{\sigma^2}{a^{2(t-1)}} \tag{1.1}$$

Clearly, the mean and variance of the GP depends completely on the three parameters μ, a, and σ^2. The inverse relation of the ratio a with the mean $E(X_t)$ explains why the trend would become monotone increasing when $a < 1$ but decreasing when $a > 1$. Comparing GP models with GLMs, equation (1.1) shows

that GP models are more general. GP models adopt both arithmetic and geometric relationships to describe changes. In fact, lots of changes in real life when repeated measurements are made are better described by geometric relationships. Common examples include the growth of a confined community of people in social sciences, of infected patients in epidemiology or of living organisms in biology. Essentially GP models separate the effects of exogenous variables from the trend effects. They identify these two sources of effects on the mean $E(X_t)$ by two parameters: the mean μ of the RP which is the baseline level of the trend and the ratio a which describes the progression of the trend over time. The models reduce to GLMs when $a = 1$.

As trend data often comes from longitudinal measurements, it is important to accommodate the serial correlation and possibly the volatility of the outcomes over time. Equation (1.1) shows that the ratio a of a GP affects not only the mean but also the variance of the GP and allows it to change over time. The ratio reveals both the direction and the strength of the trend while eliminating the random noises. It also accommodates the autocorrelation between observations as in the autoregressive (AR) model in time series. In this way, the GP model avoids the initial stage problem (Chan (5)) in the AR model and makes prediction more straightforward. The properties of GP model make it adaptive to describe a wide range of trend patterns over time.

1.2.3 Applications

For application, the GP models have been extensively used in modeling inter-arrival times with trends. For instance, Lam (31) showed that the time intervals in days between successive disasters in Great Britain during 1851 to 1962 followed a GP. Lam ((28), (29), (30)) first considered the GP models to study some replacement problems. During the fifteen years, GP models have been applied to reliability and maintenance problem in the study of optimal replacement or repairable models. For example, Lam (31) succeeded in fitting the GP model to the inter-arrival times (in hours) of the unscheduled maintenance actions for the U.S.S. Halfbeak No.3 main propulsion diesel engine. Not surprisingly, increasing attentions have been drawn on the GP models and new areas of application have been derived. They include Lam and Zhang (36) in the analysis of a two-component system arranged in series, Lam (32) and Lam and Zhang (37) in the analysis of a two-component system emerged in parallel, and Lam (33) in the study of the failure occurrence rate. Others include Lam *et al.* (39), Lam and Zhang (38), Zhang ((65), (66)), Zhang *et al.* ((68), (67)) and Sheu (54). See Lam (34) for a brief review and further reference.

1.3 Related models

1.3.1 Generalized linear mixed model (GLMM)

Nelder and Wedderburn (49) introduced the generalized linear model (GLM) to allow an exponential family of distributions for the data when they found that the linear regression (LR) model did not work well if the responses did not have linear relationships with the explanatory variables, or if the observations were serially correlated. Data from clinical trials and stock market often have these problems.

In the LR model, the responses $\boldsymbol{Y}$ should have a linear relationship with the explanatory variables $\boldsymbol{X}$ such that:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \qquad \text{where} \quad \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

In the above expression, $\boldsymbol{\beta}$ is a vector of parameters and $\boldsymbol{\epsilon}$ is the error term which follows a multivariate normal distribution with mean $\boldsymbol{0}$ and variance $\sigma^2\boldsymbol{I}$ where $\boldsymbol{I}$ is an identity matrix. Hence, if the responses $\boldsymbol{Y}$ deviate from the assumptions of the LR model, inference based on the LR model will be invalid. To overcome the problem, the GLM should be implemented. A GLM consists of three components:

1. the random component, which specifies the probability mass function (pmf) of the response variable;

2. the systematic component, which specifies a linear function of the explanatory variables; and

3. the link function, which describes a functional relationship between the systematic component and the expectation of the random component.

We define the three components in three steps in order to implement the GLM.

In the first step, a more general distribution from the exponential family should be assigned to the responses $\boldsymbol{Y}$, the random component. The exponential family of distributions for the responses $\boldsymbol{Y}$ is defined as below:

$$\begin{aligned} y_t &\overset{indept.}{\sim} f(y_t) \\ f(y_t) &= \exp\left\{\frac{A_t[\theta_t y_t - b(\theta_t)]}{\phi} + c\left(y_t, \frac{\phi}{A_t}\right)\right\} \end{aligned} \tag{1.2}$$

where $b(.)$ is a function of the canonical parameter θ_t, $c(.)$ is a function of y_t and ϕ/A_t, A_t is the prior weight and ϕ is the dispersion parameter. If ϕ is known, (1.2) represents a linear exponential family. On the other hand, if ϕ is unknown, (1.2) is called an exponential dispersion model. The GLM provides a unified model for a wide variety of discrete, continuous and censored responses that can be assumed to be independent. The exponential family includes distributions such as gamma distribution, Poisson distribution, binomial distribution and normal distribution. The distribution has a mean $\mu = E(Y) = b'(\theta)$ and a variance $Var(Y) = \phi b''(\theta)/A = \phi v(\mu)/A$ in which $v(\mu) = b''[(b')^{-1}(\mu)]$ is called

the variance function. Table 1.1 shows that different distributions have different dispersion parameters, and means and variance functions.

The second step is to define a linear function of predictors, the systematic component $\boldsymbol{X}\boldsymbol{\beta}$, so that the responses can be predicted using the estimated regression parameters $\boldsymbol{\beta}$. The linear predictor where $t = 1, \ldots, n$ can be written as:

$$\eta_t = \boldsymbol{x}_t^T \boldsymbol{\beta}$$

where $\boldsymbol{x}_t$ is a vector of the explanatory variables and $\boldsymbol{\beta}$ is the vector containing the regression parameters.

Afterwards, in the third step, a link function g should be specified to link the linear function of predictors η_t and the mean μ_t for the distribution of y_t together, that is $g(\mu_t) = \eta_t$. Since different distributions within the exponential family account for different types of data, appropriate link function should be chosen to ensure that the mean for the distribution of y_t is correctly specified. Refer to Table 1.1 for the link function g specified to different distributions within the exponential family. As a result of the three steps, the GLM is established to handle those non-normal data.

==========

Table 1.1 about here

==========

However, the GLM is still incapable of handling longitudinal data with substantial serial correlation as well as clustered data with considerable overdispersion due to the population heterogeneity. In order to solve the problems, the GLM can be extended to the generalized linear mixed model (GLMM) by the inclusion of random effects in the linear function of predictors η_t. Thus, given the random effects $\boldsymbol{u}$, the model on response y_t where $t = 1, \ldots, n$ follows an exponential family of distribution as in (1.2).

The linear function of predictors η_t where $t = 1, \ldots, n$ is defined by the explanatory variables and random effects:

$$\eta_t = \boldsymbol{x}_t^T \boldsymbol{\beta} + \boldsymbol{z}_t^T \boldsymbol{u}$$

where $\boldsymbol{u}$ is a vector of the random effect parameters and $\boldsymbol{z}_t$ is the design matrix for the unobserved random effects $\boldsymbol{u}$. The linear function of predictors would then be linked by a canonical link function $g(\mu_t)$ depending on the type of data in the model. Due to the fact that the random effects $\boldsymbol{u}$ are unobserved, a distribution is assigned to the random effects such that $\boldsymbol{u} \sim h(\boldsymbol{u})$. A common choice will be the multivariate normal distribution with mean $\mathbf{0}$ and variance $\sigma^2 \boldsymbol{I}$.

By comparing the GP model with the GLM, it is easy to find that the GLM is a special case of the GP model when $a = 1$. Hence the modeling framework and inference methodology in the GP model can be easily extended by following those of the GLMs. However, in the context of time-series count data, it is less popular

to use the GLM or the GLMM to account for the trend effect in the data because the models fail to account for the changes in the variance of the data over time. In such cases, some common time-series models like autoregressive (AR) model, moving average (MA) model and their extended models like the autoregressive conditional heteroscedasticity (ARCH) and generalized ARCH (GARCH) models, and the stochastic volatility (SV) models for continuous data can be implemented. In the next section, the development of the time-series models for count data will be discussed.

1.3.2 Time-series models for Poisson counts

A series of observations $\{W_t, t = 1, \ldots, n\}$ indexed by the time t is called a count series if W_t is a natural number. Series of counts arise typically in applications where the number of certain events during a certain time period is the object of interest (Drescher (17)). Jung *et al.* (27) stated that a successful model for time series of counts should be able to account for the pronounced dependence structure existing in most count data and the overdispersion relative to the mean of the series. Besides, in a regression context, an easy to adopt link function to incorporate some covariate effects is also a highly desirable modeling feature.

A wide variety of models have been proposed to account for time series count data. At the early twentieth century, Andrey Markov (1856-1922) proposed the

Markov chain to model some dependent events (Sheynin (55), (56)). Markov chain is a discrete-time stochastic process with the Markov property.

Definition: A Markov chain is a sequence of random variables $\{W_1, W_2, \dots\}$. The range of these variables is called the state space, and the value of W_t is the state of the process at time t. If the conditional probability distribution of W_t given the past states is a function of W_{t-1} alone, that is:

$$P(W_t = w|W_1, W_2, \dots, W_{t-1}) \equiv P(W_t = w|W_{t-1}),$$

where w is some states of the process, then the sequence is a Markov chain. The probabilities are called transition probabilities and the identity above states the Markov property. Markov chain is applicable and has been extensively used to model data in which the observations only take a few possible values, for example, the binary data. However, when the number of possible values gets very large, the Markov Chain will lose its tractability.

Recently, there has been considerable development of models for non-Gaussian time series of Poisson counts. Cox (13) introduced a system that classified the models for count series into two main categories, namely observation-driven (OD) and parameter-driven (PD). The classification is based on the theory and terminology of state-space model which consists of an observation equation and a state equation (Drescher (17)). The observation equation specifies the distribution for the outcomes. Whereas, the state equation describes the evolution of the process

defined by the state variable through time. For both OD and PD models, the observation equation is the same. It is the specification of the state equation which distinguishes between these two models. PD models induce the serial dependence by the latent state variables which evolve independently of the past outcomes, therefore it is straightforward in the interpretation of the covariate effects on the observed count series. Also, the statistical properties are easy to derive. Zeger (64) has proposed a prominent PD Poisson model with a stochastic autoregressive mean (SAM).

Definition: Let $\{W_t, t = 1, \ldots, n\}$ be a time series of counts, $\{\boldsymbol{x}_t\}$ be a vector of covariates, and $\{u_t\}$ denote a latent non-negative stochastic process. Then the conditional distribution of $w_t|(\boldsymbol{x}_t, u_t)$ is assumed to be Poisson with mean $\mu_t = \exp(\boldsymbol{x}_t^T\boldsymbol{\beta})u_t$ denoted by

$$w_t|(\boldsymbol{x}_t, u_t) \sim Poi(\exp[\boldsymbol{x}_t^T\boldsymbol{\beta}]u_t)$$

where $\boldsymbol{\beta}$ is a vector of regression parameters. The latent process u_t is typically introduced to account for the possible overdispersion and serial autocorrelation within the count data by assuming that $\lambda_t = \ln u_t$ follows a Gaussian first-order autoregressive process:

$$\lambda_t = \delta\lambda_{t-1} + \nu\epsilon_t, \quad \epsilon_t \overset{iid.}{\sim} N(0, 1).$$

The existence of the latent process u_t gives rise to a straightforward interpretation of the effects of covariates $\boldsymbol{x}_t$ on the count process W_t because the covariate effects

are formulated independent of the past states in the model. However, because of the latent variables u_t, the PD models require considerable computation for parameter estimation due to the complicated derivation of the likelihood function (Durbin and Koopman (18), Jung and Liesenfeld (26)). Another drawback is the difficulty in forecasting as the model is built on a latent process (Davis *et al.* (15)).

On the other hand, OD models specify the serial dependence by using the state variables which are expressed explicitly as functions of past outcomes. In the general framework of the OD model, it is assumed that the observation W_t given the past history $\mathcal{F}_{t-1}$ is Poisson with mean μ_t which is denoted by

$$W_t|\mathcal{F}_{t-1} \sim Poi(\mu_t).$$

In order to include covariate effects, the link function $\mu_t = \exp(\eta_t)$ is used where $\eta_t = \boldsymbol{x}_t^T\boldsymbol{\beta} + Z_t$. In this model, $\boldsymbol{x}_t$ is a vector of explanatory variables and

$$Z_t = \sum_{i=1}^{q} \theta_i e_{t-i} \tag{1.3}$$

is a sum of past errors $e_t = \dfrac{W_t - \mu_t}{\mu_t^{\lambda}}, \lambda \geq 0$ which describes the correlation structure of W_t. The class of OD models includes the discrete autoregressive moving average (DARMA) model (Chang *et al.* (11)), the integer-valued autoregressive (INARMA) models (McKenzie (43), (44)), the generalized linear autoregressive moving average (GLARMA) models (Davis *et al.* (14), (15)) and the autoregres-

sive conditional Poisson (ACP) models (Heinen (24)).

The DARMA and INARMA models are developed using the idea of the traditional AR model and MA model used in time series analysis. They have properties similar to that of the ARMA model. The DARMA model consists of probabilistic mixtures of discrete IID random variables with suitably chosen marginal distribution, while the INARMA model uses the concept of "binomial thinning" to make the dependent variable W_t takes only integer values. However, these models were not practically applicable due to its cumbersome estimation of the model parameters, hence emphases were only put on the studies of their stochastic properties (Heinen (24)).

To tackle the estimation problem, Davis *et al.* ((14), (15)) has proposed the generalized linear autoregressive moving average (GLARMA) model by making some modifications in Z_t given by (1.3) as below:

$$Z_t = \phi_1 U_{t-1} + \cdots + \phi_p U_{t-p} + \sum_{i=1}^{q} \theta_i e_{t-i}$$

where U_t is an ARMA(p, q) process with noise $\{e_t\}$. The GLARMA model can allow for the serial correlation in the count data by specifying the log of the conditional mean process as a linear function of previous counts. The main advantage of the GLARMA model was the efficient model estimation using maximum likelihood (ML) method. Moreover, covariates can be included in this model in contrast to the INARMA model. This greatly widens the applications of the

model to time-series count data.

Recently, Heinen (24) has proposed a similar group of OD models called the autoregressive conditional Poisson (ACP) model in which the counts follow a Poisson distribution with an autoregressive mean μ_t:

$$\mu_t = \omega + \sum_{i=1}^{p} \alpha_i W_{t-i} + \sum_{i=1}^{q} \beta_i \mu_{t-i}.$$

Heinen (24) proved that the ACP model exhibits overdispersion, even though it uses an equidispersed marginal distribution. This property allows the ACP model to model time series of counts successfully.

In summary, the OD models are essentially an important class of time series model for count data. Davis (14) revealed that the prediction of counts and the derivation of the likelihood function are easy for OD models. However, the statistical characteristics, like stationarity and ergodicity, are difficult to derive and the interpretation of parameters is not simple as the mean μ_t contains the past observations $\mathcal{F}_{t-1}$.

Nevertheless, the PGP model proposed in this thesis is essentially a state space model with state variable X_t which is a latent GP, evolving independently of the past outcomes over time. Therefore, it should be classified as a PD model. Unlike most PD models which are complicated and computational intensive in model implementation, the PGP model does not require complicated derivation of likelihood function. Moreover, it focuses on modeling the trend of the time

series of counts but none of the models listed above take the trend effect into account. Therefore, the PGP model and its extensions provide an important class of alternative models for modeling time series of Poisson count data with trends.

1.3.3 Stochastic volatility model

A great deal of data in economics, finance, natural science and engineering appears in the form of time series where the continuous outcomes are dependent and the nature of this dependence is of interest. For time series analysis, lots of models were proposed such as the AR model and the MA model which assume stationarity in the time series. However, it is generally acknowledged that the variability of the data in many time series is not constant over time. For instance, the volatility of many financial returns of an asset usually fluctuates periodically. Over the past two decades, two prominent classes of models, namely the generalized autoregressive conditional heteroscedasticity (GARCH) model and the stochastic volatility (SV) model, have been developed to capture the time-evolving autocorrelated volatility process and to forecast the financial market volatility which is crucial in option pricing, hedging strategies, portfolio allocation, etc.

So far, GARCH model proposed by Bollerslev (4) has been the most frequently applied class of time-varying volatility model. Nonetheless, SV model

(Taylor (60)) has drawn increasing attention and has been recognized as a viable alternative to GARCH model due to the considerable progress in the evaluation of the intractable likelihood function of the SV model. Volatility models are mostly defined by their first two moments, that is the mean and variance equation. Denote the response of a time series by $y_t, t = 1, \ldots, n$, the general form of the mean equation is given by

$$y_t \quad = \quad \mu_t + \sigma_t \epsilon_t, \qquad \epsilon_t \overset{iid.}{\sim} N(0,1)$$

where μ_t is a linear function of covariates. For the variance equation σ_t^2, the GARCH model defines the time-varying variance as a deterministic function of past squared innovations and lagged conditional variances. On the contrary, the variance in the SV model is modelled as an unobserved component that follows some stochastic process. The variance equation of the SV model is given by

$$\sigma_t^2 = \sigma^{*2} \exp(h_t)$$

where σ^{*2} is a positive scaling factor and h_t is a stochastic process.

Similar to the SV model, the PGP model also accommodate the variability in the data as its mean and variance change over time as shown in (3.3) and (3.5). Yet, GP model is more advantageous in model implementation as the likelihood function and its derivative functions are much more complicated in the SV model. Besides, the SV model focusing on modeling and forecasting volatility in the data

does not take the trend effect into account. In addition, the extension of GP model to fit different types of data is straightforward and makes it more comprehensive in different areas of applications. In the next chapter, the development of GP model is discussed in details.

Table 1.1: Characteristics of some common univariate distributions in the exponential family

	Range of y	Canonical link $g(\mu)$	Mean $b'(\theta)$	Variance function $v(\mu)$	Dispersion parameter ϕ
Normal $N(\mu, \sigma^2)$	$(-\infty, \infty)$	μ	θ	1	σ^2
Poisson $P(\mu)$	$0(1)\infty$	$\ln \mu$	$\exp(\theta)$	μ	1
Binomial $B(a, \mu)$	$\frac{0(1)a}{a}$	$\ln \frac{\mu}{1-\mu}$	$e^\theta/(1+e^\theta)$	$\mu(1-\mu)$	1
Gamma $G(\mu, \nu)$	$(0, \infty)$	$1/\mu$	$-1/\theta$	μ^2	ν^{-1}
Inverse Gaussian $IG(\mu, \sigma^2)$	$(0, \infty)$	$1/\mu^2$	$(-2\theta)^{1/2}$	μ^3	σ^2

Chapter 2

Development of Geometric Process Models

In this chapter, we will first introduce the development of the GP models. After that, we will study some common methodologies of inference which are useful in the statistical modeling for the GP models.

2.1 Extension to data with covariate effect: generalized Geometric Process model

In most cases, however, the mean of a GP may not be homogeneous, the data may exhibit covariate effects. In other words, the outcome is likely to be affected by a number of factors. For instances, the inter-arrival times between blood donations may be affected by the age or education background of the donor. Also, the number of failures of a system may depend on the humidity of the working environment. Adopting the idea of the generalized linear model (GLM), the GP model can be extended to account for such covariate effects in the outcome, and the extended model is called the generalized Geometric Process (GGP) model.

Definition: Let $X_1, X_2, \dots$ be a sequence of random variables. If there exists a positive real number a such that $\{Y_t = a^{t-1}X_t, t = 1, 2, \dots\}$ forms a stochastic process (SP), with mean $E(Y_t) = \mu_t$, then $\{X_t\}$ is called a generalized Geometric Process (GGP).

The GGP $\{X_t\}$ is also a stochastic process and a is is called the ratio of the GGP which indicates the direction and strength of the trend. However, the process $\{Y_t\}$ is no longer a RP because Y_t is not independently and identically distributed (IID). Since the RP $\{Y_t\}$ may evolve over time subject to different external effects, the GGP model can be implemented by linking a linear function of covariates η_t to the mean μ_t of some lifetime distributions for the observed data as

$$\eta_t = \beta_0 z_{0t} + \beta_1 z_{1t} + \cdots + \beta_k z_{kt}$$

where $z_{0t} = 1$ and $z_{kt}, k = 1, \ldots, q_\mu$ are covariates. For different types of data, the link function varies as in the GLM according to Table 1.1.

When comparing the GGP model with the GLM, it is not difficult to see that GLM is actually a special case of the GGP model when the ratio a is equal to 1. But, with the ratio parameter in the GGP model, it has an advantage of accounting the trend effect in the data. Therefore, the extension of the GP model to include covariate effects in the data is useful as data with trend and covariate effects are common. Ho (25) studied the daily number of infected cases of the Severe Acute Respiratory Syndrome (SARS) in Hong Kong in 2003 using the GGP model. He included the daily temperature in the mean function because it is proved that temperature is one of the crucial factors in affecting the growth of virus. Besides, Ho (25) also used the GGP model to analyze the monthly indices of goods imported in Hong Kong from January 1982 to April 2004 since the data demonstrated a seasonal effect. In the mean function, an indicator for the month

of January and February is included to account for the seasonal effect.

2.2 Extension to data with multiple trends: multiple Geometric Process model

Most events would not develop only in one direction. A new epidemic would outspread quickly in the early stage when no medicine has been invented to cure the disease. The number of infected cases would therefore keep on growing at the beginning. However, once some precautionary measures such as contact tracing, quarantine and travel advices are implemented, the epidemic becomes controllable, and the number of infected cases will drop substantially. Another example would be the change of population size when a developing country becomes a developed one. The population size will usually grow at a decreasing rate and finally stabilizes. Previous studies of the GP models only limit to data with a single trend. Chan *et al.* (10) and Ho (25) extended the models to multiple trend data and fitted separate GP to each trend. The resultant model is called the multiple GP (MGP) model.

Definition: Let $X_t, t = 1, 2, \ldots, n$ be the observed data, $T_\kappa, \kappa = 1, 2, \ldots, K$ be the turning points of the GPs and n_κ be the number of observations for the κ^{th} GP. Hence, the κ^{th} GP with n_κ observations is defined as

$$GP_\kappa = \{X_t,\ T_\kappa \leq t < T_{\kappa+1}\}, \quad \kappa = 1, \ldots, K$$

The κ^{th} GP begins with a turning point T_κ, that means

$$T_1 = 1 \qquad \text{and} \qquad T_\kappa = 1 + \sum_{j=1}^{\kappa-1} n_j, \quad \kappa = 2, \ldots, K$$

are the turning points of the GPs and we have $\sum_{\kappa=1}^{K} n_\kappa = n$. Then, we define the corresponding renewal process (RP) as

$$RP_\kappa = \{a_t^{t-T_\kappa} X_t, \;\; T_\kappa \leq t < T_{\kappa+1}\}$$

where $a_t, T_\kappa \leq t < T_{\kappa+1}$ is the ratio parameter of the κ^{th} GP. Moreover, we have

$$E(X_t) = \frac{\mu_t}{a_t^{t-T_\kappa}} \qquad \text{and} \qquad Var(X_t) = \frac{\sigma_t^2}{a_t^{2(t-T_\kappa)}}, \;\; t = 0, 1, \ldots, n_\kappa - 1$$

as the mean and variance for the κ^{th} GP. Clearly, the three parameters a_t, μ_t, and σ_t^2 where $T_\kappa \leq t < T_{\kappa+1}$ completely determine the mean and variance of X_t in the κ^{th} GP. Also, during the growing stage of the multiple trend data, a_t would be less than 1; during the stabilizing stage, a_t would be equal to 1; and during the declining stage, a_t would be greater than 1.

However, problems arise when we locate the turning points in the multiple trend data. The turning points mark the time when the observed data develops to another stage. Chan *et al.* (10) proposed a moving window method to detect the turning point in the Hong Kong SARS data. It is a computationally intensive method. Section 5.1 describes this methodology in details. Chan and Leung (9) suggested two alternative methods to determine the turning points. They include the Bayesian method and the partial likelihood method. Section 5.2.2 and 5.2.3 describe the later two methodologies in details.

One of the major objectives in fitting the MGP model is the detection of the uncertain turning points. If the observations are not sufficient or the event is still on-going, it would be difficult to locate the turning points and thus, affecting the estimation of other model parameters and the accuracy of the prediction. Nevertheless, the MGP model opens a new area of application as the GP model is no longer limited to single trend data.

2.3 Extension to type of data: binary Geometric Process model

The GP models are often used to model positive continuous data. Most studies involved the application of GP models to data of inter-arrival times. For instances, Lam (31) and Lam and Chan (35) applied the GP models to the inter-arrival times between successive disasters, and between unscheduled maintenance actions of diesel engines. Indeed, the GP models can be extended to fit other scopes of data. Binary data is one of the common types of data in our daily lives. Ho (25) and Chan and Leung (9) proposed the binary GP (BGP) model using the framework of exponential family of sampling distributions within the GLMs.

Definition: Assume that there is an underlying unobserved GP $\{X_t\}$ and we observe only the indicators of whether X_t are greater than certain cut off level b. Without the loss of generality, we set $b = 1$ and define the observed binary data $W_t = I(X_t > 1) = I(Y_t > a^{t-1})$ where $I(E)$ is an indicator function for the event

E and $\{Y_t = a^{t-1}X_t\}$ is the underlying RP. Then, we have the probability

$$P_t = P(W_t = 1) = P(X_t > 1) = P(Y_t > a^{t-1}) = 1 - P(Y_t < a^{t-1}) = 1 - F(a^{t-1}),$$

where $F(.)$ is the cumulative distribution function (cdf) of Y_t. Hence, the likelihood and log-likelihood functions for the observed data W_t are

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{n} [1 - F(a^{t-1})]^{w_t} [F(a^{t-1})]^{1-w_t}$$

and

$$\ell(\boldsymbol{\theta}) = \sum_{t=1}^{n} \{w_t \ln[1 - F(a^{t-1})] + (1 - w_t) \ln[F(a^{t-1})]\}$$

respectively where $\boldsymbol{\theta}$ is a vector of model parameters. By adopting some lifetime distributions, such as exponential, and Weibull distributions to the underlying RP $\{Y_t\}$, the BGP model can be implemented using different methods of inference. Chan and Leung (9) also the extended BGP model to include time-evolving covariate effects in the data. The resultant model is called the adaptive BGP (ABGP) model because the adaptive baseline level μ_t and ratio a_t allow a broad range of applications for the BGP model. To accommodate the time-evolving effects, a linear function of covariates is log-linked to the mean function μ_t:

$$\mu_t = \exp(\eta_{\mu t}) = \exp(\beta_{\mu 0} z_{\mu 0 t} + \beta_{\mu 1} z_{\mu 1 t} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu t}) \tag{2.1}$$

where $z_{\mu 0t} = 1$ and $z_{\mu kt},\ k = 1, \ldots, q_\mu$ are the time-evolving covariates. The underlying process $\{Y_t\}$ under the extended BGP model is no longer IID but becomes a stochastic process (SP) in general. Similarly, to account for a progression in the trend effect, the ratio a_t can be log-linked to a linear function of covariates:

$$a_t = \exp(\eta_{at}) = \exp(\beta_{a0} z_{a0t} + \beta_{a1} z_{a1t} + \cdots + \beta_{aq_a} z_{aq_a t}) \tag{2.2}$$

where $z_{a0t} = 1$ and $z_{akt},\ k = 1, \ldots, q_a$ are the time-evolving covariates that affect the progression of the trend. Furthermore, like the GP model, the BGP model can be used to model multiple trend data. Such model is called the threshold BGP (TBGP) model.

For application, the BGP models have been applied by Chan and Leung (9) to study the coal mining disasters data which contains 190 inter-arrival times between successive disasters in Great Britain. This data have been studied by Lam (31), Lam and Chan (35), Chan and Leung (8), Lam *et al.* (40) and Ho (25) in the real data analysis of the GP model. Instead of treating the inter-arrival times as some positive continuous data, Chan and Leung (9) defined a new indicator function W_t to indicate whether a disaster occurs during the t^{th} quarter within 112 years. The ABGP and TBGP models were also applied to another dataset which contains results of weekly urine drug screens, positive or negative for heroin use, from 136 patients in a methadone clinic in Western Sydney in 1986. Chan and Leung (9) demonstrated that the BGP models perform even better than the conditional logistic regression (CLR) model, a common model for binary data. The extension of the GP model to BGP model are highly contributive to the modeling of data with binary outcomes.

2.4 Implication for further developments

The development of the GP model to binary data implies that other classes of data including Poisson counts and multinomial data can also be considered. Poisson count data with trend effect are commonly found. For instance, the number of infected cases of an epidemic outbreak and the number of failures of a repairable system during a fixed period of time are some prominent examples. However, neither of these kinds of data have been studied by using the framework of the GP models before.

Furthermore, most GP models are developed for data from a single system. For data coming from multiple systems, population heterogeneity is often observed. For example, it may exist in some clinical data when there is substantial between-patient variation. For instance, different patients may react differently to the same dose of medicine in a treatment. In this case, the GP model can be extended by introducing random intercepts to the mean and ratio functions in order to account for such problem and to facilitate subject-specific inference. Nevertheless, while it is feasible to account for individual effects in the random effects model, it is more common to find that a cluster of subjects may react similarly to a treatment. In this case, the GP model can be extended to incorporate mixture and clustering effects. In this model, each subject has a probability π_l of coming from group $l, l = 1, \ldots, G$ and the group membership of each subject I_{il} which is the indicator that subject i belongs to group l can be estimated. Hence, the characteristics of the groups can be identified. The mixture model is essen-

tially a discretized version of the random effect model in which each parameter takes only G possible values. These extensions of the GP model are certainly useful in many clinical studies and widen the scope of application of the GP models considerably.

2.5 Extension to methodologies of inference

In the statistical inference of the GP models, two popular methodologies are used for parameter estimation, namely the non-parametric and parametric inference. Non-parametric (NP) inference is much simpler than parametric inference as it requires no assumption on the distribution of the observed data and hence it involves no derivation and computation of the complicated likelihood function. Examples of NP inference include the least-square error (LSE) method, and the log-LSE method which are described in Section 2.5.1.

On the contrary, in parametric inference, some lifetime distributions are assigned to the unobserved RP $\{Y_t\}$ depending on the type of data. The maximum likelihood (ML) method is one of the traditional methods of parametric inference. Another approach namely the Bayesian method has become very popular recently because it avoids the evaluation of complicated likelihood functions. In the following sections, the LSE, ML and Bayesian methods would be discussed in details.

2.5.1 Least-square error method

The NP inference is often adopted in estimating the model parameters for the GP models due to its simplicity. In the methodology, a criterion, usually the sum of squared error (SSE) of the observed outcome X_t to its mean $E(X_t)$ is defined to measure the goodness-of-fit (GOF) of a model. This criterion is independent of the distribution of the data. This explains why the method is called a NP method. The basic idea of this method is to estimate the model parameters by minimizing the SSE. Lam (31) proposed the log-LSE method and Chan and Leung (9) considered both the log-LSE and LSE methods for positive continuous data. The methods consider SSE_1 and SSE_2 respectively given by:

$$SSE_1 = \sum_{t=1}^{n} [\ln X_t - E(\ln X_t)]^2 \tag{2.3}$$

$$\text{and} \quad SSE_2 = \sum_{t=1}^{n} [X_t - E(X_t)]^2 \tag{2.4}$$

The expected value functions $E(X_t)$ and $E(\ln X_t)$ depend on the mean function μ_t, the ratio function a_t and the type of data and model. For the GP model and the PGP model when $\{X_t\}$ is a latent GP, $E(X_t) = \mu_t / a_t^{t-1}$ and $E(\ln X_t) = \ln \mu_t - (t-1) \ln a_t$ where $\ln \nu_t = E(\ln Y_t)$. Lam (31) stated that the log-LSE method for the GP model is essentially the LSE method for the linear regression (LR) model on $\ln Y_t$. He also suggested that plotting $\ln X_t$ against the time t is a useful way to check if the data $\{X_t\}$ follow a GP. On the other hand, for the BGP model, $E(X_t) = \exp(-a_t^{t-1}/\mu_t)$. And for the model type, we have $\mu_t = \mu$ and $a_t = a$ for the basic model. When covariates are included in the adaptive

models, μ_t and a_t are given by (2.1) and (2.2) respectively.

After defining the SSE_m, we minimize the $SSE_m, m = 1, 2$ by solving the score equation $SSE'_m = 0$ for the parameter estimates $\boldsymbol{\theta}$ where SSE'_m and SSE''_m are the first and second order derivatives of the SSE_m function. Since there is no closed-form solution for the parameters $\boldsymbol{\theta}$ in $SSE' = 0$, we adopt the Newton Raphson (NR) iterative method. Denote the current parameter estimates in the k^{th} iteration by $\boldsymbol{\theta}^{(k)}$ and the next updated estimates by $\boldsymbol{\theta}^{(k+1)}$. In each NR iteration, the updating procedure is:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - [SSE''_m(\boldsymbol{\theta}^{(k)})]^{-1} SSE'_m(\boldsymbol{\theta}^{(k)}) \tag{2.5}$$

The iteration continues until $\| \boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)} \|$ becomes sufficiently small. Finally, the LSE estimates $\hat{\boldsymbol{\theta}}_{LSE} = \boldsymbol{\theta}^{(k+1)}$ would be obtained.

To implement the LSE or log-LSE method, the traditional programming language Fortran, together with the subroutine $DLINRG$ in the International Mathematical and Statistical Libraries (IMSL) for evaluating matrix inverse, is usually used. IMSL is a collection of most useful mathematical and statistical subroutines such as differentiation, integration, linear algebra and random number generation, etc.

2.5.2 Maximum likelihood method

Maximum likelihood (ML) method is a traditional methodology in parametric inference. Although NP inference requires less distribution assumptions, Lam

and Chan (35) and Chan *et al.* (8) studied the GP model with lognormal and gamma distributions respectively and showed that the ML method performed better than the NP methods in parameter estimation. In order to obtain the ML estimates, we follow the algorithm:

1. Assign a model $f(\boldsymbol{y};\boldsymbol{\theta})$ to the observed data $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)^T$, and denote the density of y_t to be $f(y_t;\boldsymbol{\theta})$. For positive continuous data, lognormal and gamma distributions are adopted for the RP $\{Y_t\}$ (Lam and Chan (35), Chan *et al.* (8)). For binary data, Weibull distribution is considered for the SP $\{Y_t\}$ (Chan and Leung (9)).

2. Define the likelihood function as

$$L(\boldsymbol{\theta};\boldsymbol{y}) = f(\boldsymbol{y};\boldsymbol{\theta}) = \prod_{t=1}^{n} f(y_t;\boldsymbol{\theta})$$

   Then, the log-likelihood function is

$$\ell(\boldsymbol{\theta};\boldsymbol{y}) = \ln L(\boldsymbol{\theta};\boldsymbol{y}) = \sum_{t=1}^{n} \ln f(y_t;\boldsymbol{\theta})$$

3. Differentiate the log-likelihood function with respect to the vector of parameters $\boldsymbol{\theta}$ to get $\ell'(\boldsymbol{\theta})$, set it to be zero and solve for the ML estimates $\hat{\boldsymbol{\theta}}_{ML}$ in which the likelihood function of the observed data $\boldsymbol{y}$ would be maximized.

4. Since there is no closed-form solution for $\boldsymbol{\theta}$ in $\ell'(\boldsymbol{\theta}) = 0$, we again adopt the Newton Raphson (NR) iterative method as in the LSE method. By updating the procedure

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} - [\ell''(\boldsymbol{\theta}^{(k)})]^{-1}\ell'(\boldsymbol{\theta}^{(k)}) \tag{2.6}$$

iteratively until $\boldsymbol{\theta}^{(k+1)}$ converges for sufficiently large k, the ML estimates are given by $\hat{\boldsymbol{\theta}}_{ML} = \boldsymbol{\theta}^{(k+1)}$.

We can also derive the large sample properties for the ML estimates $\hat{\boldsymbol{\theta}}_{ML}$ using the following theorem:

Theorem:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{ML} - \boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, n^{-1}\boldsymbol{\Sigma}) \tag{2.7}$$

where $\xrightarrow{L}$ means convergence in distribution when n is large, $\boldsymbol{\Sigma}$ is the covariance matrix and $s_{uv} = -E\Big[\dfrac{\partial \ell^2(\boldsymbol{\theta})}{\partial\theta_u \partial\theta_v}\Big]^{-1}$ is the element in the u^{th} row and v^{th} column of $\boldsymbol{\Sigma}$. With these asymptotic distributions, we can construct confidence intervals and perform hypothesis testing on $\boldsymbol{\theta}$.

Similar to the LSE method, the ML method can also be implemented by writing Fortran programs with IMSL. However, when the likelihood function involves high-dimensional integration, numerical approximation is required. To avoid doing such computational intensive work, the Bayesian method is a good alternative to consider in inference methodologies.

2.5.3 Bayesian method

The Bayesian method has the advantage over the ML method that avoids the evaluation of complicated likelihood functions which may involve high-dimensional integration. It has had increasing popularity during recent years due to the advancement of computational power and the development of efficient sampling

techniques. Moreover, the appearance of the statistical software WinBUGS makes the implementation of the Bayesian methodology without tears (Spiegelhalter *et al.* (59)).

The Bayesian method requires the establishment of a joint density function $f(\boldsymbol{w}, \boldsymbol{\theta})$,

$$f(\boldsymbol{w}, \boldsymbol{\theta}) = f(\boldsymbol{w}|\,\boldsymbol{\theta}) f(\boldsymbol{\theta})$$

containing the prior density function $f(\boldsymbol{\theta})$ for the model parameter $\boldsymbol{\theta}$ and a likelihood function $f(\boldsymbol{w}|\,\boldsymbol{\theta})$ for the data $\boldsymbol{w}$. Then, the Bayes theorem is applied to determine the joint posterior density for the parameter $\boldsymbol{\theta}$ conditional on $\boldsymbol{w}$ which is given by

$$f(\boldsymbol{\theta}|\,\boldsymbol{w}) = \frac{f(\boldsymbol{w}|\,\boldsymbol{\theta}) f(\boldsymbol{\theta})}{\int f(\boldsymbol{w}|\,\boldsymbol{\theta}) f(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

In order to construct the posterior density, some prior distributions $f(\boldsymbol{\theta})$ are assigned to the parameters $\boldsymbol{\theta}$ which are treated as random variables instead of fixed in the ML method. With no specific prior information, non-informative priors with large variances are adopted. For parameters in non-restricted continuous ranges of values, normal priors are usually used. Whereas, for those restricted to positive ranges of values, gamma or Wishart priors are some possible choices, and for those parameters which represent the probability of certain events, uniform or beta priors are adopted.

Once the joint posterior density are defined, the parameter estimates $\hat{\boldsymbol{\theta}}_{BAY}$ are given by the mean or median of the posterior distribution. In the derivation of the posterior density, very often, non-standard form of distribution is resulted. Hence

the evaluation of the posterior mean analytically will be complicated and instead, sampling methods including Markov chain Monte Carlo (MCMC) method (Smith and Roberts (58), Gilks *et al.* (20)) with Gibbs sampling and Metropolis Hastings algorithm (Hastings (23), Metropolis *et al.* (45)) are used to draw sample from the posterior conditional distribution of the parameters. The sample mean or median of the posterior sample gives the parameter estimate.

The idea of the Gibbs sampling is described below. Assume that we have three parameters θ_1, θ_2 and θ_3. The joint posterior density is written as $[\theta_1, \theta_2, \theta_3]$ and the conditional density of one parameter given the other two parameters are written as $[\theta_1 | \, \theta_2, \theta_3]$, $[\theta_2 | \, \theta_1, \theta_3]$, and $[\theta_3 | \, \theta_1, \theta_2]$ respectively. Gibbs sampling enables the sampling from the high dimensional posterior distribution be reduced to sampling from some univariate conditional distributions. This greatly reduces the complexity of the sampling process. The algorithm for the implementation is illustrated below:

1. Begin at starting values of $\theta_1^{(0)}$, $\theta_2^{(0)}$, and $\theta_3^{(0)}$.
2. Draw $\theta_1^{(1)}$ from the conditional distribution $[\theta_1 | \, \theta_2^{(0)}, \theta_3^{(0)}]$.
3. Draw $\theta_2^{(1)}$ from the conditional distribution $[\theta_2 | \, \theta_1^{(1)}, \theta_3^{(0)}]$ using $\theta_3^{(0)}$ and the newly simulated $\theta_1^{(1)}$.
4. Draw $\theta_3^{(1)}$ from the conditional distribution $[\theta_3 | \, \theta_1^{(1)}, \theta_2^{(1)}]$ using $\theta_1^{(1)}$ and $\theta_2^{(1)}$.
5. Repeat Step 2 to 4 until R iterations have completed with the simulated values converged to the joint posterior density function.

Throughout the R iterations, the first B iterations are set as the burn-in period and will be discarded to ensure that convergence has reached. The remaining $(R - B)$ iterations are then used to form posterior samples for the parameter estimates. Thereafter parameters are sub-sampled or thinned from every H^{th} iteration to reduce the auto-correlation in the sample. Resulting sample will consist of $M = (R - B)/H$ realizations with their mean or median taken to be the parameter estimates. In particular, when the posterior distributions of the parameters are skewed, which is usually the case for the scale parameter σ^2, the sample median is adopted. To ensure the convergence and independence of the posterior sample, the history and auto-correlation function (ACF) plots have to be checked. A narrow horizontal band of the posterior sample running from left to right in the history plots and a sharp cut-off in the ACF plots indicate the convergence and independence of the posterior sample respectively.

The values of R, B and H vary across different PGP models in the following chapters. For some complicated method, convergence rate will be lower and serial correlation will be higher. Then we takes larger values of R, B and H.

Chapter 3

Poisson Geometric Process Model

3.1 The basic model

Traditionally, the GP models have been extensively used to model inter-arrival times with trends. Since count data with trends are often encountered in real life, we extend the GP model to Poisson count data. The resultant model is called the Poisson Geometric Process (PGP) model. The development of the PGP model to adopt count data greatly broadens the application of the GP models in different fields such as epidemiology, environmentalology, marketing, etc.

Definition: Denote the observed data by W_t with time $t = 1, \ldots, n$, we assume that W_t follows Poisson distribution with mean $\lambda_t = X_t$ where $\{X_t\}$ forms a latent GP. There exists a ratio $a > 0$ such that $\{Y_t = a^{t-1}X_t\}$ is a RP. The probability mass function (pmf) for W_t given X_t is

$$f_{ot}(w_t|\, x_t) = \frac{e^{-x_t} {x_t}^{w_t}}{w_t!}$$

where the subscript o in $f_{ot}(.)$ denotes the original version of the PGP model. In general, we write $E(X_1) = \mu_t = \mu$ and $a_t = a$.

For the basic model, which is adopted throughout this chapter, $\mu_t = e^{\beta_0}$ and $a_t = a$. The number of covariates is zero for both mean and ratio functions, that is $q_j = 0, j = a, \mu$. The vector of parameters $\boldsymbol{\theta} = (\beta_0, a)^T$ for the basic model. To

illustrate trends with different directions and rates, we consider the basic PGP model with both a_t and μ_t set to constants. We set $\mu_t = 1$ and $a_t \leq 1$ and $a_t > 1$ for cases 1 and 2 respectively. Figures 1a and 1b show the two cases of increasing and decreasing trends correspondingly.

(1) Basic PGP models with fixed ratio function $a_t = a$,

Case 1. Fixed $a_t = a \leq 1$: $E(W_t)$ increases at an increasing rate (Figure 1a). When $a = 1$, $E(W_t) = 1$ is constant over time.

Case 2. Fixed $a_t = a > 1$: $E(W_t)$ decreases at a decreasing rate (Figure 1b).

Figure 1a $E(W_t)$ in PGP model with fixed $a \leq 1$

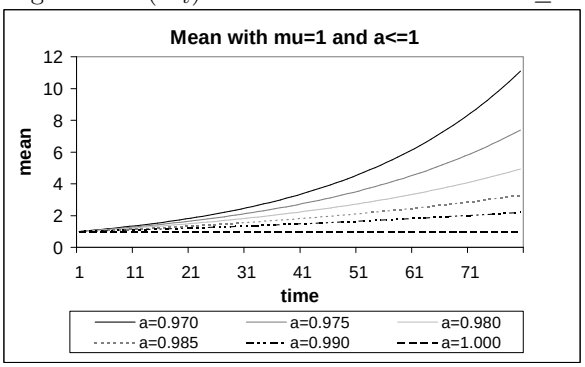
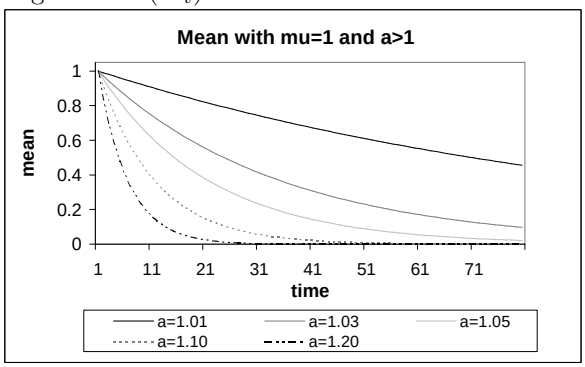


Figure 1b $E(W_t)$ in PGP model with fixed $a > 1$



Other trend patterns are considered in Chapter 4. Following the model framework of the GP model, we further assume that the RP $\{Y_t\}$ follows some lifetime distributions, like exponential, gamma, Weibull and lognormal distributions with pmf $f_{Y_t}(y)$. Since $\{X_t\}$ is unobserved, the unconditional pmf for W_t becomes:

$$f_{ot}(w_t|\,\boldsymbol{\theta}_o) = \int_0^\infty \frac{\exp(-x_t){x_t}^{w_t}}{w_t!}\, f_{Y_t}(a_t^{t-1}x_t)\, a_t^{t-1}\, dx_t \tag{3.1}$$

where $\boldsymbol{\theta}_o$ is a vector of model parameters for the distribution $f_{Y_t}(y)$ and the ratio a_t.

