

PROOF COVER SHEET

Author(s): Jennifer So Kuen Chan

Article title: Bayesian analysis of Cannabis offences using generalized Poisson geometric process model with flexible dispersion

Article no: GSCS1167211

Enclosures: 1) Query sheet
2) Article proofs

Dear Author,

1. Please check these proofs carefully. It is the responsibility of the corresponding author to check these and approve or amend them. A second proof is not normally provided. Taylor & Francis cannot be held responsible for uncorrected errors, even if introduced during the production process. Once your corrections have been added to the article, it will be considered ready for publication.


Please limit changes at this stage to the correction of errors. You should not make trivial changes, improve prose style, add new material, or delete existing material at this stage. You may be charged if your corrections are excessive (we would not expect corrections to exceed 30 changes).

For detailed guidance on how to check your proofs, please paste this address into a new browser window: <http://journalauthors.tandf.co.uk/production/checkingproofs.asp>

Your PDF proof file has been enabled so that you can comment on the proof directly using Adobe Acrobat. If you wish to do this, please save the file to your hard disk first. For further information on marking corrections using Acrobat, please paste this address into a new browser window: <http://journalauthors.tandf.co.uk/production/acrobat.asp>

2. Please review the table of contributors below and confirm that the first and last names are structured correctly and that the authors are listed in the correct order of contribution. This check is to ensure that your name will appear correctly online and when the article is indexed.

Sequence	Prefix	Given name(s)	Surname	Suffix
1.		Jennifer So Kuen	Chan	
2.		Wai Yin	Wan	

Queries are marked in the margins of the proofs, and you can also click the hyperlinks below. Content changes made during copy-editing are shown as tracked changes. Inserted text is in **red font** and revisions have a red indicator . Changes can also be viewed using the list comments function. To correct the proofs, you should insert or delete text following the instructions below, but **do not add comments to the existing tracked changes**.

AUTHOR QUERIES

General points:

- (1) **Permissions:** You have warranted that you have secured the necessary written permission from the appropriate copyright owner for the reproduction of any text, illustration, or other material in your article. Please see <http://journalauthors.tandf.co.uk/permissions/usingThirdPartyMaterial.asp>.
- (2) **Third-party content:** If there is third-party content in your article, please check that the rightsholder details for re-use are shown correctly.
- (3) **Affiliation:** The corresponding author is responsible for ensuring that address and email details are correct for all the co-authors. Affiliations given in the article should be the affiliation at the time the research was conducted. Please see <http://journalauthors.tandf.co.uk/preparation/writing.asp>.
- (4) **Funding:** Was your research for this article funded by a funding agency? If so, please insert ‘This work was supported by <insert the name of the funding agency in full>’, followed by the grant number in square brackets ‘[grant number xxxx]’.
- (5) **Supplemental data and underlying research materials:** Do you wish to include the location of the underlying research materials (e.g. data, samples or models) for your article? If so, please insert this sentence before the reference section: ‘The underlying research materials for this article can be accessed at <full link>/ description of location [author to complete]’. If your article includes supplemental data, the link will also be provided in this paragraph. See <<http://journalauthors.tandf.co.uk/preparation/multimedia.asp>> for further explanation of supplemental data and underlying research materials.
- (6) The **CrossRef database** (www.crossref.org/) has been used to validate the references. Mismatches will have resulted in a query.

QUERY NO.	QUERY DETAILS
AQ1	Please provide the missing Department name for Jennifer So Kuen Chan’s affiliation.
AQ2	The CrossRef database (www.crossref.org/) has been used to validate the references. Mismatches between the original manuscript and CrossRef are tracked in red font. Please provide a revision if the change is incorrect. Do not comment on correct changes
AQ3	Ref. “[35]” is listed in the references list but is not cited in the text. Please either cite the reference or remove it from the references list.
AQ4	Ref. “[36]” is listed in the references list but is not cited in the text. Please either cite the reference or remove it from the references list.
AQ5	Ref. “[37]” is listed in the references list but is not cited in the text. Please either cite the reference or remove it from the references list.

How to make corrections to your proofs using Adobe Acrobat/Reader

Taylor & Francis offers you a choice of options to help you make corrections to your proofs. Your PDF proof file has been enabled so that you can edit the proof directly using Adobe Acrobat/Reader. This is the simplest and best way for you to ensure that your corrections will be incorporated. If you wish to do this, please follow these instructions:

1. Save the file to your hard disk.
2. Check which version of Adobe Acrobat/Reader you have on your computer. You can do this by clicking on the “Help” tab, and then “About”.

If Adobe Reader is not installed, you can get the latest version free from <http://get.adobe.com/reader/>.

3. If you have Adobe Acrobat/Reader 10 or a later version, click on the “Comment” link at the right-hand side to view the Comments pane.

4. You can then select any text and mark it up for deletion or replacement, or insert new text as needed. Please note that these will clearly be displayed in the Comments pane and secondary annotation is not needed to draw attention to your corrections. If you need to include new sections of text, it is also possible to add a comment to the proofs. To do this, use the Sticky Note tool in the task bar. Please also see our FAQs here: <http://journalauthors.tandf.co.uk/production/index.asp>.

5. Make sure that you save the file when you close the document before uploading it to CATS using the “Upload File” button on the online correction form. If you have more than one file, please zip them together and then upload the zip file.

If you prefer, you can make your corrections using the CATS online correction form.

Troubleshooting

Acrobat help: <http://helpx.adobe.com/acrobat.html>

Reader help: <http://helpx.adobe.com/reader.html>

Please note that full user guides for earlier versions of these programs are available from the Adobe Help pages by clicking on the link “Previous versions” under the “Help and tutorials” heading from the relevant link above. Commenting functionality is available from Adobe Reader 8.0 onwards and from Adobe Acrobat 7.0 onwards.

Firefox users: Firefox’s inbuilt PDF Viewer is set to the default; please see the following for instructions on how to use this and download the PDF to your hard drive: http://support.mozilla.org/en-US/kb/view-pdf-files-firefox-without-downloading-them#w_using-a-pdf-reader-plugin



Bayesian analysis of Cannabis offences using generalized Poisson geometric process model with flexible dispersion

Jennifer So Kuen Chan^a and Wai Yin Wan^b

^aThe University of Sydney, Sydney, Australia; ^bCentre for Education Statistics and Evaluation, Sydney, Australia

ABSTRACT

This paper derives models to analyse Cannabis offences count series from New South Wales, Australia. The data display substantial overdispersion as well as underdispersion for a subset, trend movement and population heterogeneity. To describe the trend dynamic in the data, the Poisson geometric process model is first adopted and is extended to the generalized Poisson geometric process model to capture both over- and underdispersion. By further incorporating mixture effect, the model accommodates population heterogeneity and enables classification of homogeneous units. The model is implemented using Markov chain Monte Carlo algorithms via the user-friendly WinBUGS software and its performance is evaluated through a simulation study.

ARTICLE HISTORY

Received 17 March 2015
Accepted 14 March 2016

KEYWORDS

Generalized Poisson distribution; geometric process model; Markov chain Monte Carlo algorithm; mixture model; over and underdispersion

1. Introduction

This research is motivated by a cannabis data consisting of 344 yearly number of arrests for cannabis dealing and trafficking from 2001 to 2008 in 43 local government areas (LGAs) within Sydney, the metropolitan region of NSW with the highest prevalence of drug dealing and trafficking amongst all NSW divisions. The aim is to evaluate the competence of a Cannabis Cautioning Scheme (CCS) in reducing the number of cannabis users by examining the trend movement of cannabis offense counts, allowing for the differences across LGAs. As a whole, the data clearly display substantial overdispersion. However when the analysis is performed to a subset of LGAs with low levels of cannabis offences, the data exhibit low level of dispersion, ranging from equidispersion to underdispersion. As Poisson model is well known for its deficiency to describe overdispersed or underdispersed count data, we aim to derive appropriate models that can accommodate different levels of dispersion, allow for the trend movement of cannabis offense count and capture population heterogeneity.

To handle overdispersion, Greenwood and Yule [1] consider a mixed Poisson distribution with a gamma mixing distribution to result in the well-known negative binomial distribution. Since then, a number of mixing distributions are considered, including the Poisson-inverse Gaussian mixed model, [2] Poisson-lognormal mixed model [3] and Poisson-generalized gamma mixed model. [4] See Karlis and Meligkotsidou [5] for a detailed review. Wan and Chan [6] propose a Poisson mixed model using Student's t or exponential power (EP) mixing distributions. However, as these mixing distributions add extra variation to Poisson distribution and their leptokurtic shapes can handle overdispersion caused by extreme observations, these Poisson mixed distributions are overdispersed and fail to deal with underdispersion.

Compared to overdispersion, the analogous problem of underdispersion has not received much attention as it is less frequently observed in longitudinal panel count data. However due to certain

similar conditions, the cannabis offences in a subgroup of LGAs display underdispersion. Amongst the limited literature, Faddy [7] proposes the birth process model based on a generalization of a Poisson process. In the model, overdispersion or underdispersion is attained when the rate of birth process is monotone increasing or decreasing with time, respectively. Besides, alternative distributions that allow underdispersion including Conway–Maxwell–Poisson distribution,[8,9] generalized Poisson (GP) distribution,[10] double Poisson distribution,[11] weighted Poisson distribution [12] and generalized Waring distribution [13] are considered. Amongst these distributions, the GP distribution pioneered by Consul and Jain [10] receives more attention. Relative to Poisson distribution, the two-parameter GP distribution contains an additional shape parameter which allows the variance of the distribution to be more than, equal to or less than the mean when it is positive, zero or negative, respectively. Li et al. [14] propose the quasi-negative binomial (QNB) distribution with GP distribution as a limiting distribution when two of its parameters approach infinity in a ratio and this ratio ψ forms the scale parameter of GP distribution. It also gives negative binomial (NB) distribution as a special case when a shape parameter $c = 0$. Although it is more general and flexible particularly for highly skewed data, its mean, variance and higher moments do not exist if $c > 0$.

Apart from levels of dispersion, trend movement is another feature to be addressed and we adopt the geometric process (GP) model, pioneered by Lam,[15,16] to model the trend dynamics. A sequence of positive observations $\{X_t\}$ such as interarrival times follow a GP if there exists a ratio $a > 0$ such that the sequence $\{Y_t\}$ where $Y_t = a^{t-1}X_t$ becomes stationary and can be modeled by a lifetime distribution with mean μ and variance σ^2 . The parameter a indicates direction of trend movement such that $\{X_t\}$ is increasing, stationary or decreasing when $a < 1$, $a = 1$ and $a > 1$, respectively whereas μ represents the underlying mean. This two-component GP model has a nice interpretation for covariates that describe effects on the underlying stationary mean and trend movement separately. The model has been extended to different levels of measurements including binary [17] and counts [18] and applied to various fields, including industrial maintenance and replacement,[19–21] medical [17,22] and finance.[23] Moreover, the Poisson GP (PGP) model is further extended to incorporate covariate effects, trend movement, population heterogeneity and zero inflation in [18], overdispersion in [6] and multivariate counts in [24].

To derive models for the cannabis offense counts with different levels of dispersion, we advance the PGP model by replacing the Poisson data distribution with the GP distribution. The resultant model, called generalized Poisson geometric process (GPGP) model, is further extended to incorporate covariate and mixture effects to capture population heterogeneity. The model is implemented using a Bayesian approach via the Markov chain Monte Carlo (MCMC) method. Through a simulation experiment, the GPGP model shows an improvement to the robust PGP (RPGP) model of Wan and Chan [6] under different levels of dispersion.

Then two analyses are performed to analyse the policy effect in the cannabis data using the whole data which show overdispersion and a subset with underdispersion. In fitting the whole data with substantial population heterogeneity, the first analysis adopt mixture and random effects models and hence compares our proposed mixture GPGP (MGPGP) model to two conventional models, the mixture GP (MGP) model and the Poisson random effects (PRE) model, both without a GP modelling. By comparing the MGPGP model to the MGP model, we can evaluate the modelling strategy for trended data using either a time effect covariate or a GP model. On the other hand, comparing the MGPGP model to the PRE model allows us to investigate random effects model and our proposed mixture model in modelling population heterogeneity. The second analysis looks at specifically those LGAs with low arrest counts displaying underdispersion and investigates if the overall trend pattern that reviews policy effectiveness applies to these LGAs. The performance of GPGP model for under-dispersed data is evaluated and compared to random effect models with different data distributions. From the simulation study and model comparisons in data analyses, we confirm that the GPGP model and the MGPGP model are promising models for the evaluation of policy competence in cannabis data.

