Thailand Statistician
July 2025; 23(3): 460-480
http://statassoc.or.th
Contributed paper

Bayesian Estimation for {ij}-Inflated Mixture Power Series Distributions using an EM Algorithm

Amir T. Payandeh Najafabadi * and Mansoureh Sakizadeh
Department of Mathematical Sciences, Shahid Beheshti University,
G.C. Evin, Tehran, Iran.
*Corresponding author; e-mail: amirtpayandeh@sbu.ac.ir

Received: 21 August 2021
Revised: 23 December 2022
Accepted: 4 January 2023

Abstract

The purpose of this article was to illustrate how we can model different behaviors of subpopulations by introducing a mixing distribution/regression model with {ij}-inflated power series. An EM algorithm was used to estimate the parameters of the models. As a practical application, the new model has been applied to the design of an optimal rate-making system. More precisely, this article employs a number of reported claims from an Iranian third party insurance dataset, under an $\{ij\}$-inflated power series mixture distribution/regression model to estimate rate premium for such insurance contract. The numerical study shows that the $\{ij\}$-inflated negative Binomial mixture models have the potential to provide more appropriate rate premiums for policyholders under a rate-making system with four categories.

Keywords: Power Series distribution/regression, mixture model, inflated model, EM algorithm, rate-making system.

1. Introduction

Suppose Y stands for the number of the reported claim for a non-life insurance contract. In the situation that a significantly large portion of the population has zero claims, the intensity rate λ will be closed to zero and consequently, the zero-inflated poisson model becomes superior to the poisson model! Unfortunately, in the situation that $\lambda \simeq 1$ since $P(Y = 0|\lambda \simeq 1) \simeq P(Y = 1|\lambda \simeq 1)$, the zero-inflated poisson model cannot be an appropriate choice anymore! Such a situation arrives whenever different risk behaviors can be observed in subpopulations. For instance, a subpopulation reports about two claims while another subpopulation has zero risk level. A zero-inflated model assumes that the zero observations have two different origins, structural and sampling. The sampling zeros are due to the usual poisson (or negative binomial) distribution, which assumes that those zero observations, happened by chance. Zero-inflated models assume that some zeros are observed due to some specific structure in the data. For example, if the number of claimed damages is the result, some insureds may even announce zero damage even if there is a fault; these are structural attributes because they do not want to pay more premiums in the next year. Therefore, their risk behavior is assumed to be a poisson (or negative binomial) distribution that includes both zero (the sampling zeros) and non-zero counts. n contrast, a hurdle model assumes that all zeros are from one structural source. For example, consider a study of cocaine users in which a secondary outcome is the number of tobacco cigarettes smoked during the last month. In this case, it is safe to assume that only non-smokers will smoke zero cigarettes during the last month and smokers will score a positive number of cigarettes during the last month. Hence the zero observations can come from only one structural source, the non-smokers. If a subject is considered a smoker, they do not have the ability to score zero cigarettes smoked during the last month and will always score a positive number of cigarettes in a hurdle model with either truncated poisson (or truncated negative binomial) distribution. The zero-one-inflated models are another appropriate models which reflates some behavior of a phenomenon under study. Many authors employed these models, for example, Alshkaki (2016) studied some properties of zero-one inflated poisson distribution. Liu, Tang, and Xu (2021) derived the maximum likelihood and Bayesian estimations for a zero-one-inflated poisson regression model. Wieczorek, Nugent, and Hawala (2012) employed a zero-one inflated beta model to estimate the poverty rate. Nishii and Tanaka (2013) used a zero-one inflated distribution with spatial dependence to study the forest ratios.

In many practical situations, the popular classical poisson regression model has failed to model count data because of over dispersion or model assumptions cannot be strictly verified. In such cases, the power series family of distributions can be employed. Johnson, Kemp, and Kotz (2005) emphasized that the poisson distribution cannot be verified in a situation where strict requirements are available. Thus, the negative binomial distribution is regarded as an appropriate choice. To have an appropriate class of classical distributions, to model count data arrive from different practice areas such as Economics, Finance, Insurance, Ecological, etc. Rolski et al. (1999) introduced the generalized power series distributions. Surprisingly, in the practical applications, this class of distributions does not receive enough attention from authors. Most practical studies have focused on the negative binomial distribution/regression model. For example, Aryuyuen and Bodhisuwan (2013) employed the negative binomial model for heavy-tailed observation. For count data, the over-dispersed behavior has been arrived whenever *either* observing the excess of a single value more than the average, under the model, *or* the target population including several subpopulations with different behaviors. Several authors modeled the over-dispersed behavior by using the k-inflated or mixture models. For instance, the poisson mixture model (Simar, 1976), the zero-inflated negative binomial regression model (Greene, 1994 and Hall, 2000), the zero-inflated poisson (Lambert, 1992), the zero-inflated power series distribution (P. Gupta, R. Gupta, and Tripathi, 1996), k-inflated poisson mixture model (Lim, Lee, and Philip, 2014) and negative binomial mixture model (Tzougas and Frangos, 2014). Also, there are some authors that employ two or more inflated points to model the count data; Stewart (2014) used multi-points inflated poisson distribution as a comparison to the standard poisson distribution. For modelings, complete female fertility, Melkersson, and Rooth (2000) proposed a zero-and-two-inflated count data model, which accounts for a relative excess of both zero or one child. Also, Joshi (2015) considered a Swedish fertility dataset with inflated values of some particularly count. For censored mixture Topp-Leone distributions see Sindhu et al. (2019). An industrial application of mixture Gumbel distributions has beeb given by Sindhu et al. (2016). The Burr mixture distributions has been studied by Sindhu and Aslam (2014). Bayesian analysis of the censored shifted Gompertz mixture distribution, half-normal distributions and the 3-components Kumaraswamy mixture distributions have been studied by Sindhu et al. (2016), Sindhu et al. (2018) and Khalid et al. (2020), respectively. The reliability estimation of inverted Maxwell mixture model has been derived by Sindhu et al. (2019).

The rest of this article is organized as follows. Section 2 gives a brief overview of the $\{ij\}$-inflated power series mixture model and some of its properties. Section 3 elaborated on the EM algorithm to estimate the model's parameters. Section 4 discusses the applications of the $\{ij\}$-inflated power series mixture model to estimate the rate premium of a rate-making system. Section 5 employs the present model against the Iranian third-party insurance dataset to calculate the rate premium for such an insurance policy. Some theoretical elements are represented in Appendix.

2. $\{ij\}$-inflated Power Series Mixture Distribution/Regression Model

The probability mass function of a large class of count random variables can be restated as a power series distribution with the following probability mass function.

$$P(Y=y|\lambda)=\frac{\lambda^y a(y)}{g(\lambda)} \quad y\in S; \quad \lambda>0 \tag{1}$$

where $g(\lambda)$ is a given, positive, finite and differentiable function, S indicates any nonempty enumerable set of nonnegative integers, $a(y)\geq 0$ and $g(\lambda)=\sum_{y\in S}\lambda^y a(y)$.(see Heras, Moreno and Vilar-Zanón, 2018). Since the probability mass function (1) can be represented as

$$P(Y=y|\lambda) \quad = \quad \exp\{yln\lambda+lna(y)-lng(\lambda)\}.$$

The power series distributions belong to the exponential family of distributions. Moreover, cumulated distribution function and average of the power series distributions, respectively, are $F(Y=y|\lambda)=g_y(\lambda)/g(\lambda)$ and $E(Y)=\lambda g'(\lambda)/g(\lambda)$. The binomial, negative binomial, logarithmic series, and poisson distributions are, some well known, member of the power series distributions (see Table 1).

Table 1 Power series class of distribution

Distribution name	λ	$g(\lambda)$	a(y)	S
binomial(n,p)	$\frac{p}{1+p}$	$(1+\lambda)^n$	$\frac{n!}{y!(n-y)!}$	$y\in\{0,1,...,n\}$
poisson(λ)	λ	e^λ	$\frac{1}{y!}$	$y\in\{0,1,...\}$
negative binomial(r,p)	$1-p$	$(1-\lambda)^r$	$\frac{\Gamma(r+y)}{y!\Gamma(r)}$	$y\in\{0,1,...\}$
logarithmic series (λ)	$1-p$	$-ln(1-\lambda)$	$\frac{1}{y!}$	$y\in\{0,1,...\}$

The $\{ij\}$-inflated power series, says $\{ij\}-IPS$ arrives by combining a power series distribution with doubled masses at points i and j. The probability mass function for an $\{ij\}$-inflated power series distribution is given by

$$\begin{aligned} P(Y=y|\lambda) \quad = \quad & \left(p_1+(1-p_1)(1-p_2)\frac{\lambda^y a(y)}{g(\lambda)}\right) I_{\{i\}}(y) \\ & +\left(p_2(1-p_1)+(1-p_1)(1-p_2)\frac{\lambda^y a(y)}{g(\lambda)}\right) I_{\{j\}}(y) \\ & +(1-p_1)(1-p_2)\frac{\lambda^y a(y)}{g(\lambda)} I_{\{0,1,\cdots\}\backslash\{i,j\}}(y), \end{aligned} \tag{2}$$

where $I_{.}(\cdot)$ stands for indicator function and nonnegative $p_1, p_2\in[0,1]$. The above definition for the probability mass function for an $\{ij\}$-inflated power series distribution has been adopted from Razie et al. (2016).