3.2 Extension to lifetime distributions

The simplest lifetime distribution for the RP $\{Y_t\}$ is probably the one-parameter exponential distribution with the pdf

$$f(y) = \begin{cases} \frac{1}{\lambda}\, e^{-\frac{y}{\lambda}} & y \geq 0, \quad \frac{1}{\lambda} \geq 0, \\ 0 & y < 0 \end{cases}$$

where λ is called the scale parameter of the distribution. The density of the exponential distribution has a reverse J-shape, and the mean and variance of the distribution are respectively

$$E(Y) = \lambda \qquad \text{and} \qquad Var(Y) = \lambda^2.$$

Also, it has a constant hazard function given by

$$h(y|\,\lambda) = \frac{1}{\lambda}\,.$$

It is therefore a suitable model for lifetime data where a used item is considered as good as new. This is the 'memoryless' property of the exponential distribution. The exponential distribution was widely used in early work, for instance, in the reliability of the electronic components or in medical studies. In this thesis, we only consider exponential distribution because it has closed-form of cumulative

distribution function (cdf) which does not involve integration. For other distributions like gamma, Weibull and lognormal distributions can also be assigned to the RP $\{Y_t\}$. Nevertheless, the resulting pmf for the observed data W_t may involve integration which makes the estimation of parameters much more complicated. Hence, they are not considered. We assign an exponential distribution to the RP $\{Y_t\}$ in the PGP model, i.e. $Y_t \sim Exp\left(\frac{1}{\lambda_t}\right)$ with $E(Y_t) = \mu_t = \lambda_t$. From (3.1), the pmf for W_t becomes

$$f_{ot}(w_t|\,\boldsymbol{\theta}_o) = \int_0^\infty \frac{\exp(-x_t){x_t}^{w_t}}{w_t!}\left(\frac{a_t^{t-1}}{\mu_t}\right)\exp\left(-\frac{a_t^{t-1}}{\mu_t}x_t\right)\,dx_t$$

Then, the resulting pmf for W_t becomes

$$f_{ot}(w_t) = \frac{\mu_t^{-1}a_t^{t-1}}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t+1}} \tag{3.2}$$

The PGP model is essentially a state space model with state variables X_t which is a latent GP, evolving independently of the past outcomes over time. It, therefore, is classified as parameter driven (PD) model. Despite that PD models require complicated and computationally intensive model implementation, the latent GP can be integrated out by assigning an exponential distribution to Y_t. Obviously, the pmf indicates the fact that the larger the observed data W_t, the smaller is the probability. In addition, the mean and variance for W_t are

$$E_o(W_t) = \frac{\mu_t}{a_t^{t-1}} \quad \text{and} \quad Var_o(W_t) = \frac{\mu_t}{a_t^{t-1}} + \left(\frac{\mu_t}{a_t^{t-1}}\right)^2. \tag{3.3}$$

Please refer to Appendices 3.1 and 3.2 for the derivation of the pmf, mean and variance of W_t. To ensure that the Poisson counts are non-negative, we transform

the parameter μ to $\exp(\beta_0)$. It is interesting to note that the extra variability when the mean X_t of the Poisson distribution follows the GP model with pmf $f_{Y_t}(a_t^{t-1}x_t)\, a_t^{t-1}$, instead of fixed inflates the variance of the distribution for W_t so that the equidispersion property of the Poisson distribution, that is the equality of the mean and variance no longer holds. Now $Var(W_t) = E(W_t) + E(W_t)^2$ which is greater than the mean $E(W_t)$ in the PGP model. In fact, the assumption of equidispersion is often too restrictive and is violated in many real applications. The two distinctive features of the PGP model make it suitable for modeling count data with trends and more zeros.

3.3 Simplification of PGP model

In the original PGP model, we assume that the RP $\{Y_t = a_t^{t-1}X_t, t = 1, \ldots, n\}$ adopts some lifetime distributions. In order to simplify the PGP model, we simply assume that the mean λ_t of the Poisson count data W_t equal directly to the mean of the latent GP. Hence, the simplified PGP model is not a PD model. The definition of the simplified basic PGP model is given as follow.

Definition: Suppose that the observed data W_t follows a Poisson distribution with the mean λ_t equals to the mean for the unobserved GP $\{X_t\}$ such that

$$\lambda_t = E(X_t) = \frac{\mu_t}{a_t^{t-1}}$$

Then, the pmf of W_t becomes

$$f_{st}(w_t) = \frac{\exp\left(-\frac{\mu_t}{a_t^{t-1}}\right)\left(\frac{\mu_t}{a_t^{t-1}}\right)^{w_t}}{w_t!} \tag{3.4}$$

where the subscript s denotes the simplified version of the PGP model. Without the extra source of variability as in the original model, the mean and variance of W_t are equal and are given by

$$E_s(W_t) = Var_s(W_t) = \lambda_t \tag{3.5}$$

based on the properties of equidispersion of the Poisson distribution. By comparing the pmf of W_t of the simplified model with that of the original model, it is easy to see that the pmf of the simplified one would not always decrease with increasing W_t. As a result, the simplified model provides more feasibility in modeling.

For both original and simplified PGP model, it is not difficult to observe that the models have the Markov property. In probability theory, a continuous time SP $\{W_t\}$ has the Markov property if the conditional probability distribution of the future states $\{W_{t+h},\ h > 0\}$ of the process, given the present state at time t and all past states $\{W_s,\ s \leq t\}$, depends only upon the present state W_t and not on any of past states. In other words, we have

$$\Pr(w_{t+h}|\, w_s,\ \forall s \leq t) = \Pr(w_{t+h}|\, w_t,\ \forall h > 0).$$

This definition holds for a discrete time SP when t, s and h are positive integers. The PGP model, which belongs to the class of discrete time SP, has the Markov

property because the probability of an observed outcome W_t at time t, that is $f_{ot}(w_t)$ and $f_{st}(w_t)$ respectively for the original and simplified model, depends on the mean μ_t/a_t^{t-1} which in terms depends on the model parameters $\boldsymbol{\beta}_\mu$ and $\boldsymbol{\beta}_a$. Hence, the probability that $\Pr(w_{t+h}|\, w_s \forall s \leq t)$ condition on all the past history $\{W_s,\ s \leq t\}$ upto time t, will be the same as that condition only on the most recent state W_t. In other words, the pmf of the PGP model is conditionally independent of the past states given the present state.

3.4 Methodology of inference

3.4.1 Least-square error method

Least-square error (LSE) method is a kind of non-parametric methodology of inference in which the parameter estimates are obtained by minimizing the sum of squared error (SSE), as in (2.3), as below:

$$SSE \;=\; \sum_{t=1}^{n}[w_t - E(W_t)]^2 = \sum_{t=1}^{n}\left(w_t - \frac{e^{\beta_0}}{a^{t-1}}\right)^2 \tag{3.6}$$

Note that both original and simplified models have the same expectation $E(W_t)$, and hence the SSE functions are the same. Then, by solving the score function $SSE' = 0$, we can minimize the SSE and obtain the LSE estimates for a and β_0. Since there is no closed-form solution for the LSE estimates, we have to adopt the Newton Raphson (NR) iterative method and the second order derivative functions for the parameters a and β_0 are given in Appendix 3.3. The NR iterative method

can be implemented by writing Fortran program with IMSL to obtain the LSE estimates for a and β_0.

3.4.2 Maximum likelihood method

For the original PGP model, the likelihood function $L_o(\boldsymbol{\theta}_o)$ is given by

$$L_o(\boldsymbol{\theta}_o) = \prod_{t=1}^{n} f_{ot}(w_t) = \prod_{t=1}^{n} \frac{e^{-\beta_0} a^{t-1}}{(e^{-\beta_0} a^{t-1} + 1)^{w_t+1}} \tag{3.7}$$

Then, the log-likelihood function is

$$\ell_o(\boldsymbol{\theta}_o) = \sum_{t=1}^{n} (t-1) \ln a - n\beta_0 - \sum_{t=1}^{n} (w_t + 1) \ln(e^{-\beta_0} a^{t-1} + 1) \tag{3.8}$$

The first and second order derivative functions which are required in the NR method are given in Appendix 3.4.

For the simplified PGP model, the likelihood function $L_s(\boldsymbol{\theta}_s)$ and the log-likelihood function $\ell_s(\boldsymbol{\theta}_s)$ are given by

$$L_s(\boldsymbol{\theta}_s) = \prod_{t=1}^{n} f_{st}(w_t) = \prod_{t=1}^{n} \frac{\exp\left(-\frac{e^{\beta_0}}{a^{t-1}}\right)\left(\frac{e^{\beta_0}}{a^{t-1}}\right)^{w_t}}{w_t!} \tag{3.9}$$

$$\text{and} \quad \ell_s(\boldsymbol{\theta}_s) = \sum_{t=1}^{n} \left[-\frac{e^{\beta_0}}{a^{t-1}} + w_t\beta_0 - w_t(t-1)\ln a - \ln(w_t!) \right]. \tag{3.10}$$

Again, the first order and second order derivative functions as required in the NR iterative method are given in Appendix 3.4.

Having the second order derivative functions, we can also derive the limiting distributions for the ML estimators in both original and simplified PGP models.

Section 4.2.2 describes the derivation in details. Then, we can construct confidence interval and perform hypothesis testing on $\boldsymbol{\theta}_h$ by using these asymptotic distributions.

3.4.3 Bayesian method

In our research, the Bayesian method is implemented using the statistical software "WinBUGS" with version 1.4.1 (Spiegelhalter *et al.* (59)). In the original PGP model, we assign an exponential distribution to the RP $\{Y_t\}$. Here is the Bayesian hierarchy for solving the model parameters $\boldsymbol{\theta}_h$:

$$
\begin{aligned}
\text{Likelihood for the original model:} \quad w_t &\sim f_{ot}(w_t) \\
f_{ot}(w_t) &= \frac{a_t^{t-1}/\mu_t}{(a_t^{t-1}/\mu_t + 1)^{w_t+1}} \\
\mu_t &= \exp(\beta_0) \\
a_t &= a \\
\text{Likelihood for the simplified model:} \quad w_t &\sim Poi(\lambda_t) \\
\lambda_t &= \frac{\mu_t}{a_t^{t-1}} \\
\mu_t &= \exp(\beta_0) \\
a_t &= a
\end{aligned}
$$

$$
\begin{aligned}
\text{Priors for the parameters:} \quad a &\sim U(1-b, 1+b) \quad \text{where } b \text{ is a constant and } 0 < b < 1 \\
\beta_0 &\sim N(0, \sigma_b^2)
\end{aligned}
$$

Without prior information, we assign non-informative priors to the parameter. In particular, we set $b = 0.5$ and $\sigma_b^2 = 1000$ for the real data analyses in Section 3.6.

The parameters are drawn from their posterior conditional distributions $[a|\,\beta_0, \boldsymbol{W}]$ and $[\beta_0|\,a, \boldsymbol{W}]$ by Gibbs sampler where $\boldsymbol{W}$ is a vector of W_t. In both original and simplified PGP models, $R = 100000$ iterations are executed. The first $B = 5000$ iterations are set as the burn-in period and are discarded. Thereafter parameters are thinned from every 20^{th} $(H = 20)$ iteration resulting in $M = 4750$ posterior samples of a and β_0. The density plots of a and β_0 show that the distributions are almost always symmetric. Hence the Bayesian estimates are given by the sample means of these samples. The history and auto-correlation function (ACF) plots of the parameters are also examined to ensure the validity of their convergence and independence. The history and ACF plots of the parameters a and β_0 in Section 3.6.1 using the original model are given in Figures 3.1 and 3.2. They show that convergences have attained and the samples are independent. The history and ACF plots of parameters in other models are similar and hence are omitted in subsequent analyses.

==========

Figures 3.1 and 3.2 about here

==========

3.5 Simulation studies

In the simulation studies of the PGP model, the ML method is adopted for statistical inference. Bayesian method is not considered as it would be computationally intensive to handle a large amount of simulated data using WinBUGS. In addition, LSE method is not considered because we want to study the two parametric versions of the PGP model separately. We set the number of realizations $N = 200$ and the sample size in each realization $n = 50$ or 100 unless otherwise stated. Simulation procedures for the original and simplified models are different.

In the original model, for each realization, we simulate $n = 50$ or 100 observations, denoted by u_t, from a uniform distribution $U(0, 1)$ and obtain the observed counts w_t using the function

$$w_t = f_{ot}^{-1}(u_t) = \frac{\ln(a_t^{t-1}/\mu_t) - \ln u_t}{\ln(a_t^{t-1}/\mu_t + 1)} - 1$$

where $f_{ot}(w_t)$ is given by (3.2). Fortran programs are written to implement the simulation and IMSL subroutine DRNUNF is called to generate random numbers from the uniform distribution $U(0, 1)$. Whereas, in the simplified model, we simulate $n = 50$ or 100 observations from a Poisson distribution with mean $\lambda_t = \mu_t/a_t^{t-1}$ for each realization. This time another IMSL subroutine RNPOI is called to generate random numbers from the Poisson distribution with mean λ_t.

In both versions of the PGP models, we set the mean $\mu_t = \exp(\beta_0)$ and the ratio $a_t = a$ to be constant across time t. By varying the values of parameters a and β_0, we obtain different sets of simulated data. Afterwards, each realization in

the set of data is fitted to a PGP model using ML method. For each parameter θ, the reported parameter estimate $\hat{\theta} = \frac{1}{N}\sum_{j=1}^{N}\hat{\theta}_j$ is the mean of the $N = 200$ estimates $\hat{\theta}_j$, where j indicates the realization.

To assess the performance of parameter estimation, we suggest two measures to assess the accuracy of the estimates, namely and the mean squares of error (MSE_1) and the percentage error (PE). While for the precision, we suggest the standard error (SE). These measures are defined as below:

$$\begin{aligned} SE &= \sqrt{\frac{\sum_{j=1}^{N}(\hat{\theta}_j - \hat{\theta})^2}{N-1}} \\ MSE_1 &= \frac{\sum_{j=1}^{N}(\hat{\theta}_j - \theta)^2}{N} \\ PE &= \frac{\hat{\theta} - \theta}{|\,\theta\,|} \end{aligned}$$

where SE is the square root of the diagonal element that corresponds to the parameter θ in the sample variance-covariance matrix $\mathbf{\Sigma}$ and θ denotes a and β_0 in the original and simplified models. Small values of SE, MSE_1 and PE indicate the ML estimators perform satisfactory.

In both versions of the PGP models, we consider the following combinations of parameters: $\beta_0 = 0.5, 1.0, 1.5$, $a = 0.92, 0.96, 1.0, 1.04, 1.08$ and $n = 50, 100$. The simulation results are reported in Table 3.1 and Table 3.2 respectively for the original and simplified models.

==========

Table 3.1 and Table 3.2 about here

==========

In the following discussion, we write "$a \ll 1$" to mean "a is much smaller than 1" and "$a \gg 1$" to mean "a is much larger than 1". For the original model, we find that the precisions of $\hat{a}$ and $\hat{\beta}_0$ are similar across different levels of β_0 but they decrease (higher SE) with smaller n if a is fixed and larger a if n is fixed. These findings are obvious because smaller data size n implies less information in the data and larger a implies smaller W_t when t gets larger and hence the variability of W_t will be relatively larger, thus affecting the precisions of $\hat{a}$ and $\hat{\beta}_0$.

For the original and simplified versions of the PGP model, the variability of W_t increases with (1) larger β_0; (2) smaller a and larger n if $a < 1$ (increasing trend) and (3) larger a and smaller n if $a > 1$ according to (3.3) and (3.5) respectively. Hence we find that the accuracy of a in the original model decreases (higher MSE_1 and PE) if (1) n is larger (longer trend) if $a \ll 1$; (2) β_0 increases (higher initial level) if $a \ll 1$ (increasing trend) and (3) β_0 decreases (lower initial level) if $a \gg 1$ (decreasing trend). These results imply that $E(W_t)$ will either become much larger for longer increasing trend or very small for longer decreasing trend as time t goes on. Similar results hold for β_0 except for case (2) in which the accuracy of β_0 decreases with smaller β_0 (lower initial level) if $a \ll 1$ (increasing trend).

On the other hand, the precision and accuracy for both $\hat{a}$ and $\hat{\beta}_0$ in the simplified PGP models are very high (low SE, MSE_1 and PE) and their performance in SE, MSE_1 and PE are similar in general. This is due to the equidispersion property of the model. As a whole, the ML estimators perform more satisfactory when the n is larger (longer trend).

3.6 Real data analysis

We apply the PGP models to analyze two data sets: (1) coal mining disaster data in Great Britain and (2) Severe Acute Respiratory Syndrome (SARS) data in Hong Kong, which are discussed in details in Section 3.6.1.1 and Section 3.6.2.1 accordingly. For each data set, both original and simplified PGP models are applied. The parameters in each model are estimated by three methods including the LSE, the ML and the Bayesian methods. Standard errors for the parameters using the LSE and the ML methods are given by the square root of the diagonal elements in the Fisher information matrix $SSE''(\hat{\boldsymbol{\theta}}_{LSE})$ [equations (3.13) to (3.15)] and $-\ell''(\hat{\boldsymbol{\theta}}_{ML})$ [equations (3.16) to (3.18) and (3.19) to (3.21)] for the original and simplified models respectively, while those for the Bayesian method are given by the sample standard deviations of the posterior samples.

To choose the best models, we propose two goodness-of-fit (GOF) measures:

$$\text{Mean squares of error:} \qquad MSE_2 = \frac{1}{n}\sum_{t=1}^{n}(w_t - \hat{w}_t)^2 \tag{3.11}$$

$$\text{Posterior expected utility:} \qquad U = \frac{1}{n}\sum_{t=1}^{n}\ln[f_{ht}(w_t|\,\hat{\boldsymbol{\theta}}_h)] \tag{3.12}$$

where $\hat{w}_t = E_h(W_t)$, is given by (3.3) and (3.5), and the likelihood function $f_{ht}(w_t|\,\hat{\boldsymbol{\theta}}_h)$, evaluated at $\hat{\boldsymbol{\theta}}_h$, are given by (3.7) and (3.9) respectively for $h = o, s$. The quantity U is simplified from

$$U' = \frac{1}{Mn}\sum_{r=1}^{M}\sum_{t=1}^{n}\ln[f_{htr}(w_t|\,\boldsymbol{\theta}_{hr})]$$

which is originally proposed by Walker and Gutiérrez-Peña (61) for Bayesian inference. The likelihood function $f_{htr}(w_t|\,\boldsymbol{\theta}_{hr})$ is evaluated at the r^{th} set of posterior sample $\boldsymbol{\theta}_{hr}$ for the parameters $\boldsymbol{\theta}_h$ and M is the number of posterior samples. Obviously, a model with the largest U and/or the least MSE_2 is chosen to be the best model as it indicates the highest likelihood and/or the least-square errors of the data given the model.

3.6.1 Coal mining disaster data

3.6.1.1 The data

The coal mining disaster data, reported in Andrews and Herzberg (2), contains 190 inter-arrival times in days between successive disasters in Great Britain. It was originally reported by Maguire *et al.* (42) and was studied by Cox and Lewis

(12) in the analysis of trend data. Later, the data was extended so that it covers the period from 15^{th} May 1851 to 22^{nd} March 1962 inclusive of totally 40550 days or 112 years. The data has been studied by Lam (31), Lam and Chan (35), Lam *et al.* (40) and Chan *et al.* (8) using the GP models. They found that the inter-arrival times between successive disasters follow an increasing trend. Chan and Leung (9) and Ho (25) also transformed the positive continuous data to binary data, which is the indicator of disasters in each quarter of the years. and analyzed it using the BGP models.

In this analysis, we transform the inter-arrival times in days to number of coal mining disasters per year. A total number of $n = 112$ observations is obtained. The data is given in Appendix 3.5. The number of disasters fluctuates throughout the period with a decreasing trend. Moreover, the data contains no information for setting up covariate effects in the mean and/or ratio functions. Therefore, we fit both the original and simplified basic PGP models on the count data using the LSE, the ML and the Bayesian methods.

3.6.1.2 Results and comments

All the parameter estimates are statistically significant. For both original and simplified PGP model, the parameter estimates for the ratio a and initial level β_0 are quite consistent using the LSE and ML methods regardless of the versions of the model as shown in Table 3.3. On the other hand, the estimates using Bayesian method are slightly different. The estimated ratios $\hat{a}$ are greater than 1, implying

that the number of coal mining disasters decreases throughout the period. The decrease may due to the technological advancement of the coal mining machinery and the establishment of safety measures. It is noted that the LSE method gives the same results for both versions of the model because their mean $E_h(W_t)$ and thus the SSE equations are distribution invariant. Plots of the observed (W_t) and predicted $(E_h(W_t), h = o, s)$ numbers of disasters in Figure 3.3 show different decreasing trends.

==========

Table 3.3 and Figure 3.3 about here

==========

Based on the results in Table 3.3, the PGP model using the LSE method gives the best MSE_2, while the simplified model using the ML method gives the best U. Despite the similar performance according to MSE_2, the simplified models always give better U. Since the coal mining disaster data does not contain many zeros (32 zeros out of 112 observations), the original model, which is suitable for modeling data with more zeros, does not give better fit according to U. Among the simplified PGP models using different methods, the one using the ML method is chosen to be the best model as it has the second best MSE_2 and the best U, whereas the original model using the Bayesian method gives the worst fit according to both MSE_2 and U.

3.6.2 Hong Kong SARS data

3.6.2.1 The data

The Severe Acute Respiratory Syndrome (SARS), is an epidemic that has recently reported in Asia, North America and Europe. SARS was first recognized as a global health threat in mid-March 2003. The first known cases of SARS occurred in Guangdong Province, in the city of Foshan, of the People's Republic of China (PRC), on 16^{th} November 2002 and the first sufferer was a farmer. Early in the epidemic outbreak, the PRC discouraged its press from reporting on SARS and delayed in reporting its early situation of outbreak to the World Health Organization (WHO). This caused the world-wide SARS outbreak which was started on 21^{st} February 2003, when a Chinese doctor who had treated SARS cases in Guangdong stayed in the Hong Kong Metropole hotel during 12^{th} February to 2^{nd} March 2003. He transmitted the epidemic to seven persons who stayed in or visited the same floor of the hotel. The infected persons included three visitors from Singapore, two from Canada, one from the PRC, and a local Hong Kong resident. The local Hong Kong resident was believed to be the index case for an outbreak in the Prince of Wales Hospital, Hong Kong. During March to June 2003, Hong Kong was seriously affected by SARS. Various precautionary measures like contact tracing, quarantine, thermal screen to outgoing and incoming passengers, closing of schools, and public health measures such as wearing masks, frequent hand washing, avoidance of crowded places and hospitals, use of bleach, were implemented. On 23^{rd} June 2003, Hong Kong was removed from the WHO's

list of "infected areas". The detailed information for the epidemic outbreak can be found in the following web sites:

1. http://www.info.gov.hk/info/sars/eindex.htm
2. http://www.hku.hk/sars
3. http://www.hku.hk/statistics/SARS

A complete data set for four affected regions including Hong Kong used in our analysis can be downloaded from the third website. The Hong Kong SARS data contains information of the daily number of infected cases, death cases, recover cases as well as the cases in treatment from 12^{th} March to 11^{th} June 2003. To demonstrate the application of the PGP model, we target on the daily number of infected cases, with $n = 92$ observations. There is a total number of 1755 infected cases with an average of 19 daily infected cases and the average temperature is 24.4 ^oC. The daily infected cases increased rapidly at the beginning of the epidemic outbreak to reach a maximum of 80 cases on 31^{st} March 2003 and another peak of 61 cases on 10^{th} April, 2003. Then it dropped gradually until 11^{th} June. The data is given in Appendix 3.6 at the end of this chapter.

3.6.2.2 Results and comments

The parameter estimate $\hat{a} > 1$ in Table 3.4 suggests that the number of daily infected cases decreases gradually after the epidemic outbreak. Plots of the fitted outcomes $E_h(W_t), h = o, s$ and observed outcomes W_t in Figure 3.4 show different

decreasing trends. The value of $\hat{a}$ is close to 1 which means that the trend declines gently. It is because the Hong Kong government was unable to control the spreading of the SARS at the initial stage: it took a certain period of time for the implementation of various precautionary and public health measures and it took another incubation period for these measures to become effective.

==========

Table 3.4 about here

==========

In Table 3.4, we find that the ML method and Bayesian method give similar parameter estimates within each version, however, the parameter estimates for different versions are entirely different: relative to the SV of the model, $E_o(W_t)$ for the original model declines from a higher level at a higher rate. Moreover the parameter estimates using the LSE method are different from the parameter estimates of both versions: $E_h(W_t)$ declines from the lowest level at a slowest rate. For the GOF measures, the original model always has a much better (larger) U than that of the simplified model because the latter has a rather restrictive equidispersion assumption on the variance. Hence the likelihood function becomes very small when the expected value $E_s(W_t)$ is deviated from the observed count W_t which is the case during the early stage of the SARS outbreak. This explains why U is relatively worse in the simplified model. Moreover, Figure 3.4 shows that the simplified models using the ML and Bayesian methods fit the daily infected cases better at the early stage when the observed cases is higher while the original

models fit the daily infected cases better at the later stage. This again illustrates why the simplified model always has a better (smaller) MSE_2. On the other hand, it is obvious that the LSE method gives the best MSE_2 regardless of the version of the PGP model.

Nevertheless, the PGP models using the LSE method regardless of the version are selected as the best model as it has the best MSE_2 and other models give either worse U or MSE_2. The poor performance of the ML and Bayesian methods for both versions of the model may due to the fact that both of their pmf as given by (3.2) and (3.4) do not describe the data W_t well. However, this model is still not very satisfactory with a rather high $MSE_2 = 229.22$.

==========

Figure 3.4 about here

==========

3.7 Discussion

Results from the two data analyses suggest that in the process of minimizing the SSE, the LSE method tends to give the expected outcome $E_h(W_t)$ which stays close to the observed outcome throughout the whole trend. Hence, relative to the ML and Bayesian methods, a decreasing (increasing) trend using the LSE method will decrease (increase) from a lower (higher) level at a lower rate. Hence, the LSE method always gives better MSE_2 whereas the ML method tends to give

better U.

Moreover, we also observe that the original PGP model using the ML method tends to fit worse when the count W_t is larger. This characteristic can be explained from the density function (3.2) which gives a lower probability and hence a lower effect on the likelihood function for larger count. This explains why the original model using the ML method usually gives worse MSE_2. However, if the higher errors from the higher counts incur higher costs, we will have to adjust the likelihood function to reduce such errors. A possible way is to adopt a weighted likelihood function on the observed counts. This should be considered for further research. On the other hand, the density function (3.4) for the simplified model gives a lower probability and hence a lower effect on the likelihood function only when the count W_t is further away from $E(W_t)$. This illustrates why the simplified model using the ML method usually gives better MSE_2 as compared with the original model and why the parameter estimates are closer to those using the LSE method.

Finally, Figure 3.4 shows clearly that fitting a monotone trend using the basic PGP model to the SARS data is not satisfactory. In the next chapter, we consider time-varying covariate effects in the mean and ratio functions.

Appendix 3.1 Derivation of the pmf of W_t in the original PGP model

If $Y_t \sim \text{Exp}(1/\mu_t)$ with mean μ_t, the pdf of W_t is

$$\begin{aligned}
f_{ot}(w_t) &= \int_0^\infty \frac{\exp(-x_t){x_t}^{w_t}}{w_t!}\ \mu_t^{-1}a_t^{t-1}\exp(-\mu_t^{-1}a_t^{t-1}x_t)\ dx_t \\
&= \frac{\mu_t^{-1}a_t^{t-1}}{w_t!}\int_0^\infty {x_t}^{w_t}\ d\left[\frac{\exp[-x_t(\mu_t^{-1}{a_t}^{t-1}+1)]}{-(\mu_t^{-1}a_t^{t-1}+1)}\right] \\
&= \left(\frac{\mu_t^{-1}a_t^{t-1}}{w_t!}\right)\left[\frac{1}{-(\mu_t^{-1}a_t^{t-1}+1)}\right]\Big\{\left[{x_t}^{w_t}\exp[-x_t(\mu_t^{-1}{a_t}^{t-1}+1)]\right]_0^\infty - \\
&\qquad\qquad \int_0^\infty \exp[-x_t(\mu_t^{-1}{a_t}^{t-1}+1)]\ d{x_t}^{w_t}\Big\} \\
&= \left(\frac{\mu_t^{-1}a_t^{t-1}}{w_t!}\right)\left[\frac{w_t(w_t-1)}{(\mu_t^{-1}a_t^{t-1}+1)^2}\right]\int_0^\infty \exp[-x_t(\mu_t^{-1}{a_t}^{t-1}+1)]\ {x_t}^{w_t-2}\ dx_t \\
&\ \ \vdots \\
&= \left(\frac{\mu_t^{-1}a_t^{t-1}}{w_t!}\right)\left[\frac{w_t!}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t}}\right]\int_0^\infty \exp[-x_t(\mu_t^{-1}{a_t}^{t-1}+1)]\ dx_t \\
&= \left(\frac{\mu_t^{-1}a_t^{t-1}}{w_t!}\right)\left[\frac{w_t!}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t}}\right]\frac{1}{-(\mu_t a_t^{t-1}+1)}\left[\exp[-x_t(\mu_t{a_t}^{t-1}+1)]\right]_0^\infty \\
&= \frac{\mu_t^{-1}a_t^{t-1}}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t+1}}
\end{aligned}$$

It is easy to check that $\displaystyle\sum_{w_t=0}^{\infty}\frac{\mu_t^{-1}a_t^{t-1}}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t+1}} = 1$.

Appendix 3.2 Derivation of the mean and variance of W_t in the original PGP model

$$
\begin{aligned}
E(W_t) &= \mu_t^{-1}a_t^{t-1}\sum_{w_t=1}^{\infty}\frac{w_t}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t+1}} \\
&= \mu_t^{-1}a_t^{t-1}\left[\frac{1}{\mu_t^{-1}a_t^{t-1}+1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^3}+\cdots\right)\right. \\
&\qquad +\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\cdots\right)+ \\
&\qquad \left.+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^3}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\cdots\right)+\cdots\right] \\
&= \mu_t^{-1}a_t^{t-1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^3}+\cdots\right)^2 \\
&= \mu_t^{-1}a_t^{t-1}\left(\frac{\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}{1-\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}\right)^2 \\
&= \frac{\mu_t}{a_t^{t-1}}
\end{aligned}
$$

$$
\begin{aligned}
E(W_t^2) &= \mu_t^{-1}a_t^{t-1}\sum_{w_t=1}^{\infty}\frac{w_t^2}{(\mu_t^{-1}a_t^{t-1}+1)^{w_t+1}} \\
&= \mu_t^{-1}a_t^{t-1}\left[\frac{1}{\mu_t^{-1}a_t^{t-1}+1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^3}+\cdots\right)\right. \\
&\qquad +\frac{3}{(\mu_t^{-1}a_t^{t-1}+1)^2}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\cdots\right)+ \\
&\qquad \left.+\frac{5}{(\mu_t^{-1}a_t^{t-1}+1)^3}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\cdots\right)+\cdots\right] \\
&= \mu_t^{-1}a_t^{t-1}\left(\frac{\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}{1-\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}\right)\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{3}{(\mu_t^{-1}a_t^{t-1}+1)^2}\right. \\
&\qquad \left.+\frac{5}{(\mu_t^{-1}a_t^{t-1}+1)^3}+\cdots\right) \\
&= \mu_t^{-1}a_t^{t-1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}}\right)\left[\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^3}+\cdots\right)\right. \\
&\qquad +\frac{2}{\mu_t^{-1}a_t^{t-1}+1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\frac{1}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\cdots\right)+ \\
&\qquad \left.+\frac{2}{(\mu_t^{-1}a_t^{t-1}+1)^2}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}+1}+\cdots\right)+\cdots\right] \\
&= \mu_t^{-1}a_t^{t-1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}}\right)\left(\frac{\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}{1-\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}\right) \\
&\qquad \left(1+\frac{2}{\mu_t^{-1}a_t^{t-1}+1}+\frac{2}{(\mu_t^{-1}a_t^{t-1}+1)^2}+\cdots\right) \\
&= \mu_t^{-1}a_t^{t-1}\left(\frac{1}{\mu_t^{-1}a_t^{t-1}}\right)^2\left(1+\frac{\frac{2}{\mu_t^{-1}a_t^{t-1}+1}}{1-\frac{1}{\mu_t^{-1}a_t^{t-1}+1}}\right) \\
&= \frac{2+\mu_t^{-1}a_t^{t-1}}{(\mu_t^{-1}a_t^{t-1})^2} \\
Var(W_t) &= E(W_t^2)-E(W_t)^2 = \frac{a_t^{t-1}/\mu_t+1}{(a_t^{t-1}/\mu_t)^2} = E(W_t)+E(W_t)^2
\end{aligned}
$$

Appendix 3.3 Derivation of first and second order derivative functions in the NR iterative method for the PGP models using LSE method

On differentiating (3.6), the score functions for a and β_0 are

$$\begin{aligned}\frac{\partial SSE}{\partial a} &= 2\sum_{t=1}^{n} z^*_{a0t}\hat{w}_t(w_t - \hat{w}_t) \\ \frac{\partial SSE}{\partial \beta_0} &= -2\sum_{t=1}^{n} z^*_{\mu 0t}\hat{w}_t(w_t - \hat{w}_t)\end{aligned}$$

where

$$z^*_{a0t} = \frac{t-1}{a} \quad \text{and} \quad z^*_{\mu 0t} = z_{\mu 0t} = 1$$

The second order derivative functions which are required in NR iterative method are

$$\frac{\partial^2 SSE}{\partial a^2} = -2\sum_{t=1}^{n} \frac{z^*_{a0t}\hat{w}_t}{a}\left[tw_t - (2t-1)\hat{w}_t\right] \tag{3.13}$$

$$\frac{\partial^2 SSE}{\partial {\beta_0}^2} = -2\sum_{i=1}^{n} z^*_{\mu 0t}\hat{w}_t(w_t - 2\hat{w}_t) \tag{3.14}$$

$$\frac{\partial^2 SSE}{\partial a \partial \beta_0} = 2\sum_{i=1}^{n} z^*_{a0t} z^*_{\mu 0t}\hat{w}_t(w_t - 2\hat{w}_t) \tag{3.15}$$

Appendix 3.4 Derivation of first and second order derivative functions in the NR iterative method for the PGP models using ML method

In the original PGP model, by differentiating the log-likelihood function (3.8), we obtain the following first order derivative functions for $\boldsymbol{\theta}_o$.

$$\begin{aligned}
\frac{\partial \ell_o}{\partial a} &= \sum_{t=1}^{n} \frac{z^*_{a0t}(\hat{w}_t - w_t)}{\hat{w}_t + 1} \\
\frac{\partial \ell_o}{\partial \beta_0} &= -\sum_{t=1}^{n} \frac{z^*_{\mu 0t}(\hat{w}_t - w_t)}{\hat{w}_t + 1}
\end{aligned}$$

The second order derivative functions for the NR iterative method are:

$$\frac{\partial^2 \ell_o}{\partial a^2} = -\sum_{t=1}^{n} \frac{z^*_{a0t}}{a} \left[\frac{\hat{w}_t^2 - w_t + i\hat{w}_t + (t-2)\hat{w}_t w_t}{(\hat{w}_t + 1)^2} \right] \tag{3.16}$$

$$\frac{\partial^2 \ell_o}{\partial {\beta_0}^2} = -\sum_{t=1}^{n} \frac{{z^*_{\mu 0t}}^2 \hat{w}_t (w_t + 1)}{(\hat{w}_t + 1)^2} \tag{3.17}$$

$$\frac{\partial^2 \ell_o}{\partial a \partial \beta_0} = \sum_{t=1}^{n} \frac{z^*_{a0t} z^*_{\mu 0t} \hat{w}_t (w_t + 1)}{(\hat{w}_t + 1)^2} \tag{3.18}$$

For the simplified PGP model, the first order and second order derivative functions for the NR iterative method are

$$\begin{aligned}
\frac{\partial \ell_s}{\partial a} &= \sum_{t=1}^{n} z^*_{a0t}(\hat{w}_t - w_t) \\
\frac{\partial \ell_s}{\partial \beta_0} &= -\sum_{t=1}^{n} z^*_{\mu 0t}(\hat{w}_t - w_t) \\
\frac{\partial^2 \ell_s}{\partial a^2} &= -\sum_{t=1}^{n} \frac{z^*_{a0t}}{a}(t\hat{w}_t - w_t) \qquad (3.19) \\
\frac{\partial^2 \ell_s}{\partial {\beta_0}^2} &= -\sum_{t=1}^{n} {z^*_{\mu 0t}}^2 \hat{w}_t \qquad (3.20) \\
\frac{\partial^2 \ell_s}{\partial a \partial \beta_0} &= \sum_{t=1}^{n} z^*_{a0t} z^*_{\mu 0t} \hat{w}_t \qquad (3.21)
\end{aligned}$$

Appendix 3.5 Number of coal mining disasters in Great Britain during 1851-1962

Year	t	Count	Year	t	Count	Year	t	Count	Year	t	Count
1851	1	2	1879	29	3	1907	57	0	1935	85	2
1852	2	6	1880	30	3	1908	58	3	1936	86	1
1853	3	5	1881	31	2	1909	59	1	1937	87	1
1854	4	1	1882	32	5	1910	60	2	1938	88	1
1855	5	0	1883	33	2	1911	61	1	1939	89	0
1856	6	4	1884	34	2	1912	62	1	1940	90	3
1857	7	3	1885	35	3	1913	63	1	1941	91	3
1858	8	3	1886	36	4	1914	64	1	1942	92	3
1859	9	1	1887	37	3	1915	65	0	1943	93	0
1860	10	3	1888	38	1	1916	66	1	1944	94	0
1861	11	5	1889	39	3	1917	67	0	1945	95	0
1862	12	2	1890	40	2	1918	68	1	1946	96	0
1863	13	4	1891	41	2	1919	69	0	1947	97	5
1864	14	2	1892	42	1	1920	70	0	1948	98	0
1865	15	1	1893	43	1	1921	71	0	1949	99	0
1866	16	4	1894	44	1	1922	72	2	1950	100	0
1867	17	4	1895	45	1	1923	73	1	1951	101	1
1868	18	3	1896	46	3	1924	74	0	1952	102	0
1869	19	6	1897	47	0	1925	75	0	1953	103	0
1870	20	5	1898	48	0	1926	76	0	1954	104	0
1871	21	5	1899	49	1	1927	77	1	1955	105	0
1872	22	3	1900	50	0	1928	78	1	1956	106	0
1873	23	1	1901	51	1	1929	79	0	1957	107	0
1874	24	2	1902	52	1	1930	80	2	1958	108	1
1875	25	3	1903	53	0	1931	81	1	1959	109	0
1876	26	3	1904	54	0	1932	82	3	1960	110	1
1877	27	6	1905	55	3	1933	83	2	1961	111	0
1878	28	5	1906	56	1	1934	84	2	1962	112	1

Appendix 3.6 Number of daily infected SARS cases in Hong Kong during March to June 2003

Date	Count	^oC	Date	Count	^oC	Date	Count	^oC	Date	Count	^oC
12-Mar	10	17.8	4-Apr	27	23.7	27-Apr	16	25	20-May	4	26.9
13-Mar	14	19.1	5-Apr	39	20.2	28-Apr	14	25.3	21-May	1	27.2
14-Mar	5	18.2	6-Apr	42	19.4	29-Apr	15	26.1	22-May	3	27.5
15-Mar	8	20.2	7-Apr	41	21.1	30-Apr	17	25.5	23-May	2	28
16-Mar	5	22.1	8-Apr	45	21.6	1-May	11	23.5	24-May	0	28.5
17-Mar	53	23.2	9-Apr	42	19.1	2-May	11	24.1	25-May	1	27.5
18-Mar	28	20.3	10-Apr	28	18.8	3-May	10	24.6	26-May	1	27.1
19-Mar	27	16.9	11-Apr	61	21.8	4-May	8	25.5	27-May	2	26.4
20-Mar	23	15.5	12-Apr	49	25.8	5-May	8	24.9	28-May	2	25.8
21-Mar	30	16.8	13-Apr	42	26.6	6-May	9	28.1	29-May	2	26.4
22-Mar	19	17.7	14-Apr	40	23.3	7-May	8	28.5	30-May	4	28.1
23-Mar	20	17.9	15-Apr	42	21.4	8-May	7	28.7	31-May	3	28.1
24-Mar	18	18.9	16-Apr	36	21.6	9-May	6	25.5	1-Jun	3	27.3
25-Mar	26	20.2	17-Apr	29	23.5	10-May	7	25.4	2-Jun	4	27.5
26-Mar	30	21	18-Apr	30	24.9	11-May	4	25.9	3-Jun	1	27.4
27-Mar	51	23.3	19-Apr	31	25.4	12-May	5	26.3	4-Jun	1	27.6
28-Mar	58	20.6	20-Apr	22	25.4	13-May	6	27.2	5-Jun	0	27.7
29-Mar	45	19.9	21-Apr	22	25	14-May	9	28	6-Jun	2	27.8
30-Mar	60	20.9	22-Apr	32	24.5	15-May	5	28.6	7-Jun	2	27.2
31-Mar	80	24	23-Apr	24	25.8	16-May	3	29.1	8-Jun	0	27
1-Apr	75	25.5	24-Apr	30	26.5	17-May	4	29.2	9-Jun	1	25.7
2-Apr	23	25.8	25-Apr	22	27	18-May	3	26.7	10-Jun	1	25.9
3-Apr	26	26.2	26-Apr	17	26.3	19-May	1	25.4	11-Jun	1	25.7

Table 3.1: Simulation studies for the original PGP model

	$n = 50$		$n = 50$		$n = 50$		$n = 100$		$n = 100$		$n = 100$	
	$\beta_0 = 0.5$		$\beta_0 = 1.0$		$\beta_0 = 1.5$		$\beta_0 = 0.5$		$\beta_0 = 1.0$		$\beta_0 = 1.5$	
True a	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$
0.92	0.9200	0.4557	0.9202	0.9640	0.9202	1.4658	0.9293	0.8075	0.9335	1.4343	0.9387	2.0844
SE	*0.0100*	*0.3235*	*0.0097*	*0.3087*	*0.0095*	*0.2993*	*0.0038*	*0.2333*	*0.0039*	*0.2336*	*0.0040*	*0.2368*
MSE_1	*0.0001*	*0.0877*	*0.0001*	*0.0783*	*0.0001*	*0.0724*	*0.0001*	*0.1300*	*0.0002*	*0.2233*	*0.0004*	*0.3759*
PE	*-0.0001*	*-0.0885*	*0.0002*	*-0.0360*	*0.0003*	*-0.0228*	*0.0101*	*†0.6150*	*0.0147*	*†0.4343*	*0.0203*	*†0.3896*
0.96	0.9599	0.4517	0.9601	0.9603	0.9601	1.4611	0.9600	0.4838	0.9601	0.9853	0.9601	1.4848
SE	*0.0109*	*0.3373*	*0.0104*	*0.3172*	*0.0101*	*0.3045*	*0.0037*	*0.2297*	*0.0036*	*0.2195*	*0.0035*	*0.2130*
MSE_1	*0.0001*	*0.0941*	*0.0001*	*0.0824*	*0.0001*	*0.0752*	*0.0000*	*0.0384*	*0.0000*	*0.0356*	*0.0000*	*0.0331*
PE	*-0.0001*	*-0.0966*	*0.0001*	*-0.0397*	*0.0001*	*-0.0260*	*0.0000*	*-0.0324*	*0.0000*	*-0.0147*	*0.0000*	*-0.0102*
1.00	1.0001	0.4494	1.0002	0.9568	1.0002	1.4593	1.0000	0.4794	1.0001	0.9815	1.0001	1.4826
SE	*0.0128*	*0.3631*	*0.0118*	*0.3335*	*0.0111*	*0.3147*	*0.0049*	*0.2565*	*0.0041*	*0.2364*	*0.0039*	*0.2234*
MSE_1	*0.0001*	*0.1110*	*0.0001*	*0.0896*	*0.0001*	*0.0821*	*0.0000*	*0.0464*	*0.0000*	*0.0396*	*0.0000*	*0.0362*
PE	*0.0001*	*†0.1013*	*0.0003*	*-0.0432*	*0.0002*	*-0.0272*	*0.0000*	*-0.0411*	*0.0001*	*-0.0185*	*0.0001*	*-0.0116*
1.04	1.0404	0.4321	1.0408	0.9510	1.0405	1.4556	1.0401	0.4522	1.0396	0.9504	1.0397	1.4600
SE	*0.0179*	*0.4095*	*0.0155*	*0.3660*	*0.0137*	*0.3367*	*0.0095*	*0.3332*	*0.0078*	*0.2933*	*0.0067*	*0.2658*
MSE_1	*0.0003*	*0.1409*	*0.0002*	*0.1118*	*0.0002*	*0.0931*	*0.0001*	*0.0899*	*0.0000*	*0.0664*	*0.0000*	*0.0513*
PE	*0.0004*	*†-0.1357*	*0.0007*	*-0.0490*	*0.0005*	*-0.0296*	*0.0001*	*-0.0955*	*-0.0004*	*-0.0496*	*-0.0003*	*-0.0267*
1.08	1.0830	0.4268	1.0811	0.9399	1.0804	1.4412	1.0849	0.4657	1.0818	0.9544	1.0809	1.4531
SE	*0.0288*	*0.4736*	*0.0233*	*0.4145*	*0.0197*	*0.3746*	*0.0233*	*0.4355*	*0.0182*	*0.3781*	*0.0148*	*0.3378*
MSE_1	*0.0008*	*0.1757*	*0.0004*	*0.1381*	*0.0004*	*0.1249*	*0.0001*	*0.1393*	*0.0003*	*0.1097*	*0.0002*	*0.0888*
PE	*0.0028*	*†0.1465*	*0.0011*	*-0.0601*	*0.0004*	*-0.0392*	*0.0045*	*-0.0687*	*0.0016*	*-0.0456*	*0.0008*	*-0.0312*

† Absolute value of PE is greater than 10%.