The paper is presented as follows. Firstly, the motivating data of the number of arrest of cannabis users is introduced in Section 2. Then the development of GP model to our proposed GPGP and MGPGP models is described in Section 3. Section 4 discusses model implementation using MCMC algorithms and introduces model assessment criteria. Then a simulation experiment comparing the performance of GPGP and RPGP models is performed in Section 5 and the data analyses are conducted in Section 6. Lastly a brief conclusion is made in Section 7.

2. Cannabis data

Cannabis is a recreational drug with few toxic acute effects compared to other illicit drugs. Cannabis abusers and quitters experience paranoia and other withdrawal symptoms which make them feel irritated and aggressive and hence they are more prone to assaults and violent behaviours. To check the mounting crime rates, the NSW Police and Drug and Alcohol Coordination initiated the CCS in April 2000 to assist offenders to consider the legal and health ramifications of their cannabis use.[25] By providing treatment and support to offenders, the scheme is expected to reduce the number of cannabis users throughout the state but this effect still remains sceptical.[25]

To evaluate competence of the policy, we analyses the cannabis data which contains 344 yearly number of arrests for cannabis dealing and trafficking from 2001 to 2008 in 43 LGAs within Sydney, the metropolitan region of NSW with the highest prevalence of drug dealing and trafficking amongst all NSW divisions. This results in 43 time series, each containing eight yearly counts as reported in Table 1. The data are available on the official website of the NSW Bureau of Crime Statistics and Research (http://www.bocsar.nsw.gov.au/lawlink/bocsar/ll_bocsar.nsf/pages/bocsar_onlinequeries).

3. Model development

We aim to develop appropriate models for the analyses of cannabis offenses that cater for a wide range of dispersion and accommodate the trend movement from 2001 to 2008 to reveal policy effectiveness. Moreover, the data clearly exhibit cluster effects due to the fact that some LGAs such as Sydney, Bankstown, Blacktown and Campbelltown, have substantially higher number of arrests for cannabis dealing and trafficking throughout the study period. Therefore, we consider the mixture model which postulates that there exists G subgroups of LGAs each with distinct level of arrests, dispersion and trend pattern.

3.1. Trend data

Lam [15,16] first proposed to model positive continuous data with monotone trend by a monotone process called the geometric process (GP) defined as

Definition 1: Let X_1, X_2, \dots be a sequence of non-negative 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 renewal process (RP), then the stochastic process $\{X_t, t = 1, 2, \dots\}$ is called a geometric process (GP) and the real number a which can be interpreted as the unit-time discount rate from an observed process to a stationary process is called the ratio of the GP.

This stationary process $\{Y_t\}$ may follow some parametric distributions $f(y_t)$ with $E(Y_t) = \mu$ and $\text{Var}(Y_t) = \sigma^2$. Hence, the mean and variance of the GP are

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

Table 1. Dealing and trafficking of Cannabis in Sydney during 2001–2008.

		Year							
Statistical subdivision	LGA	2001	2002	2003	2004	2005	2006	2007	2008
Inner Sydney	*Botany Bay	6	2	4	2	3	1	1	2
	*Leichhardt	2	1	1	2	0	1	1	2
	Marrickville	7	5	6	4	7	8	5	5
	Sydney [†]	143	114	88	85	69	44	46	57
Eastern Suburbs	Randwick	8	3	6	8	8	2	4	1
	Waverley	4	7	13	6	12	1	2	3
	*Woollahra	2	0	0	1	0	0	0	4
St George-Sutherland	Hurstville [†]	4	8	10	3	2	3	5	13
	Kogarah	4	2	5	4	4	3	2	1
	Rockdale	5	4	6	6	5	3	2	5
	Sutherland Shire [†]	7	8	11	16	9	9	3	9
Canterbury-Bankstown	Bankstown [†]	22	17	12	20	22	13	19	15
	Canterbury [†]	18	9	13	16	17	22	17	26
Fairfield-Liverpool	Fairfield [†]	18	13	5	17	24	12	9	6
	Liverpool [†]	47	9	13	15	17	8	14	5
Outer South Western Sydney	*Camden	1	2	2	2	1	2	1	3
	Campbelltown [†]	18	22	13	16	34	35	5	16
	*Wollondilly	5	3	0	2	1	0	0	1
Inner Western Sydney	Ashfield	4	5	6	2	2	1	1	2
	*Burwood	1	2	2	5	2	1	2	4
	*Canada Bay	3	2	0	5	1	0	2	2
	*Strathfield	6	3	0	3	1	0	1	1
Central Western Sydney	Auburn	2	0	3	2	7	5	3	6
	Holroyd [†]	6	11	3	9	5	9	3	7
	Parramatta [†]	22	15	9	18	15	12	8	7
Outer Western Sydney	Blue Mountains	3	5	3	4	0	2	6	2
	Hawkesbury	9	6	4	2	7	3	2	3
	Penrith [†]	30	5	13	10	18	5	3	8
Blacktown	Blacktown [†]	30	25	26	19	14	16	13	5
Lower Northern Sydney	*Hunters Hill	0	0	2	0	1	0	0	0
	*Lane Cove	0	0	0	0	0	0	0	1
	*Mosman	2	0	1	0	4	1	0	0
	*North Sydney	4	5	1	2	2	2	1	1
	Ryde	10	11	3	3	1	4	0	1
Central Northern Sydney	*Willoughby	3	0	1	2	2	0	1	0
	Baulkham Hills [†]	11	2	2	2	5	14	12	3
	Hornsby	8	6	7	6	4	4	0	3
	*Ku-ring-gai	2	0	3	0	2	3	1	0
Northern Beaches	*Manly	3	0	1	3	1	4	0	5
	*Pittwater	1	1	6	0	1	1	0	1
	Warringah	5	10	2	3	4	2	2	7
Gosford-Wyong	Gosford [†]	9	11	7	26	6	5	3	4
	Wyong [†]	5	4	21	5	13	7	6	4
mean	Mean	11.63	8.33	7.77	8.28	8.21	6.23	4.79	5.84
	variance	512.9	306.9	190.2	187.7	148.2	81.4	63.7	88.9

*LGA selected in the second analysis.

[†] LGA classified to high-level group by MGP using MGPGP models.

respectively. The ratio parameter a allows heteroskedasticity and reveals the direction and strength of a trend movement: a GP is monotonic increasing if $a < 1$, monotonic decreasing if $a > 1$ and stationary if $a = 1$. While common time series model adds a time covariate, say $\log t$, in the mean μ to

describe a trend movement, the GP model has two components, μ and a , and hence can adopt different covariates that affect the underlying stationary mean process and trend movement separately. Adopting log link functions with no covariate in the mean, it is clear that

$$\begin{aligned} \text{GP model: } E(X_t) &= \exp[\beta_{\mu 0} - (t-1)\beta_{a0} - \beta_{a0}(t-1)\log t], \\ \text{Common model: } E(X_t) &= \exp(\beta_{\mu 0} + \beta_{\mu 1}\log t). \end{aligned} \quad (2)$$

Hence, the GP model can describe non-monotonic trend movement, same as some common time series models if they contain a quadratic term $(\log t)^2$. However, if they are expressed as Equation (2), they provide a nice interpretation in terms of the ratio a which reveals features of the trend dynamics.

Original GP model was applied to analyse interarrival times of events in system maintenance problem. To extend the GP model to longitudinal count of cannabis offenses from multiple LGAs, we assume that the count W_{it} for the i th LGA in year t , $i = 1, \dots, m$, $t = 1, \dots, n_i$ ($n = \sum_{i=1}^m n_i$) follow a Poisson distribution $f_P(\cdot | X_{it})$ with mean X_{it} forming a latent GP. Here $m = 43$ and $n_i = 8$ for the cannabis data. Then, the corresponding latent RP $\{Y_{it} = a_i^{t-1} X_{it}\}$ with mean $E(Y_{it}) = \mu_i$ can adopt some lifetime distributions. The resultant model is called the Poisson GP (PGP) model. Wan and Chan [18] consider gamma distribution $G(r, \frac{r}{\mu_i})$ for Y_{it} with the marginal probability mass function (pmf), the mean and variance for W_{it} given by

$$\begin{aligned} f(w_{it}) &= \frac{\Gamma(r + w_{it})}{w_{it}! \Gamma(r)} \frac{\left(\frac{\mu_i}{ra_i^{t-1}}\right)^{w_{it}}}{\left(1 + \frac{\mu_i}{ra_i^{t-1}}\right)^{r+w_{it}}}, \\ E(W_{it}) &= \frac{\mu_i}{a_i^{t-1}} \quad \text{and} \quad \text{Var}(W_{it}) = \frac{\mu_i}{a_i^{t-1}} + \frac{1}{r} \left(\frac{\mu_i}{a_i^{t-1}}\right)^2. \end{aligned}$$

respectively. Essentially, W_{it} follow a negative binomial (NB) distribution, approaching Poisson distribution as r tends to infinity. The PGP model is essentially a state space model with state variables X_{it} which form a latent GP.

3.2. Over- and underdispersion

Although the PGP model with mixed Poisson-gamma distribution allows overdispersion, it is still incapable to accommodate extreme outliers. Hence, Wan and Chan [6] further propose the RGP model using heavy-tailed mixing distributions for $Y_{it}^* = \ln Y_{it}$ with pdf $f(y_{it}^*)$. Then, the marginal pmf for W_{it} is

$$f(w_{it}) = \int_0^\infty \frac{\exp\left(-\frac{e^{y_{it}^*}}{a_i^{t-1}}\right) \left(\frac{e^{y_{it}^*}}{a_i^{t-1}}\right)^{w_{it}}}{w_{it}!} f(y_{it}^*) dy_{it}^*. \quad (3)$$

In particular, if Y_{it}^* has an EP distribution with pdf

$$f(y_{it}^*) = \frac{c_1}{\sigma} \exp \left\{ - \left| \frac{c_0^{1/2} (y_{it}^* - \mu_{it}^*)}{\sigma} \right|^{2/\nu} \right\}$$

where $c_0 = \Gamma(3\nu/2) / \Gamma(\nu/2)$, $c_1 = c_0^{1/2} / [\nu \Gamma(\nu/2)]$ and $\nu \in (0, 2]$ is a shape parameter which controls the kurtosis and hence the degree of overdispersion, the resultant model is named as the RGP-EP model and it can effectively downweigh the effects of extreme observations. The EP family subsumes a range of symmetric distribution such as uniform ($\nu \rightarrow 0$) with kurtosis equal to 1.8,

normal ($\nu = 1$) with kurtosis equal to 3 and double exponential ($\nu = 2$) with kurtosis of 6. To facilitate the model implementation using MCMC algorithms, the EP distribution is represented in scale mixtures of uniforms [26] and the marginal pmf for W_{it} in Equation (3) is written as

$$\int_0^\infty \frac{\exp\left(-\frac{y_{it}^*}{a_i^{t-1}}\right) \left(\frac{y_{it}^*}{a_i^{t-1}}\right)^{w_{it}}}{w_{it}!} \int_0^\infty f_U(\mu_{it}^* - \sigma \zeta_{it}^{\nu/2}, \mu_{it}^* + \sigma \zeta_{it}^{\nu/2}) f_G\left(\zeta_{it} \left| 1 + \frac{\nu}{2}, \frac{1}{2} \right.\right) d\zeta_{it} dy_{it}^*,$$

where $f_U(\cdot | b, c)$ represents a uniform distribution on the interval (b, c) , μ_{it}^* is the mean function for Y_{it}^* and ζ_{it} is a mixing parameter which controls the range of the uniform distribution. Hence the larger the ζ_{it} , the wider is the range to accommodate potential outliers and downweigh their outlying effect on inflating σ . As a result, ζ_{it} can serve as indicators for outliers in outlier diagnosis. Similarly larger ν also inflates the range to accommodate outliers. Essentially it gives heavier tail and hence more overdispersion for the data distribution. Refer to Wan and Chan [6] for more details.