The $\{ij\}-IPSM_m$ model can be arrived at by combining a {ij}-inflated power series distribution and a mixture distribution as the prior distribution for the parameter. For example, Simon (1991) derived the negative binomial distribution from the combination of poisson and gamma distributions. The following proposition shows such construction.

Proposition 1 *Suppose that for an z^{th} individual, information on count response variables $Y_{z1},...,Y_{zt}$ along with information on p covariates $X_1,...,X_p$ are available. Also suppose that Y_{zl} given parameter $\Lambda_{zl}=\lambda_{zl}$ has been distributed according to an $\{ij\}$-inflated power series with probability* *mass function given by Equation (2). Moreover, suppose that parameter λ_{zl} can be evaluated by the following regression model*

$$\ln(\lambda_{zl}) = \boldsymbol{X}_z\boldsymbol{B}_{zl} + \varepsilon_z,$$

where ε_z stands for the error of regression model, $\boldsymbol{X}_z = (1, X_{1z}, ..., X_{pz})$ and $\boldsymbol{B}_{zl} = (\beta_{0zl}, ..., \beta_{pzl})'$ is the regression coefficients vector and $u_z = \exp\{\varepsilon_z\}$ has been distributed according to a finite mixture distribution with density function $f_{U_z}(u_z) = \sum_{k=1}^{m} \varphi_k f_{U_z}^k(u_z|\alpha_k)$, where $\sum_{k=1}^{m} \varphi_k = 1$, $\varphi_k \in [0,1]$ and $\alpha_k > 0$ and $\lambda_{zl} = u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}}$ then,

$$\begin{aligned} P\left(Y_{zl} = y_{zl}|\boldsymbol{X}_z, \boldsymbol{B}_{zl}, \alpha_k, \varphi_k, p_1, p_2\right) &= p_1 I_{\{i\}}(y_{zl}) + p_2(1-p_1)I_{\{j\}}(y_{zl}) \\ &\quad + (1-p_1)(1-p_2)\sum_{k=1}^{m}\varphi_k \int_0^\infty f_{U_z}^k(u_z|\alpha_k)M(u_z)du_z, \end{aligned}$$

where $M(u_z) = \frac{a(i)u_z^i e^{i(\boldsymbol{X}_z\boldsymbol{B}_{zl})}}{g(u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}})} I_{\{i\}}(y_{zl}) + \frac{a(j)u_z^j e^{j(\boldsymbol{X}_z\boldsymbol{B}_{zl})}}{g(u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}})} I_{\{j\}}(y_{zl}) + \frac{a(y_{zl})u_z^{y_{zl}} e^{y_{zl}(\boldsymbol{X}_z\boldsymbol{B}_{zl})}}{g(u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}})} I_{\{A\}}(y_{zl})$, $I.(\cdot)$ stands for indicator function and $A = \{0, 1, \cdots\} \setminus \{i, j\}$.

Proof: The desire proof arrives by the fact that

$$\begin{aligned} P\left(Y_{zl} = y_{zl}|\mathbf{X}_z, \mathbf{B}_{zl}, \alpha_k, \varphi_k, p_1, p_2\right) &= \int_0^\infty P(Y_{zl} = y_{zl}|\mathbf{X}_z, \mathbf{B}_{zl}, \alpha_k, \varphi_k, p_1, p_2, u_z) \\ &\quad \times \sum_{k=1}^{m} \varphi_k f_{U_z}^k(u_z|\alpha_k)du_z. \ \square \end{aligned}$$

Now, suppose that all parameters, except λ_z, are given. The following provides a Bayes estimator for the parameter λ_z.

Proposition 2 *Suppose that $\boldsymbol{Y}_z = (Y_{z1}, Y_{z2}, \cdots, Y_{zt})$ given parameter $\Lambda_{zl} = \lambda_{zl}$ is distributed according to a $\{ij\}$-inflated power series distribution with probability mass function given by Equation (2). In addition, suppose that parameter Λ_{zl} can be evaluated by the following regression model $\ln(\lambda_{zl}) = \boldsymbol{X}_z\boldsymbol{B}_{zl} + \varepsilon_z$, where $\boldsymbol{B}$ is the regression coefficients vector and $u_z = e^{\varepsilon_z}$ is distributed according to a finite mixture distribution with density function $f_{U_z}(u_z) = \sum_{k=1}^{m} \varphi_k f_{U_z}^k(u_k|\alpha_k)$, where $\sum_{k=1}^{m} \varphi_k = 1$ and $\varphi_k \in [0,1]$. Then the Bayes estimator for the parameter λ_z, is given as follow*

$$\widehat{\lambda}_{z(Bayes)} = \frac{\boldsymbol{I}}{\boldsymbol{II}},$$

where $\boldsymbol{I} = \int_0^\infty e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}} u_z \prod_{l=1}^{t}((p_1 + (1-p_1)(1-p_2)I_{(y_{zl}=i)}) + (p_2(1-p_1) + (1-p_1)(1-p_2)I_{(y_{zl}=j)}) + I_A(y_z)(1-p_1)(1-p_2))\frac{u_z^{y_{zl}} e^{y_{zl}(\boldsymbol{X}_z\boldsymbol{B}_{zl})} a(y_{zl})}{g(u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}})}) \sum_{k=1}^{m} \varphi_k f_{U_z}^k(u_z|\alpha_k)du_z$, $A = \{0, 1, \cdots\} \setminus \{i, j\}$, $I.(\cdot)$ stands for indicator function and $\boldsymbol{II} = \int_0^\infty \prod_{l=1}^{t}((p_1 + (1-p_1)(1-p_2)I_{(y_{zl}=i)}) + (p_2(1-p_1) + (1-p_1)(1-p_2)I_{(y_{zl}=j)}) + I_A(y_z)(1-p_1)(1-p_2))\frac{u_z^{y_{zl}} e^{y_{zl}(\boldsymbol{X}_z\boldsymbol{B}_{zl})} a(y_{zl})}{g(u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}})}) \sum_{k=1}^{m} \varphi_k f_{U_z}^k(u_z|\alpha_k)du_z$.

Proof: Posterior distribution can be written as

$$f_{\Lambda_z=\lambda_{zl}|(Y_z,X_z)}(\lambda_{zt+1}) = \frac{\prod_{l=1}^{t} P(Y=y_{zl}|\Lambda_{zl})P(\Lambda_{zl}=e^{\mathbf{X}_z\mathbf{B}_{zl}}u_z)}{\int_0^\infty \prod_{l=1}^{t} P(Y=y_{zl}|\Lambda_{zl})P(\Lambda_{zl}=e^{\mathbf{X}_z\mathbf{B}_{zl}}u_z)du_z} = \frac{\prod_{l=1}^{t} III_l}{\int_0^\infty \prod_{l=1}^{t} IV_l(u_z)du_z},$$

where $III_l = ((p_1+(1-p_1)(1-p_2)I_{(y_{zl}=i)})+(p_2(1-p_1)+(1-p_1)(1-p_2)I_{(y_{zl}=j)})+I_A(y_z)(1-p_1)(1-p_2))II^{**}$, $IV_l(u_z) = ((p_1+(1-p_1)(1-p_2)I_{(y_{zl}=i)})+(p_2(1-p_1)+(1-p_1)(1-p_2)I_{(y_{zl}=j)})+I_A(y_z)(1-p_1)(1-p_2))II^{**}$ and $II^{**} = \frac{u_z^{y_{zl}}e^{y_{zl}(\mathbf{X}_z\mathbf{B}_{zl})}a(y_{zl})}{g(u_z e^{\mathbf{X}_z\mathbf{B}_{zl}})})\sum_{k=1}^{m}\varphi_k f_{U_z}(u_z|\alpha_k)$.

Now, the desired result arrives by

$$\widehat{\lambda}_{z(Bayes)} = \int_0^\infty e^{\mathbf{X}_z\mathbf{B}_{zl}}u_z f_{\Lambda_z=\lambda_{zl}|(Y_z,X_z)}(\lambda_z)du_z. \ \square$$

In Proposition 2, we assumed that all parameters, except λ_z, are given. In the situation that some of the parameters are unknown, one may employ an Expectation-Maximization algorithm to derive a maximum-likelihood estimate for such parameters, see Appendix for more details.