Table 3.2: Simulation studies for the simplified PGP model

	$n = 50$		$n = 50$		$n = 50$		$n = 100$		$n = 100$		$n = 100$	
	$\beta_0 = 0.5$		$\beta_0 = 1.0$		$\beta_0 = 1.5$		$\beta_0 = 0.5$		$\beta_0 = 1.0$		$\beta_0 = 1.5$	
True a	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$	$\hat{a}$	$\hat{\beta}_0$
0.92	0.9199	0.4915	0.9197	0.9834	0.9200	1.4968	0.9200	0.4964	0.9200	1.0035	0.9200	1.4969
SE	*0.0026*	*0.1122*	*0.0020*	*0.0875*	*0.0016*	*0.0680*	*0.0003*	*0.0264*	*0.0002*	*0.0205*	*0.0002*	*0.0160*
MSE_1	*0.0000*	*0.0102*	*0.0000*	*0.0070*	*0.0000*	*0.0046*	*0.0000*	*0.0007*	*0.0000*	*0.0004*	*0.0000*	*0.0003*
PE	*-0.0002*	*-0.0170*	*-0.0004*	*-0.0166*	*-0.0000*	*-0.0021*	*-0.0000*	*0.0000*	*0.0000*	*0.0035*	*-0.0000*	*-0.0021*
0.96	0.9599	0.4967	0.9600	0.9991	0.9599	1.4952	0.9600	0.4996	0.9600	0.9998	0.9599	1.4953
SE	*0.0045*	*0.1646*	*0.0035*	*0.1281*	*0.0027*	*0.0998*	*0.0010*	*0.0802*	*0.0008*	*0.0625*	*0.0006*	*0.0487*
MSE_1	*0.0000*	*0.0229*	*0.0000*	*0.0124*	*0.0000*	*0.0085*	*0.0000*	*0.0072*	*0.0000*	*0.0040*	*0.0000*	*0.0023*
PE	*-0.0001*	*-0.0066*	*0.0000*	*-0.0009*	*-0.0001*	*-0.0032*	*-0.0000*	*-0.0007*	*-0.0000*	*-0.0002*	*-0.0001*	*-0.0031*
1.00	0.9999	0.4877	0.9999	0.9922	1.0000	1.4980	1.0000	0.4931	1.0001	0.9984	1.0000	1.4989
SE	*0.0077*	*0.2184*	*0.0060*	*0.1697*	*0.0046*	*0.1318*	*0.0027*	*0.1553*	*0.0021*	*0.1206*	*0.0016*	*0.0939*
MSE_1	*0.0001*	*0.0400*	*0.0000*	*0.0233*	*0.0000*	*0.0143*	*0.0000*	*0.0249*	*0.0000*	*0.0147*	*0.0000*	*0.0089*
PE	*-0.0001*	*-0.0246*	*-0.0001*	*-0.0078*	*-0.0000*	*-0.0013*	*0.0000*	*-0.0137*	*0.0000*	*-0.0016*	*0.0000*	*-0.0007*
1.04	1.0408	0.4906	1.0401	0.9894	1.0399	1.4936	1.0411	0.5088	1.0404	0.9995	1.0405	1.5034
SE	*0.0133*	*0.2698*	*0.0102*	*0.2094*	*0.0079*	*0.1626*	*0.0078*	*0.2296*	*0.0060*	*0.1786*	*0.0047*	*0.1389*
MSE_1	*0.0002*	*0.0719*	*0.0001*	*0.0367*	*0.0001*	*0.0218*	*0.0001*	*0.0567*	*0.0000*	*0.0347*	*0.0000*	*0.0200*
PE	*0.0008*	*-0.0187*	*0.0001*	*-0.0106*	*-0.0001*	*-0.0043*	*0.0011*	*0.0176*	*0.0004*	*-0.0005*	*0.0005*	*0.0023*
1.08	1.0807	0.4691	1.0813	0.9852	1.0810	1.4999	1.0851	0.5134	1.0825	1.0047	1.0815	1.5085
SE	*0.0225*	*0.3206*	*0.0174*	*0.2479*	*0.0134*	*0.1914*	*0.0194*	*0.3005*	*0.0145*	*0.2322*	*0.0111*	*0.1798*
MSE_1	*0.0005*	*0.0939*	*0.0003*	*0.0493*	*0.0002*	*0.0310*	*0.0005*	*0.1073*	*0.0002*	*0.0627*	*0.0001*	*0.0348*
PE	*0.0006*	*-0.0618*	*0.0012*	*-0.0148*	*0.0009*	*-0.0001*	*0.0047*	*0.0268*	*0.0023*	*0.0047*	*0.0014*	*0.0057*

† Absolute value of PE is greater than 10%.

Table 3.3: Parameter estimates and standard errors for the coal mining disaster data using PGP models

Model		Original PGP				Simplified PGP			
Method	Par.	Est.	SE	MSE_2	U	Est.	SE	MSE_2	U
LSE	β_0	1.3515	*0.5256*	1.7606		1.3515	*0.5256*	1.7606	
	a	1.0175	*0.0150*			1.0175	*0.0150*		
ML	β_0	1.3722	*0.2256*	1.7618	-1.6824	1.3727	*0.1179*	1.7618	-1.5442*
	a	1.0183	*0.0040*			1.0183	*0.0025*		
BAY	β_0	1.5040	*0.2673*	1.8391	-1.6967	1.4450	*0.1287*	1.7998	-1.5568
	a	1.0240	*0.0065*			1.0220	*0.0037*		

*Model with the largest U is chosen to be the best model.

Table 3.4: Parameter estimates and standard errors for the Hong Kong SARS data using PGP models

Model		Original PGP				Simplified PGP			
Method	Par.	Est.	SE	MSE_2	U	Est.	SE	MSE_2	U
LSE	β_0	3.6911	*0.0013*	229.22*		3.6911	*0.0013*	229.22*	
	a	1.0179	*0.0438*			1.0179	*0.0438*		
ML	β_0	4.5765	*0.2497*	603.14	-3.5788	3.9322	*0.0011*	255.30	-6.6919
	a	1.0463	*0.0054*			1.0274	*0.0377*		
BAY	β_0	4.5850	*0.2505*	612.57	-3.5790	3.9320	*0.0378*	254.05	-6.6936
	a	1.0460	*0.0055*			1.0270	*0.0011*		

*Model with the smallest MSE_2 is chosen to be the best model.

Figure 3.1: History plots of a and β_0 for the coal mining disaster data using original PGP model with Bayesian method

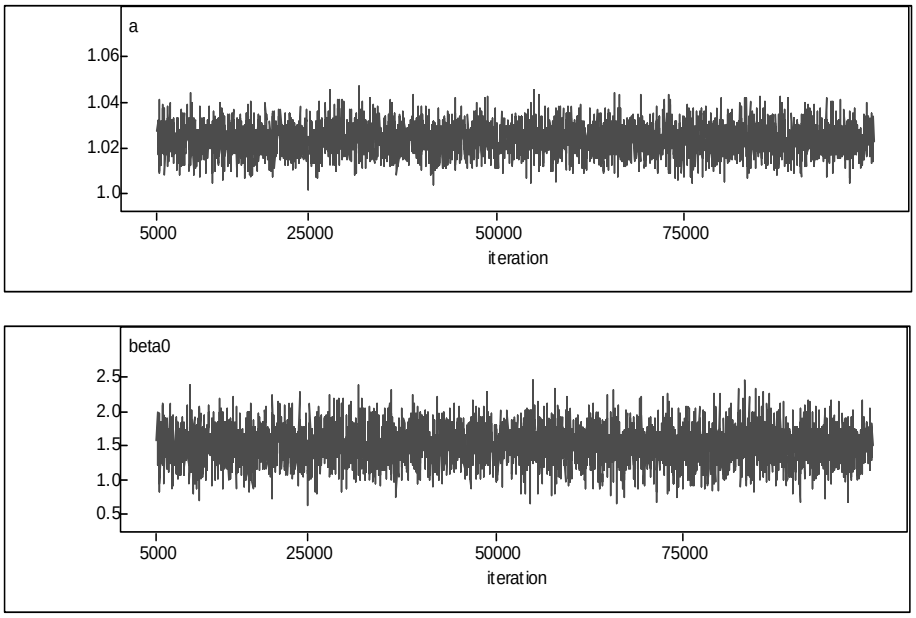


Figure 3.2: ACF plots of a and β_0 for the coal mining disaster data using original PGP model with Bayesian method

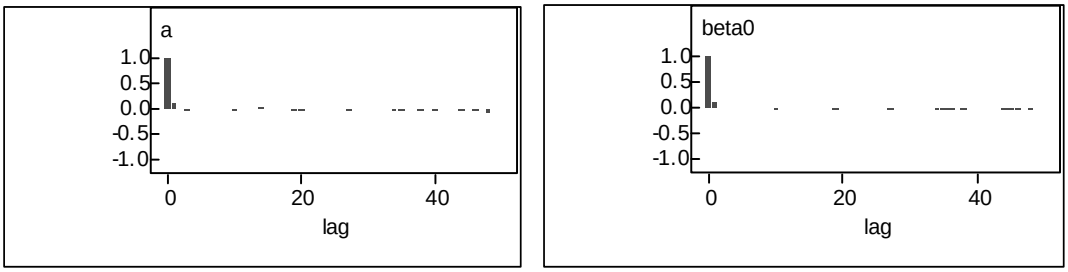
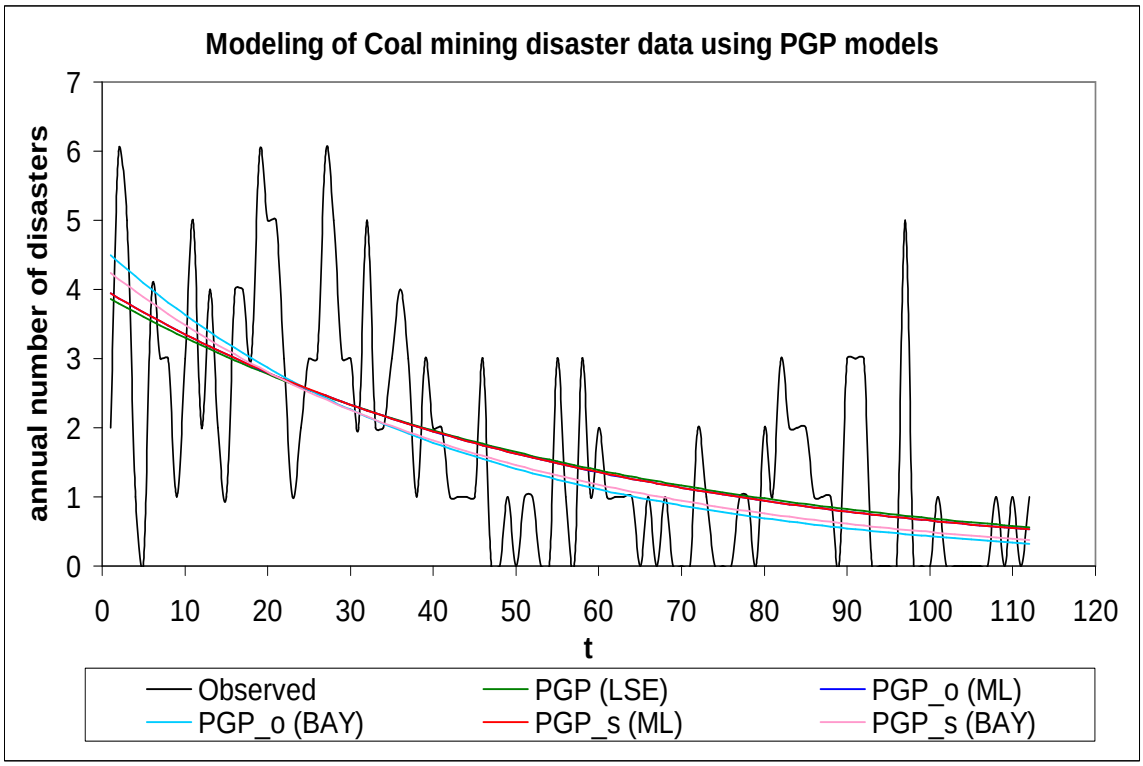
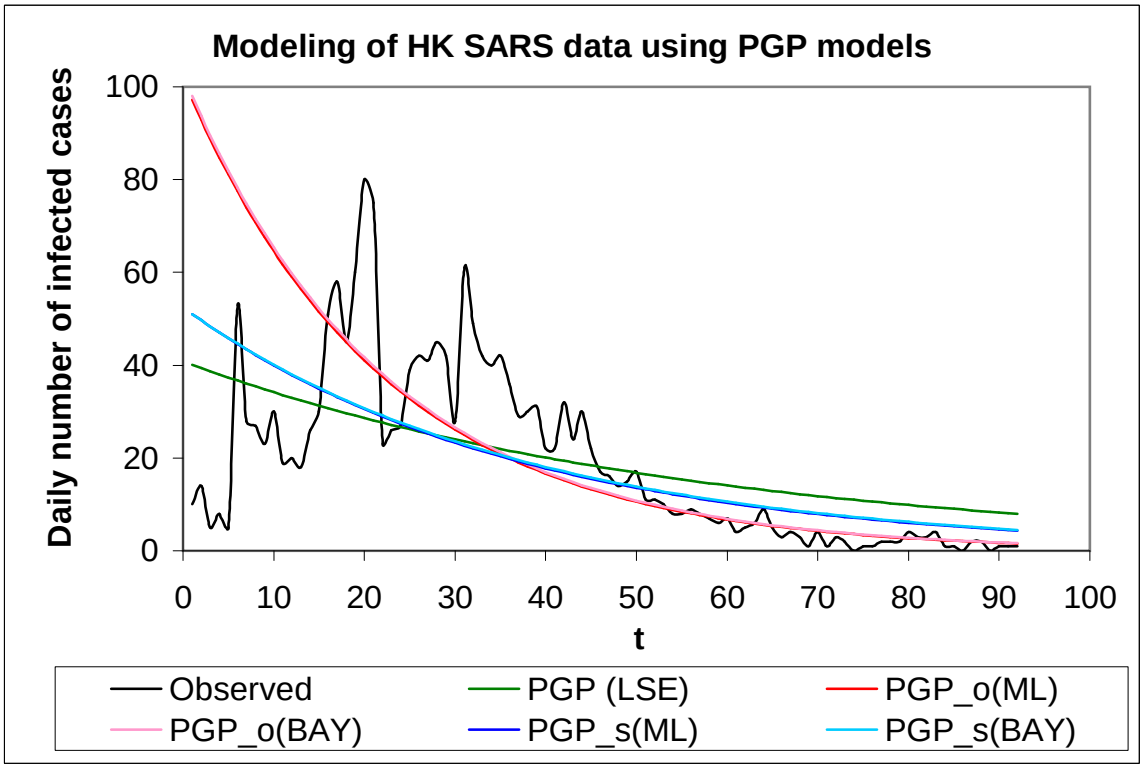


Figure 3.3: Plot of the observed w_t and expected $\hat{w}_t$ frequencies of coal mining disasters in Great Britain using PGP models



Remarks: (1) 'PGP_o' and 'PGP_s' stands for original and simplified PGP models accordingly. (2) The PGP_o (blue) and PGP_s (red) models using the ML method are overlapped as their parameter estimates are similar as shown in Table 3.3.

Figure 3.4: Plot of the observed w_t and expected $\hat{w}_t$ daily infected SARS cases in Hong Kong using PGP models



Remarks: (1) 'PGP_o' and 'PGP_s' stands for original and simplified PGP models accordingly. (2) The PGP_o models using the ML method (red) and Bayesian method (pink) are overlapped and the PGP_s models using the ML method (blue) and Bayesian method (light blue) are overlapped as their parameter estimates are similar as shown in Table 3.4.

Chapter 4

Adaptive Poisson Geometric Process Model

4.1 The model

4.1.1 The definition

The assumption that the RP follows an exponential distribution with constant mean μ over time is too restrictive because it fails to account for some time-varying effects in the data. Similar to the GP model, the PGP model can also be extended to the adaptive PGP (APGP) model by adopting a linear function of covariates log-linked to the mean function μ_t as below:

$$\mu_t = \exp(\eta_{\mu t}) = \exp(\beta_{\mu 0} z_{\mu 0 t} + \beta_{\mu 1} z_{\mu 1 t} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu t}) \tag{4.1}$$

where $z_{\mu 0t} = 1$ and $z_{\mu kt},\ k = 1, \ldots, q_\mu$ are covariates. This change makes the underlying process $\{Y_t\}$ no longer IID, instead, it becomes a stochastic process (SP) in general with $E(Y_t) = \mu_t$. Similarly, we can perform the same extension to the ratio function a_t by log-linking a linear function of covariates to the ratio function a_t as below:

$$a_t = \exp(\eta_{at}) = \exp(\beta_{a0} z_{a0t} + \beta_{a1} z_{a1t} + \cdots + \beta_{aq_a} z_{aq_a t}) \tag{4.2}$$

where $z_{a0t} = 1$ and $z_{akt},\ k = 1, \ldots, q_a$ are covariates. The vector of model parameters $\boldsymbol{\theta} = (\beta_{a0}, \beta_{a1}, \ldots, \beta_{aq_a}, \beta_{\mu 0}, \beta_{\mu 1}, \ldots, \beta_{\mu q_\mu})$.

It should be noted that the simplified APGP model in (3.4) with μ_t and a_t as given by (4.1) and (4.2) is equivalent to the standard Poisson regression model with the canonical link function given by:

$$\begin{aligned}\lambda_t = \hat{w}_t &= \exp[\beta_{\mu 0} + \beta_{\mu 1} z_{\mu 1 t} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu t} \\ &\quad -(t-1)(\beta_{a0} + \beta_{a1} z_{a1t} + \cdots + \beta_{aq_a} z_{aq_a t})] \end{aligned} \tag{4.3}$$

where λ_t is the mean of the Poisson distribution at time t, $\beta_{\mu 0}, \ldots, \beta_{\mu q_\mu}$, $\beta_{a0}, \ldots,$ β_{aq_a} are called the regression parameters and $z_{\mu 1t}, \ldots, z_{\mu q_\mu t}, \ldots, (t-1)z_{a1t}, \ldots,$ $(t-1)z_{aq_a t}$ are called the regressors or the explanatory variables.

Although, theoretically, the simplified model is equivalent to the standard Poisson regression model, the interpretation of model parameters are quite different. In the APGP model, the mean and the ratio functions can be interpreted separately. Moreover, the ratio function of the APGP model shows the trend of the progression more clearly than the standard Poisson regression model.

In the Poisson regression model, the regressors have power-multiplicative effect on the mean as given by (4.3). For example, the effect of one unit increase in $z_{\mu 1t}$ results in a multiplication of $\hat{w}_t$ by $\exp(\beta_{\mu 1})$ provided that other regressors are kept constant. However, it seems to be meaningless when we make interpretation on the regressor $(t-1)z_{akt},\ k = 1, \ldots, q_a$.

However, in the APGP model, the mean function which reveals the level of the trend and the ratio function which reveals the direction and progression of the trend are interpreted separately. Hence, the effect of one unit increase in z_{akt} results in a multiplication of the ratio function by $\exp[\beta_{ak}(t-1)]$. In this way,

we can study the progression in the ratio more explicitly.

Incorporating some time-evolving effects into the ratio function a_t means that the APGP model allows a gradual change in the direction and strength of the trend. For example, the gradual change of a_t from $a_t < 1$ to $a_t > 1$ indicates a gradual transition from a growing stage to a declining stage. The resulting pmf for the observed data W_t for the original model and simplified model are given by (3.2) and (3.4) respectively, where μ_t and a_t are given by (4.1) and (4.2) respectively. Also, the mean and variance for W_t are given by (3.3) and (3.5) respectively.

4.1.2 Types of trend pattern

Since the mean $E_h(W_t)$ of the observed data W_t depends on the ratio function a_t which could be varied, thus the APGP model can describe a variety of trend patterns with different stages of development. In addition, different forms of ratio function a_t can be obtained by incorporating different time functions. Some common time functions include $\ln t$, $1/t$, and $\sqrt{t}$. In the following, we adopt two ratio functions to demonstrate a wide variety of trend patterns.

Apart from the two cases of increasing and decreasing trends for the basic PGP model in Chapter 3, in the next six cases, for the APGP model, we use the ratio function $a_t = \exp(\beta_{a0} + \beta_{a1} \ln t)$ and mean function $\mu_t = \exp(\beta_{\mu 0})$. Then we let the two parameters β_{a0} or β_{a1} to vary. Thereafter, we adopt another ratio function $a_t = \exp(\beta_{a0} + \beta_{a1}/t)$ with the same mean function to further identify

other six possible patterns. Here are the summary for all these cases:

(2) APGP models with ratio function $a_t = \exp(\beta_{a0} + \beta_{a1} \ln t)$,

Case 3. Increasing $a_t < 1$: $E(W_t)$ increases at an increasing rate and then a decreasing rate until it reaches a maximum. Afterwards, it decreases at an increasing rate and then a decreasing rate until it drops back to a very low level. A bell-shaped pattern is obtained (Figure 2a).

Case 4. Increasing $a_t > 1$: $E(W_t)$ decreases at an increasing rate for a very short period and then at a decreasing rate (Figure 2b).

Case 5. Decreasing $a_t < 1$: $E(W_t)$ increases at an increasing rate (Figure 2c).

Case 6. Decreasing $a_t > 1$: $E(W_t)$ decreases at a decreasing rate (Figure 2d).

Case 7. Increasing a_t from $a_t < 1$ to $a_t > 1$: $E(W_t)$ increases from a certain level at a decreasing rate until it reaches a maximum. Then, it decreases at an increasing rate and then a decreasing rate until it drops back to a very low level. A partial bell-shaped pattern is obtained (Figure 2e).

Case 8. Decreasing a_t from $a_t > 1$ to $a_t < 1$: $E(W_t)$ decreases at a decreasing rate to a certain level and then increases at an increasing rate. A U-shaped pattern is observed (Figure 2f).

(3) APGP models with ratio function $a_t = \exp(\beta_{a0} + \beta_{a1}/t)$,

Case 9. Increasing $a_t < 1$: $E(W_t)$ increases at a decreasing rate and then approaches to a constant maximum level (Figure 3a).

Case 10. Increasing $a_t > 1$: $E(W_t)$ decreases at a decreasing rate, which is the same as in Case 6 (Figure 3b).

Case 11. Decreasing $a_t < 1$: $E(W_t)$ increases very slowly at the beginning and then at an increasing rate. (Figure 3c).

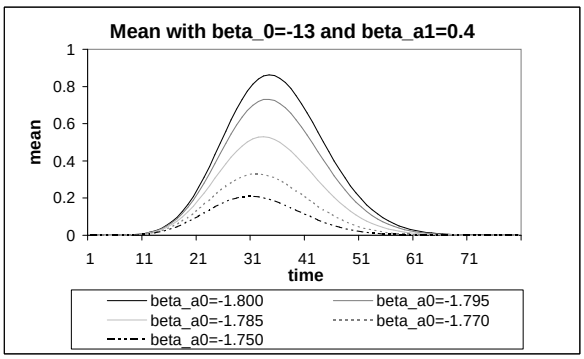
Case 12. Decreasing $a_t > 1$: $E(W_t)$ decreases rapidly and then approaches a constant minimum level (Figure 3d).

Case 13. Increasing a_t from $a_t < 1$ to $a_t > 1$: $E(W_t)$ increases at the beginning and then decreases at a decreasing rate (Figure 3e).

Case 14. Decreasing a_t from $a_t > 1$ to $a_t < 1$: $E(W_t)$ decreases rapidly at the beginning and then increases at an increasing rate. A tick pattern is observed (Figure 3f).

The extended APGP model in the ratio function can model different trend patterns and is widely applicable for Poisson count data in different fields of study.

Figure 2a $E(W_t)$ and a_t in PGP model for various β_{a0} with $a_t < 1$ & increasing ($\beta_0 = -13, \beta_{a1} = 0.4$)



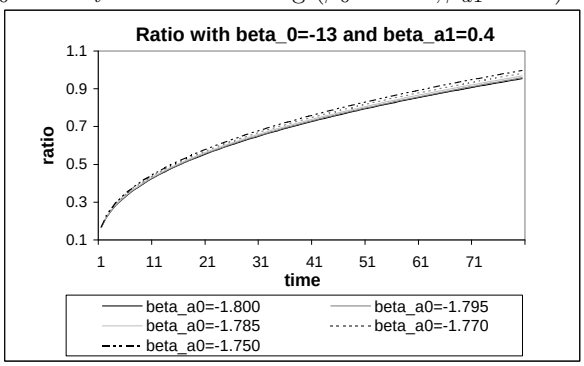
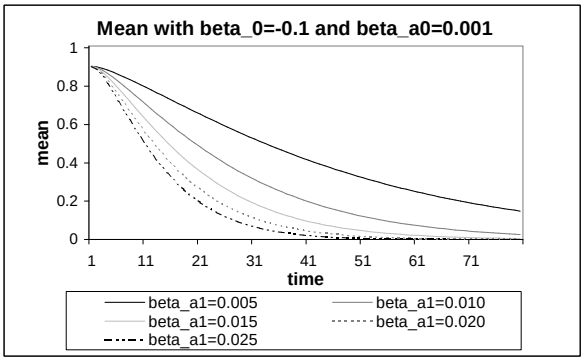


Figure 2b $E(W_t)$ and a_t in PGP model for various β_{a1} with $a_t > 1$ & increasing ($\beta_0 = -0.1, \beta_{a0} = 0.001$)



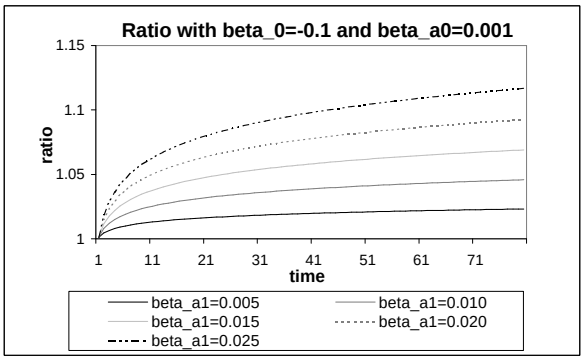
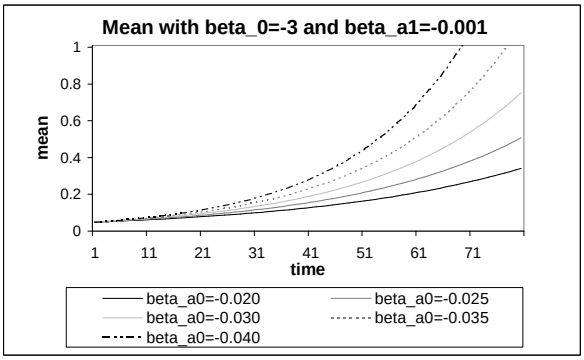


Figure 2c $E(W_t)$ and a_t in PGP model for various β_{a0} with $a_t < 1$ & decreasing ($\beta_0 = -3, \beta_{a1} = -0.001$)



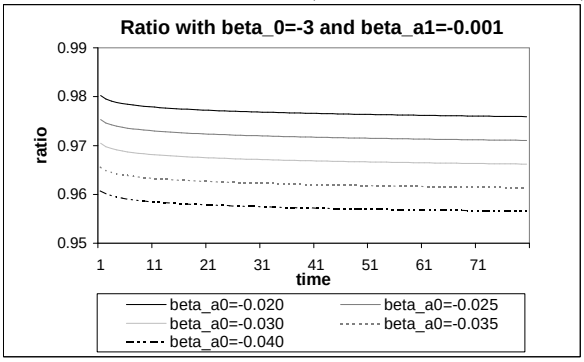
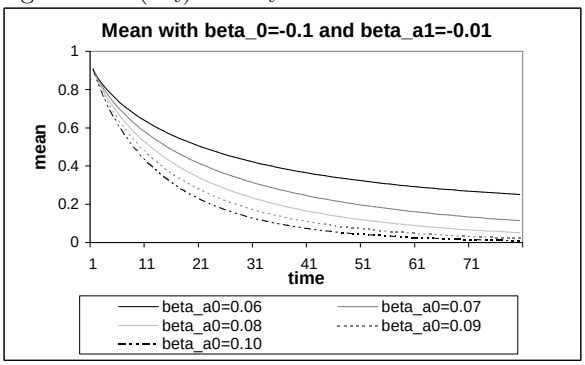


Figure 2d $E(W_t)$ and a_t in PGP model for various β_{a0} with $a_t > 1$ & decreasing ($\beta_0 = -0.1, \beta_{a1} = -0.01$)



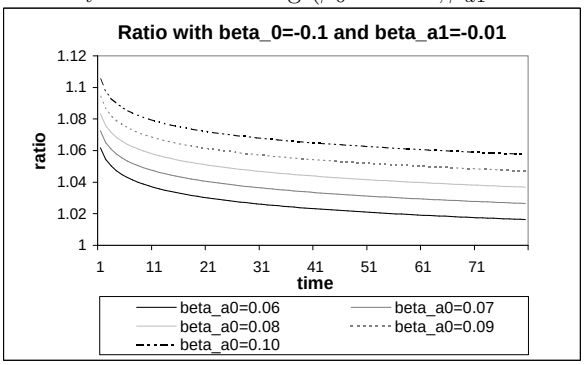
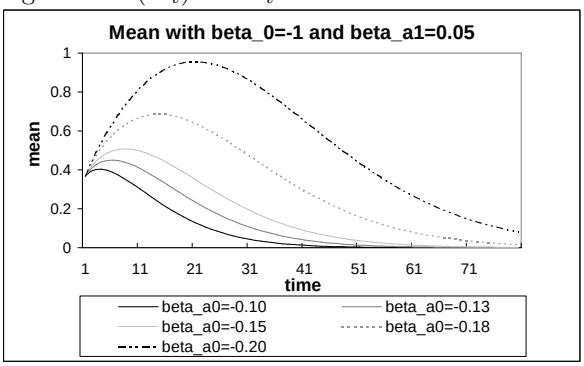


Figure 2e $E(W_t)$ and a_t in PGP model for various β_{a0} with a_t increasing from < 1 to > 1 ($\beta_0 = -1, \beta_{a1} = 0.05$)



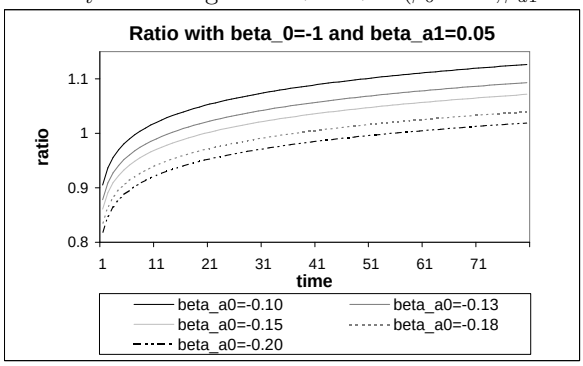
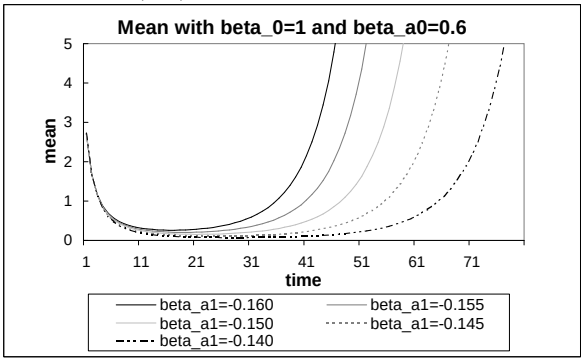


Figure 2f $E(W_t)$ and a_t in PGP model for various β_{a1} with a_t decreasing from > 1 to < 1 ($\beta_0 = 1, \beta_{a0} = 0.6$)



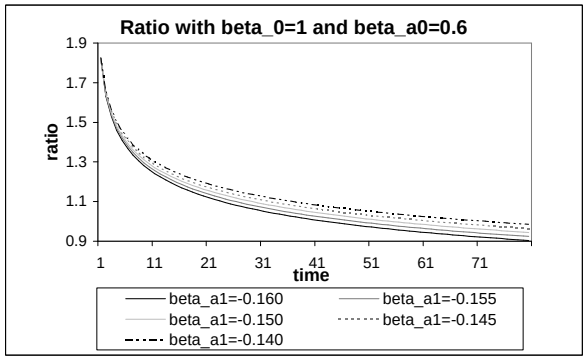
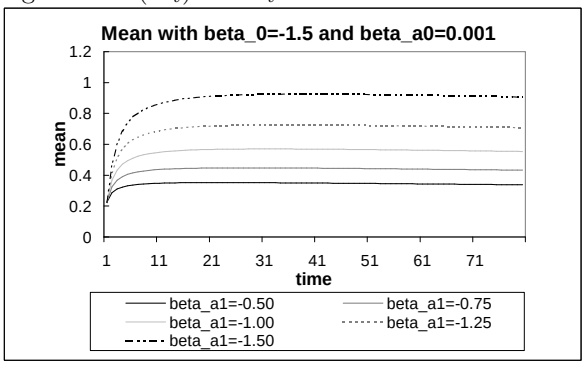


Figure 3a $E(W_t)$ and a_t in PGP model for various β_{a1} with $a_t < 1$ & increasing ($\beta_0 = -1.5, \beta_{a0} = 0.001$)



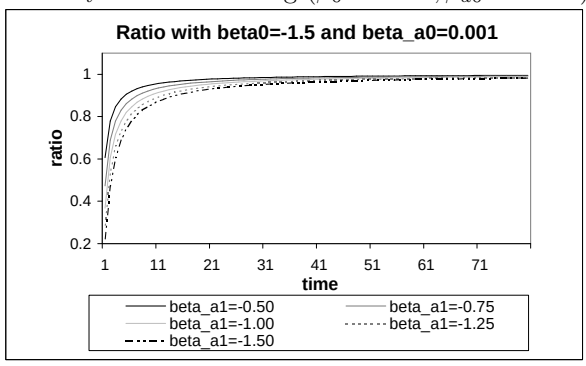
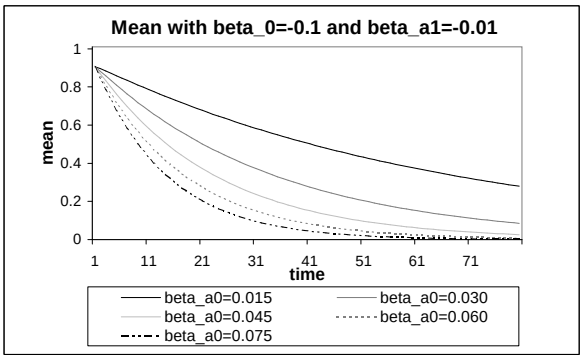


Figure 3b $E(W_t)$ and a_t in PGP model for various β_{a0} with $a_t > 1$ & increasing ($\beta_0 = -0.1, \beta_{a1} = -0.01$)



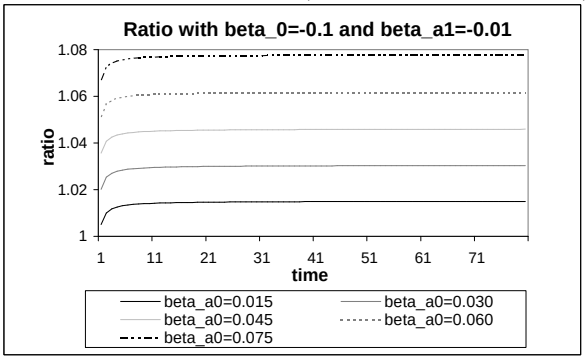
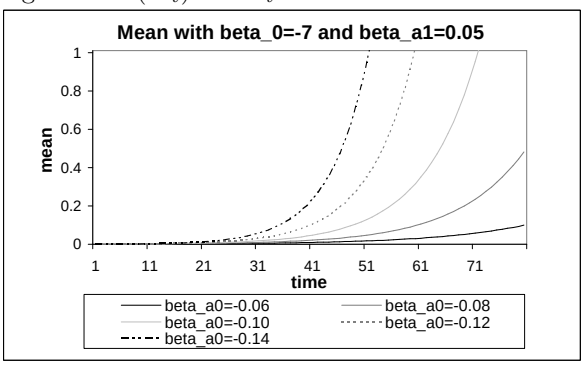


Figure 3c $E(W_t)$ and a_t in PGP model for various β_{a0} with $a_t < 1$ & decreasing ($\beta_0 = -7, \beta_{a1} = 0.05$)



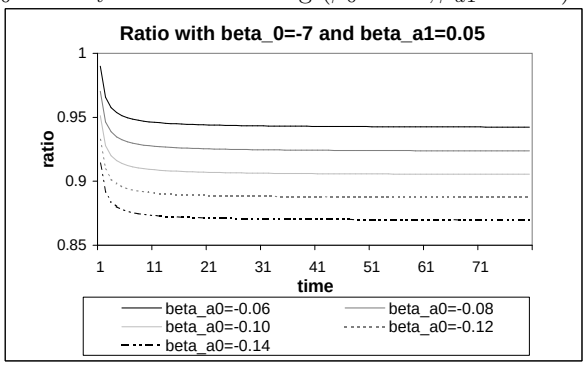
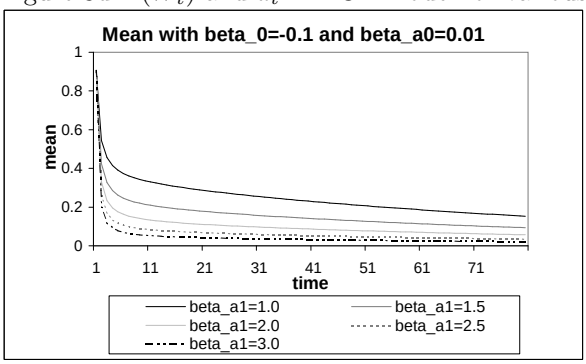


Figure 3d $E(W_t)$ and a_t in PGP model for various β_{a1} with $a_t > 1$ & decreasing ($\beta_0 = -0.1, \beta_{a0} = 0.01$)



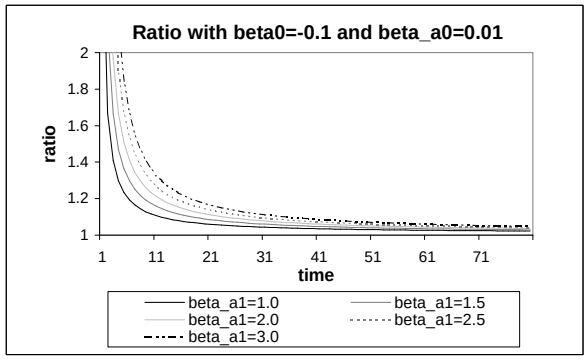
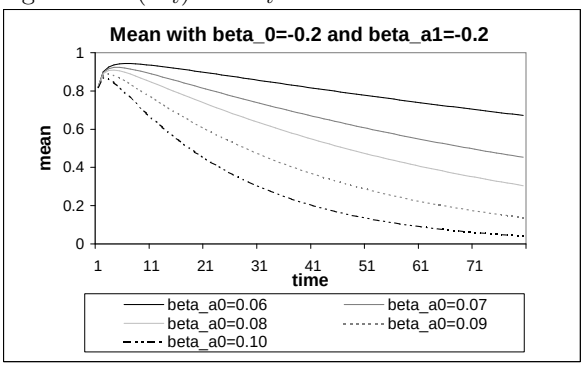


Figure 3e $E(W_t)$ and a_t in PGP model for various β_{a0} with a_t increasing from < 1 to > 1 ($\beta_0 = \beta_{a1} = -0.2$)



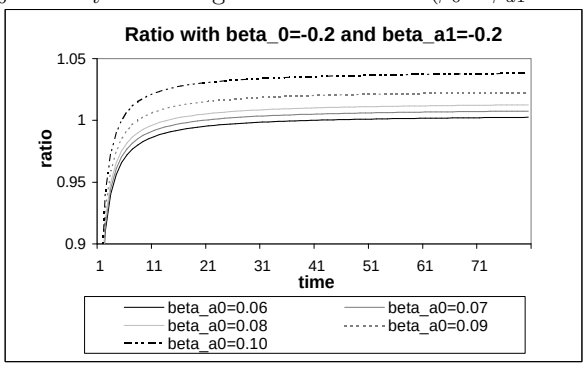
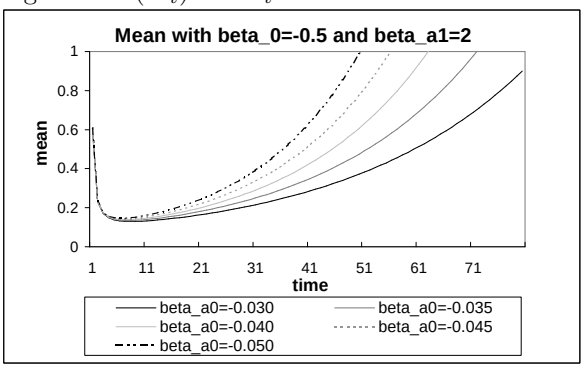
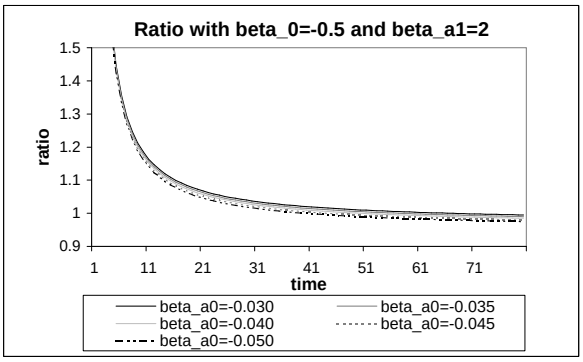


Figure 3f $E(W_t)$ and a_t in PGP model for various β_{a0} with a_t decreasing from > 1 to $< 1 (\beta_0 = -0.5, \beta_{a1} = 2)$





4.2 Methodology of inference

4.2.1 Least-square error method

Again, the SSE function for the APGP model are given by (3.6) where μ_t and a_t are given by (4.1) and (4.2) respectively. By differentiating the SSE function once and setting it to zero, we obtain the score equations. The second order derivative functions as required in the Newton Raphson (NR) iterative method are given in Appendix 4.1.

4.2.2 Maximum likelihood method

To implement the ML method, we first derive the likelihood functions $L_h(\boldsymbol{\theta}_h)$ and log-likelihood functions $\ell_h(\boldsymbol{\theta}_h)$ where $h = o, s$ for the APGP models. For the original APGP model ($h = o$),

$$L_o(\boldsymbol{\theta}_o) = \prod_{t=1}^{n} \frac{\exp(-\eta_{\mu t})\exp(\eta_{at})^{t-1}}{[\exp(-\eta_{\mu t})\exp(\eta_{at})^{t-1}+1]^{w_t+1}} \tag{4.4}$$

where $\mu_t = \exp(\eta_{\mu t})$, $a_t = \exp(\eta_{at})$ and $\eta_{\mu t}$ and η_{at} are given by (4.1) and (4.2) respectively. The log-likelihood function is

$$\begin{aligned}\ell_o(\boldsymbol{\theta}_o) = & \sum_{t=1}^{n}[(t-1)(\beta_{a0}z_{a0t}+\beta_{a1}z_{a1t}+\cdots+\beta_{aq_a}z_{aq_at}) \\ & -(\beta_{\mu 0}z_{\mu 0t}+\beta_{\mu 1}z_{\mu 1t}+\cdots+\beta_{\mu q_\mu}z_{\mu q_\mu t})] - \\ & \sum_{t=1}^{n}(w_t+1)\ln\left[\frac{\exp(\beta_{a0}z_{a0t}+\beta_{a1}z_{a1t}+\cdots+\beta_{aq_a}z_{aq_at})^{t-1}}{\exp(\beta_{\mu 0}z_{\mu 0t}+\beta_{\mu 1}z_{\mu 1t}+\cdots+\beta_{\mu q_\mu}z_{\mu q_\mu t})}+1\right]\end{aligned} \tag{4.5}$$

The first and second order derivative functions as required in the NR iterative method are given in Appendix 4.2.

For the simplified APGP model ($h = s$), the likelihood function $L_s(\boldsymbol{\theta}_s)$ and the log-likelihood function $\ell_s(\boldsymbol{\theta}_s)$ are given by

$$L_s(\boldsymbol{\theta}_s) = \prod_{t=1}^{n} \frac{\exp\left[-\dfrac{\exp(\eta_{\mu t})}{\exp(\eta_{at})^{t-1}}\right]\left[\dfrac{\exp(\eta_{\mu t})}{\exp(\eta_{at})^{t-1}}\right]^{w_t}}{w_t!} \tag{4.6}$$

$$\begin{aligned}\ell_s(\boldsymbol{\theta}_s) = & \sum_{t=1}^{n}\left[-\frac{\exp(\beta_{\mu 0}z_{\mu 0t}+\cdots+\beta_{\mu q_\mu}z_{\mu q_\mu t})}{\exp(\beta_{a0}z_{a0t}+\cdots+\beta_{aq_a}z_{aq_at})^{t-1}}+w_t(\beta_{\mu 0}z_{\mu 0t}+\cdots+\beta_{\mu q_\mu}z_{\mu q_\mu t})\right] \\ & -\sum_{t=1}^{n}[w_t(t-1)(\beta_{a0}z_{a0t}+\cdots+\beta_{aq_a}z_{aq_at})-\ln(w_t!)]\end{aligned} \tag{4.7}$$

where $\mu_t = \exp(\eta_{\mu t})$, $a_t = \exp(\eta_{at})$ and $\eta_{\mu t}$ and η_{at} are given by (4.1) and (4.2) respectively. Then, the first order and second order derivative functions required for the NR iterative method are given in Appendix 4.2. Again, the NR iterative method can be implemented by writing Fortran program with IMSL.

Moreover, we can derive the limiting distributions in (2.7) for the ML estimators in both original and simplified models by taking $w_t = E(W_t) = \hat{w}_t$ in the second order derivative functions. Hence, elements in the u^{th} row and the v^{th} column of the covariance matrix $\mathbf{\Sigma}$ in (2.7) are

$$s^{-1}_{o,uv} = -E_o\left[\frac{\partial \ell^2_o}{\partial\theta_u\partial\theta_v}\right] = \sum_{t=1}^{n} z^*_{ut} z^*_{vt} \frac{\hat{w}_t}{\hat{w}_t + 1} \tag{4.8}$$

$$\text{and} \quad s^{-1}_{s,uv} = -E_s\left[\frac{\partial \ell^2_s}{\partial\theta_u\partial\theta_v}\right] = \sum_{t=1}^{n} z^*_{ut} z^*_{vt} \hat{w}_t \tag{4.9}$$

for the original and simplified versions of the models. For the PGP model, the parameters θ_u and θ_v represent the model parameters a and β_0 respectively while z^*_{ut} and z^*_{vt} represent z^*_{a0t} and $z^*_{\mu 0t}$ respectively. For the APGP model, θ_u and θ_v may represent any of the model parameters $\beta_{jk}, j = \mu, a, k = 0, \ldots, q_j$. Accordingly, the variables z^*_{ut} and z^*_{vt} may represent any of the z^*_{akt} and $z^*_{\mu kt}$.

4.2.3 Bayesian method

The Bayesian hierarchy of the original and simplified APGP models is similar to that of the PGP model in Section 3.4.3. The Bayesian hierarchies for the APGP

models are:

$$\begin{aligned}
\text{Likelihood for the original model: } w_t &\sim f_{ot}(w_t) \\
f_{ot}(w_t) &= \frac{a_t^{t-1}/\mu_t}{(a_t^{t-1}/\mu_t + 1)^{w_t+1}} \\
\mu_t &= \exp(\beta_{\mu 0} z_{\mu 0 t} + \beta_{\mu 1} z_{\mu 1 t} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu t}) \\
a_t &= \exp(\beta_{a0} z_{a0t} + \beta_{a1} z_{a1t} + \cdots + \beta_{aq_a} z_{aq_a t})
\end{aligned}$$

$$\begin{aligned}
\text{Likelihood for the simplified model: } w_t &\sim Poi(\lambda_t) \\
\lambda_t &= \frac{\mu_t}{a_t^{t-1}} \\
\mu_t &= \exp(\beta_{\mu 0} z_{\mu 0 t} + \beta_{\mu 1} z_{\mu 1 t} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu t}) \\
a_t &= \exp(\beta_{a0} z_{a0t} + \beta_{a1} z_{a1t} + \cdots + \beta_{aq_a} z_{aq_a t}).
\end{aligned}$$

Since the parameters β_{ak} and $\beta_{\mu k}$ have non-restricted continuous ranges, normal distribution is therefore assigned to these parameters:

$$\begin{aligned}
\text{Priors for parameters: } \beta_{ak} &\sim N(0, \sigma_{ak}^2), \quad k = 0, \ldots, q_a \\
\beta_{\mu k} &\sim N(0, \sigma_{bk}^2), \quad k = 0, \ldots, q_\mu.
\end{aligned}$$

Without prior information, σ_{ak}^2 and σ_{bk}^2 are set to be 1000. To estimate the model parameters, we draw every $H^{th} = 20^{th}$ sample beginning from iteration $B + 1 = 5001$ till iteration $R = 100000$ resulting in $M = 4750$ posterior samples of β_{ak} and $\beta_{\mu k}$. The sample means of these posterior samples give the Bayesian parameter estimates. It is also essential to investigate the history and ACF plots of each model parameter to ensure their independence and convergence.