The PGP and RGP models which can deal with overdispersed data fail to model underdispersed data. Therefore, we propose to generalize the PGP model by replacing the Poisson distribution with the GP distribution introduced by Consul and Jain. [10] This model has a more flexible variance structure as it allows both overdispersion and underdispersion. The GP distribution was investigated in details in [27]. If a random variable W follows a GP distribution, its pmf is given by

$$f_{GP}(w|\psi, \lambda) = \begin{cases} \frac{\psi(\psi + \lambda w)^{w-1} \exp[-(\psi + \lambda w)]}{w!}, & w = 0, 1, \dots; \\ 0, & w > s \text{ when } \lambda < 0 \end{cases}$$

where $\psi, |\lambda| < 1$ and $s \geq 4$ is the largest natural number such that $\psi + \lambda s > 0$ when $\lambda < 0$. The two parameters ψ and λ are subject to the following constraints:

- (1) $\psi \geq 0$,
- (2) $\max\{-1, -\psi/s\} < \lambda < 1$.

The mean and variance for W are given by

$$E(W) = \frac{\psi}{1 - \lambda} \quad \text{and} \quad \text{Var}(W) = \frac{\psi}{(1 - \lambda)^3} = \frac{E(W)}{(1 - \lambda)^2}, \quad (4)$$

respectively. Obviously, the shape of GP distribution is controlled by parameter λ . A negative, zero and positive λ indicates underdispersion, equidispersion and overdispersion, respectively. Hence, this two-parameter GP distribution demonstrates a wider range of shapes and thus can fit data with different degrees of dispersion.

Under the PGP modelling framework, we assume that the count W_{it} follows a GP distribution with mean $E(W_{it}) = X_{it} = \psi_{it}/(1 - \lambda)$ which forms a latent GP and $\psi_{it} = (1 - \lambda)y_{it}/a_i^{t-1}$. Then, the latent RP is $\{Y_{it} = a_i^{t-1}X_{it}\}$ and is assumed to follow a gamma distribution $G(r, r/\mu_i)$ with mean μ_i and variance μ_i^2/r . This model is named as the generalized Poisson GP (GPGP) model. The unconditional mean $E(W_{it})$ and variance $\text{Var}(W_{it})$ using the law of total variance are given by

$$E(W_{it}) = E_x[E_w(W_{it}|X_{it})] = \frac{\mu_i}{a_i^{t-1}} \quad \text{and} \quad (5)$$

$$\text{Var}(W_{it}) = E_x[\text{Var}_w(W_{it}|X_{it})] + \text{Var}_x[E_w(W_{it}|X_{it})] = \frac{\mu_i}{a_i^{t-1}} \left[\frac{1}{(1 - \lambda)^2} + \frac{\mu_i}{a_i^{t-1}r} \right], \quad (6)$$

respectively. Clearly, μ_i and a_i controls both the mean and variance while r and λ act as the dispersion parameters as they only control the variance. Based on Equation (6), the larger the r and the more

negative the λ , the more underdispersed is the distribution of the GPGP model. Hence, by varying the two dispersion parameters, GPGP model can fit both overdispersed and underdispersed data.

3.3. Population heterogeneity

Some LGAs contain high proportions of zero counts while others do not. In view of this distinct trend pattern arising from LGAs with similar local characteristics, the GPGP model is further extended to accommodate mixture effects and time-evolving covariates.

Firstly, we extend the GPGP model to the finite MGPGP model by postulating that there are G groups of LGAs with distinct trend patterns and each LGA has a probability π_l of coming from group l , $l = 1, \dots, G$. The resultant MGPGP model is essentially a mixture of G GPGP models in which each model is associated with a pair of dispersion parameters λ_l and r_l . Without any prior information about the number of groups that cannabis users can be classified to, we fit the model with $G = 1, 2, \dots, m$ group(s) and determine the value of G by assessing certain model selection criteria such as DIC and significance of model parameters.

Secondly, to capture the possible non-monotone and covariate effects, the mean μ_{il} and ratio a_{il} for each LGAs i in group l are extended to mean function μ_{itl} and ratio function a_{itl} respectively, which are log-linked to some time-evolving covariates z_{jkit} , $j = \mu, a$; $i = 1, \dots, m$; $t = 1, \dots, n_i$; and $k = 0, 1, \dots, q_j$. For the cannabis data, we incorporate a time-evolving effect $z_{a1t} = \ln t$ in the ratio function a_{itl} but the mean function $\mu_{itl} = \mu_{il}$ as there is no LGA-specific covariate, that is,

$$\mu_{itl} = \exp(\beta_{\mu 0l}), \quad (7)$$

$$a_{itl} = \exp(\beta_{a0l} + \beta_{a1l} \ln t). \quad (8)$$

Then

$$\begin{aligned} E_l(W_{it}|I_{il} = 1) &= \frac{\mu_{itl}}{a_{itl}^{t-1}} = \frac{\exp(\beta_{\mu 0l})}{[\exp(\beta_{a0l} + \beta_{a1l} \ln t)]^{t-1}} \\ \frac{E_l(W_{it}|I_{il} = 1)}{E_l(W_{it-1}|I_{il} = 1)} &= \frac{\mu_{itl}}{\mu_{i,t-1,l}} \left[\frac{\exp(\beta_{a0l} + \beta_{a1l} \ln(t-1))}{\exp(\beta_{a0l} + \beta_{a1l} \ln t)} \right]^{t-2} \frac{1}{\exp(\beta_{a0l} + \beta_{a1l} \ln t)} \\ &= \left\{ \exp \left[\beta_{a1l} \left(\ln(t-1) - \frac{1}{t} \right) \right] \right\}^{t-2} \frac{1}{\exp(\beta_{a0l} + \beta_{a1l} \ln t)} \\ &= \exp \left\{ -\beta_{a0l} + \beta_{a1l} \left[(t-2) \ln \left(1 - \frac{1}{t} \right) - \ln t \right] \right\}. \end{aligned}$$

Here β_{a1l} can be interpreted as the change in the log of ratio a_{itl} per unit of log-time where a_{itl} is a unit-time growth (discount) rate of means from a observed count process to a stationary count process. Then conditional on group l , the marginal pmf of MGPGP model is

$$f(w_{it}) = \begin{cases} \frac{r_l \exp(-\lambda_l w_{it})}{\mu_{itl}^r \Gamma(r) w_{it}} \sum_{k=0}^{w_{it}-1} \frac{\left(\frac{1-\lambda_l}{a_{itl}^{t-1}} \right)^{k+1} (\lambda_l w_{it})^{w_{it}-1-k} \Gamma(k+r_l+1)}{(w_{it}-1-k)! k! \left(\frac{1-\lambda_l}{a_{itl}^{t-1}} + \frac{r_l}{\mu_{itl}} \right)^{k+r_l+1}} & \text{if } w_{it} \geq 1, \\ \left[\frac{r_l a_{itl}^{t-1}}{(1-\lambda_l) \mu_{itl} + r_l a_{itl}^{t-1}} \right]^r & \text{if } w_{it} = 0. \end{cases} \quad (9)$$

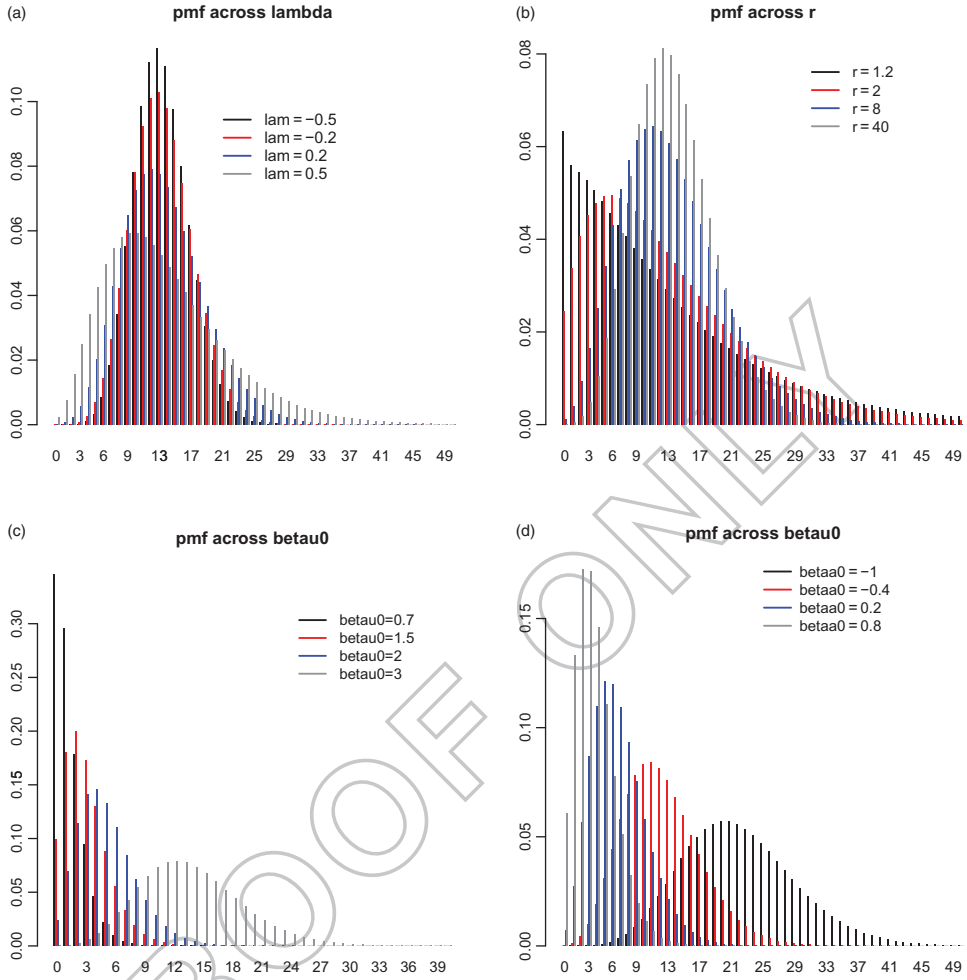


Figure 1. (a) and (b) pmfs of GPGP model with varying λ_2 and r , (c) and (d) pmfs of GPGP model with varying $\beta_{\mu 0}$ and $\beta_{\alpha 0}$.

When $\lambda_l < 0$, the marginal pmf for W_{it} is given by Equation (9) for $w_{it} = 0, 1, \dots, s_{it}$ and 0 for $w_{it} > s_{it}$. Based on Equations (5) and (6), the mean and variance conditional on group l become

$$E_l(W_{it}|I_{il} = 1) = \frac{\mu_{itl}}{a_{itl}^{t-1}} \quad \text{and} \quad \text{Var}_l(W_{it}|I_{il} = 1) = \frac{\mu_{itl}}{a_{itl}^{t-1}(1 - \lambda_l)^2} + \frac{\mu_{itl}^2}{a_{itl}^{2(t-1)}r_l}, \quad (10)$$

respectively where the latent group- l membership indicator I_{il} for W_{it} is defined as $I_{il} = 1$ if W_{it} comes from group l and zero otherwise.

Since the mixing gamma distribution adds extra variability to the GPGP model, we assess the ability of the GPGP model to allow for different levels of underdispersion and overdispersion by plotting its pmf at various parameter values with $G = 1$, $m = 1$, $q_\mu = q_a = 1$, $z_{\mu 1t} = b = 0, 1$ in Equation (7) and $z_{a 1t} = t$ in Equation (8). Hence, the mean function is $\mu_t = \exp(\beta_{\mu 0} + \beta_{\mu 1}b)$ and the ratio function is $a_t = \exp(\beta_{\alpha 0} + \beta_{\alpha 1}t)$. Fixing $b = 1$ and $t = 2$, we change the value of one model parameter at a time while keeping other parameters constant and the pmfs are displayed in Figure 1(a)–(d) with their means and variances, from Equations (5) and (6), summarized in Table 2.