3. Application to a rate-making system

Rate-making, or insurance pricing, is regarded as the determination of the rates charged by insurance companies. Rate-making is useful for ensuring those insurance companies are setting fair and adequate premiums given the competitive nature. A large number of studies have focused on rate-making systems (or Bonus-Malus systems). Heras et al. (2018) designed an alternative method based on quantile regression, which can provide more information about the loss distribution and can be used for insurance underwriting. Lange (1969) provided several mathematical tools for pricing a Bonus-Malus system. Dionne and Vanasse (1989, 1992) employed available asymmetric information under the poisson and the negative binomial regression models to determine the premium of a rate-making system. In 1995, Lemaire designed an optimal Bonus-Malus system based on a negative binomial distribution. In addition, Pinquet (1997, 1998) considered the poisson and the lognormal distributions to design an optimal Bonus-Malus system. Further, Walhin and Paris (1999) considered a Hofmanns distribution, along with a finite mixture poisson distribution, in order to evaluate the elements of a Bonus-Malus system. Furthermore, Denuit and Dhaene (2001) proposed the rate premium of a rate-making system under the exponential loss function. Lee and Frees (2016) provided a comprehensive overview of deductible rate-making and compared the advantages and disadvantages of various approaches under different parametric models. The regression approach is useful for predicting aggregate claims when deductible choices influence the frequency and severity distributions. Bermúdez and Karlis (2011) employed a Bayesian multivariate poisson model to determine the premium of a rate-making system which includes a non-ignorable correlation between different types of its claims. Further, Boucher and Inoussa (2014) introduced a new model to determine the premium of a rate-making system when a panel or longitudinal data are available. In another study, Tzougas and Frangos (2014) emphasized the Sichel distribution, along with a negative binomial distribution. Furthermore, Tzougas, Vrontos, and Frangos (2014) and Payandeh Najafabadi and Sakizadeh (2019) employed finite mixture distribution to model the frequency and severity of accidents. Payandeh

Najafabadi, Atatalab, and Omidi Najafabadi (2017) employed the ideas suggested by Payandeh Najafabadi (2010) to determine the credibility premium for a rate-making system when the number of reported claims was distributed according to a zero-inflated poisson distribution. Under a rate-making system, the pure premium of an i^{th} policyholder at $(t+1)^{\text{th}}$ year is estimated by multiplying the estimated base premium, say $BP(t+1)$, into the corresponding estimated rate premium. Based on the decision theory, viewpoint, the Bayes estimator offers an intellectual and acceptable estimation for both the rate premium $Rate(t+1)$ and the base premium $BP(t+1)$. Under the quadratic loss function, the Bayes estimators for $Rate(t+1)$ and $BP(t+1)$ can be obtained by a posterior expectation of the risk parameters of frequency and severity of reported claims at the $(t+1)^{\text{th}}$ year, respectively (Denuit et al., 2007). The following develops the Bayes estimator for $Rate(t+1)$ under an $\{ij\}-IPSM_m$ model.

Proposition 3 *Suppose that $\boldsymbol{Y}_z = (Y_{z1}, Y_{z2}, \cdots, Y_{zt})$ stands for the number of reported claims in last t years for an z^{th} policyholder. In addition, suppose that given parameter $\Lambda_{zl} = \lambda_{zl}$ random variable Y_{zl} has been distributed according to a $\{ij\}$-IPS distribution with probability mass function, given by (2). Furthermore, suppose that parameter Λ_{zl} can be evaluated by the following regression model $\ln(\lambda_{zl}) = \boldsymbol{X}_z\boldsymbol{B}_{zl} + \varepsilon_z$, where $\boldsymbol{B}_{zl}$ is the regression coefficients vector and $u_z = e^{\varepsilon_z}$ is distributed according to a finite mixture gamma distribution with density function*

$$f_{U_z}(u_z) = \sum_{k=1}^{m} \varphi_k u_z^{\alpha_k - 1} \alpha_k^{\alpha_k} e^{-\alpha_k u_z} / \Gamma(\alpha_k),$$

where $\sum_{k=1}^{m} \varphi_k = 1$ and to have $E(\varepsilon_z) = 0$, we set both parameters of gamma distributions in the finite mixture gamma distribution to be equal.Then, the Bayes estimator for the rate-making $\widehat{Rate}_z(t+1)$ of an z^{th} policyholder at $(t+1)^{th}$ year, is given by

$$\widehat{Rate}_z(t+1) = e^{\boldsymbol{X}_z\boldsymbol{B}_{zt}} \frac{\int_0^\infty u_z \prod_{l=1}^{t} V_l(u_z) du_z}{\int_0^\infty \prod_{l=1}^{t} VI_l(u_z) du_z},$$

where $I.(\cdot)$ stands for indicator function, $A^* = \{0, 1, \cdots\} \setminus \{i, j\}$, $V_l(u_z) = ((p_1 + (1-p_1)(1-p_2)I_{(y_{zl}=i)}) + (p_2(1-p_1) + (1-p_1)(1-p_2)I_{(y_{zl}=j)}) + I_{A^*}(y_z)(1-p_1)(1-p_2))II^{**}$ and $VI_l(u_z) = ((p_1+(1-p_1)(1-p_2)I_{(y_{zl}=i)}) + (p_2(1-p_1) + (1-p_1)(1-p_2)I_{(y_{zl}=j)}) + I_{A^*}(y_z)(1-p_1)(1-p_2))II^{**}$ and $II^{**} = \frac{u_z^{y_{zl}} e^{y_{zl}(\boldsymbol{X}_z\boldsymbol{B}_{zl})} a(y_{zl})}{g(u_z e^{\boldsymbol{X}_z\boldsymbol{B}_{zl}})}) \sum_{k=1}^{m} \frac{\varphi_k u_z^{\alpha_k-1} \alpha_k^{\alpha_k} e^{-\alpha_k u_z}}{\Gamma(\alpha_k)}$.

Proof: The posterior distribution of Λ_{zt+1} is given by

$$\begin{aligned} f_{\Lambda_{zt+1}=\lambda_{zl}|(Y_z,X_z)}(\lambda_{zt+1}) &= \frac{\prod_{l=1}^{t} P(Y = y_{zl}|\Lambda_{zl}) P(\Lambda_{zl} = e^{\mathbf{X}_z\mathbf{B}_{zl}} u_z)}{\int_0^\infty \prod_{l=1}^{t} P(Y = y_{zl}|\Lambda_{zl}) P(\Lambda_{zl} = e^{\mathbf{X}_z\mathbf{B}_{zl}} u_z) du_z} \\ &= \frac{\prod_{l=1}^{t} VII_l}{\int_0^\infty \prod_{l=1}^{t} VIII_l}, \end{aligned}$$

where $VII_l = ((p_1 + (1-p_1)(1-p_2)I_{(y_{zl}=i)}) + (p_2(1-p_1) + (1-p_1)(1-p_2)I_{(y_{zl}=j)}) + I_{A^*}(y_z)(1-p_1)(1-p_2))\frac{u_z^{y_{zl}} e^{y_{zl}(\mathbf{X}_z\mathbf{B}_{zl})} a(y_{zl})}{g(u_z e^{\mathbf{X}_z\mathbf{B}_{zl}})})I_{A^*}$, $VIII_l = ((p_1 + (1-p_1)(1-p_2)I_{(y_{zl}=i)}) +$

$$(p_2(1-p_1)+(1-p_1)(1-p_2)I_{(y_{zl}=j)})+I_{A^*}(y_z)(1-p_1)(1-p_2))\frac{u_z^{y_{zl}}e^{y_{zl}(\mathbf{X}_z\mathbf{B}_{zl})}a(y_{zl})}{g(u_z e^{\mathbf{X}_z\mathbf{B}_{zl}})})B(u_z)du_z$$

and $A=\sum_{k=1}^{m}\varphi_k u_z^{\alpha_k-1}\alpha_k^{\alpha_k}e^{-\alpha_k u_z}/\Gamma(\alpha_k)$ and $B(u_z)=\sum_{k=1}^{m}\varphi_k u_z^{\alpha_k-1}\alpha_k^{\alpha_k}e^{-\alpha_k u_z}/\Gamma(\alpha_k)$.

The desired result arrives by

$$\widehat{Rate}_z(t+1)=\int_0^{\infty}e^{\mathbf{X}_z\mathbf{B}_{zl}}u_z f_{\Lambda_{z(t+1)}=\lambda_{zt}|(Y_z,X_z)}(\lambda_{z(t+1)})du_z. \ \square$$

To employ the above results, other parameters should be given. But in many practical situations, some (or all) parameters are unknown, therefore, one has to employ an Expectation-Maximisation algorithm to derive a maximum-likelihood estimate for such parameters. An appendix presents such estimation under the $\{ij\}$-inflated poisson and negative binomial mixture models.

4. Numerical Application

In this section, the available data from Iranian third-party liability during 2011 were considered. After a primary evaluation, the data related to 3584 policyholders were taken into account. In this regard, two independent variables were considered as covariates. Table 2 and Table 3 represent the frequency of claims under the category of covariates. The initial information about each policyholder was obtained at the beginning of the study and such covariates were used to model the frequency of claims for evaluating the pure premium of each policyholder under a rate-making system.

Table 2 Available covariates information for each policyholder

Variable	Description
Gender	Equal to 0 for woman and 1 for man
Age	Equal to 1 for 18≤age<30 ; 2 for 30≤age<40 ; 3 for 40≤age<50 ; 4 for 50≤age

In order to find an appropriate distribution for the frequency of claims, the following bar chart, Figure 1, has been considered. As one may observe from Figure 1, with or without zero and one, we come up with a different suggestion for the distribution of the number of reported claims. Therefore, we should candidate a distribution from the class of $\{ij\}-IPSM_m$ distributions given in Table 1. Then, to estimate unknown parameters, we develop our R codes. Tables 4 and 5 indicate the maximum likelihood estimator for significant parameters of such distributions.