4.3 Simulation studies

In the simulation studies, we set the number of realizations $N = 200$ and the sample size in each realization $n = 100$. Simulation methods for the original and simplified APGP model are similar to that of the PGP model. For the APGP models, we incorporate one covariate effect in the mean and ratio functions respectively, that is $\mu_t = \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu 1t})$ and $a_t = \exp(\beta_{a0} + \beta_{a1} t)$ where $q_j = 1, j = \mu, a$. Here we set the covariate $z_{\mu 1t}$ in the mean function to be cyclic with values 0,1,2,3 in each cycle and the covariate $z_{a1t} = t$ in the ratio function. The set of parameters chosen for the baseline model are $\beta_{\mu 0} = 0.00$, $\beta_{\mu 1} = 1.00$, $\beta_{a0} = 0.01$ and $\beta_{a1} = 0.00$. We vary the value of one parameter in the set at a time to detect any change in the three GOF measures: SE for the precision, and MSE_1 and PE for the accuracy. Parameter β_{a1} is fixed at 0 in most of the cases. The results are shown in Table 4.1 and 4.2.

==========

Tables 4.1 and 4.2 about here

==========

The simulation results in Table 4.1 and 4.2 show that the performance of the ML estimators for the APGP models are generally satisfactory. Comparing the results for the original and simplified models, it is observed that the ML estimators for the simplified APGP model perform better than those of the original model. In particular, when β_{a1} varies in the original model, $\hat{\beta}_{a0}$ deviates a lot from the true value and thus results in large PE. In other words, W_t for the APGP model

is more sensitive to changes in β_{a0} and β_{a1}. It is because the ratio function a_t is raised to the power $t-1$, hence, this sensitivity of W_t becomes more pronounced when the trend is long, like $n = 100$ in this simulation study. Besides, in the original APGP model, the ML estimators of $\beta_{\mu 1}$ and β_{a0} do not perform very well when $\beta_{\mu 1}$ varies. This indicates that the poor estimation of $\beta_{\mu 1}$ would result in a poor estimation of β_{a0}. For the simplified model, the effects of the change of one parameter from the set on the resultant parameter estimates are similar but of much less magnitude.

4.4 Real data analysis

To demonstrate the application of the APGP model, we use the Hong Kong SARS data in 2003 as an example for real data analysis. In the previous chapter, we have applied the basic PGP model to this data set and found that the results are not quite satisfactory as the MSE_2 for all models are quite large indicating that the basic PGP models do not fit the data well. Hence, we modify the model in the following ways. Firstly, we set the ratio constant and include the temperature $z_{\mu 1t}$ as a covariate effect in the mean function for the APGP1 model because the prevalence of most epidemics is highly related to the temperature, and the SARS viruses show a remarkable declining survival rate with an ascending temperature. The viruses survive for at least 21 days at 4 ^{o}C but only for less than 4 days at body temperature of 37 ^{o}C.

Secondly, we further include linear time effect ($z_{a1t} = t$) in the ratio function of the APGP2 model and quadratic time effect ($z_{a2t} = t^2$) in the ratio function of the APGP3 model as we observe that an increasing trend followed by a decreasing trend of the daily number of infected cases indicating a possible change in the ratio from $a < 1$ to $a > 1$ during the SARS outbreak period. A total number of six original and simplified APGP(1,2,3) models are implemented using the LSE, the ML and the Bayesian methods.

For model assessment and selection, again we use the mean squares of error (MSE_2) and the posterior expected utility (U) with the likelihood functions $L_h(w_t|\hat{\boldsymbol{\theta}}_h)$ given by (4.4) and (4.6) respectively as the GOF measures for the two versions of the APGP models. In addition, we adopt the Akaike Information Criterion (AIC) proposed by Akaike (1) as the third GOF measure.

$$AIC = -2 \ln L_h(w_t|\hat{\boldsymbol{\theta}}_h) + 2p \tag{4.10}$$

where L_h are the likelihood functions (4.4) and (4.6) based on the parameter estimate $\hat{\boldsymbol{\theta}}_h$ and $p = q_a + q_\mu$ is the number of parameters in the model. The AIC examines the model complexity and the model goodness-of-fit. Model with many parameters will be penalized despite it provides a better fit to the data. Hence, the model with smaller AIC is more preferred.

4.4.1 Results and comments

The summary of the six APGP models each estimated by three methods are given in Table 4.3. Comparing between the original and simplified versions of

each APGP models, we find that the simplified version is worse than the original version for the APGP1 model according to U. On the other hand, the simplified model is always better than the original model according to MSE_2 as remarked in Section 3.7. However, with increasing complexity of the APGP model, the simplified version becomes better than the original version for the APGP3 model according to U. This can be explained by the fact that as the APGP model advances, the predicted daily number of infected cases $E_s(W_t)$ gets closer to the observed W_t resulting in larger and hence better U.

Comparing between each APGP model using the three methods of inference, the LSE method gives better MSE_2 whereas the ML method gives better U in general. Since the MSE_2 and U for the APGP1 models with temperature as a covariate effect in the mean function are not much improved while comparing with those of the PGP model, the model is further improved by including a linear time t as a covariate effect in the ratio function. The MSE_2 of the APGP2 models decreases dramatically and the U increases using the ML and the Bayesian method. In particular, for the simplified model, the improvement is substantial in terms of the three GOF measures. The substantial improvement is due to the incorporation of time effect in the ratio function which helps to accommodate the increasing ($a_t < 1$) and then decreasing ($a_t > 1$) trends in the SARS data as shown in Figure 4.1. Hence, the APGP2 models are more preferred than the APGP1 models.

==========

Figure 4.1 about here

==========

To further improve the modeling of the trends in the data, we consider a more general time effect, the quadratic time effect, by adding one more covariate t^2 in the ratio function. Results in Table 4.3 show a further improvement in the MSE_2 and the U for both original and simplified versions of the APGP3 model using the LSE, the ML and the Bayesian methods. As stated above, the simplified APGP3 model performs better than the original model in terms of the three GOF measures. Moreover, after penalizing for the additional parameter β_{a2} in the APGP3 model, the AIC of the APGP3 models are still generally smaller than those of the APGP2 models. Therefore, according to the three GOF measures, the simplified APGP3 model using the ML method is chosen as the best model as it has the best U and AIC, and the second best MSE_2. Figure 4.1 shows the simplified APGP3 models using different methods of inference.

For all APGP models, we incorporate the temperature as a covariate effect in the mean function. As there are improvements in the MSE_2 as compared with those for the PGP model, it implies that the temperature affects the number of daily infected SARS cases. However, the estimate $\hat{\beta}_{\mu 1}$ is positive which implies that the higher temperature is associated with higher number of daily infected cases. This association is contradictive to the fact that the SARS viruses survive for shorter period with increasing temperature. Therefore, we test for the

significance of $\hat{\beta}_{\mu 1} = 0$ in the mean function for the best model. Adopting the limiting distribution of (2.7), the asymptotic standard error $ASE_h(\hat{\beta}_{\mu 1,ML})$ using (4.8) and (4.9) are found to be 4.6328×10^{-3} and 4.6350×10^{-3} respectively. The test statistics are

$$Z_o = \frac{\hat{\beta}_{\mu 1,ML,o}}{ASE_o(\hat{\beta}_{\mu 1,ML})} = 10.3099$$

$$Z_s = \frac{\hat{\beta}_{\mu 1,ML,s}}{ASE_s(\hat{\beta}_{\mu 1,ML})} = 7.2476$$

which are highly significant. The contradiction may due to the fact that the SARS outbreak occurred in spring and early summer. During this period, the viruses spread rapidly in the initial stage and at the same time the temperature increases too. Nevertheless, the incorporation of the covariate in the mean function improves the model fit as the APGP models give better MSE_2 and U than those of the PGP models using different methods of inference.

==========

Table 4.3 about here

==========

From the real data analysis, the extension to the APGP model to accommodate covariate effects in the data greatly widen the scope of application for the PGP model since the resultant APGP model is able to describe a wide variety of trend patterns. However, the APGP model may not be suitable to model trend data which fluctuates with abrupt turns because the ratio function a_t only allows a gradual change of the ratio. In the next chapter, the PGP and APGP models will be extended to the threshold PGP (TPGP) model which is used to fit count

data with sudden changes in the trends.

Appendix 4.1 Derivation of first and second order derivative functions in the NR iterative method for the APGP models using LSE method

By differentiating the SSE function in (3.6), the first order derivative functions are:

$$\begin{aligned}\frac{\partial SSE}{\partial \beta_{ak}} &= 2\sum_{t=1}^{n} z^*_{akt}\hat{w}_t(w_t - \hat{w}_t)\\ \frac{\partial SSE}{\partial \beta_{\mu k}} &= -2\sum_{t=1}^{n} z^*_{\mu kt}\hat{w}_t(w_t - \hat{w}_t)\end{aligned}$$

where $\hat{w}_t = \dfrac{\mu_t}{a_t^{t-1}}$, $z^*_{akt} = (t-1)z_{akt}$ for $k = 0, \ldots, q_a$ and $z^*_{\mu kt} = z_{\mu kt}$ for $k = 0, \ldots, q_\mu$.

The second order derivative functions are:

$$\begin{aligned}\frac{\partial^2 SSE}{\partial {\beta_{ak}}^2} &= -2\sum_{t=1}^{n} {z^*_{akt}}^2\hat{w}_t(w_t - 2\hat{w}_t)\\ \frac{\partial^2 SSE}{\partial {\beta_{\mu k}}^2} &= -2\sum_{t=1}^{n} {z^*_{\mu kt}}^2\hat{w}_t(w_t - 2\hat{w}_t)\\ \frac{\partial^2 SSE}{\partial \beta_{ak_1}\partial \beta_{ak_2}} &= -2\sum_{t=1}^{n} z^*_{ak_1t}z^*_{ak_2t}\hat{w}_t(w_t - 2\hat{w}_t)\\ \frac{\partial^2 SSE}{\partial \beta_{ak_1}\partial \beta_{\mu k_2}} &= 2\sum_{t=1}^{n} z^*_{ak_1t}z_{\mu k_2t}\hat{w}_t(w_t - 2\hat{w}_t)\\ \frac{\partial^2 SSE}{\partial \beta_{\mu k_1}\partial \beta_{\mu k_2}} &= -2\sum_{t=1}^{n} z^*_{\mu k_1t}z^*_{\mu k_2t}\hat{w}_t(w_t - 2\hat{w}_t)\end{aligned}$$

Appendix 4.2 Derivation of first and second order derivative functions in the NR iterative method for the APGP models using ML method

For the original APGP model, on differentiation of the log-likelihood function (4.5), the first order derivative functions for $\boldsymbol{\theta}_o$ are:

$$\begin{aligned}
\frac{\partial \ell_o}{\partial \beta_{ak}} &= \sum_{t=1}^{n} \frac{z^*_{akt}(\hat{w}_t - w_t)}{\hat{w}_t + 1} \\
\frac{\partial \ell_o}{\partial \beta_{\mu k}} &= -\sum_{t=1}^{n} \frac{z^*_{\mu kt}(\hat{w}_t - w_t)}{\hat{w}_t + 1}
\end{aligned}$$

The second derivative functions as required for the NR iterative method are:

$$\begin{aligned}
\frac{\partial^2 \ell_o}{\partial \beta_{ak}^2} &= -\sum_{t=1}^{n} \frac{{z^*_{akt}}^2 \hat{w}_t(w_t + 1)}{(\hat{w}_t + 1)^2} \\
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k}^2} &= -\sum_{t=1}^{n} \frac{{z^*_{\mu kt}}^2 \hat{w}_t(w_t + 1)}{(\hat{w}_t + 1)^2} \\
\frac{\partial^2 \ell_o}{\partial \beta_{ak_1} \partial \beta_{ak_2}} &= -\sum_{t=1}^{n} \frac{z^*_{ak_1t} z^*_{ak_2t} \hat{w}_t(w_t + 1)}{(\hat{w}_t + 1)^2} \\
\frac{\partial^2 \ell_o}{\partial \beta_{ak_1} \partial \beta_{\mu k_2}} &= \sum_{t=1}^{n} \frac{z^*_{ak_1t} z^*_{\mu k_2t} \hat{w}_t(w_t + 1)}{(\hat{w}_t + 1)^2} \\
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k_1} \partial \beta_{\mu k_2}} &= -\sum_{t=1}^{n} \frac{z^*_{\mu k_1t} z^*_{\mu k_2t} \hat{w}_t(w_t + 1)}{(\hat{w}_t + 1)^2}
\end{aligned}$$

For the simplified APGP model, by differentiating (4.7), the first and second order derivative functions for $\boldsymbol{\theta}_s$ are:

$$\begin{aligned}
\frac{\partial \ell_s}{\partial \beta_{ak}} &= \sum_{t=1}^{n} z^*_{akt}(\hat{w}_t - w_t) \\
\frac{\partial \ell_s}{\partial \beta_{\mu k}} &= -\sum_{t=1}^{n} z^*_{\mu kt}(\hat{w}_t - w_t) \\
\frac{\partial^2 \ell_s}{\partial \beta_{ak}^2} &= -\sum_{t=1}^{n} {z^*_{akt}}^2 \hat{w}_t \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k}^2} &= -\sum_{t=1}^{n} {z^*_{\mu kt}}^2 \hat{w}_t \\
\frac{\partial^2 \ell_s}{\partial \beta_{ak_1} \partial \beta_{ak_2}} &= -\sum_{t=1}^{n} z^*_{ak_1 t} z^*_{ak_2 t} \hat{w}_t \\
\frac{\partial^2 \ell_s}{\partial \beta_{ak_1} \partial \beta_{\mu k_2}} &= \sum_{t=1}^{n} z^*_{ak_1 t} z^*_{\mu k_2 t} \hat{w}_t \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k_1} \partial \beta_{\mu k_2}} &= -\sum_{t=1}^{n} z^*_{\mu k_1 t} z^*_{\mu k_2 t} \hat{w}_t
\end{aligned}$$

Table 4.1: Simulation studies for the original APGP model

Parameter		$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}	β_{a1}	Parameter		$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}	β_{a1}
True value		0.00000	1.00000	0.01000	0.00000	True value		0.00000	1.00000	0.01000	0.00000
$\beta_{\mu 0}$	-1.000	-1.06532	0.98753	0.00814	0.00002	β_{a0}	-0.020	-0.04843	0.99253	-0.02148	0.00002
SE		*0.49276*	*0.14641*	*0.01989*	*0.00019*	SE		*0.37702*	*0.09904*	*0.01548*	*0.00015*
MSE_1		*0.25626*	*0.02109*	*0.00040*	*0.00000*	MSE_1		*0.14460*	*0.00996*	*0.00025*	*0.00000*
PE		*-0.06532*	*-0.01247*	†*-0.18637*	*0.00002*	PE		*-0.04843*	*-0.00747*	*-0.07417*	*0.00002*
$\beta_{\mu 0}$	0.000	-0.04675	0.99058	0.00875	0.00001	β_{a0}	-0.010	-0.04949	0.99190	-0.01155	0.00002
SE		*0.40300*	*0.11770*	*0.01693*	*0.00016*	SE		*0.38302*	*0.10279*	*0.01577*	*0.00015*
MSE_1		*0.16759*	*0.01446*	*0.00030*	*0.00000*	MSE_1		*0.15179*	*0.01077*	*0.00026*	*0.00000*
PE		*-0.04675*	*-0.00942*	‡*0.12515*	*0.00001*	PE		*-0.04949*	*-0.00810*	†*-0.15542*	*0.00002*
$\beta_{\mu 0}$	1.000	0.96507	0.98972	0.00889	0.00001	β_{a0}	0.020	-0.04648	0.98652	0.01856	0.00001
SE		*0.35781*	*0.10260*	*0.01535*	*0.00015*	SE		*0.41871*	*0.12943*	*0.01797*	*0.00018*
MSE_1		*0.12919*	*0.01086*	*0.00024*	*0.00000*	MSE_1		*0.19011*	*0.01690*	*0.00031*	*0.00000*
PE		*-0.03493*	†*-1.02750*	‡*0.11089*	*0.00001*	PE		*-0.04648*	*-0.01348*	*-0.07193*	*0.00001*
$\beta_{\mu 1}$	-0.100	-0.05902	-0.12859	0.00732	0.00004	β_{a1}	-0.0010	-0.13026	0.77855	-0.03290	-0.00040
SE		*0.51518*	*0.15802*	*0.02392*	*0.00024*	SE		*0.35955*	*0.09841*	*0.01543*	*0.00015*
MSE_1		*0.29144*	*0.02690*	*0.00062*	*0.00000*	MSE_1		*0.14839*	*0.05670*	*0.00207*	*0.00000*
PE		*-0.05902*	†*-0.28592*	†*-0.26809*	*0.00004*	PE		‡*-0.13026*	†*-0.22145*	‡*4.29035*	†*0.59908*
$\beta_{\mu 1}$	-0.050	-0.06452	-0.07708	0.00699	0.00004	β_{a1}	-0.0005	-0.05330	0.98466	0.00719	-0.00047
SE		*0.50615*	*0.15380*	*0.02334*	*0.00024*	SE		*0.38271*	*0.09971*	*0.01559*	*0.00015*
MSE_1		*0.28000*	*0.02657*	*0.00058*	*0.00000*	MSE_1		*0.14904*	*0.01006*	*0.00025*	*0.00000*
PE		*-0.06452*	†*-0.54158*	†*-0.30069*	*0.00004*	PE		*-0.05330*	*-0.01534*	†*-0.28141*	*0.06322*
$\beta_{\mu 1}$	-0.010	-0.06755	-0.03349	0.00686	0.00004	β_{a1}	0.0001	-0.04209	0.98462	0.00846	0.00012
SE		*0.49858*	*0.15045*	*0.02287*	*0.00023*	SE		*0.41225*	*0.12496*	*0.01765*	*0.00018*
MSE_1		*0.25454*	*0.02447*	*0.00053*	*0.00000*	MSE_1		*0.17921*	*0.01566*	*0.00032*	*0.00000*
PE		*-0.06755*	†*-2.34884*	†*-0.31363*	*0.00004*	PE		*-0.04209*	*-0.01538*	†*-0.15353*	†*0.15313*
$\beta_{\mu 1}$	0.050	-0.07741	0.03058	0.00649	0.00004	β_{a1}	0.0005	-0.08277	0.97732	0.00388	0.00059
SE		*0.48886*	*0.14614*	*0.02225*	*0.00004*	SE		*0.46604*	*0.15624*	*0.02445*	*0.00031*
MSE_1		*0.25339*	*0.02244*	*0.00052*	*0.00000*	MSE_1		*0.26654*	*0.02656*	*0.00067*	*0.00000*
PE		*-0.07741*	†*-0.38835*	†*-0.35102*	*0.00004*	PE		*-0.08277*	*-0.02268*	†*-0.61203*	†*0.18236*
$\beta_{\mu 1}$	0.100	-0.07298	0.08082	0.00694	0.00004	β_{a1}	0.0010	-0.14303	0.97733	-0.00381	0.00127
SE		*0.48100*	*0.14302*	*0.02176*	*0.00022*	SE		*0.53381*	*0.18277*	*0.03765*	*0.00063*
MSE_1		*0.24139*	*0.02165*	*0.00050*	*0.00000*	MSE_1		*0.33723*	*0.03290*	*0.00180*	*0.00000*
PE		*-0.07298*	†*-0.19182*	†*-0.30579*	*0.00004*	PE		‡*0.14303*	*-0.02267*	†*-1.38135*	†*0.26707*

$PE = \hat{\theta} - \theta$ when $\theta = 0$.

† Absolute value of PE is greater than 10%.

Table 4.2: Simulation studies for the simplified APGP model

Parameter		$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}	β_{a1}	Parameter		$\beta_{\mu 0}$	$\beta_{\mu 1}$	β_{a0}	β_{a1}
True value		0.00000	1.00000	0.01000	0.00000	True value		0.00000	1.00000	0.01000	0.00000
$\beta_{\mu 0}$	-1.000	-1.02242	1.00821	0.01040	-0.00000	β_{a0}	-0.020	-0.00315	1.00084	-0.02013	0.00000
SE		*0.30513*	*0.09513*	*0.01013*	*0.00010*	SE		*0.11127*	*0.02572*	*0.00345*	*0.00003*
MSE_1		*0.11480*	*0.01109*	*0.00010*	*0.00000*	MSE_1		*0.01367*	*0.00064*	*0.00001*	*0.00000*
PE		*-0.02242*	*0.00821*	*0.03989*	*-0.00000*	PE		*-0.00315*	*0.00084*	*-0.00636*	*0.00000*
$\beta_{\mu 0}$	0.000	-0.00680	0.99949	0.00944	0.00001	β_{a0}	-0.010	-0.00199	1.00105	-0.00997	-0.00000
SE		*0.18374*	*0.05722*	*0.00614*	*0.00006*	SE		*0.13234*	*0.03490*	*0.00418*	*0.00004*
MSE_1		*0.04020*	*0.00388*	*0.00004*	*0.00000*	MSE_1		*0.01698*	*0.00123*	*0.00002*	*0.00000*
PE		*-0.00680*	*-0.00051*	*-0.05648*	*0.00001*	PE		*-0.00199*	*0.00105*	*0.00387*	*-0.00000*
$\beta_{\mu 0}$	1.000	0.99300	1.00055	0.00981	0.00000	β_{a0}	0.020	-0.00207	0.99910	0.01985	0.00000
SE		*0.11141*	*0.03472*	*0.00372*	*0.00004*	SE		*0.21136*	*0.06901*	*0.00741*	*0.00008*
MSE_1		*0.01301*	*0.00117*	*0.00002*	*0.00000*	MSE_1		*0.04935*	*0.00494*	*0.00005*	*0.00000*
PE		*-0.00400*	*0.00055*	*-0.01877*	*0.00000*	PE		*-0.00207*	*-0.00090*	*-0.00761*	*0.00000*
$\beta_{\mu 1}$	-0.100	-0.00816	-0.09561	0.01158	-0.00010	β_{a1}	-0.0010	-0.00485	0.98987	0.00977	-0.00099
SE		*0.37623*	*0.12262*	*0.01808*	*0.00019*	SE		*0.06827*	*0.00219*	*0.00162*	*0.00001*
MSE_1		*0.13642*	*0.01509*	*0.00029*	*0.00000*	MSE_1		*0.00407*	*0.01000*	*0.00000*	*0.00000*
PE		*-0.00816*	*0.04387*	*† 0.15842*	*-0.00010*	PE		*-0.00485*	*-0.01013*	*-0.02273*	*0.01072*
$\beta_{\mu 1}$	-0.050	-0.01181	-0.04698	0.01132	-0.00001	β_{a1}	-0.0005	-0.00267	0.99812	0.00977	-0.00050
SE		*0.36720*	*0.11784*	*0.01747*	*0.00018*	SE		*0.10424*	*0.01691*	*0.00307*	*0.00002*
MSE_1		*0.13667*	*0.01368*	*0.00029*	*0.00000*	MSE_1		*0.00761*	*0.00023*	*0.00000*	*0.00000*
PE		*-0.01181*	*0.06034*	*†0.13236*	*-0.00001*	PE		*-0.00267*	*-0.00188*	*-0.02347*	*-0.00348*
$\beta_{\mu 1}$	-0.010	-0.00589	-0.00551	0.01147	-0.00001	β_{a1}	0.0001	-0.00976	1.00398	0.01021	0.00010
SE		*0.35887*	*0.11422*	*0.01696*	*0.00018*	SE		*0.19934*	*0.06404*	*0.00697*	*0.00008*
MSE_1		*0.12787*	*0.01274*	*0.00026*	*0.00000*	MSE_1		*0.04172*	*0.00415*	*0.00004*	*0.00000*
PE		*-0.00589*	*† 0.44948*	*†0.14663*	*-0.00001*	PE		*-0.00976*	*0.00398*	*0.02125*	*-0.02544*
$\beta_{\mu 1}$	0.050	-0.00805	0.05408	0.01140	-0.00001	β_{a1}	0.0005	-0.02496	1.00170	0.00788	0.00054
SE		*0.34768*	*0.10924*	*0.01622*	*0.00017*	SE		*0.24429*	*0.08141*	*0.01139*	*0.00017*
MSE_1		*0.11834*	*0.01236*	*0.00023*	*0.00000*	MSE_1		*0.06425*	*0.00569*	*0.00013*	*0.00000*
PE		*-0.00805*	*0.08150*	*†0.13996*	*-0.00001*	PE		*-0.02496*	*0.00172*	*† -0.21196*	*0.07746*
$\beta_{\mu 1}$	0.100	-0.01637	0.10737	0.01087	-0.00001	β_{a1}	0.0010	-0.05439	1.00524	0.00516	0.00111
SE		*0.33815*	*0.10507*	*0.01559*	*0.00016*	SE		*0.28169*	*0.09411*	*0.01795*	*0.00036*
MSE_1		*0.11522*	*0.01161*	*0.00021*	*0.00000*	MSE_1		*0.08799*	*0.00949*	*0.00035*	*0.00000*
PE		*-0.01637*	*0.07372*	*0.08714*	*-0.00001*	PE		*-0.05439*	*0.00524*	*† -0.48433*	*†0.10817*

$PE = \hat{\theta} - \theta$ when $\theta = 0$.

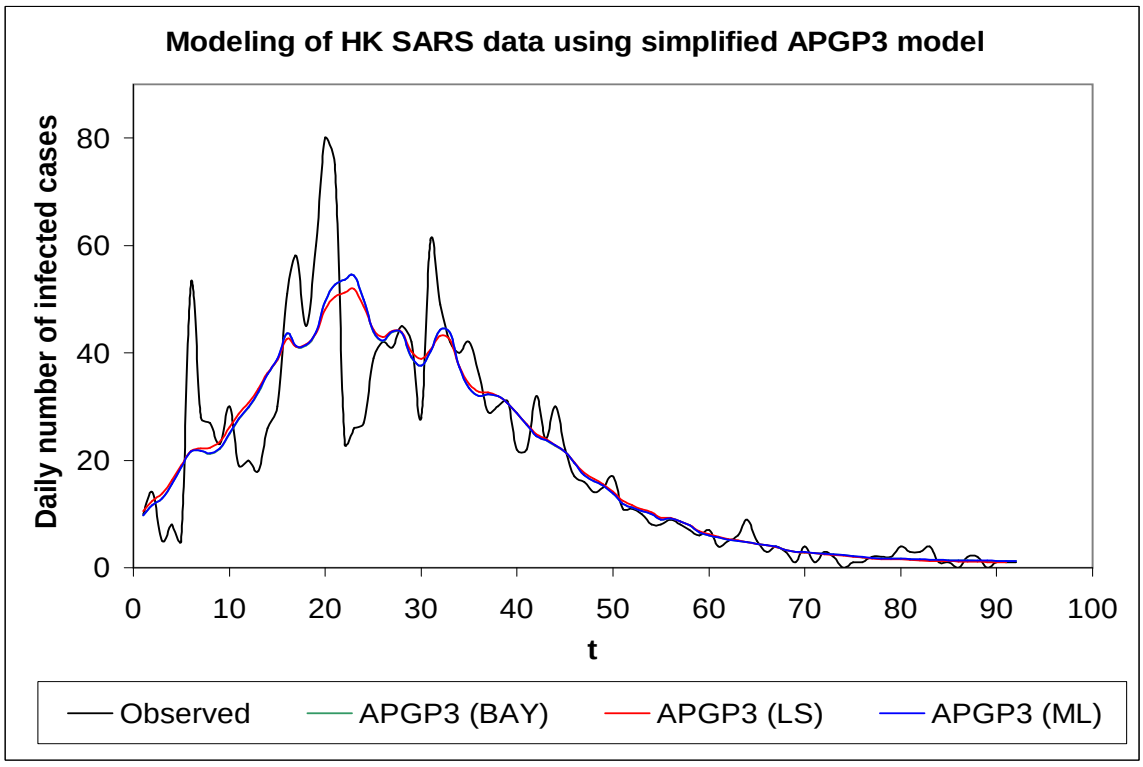
† Absolute value of PE is greater than 10%.

Table 4.3: Parameter estimates and standard errors for the Hong Kong SARS data using APGP models

Method		LSE			ML			Bay		
Model	Par.	Est.	SE	GOF	Est.	SE	GOF	Est.	SE	GOF
Original	$\beta_{\mu 0}$	2.1128	*0.2980*	213.976	2.7995	*1.1216*	514.672	2.4740	*1.0610*	524.986
APGP1	$\beta_{\mu 1}$	0.0824	*0.0149*		0.0895	*0.0560*	*-3.5649*	0.1065	*0.0533*	*-3.5654*
	a	1.0283	*0.0023*		1.0561	*0.0083*	661.944	1.0580	*0.0082*	662.040
Simplified	$\beta_{\mu 0}$	2.1128	*0.2980*	213.976	2.3691	*0.2145*	237.354	2.3710	*0.2132*	235.836
APGP1	$\beta_{\mu 1}$	0.0824	*0.0149*		0.0813	*0.0108*	*-6.3869*	0.0812	*0.0107*	*-6.3893*
	a	1.0283	*0.0023*		1.0375	*0.0017*	1181.18	1.0370	*0.0017*	1181.62
Original	$\beta_{\mu 0}$	2.1213	*0.2380*	80.5779	2.9566	*1.1118*	120.984	2.5990	*1.0350*	117.425
APGP2	$\beta_{\mu 1}$	0.0271	*0.0115*		0.0136	*0.0002*	*-3.4860*	0.0333	*0.0545*	*-3.4868*
	β_{a0}	-0.0963	*0.0106*		-0.0312	*0.0226*	649.433	-0.0282	*0.0227*	649.567
	β_{a1}	0.0020	*0.0002*		0.0008	*0.0002*		0.0008	*0.0002*	
Simplified	$\beta_{\mu 0}$	2.1213	*0.2380*	80.5779	2.3195	*0.2120*	89.3240	2.3160	*0.2106*	89.2882
APGP2	$\beta_{\mu 1}$	0.0271	*0.0115*		0.0299	*0.0109*	*-3.7294*	0.0300	*0.0108*	*-3.7294*
	β_{a0}	-0.0963	*0.0106*		-0.0635	*0.0053*	694.207	-0.0637	*0.0053*	694.210
	β_{a1}	0.0020	*0.0002*		0.0014	*0.0001*		0.0014	*0.0001*	
Original	$\beta_{\mu 0}$	1.9149	*0.2550*	78.1419	1.3832	*1.2319*	80.7492	1.2200	*1.1900*	84.9623
APGP3	$\beta_{\mu 1}$	0.0248	*0.0115*		0.0478	*0.0610*	*-3.4364*	0.0583	*0.0595*	*-3.4368*
	β_{a0}	-0.1401	*0.0204*		-0.1492	*0.0429*	642.289	-0.1480	*0.0430*	642.373
	β_{a1}	0.0039	*0.0007*		0.0043	*0.0011*		0.0043	*0.0011*	
	β_{a2}	-0.00002	*0.00001*		-0.00003	*0.00001*		-0.00003	*0.00001*	
Simplified	$\beta_{\mu 0}$	1.9149	*0.2550*	78.1419	1.6787	*0.2212*	78.6224	1.6710	*0.2230*	78.6387
APGP3	$\beta_{\mu 1}$	0.0248	*0.0115*		0.0336	*0.0106*	*-3.2746**	0.0339	*0.0107*	*-3.2747*
	β_{a0}	-0.1401	*0.0204*		-0.1478	*0.0109*	612.523*	-0.1477	*0.0109*	612.537
	β_{a1}	0.0039	*0.0007*		0.0041	*0.0003*		0.0041	*0.0003*	
	β_{a2}	-0.00002	*0.00001*		-0.00002	*3* $\times 10^{-6}$		-0.00002	*3* $\times 10^{-6}$	

The chosen model is indicated by '*'. The three GOF measures are MSE_2, U (in italic) and AIC respectively.

Figure 4.1: Plot of the observed w_t and expected $\hat{w}_t$ daily infected SARS cases in Hong Kong using simplified APGP3 model



Remark: The simplified APGP3 models using the ML method (blue) and Bayesian method (green) are overlapped as their parameter estimates are similar as shown in Table 4.3.

Chapter 5

Threshold Poisson Geometric Process Model

5.1 The model

Another useful extension of the PGP model would be the threshold PGP (TPGP) model. Different from the APGP model where a single PGP model is considered, the TPGP model allows the modeling of a multiple trend data by using a series of PGP and/or APGP models. Although the APGP model with ratio function a_t allows a gradual change in the progression of trend for a multiple trend data, the change is usually smooth and continuous. However, if there are sudden changes in the trend or if the trend fluctuates with many abrupt turns, the APGP model may not be the best model to account for such situation. Instead, we can adopt the idea of the multiple Geometric Process (MGP) model in Ho (25) and Chan *et al.* (10) to extend the PGP model and the resultant model is called the threshold PGP (TPGP) model. The model can be used to describe some discontinuous trends and trends with abrupt patterns.

Definition: Let the observed data be W_t with time $t = 1, \ldots, n$. Suppose that there are K turning points $T_\kappa, \kappa = 1, \ldots, K$ for the K trends in the data, and n_κ is the number of observations in the κ^{th} trend modeled by the κ^{th} PGP. A turning point marks the time when a trend in the data changes its strength and/or

directions. Hence we write

$$PGP_\kappa = \{W_t,\ T_\kappa \le t < T_{\kappa+1}\}, \quad \kappa = 1, \ldots, K$$

for the κ^{th} PGP which begins with a turning point T_κ. In other words,

$$T_1 = 1 \qquad \text{and} \qquad T_\kappa = 1 + \sum_{j=1}^{\kappa-1} n_j, \quad \kappa = 2, \ldots, K$$

are the turning points of the PGPs and $\sum_{\kappa=1}^{K} n_\kappa = n$. Then, we define the corresponding latent GP and RP as

$$\begin{aligned} GP_\kappa &= \{X_t,\ T_\kappa \le t < T_{\kappa+1}\} \\ \text{and} \quad RP_\kappa &= \{Y_t = a_t^{t-T_\kappa} X_t,\ T_\kappa \le t < T_{\kappa+1}\}, \quad \kappa = 1, \ldots, K \end{aligned}$$

where $a_t > 0$ is the ratio parameter of the κ^{th} GP and $T_{K+1} = n+1$. By assigning an exponential distribution to $\{Y_t\}$ with mean $\mu_t = \lambda_t$ to the RPs in the TPGP model, the pmfs for the observed data W_t would be

$$f_{ot}(w_t) = \frac{a_t^{t-T_\kappa}/\mu_t}{(a_t^{t-T_\kappa}/\mu_t + 1)^{w_t+1}} \tag{5.1}$$

$$\text{and} \quad f_{st}(w_t) = \frac{\exp\left(-\frac{\mu_t}{a_t^{t-T_\kappa}}\right)\left(\frac{\mu_t}{a_t^{t-T_\kappa}}\right)^{w_t}}{w_t!} \tag{5.2}$$

for the original and simplified model respectively where $T_\kappa \le t < T_{\kappa+1}$. Based on the pmfs of W_t, we can derive

$$\text{Original TPGP model:} \quad E_o(W_t) = \frac{\mu_t}{a_t^{t-T_\kappa}} \quad \text{and} \quad Var_o(W_t) = \frac{\mu_t}{a_t^{t-T_\kappa}} + \left(\frac{\mu_t}{a_t^{t-T_\kappa}}\right)^2$$

$$\text{Simplified TPGP model:} \quad E_s(W_t) = Var_s(W_t) = \frac{\mu_t}{a_t^{t-T_\kappa}}$$

as the mean and variance for the κ^{th} PGP.

To further extend the application of the TPGP model, we adopt a linear function of covariates log-linked to the mean function or the ratio function

$$\mu_t = \exp(\eta_{\mu t}) = \exp(\beta_{\mu 0\kappa} z_{\mu 0t} + \beta_{\mu 1\kappa} z_{\mu 1t} + \ldots + \beta_{\mu q_\mu \kappa} z_{\mu q_\mu t}) \quad (5.3)$$

$$\text{and} \quad a_t = \exp(\eta_{at}) = \exp(\beta_{a0\kappa} z_{a0t} + \beta_{a1\kappa} z_{a1t} + \ldots + \beta_{aq_a\kappa} z_{aq_a t}) \quad (5.4)$$

where $T_\kappa \leq t < T_{\kappa+1}$, $z_{j0t} = 1$ and $z_{jkt}, k = 1, \ldots, q_j, j = \mu, a$ are covariates, to accommodate some time-evolving effects as in the APGP model. The resultant model is called the threshold adaptive PGP (TAPGP) model with $\hat{w}_t = E(W_t) = \mu_t / a_t^{t-T_\kappa}$ for the κ^{th} PGP.

5.2 Trend identification

For the estimation of the number and location of the turning point(s), three methods have been proposed. They included the moving window technique, the Bayesian method and the partial likelihood method. Chan *et al.* (10) adopted the moving window technique to identify the turning points for the daily infected cases of SARS in Hong Kong. In this method, GP models are fitted $n - L + 1$ times to subsets of data as windows with width L (from observations $t = i$ to $t = i + L - 1$) move across the data from $i = 1$, $i = 2$ to $i = N - L + 1$. Turning points are detected when the ratio a_i changes from $a_i < 1$ to $a_i > 1$ or vice versa. Depending on the nature of data, the method may detect more than one turning point. Different values of L will result in different sets of turning points.

A small L can identify the changes of the ratio more accurately, however, noises may also be captured. However, a large L may average the changes of the ratio too much leading to a loss of information. The window width L and the number of turning points K are chosen according to some model selection measures in real data analysis. Chan *et al.* (10) suggested to use the adjusted mean-square error (AMSE)

$$AMSE = c(P) + \frac{1}{n}\sum_{t=1}^{n}[W_t - E(W_t)]^2, \quad \text{for} \quad T_\kappa \leq t < T_{\kappa+1}$$

which is the sum of the mean-square error (MSE) and a penalized term of a scaler multiple of the number of parameters P for the GP model. They took $c = \frac{1}{2}\ln(\bar{w})$ where $\bar{w} = \frac{1}{n}\sum_{t=1}^{n} W_t$ so that higher average level of $\bar{w}$ leads to heavier penalty. This is because data with higher level of $\bar{w}$ usually have more fluctuations or turns in W_t, some of which may due to white noise. The use of heavier penalty for data with higher level of $\bar{w}$ reduces the number of turns and hence makes the model less complicated. However, we will not consider this technique in the thesis because it is too computationally intensive.

In this thesis, we will use the other two methods, namely the Bayesian method and the partial likelihood or LSE method. In the Bayesian analysis, the turning points are set as the model parameters and would be estimated from the data. Usually, noninformative priors, say $T_\kappa \sim U(T_{\kappa-1} + b, n)$ where b is the minimum width for any trend, are adopted for T_κ. In the partial maximum likelihood method say, the turning points are located by some searching methods which consist of three parts: (1) location of turning points, (2) estimation of

parameters and (3) selection of turning points by maximal U method or minimal MSE method where U is a goodness-of-fit (GOF) measures for model assessment proposed by Walker and Gutiérrez-Peña (61). TPGP or TAPGP model with the largest value of U or minimal value of MSE is preferred as it implies higher likelihood or lower discrepancy with the observed data. Then other model parameters including $\beta_{ak\kappa}$ and $\beta_{\mu k\kappa}$ are estimated by using ML method (Chan and Leung (9)). Similarly, the partial LSE method adopts the same idea, but another GOF measures, namely the mean squares of error (MSE) is used instead of U. The next section will discuss the three methodologies of inference implemented in the TAPGP model.

5.3 Methodology of inference

5.3.1 Partial least-square error method

The partial LSE method involves three parts: (1) location of turning points T_κ by plotting, (2) estimation of parameters by LSE method and (3) selection of turning points by minimal MSE method. The quantity MSE is a GOF measure and is defined as

$$MSE = \frac{1}{n}\sum_{\kappa=1}^{K}\sum_{t=T_\kappa}^{T_{\kappa+1}-1}\left(w_t - \frac{\mu_t}{a_t^{t-T_\kappa}}\right)^2 \tag{5.5}$$

where μ_t and a_t are given by (5.3) and (5.4) respectively. Obviously, the smaller the MSE, the smaller is the deviation between the observation and its expected

value. Hence, the set of turning points with the smallest MSE is chosen.

In the partial LSE method, firstly, we plot the data to identify the ranges of time t where turning points lie. We consider all combinations of the K turning points, one from each range. From these combinations of K turning points, secondly, we implement the LSE method as in Section 4.2.1 to find the LSE estimates for the vectors of parameters $\boldsymbol{\theta}_{h\kappa} = (\beta_{\mu 0\kappa}, \ldots, \beta_{\mu q_\mu \kappa}, \beta_{a0\kappa}, \ldots, \beta_{aq_a\kappa})^T, h = o, s, \kappa = 1, \ldots, K$ for each APGP model fitted to each PGP separately, and MSE in (5.5) is calculated for the fitted model. Thirdly, we locate one set of turning points $T_\kappa, \kappa = 1, \ldots, K$ which gives the minimal MSE and the corresponding $\boldsymbol{\theta}_{h\kappa}$ is taken to be $\hat{\boldsymbol{\theta}}_h$ which is the model parameter estimates for the TAPGP model.

5.3.2 Partial likelihood method

Similar to the partial LSE method, the partial likelihood method also consists of three parts. The posterior expected utility U used in this method is similar to that used in the PGP and APGP models and is defined as

$$U = \frac{1}{n} \sum_{\kappa=1}^{K} \ln[L_{h,\kappa}(w_t|\, \hat{\boldsymbol{\theta}}_h)] \tag{5.6}$$

where $L_{h,\kappa}(w_t|\, \hat{\boldsymbol{\theta}}_h) = \prod_{t=T_\kappa}^{T_{\kappa+1}-1} f_{ht}(w_t)$ and $f_{ht}(w_t)$ is given by (5.1) and (5.2) respectively for the original and simplified model. Again, the partial likelihood method involves three parts as in Section 5.3.1 except that we use U as the GOF measure. Finally the set of turning points $\{T_\kappa, \kappa = 1, \ldots, K\}$ which gives the maximum

U is selected and the corresponding $\boldsymbol{\theta}_{h\kappa}$ is taken to be the model parameter estimates $\hat{\boldsymbol{\theta}}_h$.

5.3.3 Bayesian method

The Bayesian hierarchy of the original and simplified TAPGP model for solving the model parameters $\boldsymbol{\theta}$ are

$$\begin{aligned}
\text{Likelihood for the original model: } w_t &\sim f_{ot}(w_t) \\
f_{ot}(w_t) &= \frac{a_t^{t-T_\kappa}/\mu_t}{(a_t^{t-T_\kappa}/\mu_t+1)^{w_t+1}}, \quad T_\kappa \le t < T_{\kappa+1} \\
\mu_t &= \exp(\beta_{\mu 0\kappa} z_{\mu 0t} + \beta_{\mu 1\kappa} z_{\mu 1t} + \cdots + \beta_{\mu q_\mu \kappa} z_{\mu q_\mu t}) \\
a_t &= \exp(\beta_{a0\kappa} z_{a0t} + \beta_{a1\kappa} z_{a1t} + \cdots + \beta_{aq_a\kappa} z_{aq_a t})
\end{aligned}$$

$$\begin{aligned}
\text{Likelihood for the simplified model: } w_t &\sim Poi(\lambda_t) \\
\lambda_t &= \frac{\mu_t}{a_t^{t-T_\kappa}}, \quad T_\kappa \le t < T_{\kappa+1} \\
\mu_t &= \exp(\beta_{\mu 0\kappa} z_{\mu 0t} + \beta_{\mu 1\kappa} z_{\mu 1t} + \cdots + \beta_{\mu q_\mu \kappa} z_{\mu q_\mu t}) \\
a_t &= \exp(\beta_{a0\kappa} z_{a0t} + \beta_{a1\kappa} z_{a1t} + \cdots + \beta_{aq_a\kappa} z_{aq_a t}).
\end{aligned}$$

Unlike the partial likelihood and LSE methods, the turning points T_κ are treated as one of the model parameters. Since the values for the parameters T_κ are restricted to some positive continuous ranges, uniform prior distribution is therefore

assigned to these parameters.

$$\begin{aligned} \text{Priors for parameters: } \beta_{ak\kappa} &\sim N(0, \sigma_a^2), \quad k = 0, \ldots, q_a, \kappa = 1, \ldots, K \\ \beta_{\mu k\kappa} &\sim N(0, \sigma_b^2), \quad k = 0, \ldots, q_\mu, \kappa = 1, \ldots, K \\ T_\kappa &\sim U(T_{\kappa-1} + b, n), \quad \kappa = 2, \ldots, K \end{aligned}$$

where $T_1 = 1$, b is the minimum width for any trend and n is the total number of observations.

After executing $R = 100000$ iterations and discarding the first $B = 5000$ iterations as the burn-in period, we draw every $H^{th} = 20^{th}$ sample thereafter resulting in $M = 4750$ posterior samples of T_κ, $\beta_{ak\kappa}$ and $\beta_{\mu k\kappa}$. The Bayesian parameter estimates are given by the sample means of these M posterior samples. Lastly, we check the history plots and ACF plots for all parameters to ensure their convergence and independence.

Simulation studies are not conducted because of the computationally intensive search method for the turning points in the partial ML and partial LSE methods.

5.4 Real data analysis

In Chapter 4, we have analyzed the Hong Kong SARS data using the APGP models. However, the model only allows a gradual change in the progression of the trend but the data shows a clear growing and declining stages. Chan *et al.* (10) have fitted the data using the GP model for positive continuous data and

located the turning point to lie on 13^{rd} to 14^{th} April, 2003 ($T_2 = 33, 34$) using the LSE and log-LSE methods respectively with moving window technique. Hence, we first consider a two-trend TPGP ($TPGP_2$) model using the partial LSE, the partial likelihood and the Bayesian methods.