Based on Figure 1 and Table 2, it is clear that parameters $\beta_{\mu 0}$ and $\beta_{\alpha 0}$ control both location and degree of dispersion of the distribution. A larger $\beta_{\mu 0}$ and smaller $\beta_{\alpha 0}$ lead to larger mean and variance

Table 2. Moments of marginal pmfs for GPGP model under a set of floating parameters with fixed values of $\lambda = 0.2, r = 30, \beta_{\mu 0} = 3, \beta_{\mu 1} = -0.5, \beta_{a0} = -0.5, \beta_{a1} = 0.2$.

floating par	$E(W_t)$	$\text{Var}(W_t)$	Floating par	$E(W_t)$	$\text{Var}(W_t)$	Floating par	$E(W_t)$	$\text{Var}(W_t)$	Floating par	$E(W_t)$	$\text{Var}(W_t)$
$\lambda = -0.5$	13.464	12.026	$r = 1.2$	13.464	172.097	$\beta_{\mu 0} = 0.7$	1.350	2.170	$\beta_{a0} = -1.0$	22.198	51.109
$\lambda = -0.2$	13.464	15.392	$r = 2.0$	13.464	111.673	$\beta_{\mu 0} = 1.5$	3.004	4.995	$\beta_{a0} = -0.4$	12.182	23.982
$\lambda = 0.2$	13.464	27.079	$r = 8.0$	13.464	43.696	$\beta_{\mu 0} = 2.0$	4.953	8.557	$\beta_{a0} = 0.2$	6.686	11.937
$\lambda = 0.5$	13.464	59.897	$r = 40.0$	13.464	25.569	$\beta_{\mu 0} = 3.0$	13.464	27.079	$\beta_{a0} = 0.8$	3.669	6.182

of the resultant distribution. On the other hand, both r and λ serve as dispersion parameters. Without altering the mean, a smaller r or larger λ contributes to a larger variance for the distribution. Note that since the gamma distribution of the underlying stochastic process $\{Y_t\}$ introduces extra variation to the GP distribution, a negative λ does not necessarily reveal an underdispersion as shown in Table 2. For example, overdispersion ($E(W_t) < \text{Var}(W_t)$) is retained even when $\lambda = -0.2$ because the variance is controlled by the factor $1/(1 - \lambda)^2 + \mu_t/a_t^{t-1}r$ based on Equation (6). Generally speaking, the more negative the λ , the more underdispersed is the distribution. In summary, by varying the model parameters, GPGP model can fit count data with different degrees of dispersion, ranging from underdispersion due to homogeneity to overdispersion due to some outlying observations.

4. Bayesian inference

For the statistical inference, maximum likelihood (ML) method [28] or moment method [10] has been used since the model has tractable likelihood function. For random effects model, Yau et al. [29] adopt the expectation-maximization (EM) method. However for the MGPGP model with a large set of latent parameters $\{y_{itl}, I_{itl}\}$, a favourable choice is the Bayesian approach via MCMC algorithms. The key advantage is that it augments the parameter space to include all latent parameters through a hierarchical structure and that numerical optimisation is not required. The model can be easily implemented using the popular WinBUGS (Bayesian analysis Using Gibbs Sampling) package.[30]

4.1. MCMC algorithms

The hierarchical structure of GPGP model under the Bayesian framework is stated as follows:

$$w_{it} \sim I_{i1} \text{GP} \left(\frac{y_{it1}}{a_{it1}^{t-1}}, \lambda_1 \right) + \dots + I_{iG} \text{GP} \left(\frac{y_{itG}}{a_{itG}^{t-1}}, \lambda_G \right),$$

$$y_{itl} \sim G \left(r_l, \frac{r_l}{\mu_{itl}} \right),$$

where μ_{itl} and a_{itl} are given by Equations (7) and (8). We further assume the missing data:

$$(I_{i1}, \dots, I_{iG})^T \sim \text{multinomial}(1, \pi_1, \dots, \pi_G). \quad (11)$$

In order to construct the posterior density, some non-informative prior distributions are assigned to the model parameters as follows:

$$\beta_{jkl} \sim N(0, \tau_{jkl}^2), \quad j = \mu, a; k = 0, 1, \dots, q; l = 1, \dots, G, \quad (12)$$

$$r_l \sim G(d_l, e_l), \quad (13)$$

$$\lambda_l \sim U(-1, 1), \quad (14)$$

$$(\pi_1, \dots, \pi_G)^T \sim \text{Dir}(\alpha_1, \dots, \alpha_G), \quad (15)$$

where $G(d_l, e_l)$ denotes gamma distribution with shape parameter d_l and scale parameter e_l , $U(b, c)$ denotes uniform distribution on the interval (b, c) and $Dir(\alpha)$ represents a Dirichlet distribution, conjugate to multinomial distribution, with parameters $\alpha = (\alpha_1, \dots, \alpha_G)$. In case of a 2-group ($G=2$) mixture model, Equation (11) can be simplified to $I_{i1} \sim \text{Bernoulli}(\pi_1)$, $I_{i2} = 1 - I_{i1}$ and Equation (15) becomes a uniform prior $f_U(0, 1)$ for π_1 with $\pi_2 = 1 - \pi_1$. For the data analyses and simulation studies, the hyperparameters τ_{jkl} are set to be 1000, d_l and e_l to be 0.1 and α_l to be 1 to obtain **non-informative** priors. Moreover, we specify $d_l = 0.1$, $e_1 = 0.05$ and $e_2 = 0.1$ in Equation (13). Note that to avoid label switching, we set $\beta_{\mu 01} > \beta_{\mu 02}$ so that $l=1$ indicates the **high level** and $l=2$ the **low level**.

According to Bayes' theorem, the posterior density is proportional to the joint densities of complete data likelihood and prior probability distributions. For the MGPGP model, the complete data likelihood function $L(\theta)$ for the observed data w_{it} and missing data $\{y_{itl}, I_{itl}\}$ is derived as

$$L(\theta) = \prod_{i=1}^m \prod_{l=1}^G \left\{ \pi_l \prod_{t=1}^{n_i} f_{\text{GPD}}(w_{it} | \beta_{al}, \lambda_l, y_{itl}) f_G(y_{itl} | \beta_{\mu l}, r_l) \right\}^{I_{il}}. \quad (16)$$

The vector of model parameters $\theta = (\beta^T, \lambda^T, r^T, \pi^T)^T$ where $\beta = (\beta_{\mu 1}^T, \dots, \beta_{\mu G}^T, \beta_{a1}^T, \dots, \beta_{aG}^T)^T$, $\beta_{jl} = (\beta_{j0l}, \beta_{j1l}, \dots, \beta_{jq_l l})^T$ for $j = \mu, a$; $l = 1, \dots, G$, $\lambda = (\lambda_1, \dots, \lambda_G)^T$, $r = (r_1, \dots, r_G)^T$ and $\pi = (\pi_1, \dots, \pi_G)^T$. The joint posterior density of the complete data likelihood and prior densities is expressed as follows:

$$f(\beta, \lambda, r, \pi | w, I, y) \propto L(\theta) \left(\prod_{j=\mu, a} \prod_{k=0}^{q_j} \prod_{l=1}^G f_N(\beta_{jkl} | 0, \tau_{jkl}^2) \right) \left(\prod_{l=1}^G f_G(r_l | d_l, e_l) \right) \\ \times \left(\prod_{l=1}^G f_U(\lambda_l | -1, 1) \right) f_{\text{Dir}}(\pi | \alpha),$$

where $w = (w_{11}, w_{12}, \dots, w_{mn_m})^T$, $I = (I_{11}, I_{12}, \dots, I_{mG})^T$, $y = (y_{111}, y_{121}, \dots, y_{mn_m G})^T$, $L(\theta)$ is given by (16) and the priors are given by Equations (12)–(15). In Gibbs sampling, the parameters $\Theta = (\beta, \lambda, r, \pi, y, I)$ including model parameters θ and missing observations y and I are simulated iteratively from their univariate conditional posterior distributions which are listed in the appendix.

Theories for Bayesian computation state that the empirical distribution of this sequence of simulated parameters converges towards the posterior distribution of the unknown parameters. The posterior distribution can thus be approximated on the basis of these simulated values. The MCMC algorithms are implemented using WinBUGS where three chains of **25,000** iterations each are executed for each model and the first 5000 iterations are discarded as the burn-in period. Thereafter, parameters are sub-sampled from every 10^{th} iteration to reduce the auto-correlation in the sample. This results in $M = 6000$ simulated posterior samples for each parameter and parameter estimates are given by their sample means. History plots, **Brooks–Gelman–Rubin** (BGR) diagnostic plots [31] and auto-correlation function (ACF) plots of each parameter are examined to ensure convergence and independence amongst the parameters.

4.2. Model selection criterion

For model selection, we adopt the deviance information criterion (DIC) [32] defined as

$$\text{DIC} = \overline{D(\theta)} + p_D,$$

which is the sum of a measure for model fit and a penalty for model complexity. The model fit is measured by the deviance:

$$D(\boldsymbol{\theta}) = -2 \log f(\mathbf{y}|\boldsymbol{\theta})$$

where $\boldsymbol{\theta}$ denotes the parameter vector, \mathbf{y} is the observed data and f is the likelihood function. The posterior mean deviance $\overline{D(\boldsymbol{\theta})}$ is evaluated as follows:

$$E_{\boldsymbol{\theta}|\mathbf{y}}[D(\boldsymbol{\theta})] = \overline{D(\boldsymbol{\theta})},$$

whereas the penalty for model complexity is $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$. DIC is very useful in measuring the model fit of complicated hierarchical models where the number of parameters is not well defined and hence the calculation of AIC and BIC fails. Applying the idea of Celeux et al. [34] and considering I_{il} as missing data, the DIC for the MGPGP model is defined as

$$\begin{aligned} & -\frac{4}{M} \sum_{j=1}^M \sum_{i=1}^m \sum_{l=1}^G I'_{il}{}^{(j)} \left[\ln \pi_l^{(j)} + \sum_{t=1}^{n_i} \ln \{f_{GP}(w_{it} | y_{itl}^{(j)}, \boldsymbol{\beta}_{al}^{(j)}, \lambda_l^{(j)})\} \right] \\ & + 2 \sum_{i=1}^m \sum_{l=1}^G \bar{I}'_{il} \left[\ln \bar{\pi}_l + \sum_{t=1}^{n_i} \ln \{f_{GP}(w_{it} | \bar{y}_{itl}, \bar{\boldsymbol{\beta}}_{al}, \bar{\lambda}_l)\} \right] \end{aligned} \quad (17)$$

where $\boldsymbol{\theta}^{(j)}$ and $\bar{\boldsymbol{\theta}}$ represent the j^{th} posterior sample and posterior mean of parameter $\boldsymbol{\theta}$,

$$I'_{il}{}^{(j)} = \frac{\pi_l^{(j)} \prod_{t=1}^{n_i} \{f_{GP}(w_{it} | y_{itl}^{(j)}, \boldsymbol{\beta}_{al}^{(j)}, \lambda_l^{(j)})\}}{\sum_{l'=1}^G \pi_{l'}^{(j)} \prod_{t=1}^{n_i} \{f_{GP}(w_{it} | y_{itl'}^{(j)}, \boldsymbol{\beta}_{al'}^{(j)}, \lambda_{l'}^{(j)})\}},$$

and \bar{I}'_{il} is defined in a similar way by replacing $y_{itl}^{(j)}$ and $\boldsymbol{\theta}^{(j)}$ with \bar{y}_{itl} and $\bar{\boldsymbol{\theta}}$. The best model should have the smallest DIC indicating the best model fit after accounting for model complexity.

5. Simulation study

The objective of the study is to assess the competence of the GPGP model under different levels of dispersion and to compare its performance with the RGP-EP model. In the study, data are simulated from the GPGP model in Equation (9) with $G=1$ (hence dropping the subscript l) under a general situation with non-constant mean function $\mu_{it} = \exp(\beta_{\mu 0} + \beta_{\mu 1} z_{\mu it})$ and ratio function $a_{it} = \exp(\beta_{a0} + \beta_{a1} \ln t)$ where $q_{\mu} = q_a = 1$ and the true parameters $\beta_{\mu 0} = -2.0$, $\beta_{\mu 1} = -0.5$, $\beta_{a0} = -0.05$, $\beta_{a1} = 0.005$, $\lambda = -0.99$ and $r=30$. The data contain $m=80$ time series of length $n_i = 8$ from $m_0 = 40$ LGAs with $z_{\mu it} = 0$ (control group) and another $m_1 = 40$ LGAs with $z_{\mu it} = 1$ (treatment group). It exhibits underdispersion across time as revealed by the first row under mean and variance in Table 3 and is fitted to both GPGP and RGP-EP models. Results in Table 4 show that the GPGP model consistently gives a better model fit since it accommodates underdispersion while the RGP-EP model fails to do so.