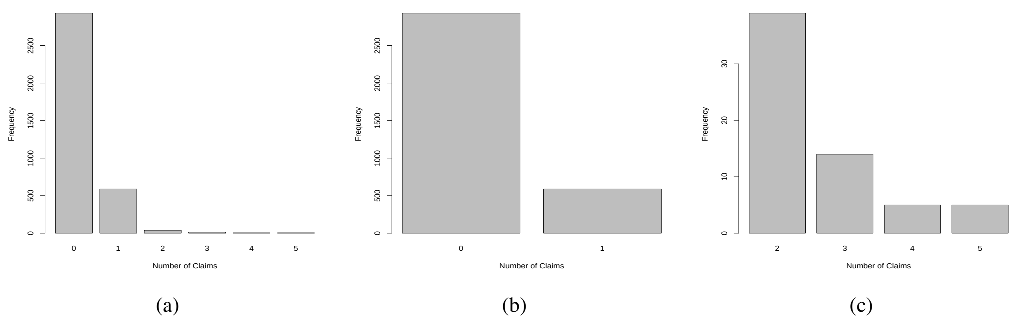

Figure 1 Part (a): Bar chart for all observation; Part (b): Bar chart for zero and one reported claims; Part (c): Bar chart for more than one reported claims.

Now, those covariates which may influence on response variable for each regression model were found by using a backward elimination selection method. Table 6 shows the result of the backward selection method for frequency of accidents.

Table 3 Descriptive statistics for frequency of claims under category of covariate

Category	Count	Mean	S.D.	Skewness	Kurtosis
Total	3583	0.21	0.01	3.48	18.97
Gender: Woman	601	0.17	0.22	4.96	31.95
Gender: Man	2982	0.22	0.01	3.10	15.49
18<Age<30	721	0.31	0.22	2.66	11.61
30<Age<40	951	0.16	0.15	4.07	25.66
40<Age<50	1312	0.18	0.13	3.74	20.77
Age>50	599	0.25	0.22	3.49	20.44

Table 4 Estimation for parameters on various model for frequency

	NB	$\{0\}NBM_1$	$\{1\}NBM_1$	$\{2\}NBM_1$	$\{-\}NBM_2$	$\{12\}NBM_2$	$\{10\}NBM_2$	$\{20\}NBM_2$
df	2	3	3	3	5	7	7	7
mu1	0.21	0.21	0.10	0.21	0.20	0.20	0.18	0.24
sigma1	0.74	0.69	9.28	0.69	0.00	4.42	5.24	0.50
p11	-	0.00	0.12	0.00	-	0.13	0.02	0.00
p12	-	1.00	0.88	1.00	0.82	0.41	0.46	0.87
mu2	-	-	-	-	0.26	0.00	0.04	0.00
sigma2	-	-	-	-	4.59	1.03	0.00	1.03
p21	-	-	-	-	-	0.00	0.09	0.00
p22	-	-	-	-	0.18	0.46	0.43	0.13

Table 5 Estimation for parameters on various model for frequency

	POI	$\{0\}POIM_1$	$\{1\}POIM_1$	$\{2\}POIM_1$	$\{-\}POIM_1$	$\{01\}POIM_1$	$\{02\}POIM_1$	$\{12\}POIM_1$
df	1	2	2	2	3	5	5	5
mu1	0.21	0.29	0.21	0.21	0.50	0.30	0.21	0.00
p11	-	0.29	0.00	0.00	-	0.00	0.00	0.05
p12	-	-	-	-	0.23	0.70	0.89	0.81
mu2	-	-	-	-	0.12	0.00	2.78	1.47
p21	-	-	-	-	-	0.03	0.10	0.10
p22	-	-	-	-	0.77	0.27	0.01	0.04

Table 6 Regression coefficients for various model for frequency

Distribution	DF	Intercept B10	Gender B11	Age B12	sigma1	P11	P12	Intercept B20	Gender B21	Age B22	sigma2	P21	P22
NB	4	0.96-	-0.30	-0.10	0.66	-	-	-	-	-	-	-	-
$\{0\}NBM_1$	5	-0.96	-0.30	-0.10	0.66	0.00	1.00	-	-	-	-	-	-
$\{1\}NBM_1$	5	-1.95	-0.04	-0.11	8.85	0.12	0.88	-	-	-	-	-	-
$\{2\}NBM_1$	5	-0.96	-0.30	-0.10	0.66	0.00	1.00	-	-	-	-	-	-
$\{-\}NBM_2$	8	-72.14	14.66	14.17	0.00		0.19	-0.19	-0.61	-0.24	0.47	-	0.81
$\{10\}NBM_2$	10	-1.34	0.02	-0.10	4.17	0.12	0.41	36.04-	0.00	0.00	1.03	0.00	0.46
POI	3	0.93-	-0.31	-0.11	-	-	-	-	-	-	-	-	-
$\{0\}POIM_1$	4	-0.68	-0.29	-0.10	-	0.26	0.74	-	-	-	-	-	-
$\{1\}POIM_1$	4	-0.93	-0.31	-0.11	-	0.00	1.00	-	-	-	-	-	-
$\{2\}POIM_1$	4	-0.93	-0.31	-0.11	-	0.00	1.00	-	-	-	-	-	-
$\{-\}POIM_2$	7	-0.45	-0.22	-0.21	-	-	0.66	-41.46	-12.87	13.27	-	-	0.34
$\{01\}POIM_2$	9	-73.04	35.05	0.90	-	0.00	0.29	34.08	-34.73	0.22-	-	0.00	0.71

In this section $\{ij\} - IPSM_2$ is used as follow:

$$\{ij\} - IPSM_2 \sim \begin{cases} p_1 + (1-p_1)(1-p_2)\left(\varphi_1 f_1(i) + \varphi_2 f_2(i)\right) & y_{zl} = i \\ p_2(1-p_1) + (1-p_1)(1-p_2)\left(\varphi_1 f_1(j) + \varphi_2 f_2(j)\right) & y_{zl} = j \\ (1-p_1)(1-p_2)\left(\varphi_1 f_1(y_{zl}) + \varphi_2 f_2(y_{zl})\right) & y_{zl} \neq i, y_{zl} \neq j \end{cases}$$

where $A = \{0, 1, \cdots\} \setminus \{i, j\}$, $I_{\cdot}(\cdot)$ stands for indicator function, $f(y_z \mid \boldsymbol{\theta})$ is poisson and negative binomial distribution and the vector $\boldsymbol{\theta}$ represents all of unknown own parameters. for comparing, one point inflated "$\{i\}$", tow point inflated "$\{ij\}$" and power series without inflated point "$\{-\}$" are considered. for example:

$$\{20\} - NBM_2 \sim p_{11} I_{\{2\}}(y_z) + p_{12} I_{\{0\}}(y_z) + \sum_{k=1}^{2} p_{k2} I_A(y_z) f(y_z \mid \boldsymbol{\theta}),$$

where $A = \{0, 1, \cdots\} \setminus \{i, j\}$, $I_{\cdot}(\cdot)$ stands for indicator function, $f(y_z \mid \boldsymbol{\theta})$ is negative binomial distribution.
In parameter estimation, in negative binomial case with parameter (N, λ_k) and $\lambda_k \sim Gamma(\alpha_k)$, $\alpha_k = \mu_k, N_k = \sigma_k$, also if $\lambda_k = \mathbf{X}_k \mathbf{B}_k + \varepsilon_k$ then tow regression coefficients gender and age are considered. In poisson case with parameter (λ_k) and $\lambda_k \sim Gamma(\alpha_k)$, $\alpha_k = \mu_k$, also if $\lambda_k = \mathbf{X}_k \mathbf{B}_k + \varepsilon_k$ then tow regression coefficients gender and age are considered.

4.1. Model comparison

In order to obtain an appropriate model for a given rate-making system, in the first step, we focus on the $\{ij\} - IPSM_m$, given in Table 1. Then, two evaluation approaches were employed to compare the result of regression/ distribution models.
More precisely,

1. in the first approach, each fitted distribution is employed 200 times to simulate 8874 data in order to study the performance of count distributions. Then, using the mean square error, say MSE, criteria, the stimulated data were compared with the observed data. Tables 7 to 13 indicate the result of such a comparison.

2. in the second approach, the Akaike Information Criterion, say AIC, the Global Deviance, say GD, and the Schwarz Bayesian Information Criterion, say SBIC, are used to compare regression/distribution models for frequency of claims. Table 14 represents the result of the comparison.