However, some abrupt turns are observed during the growing stage when the number of daily infected cases increases or decreases dramatically. For instance, the number of cases increases rapidly from 10 cases at the beginning to 80 cases on 31^{st} March ($t = 20$). This period may be regarded as the first phase of the growing stage when some infection controls were not yet enforced or become effective. The measures were first enforced on 29^{th} March requiring all incoming travelers at the airport, borders and ports to complete a health declaration form. Schools were closed to reduce probable transmission routes. On 2^{nd} April, WHO first issued a travel advice to Hong Kong. These measures indicated that the epidemic had been spreading rapidly during that time. The second phase of the growing stage seems to occur around 1^{st} April ($t = 21$) till 10^{th} April ($t = 30$) when the second peak of 61 cases was reported. Allowing a maximum incubation period of 10 days after the first set of infection measures (on 29^{th} March), the control measures, while interacting with other regional and global measures, took 3 to 4 more days to be effective. Hence, the number of daily infected cases during this phase still increased but at a slower rate.

From 10^{th} April, when the home quarantine for 10 days was set up to monitor and treat those who have contacts to confirmed SARS patients, the number of

daily infected cases declined gradually. Then from 17^{th} April, all passengers departing Hong Kong at the airport were required to take their body temperature before check-in and those with fever or symptoms suggestive of SARS should not board a plane. Until 23^{rd} May, WHO removed the travel advice to Hong Kong and further on 23^{rd} June, Hong Kong was removed from the list of areas with recent local transmission. To model such special trend pattern, we also fit a three-trend TPGP ($TPGP_3$) model and a TAPGP ($TAPGP_3$) model, by accommodating the temperature effect in the mean function, using three estimation methods.

For the partial LSE and the partial likelihood methods, we locate ranges for the possible turning points with the help of Figure 5.1. Obviously, one neighborhood of the possible turning point lies around 31^{st} March ($t = 20$) and another neighborhood of the possible turning point lies around 10^{th} April ($t = 30$). For the $TPGP_3$ and $TAPGP_3$ models, there are two turning points and hence different combination of turning points are attempted. Finally turning point or set of turning points are selected according to the minimal MSE method and the maximal U method. For the Bayesian method, the turning point(s) is/are random and included in the parameter estimation. Similar to the APGP model, the three GOF measures: MSE, U and AIC are again used for model selection. The results of the original and simplified versions of the TPGP and TAPGP models are summarized in Table 5.1.

5.4.1 Results and comments

In general, across different estimation methods, models using the LSE method always give the smallest MSE while those using the ML method give the largest U. Moreover, comparing across the two versions of TPGP or TAPGP models, the original versions of those less complicated models have a better U but they have worse U for those more complicated models, for example, the $TPGP_3$ model. The turning point T_2 is found to be similar or the same across different estimation methods, and it is roughly located between 31^{st} March ($t = 20$) and 2^{nd} April ($t = 22$). For the first trend ($\kappa = 1$), all three methods give quite consistent estimate of $\hat{a}_1 \approx 0.91 < 1$ (increase by 9% daily). This means that the viruses spread rapidly during the initial stage of the SARS outbreak. For the second trend ($\kappa = 2$), all $\hat{a}_2 \approx 1.04$ (decrease by 4% daily) which indicates that the number of daily infected cases only declines gradually after the peak period.

As there may exist more than one turning point in the data, we also fit a three-trend TPGP ($TPGP_3$) model and compare it with the $TPGP_2$ model. Results reveal a great improvement in MSE and U for the simplified models. Specifically, the increase in U for the simplified model is significant and the U becomes larger than that of the original model. At the same time, the AIC which is associated with U also improves in a great deal despite the increasing model complexity with three more parameters T_3, a_3 and $\beta_{\mu 03}$ in the simplified $TPGP_3$ model. Thereafter, we incorporate the temperature as the covariate in the mean function in the $TAPGP_3$ model using the simplified version. Results reveal that the three

GOF measures further improve though three more parameters $\beta_{\mu 1\kappa}, \kappa = 1, 2, 3$ are included in the $TAPGP_3$ model.

Therefore, the simplified $TAPGP_3$ model using the ML method is chosen to be the best model because of its best U and AIC, and the second best MSE among other TPGP and TAPGP models. According to the best model, there are two phases during the growing stages which are started at the beginning and on 2^{nd} April respectively. In the first phase, the number of daily infected cases increased from 11 cases by 7.08% each day and those in the second phase were 26 cases and 7.24% correspondingly. Obviously, the increasing rates of the two phases were close but the first phase starts at a lower initial level and has a longer trend. Afterwards, starting from 12^{th} April ($T_3 = 32$), the number of daily infected cases dropped from 48 cases by 7.32% each day. However, the turning points from the best model are different from those reported in Chan *et al.* (10). Figure 5.1 shows that $T_2 = 22$ marked the change of rates of growth only. The maximum number of 80 cases is too high to detect another growing phase right after it. Besides, regarding the covariate effect, the temperature generally showed a positive relationship with the number of daily infected cases throughout the two phases of growing stage and the declining stage. Figure 5.1 plots the simplified TPGP and TAPGP models using the ML method.

==========

Table 5.1 and Figure 5.1 about here

==========

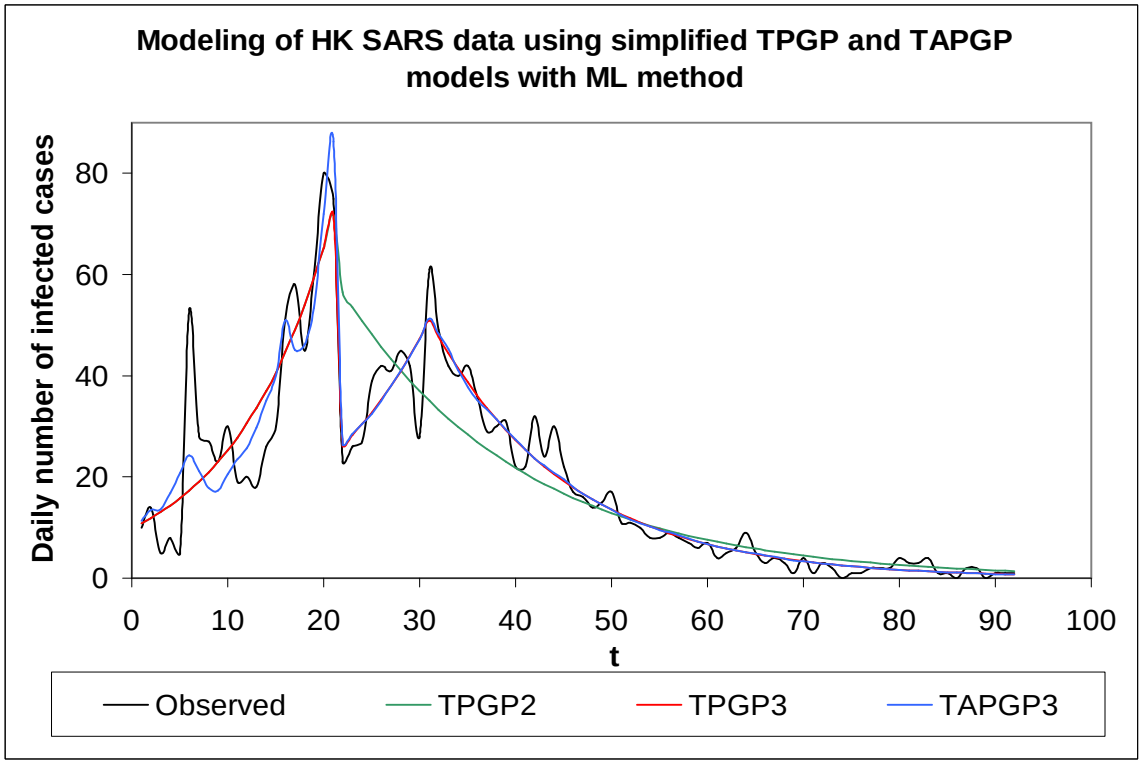
Comparing the best $TAPGP_3$ model with the best APGP model in Chapter 4, we find that the performance of the $TAPGP_3$ model is much better as shown by the MSE, the U and the AIC. Hence, we conclude that the TPGP or TAPGP models are more suitable for fitting the SARS data as they can deal with the abrupt turns in the trend pattern. After all, the TPGP or TAPGP models also have their weaknesses. First, when the number of trends becomes large, the number of parameters will increase substantially and at the same time we need to consider a large number of combinations of turning points in the partial LSE and the partial likelihood methods. Then the selection of the turning points will be computationally intensive. Secondly, sometimes the number and the location of turning points are quite difficult to estimate from the observed data particularly if it has many abrupt turns. Hence, for further model development, we will explore other methods which can estimate the number and location of turning points more efficiently.

Table 5.1: Parameter estimates and standard errors for the Hong Kong SARS data using TPGP models

Method		LSE			ML			Bay		
Model	Par.	Est.	SE	GOF	Est.	SE	GOF	Est.	SE	GOF
Original	$\beta_{\mu 01}$	2.2449	*0.0783*	71.9347	2.4498	*0.4614*	101.629	2.6790	*0.5145*	107.481
$TPGP_2$	$\beta_{\mu 02}$	3.8649	*0.0460*		4.3532	*0.2477*	*-3.4479*	4.1650	*0.5031*	*-3.4498*
	a_1	0.9015	*0.0043*		0.9177	*0.0401*	644.409	0.9396	*0.0583*	644.771
	a_2	1.0414	*0.0027*		1.0635	*0.0069*		1.0650	*0.0089*	
	T_2	22			20			23.79	*8.6060*	
Simplified	$\beta_{\mu 01}$	2.2449	*0.0783*	71.9347	2.3779	*0.0996*	80.5290	2.4300	*0.1154*	85.3676
$TPGP_2$	$\beta_{\mu 02}$	3.8649	*0.0460*		4.0340	*0.0453*	*-3.4941*	4.1670	*0.1000*	*-3.5157*
	a_1	0.9015	*0.0043*		0.9096	*0.0063*	652.909	0.9157	*0.0094*	656.882
	a_2	1.0414	*0.0027*		1.0544	*0.0021*		1.0550	*0.0022*	
	T_2	22			22			20.05	*1.2990*	
Original	$\beta_{\mu 01}$	2.2449	*0.0783*	39.7879	2.4116	*0.4342*	41.1708	2.6220	*0.5966*	118.229
$TPGP_3$	$\beta_{\mu 02}$	3.2842	*0.0042*		3.2704	*0.5857*	*-3.4337*	3.7450	*4.1760*	*-3.4689*
	$\beta_{\mu 03}$	3.8628	*0.0569*		3.7266	*0.2634*	647.796	3.7130	*1.9000*	654.277
	a_1	0.9015	*0.0043*		0.9121	*0.0336*		0.9460	*0.1457*	
	a_2	0.9318	*0.0062*		0.9314	*0.0928*		1.0550	*0.1744*	
	a_3	1.0720	*0.0062*		1.0691	*0.0091*		1.0650	*0.1352*	
	T_2	22			22			18.78	*11.780*	
	T_3	32			33			33.06	*28.330*	
Simplified	$\beta_{\mu 01}$	2.2449	*0.0783*	39.7879	2.3779	*0.0996*	40.2427	2.3750	*0.1008*	40.6434
$TPGP_3$	$\beta_{\mu 02}$	3.2842	*0.0042*		3.2700	*0.1064*	*-2.9245*	3.2830	*0.1271*	*-2.9303*
	$\beta_{\mu 03}$	3.8628	*0.0569*		3.8649	*0.0551*	554.115	3.9420	*0.0985*	555.176
	a_1	0.9015	*0.0043*		0.9096	*0.0063*		0.9095	*0.0063*	
	a_2	0.9318	*0.0062*		0.9294	*0.0170*		0.9430	*0.0230*	
	a_3	1.0720	*0.0062*		1.0725	*0.0033*		1.0730	*0.0034*	
	T_2	22			22			21.51	*0.2881*	
	T_3	32			32			31.08	*1.0720*	
Simplified	$\beta_{\mu 01}$	1.2658	*0.1734*	34.7891	1.1418	*0.3237*	35.1065	1.1330	*0.3236*	36.5961
$TAPGP_3$	$\beta_{\mu 02}$	2.7212	*0.2737*		3.1865	*0.7219*	*-2.8348**	4.6080	*1.2020*	*-3.3918*
	$\beta_{\mu 03}$	3.8471	*0.6012*		3.6657	*0.5957*	543.601*	3.6040	*0.7185*	545.706
	$\beta_{\mu 11}$	0.0674	*0.0102*		0.0724	*0.0178*		0.0727	*0.0177*	
	$\beta_{\mu 12}$	0.0230	*0.0110*		0.0035	*0.0294*		-0.0539	*0.0457*	
	$\beta_{\mu 13}$	0.0007	*0.0247*		0.0082	*0.0245*		0.0177	*0.0320*	
	a_1	0.9302	*0.0053*		0.9292	*0.0078*		0.9292	*0.0077*	
	a_2	0.9195	*0.0083*		0.9276	*0.0227*		0.9746	*0.0660*	
	a_3	1.0720	*0.0064*		1.0732	*0.0039*		1.0740	*0.0040*	
	T_2	22			22			21.50	*0.2886*	
	T_3	32			32			29.72	*1.5870*	

The chosen model is indicated by '*'. The three GOF measures are MSE, U (in italic) and AIC respectively.

Figure 5.1: Plot of the observed w_t and expected $\hat{w}_t$ daily infected SARS cases in Hong Kong using simplified TPGP and TAPGP models with ML method



Chapter 6

Mixture Poisson Geometric Process Model

6.1 The model

So far the GP models are applied to a single system to study the inter-arrival times of the operating system such as the number of infected cases of an epidemic and the inter-arrival times of coal-mining disasters in a region, etc. However, we often encounter multiple systems in daily life. A prominent example will be the longitudinal measurements of a large number of patients each undergoing similar treatments in clinical trials. Observations from patients with different personal characteristic and undergoing different treatments exhibit considerable population heterogeneity and clustering effects. As a result, observations from different patients show different trend patterns. Chan *et al.* ((6), (7)) adopted a mixture model using expectation maximization (EM) method to study the effect of methadone treatment in reducing multiple drug use for patients at a Sydney clinic in 1986. Later, Chan and Leung (9) proposed a mixture BGP model to study the data using ML and Bayesian methods. The model describes the probability of heroin taking which varies across different groups of patients: heavy user or light user. To be specific, the model accounts for different heroin-taking patterns among patients, incorporates covariate effects to allow for patients' characteristics and at the same time classifies patients into different heroin-taking groups. The

proposed model is an extension of the GP model to binary data that incorporate covariate and mixture effects in the mean and ratio functions.

Let W_{it} denote the observed event count for subject $i, i = 1, \ldots, m$ at time $t, t = 1, \ldots, n_i$ and $n = \sum_{i=1}^{m} n_i$. The proposed MPGP and MAPGP models assume that there are G groups of subjects who have different trend patterns and each subject has a probability π_l of coming from group $l, l = 1, \ldots, G$. The resultant model is essentially a discrete random effects model such that $\boldsymbol{\beta}_j = \boldsymbol{\beta}_{j\,l}, j = \mu, a$ with probability π_l. The parameters $\boldsymbol{\beta}_{j\,l} = (\beta_{j0l}, \beta_{j1l}, \ldots, \beta_{jq_jl})^T$ which is a vector of parameters β_{jkl}, where $j = \mu, a$, $k = 0, \ldots, q_j$ and $l = 1, \ldots, G$. The probability mass function (pmf) for subject i with group membership l would be:

$$\text{Original:} \quad f_{o,\,itl}(w_{it}) \;=\; L_{o,\,itl} = \frac{a_{itl}^{t-1}/\mu_{itl}}{(a_{itl}^{t-1}/\mu_{itl}+1)^{w_{it}+1}} \tag{6.1}$$

and

$$\text{Simplified:} \quad f_{s,\,itl}(w_{it}) \;=\; L_{s,\,itl} = \frac{\exp\left(-\dfrac{\mu_{itl}}{{a_{itl}}^{t-1}}\right)\left(\dfrac{\mu_{itl}}{{a_{itl}}^{t-1}}\right)^{w_{it}}}{w_{it}!} \tag{6.2}$$

respectively for the original ($h = o$) and simplified ($h = s$) mixture models. And, the mean and ratio functions, μ_{itl} and a_{itl}, are given correspondingly by:

$$\mu_{itl} \;=\; \exp(\eta_{\mu itl}) = \exp(\beta_{\mu 0l} + \beta_{\mu 1l} z_{\mu 1it} + \cdots + \beta_{\mu q_\mu l} z_{\mu q_\mu it}) \tag{6.3}$$

$$a_{itl} \;=\; \exp(\eta_{aitl}) = \exp(\beta_{a0l} + \beta_{a1l} z_{a1it} + \cdots + \beta_{aq_a l} z_{aq_a it}) \tag{6.4}$$

where $z_{jkit}, j = \mu, a$, $k = 1, \ldots, q_j, i = 1, \ldots, m$ are the time-evolving covariate effects at time $t = 1, \ldots, n_i$. Since the group membership of each subject is unobserved, the likelihood function for the observed data $\{W_{it}\}$, denoted by $L_h(\boldsymbol{\theta}_h)$,

involves conditional expectation taken over latent group l.

$$L_h(\boldsymbol{\theta}_h) = \prod_{i=1}^{m} \left[\sum_{l=1}^{G} \pi_l \left(\prod_{t=1}^{n_i} L_{h,itl} \right) \right] \tag{6.5}$$

and the log-likelihood function $\ell_h(\boldsymbol{\theta}_h)$ is given by:

$$\ell_h(\boldsymbol{\theta}_h) = \sum_{i=1}^{m} \ln \left[\sum_{l=1}^{G} \pi_l \left(\prod_{t=1}^{n_i} L_{h,itl} \right) \right]. \tag{6.6}$$

where $L_{h,itl}$ are given by (6.1) and (6.2) respectively. This mixture model has a tractable likelihood function which does not involve integration. Hence, a full likelihood analysis is possible. The vector of model parameters is $\boldsymbol{\theta}_h = (\boldsymbol{\theta}_{h1}^T, \ldots, \boldsymbol{\theta}_{hG}^T, \pi_1, \ldots, \pi_{G-1})^T$ where $\boldsymbol{\theta}_{hl}$ are vectors of parameters $(\boldsymbol{\beta}_{\mu 0\, l}, \ldots, \boldsymbol{\beta}_{\mu q_\mu l}$, $\boldsymbol{\beta}_{a0l}, \ldots, \boldsymbol{\beta}_{aq_a l})^T$.

6.2 Methodology of inference

For model implementation, we investigate three methodologies of inference which are adopted under different situations. They include the classical maximum likelihood (ML) method, the Bayesian method and the expectation maximization (EM) method. However, the model implementation is much more complicated due to the inclusion of mixture effects. Because of the unobserved group membership for each subject, conditional expectation is taken in the likelihood function leading to considerable complication in the differentiation for the NR iterative procedures. Moreover, the LSE method is not feasible for the MAPGP model because of the inability to estimate the model parameters.

6.2.1 Maximum likelihood method

We assume that there are only two groups ($l = 2$) of subjects having different trend patterns with no time-varying effect in the ratio function, that is $a_{itl} = a_l$. The classical ML method would be adopted for parameter estimation. The likelihood function in (6.5) for the original MAPGP model ($h = o$) would become:

$$L_o(\boldsymbol{\theta}_o) = \prod_{i=1}^{m} \left\{ \pi_1 \left[\prod_{t=1}^{n_i} \frac{a_1^{t-1}/\mu_{it1}}{(a_1^{t-1}/\mu_{it1} + 1)^{w_{it}+1}} \right] + \pi_2 \left[\prod_{t=1}^{n_i} \frac{a_2^{t-1}/\mu_{it2}}{(a_2^{t-1}/\mu_{it2} + 1)^{w_{it}+1}} \right] \right\}$$

where μ_{itl} is given by (6.3), $\pi_2 = 1 - \pi_1$ and $0 < \pi_1 \leq 1$. For the simplified MAPGP model ($h = s$), the likelihood function in (6.5) is as follows:

$$\begin{aligned} L_s(\boldsymbol{\theta}_s) \;=\; \prod_{i=1}^{m} \Bigg\{ & \pi_1 \left[\prod_{t=1}^{n_i} \frac{\exp\left(-\frac{\mu_{it1}}{{a_{it1}}^{t-1}}\right)\left(\frac{\mu_{it1}}{{a_{it1}}^{t-1}}\right)^{w_{it}}}{w_{it}!} \right] \\ & + \pi_2 \left[\prod_{t=1}^{n_i} \frac{\exp\left(-\frac{\mu_{it2}}{{a_{it2}}^{t-1}}\right)\left(\frac{\mu_{it2}}{{a_{it2}}^{t-1}}\right)^{w_{it}}}{w_{it}!} \right] \Bigg\} \end{aligned}$$

where μ_{itl} is again given by (6.3), $\pi_2 = 1 - \pi_1$ and $0 < \pi_1 \leq 1$. To allow for the constraint, we may transform it into $\pi_1 = \dfrac{e^{\delta}}{1 + e^{\delta}}$ where δ has an unlimited range. However, our experience shows that π_1 always satisfies $0 < \pi_1 \leq 1$ even without the transformation because the group l probability is far away from both 0 and 1. We then derive the first order and second order derivative functions for both original and simplified MAPGP models in Appendix 6.1 so that the NR iterative method can be implemented for solving the parameters.

For $l \geq 3$, the parameters $\pi_l \geq 0$ will have to satisfy the constraint: $\sum_l \pi_l = 1$, $0 < \pi_l \leq 1$ and $\pi_G = 1 - \sum_{l=1}^{G-1} \pi_l$ which is more complicated than the case when $l = 2$. Indeed, there are several ways to allow for the constraints in the classical ML method, amongst some methods stating the constraints explicitly. Without the constraint, the likelihood function and its derivative functions will be very sensitive to a small change in the updated parameters. Hence, the NR iterative method would become unstable when some of the π_l are negative. To solve for the constraints problem, we propose an alternative methodology called the expectation maximization (EM) method so that a multi-group mixture model can be implemented.

6.2.2 Expectation maximization method

The EM method was proposed by Dempster *et al.* (16) to handle some missing values in the incomplete data. The data is incomplete in the sense that there are some hidden parameters which are not systematically observed. This method is often applied to mixture models and random intercept models when there are latent group memberships and random effects (Gilks *et al.* (20)). As compared to the classical ML method where the conditional expectation (E-step) and maximization (M-step) are done simultaneously, the E-step and M-step are done separately and alternatively in iterations in the EM method.

Despite its slow convergence rate, the method is able to handle mixture model with several latent groups. Individual latent group membership can be estimated

in the E-step. This is an advantage over the classical ML method. The EM method consists of two steps: E-step and M-step. In the E-step, the missing group membership are estimated by taking conditional expectation using the observed data and current parameter estimates. Let $I_{hil}, h = o, s$ be the indicator that subject i belongs to group $l, l = 1, \ldots, G$ such that $I_{hil} = 1$ if the subject i belongs to group l and $I_{hil} = 0$ otherwise. Note that I_{hil} are unobserved since the group membership of each subject is unknown and it is estimated by taking conditional expectation:

$$E_o(I_{oil}|\,\hat{\boldsymbol{\theta}}_o, \boldsymbol{w}_i) = \hat{I}_{oil} = \frac{\pi_l \prod_{t=1}^{n_i} \frac{a_{itl}^{t-1}/\mu_{itl}}{(a_{itl}^{t-1}/\mu_{itl}+1)^{w_{it}+1}}}{\sum_{l'=1}^{G} \pi_{l'} \left[\prod_{t=1}^{n_i} \frac{a_{itl'}^{t-1}/\mu_{itl'}}{(a_{itl'}^{t-1}/\mu_{itl'}+1)^{w_{it}+1}}\right]} \tag{6.7}$$

and

$$E_s(I_{sil}|\,\hat{\boldsymbol{\theta}}_s, \boldsymbol{w}_i) = \hat{I}_{sil} = \frac{\pi_l \prod_{t=1}^{n_i} \frac{\exp\left(-\mu_{itl}/{a_{itl}}^{t-1}\right)\left(\mu_{itl}/{a_{itl}}^{t-1}\right)^{w_{it}}}{w_{it}!}}{\sum_{l'=1}^{G} \pi_{l'} \left[\prod_{t=1}^{n_i} \frac{\exp\left(-\mu_{itl'}/{a_{itl'}}^{t-1}\right)\left(\mu_{itl'}/{a_{itl'}}^{t-1}\right)^{w_{it}}}{w_{it}!}\right]} \tag{6.8}$$

evaluated at current parameter estimates $\hat{\boldsymbol{\theta}}_h$. In the M-step, the likelihood function conditioning on the missing I_{hil}, called the "complete-data" likelihood, is:

$$L_{ch}(\boldsymbol{\theta}_h) = \prod_{i=1}^{m} \prod_{l=1}^{G} \left(\pi_l \prod_{t=1}^{n_i} L_{h,itl}\right)^{I_{hil}},$$

where $h = o, s$ denotes the original and simplified PGP model and $L_{h,itl}$ are given by (6.1) and (6.2) accordingly. Hence the log-likelihood functions for the two

MAPGP models are respectively:

$$\ell_{co}(\boldsymbol{\theta}_o) = \sum_{l=1}^{G} \ln \pi_l \left(\sum_{i=1}^{m} I_{oil} \right) + \sum_{i=1}^{m} \sum_{l=1}^{G} \sum_{t=1}^{n_i} I_{oil} \left[-\ln \mu_{itl} + (t-1) \ln a_l - (w_{it}+1) \ln(a_l^{t-1}/\mu_{itl} + 1) \right] \quad (6.9)$$

and

$$\ell_{cs}(\boldsymbol{\theta}_s) = \sum_{l=1}^{G} \ln \pi_l \left(\sum_{i=1}^{m} I_{sil} \right) + \sum_{i=1}^{m} \sum_{l=1}^{G} \sum_{t=1}^{n_i} I_{sil} \left[-a_l^{-(t-1)} \mu_{itl} + w_{it} \ln \mu_{itl} - w_{it}(t-1) \ln a_l - \ln(w_{it}!) \right] \quad (6.10)$$

where μ_{itl} is given by (6.3) and $a_{itl} = a_l$ in general. Then I_{hil} is replaced by $\hat{I}_{hil}$ in (6.9) and (6.10) respectively and the resulting complete-data log-likelihood function is maximized to obtain the updated parameter estimates. The first and second order derivative functions of the log-likelihood function in the M-step as required in the NR iterative method are given in Appendix 6.2.

The procedures iterate between the E-step and M-step until convergence is attained. As NR iterative method is used in the M-step, it requires iterations within iterations and becomes more computationally intensive. Clearly, the M-step in the EM algorithm involves much less complicated log-likelihood functions and derivative functions than the classical ML method. Hence, it is more preferred in the multi-group MAPGP model. However, there are still some shortcomings of the EM method, for instance, our experience shows that the EM method is often sensitive to starting values of the parameter estimates. Alternatively, we propose the Bayesian method for the inference of MAPGP model. Moreover, Bayesian estimates are also needed as the starting points in the EM method.

6.2.3 Bayesian method

To implement the Bayesian method, we set up the following Bayesian hierarchy for the MAPGP model:

$$\begin{aligned}
\text{Likelihood for the original model: } w_{it} &\sim f_{o,it}(w_{it}) \\
f_{o,it}(w_{it}) &= I_{oi1} f_{o,it1}(w_{it}) + \cdots + I_{oiG} f_{o,itG}(w_{it}) \\
f_{o,itl}(w_{it}) &= \frac{\lambda_{itl}^{-1}}{(\lambda_{itl}^{-1}+1)^{w_{it}+1}} \\
\lambda_{itl} &= \frac{\mu_{itl}}{{a_{itl}}^{t-1}} \\
\text{Likelihood for the simplified model: } w_{it} &\sim Poi(\lambda_{it}) \\
\lambda_{it} &= I_{si1}\lambda_{it1} + \cdots + I_{siG}\lambda_{itG} \qquad (6.11) \\
\lambda_{itl} &= \frac{\mu_{itl}}{{a_{itl}}^{t-1}}
\end{aligned}$$

where μ_{itl} is given by (6.3) and $a_{itl} = a_l$ or is given by (6.4) in general. Priors for the model parameters $\boldsymbol{\theta}_h$:

$$\begin{aligned}
a_l &\sim U(\alpha_1, \alpha_2) \text{ or } \beta_{akl} \sim N(0, \sigma_{akl}^2),\ k = 0, 1, \ldots, q_a; l = 1, \ldots, G \\
\beta_{\mu kl} &\sim N(0, \sigma_{\mu kl}^2),\ k = 0, 1, \ldots, q_\mu;\ l = 1, \ldots, G \\
\pi_l &\sim U(0,1),\ l = 1, \ldots, G-1,\ \sum_{l=1}^{G} \pi_l = 1 \\
(I_{hi1}, \ldots, I_{hiG}) &\sim \text{Multinomial}(\pi_1, \ldots, \pi_G, 1) \qquad (6.12)
\end{aligned}$$

where $U(\alpha_1, \alpha_2)$ is a uniform distribution with range (α_1, α_2). In the real data analysis of the MAPGP model, we set $(\alpha_1, \alpha_2) = (0.5, 2.5)$ for the bladder cancer data in Section 6.3.1. Whereas for the blood donation data in Section 6.3.2, we set $(\alpha_1, \alpha_2) = (0.95, 150)$ for female donors, $(\alpha_1, \alpha_2) = (0.95, 5000)$ for male

donors. To sample π_l from $U(0,1)$ subject to $\sum_{l=1}^{G} \pi_l = 1$, we draw $b_l \sim U(0,1),\ l = 1,\ldots,G-1$ and set $\{c_l = b_{(l)}\}$ the rank order set of $\{b_l\}$ arranged in ascending order. Then $\pi_l = d_{(l)}$, where $d_l = c_l - c_{l-1},\ l = 1,\ldots,G$, $c_0 = 0$ and $c_G = 1$.

The parameters are drawn from their posterior conditional distributions $[\boldsymbol{\beta}_\mu|\,\boldsymbol{a}, \boldsymbol{\pi}, \boldsymbol{I}, \boldsymbol{w}]$, $[\boldsymbol{a}|\,\boldsymbol{\pi}, \boldsymbol{I}, \boldsymbol{\beta}_\mu, \boldsymbol{w}]$, $[\boldsymbol{\pi}|\,\boldsymbol{I}, \boldsymbol{\beta}_\mu, \boldsymbol{a}, \boldsymbol{w}]$ and $[\boldsymbol{I}|\,\boldsymbol{\beta}_\mu, \boldsymbol{a}, \boldsymbol{\pi}, \boldsymbol{w}]$ by Gibbs sampler where $\boldsymbol{\beta}_\mu\,\boldsymbol{a}, \boldsymbol{\pi}, \boldsymbol{I}, \boldsymbol{w}$ denote the vectors of $\beta_{\mu kl}, a_l, \pi_l, I_{hil}, w_{it}$ respectively. For each model, $R = 33000$ iterations are executed and the first $B = 5000$ iterations are set as the burn-in period to ensure that convergence has reached. Thereafter parameters are sub-sampled from every 20^{th} iteration ($H = 20$) to reduce the auto-correlation in the sample. Each resulting sample contains $M = 1400$ realizations and the estimates of the parameters are given by the sample means or medians of these samples. To check for convergence and independence, history and auto-correlation function (ACF) plots of each parameter in the WinBUGS output are examined.

It is noted that for the MAPGP model, simulation studies are not conducted as the likelihood function involves the unobserved indicator I_{hil} which shows the group membership l for subject i. Moreover, it is computationally intensive to simulate longitudinal data with trends for hundreds of times. Hence, it is infeasible to conduct simulation studies for the mixture models.

6.3 Real data analysis

The application of the MAPGP models are illustrated using two examples: (1) the bladder cancer data and (2) the blood donation data in Hong Kong. Information of the two data sets is given in details in the following sections. For the first data set, we fit a two-group MAPGP models using the classical ML and the Bayesian methods. Whereas, for the blood donation data, we fit a two-group and a three-group MPGP and MAPGP models using the EM and Bayesian methods.

Again, for the model assessment and selection, we consider the MSE and the U as the GOF measures for the MAPGP models:

$$MSE = \frac{1}{n}\sum_{i=1}^{m}\sum_{l=1}^{G} I(l=g_i)\sum_{t=1}^{n_i}(w_{it}-\hat{w}_{itl})^2 \tag{6.13}$$

$$U = \frac{1}{n}\sum_{i=1}^{m}\ln\left[\sum_{l=1}^{G}\pi_l\left(\prod_{t=1}^{n_i}L_{h,itl}\right)\right] \tag{6.14}$$

where g_i, defined as $\max_l(\hat{I}_{\nu,il}) = \hat{I}_{\nu,ig_i}$, is the group membership for subject i, $\nu = ml, e, b$ indicates the classical ML, the EM and Bayesian methods respectively and the likelihood functions $L_{h,itl}$, $h = o, s$, defined in (6.1) and (6.2) respectively, are evaluated at $\hat{\boldsymbol{\theta}}_h$. Also, $\hat{w}_{itl} = \mu_{itl}/a_{itl}^{t-1}$ where μ_{itl} and a_{itl} are given by (6.3) and (6.4) respectively. These two GOF measures are different from those in the PGP, the APGP and the TPGP models as they involve the indicator $I(l = g_i)$ in both functions.

6.3.1 Bladder cancer data

6.3.1.1 The data

The bladder cancer data comes from a bladder cancer study conducted by the Veterans Administration Cooperative Urological Research Group (Hand *et al.* (22)). In the study, a group of patients who had superficial bladder tumours were selected to enter the trial. After removing the tumours inside their bodies, the patients were assigned to one of the treatment groups: placebo, an inactive substance having no therapeutic value (treatment 1), thiotepa (treatment 2) and pyridoxine (treatment 3). At subsequent visits, new tumours were removed and the treatment continued. Visits were made quarterly for 36 months, but there were many missing appointments for the patients. The data set were reported in (22). It contains 82 patients with 46 patients underwent the placebo ($z_{\mu 1it} = 1$) treatment while the rest underwent the thiotepa treatment ($z_{\mu 1it} = 2$). For each patient i, the accumulated number of tumours found was recorded after each quarterly visit. Different patients have different number of observations n_i. The maximum n_i is 12 which means that there are no missing appointments during these 36 months. Outcome variable is the number of new tumours identified in the t^{th} visit. There are $n = 571$ observed data and 413 missing data. Among the 571 observations, $n_0 = 459$ (out of 571) are zero. The data is given in Appendix 6.3 for reference.

6.3.1.2 Objectives

The objective of fitting a two-group MAPGP model for the bladder cancer data is to investigate the treatment effect across treatment groups. We assume that there are two groups of patients with different trends in the number of new tumours identified in each visit t. Those patients with no or low tumour counts usually show a non-increasing trend of counts. This means that they would be classified into group l say, with a lower initial level $\mu_{itl} = \exp(\beta_{\mu 0l} + \beta_{\mu 1l} z_{\mu 1it})$ and/or a higher declining rate $a_l \gg 1$. On the other hand, patients with high tumour counts usually show a non-decreasing trend of counts as indicated by a higher initial level μ_{itl} and/or a lower declining rate $a_l > 1$ or even a growth rate $a_l < 1$.

To investigate how the treatment groups and the tumour-level groups are related, we first classify patients in the low- or high-level tumour groups by looking at the indicator of group membership I_{hil} for each patient i as defined in (6.7) and (6.8). Then, we can study the proportion of each treatment in the low-level and high-level of tumour groups. Similarly, we can also look at the proportion of patients with a low-level or high-level of tumours in each treatment group. We expect that the thiotepa treatment will associate with higher proportion of patients with a low-level of tumours.

Whereas across the treatment groups as indicated by $z_{\mu 1it} = 1$ or 2 within the tumour group l in the MAPGP models, we determine the effectiveness of the two treatments: placebo and thiotepa, by comparing the tumour counts of the two treatment groups of patients in each tumour group ($l = 1, 2$). Again, the

more effective the treatment, the lower is the tumour count with a non-increasing trend. Thus, a negative $\beta_{\mu 1l}, l = 1, 2$ indicates that the thiotepa treatment is more effective than the placebo treatment for both low- and high-level tumour groups of patients, and vice versa.

6.3.1.3 Results and comments

The summary for the four original and simplified MAPGP models using the classical ML and the Bayesian methods are given in Table 6.1. Fitting the two-group original and simplified MAPGP models for the 82 patients, it is found that the parameter estimate $\hat{a}_l, l = 1, 2$ in both versions of MAPGP models using the classical ML and Bayesian methods are quite close to each other especially for $\hat{a}_2$. In general, the fact that $\hat{a}_1 > \hat{a}_2 > 1$ reveals that the number of new tumour counts for both groups of patients are dropping, however the decline rate for group 1 is a bit faster. For the parameter estimates $\hat{\beta}_{\mu 0l}$ and $\hat{\beta}_{\mu 1l}$, the original MAPGP model yields different results using the classical ML and the Bayesian methods, while the simplified model has quite similar results using the two methods. Nonetheless, the estimates are consistent in the sign with $\hat{\beta}_{\mu 01}$ being negative while $\hat{\beta}_{\mu 02}$ being positive and they are marginally insignificant in general. In other words, those patients in group 2 with a higher initial level and a lower decline rate, have a higher level of tumour counts. Therefore, the MAPGP models essentially classify the patients into two groups in which group 1 contains those patients who have lower counts of tumours throughout the trend, and group

2 contains those patients whose tumour counts remain in a considerably high level as $\hat{a}_2$ is close to 1.

Moreover, the parameter $\hat{\beta}_{\mu 1l}, l = 1, 2$ in all the MAPGP models are always negative with $\hat{\beta}_{\mu 11} < \hat{\beta}_{\mu 12} < 0$, this implies that the thiotepa treatment ($z_{\mu 1it} = 2$) suppresses the growth of new tumours in general for the patients in both groups, but its effect is more among the patients with lower level of tumours. However, as $\hat{\beta}_{\mu 1l}$ for all the four MAPGP models are insignificant, these effects may not be significant. This is possibly due to the small size in each combination of groups: 23, 23, 24, 12 for those receive the placebo treatment in group 1 ($l = 1, v = 1$), the placebo treatment in group 2 ($l = 2, v = 1$), the thiotepa treatment in group 1 ($l = 1, v = 2$) and the thiotepa treatment in group 2 ($l = 2, v = 2$). For patients with low levels of tumours, the levels of tumours may be too low to detect a significance difference between the two treatments. On the other hand, for patients with high level of tumours, the observed counts fluctuate a lot over time regardless of the treatment received through the trend is still decreasing in general. The situation is more obvious for the thiotepa treatment. The substantial amount of noise in the data is possibly due to the high variability in the data and the small sample size again. These effects lead to insignificant result in the detection of treatment effect among patients with high level of tumours. Moreover, these fluctuations also add the difficulties of detecting a monotone trend according to the model assumption. As a result, a generally stationary is detected as shown by $\hat{a}_2 \approx 1$

==========

Table 6.1 about here

==========

For model selection, we study the performance of the four models using U and MSE. Similar to the results in the analysis of the SARS data in Hong Kong, the original MAPGP models again give a better (larger) U but a worse (larger) MSE than the simplified models. Since the bladder cancer data contains many zero observations, the original model, which is suitable for modeling trend data with more zeros based on its pmf as given by (6.1), therefore gives a better U. However, as discussed in Section 3.7, the original MAPGP models always give a worse MSE because they tend to fit the trend data worse when the count W_{it} is large. Meanwhile, the bladder cancer data does not contain many observations with large values, the performance of the original model, hence, is not such worse in terms of MSE. The differences in the MSE are only about 0.02 using the classical ML and the Bayesian methods as shown in Table 6.1. Due to the best performance of U and an acceptable level of MSE in the original MAPGP model using the classical ML method, it is chosen as the best model for the bladder cancer data. Based on the best model, Figure 6.1 shows the observed average tumour counts w_{tlv} and the expected average tumour counts $\hat{w}_{tlv}$ for $(l, v) = (1,1)$, $(2,1)$, $(1,2)$ and $(2,2)$ during the 12 visits, as given by

$$w_{tlv} = \frac{\sum_{i=1}^{m} \hat{I}_{ml,il}\, I(z_i = v)\, w_{it}}{\sum_{i=1}^{m} \hat{I}_{ml,il}\, I(z_i = v)} \quad \text{and} \quad \hat{w}_{tlv} = \frac{\sum_{i=1}^{m} \hat{I}_{ml,il}\, I(z_i = v)\, \frac{\mu_{itl}}{a_l^{t-1}}}{\sum_{i=1}^{m} \hat{I}_{ml,il}\, I(z_i = v)}$$

across the time $t = 1, \ldots, 12$, where $i = 1, \ldots, 82$ for all patients and $\hat{I}_{ml,il}$ are given by (6.7).

==========

Figure 6.1 about here

==========

We observe in the figure that w_{t1v} and $\hat{w}_{t1v}$ are close in values for patients with low level of tumours ($l = 1$). On the contrary, as w_{t2v} for patients with high level of tumours fluctuates vigorously across each visit t, the MAPGP model fits an almost stationary trend by averaging the fluctuations, hence $\hat{w}_{t2v}$ are almost constant. Note that the observed counts w_{t1v} and w_{t2v} may not reveal the general trend of tumour counts for patients from the low-level and high-level tumour groups as they are subjective to high variability in the data and are calculated based on limited number of observations. This problem is serious especially for the high-level tumour group. Of the 23 patients in the group ($l = 2, v = 1$), the observed counts w_{t21} are averaged over 20, 17, 19, 16, 15, 17, 12, 17, 15, 13, 6 and 9 observations respectively for $t = 1, \ldots, 12$. While for the group ($l = 2, v = 2$) with 12 patients, the number of observations are 9, 10, 7, 7, 6, 8, 6, 6, 6, 5, 3 and 2 accordingly in the calculation of w_{t22}. The small number of observations, especially at the later visits ($t = 9, 10, 11, 12$), explains why the MAPGP model cannot capture the noise in the fluctuation of the observed trend for the high-level tumour group ($l = 2$) when the variation of the observations is high (up to nine tumour counts).

Moreover, Figure 6.1 reveals different treatment effects within and across tumour group l. In addition, the parameter $\hat{\pi}_1 = 55.12\%$ in the best model, this shows that there are more patients with low-level tumour counts. To be more specific, most patients from group $l, l = 1$ or 2 are clearly classified with $I_{ml,il} > 0.8$ except seven patients. Patients 1, 2, 54 who have only one zero observation are classified as having low level of tumours with $I_{ml,i1} \approx 0.7$. Besides, patient 78 who has one new tumour found in the first and the eighth visits and patient 4 who has one found in the second visit are also classified as having low level of tumours with $I_{ml,i1} \approx 0.65$. On the other hand, patients 8 and 56 with two new tumours found in the second and sixth visits accordingly and the rest of data missing or being zeros, are classified as having high level of tumours ($I_{ml,i2} \approx 0.63$). These findings give evidence that patients with many missing data or with many zeros tend to be classified unclearly.

Nevertheless, as there is insufficient information from the data with small sample size and many missing data, the detection of trend patterns among the patients becomes difficult, when comparing with the blood donation study with a much larger sample size of donors in the next section. Also, with the substantial number of zero observations in the data, the empirical distribution will depart greatly from the assumption of Poisson distribution. Moreover, the zeros also contribute substantial noise in the detection of monotone trends. These problems in fitting the MAPGP models to the bladder cancer data motivate us to study the zero-inflated PGP models in the next chapter.

6.3.2 Blood donation data

6.3.2.1 Objectives

Blood is vital to human lives. In all types of operations, blood is essential to save the lives of victims who get seriously injured in accidents, or who suffer from catastrophic diseases. Since there is no substitute for blood, the supply of blood relies solely on the voluntary, non-remunerated donors. According to the Hong Kong Red Cross Blood Transfusion Service (BTS), an average of 190,000 to 200,000 donation is required annually to meet the territory's demand. To ensure a stable and adequate blood supply, it is necessary to maintain a reasonably sized pool of blood donors by retaining previous blood donors and recruiting new donors to compensate for the dropouts. However, with an ageing population in the society and more stringent donor selection criteria, it becomes more difficult to recruit new donors than to encourage previous donors to return and become regular donors. The latter has the advantage that previous donors are in general safer and more committed. Moreover, they are usually younger and if committed, can form a stable source of blood supply.

The objective of studying the blood donation data is to investigate individual blood donation pattern among Chinese first-time donors for a blood donor retention program. So far, researches on blood donation, especially among the Chinese community, are very limited. As attitude on blood donation varies across culture and health care system, researches that focus particularly on the Chinese community are necessary and important. For example, the maximum number of

donations for male and female in Hong Kong is three and four per year respectively whereas in the US, they can donate every 8 weeks. Among the limited studies in blood donation, Ownby *et al.* (50) examined the characteristics of first time donors and analyzed the return times by Cox proportional hazards regression. It was shown that the donation patterns were related to the donation interval between the first two donations, age and education level of donors. This finding indicated that the inter-donation time between the first two donations and the demographic statistics were useful predictors.

In addition, Schreiber *et al.* (53) pointed out that first year donation frequency was significantly and positively correlated with long-term donor return even when demographic variables were adjusted. Furthermore, Yu *et al.* (63) used cluster analysis to classify donors according to their donation patterns and then developed classification rules for the classification of donors using decision tree. Many of these studies, for example the Yu *et al.* (63) study, attempted to relate donation pattern to covariates. However, none of them modeled donation pattern as well as its relationship with covariates simultaneously in one model. Hence, they fell short of controlling for the trend effects in different donation patterns. Ideally, both covariate effects and trend effects should be accounted for in one single model.

Therefore, we propose the MAPGP model on donation frequency that identifies different donation patterns and classifies donors according to their donation patterns. At the same time, by incorporating some personal characteristics of the

donors as covariate effects in the model, we can study the relationship of different characteristics with the donation frequency. This information together with the estimated group memberships help to identify the characteristics of each donor group giving valuable insights to health care organizations in designing different donor promotion strategies and retention programs for the Chinese community.

6.3.2.2 The data

The blood donation data contains data of the whole blood donors, aged 18 or above when their first time blood donation occurred between January 2000 and May 2001 in Hong Kong. The data set were extracted from the BTS. Donors were followed longitudinally for a minimum of four years. Since people under age 18 can donate blood only twice per year, those donors whose first donation was taken under that age were omitted. As the needs for different blood types are different and the O positive blood is regarded as the universal blood, analyses are done specifically on the O positive blood donors. Donors with other blood types can be analyzed similarly. Moreover, as the maximum number of donation per year for male and female are different, the donation patterns for male and female donors are studied separately. After screening the targeted donors, the studied data set contains 3904 (46%) male and 4561 (54%) female donors. Some important information for each donor include sex, age, donation times and donation venues.