Without altering the overall trend, two more data sets with equidispersion and overdispersion are then generated by adding a constant 9 and a constant 15, respectively to all observations of the last LGA in both control group ($i=40$) and treatment group ($i=80$). Hence, each line in Table 3 reports the trend from a single data set of 80 LGAs. With equidispersed data, the parameters of the mean and ratio functions for the RGP-EP model do not alter much since the trend movement remains unchanged and the mean levels only increase slightly. As expected, the shape parameter ν for $\{Y_{it}^* = \ln Y_{it}\}$ has changed sharply from 0.15 to 0.88 and the scale parameter σ has increased moderately. These imply that the distribution of the latent $\{Y_{it}^*\}$ becomes more heavy-tailed to accommodate the outliers. For the GPGP model, again parameters in the mean and ratio functions remain nearly

Table 3. Mean and variance of three simulated data sets in the simulation study.

		Time								Average over time
Data		1	2	3	4	5	6	7	8	
Underdispersion	mean	2.750	3.088	3.013	3.213	3.300	3.175	3.075	3.075	3.086
Equidispersion		2.975	3.313	3.238	3.438	3.525	3.400	3.300	3.300	3.311
Overdispersion		3.125	3.463	3.388	3.588	3.675	3.550	3.450	3.450	3.461
Underdispersion	variance	1.658	1.777	1.354	1.258	1.428	1.437	1.665	1.412	1.499
Equidispersion		3.772	3.509	3.120	3.616	3.063	3.585	3.630	2.922	3.402
Overdispersion		7.402	6.885	6.519	7.410	6.374	7.238	7.162	6.149	6.892

Table 4. Parameter estimates, *SE* and *DIC* for GPGP and RPGP-EP models under different degrees of dispersion.

Model	Parameter	Underdispersion		Equidispersion		Overdispersion	
		Estimate	SE	Estimate	SE	Estimate	SE
GPGP	β_{a0}	-0.0669	0.0194	-0.0389	0.0513	-0.0733	0.0283
	β_{a1}	0.0070	0.0024	0.0040	0.0062	0.0080	0.0033
	$\beta_{\mu 0}$	1.1430	0.0368	1.2255	0.0606	1.2365	0.0516
	$\beta_{\mu 1}$	-0.2212	0.0312	-0.2037	0.0413	-0.1915	0.0451
	λ	-0.8792	0.1019	-0.9651	0.0432	-0.9701	0.0422
	r	19.3500	8.4734	6.1110	0.4832	4.0845	0.2801
	DIC	1893.60		1943.31	(+2.6%)	1966.24	(+1.2%)
RPGP-EP	β_{a0}	-0.1809	0.0254	-0.1111	0.0228	-0.1244	0.0368
	β_{a1}	0.0188	0.0038	0.0115	0.0032	0.0127	0.0053
	$\beta_{\mu 0}$	1.0380	0.0064	1.0950	0.0125	1.0880	0.0093
	$\beta_{\mu 1}$	-0.4143	0.0157	-0.1672	0.0242	-0.2892	0.0175
	ν	0.1462	0.0731	0.8775	0.1264	1.9810	0.0161
	σ	0.1090	0.0059	0.2005	0.0067	0.1189	0.0039
	DIC	2206.33		2435.34	(+10.4%)	2449.17	(+0.6%)

stagnant except for the shape parameter r of $\{Y_{it}\}$ which drops abruptly to allow for the inflated variance due to the enlarged observations. The dispersion parameter λ in the GPGP model however does not increase but instead becomes slightly more negative, due to the fact that the variance of the distribution is controlled by the factor $1/(1 - \lambda)^2 + \mu_t/a_t^{t-1}r$, which involves both λ and r as shown in Equation (6). This reveals that the shape parameter r of the gamma distribution for $\{Y_{it}\}$ has a stronger effect on the level of dispersion than the dispersion parameter λ in the GP distribution for the outcome $\{W_{it}\}$. A decrease in r ‘absorbs’ the outlying effect by increasing the tail of the gamma distribution. Similarly, the RPGP-EP model downweights the outlying effect by increasing the shape parameter ν of the EP distribution for $\{Y_{it}^*\}$. Comparing the model fit, the DIC shows a further larger percentage increase, 10.4% versus 2.6% for the RPGP-EP and GPGP models, respectively, from underdispersed to equidispersed data. This indicates that the GPGP model still performs better with equidispersed data.

Lastly, in the case of overdispersion, the parameters show similar movements with a dramatic increase in ν in the RPGP-EP model (0.88 to 1.98) and a further decrease in r in the GPGP model (6.11 to 4.09). However, the percentages increase in DIC, 1.2% versus 0.6% for GPGP and RPGP-EP models, respectively, are less as compared to the case of equidispersion because they both can handle overdispersed data by increasing the tailedness of the distribution for the underlying stochastic process $\{Y_{it}\}$. In the RPGP-EP model, the outlying effect is downweighed by widening the interval of the uniform distribution for $\{Y_{it}^*\}$ whereas in the GPGP model, it is accommodated by magnifying the variance of the gamma distribution for $\{Y_{it}\}$.

In summary, the simulation study reveals good performance of the GPGP model as it can fit data with different degrees of dispersion by varying the two model parameters λ and r . Hence, the GPGP and MGPGP models are fitted to the cannabis data to evaluate the policy competence.

6. Data analysis

6.1. Analysis of all LGAs

Because of the heterogeneity across all LGAs, we fit the MGPGP model, beginning with a 1-group MGPGP model ($G = 1$). Obviously, it fails to explain the population heterogeneity leaving a large deviation between the observed and predicted counts for some LGAs with high occurrences of cannabis dealing and trafficking. This poor fit is revealed from the larger DIC (2076.76) of the 1-group MGPGP model when compared to that of the 2-group MGPGP model ($DIC = 1815.81$). However, a 3-group MGPGP model ($G = 3$) overfits the data with one cluster actually degenerating. Therefore, we conclude that the 2-group MGPGP model ($G = 2$) is the most appropriate model for the data. In addition, we find that the parameters β_{a1l} , $l = 1, 2$ in the ratio functions are insignificant (95% credible interval (CI) for β_{a1l} contains zero) indicating that the trends are indeed monotone in both groups. Hence, we drop the redundant time-evolving covariate $\ln t$ in the ratio functions and they become $a_{il} = \exp(\beta_{a0l})$.

For model comparison, we further fit two conventional models, the 2-group mixture generalized Poisson (MGP) model and the Poisson random effects (PRE) model, both without a GP model, defined as

MGP: $W_{it} \sim GP(\psi_{it}^*, \lambda_i)$; $\psi_{it}^* = (1 - \lambda_i) \exp(\beta_{\mu 0l} + \beta_{\mu 1l} \log t)$ if $I_{il} = 1$, (18)

PRE: $W_{it} \sim Poi(\mu_{it}^*)$; $\mu_{it}^* = \exp(\beta_{\mu 0} + \beta_{\mu 1} \log t + u_i)$; $u_i \sim N(0, \sigma^2)$. (19)

In these two models, parameters β_{1l} and $\beta_{\mu 1}$ account for the time effect similar to parameter β_{a0l} in the ratio function of the GP model. However, negative signs of β_{1l} and $\beta_{\mu 1}$ indicate a decreasing trend whereas a decreasing trend in the GP model is indicated by a positive sign for β_{a0l} . Table 5 summarizes the parameter estimates, standard errors (SE), Monte Carlo error (MCE) and DIC for the MGP, PRE, and MGPGP models.

To evaluate the properties of these estimates, the history plots of three MCMC chains for the parameters of MGPGP model are presented in Figure 2. The plots clearly demonstrate convergence and good mixing of the chains. Moreover, the BGR diagnostic plots also show convergence and the autocorrelation function (ACF) plots display low autocorrelation across iterations for all parameters. The BGR diagnostic plots and ACF plots for all models and the history plots for MGP and PRE models are available upon requests.

For the MGPGP model, the low-level group exhibits a monotone decreasing trend as the ratio parameter β_{a02} (95% CI is [0.0350, 0.1522]) is significantly greater than zero. On the other hand, the high-level group demonstrates a stationary trend since β_{a01} is marginally insignificant (95% CI for $\beta_{\mu 11}$ is [-0.0040, 0.1138]) possibly due to the high within-group variability.

Table 5. Parameter estimates, SE, MCE and DIC for MGP, PRE and MGPGP models using the Cannabis data.

Group	MGP model					PRE model				MGPGP model			
	Par	Estimate	SE	MCE		Par	Estimate	SE	MCE	Par	Estimate	SE	MCE
l = 1 high-level	$\beta_{\mu 01}$	2.7570	0.1278	0.0018		$\beta_{\mu 0}$	1.8080	0.1781	0.0033	$\beta_{\mu 01}$	3.0000	0.1646	0.0079
	$\beta_{\mu 11}$	-0.1929	0.0809	0.0011		$\beta_{\mu 1}$	-0.3262	0.0276	0.0003	β_{a01}	0.0547	0.0297	0.0009
	λ_1	0.7223	0.0220	0.0003		σ	1.1250	0.1343	0.0017	λ_1	0.7247	0.0269	0.0005
										r_1	32.05	17.70	0.3115
l = 2 low-level	π_1	0.5544	0.0744	0.0011						π_1	0.3626	0.0849	0.0022
	$\beta_{\mu 02}$	0.9629	0.1208	0.0015						$\beta_{\mu 02}$	1.2810	0.1285	0.0031
	$\beta_{\mu 12}$	-0.3274	0.0912	0.0011						β_{a02}	0.0924	0.0297	0.0006
	λ_2	0.0039	0.0039	0.0005						λ_2	0.0017	0.0018	0.0000
										r_2	2.1400	0.5005	0.0106
	DIC	1930					1740				1819		

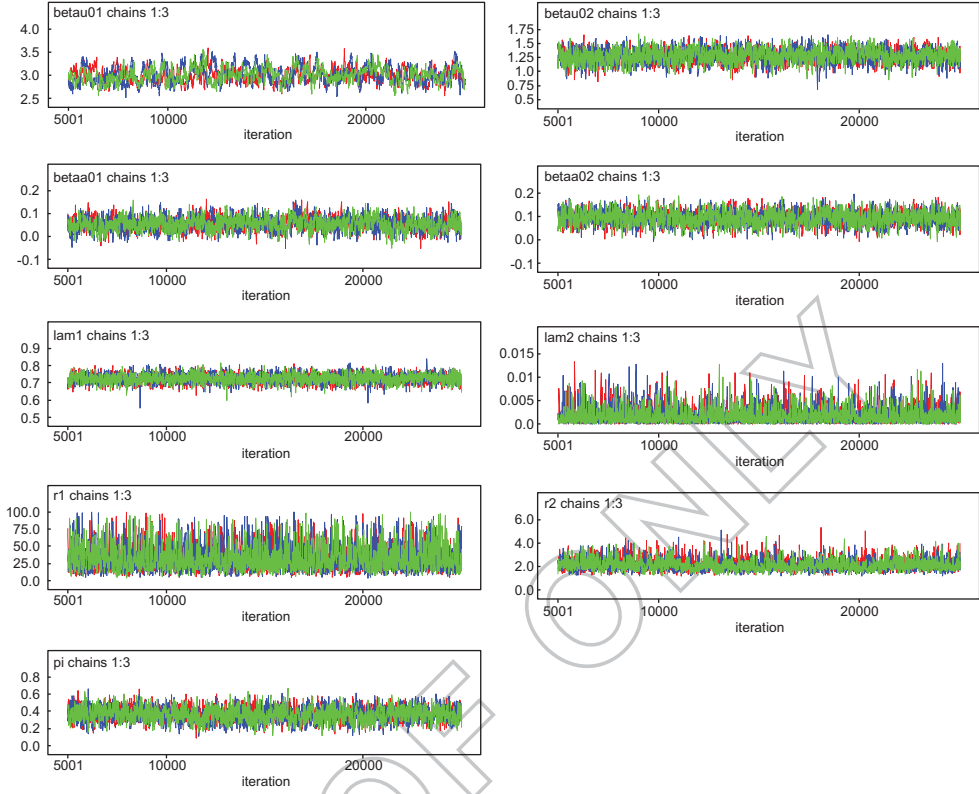


Figure 2. History plots of model parameters for the MGPGP model.