Table 7 MSE of frequency for **total** observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	2932	22433.68	2932.00	2932.00	2932.00	389385.81	1.82501E+34	522171.86	4.71765E+34
1	588	1834.04	0.00	0.00	0.00	65125.28	3.65999E+33	89613.38	9.46104E+33
2	39	22.89	39.00	39.00	39.00	3537.65	2.42754E+32	5019.82	6.27518E+32
3	14	0.77	56.00	56.00	56.00	1017.25	8.71425E+31	1498.32	2.25263E+32
4	5	7.61	45.00	45.00	45.00	283.06	3.11223E+31	436.66	8.0451E+31
5	5	24.95	80.00	80.00	80.00	212.82	3.11223E+31	348.21	8.0451E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	2932	129.73	722.80	129.73	129.73	5856.46	3487.19	35809.96	14944.87
1	588	366.65	149.06	366.65	366.65	100.44	4.82	3659.67	930.09
2	39	124.91	88.16	124.91	124.91	13.42	32.26	87.14	2.59
3	14	108.95	87.74	108.95	108.95	35.25	51.04	3.43	7.71
4	5	71.81	61.37	71.81	71.81	33.45	42.32	1.28	15.18
5	5	114.70	101.41	114.70	114.70	64.32	76.42	11.33	37.60

Table 8 MSE of frequency for **Gender=Man** observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	2405	18401.43	2405.00	2405.00	2405.00	319397.30	1.49698E+34	428316.28	3.86969E+34
1	527	1643.77	0.00	0.00	0.00	58369.08	3.28029E+33	80316.75	8.47953E+33
2	33	19.37	33.00	33.00	33.00	2993.40	2.05407E+32	4247.54	5.30977E+32
3	11	0.60	44.00	44.00	44.00	799.27	6.84691E+31	1177.25	1.76992E+32
4	3	4.57	27.00	27.00	27.00	169.84	1.86734E+31	262.00	4.82706E+31
5	3	14.97	48.00	48.00	48.00	127.69	1.86734E+31	208.93	4.82706E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	2405	106.41	592.88	106.41	106.41	4803.82	2860.40	29373.45	12258.67
1	527	328.61	133.60	328.61	328.61	90.02	4.32	3280.01	833.60
2	33	105.69	74.60	105.69	105.69	11.36	27.29	73.73	2.19
3	11	85.60	68.94	85.60	85.60	27.69	40.10	2.69	6.06
4	3	43.08	36.82	43.08	43.08	20.07	25.39	0.77	9.11
5	3	68.82	60.84	68.82	68.82	38.59	45.85	6.80	22.56

Table 9 MSE of frequency for **Gender=Woman** observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	527	4032.25	527.00	527.00	527.00	69988.51	3.28029E+33	93855.58	8.47953E+33
1	61	190.27	0.00	0.00	0.00	6756.19	3.79692E+32	9296.63	9.81502E+32
2	6	3.52	6.00	6.00	6.00	544.25	3.73468E+31	772.28	9.65412E+31
3	3	0.16	12.00	12.00	12.00	217.98	1.86734E+31	321.07	4.82706E+31
4	2	3.05	18.00	18.00	18.00	113.23	1.24489E+31	174.67	3.21804E+31
5	2	9.98	32.00	32.00	32.00	85.13	1.24489E+31	139.28	3.21804E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	527	23.32	129.92	23.32	23.32	1052.65	626.79	6436.51	2686.20
1	61	38.04	15.46	38.04	38.04	10.42	0.50	379.66	96.49
2	6	19.22	13.56	19.22	19.22	2.07	4.96	13.41	0.40
3	3	23.35	18.80	23.35	23.35	7.55	10.94	0.73	1.65
4	2	28.72	24.55	28.72	28.72	13.38	16.93	0.51	6.07
5	2	45.88	40.56	45.88	45.88	25.73	30.57	4.53	15.04

Table 10 MSE of frequency for $18 \leq$ Age < 30 observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	527	4032.25	527.00	527.00	527.00	69988.51	3.28029E+33	93855.58	8.47953E+33
1	176	548.96	0.00	0.00	0.00	19493.28	1.09551E+33	26823.05	2.83187E+33
2	11	6.46	11.00	11.00	11.00	997.80	6.84691E+31	1415.85	1.76992E+32
3	4	0.22	16.00	16.00	16.00	290.64	2.48979E+31	428.09	6.43608E+31
4	2	3.05	18.00	18.00	18.00	113.23	1.24489E+31	174.67	3.21804E+31
5	1	4.99	16.00	16.00	16.00	42.56	6.22447E+30	69.64	1.60902E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	527	23.32	129.92	23.32	23.32	1052.65	626.79	6436.51	2686.20
1	176	109.75	44.62	109.75	109.75	30.06	1.44	1095.41	278.39
2	11	35.23	24.87	35.23	35.23	3.79	9.10	24.58	0.73
3	4	31.13	25.07	31.13	31.13	10.07	14.58	0.98	2.20
4	2	28.72	24.55	28.72	28.72	13.38	16.93	0.51	6.07
5	1	22.94	20.28	22.94	22.94	12.86	15.28	2.27	7.52

Table 11 MSE of frequency for $30 \leq$ Age < 40 observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	819	6266.43	819.00	819.00	819.00	108767.73	5.09784E+33	145859.06	1.31779E+34
1	120	374.29	0.00	0.00	0.00	13290.87	7.46936E+32	18288.44	1.93082E+33
2	7	4.11	7.00	7.00	7.00	634.96	4.35713E+31	900.99	1.12631E+32
3	3	0.16	12.00	12.00	12.00	217.98	1.86734E+31	321.07	4.82706E+31
4	1	1.52	9.00	9.00	9.00	56.61	6.22447E+30	87.33	1.60902E+31
5	1	4.99	16.00	16.00	16.00	42.56	6.22447E+30	69.64	1.60902E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	819	36.24	201.90	36.24	36.24	1635.89	974.08	10002.85	4174.57
1	120	74.83	30.42	74.83	74.83	20.50	0.98	746.87	189.81
2	7	22.42	15.82	22.42	22.42	2.41	5.79	15.64	0.46
3	3	23.35	18.80	23.35	23.35	7.55	10.94	0.73	1.65
4	1	14.36	12.27	14.36	14.36	6.69	8.46	0.26	3.04
5	1	22.94	20.28	22.94	22.94	12.86	15.28	2.27	7.52

Table 12 MSE of frequency for $40 \leq$ Age < 50 observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	1115	8531.22	1115.00	1115.00	1115.00	148078.17	6.94028E+33	198574.91	1.79406E+34
1	176	548.96	0.00	0.00	0.00	19493.28	1.09551E+33	26823.05	2.83187E+33
2	13	7.63	13.00	13.00	13.00	1179.22	8.09181E+31	1673.27	2.09173E+32
3	5	0.27	20.00	20.00	20.00	363.30	3.11223E+31	535.11	8.0451E+31
4	2	3.05	18.00	18.00	18.00	113.23	1.24489E+31	174.67	3.21804E+31
5	1	4.99	16.00	16.00	16.00	42.56	6.22447E+30	69.64	1.60902E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	1115	49.33	274.87	49.33	49.33	2227.13	1326.13	13618.04	5683.33
1	176	109.75	44.62	109.75	109.75	30.06	1.44	1095.41	278.39
2	13	41.64	29.39	41.64	41.64	4.47	10.75	29.05	0.86
3	5	38.91	31.34	38.91	38.91	12.59	18.23	1.22	2.76
4	2	28.72	24.55	28.72	28.72	13.38	16.93	0.51	6.07
5	1	22.94	20.28	22.94	22.94	12.86	15.28	2.27	7.52

Table 13 MSE of frequency for Age ≥ 50 observation under count distributions given by Table 2

claim	frequency	NB	$\{0\}-NBM_1$	$\{1\}-NBM_1$	$\{2\}-NBM_1$	$\{-\}-NBM_2$	$\{12\}-NBM_2$	$\{10\}-NBM_2$	$\{20\}-NBM_2$
0	471	3603.77	471.00	471.00	471.00	62551.40	2.93172E+33	83882.31	7.57848E+33
1	116	361.82	0.00	0.00	0.00	12847.84	7.22038E+32	17678.83	1.86646E+33
2	8	4.70	8.00	8.00	8.00	725.67	4.97957E+31	1029.71	1.28722E+32
3	2	0.11	8.00	8.00	8.00	145.32	1.24489E+31	214.05	3.21804E+31
4	0	0.00	0.00	0.00	0.00	0.00	0	0.00	0
5	2	9.98	32.00	32.00	32.00	85.13	1.24489E+31	139.28	3.21804E+31
claim	frequency	POI	$\{0\}-POIM_1$	$\{1\}-POIM_1$	$\{2\}-POIM_1$	$\{-\}-POIM_2$	$\{01\}-POIM_2$	$\{02\}-POIM_2$	$\{12\}-POIM_2$
0	471	20.84	116.11	20.84	20.84	940.79	560.19	5752.55	2400.76
1	116	72.33	29.41	72.33	72.33	19.82	0.95	721.98	183.49
2	8	25.62	18.08	25.62	25.62	2.75	6.62	17.87	0.53
3	2	15.56	12.53	15.56	15.56	5.04	7.29	0.49	1.10
4	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	2	45.88	40.56	45.88	45.88	25.73	30.57	4.53	15.04

Table 14 The GD , the AIC and the SBIC

model	DF	GD	AIC	SBC	DF	GD	AIC	SBC
NB	4	3990.50	3998.50	4023.24	2	4003.72	4007.72	4020.09
$\{0\}-NBM_1$	5	3990.50	4000.50	4031.42	3	4003.72	4009.72	4028.27
$\{1\}-NBM_1$	5	3948.27	3958.27	3989.19	3	3950.06	3956.06	3974.62
$\{2\}-NBM_1$	5	3990.50	4000.50	4031.42	3	4003.72	4009.72	4028.27
$\{-\}-NBM_2$	8	3966.24	3984.24	4039.90	5	3971.31	3981.31	4012.23
$\{10\}-NBM_2$	10	3948.46	3970.46	4038.48	7	3950.03	3964.03	4007.32
POI	3	4026.71	4032.71	4051.26	1	4042.62	4044.62	4050.81
$\{0\}-POIM_1$	4	4016.71	4024.71	4049.45	2	4029.72	4033.72	4046.09
$\{1\}-POIM_1$	4	4026.71	4034.71	4059.44	2	4042.62	4046.62	4058.99
$\{2\}-POIM_1$	4	4026.71	4034.71	4059.44	2	4042.62	4046.62	4058.99
$\{-\}-POIM_2$	7	3993.23	4007.23	4050.52	3	4007.01	4013.01	4031.57
$\{01\}-POIM_2$	9	3934.65	3952.65	4008.31	5	4029.72	4039.72	4070.64

The AIC, SBIC, and GD are three measures to select an appropriate model among a set of candidate models. A preferred model is one that has the minimum AIC, SBIC, or GD, as shown in Table14, the AIC, SBIC, or GD measure indicates that $\{01\}$-inflated negative binomial mixture model, say $\{01\} - NBM_2$, provides better fit compared to other models.