To understand the behavior of donors in blood donation, a practical way is to investigate their blood donation patterns. To describe the pattern, one issue

arises on whether the time between donations or frequency of donations should be used for the outcome variable. Inter-arrival time is often used in survival analysis when the lifetime of certain event is the focus. Whereas, event frequency is often used when the intensity of the occurrence of the event is the focus. To investigate strategies to maintain a stable blood supply in the BTS, we choose the annual donation frequency (ADF) during a surveillance period rather than the inter-donation time to be the outcome variable. We believe that the loss of information in converting donation times to ADFs is light because the time interval for the frequency count, one year in this analysis, is sufficiently short. We write W_{it} to denote the ADF for donor i during the t^{th} year. In addition, the ADF is easier to interpret and model than the inter-donation time in survival analysis. Moreover, the MAPGP model on ADF can identify donation patterns more efficiently as the ADF reveals donation patterns more directly than the inter-donation time. The application of the MAPGP model to study blood donation patterns is very unique in the field because there are few studies to investigate donation patterns for the first-time donors using donation frequency (Wu and Schreiber (62), Schreiber *et al.* (53)).

Indeed, the personal characteristics of donors help to explain the donation pattern and thus the donor's behavior in blood donation. From the experience of the American blood donation studies, variables such as occupation, education level and transfusion history are significant in forecasting the donors' return behavior (Wu and Schreiber (62)). However, these information are not provided in

the present database. Instead, we consider two covariates in the MAPGP model. The first one is the donor's age during the first donation $z_{\mu 1it}$, which is a common explanatory variable. It reveals, in this data, the willingness to give blood more frequently as experience shows that young donors are more likely to be regular donors. Table 6.2 shows the age distribution of the male and female donors and it shows that most donors come from the age group 20-29 with 48.4% for male and 41.6% for female. The second covariate is the "venue for donation" and it indicates donors' intention to donate blood. We set $z_{\mu 2it} = 1$ to indicate active (A) donors if they donate blood in the health care related centers including the blood donation centers. Otherwise, we set $z_{\mu 2it} = 2$ to indicate (P) passive donors if they donate blood in other venues including donation vehicles and sponsor sites such as high schools, colleges, universities, corporates, police stations, fire stations, government offices, custom duty offices, churches and places for societies and organizations. If a donor donates blood in both types of donation venue, the type of venue which is visited more frequently will be taken to represent his/her donation venue. Note that $z_{\mu kit}$ is constant over t. Therefore, we simply write $z_{\mu kit} = z_{\mu ki}, k = 1, 2$ for donor i. The percentage of active donors are close to half for both male and female donors. Table 6.2 shows that active donors have higher ADFs.

==========

Table 6.2 about here

==========

6.3.2.3 Model fitting

To analyze the blood donation data, we fit separately the whole set of male and female O positive donors instead of a random training sample of donors for each MAPGP model as they will give similar results if the sample size of the training sample is large enough. For model fitting, we first consider the two-group MPGP ($MPGP_2$) models using both original and simplified versions by setting the mean function $\mu_{itl} = \exp(\beta_{\mu 0l})$ and the ratio function $a_{itl} = a_{0l}$ respectively. Then, we incorporate the two covariate effects: age $z_{\mu 1i}$ and intention $z_{\mu 2i}$ into the mean function μ_{itl} as given by (6.3) while the ratio function is still given by $a_{itl} = a_{0l}$ or by (6.4) in the two-group MAPGP ($MAPGP_2$) models. By incorporating the time covariate $z_{a1it} = t$ in the ratio function, the $MAPGP_2$ model can describe a wider variety of donation patterns among the donors.

Unlike the bladder cancer study, we adopt the EM method and the Bayesian method for model implementation because the classical ML method fails to estimate the model parameters when $l \geq 3$ for the reason stated in Section 6.2.1. For the model assessment, we again use the MSE in (6.13) and the U in (6.14) as the GOF measures.

Comparing the two covariate effects in the mean function, we find that the age effect is insignificant. Therefore, we only fit the MAPGP model with intention $z_{\mu 2i}$ as the only covariate effect, that is $\mu_{itl} = \exp(\beta_{\mu 0l} + \beta_{\mu 1l} z_{\mu 2i})$. After refitting the $MAPGP_2$ model, we observe that the simplified $MPGP_2$ and $MAPGP_2$ models always perform better than the original models according to both MSE

and U. Thus, we further fit the three-group simplified:

1. MPGP ($MPGP_3$) model ($\mu_{itl} = \exp(\beta_{\mu 0l})$; $a_{itl} = a_{0l}$);
2. MAPGP ($MAPGP_{3a}$) model ($\mu_{itl} = \exp(\beta_{\mu 0l} + \beta_{\mu 1l} z_{\mu 2i})$; $a_{itl} = a_{0l}$);
3. MAPGP ($MAPGP_{3b}$) model ($\mu_{itl} = \exp(\beta_{\mu 0l})$; $a_{it1} = \exp(\beta_{a01} + \beta_{a11} t)$; $a_{itl} = a_{0l}, l = 2, 3$) and
4. MAPGP ($MAPGP_{3c}$) model ($\mu_{itl} = \exp(\beta_{\mu 0l} + \beta_{\mu 1l} z_{\mu 2i})$; $a_{it1} = \exp(\beta_{a01} + \beta_{a11} t)$; $a_{itl} = a_{0l}, l = 2, 3$).

using the EM and the Bayesian methods.

It should be noted that the MAPGP models incorporate $z_{a1it} = t$ in the ratio function only for the female donors in group 1. For other groups of female donors and all groups of male donors, the parameter β_{a1l} is not significant and hence the ratio function (6.4) is not considered. In other words, the $MAPGP_{3b}$ and $MAPGP_{3c}$ models are not adopted for female donors in group 2 and 3 as well as male donors. Table 6.3 shows that for female donors, the best model according to both of U and MSE is the simplified $MAPGP_{3c}$ model using the Bayesian approach. Whereas for male donors, Table 6.4 indicates that the best model according to the two GOF measures is the simplified $MAPGP_{3a}$ model using the Bayesian approach too.

==========

Tables 6.3 and 6.4 about here

==========

6.3.2.4 Interpretation of best model

Based on the best models, Figure 6.2 for female donors and Figure 6.3 for male donors plot the observed ADF w_{tlv} and expected ADF $\hat{w}_{tlv}$ averaged over donor i, as given by

$$w_{tlv} = \frac{\sum_{i=1}^{m} \hat{I}_{b,il}\, I(z_{\mu 2i} = v)\, w_{it}}{\sum_{i=1}^{m} \hat{I}_{b,il}\, I(z_{\mu 2i} = v)} \quad \text{and} \quad \hat{w}_{tlv} = \frac{\sum_{i=1}^{m} \hat{I}_{b,il}\, I(z_{\mu 2i} = v)\, \frac{\mu_{itl}}{a_{itl}^{t-1}}}{\sum_{i=1}^{m} \hat{I}_{b,il}\, I(z_{\mu 2i} = v)}$$

across the time $t = 1, \ldots, 4$ for the active ($v = 1$) and passive ($v = 2$) donors in groups $l = 1, 2, 3$. From the two figures, w_{tlv} and $\hat{w}_{tlv}$ are closed in values, showing that the model fits are good. Moreover, the two figures also reveal different donation patterns across group l: the ADF always drops ($a_{0l} > 1$) from an initial level which decreases across group l ($\beta_{\mu 0l}$ decreases across l) and its rate of drop increases across group l (a_{0l} increases across l) except for the female donors in group 1.

Clearly, group 1 consists of committed (C) donors who have much higher levels of ADF. For the female donors, the ADF drops at $t = 1$ and rises up gradually at $t = 2, 3$ ($\beta_{a11} < 0$ and is significant). For male donors, the ADF drops at $t = 1, 3$ and rises up at $t = 2$ showing no clear rebound trend for the ADF. Donors in group 2 are dropout (D) donors since the ADF drops from a lower level slowly and continuously. Lastly, the ADF for donors in group 3 drops rapidly from one to zero at $t = 2$. Hence, they are called one-time (O) donors. Figures 6.2 and 6.3 show that the ADFs for the D and O donors are similar in level at $t = 1$ but the ADF for the O donors drops straight to zero at $t = 2$. Moreover, the

negative $\beta_{\mu 1l}$ shows that the active donors have higher levels of ADF but such difference reduces across the C, D and O groups. Male donors also have higher levels of ADF (higher $\beta_{\mu 0l}$) than female donors in general. One possible reason to explain this phenomenon is that male donors can donate blood more frequently than female donors within a year. All the model parameters, for female and male donors, are significant.

==========

Figures 6.2 and 6.3 about here

==========

We also study the model fit of the best model by considering the observed proportions p_{tluv} and the expected probabilities $f_{s,itl}(u|z_{\mu 2i} = v)$ in (6.2). Figures 6.4, 6.5 and 6.6 plot, when $t = 1, \ldots, 4$ and $v = 1, 2$, the observed proportions

$$p_{tluv} = \frac{\sum_{i=1}^{m} \hat{I}_{b,il}\, I(z_{\mu 2i} = v)\, I(w_{it} = u)}{\sum_{i=1}^{m} \hat{I}_{b,il}\, I(z_{\mu 2i} = v)}$$

and the expected probabilities $f_{s,itl}(u|z_{\mu 2i} = v)$ for each female donor group $l, l = 1, 2, 3$ across different levels of the observed ADF $u = 0, 1, \ldots, 7$ using the best model. Most of the expected probabilities $f_{s,itl}(u|z_{\mu 2i} = v)$ are close to the observed proportions p_{tluv} except for $t = 1$ in the O group.

==========

Figures 6.4, 6.5 and 6.6 about here

==========

Obviously, the MPGP and MAPGP models allow W_{i1} to be zero but actually

$W_{i1} \neq 0$. To deal with this problem, we may set W_{i1} to be 1 less than the observed ADF so that W_{i1} indicates the ADF after the first donation. However, the first donation period will then be less than a year because the female and male donors will have to wait for at least 4 and 3 weeks respectively for the second donation. Alternatively we may consider a truncated Poisson distribution for W_{i1}. Then the MPGP and MAPGP models should adopt a different set of parameters for $t = 1$, leading to a substantial increase in the number of parameters and hence complicating the model considerably. As a result, no adjustment is made to account for the non-zero W_{i1}. Although the model-fit may be worsened, Figures 6.4 to 6.6 show that the model parameters can be adjusted to give a lower probability for $W_{i1} = 0$. The only exception is for the O donors ($l = 3$) because when $t = 1$, the observed proportions, $p_{130v} = 0$, $p_{131v} \simeq 1$ and $p_{132v} \simeq 0$, changes too rapidly across $u = 0, 1, 2$ for a satisfactory model fit. For male donors, the observed proportions p_{tluv} and the expected probabilities $f_{s,itl}(u|z_{\mu 2i} = v)$ are close to those of female donors and hence the plots of p_{tluv} and $f_{s,itl}(u|z_{\mu 2i} = v)$ are omitted.

6.3.2.5 Classification

A donor i is classified into group g_i if $\max_l(\hat{I}_{b,il}) = \hat{I}_{b,ig_i}$ where the group membership indicator $\hat{I}_{b,il}$ is the posterior mean of I_{sil} in (6.11) using the Bayesian method. Another set of group membership indicators $\hat{I}_{e,il} = \hat{I}_{sil}$ in (6.8) using the EM method gives another set of classification $g_{e,i}$ of which 98% (97% for the male

donors) agree with g_i. Of the 4561 female donors, 420 (9%), 1233 (27%) and 2908 (64%) are classified into the C, D and O groups respectively, whereas among the 3901 male donors, the corresponding counts for the C, D and O groups are 254 (7%), 1008 (26%) and 2642 (68%) respectively.

We further classify the female donors into the "clear" group ($q_i = 1$) if $\hat{I}_{b,ig_i} \geq 0.8$ (0.85 for male donors) and into the "unclear" group ($q_i = 0$) if otherwise. For female donors, about 41%, 25% and 57% of the C, D and O donors show some partial group memberships in the D, C and D groups respectively and so they are called the CD, DC and OD donors accordingly. The corresponding percentages for male donors are 26%, 8% and 40% respectively. Figures 6.7 and 6.9 show the cell frequency, the predicted total ADF, the observed total ADF and the observed total ADF in the first 2 years which are defined as

$$n_{lcv} = \sum_{i=1}^{m} I(g_i = l)\, I(q_i = c)\, I(z_{\mu 2i} = v), \tag{6.15}$$

$$\widehat{w}_{lcv} = \frac{1}{n_{lcv}} \sum_{i=1}^{m} \sum_{t=1}^{4} I(g_i = l)\, I(q_i = c)\, I(z_{\mu 2i} = v)\, \frac{\mu_{itl}}{a_{itl}^{t-1}}, \tag{6.16}$$

$$w_{lcv} = \frac{1}{n_{lcv}} \sum_{i=1}^{m} \sum_{t=1}^{4} I(g_i = l)\, I(q_i = c)\, I(z_{\mu 2i} = v)\, w_{it}, \tag{6.17}$$

$$w^*_{lcv} = \frac{1}{n_{lcv}} \sum_{i=1}^{m} \sum_{t=1}^{2} I(g_i = l)\, I(q_i = c)\, I(z_{\mu 2i} = v)\, w_{it}. \tag{6.18}$$

respectively where $l = 1, 2, 3$, $c = 0, 1$ and $v = 1, 2$. The two figures show that w_{lcv} and w^*_{lcv} decrease across the donor groups C, CD, DC, D, OD and O for both of the male and female donors. Within each donor group, the active donors have higher levels of ADF than those of the passive donors. In general,

the observed total ADF w_{lcv} agrees with the predicted total ADF $\hat{w}_{lcv}$ for the classes C, D and O cross-classified with donor type (active and passive) except that $\hat{w}_{lcv} < w_{lcv}$ consistently for the D donors. This is because many D donors show partial membership in the C group which has a much higher level of ADF leading to a larger observed w_{lcv}.

==========

Figures 6.7 and 6.9 about here

==========

6.3.2.6 Prediction

As donors need time to develop blood donation as a habit, the longer the observed period, the more accurate is the prediction on donors' donation patterns. On the other hand, the observed period for prediction should be as short as possible and less than the surveillance period of four years for management efficiency. As the D and O donors have similar W_{i1} but W_{i2} drops straight to zero among the O donors, it is necessary to set the observed period for prediction to beyond one year. Therefore, we set it to be two years.

The membership indicator for donor i in group l based only on the ADF of the first two years, W_{it}, $t = 1, 2$, is

$$\hat{I}^*_{b,il} = E_s(I_{sil} |\ \boldsymbol{\theta}_s, \boldsymbol{w}_i) = \left(\pi_l \prod_{t=1}^{2} L_{s,itl} \right) / \left[\sum_{l'=1}^{G} \pi_{l'} \left(\prod_{t=1}^{2} L_{s,itl'} \right) \right] \tag{6.19}$$

where $L_{s,itl}$ in (6.2) is evaluated at the vector of parameter estimates $\hat{\boldsymbol{\theta}}_s$. Then a

donor i is predicted to be classified into group $g^*_{b,i}$ if $\max_l(\hat{I}^*_{b,il}) = \hat{I}^*_{b,ig^*_{b,i}}$. The new groups are denoted by C*, D* and O* respectively for the committed, drop-out and one-time groups.

Table 6.5 shows that the predictive performance of $g^*_{b,i}$ based on W_{it}, $t = 1, 2$ is satisfactory: at least 74% of the female donors (84% for the male donors) who are classified into donor groups C*, D* and O* actually come from groups C, D and O respectively, that is $g_i = g^*_{b,i}$. However, the classification performance of g_i based on W_{it}, $t = 1, \ldots, 4$ is less satisfactory for the C and D donors: only 38% and 59% of the female C and D donors belong to the C* and D* groups respectively. For male donors, the corresponding percentages are 54% and 57% respectively.

==========

Table 6.5 about here

==========

To explain the poorer classification performance of g_i, we look at the average ADF when $t = 1, \ldots, 4$ among the 180 C female donors and the 448 D female donors who are classified into the D* and O* groups respectively and their values are 1.8, 1.29, 1.36, 1.49 and 1.32, 0, 0.68, 0.53 accordingly. It is clear that the ADFs for the C and D donors are lower and closer to those of the D and O female donors respectively when $t = 1, 2$ but they increase after $t = 2$. These explain why these two groups of donors based on W_{it}, $t = 1, 2$ is easily misclassified. Similar result holds for the male donors. There are 93 C male donors and 406

D male donors who are classified into the D* and O* groups respectively. Their ADFs when $t = 1, \ldots, 4$ are 2.26, 1.52, 1.95, 2.01 and 1.38, 0.23, 0.17, 0.16 correspondingly. Regarding our aim to identify the D donors who have the potential to become C donors and the C donors who can donate more frequently based on W_{it}, $t = 1, 2$, the predictive performance of $g^*_{b,i}$ is more important.

We further classify the female donors into the "clear" group ($q^*_i = 1$) if $\hat{I}^*_{b,ig^*_i} \geq 0.7$ (0.8 for male donors) and into the "unclear" group ($q^*_i = 0$) if otherwise. Figures 6.8 and 6.10 show the classification of $g^*_{b,i}$ into different donor groups where n_{ljv}, w_{ljv} and w^*_{ljv} are calculated using (6.15) to (6.18) and $g^*_{b,i}$. The percentage of correct classification p^*_{lcv} is defined as

$$p^*_{lcv} = \frac{\sum_{i=1}^{m} I(g_i = l)\, I(g^*_{b,i} = l)\, I(q^*_i = c)\, I(z_{\mu 2i} = v)}{\sum_{i=1}^{m} I(g^*_{b,i} = l)\, I(q^*_i = c)\, I(z_{\mu 2i} = v)}.$$

Again the two figures show that w_{lcv} and w^*_{lcv} decrease across the donor groups C*, CD*, DC*, D*, OD* and O* for both of the male and female donors but they are higher than the corresponding values in Figures 6.7 and 6.9 based on g_i. Moreover, the counts n_{lcv}, $l = 1, 2$ are lower than those in Figures 6.7 and 6.9. The reason is obvious. Observing only W_{i1} and W_{i2}, a donor needs to have a higher level of W_{it} in order to be classified into the C* or D* groups since unclear group memberships ($\hat{I}^*_{b,ig^*_i} < 0.7$ and 0.8 for female and male donors respectively) are more common. As a result, the classification based on $g^*_{b,i}$ identifies the C*, CD*, DC*, D* donors with higher levels of ADF.

==========

Figures 6.8 and 6.10 about here

==========

For female donors, Figure 6.8 show that our target donors will be those (1) 197 CD* donors (active or passive) who donate 2.5 times during the third and fourth years, (2) 816 DC*, D* and OD* active donors who donate 0.8 times and (3) 393 DC* and D* passive donors who donate 0.8 times. They account for 31% of all female donors but they give 1448 donations, 68% of all donations during the third and fourth year. If blood donor retention program is successful in retaining them during the fifth year, they will donate further 2.1, 0.1 and 0.1 times giving a total of 473 donations, as predicted using the best model with $t = 5$. Similar results for male donors are given in Table 6.6. Hence, by assuming that the donation patterns continue to the fifth year, future donations from target donors (female and male) during year 3-4 and 3-5 will account for 14.5% and 15.7% respectively of all donations ($t = 1, \ldots, 5$).

==========

Table 6.6 about here

==========

After all, because of the high ownership and usage of mobile phones and computers in Hong Kong, promotional campaigns, including sending email and Short Message Service (SMS) reminders to these target donors, are surely some useful strategies to increase their donation frequencies especially those D donors.

Also, it is more effective to encourage longer term commitment to blood donation than to recruiting new donors. Nevertheless, as many personal information and characteristics of the donors are unavailable in the blood donation data, further prospective studies with information such as occupation, education level, and previous transfusion history should help to build a better model and to validate the current donor retention strategies.

Appendix 6.1 Derivation of first and second order derivative functions in the NR iterative method for the MAPGP models using classical ML method

The first and second order derivative functions for the MAPGP model in the NR iterative method are, in general, given by:

$$\begin{aligned}
\frac{\partial \ell_h}{\partial \theta_{jkl}} &= \sum_{i=1}^{m} \frac{\pi_l L_{h,il}}{L_{h,i}} \left(\sum_{t=1}^{n_i} \frac{\partial \ln L_{h,itl}}{\partial \theta_{jkl}} \right) \\
\frac{\partial^2 \ell_h}{\partial \theta_{jkl}^2} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{h,il}}{L_{h,i}} \left(\sum_{t=1}^{n_i} \frac{\partial^2 \ln L_{h,itl}}{\partial \theta_{jkl}^2} \right) + \frac{\pi_l L_{h,il}}{L_{h,i}} \left(\sum_{t=1}^{n_i} \frac{\partial \ln L_{h,itl}}{\partial \theta_{jkl}} \right)^2 - \right. \\
&\qquad \left. \left[\frac{\pi_l L_{h,il}}{L_{h,i}} \left(\sum_{t=1}^{n_i} \frac{\partial \ln L_{h,itl}}{\partial \theta_{jkl}} \right) \right]^2 \right\}
\end{aligned}$$

where

$$\ell_h = \sum_{i=1}^{m} \ln \left(\pi_1 \prod_{t=1}^{n_i} L_{h,it1} + \pi_2 \prod_{t=1}^{n_i} L_{h,it2} \right) = \sum_{i=1}^{m} \ln(\pi_1 L_{h,i1} + \pi_2 L_{h,i2}) = \sum_{i=1}^{m} \ln L_{h,i},$$

$$L_{h,il} = \prod_{t=1}^{n_i} L_{h,itl} \quad \text{and} \quad L_{h,i} = \pi_1 L_{h,i1} + \pi_2 L_{h,i2},$$

$L_{h,itl}$ is given by (6.1) and (6.2) for the original and simplified versions of the MAPGP models. θ_{jkl} represents the parameters for group l. In particular, we obtain the following derivative functions for the original MAPGP model:

$$
\begin{aligned}
\frac{\partial \ell_o}{\partial a_l} &= \sum_{i=1}^{m} \frac{\pi_l L_{o,\,il}}{L_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z^*_{atl}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \\
\frac{\partial \ell_o}{\partial \beta_{\mu k\, l}} &= -\sum_{i=1}^{m} \frac{\pi_l L_{o,\,il}}{L_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z_{\mu kit}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \\
\frac{\partial \ell_o}{\partial \pi_1} &= \sum_{i=1}^{m} \frac{L_{o,\,i1} - L_{o,\,i2}}{L_{o,\,i}} \\
\frac{\partial^2 \ell_o}{\partial a_l^2} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{o,\,il}}{L_{o,\,i}} \left[-\sum_{t=1}^{n_i} \left(\frac{z^*_{atl}}{a_l} \right) \frac{(\hat{w}_{itl} + 1)(\hat{w}_{itl} - w_{it}) + (t-1)\hat{w}_{itl}(w_{it} + 1)}{(\hat{w}_{itl} + 1)^2} \right] + \right. \\
&\qquad \left. \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L^2_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z^*_{atl}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right]^2 \right\} \\
\frac{\partial^2 \ell_o}{\partial \beta^2_{\mu kl}} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{o,\,il}}{L_{o,\,i}} \left[-\sum_{t=1}^{n_i} \frac{z^2_{\mu kit}\hat{w}_{itl}(w_{it} + 1)}{(\hat{w}_{itl} + 1)^2} \right] + \right. \\
&\qquad \left. \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L^2_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z_{\mu kit}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right]^2 \right\} \\
\frac{\partial^2 \ell_o}{\partial a_l \partial \beta_{\mu kl}} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{o,\,il}}{L_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z^*_{atl} z_{\mu kit}\hat{w}_{itl}(w_{it} + 1)}{(\hat{w}_{itl} + 1)^2} \right] - \right. \\
&\qquad \left. \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L^2_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z^*_{atl}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \left[\sum_{t=1}^{n_i} \frac{z_{\mu kit}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \right\} \\
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k_1 l} \partial \beta_{\mu k_2 l}} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{o,\,il}}{L_{o,\,i}} \left[-\sum_{t=1}^{n_i} \frac{z_{\mu k_1 it} z_{\mu k_2 it}\hat{w}_{itl}(w_{it} + 1)}{(\hat{w}_{itl} + 1)^2} \right] + \right. \\
&\qquad \left. \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L^2_{o,\,i}} \left[\sum_{t=1}^{n_i} \frac{z_{\mu k_1 it}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \left[\sum_{t=1}^{n_i} \frac{z_{\mu k_2 it}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \right\} \\
\frac{\partial^2 \ell_o}{\partial a_1 \partial a_2} &= -\sum_{i=1}^{m} \left\{ \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L_{o,\,i^2}} \left[\sum_{t=1}^{n_i} \frac{z^*_{at1}(\hat{w}_{it1} - w_{it})}{\hat{w}_{it1} + 1} \right] \left[\sum_{t=1}^{n_i} \frac{z^*_{at2}(\hat{w}_{it2} - w_{it})}{\hat{w}_{it2} + 1} \right] \right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k_1 1} \partial \beta_{\mu k_2 2}} &= -\sum_{i=1}^{m} \left\{ \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L_{o,\,i}^2} \left[\sum_{t=1}^{n_i} \frac{z_{\mu k_1 it}(\hat{w}_{it1} - w_{it})}{\hat{w}_{it1} + 1} \right] \left[\sum_{t=1}^{n_i} \frac{z_{\mu k_2 it}(\hat{w}_{it2} - w_{it})}{\hat{w}_{it2} + 1} \right] \right\} \\
\frac{\partial^2 \ell_o}{\partial a_{l_1} \partial \beta_{\mu k l_2}} &= \sum_{i=1}^{m} \left\{ \frac{\pi_1 L_{o,\,i1} \pi_2 L_{o,\,i2}}{L_{o,\,i}^2} \left[\sum_{t=1}^{n_i} \frac{z^*_{atl_1}(\hat{w}_{itl_1} - w_{it})}{\hat{w}_{itl_1} + 1} \right] \left[\sum_{t=1}^{n_i} \frac{z_{\mu k i t}(\hat{w}_{itl_2} - w_{it})}{\hat{w}_{itl_2} + 1} \right] \right\} \\
\frac{\partial^2 \ell_o}{\partial a_l \partial \pi_1} &= \sum_{i=1}^{m} I_\pi(l) \frac{L_{o,\,i1} L_{o,\,i2}}{L_{o,\,i}^2} \left[\sum_{t=1}^{n_i} \frac{z^*_{atl}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \\
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k l} \partial \pi_1} &= -\sum_{i=1}^{m} I_\pi(l) \frac{L_{o,\,i1} L_{o,\,i2}}{L_{o,\,i}^2} \left[\sum_{t=1}^{n_i} \frac{z_{\mu k i t}(\hat{w}_{itl} - w_{it})}{\hat{w}_{itl} + 1} \right] \\
\frac{\partial^2 \ell_o}{\partial \pi_1^2} &= -\sum_{i=1}^{m} \frac{(L_{o,\,i1} - L_{o,\,i2})^2}{L_{o,\,i}^2}
\end{aligned}
$$

where $\hat{w}_{itl} = \dfrac{\mu_{itl}}{a_l^{t-1}}$, μ_{itl} is given by (6.3), $z^*_{atl} = \dfrac{t-1}{a_l}$, and $I_\pi(l) = \begin{cases} 1 & l = 1, \\ -1 & l = 2 \end{cases}$.

For the simplified MAPGP model, the derivative functions are given by:

$$
\begin{aligned}
\frac{\partial \ell_s}{\partial a_l} &= \sum_{i=1}^{m} \frac{\pi_l L_{s,\,il}}{L_{s,\,i}} \left[\sum_{t=1}^{n_i} z^*_{atl}(\hat{w}_{itl} - w_{it}) \right] \\
\frac{\partial \ell_s}{\partial \beta_{\mu k\, l}} &= -\sum_{i=1}^{m} \frac{\pi_l L_{s,\,il}}{L_{s,\,i}} \left[\sum_{t=1}^{n_i} z_{\mu k i t}(\hat{w}_{itl} - w_{it}) \right] \\
\frac{\partial \ell_s}{\partial \pi_1} &= \sum_{i=1}^{m} \frac{L_{s,\,i1} - L_{s,\,i2}}{L_{s,\,i}} \\
\frac{\partial^2 \ell_s}{\partial a_l^2} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{s,\,il}}{L_{s,\,i}} \left[-\sum_{t=1}^{n_i} \frac{z^*_{atl}(t\hat{w}_{itl} - w_{it})}{a_l} \right] + \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L_{s,\,i}^2} \left[\sum_{t=1}^{n_i} z^*_{atl}(\hat{w}_{itl} - w_{it}) \right]^2 \right\} \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k l}^2} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{s,\,il}}{L_{s,\,i}} \left(-\sum_{t=1}^{n_i} z^2_{\mu k i t} \hat{w}_{itl} \right) + \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L_{s,\,i}^2} \left[\sum_{t=1}^{n_i} z_{\mu k i t}(\hat{w}_{itl} - w_{it}) \right]^2 \right\}
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell_s}{\partial a_l \partial \beta_{\mu k l}} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{s,\,il}}{L_{s,\,i}} \left(\sum_{t=1}^{n_i} z^*_{atl} z_{\mu k i t} \hat{w}_{itl} \right) - \right. \\
&\quad \left. \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L^2_{s,\,i}} \left[\sum_{t=1}^{n_i} z^*_{atl} (\hat{w}_{itl} - w_{it}) \right] \left[\sum_{t=1}^{n_i} z_{\mu k i t} (\hat{w}_{itl} - w_{it}) \right] \right\} \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k_1 l} \partial \beta_{\mu k_2 l}} &= \sum_{i=1}^{m} \left\{ \frac{\pi_l L_{s,\,il}}{L_{s,\,i}} \left[- \sum_{t=1}^{n_i} z_{\mu k_1 i t} z_{\mu k_2 i t} \hat{w}_{itl} \right] + \right. \\
&\quad \left. \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L^2_{s,\,i}} \left[\sum_{t=1}^{n_i} z_{\mu k_1 i t} (\hat{w}_{itl} - w_{it}) \right] \left[\sum_{t=1}^{n_i} z_{\mu k_2 i t} (\hat{w}_{itl} - w_{it}) \right] \right\} \\
\frac{\partial^2 \ell_s}{\partial a_1 \partial a_2} &= - \sum_{i=1}^{m} \left\{ \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L_{s,\,i^2}} \left[\sum_{t=1}^{n_i} z^*_{at1} (\hat{w}_{it1} - w_{it}) \right] \left[\sum_{t=1}^{n_i} z^*_{at2} (\hat{w}_{it2} - w_{it}) \right] \right\} \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k_1 1} \partial \beta_{\mu k_2 2}} &= - \sum_{i=1}^{m} \left\{ \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L^2_{s,\,i}} \left[\sum_{t=1}^{n_i} z_{\mu k_1 i t} (\hat{w}_{it1} - w_{it}) \right] \left[\sum_{t=1}^{n_i} z_{\mu k_2 i t} (\hat{w}_{it2} - w_{it}) \right] \right\} \\
\frac{\partial^2 \ell_s}{\partial a_{l_1} \partial \beta_{\mu k l_2}} &= - \sum_{i=1}^{m} \left\{ \frac{\pi_1 L_{s,\,i1} \pi_2 L_{s,\,i2}}{L^2_{s,\,i}} \left[\sum_{t=1}^{n_i} z^*_{atl_1} (\hat{w}_{itl_1} - w_{it}) \right] \left[\sum_{t=1}^{n_i} z_{\mu k i t} (\hat{w}_{itl_2} - w_{it}) \right] \right\} \\
\frac{\partial^2 \ell_s}{\partial a_l \partial \pi_1} &= \sum_{i=1}^{m} I_\pi(l) \frac{L_{s,\,i1} L_{s,\,i2}}{L^2_{s,\,i}} \left[\sum_{t=1}^{n_i} z^*_{atl} (\hat{w}_{itl} - w_{it}) \right] \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k l} \partial \pi_1} &= - \sum_{i=1}^{m} I_\pi(l) \frac{L_{s,\,i1} L_{s,\,i2}}{L^2_{s,\,i}} \left[\sum_{t=1}^{n_i} z_{\mu k i t} (\hat{w}_{itl} - w_{it}) \right] \\
\frac{\partial^2 \ell_s}{\partial \pi_1^2} &= - \sum_{i=1}^{m} \frac{(L_{s,\,i1} - L_{s,\,i2})^2}{L^2_{s,\,i}}
\end{aligned}
$$

Appendix 6.2 Derivation of first and second order derivative functions in the M-step for the MAPGP models using EM method

For the original MAPGP model, the derivative functions are given by:

$$\begin{aligned}
\frac{\partial \ell_o}{\partial a_l} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{oil}\left[\frac{z^*_{atl}(\hat{w}_{itl}-w_{it})}{\hat{w}_{itl}+1}\right] \\
\frac{\partial \ell_o}{\partial \beta_{\mu kl}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{oil}\left[\frac{z_{\mu kit}(\hat{w}_{itl}-w_{it})}{\hat{w}_{itl}+1}\right] \\
\frac{\partial \ell_o}{\partial \pi_l} &= \frac{1}{\pi_l}\sum_{i=1}^{m} I_{oil} \\
\frac{\partial^2 \ell_o}{\partial a_l^2} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{oil}\left[\left(\frac{z^*_{atl}}{a_l}\right)\frac{(\hat{w}_{itl}+1)(\hat{w}_{itl}-w_{it})+(t-1)\hat{w}_{itl}(w_{it}+1)}{(\hat{w}_{itl}+1)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k_1 l}\partial \beta_{\mu k_2 l}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{oil}\left[\frac{z_{\mu k_1 it}z_{\mu k_2 it}\hat{w}_{itl}(w_{it}+1)}{(\hat{w}_{itl}+1)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial a_l \partial \beta_{\mu kl}} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{oil}\left[\frac{z^*_{atl}z_{\mu kit}\hat{w}_{itl}(w_{it}+1)}{(\hat{w}_{itl}+1)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial \pi_l^2} &= -\frac{1}{\pi_l^2}\sum_{i=1}^{m} I_{oil} \\
\frac{\partial^2 \ell_o}{\partial a_{l_1}\partial a_{l_2}} &= \frac{\partial^2 \ell_o}{\partial a_{l_1}\partial \beta_{\mu k l_2}} = \frac{\partial^2 \ell_o}{\partial \pi_{l_1}\partial \pi_{l_2}} = \frac{\partial^2 \ell_o}{\partial a_l \partial \pi_l} = \frac{\partial^2 \ell_o}{\partial \beta_{\mu kl}\partial \pi_l} = \frac{\partial^2 \ell_o}{\partial a_{l_1}\partial \pi_{l_2}} = \frac{\partial^2 \ell_o}{\partial \beta_{\mu k l_1}\partial \pi_{l_2}} = 0
\end{aligned}$$

where $z^*_{atl} = \dfrac{t-1}{a_l}$ and $\hat{w}_{itl} = \dfrac{\mu_{itl}}{a_l^{t-1}}$.

For the simplified MAPGP model, the derivative functions are given by:

$$\begin{aligned}
\frac{\partial \ell_s}{\partial a_l} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{sil}\left[z^*_{atl}(\hat{w}_{itl} - w_{it})\right] \\
\frac{\partial \ell_s}{\partial \beta_{\mu kl}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{sil}\left[z_{\mu kit}(\hat{w}_{itl} - w_{it})\right] \\
\frac{\partial \ell_s}{\partial \pi_l} &= \frac{1}{\pi_l}\sum_{i=1}^{m} I_{sil} \\
\frac{\partial^2 \ell_s}{\partial a_l^2} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{sil}\left[\frac{z^*_{atl}}{a_l}(t\hat{w}_{itl} - w_{it})\right] \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k_1 l}\partial \beta_{\mu k_2 l}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{sil}\left[z_{\mu k_1 it} z_{\mu k_2 it}\hat{w}_{itl}\right] \\
\frac{\partial^2 \ell_s}{\partial a_l \partial \beta_{\mu kl}} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{sil}\left[z^*_{atl} z_{\mu kit}\hat{w}_{itl}\right] \\
\frac{\partial^2 \ell_s}{\partial \pi_l^2} &= -\frac{1}{\pi_l^2}\sum_{i=1}^{m} I_{sil} \\
\frac{\partial^2 \ell_s}{\partial a_{l_1}\partial a_{l_2}} &= \frac{\partial^2 \ell_s}{\partial a_{l_1}\partial \beta_{\mu k l_2}} = \frac{\partial^2 \ell_s}{\partial \pi_{l_1}\partial \pi_{l_2}} = \frac{\partial^2 \ell_s}{\partial a_l \partial \pi_l} = \frac{\partial^2 \ell_s}{\partial \beta_{\mu kl}\partial \pi_l} = \frac{\partial^2 \ell_s}{\partial a_{l_1}\partial \pi_{l_2}} = \frac{\partial^2 \ell_s}{\partial \beta_{\mu k l_1}\partial \pi_{l_2}} = 0
\end{aligned}$$

Appendix 6.3 Accumulated number of tumours found in 82 patients taking placebo and thiotepa treatments in the bladder cancer study

Placebo:

	Visit (months)											
Patient	3	6	9	12	15	18	21	24	27	30	33	36
1	0	-	-	-	-	-	-	-	-	-	-	-
2	-	0	-	-	-	-	-	-	-	-	-	-
3	0	-	0	-	-	-	-	-	-	-	-	-
4	0	1	1	-	-	-	-	-	-	-	-	-
5	0	-	0	-	0	-	-	-	-	-	-	-
6	0	-	0	2	5	5	-	-	-	-	-	-
7	0	-	-	-	0	0	-	-	-	-	-	-
8	-	2	2	2	-	2	-	-	-	-	-	-
9	0	0	6	-	9	9	-	9	-	-	-	-
10	8	-	8	8	16	16	-	24	-	-	-	-
11	1	1	2	2	2	2	10	10	-	-	-	-
12	0	0	0	0	0	0	-	0	-	-	-	-
13	0	8	15	15	20	-	-	27	-	-	-	-
14	1	1	1	1	2	2	2	5	-	-	-	-
15	8	8	-	8	8	8	8	8	8	-	-	-
16	4	4	-	-	-	-	-	-	12	-	-	-
17	0	0	-	0	0	-	0	0	0	-	-	-
18	-	0	-	-	-	-	-	3	3	-	-	-
19	-	0	0	-	-	0	-	-	-	0	-	-
20	0	0	-	-	0	-	-	-	-	0	-	-
21	0	0	0	0	-	0	-	0	-	0	-	-
22	4	-	4	-	-	6	10	10	10	10	-	-
23	1	4	7	10	10	10	10	10	13	13	-	-
24	0	0	0	0	-	-	0	-	2	3	-	-
25	0	0	0	2	5	5	-	6	6	6	-	-
26	0	0	0	0	0	0	0	-	0	-	0	-
27	0	-	0	-	0	-	0	-	-	0	0	-
28	0	0	0	0	0	0	0	0	-	0	-	0
29	-	-	0	-	-	-	0	-	-	8	-	8
30	0	-	0	-	-	0	-	0	-	0	-	0
31	0	-	8	-	8	10	15	16	16	16	-	16
32	0	0	0	0	0	9	-	10	10	12	13	13
33	0	0	0	0	0	0	0	0	0	0	-	0
34	3	3	3	-	3	-	3	-	3	3	3	3
35	0	1	1	1	1	1	1	1	-	-	1	1
36	5	8	12	-	12	12	12	12	12	12	12	12
37	0	0	1	4	-	4	5	5	9	12	12	12
38	0	0	0	0	0	1	1	1	1	1	1	1
39	-	0	0	0	-	0	-	0	-	0	-	0
40	0	0	-	0	-	0	-	0	-	0	-	1
41	0	-	0	0	0	1	1	1	1	1	1	1
42	7	-	-	7	9	-	-	-	9	9	-	9
43	0	0	0	0	0	0	0	-	0	-	-	0
44	1	1	1	1	4	4	-	8	8	11	15	15
45	0	3	3	3	7	7	9	9	14	14	14	14
46	1	1	4	10	-	14	15	15	15	15	16	16

Thiotepa:

	Visit (months)											
Patient	3	6	9	12	15	18	21	24	27	30	33	36
47	-	8	-	-	-	-	-	-	-	-	-	-
48	0	0	0	-	-	-	-	-	-	-	-	-
49	-	0	0	-	-	-	-	-	-	-	-	-
50	0	-	-	0	-	-	-	-	-	-	-	-
51	1	1	1	1	1	-	-	-	-	-	-	-
52	7	14	16	16	16	16	-	-	-	-	-	-
53	-	0	-	0	-	2	-	-	-	-	-	-
54	-	-	-	-	-	0	-	-	-	-	-	-
55	2	-	-	-	-	2	-	-	-	-	-	-
56	0	0	0	0	0	2	2	-	-	-	-	-
57	0	0	0	-	0	0	0	-	-	-	-	-
58	0	-	0	0	-	0	0	0	-	-	-	-
59	-	-	0	0	-	0	-	0	-	-	-	-
60	0	0	0	0	0	0	0	0	-	-	-	-
61	0	2	2	5	-	-	6	6	6	-	-	-
62	0	1	1	1	1	-	1	-	1	-	-	-
63	2	-	-	-	-	-	-	-	-	2	-	-
64	0	0	0	-	-	-	0	0	3	3	3	6
65	0	0	0	0	0	0	0	0	0	-	-	-
66	0	1	1	1	4	4	4	7	10	10	18	27
67	0	0	0	0	0	0	2	3	5	-	8	-
68	-	0	-	-	0	0	-	3	5	6	-	-
69	0	0	0	0	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0	0	0	0	0
71	1	1	-	1	-	1	1	1	2	2	2	2
72	-	-	0	-	-	-	-	0	-	-	-	-
73	2	2	2	2	2	2	3	5	6	6	-	-
74	0	0	0	0	0	-	0	0	0	-	0	-
75	1	1	1	1	1	1	1	1	1	-	1	-
76	0	0	0	0	0	0	0	0	0	0	0	0
77	-	-	-	0	0	-	-	0	-	0	-	-
78	1	-	1	1	-	-	1	2	-	2	2	2
79	0	-	-	-	-	-	0	0	-	-	0	-
80	0	0	0	0	0	0	0	0	0	0	0	-
81	0	0	0	0	0	0	0	0	0	0	0	0
82	0	0	0	0	0	0	0	0	0	0	0	0

Table 6.1: Parameter estimates and standard errors for the bladder cancer data using MAPGP models

MAPGP								
	ML				BAY			
GOF		***U***	***MSE***			***U***	***MSE***	
Estimate	π_l	$\beta_{\mu 0l}$	$\beta_{\mu 1l}$	a_l	π_l	$\beta_{\mu 0l}$	$\beta_{\mu 1l}$	a_l
Original								
GOF		**−0.8998***	**2.2632**			**−0.9000**	**2.2639**	
$l = 1$	0.5512	-1.9140	-0.2048	1.2459	0.5369	-1.9380	-0.2825	1.2590
	0.0615	*0.9903*	*0.6150*	*0.1437*	*0.0625*	*1.5780*	*1.3100*	*0.1353*
$l = 2$	0.4488	0.5796	-0.1862	1.0110	0.4631	0.6167	-0.2219	1.0130
	0.0615	*0.2863*	*0.2082*	*0.0256*	*0.0625*	*0.2956*	*0.2176*	*0.0253*
Simplified								
GOF		**−1.1357**	**2.2453**			**−1.1359**	**2.2453**	
$l = 1$	0.5862	-1.9212	-0.1719	1.2020	0.5821	-1.8680	-0.2369	1.2210
	0.0570	*0.8702*	*0.5084*	*0.1752*	*0.0565*	*0.8890*	*0.5581*	*0.1392*
$l = 2$	0.4138	0.5728	-0.1205	1.0181	0.4179	0.5945	-0.1398	1.0190
	0.0570	*0.2173*	*0.1302*	*0.0237*	*0.0565*	*0.1962*	*0.1343*	*0.0174*

*Model with the largest U is chosen to be the best model.

Table 6.2: Basic descriptive statistics for the blood donation data

		Male		Female	
		Count	%	Count	%
Count		3904		4561	
Age	< 20	661	16.9	767	16.8
	20-29	1889	48.4	1896	41.6
	30-39	787	20.2	1101	24.1
	40-49	438	11.2	633	13.9
	≥ 50	129	3.3	164	3.6
Intention	Active	2016	51.6	2286	50.1
	Passive	1888	48.4	2275	49.9
Return	1st year	896	23.0	1060	23.2
	2nd year	371	9.5	538	11.8
	3rd year	193	4.9	230	5.0
	4th year	112	2.9	137	3.0
	no return	2332	59.7	2596	56.9
ADF	1	2332	59.7	2596	56.9
	2	646	16.5	810	17.8
	3	326	8.4	468	10.3
	4	203	5.2	253	5.5
	≥ 5	397	10.2	434	9.5
Total ADF per donor	Active	2.57		2.40	
	Passive	1.87		1.82	
	Overall	2.23		2.11	

Table 6.3: Parameter estimates and standard errors for female donors

	EM				Bay				EM					Bay				
GOF		***U***	***M***SE			***U***	***M***SE			***U***	***M***SE				***U***	***M***SE		
Est.	π_l	$\beta_{\mu 0l}$	a_{0l}†	β_{a1l}	π_l	$\beta_{\mu 0l}$	a_{0l}†	β_{a1l}	π_l	$\beta_{\mu 0l}$	$\beta_{\mu 1l}$	a_{0l}†	β_{a1l}	π_l	$\beta_{\mu 0l}$	$\beta_{\mu 1l}$	a_{0l}†	β_{a1l}
	Original $MPGP_2$								Original $MAPGP_2$									
GOF		**-0.853**	**0.263**			**-0.850**	**0.267**			**-0.851**	**0.259**				**-0.851**	**0.262**		
C	32	0.326	1.331		36	0.305	1.374		35	0.771	-0.322	1.359		37	0.769	-0.328	1.383	
		0.029	*0.022*			*0.029*	*0.026*			*0.061*	*0.037*	*0.022*			*0.064*	*0.040*	*0.025*	
O	68	0.133	8.248		64	0.136	10.43		65	0.370	-0.155	9.489		63	0.358	-0.146	10.68	
		0.024	*0.363*			*0.026*	*1.077*			*0.072*	*0.045*	*0.463*			*0.078*	*0.049*	*1.104*	
	Simplified $MPGP_2$								Simplified $MAPGP_2$									
GOF		**-0.744**	**0.258**			**-0.744**	**0.258**			**-0.741**	**0.253**				**-0.741**	**0.253**		
C	28	0.404	1.310		28	0.405	1.306		31	0.828	-0.312	1.338		29	0.841	-0.313	1.317	
		0.020	*0.017*			*0.021*	*0.020*			*0.043*	*0.027*	*0.017*			*0.051*	*0.033*	*0.020*	
O	72	0.134	7.851		72	0.134	7.722		69	0.345	-0.139	9.259		71	0.354	-0.145	8.148	
		0.016	*0.296*			*0.016*	*0.625*			*0.049*	*0.031*	*0.040*			*0.048*	*0.031*	*0.700*	
	Simplified $MPGP_3$								Simplified $MAPGP_{3a}$									
GOF		**-0.737**	**0.196**			**-0.737**	**0.193**			**-0.737**	**0.191**				**-0.735**	**0.191**		
C	6	0.726	1.117		7	0.712	1.127		9	0.979	-0.167	1.102		8	1.029	-0.262	1.126	
		0.035	*0.022*			*0.042*	*0.024*			*0.075*	*0.048*	*0.023*			*0.087*	*0.061*	*0.026*	
D	32	0.268	1.649		36	0.242	1.774		35	0.702	-0.303	1.639		36	0.681	-0.299	1.809	
		0.021	*0.025*			*0.227*	*0.062*			*0.048*	*0.031*	*0.024*			*0.062*	*0.041*	*0.075*	
O	61	0.125	18.04		57	0.122	36.73		56	0.291	-0.108	16.82		56	0.260	-0.092	39.05	
		0.018	*1.251*			*0.019*	*22.01*			*0.054*	*0.035*	*1.221*			*0.063*	*0.040*	*23.12*	
	Simplified $MPAGP_{3b}$								Simplified $MAPGP_{3c}$									
GOF		**-0.736**	**0.195**			**-0.736**	**0.194**			**-0.736**	**0.227**				‡**-0.734**	**0.194**		
C	8	0.754	†0.597	-0.117	7	0.730	†0.631	-0.127	28	0.981	-0.332	†0.915	-0.163	12	1.106	-0.312	†0.666	-0.134
		0.034	*0.082*	*0.021*		*0.042*	*0.088*	*0.022*		*0.045*	*0.029*	*0.057*	*0.015*		*0.076*	*0.051*	*0.087*	*0.021*
D	30	0.256	1.679		36	0.233	1.922		20	0.617	-0.273	2.594		36	0.658	-0.290	2.034	
		0.022	*0.027*			*0.023*	*0.079*			*0.073*	*0.048*	*0.075*			*0.062*	*0.040*	*0.100*	
O	62	0.125	17.34		57	0.117	62.14		52	0.230	-0.079	71.83		52	0.243	-0.084	76.02	
		0.018	*1.172*			*0.020*	*33.54*			*0.061*	*0.039*	*11.57*			*0.062*	*0.039*	*36.07*	

† β_{a0l} for $MAPGP_{3b}$ and $MAPGP_{3c}$ models with adaptive ratio function.