Classification of LGAs is made using the posterior means of the group membership indicators \hat{I}_{il} . In particular, LGA i is classified into group l if \hat{I}_{il} is a maximum across l , that is, $\hat{I}_{il} = \max_l \hat{I}_{il}$. Results show that classifications under both MGP and MGPGP models are identical, namely, 15 LGAs, indicated by ‘†’ in Table 1, are categorized to the high-level group ($l = 1$). As expected, Sydney which is the central business district with a high concentration of pubs and night clubs belongs to this group. The categorization implies that the arrests for cannabis dealing and trafficking are concentrated in the inner and western part of Sydney. These results concur with the findings in [25]. The affinity of the classification may possibly be due to the fact that both models capture the population heterogeneity by incorporating mixture effects. Moreover, this classification can also be obtained using the PRE model by selecting 15 LGAs with highest random variables u_i . Since the LGAs with outlying observations are classified into the high-level group ($l = 1$), this group thus exhibits a higher degree of overdispersion in contrast to the low-level group ($l = 2$) which comprises of LGAs with low incidence counts.

To verify this, we compare the mean and variance over time for the MGP, PRE and MGPGP models. The predicted mean $\hat{E}_l(W_t)$ and variance $\hat{V}ar_l(W_t)$ for the MGPGP model are given by Equation (10) with $a_{tl} = \exp(\beta_{a0l})$ and $\mu_{tl} = \exp(\beta_{\mu 0l})$ and those for MGP model are given by Equation (4). For the PRE model, the approximated pmf $\hat{f}_t(w)$ is first obtained by

$$\hat{f}_t(w) = \sum_{i=1}^{M_s} f_P(w|\mu_{it}) = \sum_{i=1}^{M_s} \frac{\exp(-\mu_{it}) \mu_{it}^w}{w!}, \quad w = 0, 1, \dots, \infty,$$

where μ_{it} in Equation (19) depends on u_i , $u_i \sim N(0, \sigma^2)$, $i = 1, \dots, M_s$ and M_s is the number of simulated LGAs. Then the predicted mean and variance are calculated using:

$$\hat{E}(W_t) = \sum_{w=0}^{\infty} w \hat{f}_t(w) \quad \text{and} \quad \hat{\text{Var}}(W_t) = \left[\sum_{w=0}^{\infty} w^2 \hat{f}_t(w) \right] - [\hat{E}(W_t)]^2. \quad (20)$$

The observed mean $E_l(W_t)$ and variance $\text{Var}_l(W_t)$ are given by the weighted mean \bar{W}_{il} and variance S_{il}^2 over the subsets $\mathcal{W}_t = \{W_{it'} : t' = t\}$ using weights \hat{I}_{il} ($\hat{I}_{i2} = 1 - \hat{I}_{i1}$). To illustrate their distinct trend patterns, Figure 3 plots the observed means and variances and predicted means and variances across time for the three models. We note that the observed values are different across MGP and MGPGP models at both high and low levels due to the different weighting \hat{I}_{il} . However as they are very similar, only the trend lines for the MGPGP model are plotted. Clearly, the degree of overdispersion is substantially higher in the high-level group since some extreme counts ($w_{4,1} = 155$, $w_{4,2} = 114$) in Sydney, one of the LGAs in this group, inflate the variability of the counts dramatically. However, as both MGP and MGPGP models can capture overdispersion, the predicted variances of the high-level group are underestimated in the first two years but are getting closer to the observed variances in the next six years. Results for the PRE model are similar in terms of trend movement but both the mean and variance lie between those of the high- and low-level groups.

Figure 3 shows a declining trend in both high-level and low-level groups with a relatively slower decline rate in the low-level group despite the downward trend of the high-level group is marginally insignificant for the MGPGP model. Moreover, the close agreement between the observed and predicted mean trends proves the adequacy of the models to fit the panel data. Overall, these results give statistical evidence that the Cannabis Caution Scheme is associated with a drop in the number of arrests for cannabis dealing and trafficking in NSW. In other words, our results imply that the Cannabis Caution Scheme is effective in reducing the number of cannabis users within the Sydney metro area. This finding also agrees with the 2007 National Drug Strategy Household Survey [33] that the proportion of the population aged 14 years and over who had used cannabis in the last 12 months has fallen from 15.8% in 2001 to 14.4% in 2004 and further reduced to 11.6% in 2007.

To compare model-fit, the MGPGP model clearly outperforms the MGP model without a GP model. Close investigation reveals that the state variables y_{itl} are enlarged to allow larger means for outliers whereas the mean μ_{itl} in the MGP model is constant for a given time t and group l . This confirms the advantage of adopting a GP model for trend data. However, the MGPGP model does not outperform the PRE model according to DIC. The MGPGP model has the capability to model underdispersion while the PRE model fails. Nevertheless, this advantage may be partly offset by its increase in the effective numbers of parameters p_D (98.0 vs. 42.3) because the state variables Y_{itl} for modelling population heterogeneity in the MGPGP model are more abundant than those u_i in the PRE model. Moreover, the MGPGP model has a separate ratio function (8) to describe the dynamic of trend movement whereas the PRE model allows the trend movement by a log-time covariate in the mean function (19). We further display the performance of the MGPGP model by comparing its observed pmf $f_{il}(w)$ and fitted pmf $\hat{f}_{il}(w)$ at two chosen times $t = 1$ and $t = 5$ in Figures 4. They show that the distribution of the MGPGP model can describe the mode of the empirical distribution well particularly for the low-level group.

In summary, the findings of the significant downward trend for cannabis dealing and trafficking and the two identified groups of LGAs certainly provide useful information for the management of CCS and allocation of police force to bring down the cannabis use especially in those LGAs in the high-level group with high number of cannabis offences.

6.2. Analysis of LGAs with low counts

While a policy effect is detected among LGAs with a high-level cannabis offenses, it is unsure if the result applies to those 'peaceful' LGAs with a low level of cannabis offenses. In light of this, the second

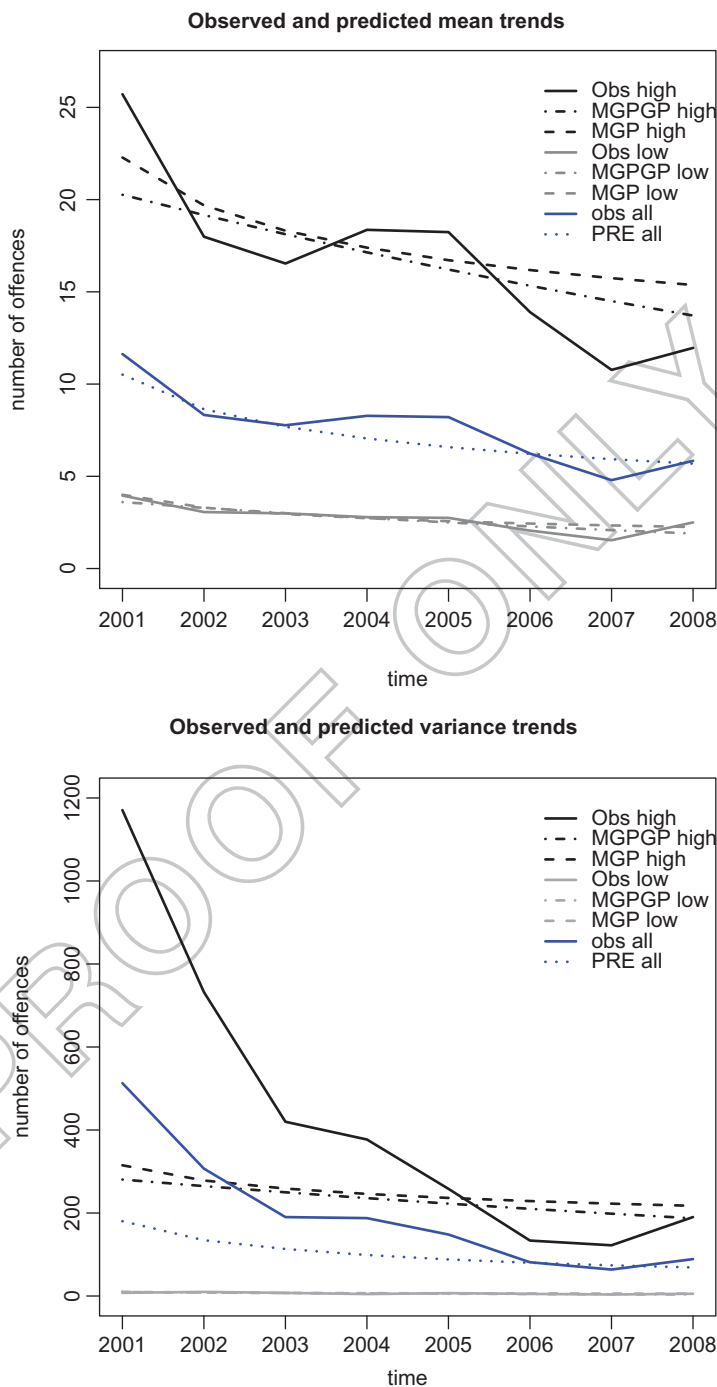


Figure 3. Observed and fitted mean and variance for MGPGP and MGP models using complete data.

analysis is conducted on 16 LGAs with lowest counts as indicated by their lowest \hat{u}_i and zero \hat{I}_{i1} with an aim to investigate if the downward trend still applies to these LGAs. Apparently, a zero-inflated distribution may be suggested as this data contain a substantial proportion of zeros. However, the

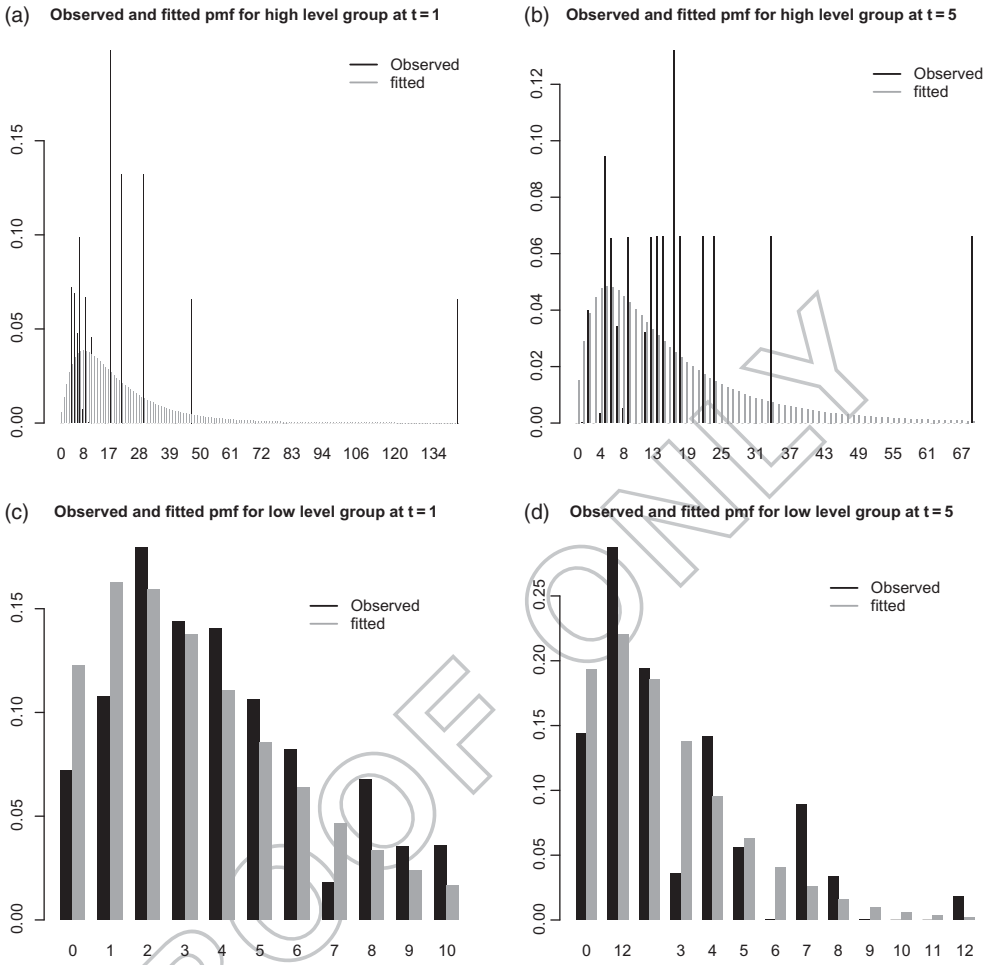


Figure 4. (a) and (b) Pmfs of high-level group for MGPGP model at $t = 1$ and $t = 5$, (c) and (d) Pmfs of low-level group for MGPGP model at $t = 1$ and $t = 5$.