4.2. The rate premium for the Rate-making system

In order to demonstrate the practical application of the present findings, the rate premium was calculated for the set of well-fitted distributions/regression models. Since we were interested in the differences between rate premiums of various classes, we set the rate premium for a new policyholder equal to 1 unit, at t = 0. In addition, four different categories were considered as shown in Table 15.

Table 15 Categories which considered to evaluate rate premiums under well fitted models

Category	Description
A_1	Whenever, chosen policyholder is old woman at age 56.
A_2	Whenever, chosen policyholder is young man at age 27.
A_3	Whenever, chosen policyholder is man at age 43.
A_4	Whenever, chosen policyholder is young woman at age 21.

Furthermore, the well-fitted models were implemented to calculate the rate premium for four categories A_1,A_2,A_3 and A_4 (Table 15). In this regard, an approach is suggested based on emphasizing the exact number of reported claims for each year in the history of the policyholder. The best fit model for data among the set of candidate models is $\{01\} - NBM_2$, so the rate premium under it for four categories, given in Table 15, is represented in Table 16.

Table 16 The rate premium for three categories A_1, A_2, A_3 and A_4

model year	Number of cumulated claims up yo this year(k)	A_1	A_2	A_3	A_4
t=1	k=0	1.00	1.00	1.00	1.00
	k=1	1.00	1.00	1.00	1.00
	k=2	1.42	1.40	1.42	1.40
	k=3	1.65	1.63	1.65	1.63
	k=4	1.88	1.86	1.88	1.86
	k=5	2.11	2.09	2.10	2.08

For the given policyholders belong to categories in Table 15, based on Table 16 result, the rate premium under $\{01\} - NBM_2$ model is 0.99 units whenever such policyholders do not report any claim in the first year and it is one unit when they report one claim. Indeed there is no difference between the 4 categories. In the situation that such policyholders report more than one claims, the young man at age 27 and a young woman at age 21 should pay less rate premium unit than an old woman at age 56 and a man at age 43.

Acknowledgements

Authors would like to thank anonymous reviewers for their constructive comments.

References

Alshkaki RSA. On the zero-one inflated poisson distribution. J Stat Distrib Appl. 2016; 2(4): 42-48.

Aryuyuen S, Bodhisuwan W. The negative binomial-generalized exponential (NB-GE) distribution. Appl Math Sci. 2013; 7(22): 1093-105.

Bermúdez L, Karlis D. Bayesian multivariate poisson models for insurance ratemaking. Insur Math Econ. 2011; 48(2), 226-36.

Boucher JP, Inoussa R. A posteriori ratemaking with panel data. Austin Bull. 2014; 44(3): 587-612.

Denuit M, Dhaene J. Bonus-Malus scales using exponential loss functions. Bl atter der DGVFM. 2001; 25(1): 13-27.

Denuit M, Maréchal X, Pitrebois S, Walhin JF. Actuarial modelling of claim counts: Risk classification, credibility and bonus-malus systems. New York: John Wiley & Sons; 2007.

Dionne G, Vanasse C. A generalization of automobile insurance rating models: the negative binomial distribution with a regression component. Austin Bull. 1989; 19(2): 199-212.

Dionne G, Vanasse C. Automobile insurance ratemaking in the presence of asymmetrical information. J Appl Econ. 1992; 7(2): 149-165.

Greene WH. Accounting for excess zeros and sample selection in poisson and negative binomial regression models. Technical Report [No. EC-94-10]. Department of Economics, Stern School of Business: New York University.

Gupta PL, Gupta RC, Tripathi RC. Analysis of zero-adjusted count data. Comput Stat Data Anal. 1996; 23(2): 207-218.

Hall DB. Zero-inflated poisson and binomial regression with random effects: a case study. J Biom. 2000; 56(4): 1030-1039.

Heras A, Moreno I, Vilar-Zanón JL. An application of two-stage quantile regression to insurance ratemaking. Scand Actuar J. 2018; 9: 1-17.

Johnson NL, Kemp AW Kotz S. Univariate Discrete Distributions (Vol. 444). New York: John Wiley & Sons; 2007.

Joshi R. A Generalized Inflated Geometric Distribution. PhD [dissertation]. Marshal university; 2015.

Khalid M, Aslam M, Sindhu TN. Bayesian analysis of 3-components Kumaraswamy mixture model: quadrature method vs. Importance sampling. Alexandria Eng J. 2020; 59(4): 2753-2763.

Lange JT. Application of a mathematical concept of risk to property-liability insurance ratemaking. J Risk Insur. 1969; 38(2): 383-391.

Lee GY, Frees EW. General Insurance Deductible Ratemaking with Applications to the Local Government Property Insurance Fund. Available from: https://www.soa.org/Files/Research/Projects/research-2016-gi-deductible-ratemaking.pdf.

Lemaire, J. (1995). Bonus-malus systems in automobile insurance. New York: Springer science & business media; 1995.

Lim HK, Li WK, and Philip LH. Zero-inflated poisson regression mixture model. Comput Stat Data Anal. 2014; 71: 151-158.

Liu W, Tang Y, and Xu A. Zero-and-one-inflated poisson regression model. Stat Pap (Berl). 2021; 62(2):, 915-934.

McLachlan G, and Krishnan N. The EM algorithm and Extensions. New York: Wiley; 1997.

Melkersson M, and Rooth D. Modeling female fertility using inflated count data models. J Popul Econ. 2000; 13: 189-203.

Nishii R, and Tanaka S. Modeling and inference of forest coverage ratio using zero-one inflated distributions with spatial dependence. Environ Ecol Stat. 2013; 20(2): 315-336.

Payandeh Najafabadi, AT. A new approach to the credibility formula. Insur Math Econ. 2010; 46(2): 334-338.

Payandeh Najafabadi, AT, Atatalab F, and Omidi Najafabadi M. Credibility premium for rate-making systems. Commun Stat Theory Methods. 2017; 46(1): 415-426.

Payandeh Najafabadi AT, and Sakizadeh M. Designing an Optimal Bonus–Malus System Using the Number of Reported Claims, Steady-State Distribution, and Mixture Claim Size Distribution. Int J Ind Syst Eng. 2019; 32(3): 304-331.

Pinquet, J. Allowance for cost of claims in bonus-malus systems. Austin Bull. 1997; 27(1): 33-57

Pinquet, J. (1998). Designing optimal bonus-malus systems from different types of claims. Austin Bull. 1998; 28(2): 205-220.

Razie F, Bahrami Samani E, and Ganjali M. A latent variable model for analyzing mixed longitudinal $\{k, l\}$-inflated count and ordinal responses. J Appl Stat. 2016; 43(12), 2203-2224.

Rigby RA, and Stasinopoulos DM. The GAMLSS project: a flexible approach to statistical modelling. In New trends in statistical modelling: Proceedings of the 16th international workshop on statistical modelling. University of Southern Denmark; 2001, p. 345-355.

Simar, L. Maximum likelihood estimation of a compound poisson process. Ann Stat. 1976; 4(6): 1200-1209.

Simon, LJ. Fitting negative binomial distributions by the method of maximum likelihood. Proc. Casual. Actuar. Soc. 1961; 48: 45-53.

Sindhu TN, and Aslam M. Preference of prior for Bayesian analysis of the mixed Burr type X distribution under type I censored samples. Pakistan J Stat Operat Res. 2014; 10(1): 17-39.

Sindhu TN, Aslam M, and Hussain Z. A simulation study of parameters for the censored shifted Gompertz mixture distribution: A Bayesian approach. J stat Manag Syst. 2016; 19(3): 423-450.

Sindhu TN, Hussain Z, and Aslam M. On the Bayesian analysis of censored mixture of two Topp-Leone distribution. Sri Lankan J Appl Stat. 2019; 19(1): 13-20.

Sindhu TN, Riaz M, Aslam M, and Ahmed Z. Bayes estimation of Gumbel mixture models with industrial applications. Trans Ins Meas Control. 2016; 38(2): 201-214.