‡ U for the best model.

Table 6.4: Parameter estimates and standard errors for male donors

	EM			Bay			EM				Bay			
GOF		***U***	***M***SE		***U***	***M***SE		***U***	***M***SE			***U***	***M***SE	
Est.	π_l	$\beta_{\mu 0l}$	a_{0l}	π_l	$\beta_{\mu 0l}$	a_{0l}	π_l	$\beta_{\mu 0l}$	$\beta_{\mu 1l}$	a_{0l}	π_l	$\beta_{\mu 0l}$	$\beta_{\mu 1l}$	a_{0l}
		Original $MPGP_2$						Original $MAPGP_2$						
GOF		**-0.855**	**0.390**		**-0.855**	**0.391**		**-0.852**	**0.380**			**-0.852**	**0.380**	
C	28	0.465	1.259	28	0.460	1.266	30	1.065	-0.439	1.283	29	1.076	-0.445	1.281
		0.033	*0.023*		*0.035*	*0.026*		*0.068*	*0.042*	*0.023*		*0.073*	*0.046*	*0.026*
O	72	0.147	9.518	72	0.149	10.13	70	0.299	-0.099	11.09	71	0.296	-0.098	10.95
		0.025	*0.476*		*0.026*	*1.180*		*0.075*	*0.048*	*0.615*		*0.076*	*0.049*	*1.289*
		Simplified $MPGP_2$						Simplified $MAPGP_2$						
GOF		**-0.757**	**0.331**		**-0.757**	**0.332**		**-0.757**	**0.377**			**-0.753**	**0.330**	
C	15	0.705	1.157	15	0.709	1.156	28	1.086	-0.417	1.282	17	1.279	-0.451	1.171
		0.025	*0.017*		*0.030*	*0.018*		*0.045*	*0.029*	*0.016*		*0.056*	*0.038*	*0.018*
O	85	0.129	4.508	85	0.129	4.493	72	0.305	-0.100	12.35	83	0.346	-0.142	5.010
		0.016	*0.113*		*0.016*	*0.190*		*0.052*	*0.034*	*0.669*		*0.047*	*0.030*	*0.247*
		Simplified $MPGP_3$						Simplified $MAPGP_{3a}$						
GOF		**-0.737**	**0.247**		**-0.736**	**0.249**		**-0.735**	**0.255**			‡**-0.734**	**0.245**	
C	6	0.999	1.075	7	0.998	1.084	9	1.250	-0.281	1.097	7	1.337	-0.287	1.080
		0.033	*0.020*		*0.037*	*0.021*		*0.061*	*0.041*	*0.018*		*0.076*	*0.058*	*0.021*
D	27	0.313	1.625	34	0.274	1.810	33	0.634	-0.253	1.933	34	0.674	-0.270	1.834
		0.024	*0.028*		*0.024*	*0.044*		*0.056*	*0.035*	*0.036*		*0.065*	*0.042*	*0.043*
O	67	0.145	24.71	60	0.137	4691	58	0.257	-0.080	234.1	59	0.256	-0.081	4874
		0.018	*2.107*		*0.020*	*3091*		*0.061*	*0.039*	*69.88*		*0.060*	*0.038*	*2968*

‡ U for the best model.

Table 6.5: Classification of predicted group using the best model

Classification	Female					Male				
Using 2 year's obs.	Using 4 year's obs.			Total	Correct	Using 4 year's obs.			Total	Correct
Group	C	D	O	freq.	prediction	C	D	O	freq.	prediction
C*	157	56	0	213	74%	136	26	0	162	84%
D*	†180	729	0	909	80%	†93	576	0	669	86%
O*	83	‡448	2908	3439	85%	25	‡406	2642	3073	86%
Total frequency	418	1235	2908	4561		254	1008	2642	3904	
Correct classification	38%	59%	100%			54%	57%	100%		

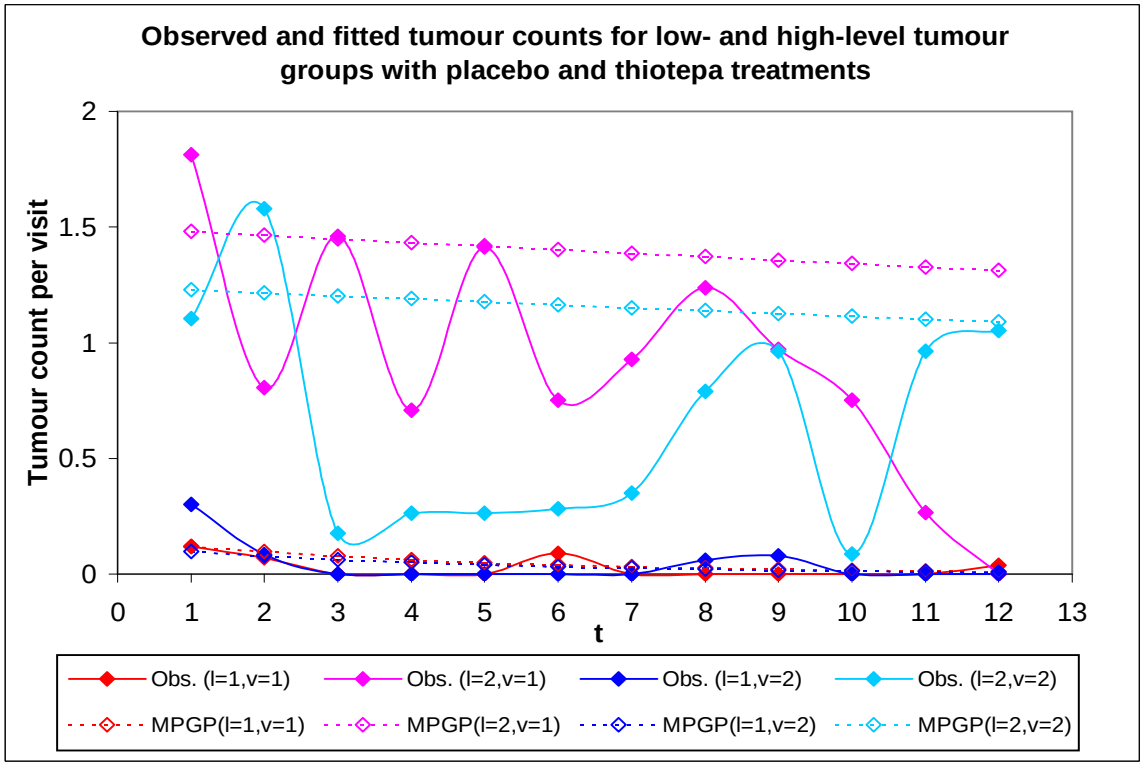
† The C donors are classified into D* group using $W_{it},\ t = 1, 2$.

‡ The D donors are classified into O* group using $W_{it},\ t = 1, 2$.

Table 6.6: Target donors for the promotional programs

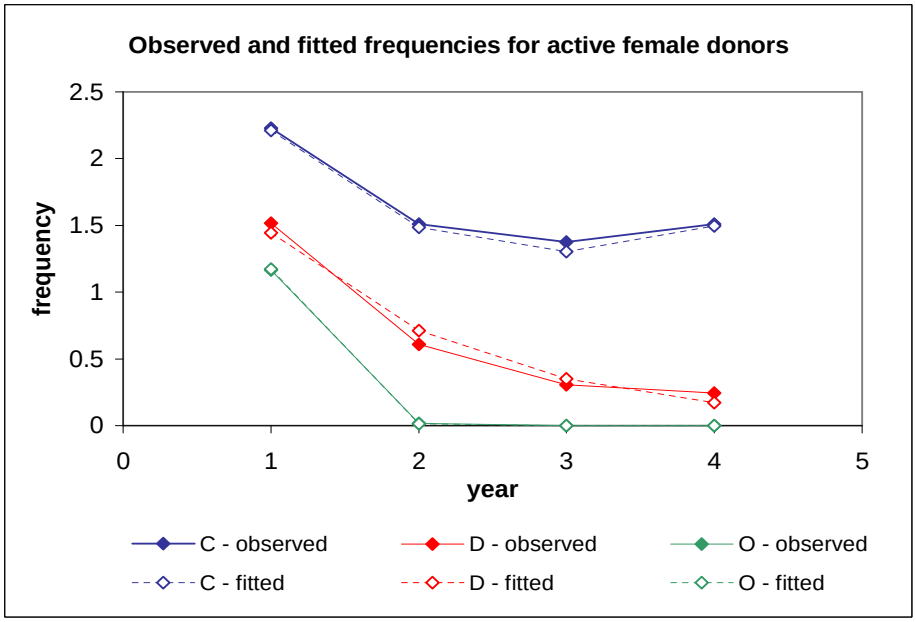
		Female						Male					
Target	Target	Count	DF	DF	TDF	DF	TDF	Count	DF	DF	TDF	DF	TDF
donor	group	(%)	1-2	3-4	3-4	5	5	(%)	1-2	3-4	3-4	5	5
C	CD*	197 (4)	5.1	2.5	495	2.1	405	96 (2)	6.0	3.6	341	2.0	192
D active	DC*, D*, OD*	816 (18)	2.6	0.8	654	0.1	44	636 (16)	2.8	0.9	595	0.1	45
D passive	DC*, D*	393 (9)	2.4	0.8	299	0.1	25	329 (8)	2.5	0.9	282	0.1	33
Total		1406 (31)			1448		473	1061 (27)			1218		271

Figure 6.1: Plot of the observed w_{tlv} and expected $\hat{w}_{tlv}$ tumour counts for the bladder cancer patients using original MAPGP model with classical ML method



Remark: The fluctuation of w_{t2v} for high-level tumour group ($l = 2$) is possibly due to the small number relative to the large variation of the observations in each visit t for the group ($l = 2, v = 1$) and ($l = 2, v = 2$). For the former group with 23 patients, the calculation of w_{t21} are averaged over 20, 17, 19, 16, 15, 17, 12, 17, 15, 13, 6 and 9 observations with the variation of 8, 8, 8, 6, 8, 9, 8, 8, 8, 8, 4 and 0 tumour counts among patients for $t = 1, \ldots, 12$ respectively. Whereas for the latter group with 12 patients, the numbers are 9, 10, 7, 7, 6, 8, 6, 6, 6, 5, 3, 2 correspondingly with the variation of 7, 7, 2, 3, 3, 2, 2, 3, 3, 1, 8 and 6 tumour counts.

Figure 6.2: Plot of the observed w_{tlv} and expected $\hat{w}_{tlv}$ donation frequencies for female donors using simplified $MAPGP_{3c}$ model with Bayesian method



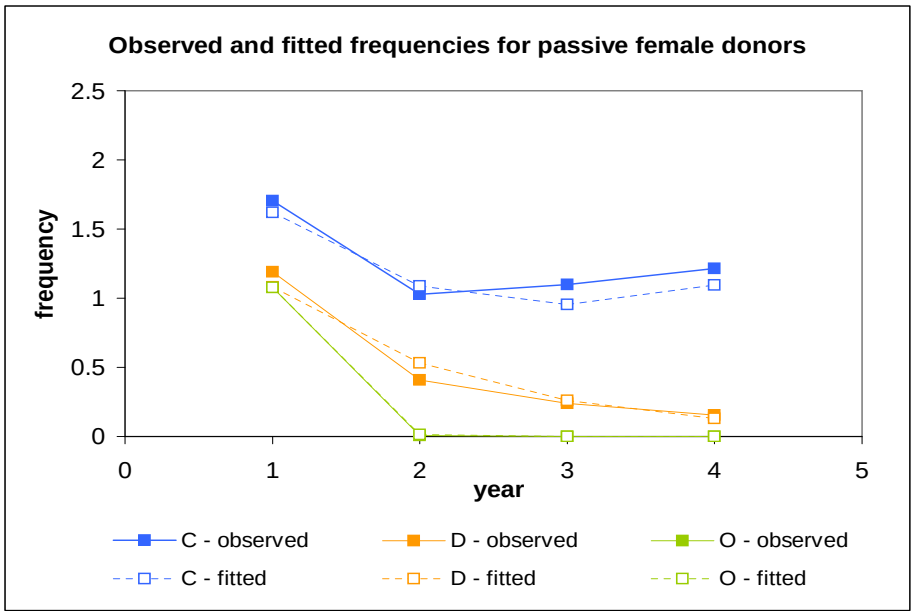
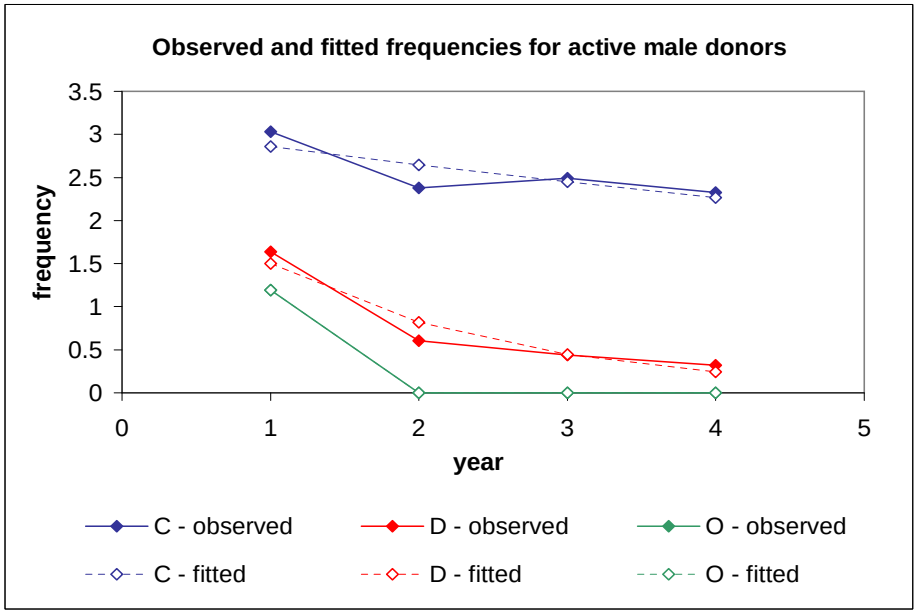


Figure 6.3: Plot of the observed w_{tlv} and expected $\hat{w}_{tlv}$ donation frequencies for male donors using simplified $MAPGP_{3a}$ model with Bayesian method



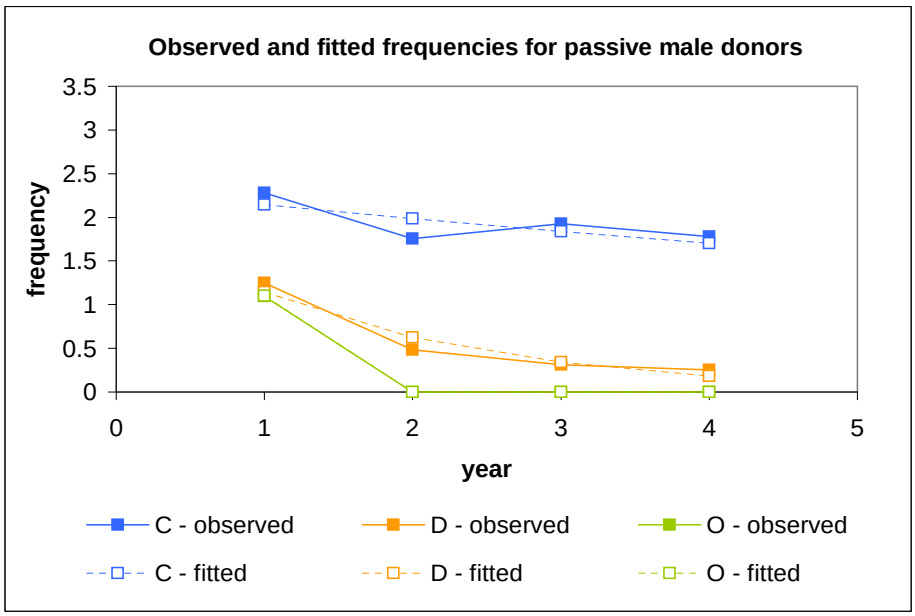


Figure 6.4: Observed proportion p_{t1uv} and expected probability $f_{s,t1}(u|z_{\mu 2i} = v)$ across frequency u among committed female donors

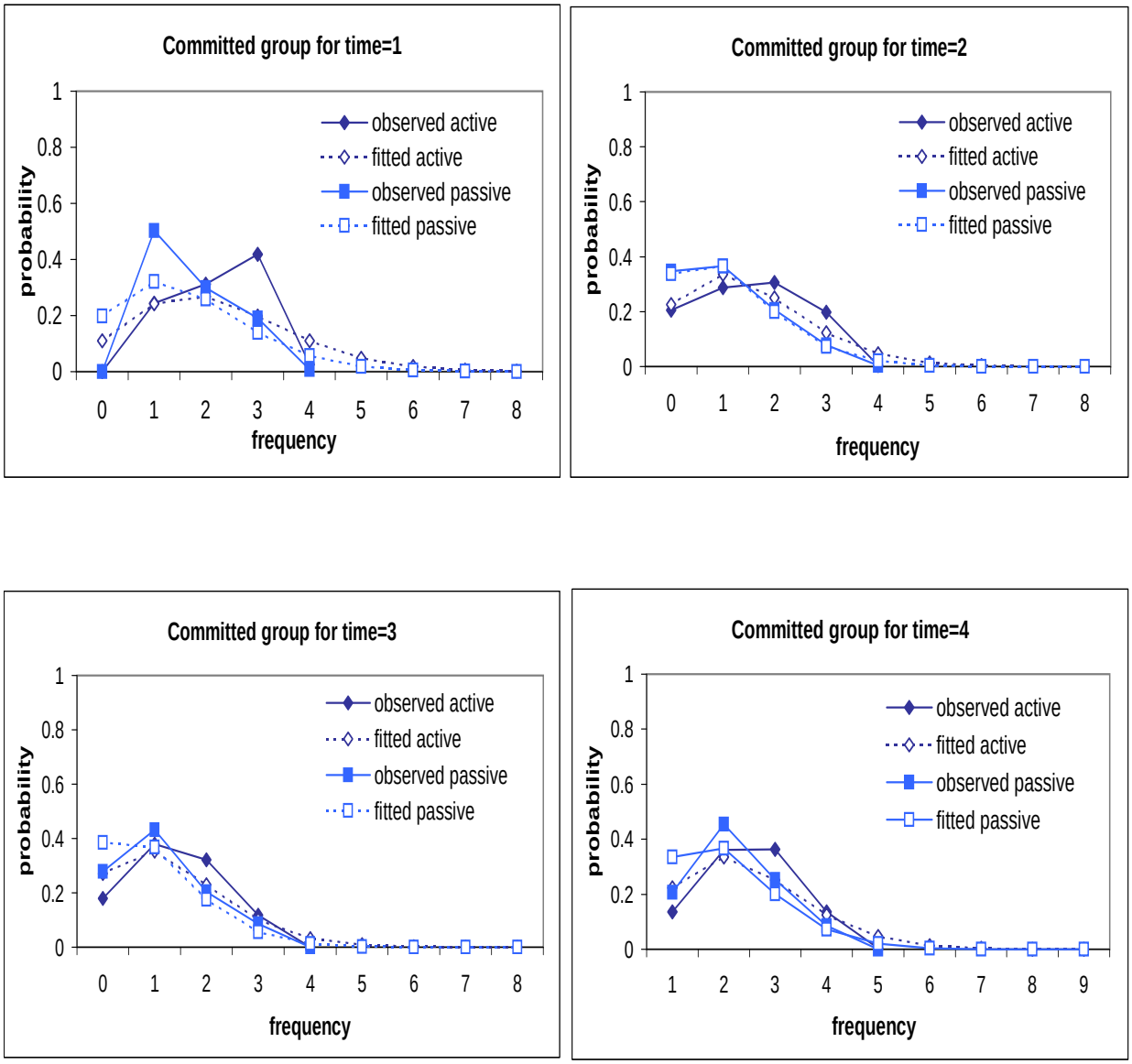


Figure 6.5: Observed proportion p_{t2uv} and expected probability $f_{s,t2}(u|z_{\mu 2i} = v)$ across frequency u among dropout female donors

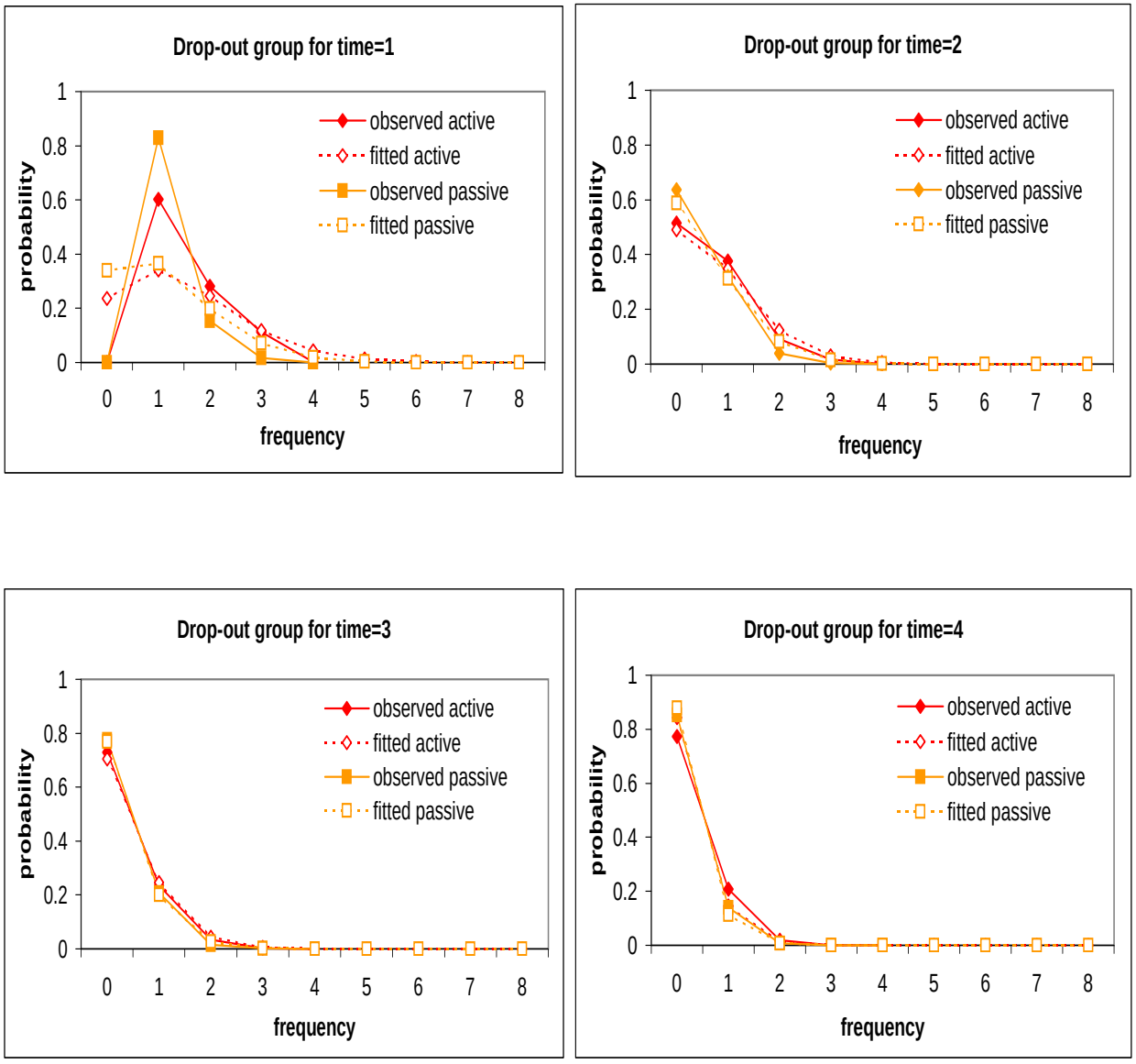


Figure 6.6: Observed proportion p_{t3uv} and expected probability $f_{s,t3}(u|z_{\mu 2i} = v)$ across frequency u among one-time female donors

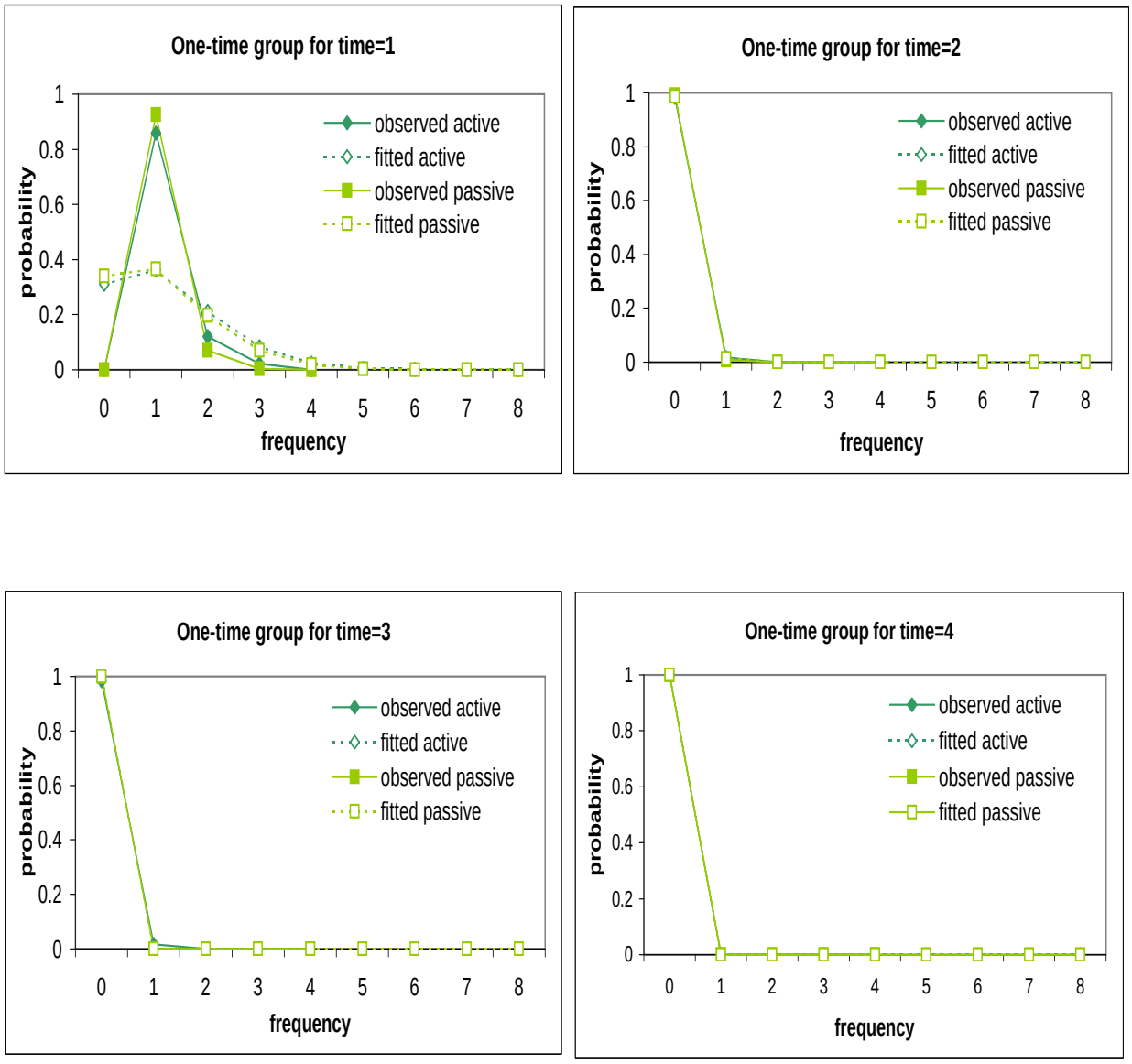
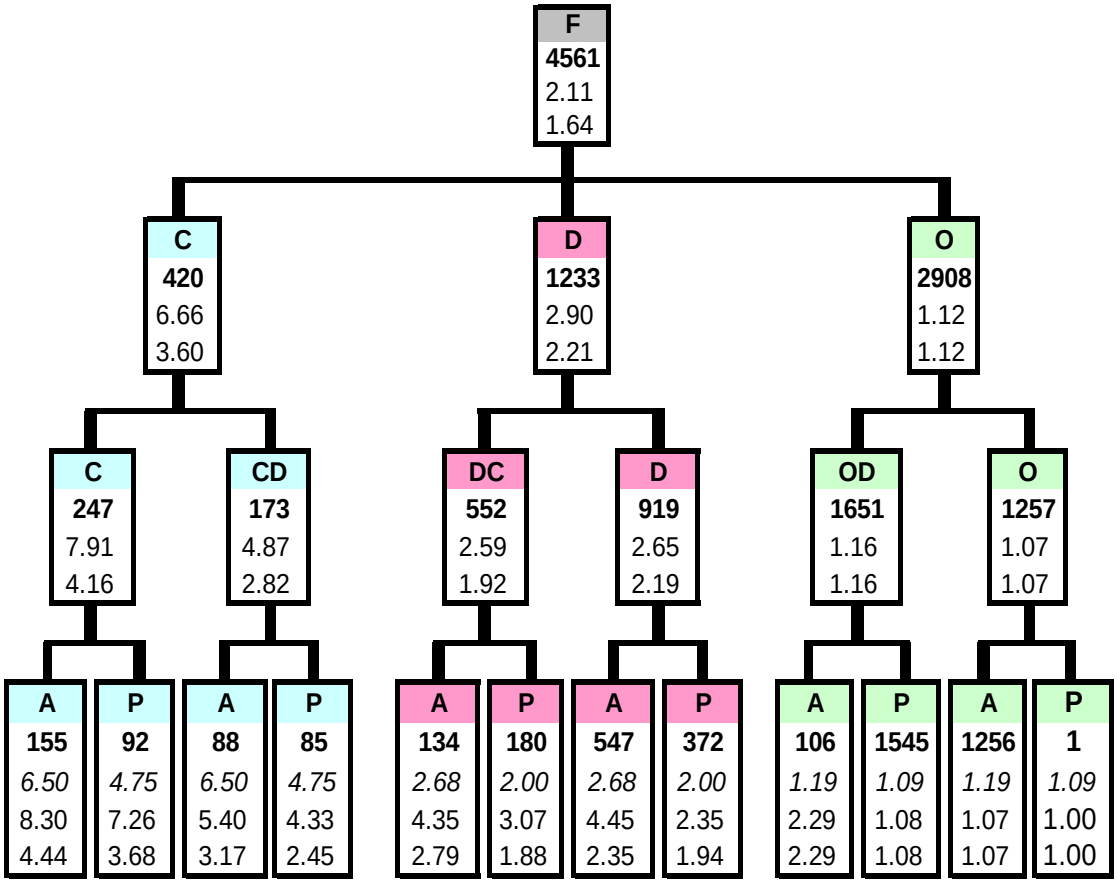
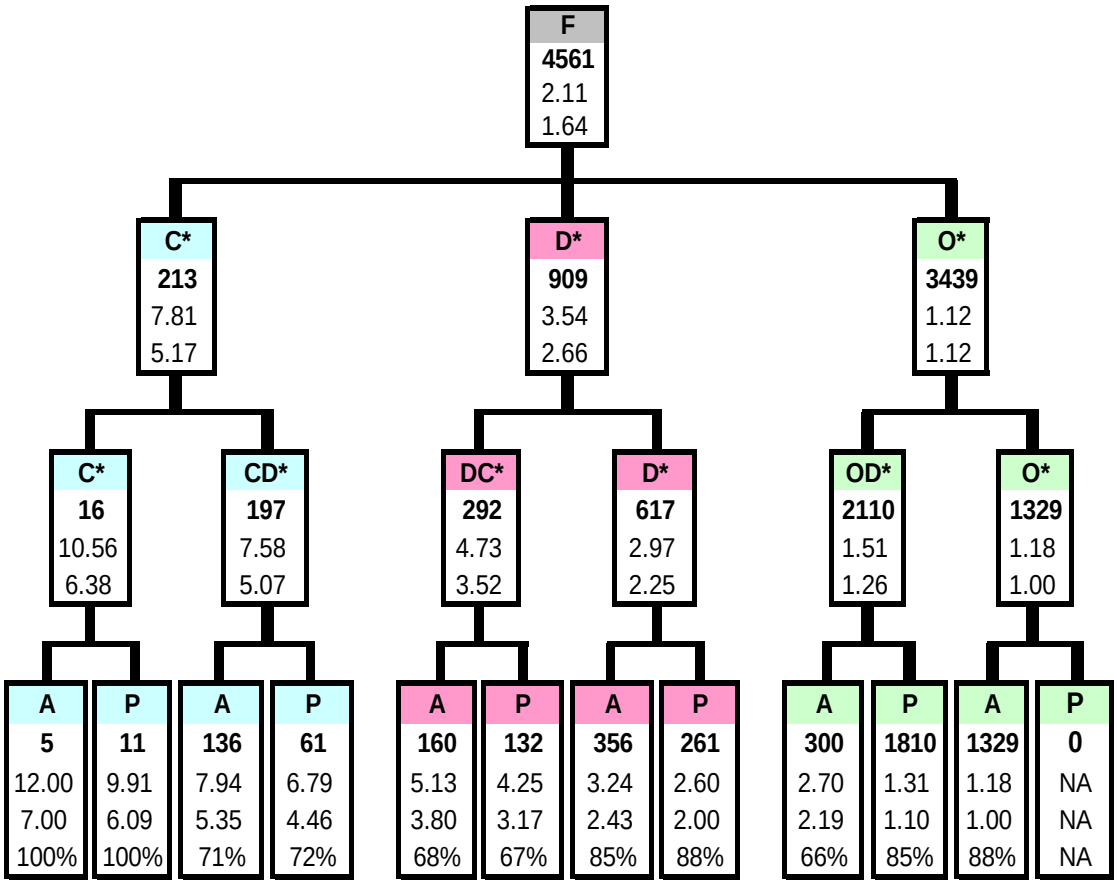


Figure 6.7: Classification of female donors based on four year donation frequencies



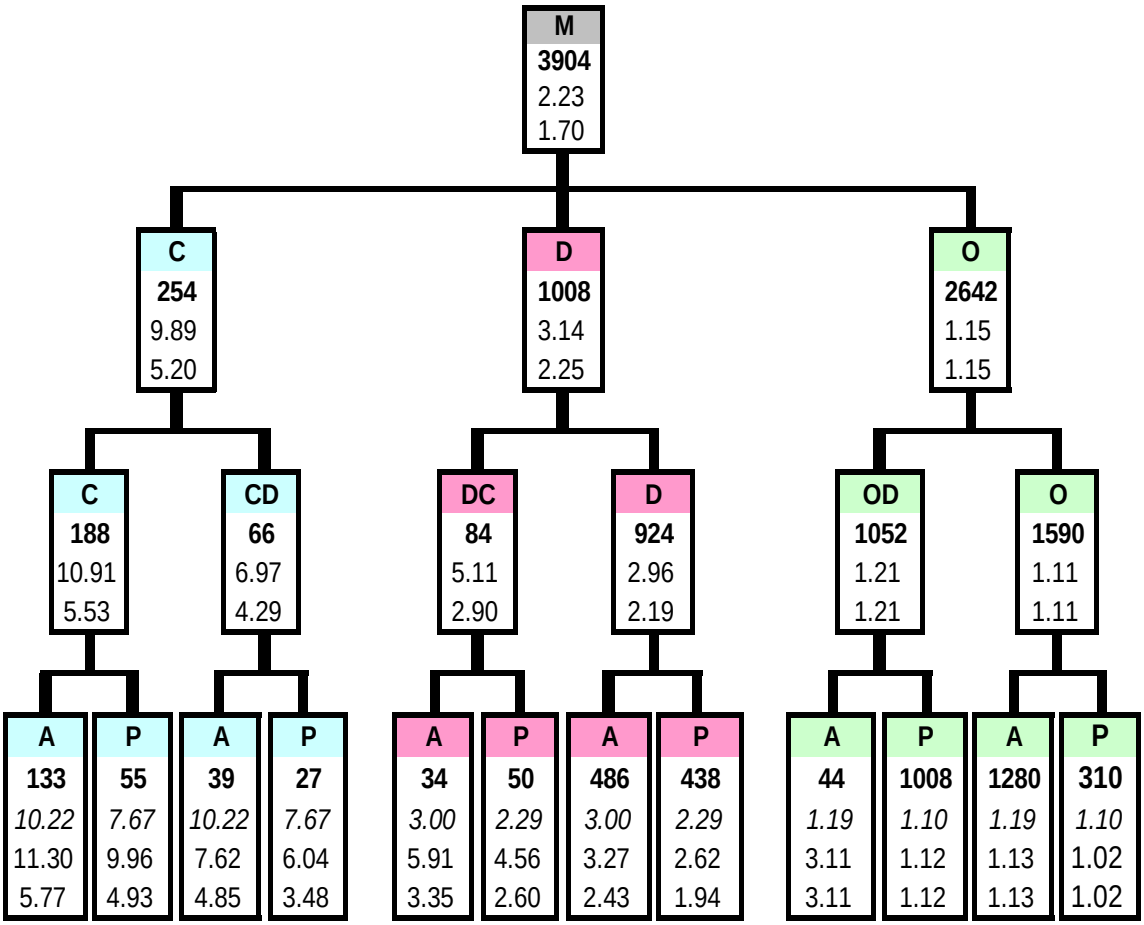
Each cell contains respectively the frequency n_{lcv} (bold), the predicted total ADF $\hat{w}_{lcv}$ (italic), the observed total ADF w_{lcv} and the observed total ADF w^*_{lcv} in the first two years.

Figure 6.8: Classification of female donors based on two year donation frequencies



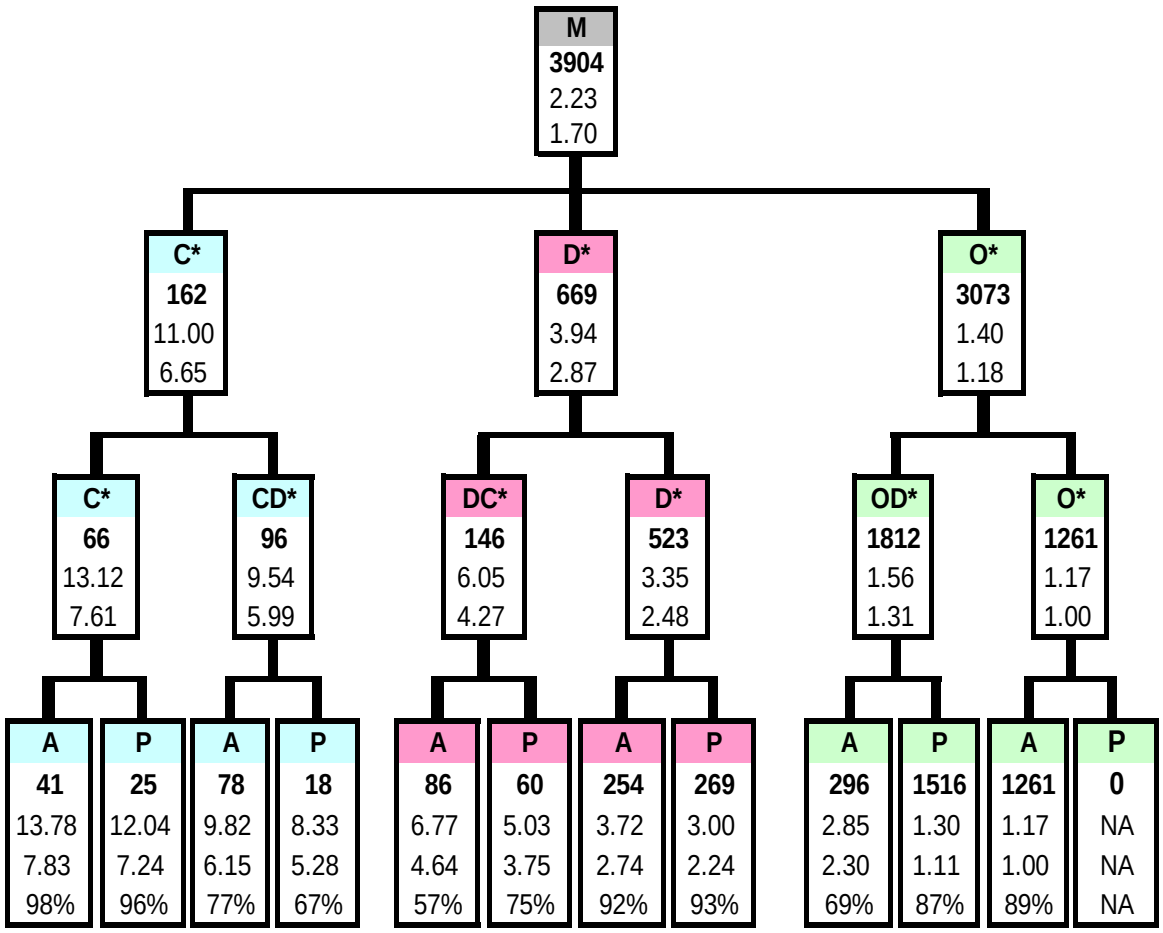
Each cell contains respectively the frequency n_{lcv} (bold), the observed total ADF w_{lcv}, the observed total ADF w^*_{lcv} in the first two years and the percentage of correct classification p^*_{lcv}.

Figure 6.9: Classification of male donors based on four year donation frequencies



Each cell contains respectively the frequency n_{lcv} (bold), the predicted total ADF $\hat{w}_{lcv}$ (italic), the observed total ADF w_{lcv} and the average observed total ADF w^*_{lcv} in the first two years.

Figure 6.10: Classification of male donors based on two year donation frequencies



Each cell contains respectively the frequency n_{lcv} (bold), the observed total ADF w_{lcv}, the observed total ADF w^*_{lcv} in the first two years and the percentage of correct classification p^*_{lcv}.

Chapter 7

Zero-Inflated Poisson Geometric Process Model

7.1 Introduction

When we deal with repeated measurements of count data, the situation of excess zeros sometimes arises. Here, the term "excess zeros" implies that the incidence of zero counts is greater than expected for the Poisson distribution. We have to take these excess zeros into account because there maybe different interpretations for the zero count. For instances, the number of times a subject used medical services per month during the past ten years may contain many zero observations, either because they have not suffered from any disease by chance or they cannot afford the expensive payment for visiting a doctor. Another example is the daily frequencies of traffic accidents in different road intersections. Some intersections are recognized as the black spot while some may have no car accident record for years, but the frequencies are highly correlated with the total traffic volume, the road width, and the traffic signal cycle time, etc. Therefore, it is important to distinguish between structural zeros, which are inevitable, and sampling zeros, which occur by chance (Ridout (51)).

7.1.1 Previous researches

In recent years, increasing researches have been done on zero-inflated count data. So far, two main types of models are developed, namely the hurdle model and the zero-inflated Poisson (ZIP) model. The hurdle model was originally proposed by Mullahy (48) in the econometric study. It is essentially a two-part model in which one part is a binary model measuring whether the outcome W_{it} overcomes the hurdle, while another part is a truncated model explaining those outcomes which passed the hurdle. In the study of zero-inflated count data, the hurdle is set to zero and a discrete distribution, like Poisson distribution or negative binomial distribution, are assigned to the truncated model. The general definition of hurdle model is given as follow:

Definition: For the outcome $\boldsymbol{W}$, denote w_i to be the observation for subject $i, i = 1, \ldots, m$. Assume that the binary model is governed by a pmf g_1 and that $\{W_i | W_i > 0\}$ follows a truncated-at-zero pmf g_2, the complete distribution of the hurdle model is then given by:

$$P(W_i = 0) = g_1(0) \tag{7.1}$$

$$\text{and} \qquad P(W_i = j) = [1 - g_1(0)] \frac{g_2(j)}{1 - g_2(0)}, \quad j = 1, 2, \ldots \tag{7.2}$$

An alternative approach to model zero-inflated count data would be the zero-inflated count model pioneered by Lambert (41). He assumed that the count data comes from a mixture of a Poisson distribution and a distribution degenerated at zero, the resultant model is called the zero-inflated Poisson (ZIP) model with definition as follows:

Definition: Assume that for subject i, the probability of a zero outcome, that is $P(W_i = 0) = 1 - \phi_i$, and the probability of a non-zero outcome which comes from a Poisson distribution with mean λ_i is ϕ_i, then the probability distribution for the ZIP model is:

$$P(W_i = 0) = (1 - \phi_i) + \phi_i e^{-\lambda_i} \tag{7.3}$$

$$\text{and} \qquad P(W_i = j) = \phi_i \frac{e^{-\lambda_i} \lambda_i^j}{j!}, \quad j = 1, 2, \ldots \tag{7.4}$$

This ZIP model aligns with the two-group mixture model for the zero observation $W_i = 0$. Together with the hurdle model, these two major models for zero-inflated count data have been applied in a diversity of areas. For instances, Lambert (41) studied the manufacturing defects using the ZIP model. Miaou (46) analyzed the relationship between truck accidents and geometric design of road sections using the hurdle model. Shonkwiler and Shaw (57) used the hurdle model in recreation demand analysis. Ridout (51) compared the hurdle and the ZIP models using some biological examples in horticultural research. Moreover, extensions have been made to these models such as incorporating random effects into the models (Min and Agresti (47)), developing a semi-parametric hurdle model to improve its robustness (Gurmu (21)), etc.