distribution makes an explicit assumption that the zeros come from two sources, a degenerated distribution at zero and a Poisson distribution and such assumption may be difficult to justify in practice. On the other hand, the excessive zeros do give rise to low level of dispersion (ranging from near equidispersion to underdispersion at times $t = 5$ and 7) which can be captured by a GP distribution with a shape parameter to reveal the level of dispersion. Hence, a GP distribution is adopted and its capability in modelling underdispersion is compared to other models with different data distributions, namely, Poisson (P), NB and QNB and with different model types, namely, geometric process (GP) and random effects (RE) models. The five models we consider are PGP, NBGP, GP, GP and QNBRE. We adopt dynamic shape parameters λ_t for the first three GP models (except for the PGP model with no shape parameter) to allow flexibility in modelling dispersion. For the last two RE models, we fix λ and include random variables u_i in the mean function. The five models are summarized as below:

$$\text{PGP} : W_{it} \sim \text{Poi}(X_{it}) \quad \text{where } X_{it} = \frac{Y_{it}}{a^{t-1}} \quad \text{and} \quad Y_{it} = X_{it}a^{t-1} \sim G(r, \frac{r}{\mu}),$$

NBGP : $W_{it} \sim \text{NB}(p_{it}, \lambda_t)$ where $p_{it} = \frac{Y_{it}}{Y_{it} + a^{t-1}\lambda_t}$ and $Y_{it} = X_{it}a^{t-1} \sim G(r, \frac{r}{\mu})$,

GPGP : $W_{it} \sim \text{GP}(\psi_{it}, \lambda_t)$ where $\psi_{it} = \frac{(1 - \lambda_t)Y_{it}}{a^{t-1}}$ and $Y_{it} = X_{it}a^{t-1} \sim G(r, \frac{r}{\mu})$,

GPRE : $W_{it} \sim \text{GP}(\psi_{it}, \lambda)$ where $\psi_{it} = (1 - \lambda)\mu_{it}^*$ and $u_i \sim N(0, \sigma^2)$,

QNBRE : $W_{it} \sim \text{QNB}(b_{it}, \lambda, \alpha)$ where $b_{it} = \frac{\alpha(1 - \lambda)}{\mu_{it}^*}$ and $u_i \sim N(0, \sigma^2)$,

where $a = \exp(\beta_0)$, $\mu = \exp(\beta_1)$, $\mu_{it}^* = \exp(\beta_0 + \beta_1 \log t + u_i)$ and the pmf of $\text{QNB}(b, \lambda, \alpha)$ ($\lambda = c$ in Li et al., 2011) is given by

$$f(w) = \begin{cases} \frac{\Gamma(w + \alpha)}{w! \Gamma(\alpha)} \frac{1}{1 + \lambda w} \left(\frac{1 + \lambda w}{1 + b + \lambda w} \right)^w \left(\frac{b}{1 + b + \lambda w} \right)^\alpha, & w = 0, 1, \dots \\ 0, & w > m \text{ if } \lambda < 0, \end{cases}$$

where $\alpha > 0$, $b > 0$, $m > [-1/\lambda]$ and m is the largest positive integer for $1 + m\lambda \geq 0$. The parameterization for the GP models is obtained by setting $E(W_{it}|X_{it}) = X_{it}$ and for the GPRE model by setting $E(W_{it}|u_i) = \mu_{it}^*$. Parameter estimates and DIC for the five models are presented in Table 6. Moreover, the marginal means and variances for the three GP models which are given by

PGP: $E(W_{it}) = \frac{\mu}{a^{t-1}}$ and $\text{Var}(W_{it}) = \frac{\mu}{a^{t-1}} + \frac{\mu^2}{ra^{2(t-1)}}$,

NBGP: $E(W_{it}) = \frac{\mu}{a^{t-1}}$ and $\text{Var}(W_{it}) = \frac{(1 + r)\mu^2 + r\lambda_t a^{t-1}\mu}{r\lambda_t a^{2(t-1)}} + \frac{\mu^2}{ra^{2(t-1)}}$,

and Equations (5) and (6), replacing λ by λ_t , for the GPGP model are plotted in Figure 5 to show the trend movements. Results depict that all the means and variances show decreasing trends, with similar trends for means but very different trends for variances across models. All trends show overdispersion except for $t = 5$ under the GPGP model. This demonstrates the capability of the GPGP model to model underdispersion. Moreover, Table 6 shows that the GPGP model provides the best model fit according to DIC because only GP distribution capture underdispersion. Furthermore, the PGP model outperforms the NBGP model possibly because it suits less overdispersed data. In comparing between the GPRE and QNBRE models, we note that the shape parameter $\hat{\lambda} > 0$ in the QNBRE model implying that the mean and variance do not exist. This explains why the GPRE model outperforms the QNBRE model. Lastly, the GPGP model also outperforms the GPRE model showing that the geometric process model with dynamic shape parameters describes the dynamics of trend movement and underdispersion across time better than the random effects model. In summary, the downward trend of cannabis offences still applies to those peaceful LGAs (all CIs for β_{a0} exclude 0) confirming a policy effect in these LGAs.

Table 6. Parameter estimates, SE, MCE and DIC for PGP, NBGP, GPGP, GPRE and QNBRE models using low level cannabis data.

Par	PGP			NBGP			GPGP			GPRE			QNBRE		
	est	SE	MCE	est	SE	MCE	est	SE	MCE	est	SE	MCE	est	SE	MCE
β_0	0.687	0.151	0.0021	0.699	0.158	0.0030	0.757	0.161	0.0026	0.739	0.203	0.0025	0.583	0.228	0.0032
β_1	0.087	0.038	0.0005	0.090	0.040	0.0010	0.100	0.042	0.0007	-0.332	0.119	0.0016	-0.305	0.116	0.0017
λ	-	-	-	varied	-	-	varied	-	-	0.141	0.062	0.0008	0.050	0.040	0.0006
α	-	-	-	-	-	-	-	-	-	-	-	-	9.807	4.797	0.0643
r	3.949	2.445	0.0335	7.378	4.475	0.0621	3.201	2.357	0.0328	-	-	-	-	-	-
σ^2	-	-	-	-	-	-	-	-	-	0.202	0.157	0.0012	0.152	0.137	0.0019
DIC	407			416			384			408			413		

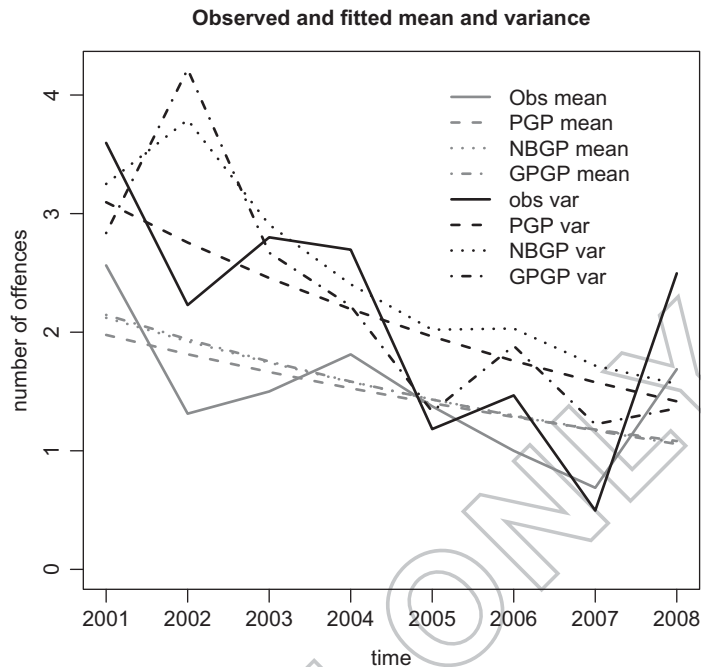


Figure 5. Observed and fitted mean and variance for PGP, NBGP and GPGP models using the clear low-level data.

7. Conclusion

To analyse the cannabis data, we propose the GPGP model which relaxes the restrictive overdispersion assumption in the PGP [18] and RPGP [6] models. The GPGP model adopts the more general GP distribution to account for both underdispersed and overdispersed data. The model is further extended to accommodate non-monotone trends and covariate and cluster effects. By varying different model parameters, the mean and variance of the GPGP model are reported in Table 2 and their pmfs are illustrated in Figure 1(a)–(d). Results in Table 2 show that the model can accommodate data with trends and different degrees of dispersion. The GPGP model can be implemented using MCMC algorithms via the user friendly Bayesian software WinBUGS.

In order to verify the model competence of the GPGP model over the RPGP-EP model, we conduct a simulation experiment to study the model performance under three situations (1) underdispersion; (2) equidispersion and (3) overdispersion. Our findings highlight that while the two models are comparable and perform satisfactorily in the case of overdispersion, the GPGP model outperforms the RPGP-EP model in handling underdispersed data.