Sindhu TN, Hussain Z, and Aslam M. Parameter and reliability estimation of inverted Maxwell mixture model. J stat Manag Syst. 2019; 22(3): 459-493.

Sindhu TN, Khan HM, Hussain Z, and Al-Zahrani, B. Bayesian inference from the mixture of half-normal distributions under censoring. J Natl Sci Found. 2018; 46(4): 587-600.

Tzougas G, and Frangos N. The design of an optimal bonus-malus system based on the Sichel distribution. Mod Probl Insur Math. 2014; 1: 239-260.

Tzougas G, Vrontos S, and Frangos N. Optimal bonus-malus systems using finite mixture models.Austin Bull. 2014; 44(2): 417-444.

Walhin JF, and Paris J. Using Mixed poisson Processes in Connection with Bonus-Malus Systems. Austin Bull. 1999; 29(1): 81-99.

Wieczorek J, Nugent C, and Hawala S. A Bayesian zero-one inflated beta model for small area shrinkage estimation. In Proceedings of the 2012 Joint Statistical Meetings, American Statistical Association, Alexandria, VA; 2012.

Appendix: An EM algorithm approach to estimate unknown parameters

In Proposition 2 and Proposition 3, we assumed that all parameters, except λ_z, are given. But in many practical situations, some (or all) parameters are unknown, therefore, one has to employ an Expectation-Maximization algorithm to derive a maximum-likelihood estimate for such parameters.

This appendix, based on Rigby and Stasinopoulos (2009) method, presents such estimation under the $\{ij\}$-inflated poisson and negative binomial mixture models.

The Expectation-Maximization, say EM, the algorithm is considered as a method for finding the maximum-likelihood estimates for unknown parameters when the related data are incomplete, include missing data points, or have unobserved (hidden) latent variables. Furthermore, the EM algorithm can be implemented to estimate latent variables like those coming from mixture distributions. In addition, it is regarded as an iterative way to approximate the maximum likelihood function. However, the maximum likelihood estimation fails to work well, especially for incomplete data sets although it can find the best fit model for a set of data. The EM algorithm works by choosing random values for the missing data points and using those guesses to estimate a second set of data. The new values are used to create a better prediction for the first set, and the process continues until the algorithm converges to a fixed point (McLachlan and Krishnan, 1997). Now, suppose that the number of components, m, is given, and $\nu_z = (\nu_{z1}, \nu_{z2}, \cdots, \nu_{zm})$ stands for the latent vector of component indicator variables, where $\nu_{zk} = 1$ for $z = 1, \cdots, n$ and $k = 1, \cdots, m$, whenever observation z comes from the k^{th} component. Otherwise, $\nu_{zk} = 0$. Therefore, we assume that each observation is arrived from one of the m components, although the related component is unobservable and should be considered as the missing data.

$$\begin{aligned}
&\iota_c(\boldsymbol{\theta}|y_z,\nu_z,\mathbf{X}_z) = \\
&\ln\prod_{z=1}^{n}\left(\left(\left(p_1+(1-p_1)(1-p_2)I_{\{i\}}(y_z)\right)\right)^{\nu_{z1}}\left(\left(p_2(1-p_1)+(1-p_1)(1-p_2)I_{\{j\}}(y_z)\right)\right)^{\nu_{z2}}III^{**}\right) \\
=\;& \sum_{z=1}^{n}\nu_{z1}I_{\{i\}}(y_z)\ln\left(p_1+(1-p_1)(1-p_2)\right)+\sum_{z=1}^{n}\nu_{z2}I_{\{j\}}(y_z)\ln\left(p_2(1-p_1)+(1-p_1)(1-p_2)\right) \\
&+\sum_{z=1}^{n}\sum_{k=1}^{m}I_{A*}(y_z)\nu_{zk}\ln\left((1-p_1)(1-p_2)\right)\varphi_k A(k,z)\right),
\end{aligned}$$

where $A^*=\{0,1,\cdots\}\setminus\{i,j\}$, $I_{\cdot}(\cdot)$ stands for indicator function, $III^{**}=\prod_{k=1}^{m}(I_{A*}(y_z)(1-p_1)(1-p_2))\varphi_k A(k,z))^{\nu_{zk}}$, $A(k,z)=\int_0^\infty \frac{u_z^{y_z}e^{y_z(\mathbf{X}_z\mathbf{B}_z)}a(y_z)}{g(u_z e^{\mathbf{X}_z\mathbf{B}_z})}f_{U_z}(u_z|\alpha_k)du_z$, and the vector of parameter $\boldsymbol{\theta}=(\mathbf{B},\boldsymbol{\alpha},\boldsymbol{\varphi},p_1,p_2)$ represents all unknown parameters, $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,...,\alpha_m)$, $\mathbf{B}=(\mathbf{B}_1,...,\mathbf{B}_n)$ and $\boldsymbol{\varphi}=(\varphi_1,\varphi_2,...,\varphi_m)$

E-step:

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{(r)} &= (\widehat{\mathbf{B}}^{(r)},\widehat{\boldsymbol{\alpha}}^{(r)},\widehat{\boldsymbol{\varphi}}^{(r)},\widehat{p_1}^{(r)},\widehat{p_2}^{(r)}) \\
\widehat{\nu}_{zk}^{(r+1)} &= E\left[\nu_{zk}|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right]=1\times P\left(\nu_{zk}=1|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)+0\times P\left(\nu_{zk}=0|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right) \\
&= \frac{f\left(y_z|\nu_{zn}=1,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)P\left(\nu_{zn}=1|\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)}{f\left(y_z|\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)} \\
&= \frac{\sum_{k=1}^{m}I_{A*}(y_z)\widehat{p_3}^{(r)}\widehat{\varphi}_k^{(r)}\int_0^\infty\frac{u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r)})}a(y_z)}{g(u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r)})})}f_{U_z}^k(u_z|\widehat{\alpha}_k^{(r)})du_z}{\widehat{p_1}^{(r)}I_{\{i\}}(y_z)+\widehat{p_2}^{(r)}I_{\{j\}}(y_z)+\sum_{k=1}^{m}I_{A*}(y_z)\widehat{p_3}^{(r)}\widehat{\varphi}_k^{(r)}\int_0^\infty\frac{u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r)})}a(y_z)}{g(u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r)})})}f_{U_z}^k(u_z|\widehat{\alpha}_k^{(r)})du_z}
\end{aligned}$$

M-step:

$$\begin{aligned}
\mathbf{Q} &= E\left[\iota_c|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right] \\
&= \sum_{z=1}^{n}\widehat{\nu}_{z1}^{(r+1)}I_{\{i\}}(y_z)\ln\left(\widehat{p_1}^{(r)}\right)+\sum_{z=1}^{n}\widehat{\nu}_{z2}^{(r+1)}I_{\{j\}}(y_z)\ln\left(\widehat{p_2}^{(r)}\right) \\
&\quad+\sum_{z=1}^{n}\sum_{k=1}^{m}\widehat{\nu}_{zk}^{(r+1)}I_{A*}(y_z)\ln\left(\widehat{p_3}^{(r)}\widehat{\varphi}_k^{(r)}\int_0^\infty\frac{u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r)})}a(y_z)}{g(u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r)})})}f_{U_z}^k(u_z|\widehat{\alpha}_k^{(r)})du_z\right)
\end{aligned}$$

where $\widehat{\boldsymbol{\theta}}^{(r+1)}=(\widehat{\mathbf{B}}^{(r+1)},\widehat{\boldsymbol{\alpha}}^{(r+1)},\widehat{\boldsymbol{\varphi}}^{(r+1)},\widehat{p_1}^{(r+1)},\widehat{p_2}^{(r+1)})$ have been arrived by solving the following equation

$$\frac{\partial\mathbf{Q}}{\partial\alpha_k}=0;\qquad\frac{\partial\mathbf{Q}}{\partial\mathbf{B}_z}=0;\qquad\frac{\partial\mathbf{Q}}{\partial\varphi_k}=0;\qquad\frac{\partial\mathbf{Q}}{\partial p_1}=0;\qquad\frac{\partial\mathbf{Q}}{\partial p_2}=0$$

the above equation can not solve explicitly, therefore, one has to employ the following recursive approach

$$\widehat{\alpha}_k^{(r+1)}=\widehat{\alpha}_k^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial\alpha_k^2}\right)\frac{-\partial\mathbf{Q}}{\partial\alpha_k^2};\qquad\widehat{\mathbf{B}}_z^{(r+1)}=\widehat{\mathbf{B}}_z^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial\mathbf{B}_z^2}\right)\frac{-\partial\mathbf{Q}}{\partial\mathbf{B}_z^2};$$

$$\widehat{\varphi}_k^{(r+1)}=\widehat{\varphi}_k^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial\varphi_k^2}\right)\frac{-\partial\mathbf{Q}}{\partial\varphi_k^2};\qquad\widehat{p_1}^{(r+1)}=\widehat{p_1}^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial p_1^2}\right)\frac{-\partial\mathbf{Q}}{\partial p_1^2};$$

$$\widehat{p_2}^{(r+1)}=\widehat{p_2}^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial p_2^2}\right)\frac{-\partial\mathbf{Q}}{\partial p_2^2}.$$

Then

$$\begin{aligned}\iota_c^{(r+1)} &= \sum_{z=1}^{n}\widehat{\nu}_{z1}^{(r+1)}I_{\{i\}}(y_z)\ln\left(\widehat{p_1}^{(r+1)}\right)+\sum_{z=1}^{n}\widehat{\nu}_{z2}^{(r+1)}I_{\{j\}}(y_z)\ln\left(\widehat{p_2}^{(r+1)}\right)\\ &+\sum_{z=1}^{n}\sum_{k=1}^{m}\widehat{\nu}_{zk}^{(r+1)}I_{A*}(y_z)\ln\left(p_3{}^{(r+1)})\int_0^\infty\frac{u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r+1)})}a(y_z)}{g(u_z^{y_z}e^{y_z(\mathbf{X}_z\widehat{\mathbf{B}}_z^{(r+1)})})}f_{U_z}^k(u_z|\widehat{\alpha}_k^{(r+1)})du_z\right)\end{aligned}$$

To see how one may employ the above EM algorithm in practice, in the rest of this appendix, we estimate unknown parameters under two the $\{ij\}$-inflated poisson and negative binomial mixture models.