Despite the many researches on the zero-inflated count models, few works have been done on the modeling of clustered observations in longitudinal data (Min and Agresti (47)). Hence we propose, in this thesis, a zero-inflated PGP (ZIPGP) model for longitudinal measurements to model excess zero observations in the count data.

7.1.2 The model

Adopting the idea of the hurdle model, we extend the PGP model into a two-part ZIPGP model. Suppose there is a probability $\phi_h, h = o, s$ that the observed count W_{it} of subject i at time t is zero, we denote the indicator of the event "$W_{it} = 0$" by $I_{h,it0} = I_h(W_{it} = 0)$, then $I_{h,it1} = 1 - I_{h,it0}$ indicates the event "$W_{it} > 0$". Moreover, we let $n_0 = \sum_{i=1}^{m} \sum_{t=1}^{n_i} I_{h,it0}$ be the total number of zero observations and $n_1 = n - n_0$, then we can write the pmf for the original ($h = o$) and simplified ($h = s$) ZIPGP models as:

$$f_{h,it}(w_{it}) = \begin{cases} \phi_h & w_{it} = 0, \\ (1 - \phi_h)\dfrac{L_{h,it}}{1 - L_{h,it0}} & w_{it} > 0 \end{cases} \tag{7.5}$$

where $L_{h,it0}$ and $L_{h,it}$ for the original and simplified models are given respectively by:

$$\text{Original model: } L_{o,it0} = f_{o,it}(0) = Pr(W_{it} = 0) = \frac{a_t^{t-1}/\mu_t}{(a_t^{t-1}/\mu_t + 1)}, \tag{7.6}$$

$$L_{o,it} = f_{o,it}(w_{it}) = Pr(W_{it} = w_{it}) = \frac{a_{it}^{t-1}/\mu_{it}}{(a_{it}^{t-1}/\mu_{it} + 1)^{w_{it}+1}}, \tag{7.7}$$

$$\text{Simplified model: } L_{s,it0} = f_{s,it}(0) = Pr(W_{it} = 0) = \exp\left(-\frac{\mu_{it}}{a_{it}^{t-1}}\right), \tag{7.8}$$

$$L_{s,it} = f_{s,it}(w_{it}) = Pr(W_{it} = w_{it}) = \frac{\exp\left(-\dfrac{\mu_{it}}{a_{it}^{t-1}}\right)\left(\dfrac{\mu_{it}}{a_{it}^{t-1}}\right)^{w_{it}}}{w_{it}!}, \tag{7.9}$$

and the ratio function a_{it} and the mean function μ_{it} are given by:

$$\mu_{it} = \exp(\eta_{\mu it}) = \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu 1 it} + \cdots + \beta_{\mu q_\mu} z_{\mu q_\mu it}) \tag{7.10}$$

$$a_{it} = \exp(\eta_{ait}) = \exp(\beta_{a0} + \beta_{a1} z_{a1it} + \cdots + \beta_{aq_a} z_{aq_a it}) \tag{7.11}$$

correspondingly where $z_{jkit}, j = \mu, a$, $k = 1, \ldots, q_j, i = 1, \ldots, m$ are the time-evolving covariate effects at time $t = 1, \ldots, n_i$. Then, the conditional mean $E_h(W_{it})$ for the ZIPGP model is given by:

$$E_h(W_{it}) = \begin{cases} 0 & w_{it} = 0, \\ \hat{w}_{it} & w_{it} > 0 \end{cases} \tag{7.12}$$

where $\hat{w}_{it} = \dfrac{\mu_{it}}{a_{it}^{t-1}}$. Thereafter, we can derive the likelihood function $L_h(\boldsymbol{\theta}_h)$ for the observed data W_{it} which involves the indicator $I_{h,it0}$ and the probability ϕ_h as:

$$L_h(\boldsymbol{\theta}_h) = \prod_{i=1}^{m} \prod_{t=1}^{n_i} \phi_h^{I_{h,it0}} \left[(1 - \phi_h) \frac{L_{h,it}}{1 - L_{h,it0}} \right]^{I_{h,it1}} \tag{7.13}$$

and the log-likelihood function $\ell_h(\boldsymbol{\theta}_h)$ is:

$$\ell_h(\boldsymbol{\theta}_h) = n_0 \ln \phi_h + n_1 \ln(1 - \phi_h) + \sum_{i=1}^{m} \sum_{t=1}^{n_i} I_{h,it1} [\ln L_{h,it} - \ln(1 - L_{h,it0})] \tag{7.14}$$

where $L_{h,it0}$ are given by (7.6) and (7.8), and $L_{h,it}$ are given by (7.7) and (7.9). The log-likelihood functions are not complicated and we can implement the classical ML method or the Bayesian method to estimate the parameters $\boldsymbol{\theta}_h$. For model extension, we focus on the incorporation of covariate effects in the mean function, the resultant model is called the adaptive ZIPGP (AZIPGP) model with the vector of model parameters $\boldsymbol{\theta}_h = (\beta_{\mu 0}, \ldots, \beta_{\mu q_\mu}, a, \phi)^T$. The AZIPGP model is in fact a simple extension from the APGP model by multiplying $\dfrac{(1 - \phi_h)}{1 - L_{h,it0}}$ to the likelihood function $L_{h,it}$ when $W_{it} \geq 1$ and setting $L_{h,it} = \phi_h$ when $W_{it} = 0$.

7.2 Methodology of inference

In this chapter, we assume that there are covariate effect only in the mean function. In other words, $a_t = a$ in the AZIPGP model. The classical ML and the Bayesian methods are considered for model implementation.

7.2.1 Maximum likelihood method

To implement the classical ML method for parameter estimation, we derive the likelihood functions $L_h(\boldsymbol{\theta}_h)$ for the original and simplified AZIPGP models based on the equation (7.13) as follows:

$$L_o(\boldsymbol{\theta}_o) = \prod_{i=1}^{m}\prod_{t=1}^{n_i} \phi_o^{I_{o,it0}} \left[(1-\phi_o)\frac{\hat{w}_{it}^{(w_{it}-1)}}{(\hat{w}_{it}+1)^{w_{it}}}\right]^{I_{o,it1}} \tag{7.15}$$

$$\text{and} \quad L_s(\boldsymbol{\theta}_s) = \prod_{i=1}^{m}\prod_{t=1}^{n_i} \phi_s^{I_{s,it0}} \left[(1-\phi_s)\frac{e^{-\hat{w}_{it}}\hat{w}_{it}^{w_{it}}}{(1-e^{-\hat{w}_{it}})w_{it}!}\right]^{I_{s,it1}} \tag{7.16}$$

where $\hat{w}_{it} = \dfrac{\mu_{it}}{a^{t-1}}$ and μ_{it} is given by (7.12). Then, the log-likelihood functions $l_h(\boldsymbol{\theta}_h)$ would become

$$\begin{aligned}\ell_o(\boldsymbol{\theta}_o) &= n_0 \ln\phi_o + n_1 \ln(1-\phi_o) + \\ &\quad \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{o,it1}[(w_{it}-1)\ln\hat{w}_{it} - w_{it}\ln(\hat{w}_{it}+1)]\end{aligned} \tag{7.17}$$

and

$$\begin{aligned}\ell_s(\boldsymbol{\theta}_s) &= n_0 \ln\phi_s + n_1 \ln(1-\phi_s) + \\ &\quad \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{s,it1}[-\hat{w}_{it} + w_{it}\ln\hat{w}_{it} - \ln(1-e^{-\hat{w}_{it}}) - \ln w_{it}!]\end{aligned} \tag{7.18}$$

respectively. Afterwards, the first order and second order derivative functions for both original and simplified AZIPGP models can be derived and are given in Appendix 7.1 for the implementation of the NR iterative method to solve the model parameters.

7.2.2 Bayesian method

The Bayesian hierarchy of the original and simplified AZIPGP models for solving the parameters $\boldsymbol{\theta}_h, h = o, s$ are given by:

$$\begin{aligned}
\text{Likelihood for the original model: } w_{it} &\sim f_{o,it}(w_{it}) \\
f_{o,it}(w_{it}) &= I_{o,it0} + I_{o,it1} g_{o,it}(w_{it}) \\
g_{o,it}(w_{it}) &= \frac{\lambda_{it}^{(w_{it}-1)}}{(\lambda_{it}+1)^{w_{it}}} \\
\lambda_{it} &= \frac{\mu_{it}}{{a_{it}}^{t-1}} \\
\text{Likelihood for the simplified model: } w_{it} &\sim f_{s,it}(w_{it}) \\
f_{s,it}(w_{it}) &= I_{s,it0} + I_{s,it1} g_{s,it}(w_{it}) \\
g_{s,it}(w_{it}) &= \frac{e^{-\lambda_{it}} \lambda_{it}^{w_{it}}}{(1-e^{-\lambda_{it}}) w_{it}!} \\
\lambda_{it} &= \frac{\mu_{it}}{{a_{it}}^{t-1}}
\end{aligned}$$

where μ_{it} is given by (7.10) and $a_{it} = a$ or is given by (7.11) in general.

Priors for the model parameters $\boldsymbol{\theta}_h$:

$$\begin{aligned} a &\sim U(1-b, 1+b) \quad \text{or} \quad \beta_{ak} \sim N(0, \sigma_{ak}^2),\ k = 0, 1, \ldots, q_a \\ \beta_{\mu k} &\sim N(0, \sigma_{\mu k}^2),\ k = 0, 1, \ldots, q_\mu \\ \phi_h &\sim U(0, 1) \\ I_{h,it0} &\sim \text{Bernoulli}(\phi_h) \end{aligned}$$

The values for b, σ_{ak}^2 and $\sigma_{\mu k}^2$ are similar to those stated in Section 3.4.3 and Section 4.2.3. Besides, the procedures for the implementation of the Bayesian method for both original and simplified AZIPGP models are the same as those for the APGP models. Please refer to Section 4.2.3 for further details.

It is noted that simulation studies are not conducted because this model is in fact a simple extension from the APGP model as shown from the likelihood function (7.13).

7.3 Real data analysis

Due to the problem of "excess zeros" in the bladder cancer data, we fit the data to the AZIPGP models using the classical ML and Bayesian methods. Again for the model assessment and selection, we consider the MSE_z and the U_z as the

GOF measures and they are defined as:

$$MSE_z = \frac{1}{n}\sum_{i=1}^{m}\sum_{t=1}^{n_i}(w_{it}-\hat{w}_{it})^2 \tag{7.19}$$

$$U_z = \frac{1}{n}\sum_{i=1}^{m}\sum_{t=1}^{n_i}\ln\left\{\phi_h^{I_{h,it0}}\left[(1-\phi_h)\frac{L_{h,it}}{1-L_{h,it0}}\right]^{1-I_{h,it0}}\right\} \tag{7.20}$$

where $L_{h,it}$ and $L_{h,it0}, h = o, s$ are given by (7.7), (7.9), (7.6), and (7.8) respectively, and $I_{h,it0}$ is the indictor of zero observation for each count W_{it}. Also, $\hat{w}_{it}$ in the MSE_z is 0 if $w_{it} = 0$ and is μ_{it}/a_{it}^{t-1} otherwise.

7.3.1 Objectives

In Chapter 6, we state that excess zeros add considerable noise to the detection of a monotone trend. However, will removing the zeros remove the information in the detection of a trend? We believe that the focus of the study is on the trend of new tumours for patients in each visit. A zero adds no information as it might just indicate that "new tumour is not yet found". Results in Section 7.3.2 show that the thiotepa treatment effect becomes more significant when there are less fluctuations in counts due to the removal of zeros when fitting the AZIPGP models.

The objective of fitting a AZIPGP model to the data is to examine the treatment effect to the patients allowing for the problem of excess zero observations. Of the 571 observations in the bladder cancer data, over 80% (459) are zeros. As mentioned in Section 6.3.1.3, these zeros contribute substantial noise in the detection of monotone trends. Therefore, we propose the AZIPGP model so

that those outcomes which cross the hurdle ($W_{it} > 0$) are explained by the zero-truncated APGP model. In the model, the treatment effect is again allowed for as a covariate in the mean function μ_{it} but as the first stage of development for the AZIPGP model, we do not consider mixture effects in the model. The model could be easily extended to include mixture effects.

In the model, a large $a_{it} = a$ and a small $\mu_{it} = \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu 1 it})$ reveal that the level of tumours is low in general and hence the growth of new tumours is prohibited in the patients. Meanwhile, we can also compare the thiotepa treatment effect with the placebo by looking at the parameter estimate $\hat{\beta}_{\mu 1}$. A negative value suggests that the thiotepa treatment ($z_{\mu 1 it} = 2$) relative to the placebo treatment ($z_{\mu 1 it} = 1$) will slow down the growth of new tumours for the patients.

7.3.2 Results and comments

Based on the results in Table 7.1, the parameter estimates using the classical ML and the Bayesian methods are generally consistent for the original and simplified AZIPGP models respectively. After allowing for the excess zeros, the parameter estimate $\hat{a}$ are smaller than but close to 1. In other words, the number of new tumours found in the patients is rising slowly in each visit. This result contradicts to the finding in Section 6.3.1.3 but is reasonable because the excess zeros actually pull down the rising trend to be a stationary or decreasing one. Moreover, the negative and significant (marginally insignificant for the original model) $\hat{\beta}_{\mu 1}$ suggests that the thiotepa treatment ($z_{\mu i t} = 2$) relative to the placebo will

suppress the growth of tumours for the patients with bladder cancer. This result matches in direction with the findings when the MAPGP models are used but $\hat{\beta}_{\mu 1}$ becomes much more significant now showing that the truncation of zeros in the AZIPGP model helps to detect treatment effect more clearly.

==========

Table 7.1 about here

==========

It is not surprising to find that the original models give a better U_z but a worse MSE_z as compared to the simplified models after the zero observations are truncated in the AZIPGP model. The reason is that regardless of the zero observations, the density of the original model as shown in (7.15) is more even with larger variance. In the bladder cancer data, about 50% of the non-zero observations are smaller than or equal to 2, while about 20% are greater than or equal to 5 with a maximum value 9. This shows that the data distribution is quite even. Therefore, the original AZIPGP model gives a better U and tends to fit an increasing trend that increases from a lower level at a higher rate. However, as most non-zero observations are less than 3, the MSE_z of the original model becomes relatively larger for larger t when comparing to that of the simplified model which tends to fit an increasing trend that increases from a higher level at a lower rate.

Nevertheless, the simplified AZIPGP model using the classical ML method is chosen as it has the best MSE_z. Also, the improvement in the MSE_z outweighs

the decrement in U_z when comparing it with the original model using the classical ML and the Bayesian methods. It also performs better than the simplified model using the Bayesian method in terms of the two GOF measures. The observed average tumour counts w_{tv} and expected average tumour counts $\hat{w}_{tv}$ averaged over the 46 patients receiving the placebo treatment ($v = 1$) and 36 patients receiving the thiotepa treatment ($v = 2$) at the beginning using the best model are given by Figure 7.1.

==========

Figure 7.1 about here

==========

Figure 7.1 shows that for both treatment groups, the observed average tumour counts w_{tv} go up and down vigorously and randomly, with w_{t1} generally higher than w_{t2} in general. Yet, the expected average tumour counts $\hat{w}_{tv}$ do not capture the noise in the fluctuation of the trend and remain nearly constant across time t. This phenomenon is similar to that of the patients with higher level of tumour counts in group $l = 2$ in the MAPGP model and may also be explained by the small number of non-zero observations and the high variability in the data. Moreover, the figure again reveals that the thiotepa treatment ($v = 2$) relative to the placebo will inhibit the growth of new tumours in the patients as $\hat{w}_{t1} > \hat{w}_{t2}$ in each visit t.

So far, we have studied the bladder cancer data using the MAPGP and AZIPGP models and they have the same conclusion that the thiotepa treat-

ment is better for bladder cancer patients. However, as a difference, we find that the parameter estimates $\hat{a}_l$ are generally larger than 1 in the MAPGP models whereas they are slightly smaller than 1 in the AZIPGP models. This gives us an insight that the tumour counts are actually increasing slowly rather than dropping after the truncation of the zero observations. This insight gives a rather different point of view that regardless of whether thiotepa or placebo treatment a patient receives, the growth of tumours will increase in long run but the tumours of those patients undergoing thiotepa treatment maintain at a relatively lower level. Nevertheless, the discrepancy in the parameter estimates of the ratio between the MAPGP and AZIPGP models may be clarified if there are less missing appointments for the patients and larger sample of patients in the bladder cancer study.

Appendix 7.1 Derivation of first and second order derivative functions in the NR iterative method for the AZIPGP models using classical ML method

For the original AZIPGP model, the derivative functions are obtained by differentiating the log-likelihood function $\ell_o(\boldsymbol{\theta}_o)$ in equation (7.17) and are given by:

$$\begin{aligned}
\frac{\partial \ell_o}{\partial a} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{o,it1}\left[\frac{z^*_{a0t}(\hat{w}_{it}-w_{it}+1)}{\hat{w}_{it}+1}\right] \\
\frac{\partial \ell_o}{\partial \beta_{\mu k}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{o,it1}\left[\frac{z_{\mu kit}(\hat{w}_{it}-w_{it}+1)}{\hat{w}_{it}+1}\right] \\
\frac{\partial \ell_o}{\partial \phi_o} &= \frac{n_0}{\phi_o}-\frac{n_1}{1-\phi_o} \\
\frac{\partial^2 \ell_o}{\partial a^2} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{o,it1}\left[\left(\frac{t}{a}\right)\frac{z^*_{a0t}\hat{w}_{it}w_{it}}{(\hat{w}_{it}+1)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial \beta_{\mu k_1}\partial \beta_{\mu k_2}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{o,it1}\left[\frac{z_{\mu k_1 it}z_{\mu k_2 it}\hat{w}_{itl}w_{it}}{(\hat{w}_{it}+1)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial a \partial \beta_{\mu k}} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{o,it1}\left[\frac{z^*_{a0t}z_{\mu kit}\hat{w}_{itl}w_{it}}{(\hat{w}_{it}+1)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial \phi_o^2} &= -\left[\frac{n_0}{\phi_o^2}+\frac{n_1}{(1-\phi_o)^2}\right] \\
\frac{\partial^2 \ell_o}{\partial a \partial \phi_o} &= \frac{\partial^2 \ell_o}{\partial \beta_{\mu k}\partial \phi_o}=0
\end{aligned}$$

where $z^*_{a0t}=\dfrac{t-1}{a}$, $\hat{w}_{it}=\dfrac{\mu_{it}}{a^{t-1}}$ and μ_{it} is given by (7.10).

For the simplified AZIPGP model, by differentiating the log-likelihood function $\ell_s(\boldsymbol{\theta}_s)$ in equation (7.18) once and twice, the derivative functions are given by:

$$
\begin{aligned}
\frac{\partial \ell_s}{\partial a} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{s,it1} z^*_{a0t}\left[\hat{w}_{it} - w_{it} + \frac{\hat{w}_{it}}{e^{\hat{w}_{it}-1}}\right] \\
\frac{\partial \ell_s}{\partial \beta_{\mu k}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{s,it1} z_{\mu k it}\left[\hat{w}_{it} - w_{it} + \frac{\hat{w}_{it}}{e^{\hat{w}_{it}-1}}\right] \\
\frac{\partial \ell_s}{\partial \phi_s} &= \frac{n_0}{\phi_s} - \frac{n_1}{1-\phi_s} \\
\frac{\partial^2 \ell_s}{\partial a^2} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} I_{s,it1}\frac{z^*_{a0t}}{a}\left[\frac{t\hat{w}_{it}}{1-e^{-\hat{w}_{it}}} - \frac{(t-1)\hat{w}^2_{it}e^{-\hat{w}_{it}}}{(1-e^{-\hat{w}_{it}})^2} - w_{it}\right] \\
\frac{\partial^2 \ell_s}{\partial \beta_{\mu k_1}\partial \beta_{\mu k_2}} &= -\sum_{i=1}^{m}\sum_{t=1}^{n_i} \frac{I_{s,it1} z_{\mu k_1 it} z_{\mu k_2 it}\hat{w}_{it}}{1-e^{-\hat{w}_{it}}}\left(1 - \frac{\hat{w}_{it}}{e^{\hat{w}_{it}}-1}\right) \\
\frac{\partial^2 \ell_s}{\partial a \partial \beta_{\mu k}} &= \sum_{i=1}^{m}\sum_{t=1}^{n_i} \frac{I_{s,it1} z^*_{a0t} z_{\mu k it}\hat{w}_{it}}{1-e^{-\hat{w}_{it}}}\left(1 - \frac{\hat{w}_{it}}{e^{\hat{w}_{it}}-1}\right) \\
\frac{\partial^2 \ell_s}{\partial \phi_s^2} &= -\left[\frac{n_0}{\phi_s^2} + \frac{n_1}{(1-\phi_s)^2}\right] \\
\frac{\partial^2 \ell_s}{\partial a \partial \phi_s} &= \frac{\partial^2 \ell_s}{\partial \beta_{\mu k}\partial \phi_s} = 0
\end{aligned}
$$

In particular, the parameter ϕ_h can be solved using the equation $\dfrac{\partial \ell_h}{\partial \phi_h} = 0$. Thus, $\hat{\phi}_h$ is obtained directly by:

$$
\frac{n_0}{\phi_h} - \frac{n_1}{1-\phi_h} = 0 \quad \Rightarrow \quad \phi_h = \frac{n_0}{n}
$$

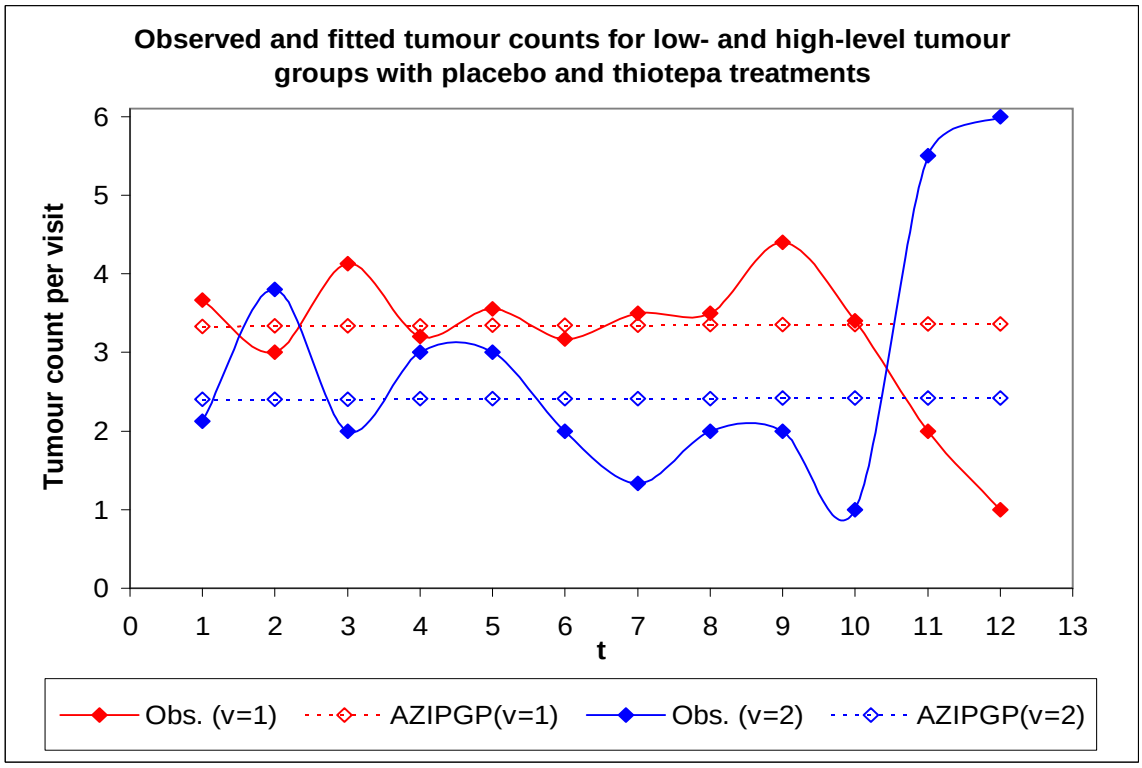
which is actually the proportion of zero observations in the count data W_{it}.

Table 7.1: Parameter estimates and standard errors for the bladder cancer data using AZIPGP models

AZIPGP								
	ML				BAY			
GOF		$\boldsymbol{U_z}$	$\boldsymbol{MSE_z}$			$\boldsymbol{U_z}$	$\boldsymbol{MSE_z}$	
Estimate	ϕ	$\beta_{\mu 0}$	$\beta_{\mu 1}$	a	ϕ	$\beta_{\mu 0}$	$\beta_{\mu 1}$	a
Original								
GOF		**−0.88232**	**1.2564**			**−0.88233**	**1.2531**	
	0.8039	1.2908	-0.4077	0.9953	0.8025	1.2990	-0.4049	0.9971
	0.0007	*0.3556*	*0.2510*	*0.0285*	*0.0168*	*0.3677*	*0.2569*	*0.0338*
Simplified								
GOF		**−0.91213**	**1.0654***			**−0.91215**	**1.0671**	
	0.8039	1.5315	-0.3276	0.9992	0.8029	1.5370	-0.3359	0.9998
	0.0007	*0.1881*	*0.1335*	*0.0174*	*0.0170*	*0.1896*	*0.1344*	*0.0175*

*Model with the smallest MSE_z is chosen to be the best model.

Figure 7.1: Plot of the the observed w_{tv} and expected $\hat{w}_{tv}$ tumour counts for the bladder cancer patients using simplified AZIPGP model with classical ML method



Remarks: (1) The fluctuation of w_{tv} for the two treatment groups is due to the large variation in the non-zero observations relative to the small number of non-zero observations in each visit t.

(2) An unnatural fluctuation of w_{tv} is observed at later visits ($t = 11, 12$) when the number of non-zero observations further drops to 3 and 1 for the placebo treatment group ($v = 1$) with 46 patients, and 2 and 2 for the thiotepa treatment group ($v = 2$) with 36 patients.

(3) The exceptionally high $w_{11,2}$ and $w_{12,2}$ are due to two strange observations from patient 66 with 8 and 9 tumours found in the 11^{th} and 12^{th} visits respectively relative to the small number of non-zero observations.

Chapter 8

Summary

8.1 Overview

A time series of count data which exhibits monotone increasing or decreasing trend patterns is commonly found in epidemiology, engineering, marketing, natural sciences, etc. The list of areas in which time series of count data are observed and analyzed is endless. To model such data, we propose the Poisson Geometric Process (PGP) model, which is extended from the GP model first proposed by Lam ((28), (29)) to model directly the monotone trend by a monotone process. Previous GP models limit to the modeling of positive continuous data such as the inter-arrival times of a series of events generated from a system. Recently, the GP model has been extended to model binary data, the resultant binary GP (BGP) model (Chan and Leung (9), Ho (25)) is useful as binary data is one of the common types of data in our daily lives. In this research, we aim at extending the GP model to Poisson count data with trends. Through simulation studies and real life applications in different fields, the PGP model and its extensions are demonstrated using different methodologies of inference.

As reviewed in Chapter 1, there exists two major categories of models, namely the observation-driven (OD) and parameter-driven (PD) models. In the OD models, the prediction of counts and the derivation of the likelihood function are

straightforward. However, difficulties arise when we interpret the model parameters and derive the statistical characteristics of the data. On the contrary, the PD models require considerable computation for parameter estimation due to the complicated derivation of the likelihood function. Besides, there exists difficulties in forecasting as the model is built on a latent process. The PGP model with original version is essentially a PD model with state variable X_t which is a latent GP, evolving independently of the past outcomes over time. Unlike most PD models, PGP model does not require complicated derivation of likelihood function. Moreover, it focuses on modeling the trend of the time series of counts whereas none of the models for time series of count data do so. Therefore, the PGP model provides an important class of alternative models for modeling time series of Poisson count data with trends.

In Chapter 3, two versions (original and simplified) of the PGP model are derived. The main difference between these two versions is that an exponential distribution is assigned to the renewal process (RP) $\{Y_t\}$ in the original version. Hence, the simplified PGP model is not a PD model. In terms of properties, equations (3.3) and (3.5) show that the variability of the count W_t is inflated in the original version which makes it suitable for modeling count data with trends and more zeros. In general, equations (3.3) and (3.5) also reveal that the ratio parameter a_t of a PGP model affects both mean and variance and allows them to change over time. Hence, the ratio parameter plays a significant role in the whole modeling methodologies and it reduces the PGP models to the generalized

linear mixed models (GLMMs) when $a_t = 1$. Another important parameter is μ_t which indicates the baseline level of the trend in the model.

Following the framework of the exponential family of distributions within the GLMMs, PGP models are extended to the adaptive PGP (APGP) model in Chapter 4 by incorporating time-evolving covariate effects in either or both mean function μ_t and ratio function a_t, by log-linking a linear function of covariates $z_{jkt}, j = \mu, a, k = 1, \ldots, q_j, t = 1, \ldots, n$ to the mean and/or ratio functions. In this way, the underlying process $\{Y_t\}$ is no longer identically and independently distributed (IID), instead, it becomes a stochastic process (SP) in general. The APGP model is so called as the ratio function allows a gradual change in both the direction and strength of the trend. For example, the gradual change of a_t from $a_t < 1$ to $a_t > 1$ indicates a gradual transition from a growing stage to a declining stage. Accompanying with the mean function, the APGP model is able to describe a great variety of trend patterns and hence allows a broad range of applications.

However, as the APGP models fail to accommodate sudden changes or abrupt turns in the trend, the threshold PGP (TPGP) model is proposed in Chapter 5 to capture multiple trend data that exhibits a growing, developing and declining stages of some processes. More specifically, it is suitable for modeling those trend data with discontinuous trends and abrupt patterns. In the TPGP and TAPGP models, the partial LSE and partial likelihood methods are adopted to determine the number and the location of the turning point(s). However, when the number

of trends becomes large, the number of possible combinations of turning points also increases substantially. Moreover, sometimes the number and the location of turning points are quite difficult to estimate from the observed data particularly if the trend has many abrupt turns. Hence, for further model improvement, other methods which can estimate the number and location of turning points more efficiently can be explored.

Different from the APGP model where a single PGP model is considered, the TPGP and/or TAPGP models allow the modeling of multiple trend data by using a series of PGP and/or APGP models. This gives us an insight to consider the application of PGP models to a parallel system. In Chaper 6, we extend the PGP model to longitudinal count data with trends. Since observations from longitudinal measurements often exhibit considerable population heterogeneity and clustering effects, we propose the mixture PGP (MPGP) model and the adaptive MPGP (MAPGP) model to identify the trend patterns for different subjects as well as classifying the subjects into groups by estimating the individual latent group membership. Due to the considerable complication of constraints in the probabilities of the group memberships when there are more than two groups in the model, an alternative estimation method namely the expectation maximization (EM) method is adopted in the implementation of the MPGP and MAPGP models. Constraints also exist for the two-group model, that is $p_1+p_2 = 1$ and $p_i \geq 0, i = 1, 2$. Even though we may set $p_2 = 1 - p_1$, we still have constraints $p_i \geq 0, i = 1, 2$. However, such constraint is more simple. The

EM method involves separate expectation (E) and maximization (M) steps and hence the convergence rate is slow. However, the log-likelihood function and its derivative functions in the M-step are much less complicated than the classical ML method. Hence, it is more preferred in the multi-group MAPGP model.

In Chapter 7, we further extend the PGP model to zero-inflated PGP (ZIPGP) model because when we deal with repeated measurements of count data, the situation of excess zeros often occurs in diverse applications. This situation is also observed in one of the real life applications of the MAPGP model, the bladder cancer data. As the excess zeros contribute substantial noise in the detection of monotone trends and may affect the estimation of the ratio a_{it} and the baseline level of the trend μ_{it} in general, the ZIPGP model, which is essentially a hurdle model, is proposed to explain those non-zero outcomes which can pass the hurdle ($W_t = 0$). For further model development, a zero-inflated mixture PGP (ZIMPGP) model can be investigated such that the excess zeros problem and the clustering effect in the data are taken into account at the same time.

As a whole, the PGP model takes into account a rather pronounced dependency structure and the extra variation relative to the mean of the time series. Moreover, the extension of the model and the model implementation are also straightforward and highly applicable. These no doubt make the PGP model outstanding among other time-series models for count data.

8.2 Further research

Despite of the substantial development of GP models in recent years, there are still many areas to explore for the GP model.

8.2.1 Extension to data with random effects

The inclusion of random effects into the model accounts for the over-dispersion and accommodates population heterogeneity inherent from clustering. We may consider a random effect model in which a random intercept and/or random coefficients are added to the linear function of predictors log-linked to the mean of $\{Y_t\}$. Most of the current works assume a normal distribution for the random effects. In practice, a wider class of random effects distributions that include heavy-tailed and skewed distributions may be more appropriate. Examples include the scale mixtures of normal (SMN) distributions which contain student-t and exponential power distributions and the scale mixtures of beta (SMB) which can be skewed.

8.2.2 Extension to bivariate and multivariate data

Previous GP models focused only on univariate data. In fact, we can investigate the possibility of extending the GP model to bivariate and multivariate data as they are common in our fields of studies, say in the medical science.

8.2.3 Robust modeling

Lam and Chan (35) and Chan *et al.* (8) considered the lognormal and gamma distributions for the SP $\{Y_t\}$ respectively. Although these distributions have been widely employed in modeling inter-arrival times or life-spans, there are a host of alternative distributions for $\{Y_t\}$. The SMN distributions are symmetric and can be more platykurtic and leptokurtic than the normal distribution. Moreover, some members in this class have infinite variance, for example, the Cauchy and stable distributions, and hence are suitable for modeling diversified effects. As the density functions can be written into scale mixture forms, model implementation are facilitated via hierarchical structures and the use of MCMC methods and Gibbs sampling. Outliers are automatically down-weighted by the mixing parameters in the scale mixtures to achieve model robustness. Other scale mixtures distributions including the scale mixtures of uniform (SMU) and SMB distributions are other possible options. Then sensitivity analysis can be performed to compare different choices of distributions and to identify some robust choices of distributions for $\{Y_t\}$.

8.2.4 Derivation of large sample distributions for model parameters

Large sample properties of the estimators using both non-parametric and parametric methods for different types of data can be studied by deriving their large

sample distributions. Knowing the asymptotic distributions, we can then construct confidence intervals and perform hypothesis tests on the model parameters.

8.2.5 New applications

Last but not least, apart from the extensions and improvements of the GP model, real applications play a significant role too. GP model can model data with different trend patterns more efficiently and give a better interpretation of the data. Hence, to further widen the scope of the applications of the GP models, we need to seek new applications in health science, social science, insurance, finance, engineering, etc. In this way, the GP models can become more beneficial to the communities as a whole.

References

[1] Akaike, Hirotugu, (1974), A new look at the statistical model indentification, *IEEE Transactions on Automatic Control*, **19(6)**, 716-723.

[2] Andrews, D.F. and Herzberg, A.M., (1985), *Data*, Springer, New York, 1985.

[3] Ascher, H. and Feingold, H. (1981), *Repairable Systems Reliability*, Marcel Dekker, New York.

[4] Bollerslev, T. (1986), Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, **31**, 307-327.

[5] Chan, J.S.K. (2000), Initial stage problem in autoregressive binary regression, *The Statistician*, **49**, 495-502.

[6] Chan, J.S.K., Kuk, A.Y.C. and Bell, J. (1997), A Likelihood approach to analysing longitudinal bivariate binary data, *Biometrical Journal*, **39**, No.4, 409-421.

[7] Chan, J.S.K., Kuk, A.Y.C., Bell, J. and McGilchrist, C. (1998), The analysis of methadone clinic data using marginal and conditional logistic models with mixture or random effects. *Australian and New Zealand Journal of Statistics*, **40**, No.1, 1-10.

[8] Chan, J.S.K., Lam, Y. and Leung, D.Y.P. (2004), Statistical inference for geometric processes with gamma distributions, *Computation Statistics and Data Analysis*, **47**, 565-581.

[9] Chan, J.S.K. and Leung, D.Y.P. (2006), A new approach to the modeling

of longitudinal binary data with trend: the Binary Geometric Process model. Submitted for publication.

[10] Chan, J.S.K., Yu, P.L.H., Lam, Y. and Ho, A.P.K. (2006), Modeling SARS data using threshold geometric process. *Statist. Med. 2006*, **25**, 1826-1839.

[11] Chang, T.J., Kavvas, M.L. and Delleur, J.W. (1984), Daily precipitation modeling by discrete autoregressive moving average processes, *Water Resources Research*, **20**, 565-580.

[12] Cox, D.R. and Lewis, P.A. (1966), *The Statistical Analysis of Series of Events*, Mathuen, London.

[13] Cox, D.R. (1981), Statistical Analysis of Time Series: Some Recent Developments, *Scandinavian Journal of Statistics*, **8**, 93-115.

[14] Davis, R.A., Dunsmuir, W.T.M. and Wang, Y. (1999), Modelling time series of counts, *Asymptotics, nonparametrics, and time series* (Subir Ghosh, editor), Marcel Dekker, New York, 63-114.

[15] Davis, R.A., Dunsmuir, W.T.M. and Streett, S.B. (2003), Observation-driven models for Poisson counts, *Biometrika*, **90**, 777-790.

[16] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society B*, **39**, 599-613.

[17] Drescher, D. (2005), Alternative distributions for observation driven count series models, *Economics Working Paper No. 2005-11*, Christian-

Albrechts-Universität, Kiel, Germany.

[18] Durbin, J. and Koopman, S.J. (2000), Time series analysis of non-Gaussian observations based on state space models from both classical and Baye-sain perspectives, *Journal of Royal Statististical Society, Series B*, **62**, 3-56.

[19] Feller, A. (1949), *Trans. Amer. Math. Soc.*, **67**, 98-119.

[20] Gilks, W.R., Richardson, S. and Spiegelhalter, D.J. (1996), *Markov Chain Monte Carlo in Practice*, Chapman and Hall, UK.

[21] Gurmu, S. (1997), Semi-parametric estimation of hurdle regression models with an application to Medicaid utilization, *Journal of Applied Econo-metrics*, **12**, 225-242.

[22] Hand, D.J., Daly, F., Lunn, A.D., McConway, K.J. and Ostrowski, E. (1994), *A Handbook of Small Data Sets*, Chapman and Hall, London.

[23] Hastings, W.K. (1970), Monte Carlo sampling methods using Markov Chains and their applications, *Biometrika*, **57**, 97-109.

[24] Heinen, A. (2003) Modelling time series count data: An autoregressive conditional Poisson model, *Core discussion paper No. 2003-63.*

[25] Ho, A.P.K. (2005), Parametric and Non-parametric inference for geometric process, M. Phil. Thesis, Department of Statistics and Actuarial Science, The University of Hong Kong.

[26] Jung, R.C. and Liesenfeld, R. (2001), Estimating time series models for count data using efficient importance sampling, *Allgemeines Statistisches*

Archiv, **85**, 387-407.

[27] Jung, R.C., Kukuk, M. and Liesenfeld, R. (2005), Time Series of Count Data: Modelling and Estimation, *Economics Working Paper No. 2005-08*, Christian-Albrechts-Universität, Kiel, Germany.

[28] Lam, Y. (1988), A note on the optimal replacement problem, *Advances in Applied Probability*, **20**, 479-482.

[29] Lam, Y. (1988), Geometric process and replacement problem, *Acta Mathematicae Applicatae Sinica* **4**, 366-377.

[30] Lam, Y. (1992), Optimal geometric process replacement model, *Acta Mathematicae Applicatae Sinica* **8**, 73-81.

[31] Lam, Y. (1992), Nonparametric inference for geometric processes, *Communication Statistics: Theory and Methodology*, **21**, 2083-2105.

[32] Lam, Y. (1995), Calculating the rate of occurrence of failures for continuous-time Markov chains with application to a two-component parallel system, *Journal of Operation Research Society*, **45**, 528-536.

[33] Lam, Y. (1997), The rate of occurence of failures, *J. Appl. Prob.*, **34**, 234-247.

[34] Lam, Y. (2004), Geometric process, *The Encyclopedia of Statistical Sciences, 2nd edition*, Balakrishnan, N., Read, C., Kotz, S. and Vidakovic, B. ed. John Wiley & Sons, Inc., New York.

[35] Lam, Y. and Chan, J.S.K. (1998), Statistical inference for geometric processes with lognormal distribution, *Computational Statistics and Data Analysis*, **27**, 99-112.

[36] Lam, Y. and Zhang, Y.L. (1996), Analysis of a two-component series system with a geometric process model, *Naval Research Logistics*, **43**, 491-502.

[37] Lam, Y. and Zhang, Y.L. (1996), Analysis of a parallel system with two different units, *Acta Mathematicae Applicatae Sinica*, **12**, 408-417.

[38] Lam, Y. and Zhang, Y.L. (2003), A geometric process maintenance model for a deteriorating system under a random environment, *IEEE Trans. Reliab.*, **52**, 83-89.

[39] Lam, Y., Zhang, Y.L. and Zheng, Y.H. (2002), A geometric process equivalent model for a multistate degenerative system, *European Journal of Operational Research*, **142**, 21-29.

[40] Lam, Y., Zhu, L.X., Chan, J.S.K. and Liu, Q. (2004), Analysis of data from a series of events by a Geometric Process model, *Acta Mathematica Applicatae Sinica, English Series*, **20**, 263-282.

[41] Lambert, D. (1992), Zero-inflated Poisson regression, with an application to defects in manufacturing, *Technometrics*, **34**, 1-14.

[42] Maguire, B.A., Pearson, E.S. and Wynn, A.H.A. (1952), The time intervals between industrial accidents, *Biometrika*, **39**, 168-180.

[43] McKenzie, E. (1985), Some simple models for discrete variate time series. *Water Resource Bulletin*, **21**, 645-650

[44] McKenzie, E. (2003), Discrete variate time series. In Shanbhag, D.N. and Rao, C.R. (Eds.) *Handbook of Statistics*, **21**, 573-606, Elsevier, Amsterdam.

[45] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N. and Teller, A.H. (1953), Equations of State calculations by fast computing machines, *Journal of Chemical Physics*, **21**, 1087-1091.

[46] Miaou, S.-P. (1994), The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions, *Accident Analysis and Prevention*, **26**, 471-482.

[47] Min, Y. and Agresti, A. (2005), Random effect models for repeated measures of zero-inflated count data, *Statistical Modelling 2005*, **5**, 1-19.

[48] Mullahy, J. (1986), Specification and testing of some modified count data models, *Journal of Econometrics*, **33**, 341-365.

[49] Nelder, J.A. and Wedderburn, R.W.M. (1972), Generalized linear models, *Journal of the Royal Statistical Society A*, **135**, 370-384.

[50] Ownby, H.E., Kong, F. and Watanabe, K. (1999), Analysis of donor return behavior. Retrovirus Epidemiology Donor Study, *Transfusion*, **39(10)**,1128-35.

[51] Ridout, M., Demétrio C. and Hinde, J. (1998), Models for count data with many zeros, *International Biometric Conference*, Cape Town, December 1998.

[52] Schreiber, G.B., Sanchez, A.M., Glynn, S.A. and Wright, D.J. (2003), Increasing blood availability by changing donation patterns, *Transfusion*, **43(5)**, 591-7.

[53] Schreiber, G.B., Sharma, U.K., Wright, D.J. (2005), First year donation

patterns predict long-term commitment for first-time donors, *Vox Sang*, **88(2)**,114-21.

[54] Sheu, S.H. (1999), Extended optimal replacement model for deteriorating systems, *European Journal of Operational Research*, **112**, 503-516.

[55] Sheynin, O.B. (1988), A A Markov's Work on Probability, *Archive for History of Exact Science*, **39**, 337-377.

[56] Sheynin, O.B. (1989), Errata for the contribution: 'A A Markov's work on probability', *Archive for History of Exact Science*, **40(4)**, 387.

[57] Shonkwiler, J. and Shaw, W. (1996), Hurdle count-data models in recreation demand analysis, *Journal of Agricultural and Resource Economics*, **21**, 210-219.

[58] Smith, A.F.M. and Roberts, G.O. (1993), Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *Journal of the Royal Statistical Society B*, **55**, 3-23.

[59] Spiegelhalter, D., Thomas, A. and Best, N. (2000), Bayesian inference using Gibbs sampling, Window version WinBUGS, from http://www.mrc-bsu.cam.ac/bugs.

[60] Taylor, S.J. (1986), *Modelling Financial Time Series*, John Wiley: Chichester.

[61] Walker, S. G. and Gutiérrez-Peña, E. (1999), Robustifying Bayesian Procedures, *Bayesian Statistics*, **6**, 685-710.

[62] Wu, Y., Glynn, S.A. and Schreiber, G.B. (2001), First-time blood donors:

demographic trends, *Transfusion*, **41(3)**,360-4.

[63] Yu, P.L.H., Chung, K.H., Lee, C.K., Lin, C.K. and Chan, J.S.K. (2006), Predicting drop-out and committed first time blood donors: the case in Hong Kong Chinese, submitted for publication.

[64] Zeger, S.L. (1988), A regression model for time series of counts, *Biometrika*, **75**, 621-629.

[65] Zhang, Y.L. (1999), An optimal geometric process model for a cold standby repairable system, *Reliability Engineering and Systems Safety*, **63**, 107-110.

[66] Zhang, Y.L. (2002), A geometric-process repair-model with good-as-new preventive repair, *IEEE Trans. Reliab.*, **51**, 223-228.

[67] Zhang, Y.L., Yam, R.C.M. and Zuo, M.J. (2001), Optimal replacement policy for a deteriorating production system with preventive maintenance, *Int. J. Syst. Sci.*, **32**, 1193-1198.

[68] Zhang, Y.L., Yam, R.C.M. and Zuo, M.J. (2002), Optimal replacement policy for multistate repairable system, *Journal of Operational Research Society*, **53**, 336-341.