Lastly, the analysis of cannabis offences data demonstrates the applicability of the MGPGP model to handle cluster effects, trend movements and overdispersion. The performance of MGPGP model is compared to two conventional models, namely the mixture generalized Poisson (MGP) and Poisson random effects (PRE) models in the first analysis. Results show that the MGPGP model outperforms the MGP models and provides comparable model-fit to the PRE model for this cannabis data. Focusing only on underdispersed data in the second analysis, the GPGP model provides better performance to other models with Poisson, NB and QNB distributions and with random effects modelling. In summary, the proposed GPGP and MGPGP models offer greater model flexibility to detect the decreasing trend of cannabis offences and hence a policy effect in different LGAs. In general, the models allow the study of treatment effectiveness in many practical problems by analysing the trend of treatment effects over time.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- [1] Greenwood M, Yule G. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents. *J R Statist Soc A*. 1920;83:255–279.
- [2] Holla M. On a Poisson-inverse Gaussian distribution. *Metrika*. 1966;11:115–121.
- [3] Blumer MG. On the fitting the Poisson-lognormal distribution to species abundance data. *Biometrics*. 1974;30:101–110.
- [4] Albrecht P. Laplace transforms, Mellin transforms and mixed Poisson processes. *Scand Actuar J*. 1984;11:58–64.
- [5] Karlis D, Meligkotsidou L. Multivariate Poisson regression with covariance structure. *Stat Comput*. 2005;15:255–265.
- [6] Wan W-Y, Chan JS-K. Bayesian analysis of robust Poisson geometric process model using heavy-tailed distributions. *Comput Stat Data Anal*. 2011;55:687–702.
- [7] Faddy MJ. Extended Poisson process modelling and analysis of count data. *Biom J*. 1997;39:431–440.
- [8] Conway RW, Maxwell WL. A queuing model with state dependent service rates. *J Ind Eng*. 1962;12:132–136.
- [9] Grunwald GK, Bruce SL, Jiang L, Strand M, Rabinovitch N. A statistical model for under- or overdispersed clustered and longitudinal count data. *Biomet J*. 2011;53:578–594.
- [10] Consul PC, Jain GC. A Generalization of the Poisson Distribution. *Technometrics*. 1973;15:791–799.
- [11] Efron B. Double exponential families and their use in generalized linear regression. *J Amer Statist Assoc*. 1986;81:709–721.
- [12] Ridout MS, Besbeas P. An empirical model for underdispersed count data. *Statist Model*. 2004;4:77–89.
- [13] Rodríguez-Avi J, Conde-Sánchez A, Sáez-Castillo AJ, Olmo-Jiménez MJ, Martínez-Rodríguez AM. A generalized Waring regression model for count data. *Comput Stat Data Anal*. 2009;53:3717–3725.
- [14] Li S, Yang F, Famoye F, Lee C, Black D. Quasi-negative binomial distribution: Properties and applications. *Comput Stat Data Anal*. 2011;55:2363–2371.
- [15] Lam Y. Geometric process and replacement problem. *Acta Math Appl Sin*. 1988;4:366–377.
- [16] Lam Y. A note on the optimal replacement problem. *Adv Appl Probab*. 1988;20:479–482.
- [17] Chan JSK, Leung DYP. Binary geometric process model for the modelling of longitudinal binary data with trend. *Comput Stat*. 2010;25:505–536.
- [18] Wan W-Y, Chan JSK. A new approach for handling longitudinal count data with zero-inflation and overdispersion: Poisson geometric process model. *Biom J*. 2009;51:556–570.
- [19] Chan JSK, Lam Y, Leung DYP. Statistical inference for geometric processes with gamma distributions. *Comput Stat Data Anal*. 2004;47:565–581.
- [20] Lam Y, Chan JSK. Statistical inference for geometric processes with lognormal distribution. *Comput Stat Data Anal*. 1998;27:99–112.
- [21] Lam Y, Zhu LX, Chan JSK, Liu Q. Analysis of data from a series of events by a geometric process model. *Acta Math Appl Sin English Ser*. 2004;20:263–282.
- [22] Chan JSK, Yu PLH, Lam Y, Ho APK. Modelling SARS data using threshold geometric process. *Stat Med*. 2006;25:1826–1839.
- [23] Chan JSK, Lam CPY, Yu PLH, Choy STB. Bayesian conditional autoregressive geometric process model for range data. *Comput Stat Data Anal*. 2012;56:3006–3019.
- [24] Chan JSK, Wan WY. Multivariate generalized Poisson geometric process model with scale mixture of normal distributions. *J Multivariate Anal*. 2014;127:72–84.
- [25] Baker J, Gor D. The Cannabis Cautioning Scheme three years on: an implementation and outcome evaluation. Drug and alcohol coordination, Sydney: NSW Police; 2004.
- [26] Walker SG, Gutiérrez-Peña E. Robustifying Bayesian procedures. In: Bernardo JM, Dawid AP, Smith AFM, editors. *Bayesian statistics*. Vol. 6, New York: Oxford; 1999. p. 685–710.
- [27] Consul PC. Generalized Poisson distributions: properties and applications. New York: Marcel Dekker; 1989.
- [28] Bae S, Famoye F, Wulu JT, Bartolucci AA, Singh KP. A rich family of generalized Poisson regression models with applications. *Math Comput Simul*. 2005;69:4–11.
- [29] Yau KKW, Wang K, Lee AH. Zero-inflated negative binomial mixed regression modelling of over-dispersed count data with extra zeros. *Biom J*. 2003;45:437–452.
- [30] Spiegelhalter D, Thomas A, Best NG, Lunn D. Bayesian inference using Gibbs sampling for Window version (WinBUGS) version 1.4.3. The University of Cambridge; 2004. Available from: www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml.
- [31] Brooks S, Gelman A. General methods for monitoring convergence of iterative simulations. *J Comput Graph Stat*. 1998;7:434–456.

- [32] Spiegelhalter DJ, Best NG, Carlin BP, Van der Linde A. Bayesian measures of model complexity and fit (with discussion). J R Statist Soc B. 2002;64:583–616.
- [33] Australian Institute of Health and Welfare. National Drug Strategy Household Survey: first results. Drug Statistics Series number 20. Cat. no. PHE 98. Canberra: AIHW; 2008.
- [34] Celeux G, Forbes F, Robert CP, Titterton DM. Deviance information criteria for missing data models. Bayesian Anal. 2006;1:651–674.
- [35] Gilks WR, Richardson S, Spiegelhalter DJ. Markov Chain Monte Carlo in Practice. London: Chapman and Hall; 1996.
- [36] Lambert D. Zero inflated Poisson regression with an application to defects in manufacturing. Technometrics. 1992;34:1–14.
- [37] Schwarz GE. Estimating the dimension of a model. Ann Stat. 1978;6:461–464.

Appendix

A. The univariate full conditional posterior densities for each of the unknown model parameters are given by

$$f(\beta_{\mu kl}|\text{rest}) \propto \exp \left[- \sum_{i=1}^m I_{il} \sum_{t=1}^{n_i} \left(r_l \beta_{\mu kl} z_{\mu kit} + \frac{r_l y_{itl}}{\mu_{itl}} \right) - \frac{\beta_{\mu kl}^2}{2\tau_{\mu kl}^2} \right];$$

$$f(\beta_{akl}|\text{rest}) \propto$$

$$\prod_{i=1}^m \prod_{t=1}^{n_i} \left\{ \left[\frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} + \lambda_l w_{itl} \right]^{w_{itl}-1} \exp \left[- \frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} - \beta_{akl} z_{akit}(t-1) \right] \right\}^{I_{il}} \exp \left(- \frac{\beta_{akl}^2}{2\tau_{akl}^2} \right),$$

$$\text{restricted to } \beta_{akl} < \min_{i,t} \left\{ \left[\frac{1}{t-1} \log \left(\frac{-y_{itl}(1-\lambda_l)}{s_{itl}\lambda_l} \right) - \xi_{kitl} \right] / z_{akit} \right\} \text{ if } \lambda_l < 0$$

$$\text{where } \xi_{kitl} = \beta_{a0l} + \beta_{a1l} z_{a1it} + \dots + \beta_{a,k-1,l} z_{a,k-1,it} + \beta_{a,k+1,l} z_{a,k+1,it} + \dots + \beta_{a,q,l} z_{a,qit}$$

$$f(y_{itl}|\text{rest}) \propto \left\{ y_{itl}^r \left[\frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} + \lambda_l w_{itl} \right]^{w_{itl}-1} \exp \left[- \frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} - \frac{r_l y_{itl}}{\mu_{itl}} \right] \right\}^{I_{il}},$$

$$\text{restricted to } y_{itl} > - \frac{s_{itl}\lambda_l a_{itl}^{t-1}}{1-\lambda_l} \text{ if } \lambda_l < 0;$$

$$f(\lambda_l|\text{rest}) \propto \prod_{i=1}^m \prod_{t=1}^{n_i} \left\{ (1-\lambda_l) \left[\frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} + \lambda_l w_{itl} \right]^{w_{itl}-1} \exp \left[- \frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} - \lambda_l w_{itl} \right] \right\}^{I_{il}},$$

$$\text{restricted to } \max_{i,t; y_{itl} < a_{itl}^{t-1} s_{itl}} \left(\left(1 - \frac{s_{itl} a_{itl}^{t-1}}{y_{itl}} \right)^{-1} \right) < \lambda_l < \min_{i,t; y_{itl} > a_{itl}^{t-1} s_{itl}} \left(\left(1 - \frac{s_{itl} a_{itl}^{t-1}}{y_{itl}} \right)^{-1} \right);$$

$$f(r_l|\text{rest}) \propto \prod_{i=1}^m \prod_{t=1}^{n_i} \left[\frac{r_l^r y_{itl}^r}{\mu_{itl}^r \Gamma(r_l)} \exp \left(- \frac{r_l y_{itl}}{\mu_{itl}} \right) \right]^{I_{il}} r_l^{d_l-1} \exp(-r_l e_l);$$

$$f(\pi_l|\text{rest}) \propto \pi_l^{\sum_{i=1}^m I_{il} + \alpha_l - 1};$$

$$f(I_i|\text{rest}) \propto \prod_{i=1}^G \pi_i'^{I_{il}} = \text{Multinomial}(1, \pi_{i1}', \dots, \pi_{iG}'),$$

$$\text{where } \pi_{il}' = \frac{\pi_l \prod_{t=1}^{n_i} \left(\frac{r_l y_{itl}^r}{\mu_{itl}^r \Gamma(r_l)} \frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} \left[\frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} + \lambda_l w_{itl} \right]^{w_{itl}-1} e^{-\left(\frac{y_{itl}(1-\lambda_l)}{a_{itl}^{t-1}} + \lambda_l w_{itl} + \frac{r_l y_{itl}}{\mu_{itl}} \right)} \right)}{\sum_{l'=1}^G \left\{ \pi_{l'} \prod_{t=1}^{n_i} \left(\frac{r_{l'} y_{itl'}^r}{\mu_{itl'}^r \Gamma(r_{l'})} \frac{y_{itl'}(1-\lambda_{l'})}{a_{itl'}^{t-1}} \left[\frac{y_{itl'}(1-\lambda_{l'})}{a_{itl'}^{t-1}} + \lambda_{l'} w_{itl'} \right]^{w_{itl'}-1} e^{-\left(\frac{y_{itl'}(1-\lambda_{l'})}{a_{itl'}^{t-1}} + \lambda_{l'} w_{itl'} + \frac{r_{l'} y_{itl'}}{\mu_{itl'}} \right)} \right\}}.$$

B. The WinBUGS code for MGPGP model in Table 5 is given by

```

1072 Model
1073 { for (i in 1:N) {
1074   one[i] <- 1
1075   one[i] ~ dbern(f[i])
1076
1077   ey1[i] <- exp(betau01)
1078   yscale1[i] <- r1/ey1[i]
1079   y1[i] ~ dgamma(r1,yscale1[i])
1080   x1[i] <- y1[i]/pow(exp(betaa01),time[i]-1)
1081   phi1[i] <- x1[i]*(1-lam1)
1082   term1[i] <- phi1[i]+w[i]*lam1
1083   f1[i] <- step(term1[i])*phi1[i]*pow(term1[i],w[i]-1)*exp(-1*term1[i])/exp(logfact(w[i]))
1084
1085   ey2[i] <- exp(betau02)
1086   yscale2[i] <- r2/ey2[i]
1087   y2[i] ~ dgamma(r2,yscale2[i])
1088   x2[i] <- y2[i]/pow(exp(betaa02),time[i]-1)
1089   x2r[i] <- x2[i]*(1-mgp[pat[i]])
1090   phi2[i] <- x2[i]*(1-lam2)
1091   term2[i] <- phi2[i]+w[i]*lam2
1092   f2[i] <- step(term2[i])*phi2[i]*pow(phi2[i],w[i]-1)*exp(-1*term2[i])/exp(logfact(w[i]))
1093   f[i] <- mgp[pat[i]]*f1[i]+(1- mgp[pat[i]])*f2[i] }
1094
1095 for (i in 1:T) {
1096   mgp[i] ~ dbern(pi) }
1097
1098   lam1 ~ dunif(-1,1)
1099   r1 ~ dgamma(0.1,0.05)I(0,100)
1100   betau01 ~ dnorm(0,0.1)
1101   betaa01 ~ dnorm(0,0.1)
1102   lam2 ~ dunif(0,1)
1103   r2 ~ dgamma(0.1,0.1)I(0,100)
1104   betau02 ~ dnorm(0,0.1)
1105   betaa02 ~ dnorm(0,0.1)
1106   pi ~ dunif(0,1)
1107 }
1108
1109 #starting values list(betau01=2.96,betaa01=0.0455,lam1=0.7238,r1=29.6,betau02=1.291,
1110   betaa02=0.0915,lam2=0.1, r2=2.7,pi=0.353, mgp=c(0,0,0,1,0,0,0,0,0,0, 1,1,1,1,1,0,1,0,0,0,
1111   0,0,1,1,1,0,1,1,1,0, 0,0,0,0,0,1,1,0,0,0, 1,1,1))
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122

```