Estimation under the $\{ij\}$-inflated poisson mixture model

$$\begin{aligned}P(Y=y|\lambda) &= \left(p_1+(1-p_1)(1-p_2)I_{\{i\}}(y)\right)+\left(p_2(1-p_1)+(1-p_1)(1-p_2)I_{\{j\}}(y)\right)\\ &+(1-p_1)(1-p_2))\,\frac{e^{-\lambda}\lambda^y}{y!}\\ P(Y_{zl}=y_{zl}|\lambda) &= \int_0^\infty P(Y_{zl}=y_{zl}|\boldsymbol{\theta},u_z)f_{U_z}(u_z)du_z\\ &= \left(p_1+(1-p_1)(1-p_2)I_{\{i\}}(y_{zl})\right)+\left(p_2(1-p_1)+(1-p_1)(1-p_2)I_{\{j\}}(y_{zl})\right)\\ &+\sum_{k=1}^{m}I_{A*}(y_{zl})\,(1-p_1)(1-p_2))\,\varphi_k\frac{\alpha_k^{\alpha_k}d_{zl}^{y_{zl}}}{\Gamma(\alpha_k)y_{zl}!}\int_0^\infty e^{-u_z(d_{zl}+\alpha_k)}u_z^{(\alpha_k+y_{zl}-1)}du_z\\ &= \left(p_1+(1-p_1)(1-p_2)I_{\{i\}}(y_{zl})\right)+\left(p_2(1-p_1)+(1-p_1)(1-p_2)I_{\{j\}}(y_{zl})\right)\\ &+\sum_{k=1}^{m}I_{A*}(y_{zl})\,(1-p_1)(1-p_2))\,\varphi_k\frac{\alpha_k^{\alpha_k}d_{zl}^{y_{zl}}}{\Gamma(\alpha_k)y_{zl}!}\frac{\Gamma(\alpha_k+y_{zl})}{(d_{zl}+\alpha_k)^{\alpha_k+y_{zl}}}\\ &= \left(p_1+(1-p_1)(1-p_2)I_{\{i\}}(y_{zl})\right)+\left(p_2(1-p_1)+(1-p_1)(1-p_2)I_{\{j\}}(y_{zl})\right)\\ &+\sum_{k=1}^{m}I_{A*}(y_{zl})\,(1-p_1)(1-p_2))\,\varphi_k\frac{(\alpha_k+y_{zl}-1)!}{y_{zl}!(\alpha_k-1)!}\left(\frac{\alpha_k}{d_{zl}+\alpha_k}\right)^{\alpha_k}\left(\frac{d_{zl}}{d_{zl}+\alpha_k}\right)^{y_{zl}},\end{aligned}$$

where $A^*=\{0,1,\cdots\}\setminus\{i,j\}$, $I_\cdot(\cdot)$ stands for indicator function and $d_{zl}=\mathbf{X}_z\mathbf{B}_{zl}$.
Now using the Multinomial distribution for the unobservable vector ν_i, the complete data log-likelihood function, for, the $\{ij\}-IPSM_m$ regression, can be written as the following, see Rigby and Stasinopoulos (2009) for more information. from results from Proposition 1, the EM algorithm employs the following steps.

$$\begin{aligned}\iota_c(\boldsymbol{\theta}|y_z,\nu_z,\mathbf{X}_z) &= \ln\Pi_{z=1}^{n}\left(p_1I_{\{i\}}(y_z)\right)^{\nu_{z1}}\times\left(p_2I_{\{j\}}(y_z)\right)^{\nu_{z2}}\\ &\times\Pi_{k=1}^{m}\left(I_{A^*}(y_z)p_3\varphi_k\frac{(\alpha_k+y_z-1)!}{y_z!(\alpha_k-1)!}\left(\frac{\alpha_k}{d_z+\alpha_k}\right)^{\alpha_k}\left(\frac{d_z}{d_z+\alpha_k}\right)^{y_z}\right)^{\nu_{zk}}\\ &= \sum_{z=1}^{n}\nu_{z1}I_{\{i\}}(y_z)\ln(p_1)+\sum_{z=1}^{n}\nu_{z2}I_{\{j\}}(y_z)\ln(p_2)\\ &+\sum_{z=1}^{n}\sum_{k=1}^{m}\nu_{zk}I_{A^*}(y_z)\ln\left(p_3\varphi_k\frac{(\alpha_k+y_z-1)!}{y_z!(\alpha_k-1)!}\left(\frac{\alpha_k}{d_z+\alpha_k}\right)^{\alpha_k}\left(\frac{d_z}{d_z+\alpha_k}\right)^{y_z}\right)\end{aligned}$$

where $A^*=\{0,1,\cdots\}\setminus\{i,j\}$, $I_\cdot(\cdot)$ stands for indicator function and $\boldsymbol{\theta}=(\mathbf{B},\boldsymbol{\alpha},\boldsymbol{\varphi},p_1,p_2)$ and $d_z=\mathbf{X}_z\mathbf{B}_z$

E-step:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{(r)} &= (\widehat{\boldsymbol{\alpha}}^{(r)},\widehat{\mathbf{B}}^{(r)},\widehat{\boldsymbol{\varphi}}^{(r)},\widehat{p_1}^{(r)},\widehat{p_2}^{(r)})\\ \widehat{\nu}_{zk}^{(r+1)}\end{aligned}$$

$$
\begin{aligned}
&= E\left[\nu_{zk}|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right] = 1\times P\left(\nu_{zk}=1|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right) + 0\times P\left(\nu_{zk}=0|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)\\
&= \frac{f\left(y_z|\nu_{zn}=1,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right) P\left(\nu_{zn}=1|\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)}{f\left(y_z|\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right)}\\
&= \frac{I_{A^*}(y_z)\widehat{p_3}^{(r)}\widehat{\varphi}_k^{(r)}\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_z}}{\widehat{p_1}^{(r)}I_{\{i\}}(y_z)+\widehat{p_2}^{(r)}I_{\{j\}}(y_z)+\sum_{k=1}^{m}I_{A^*}(y_z)\widehat{p_3}^{(r)}\widehat{\varphi}_k^{(r)}\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_z}}
\end{aligned}
$$

M-step:

$$
\begin{aligned}
\mathbf{Q} &= E\left[\iota_c|y_z,\mathbf{X}_z,\widehat{\boldsymbol{\theta}}^{(r)}\right]\\
&= \sum_{z=1}^{n}\widehat{\nu}_{z1}^{(r+1)}I_{\{i\}}(y_z)\ln\left(\widehat{p_1}^{(r)}\right)+\sum_{z=1}^{n}\widehat{\nu}_{z2}^{(r+1)}I_{\{j\}}(y_z)\ln\left(\widehat{p_2}^{(r)}\right)\\
&\quad+\sum_{z=1}^{n}\sum_{k=1}^{m}\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left(\widehat{p_3}^{(r)}\varphi_k\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_z}\right)
\end{aligned}
$$

where $\widehat{\boldsymbol{\theta}}^{(r+1)} = (\widehat{\mathbf{B}}^{(r+1)}, \widehat{\boldsymbol{\alpha}}^{(r+1)}, \widehat{\boldsymbol{\varphi}}^{(r+1)}, \widehat{p_1}^{(r+1)}, \widehat{p_2}^{(r+1)})$ have been arrived by solving the following equation.

$$
\begin{aligned}
\frac{\partial\mathbf{Q}}{\partial\alpha_k} &= \sum_{z=1}^{n}\sum_{k=1}^{m}\frac{\partial}{\partial\alpha_k}\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left(\widehat{p_3}^{(r)}\varphi_k\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_z}\right)=0\\
\frac{\partial\mathbf{Q}}{\partial\mathbf{B}_z} &= \sum_{z=1}^{n}\sum_{k=1}^{m}\frac{\partial}{\partial\beta_k}\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left(\widehat{p_3}^{(r)}\varphi_k\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_z}\right)=0\\
\frac{\partial\mathbf{Q}}{\partial\varphi_k} &= \sum_{z=1}^{n}\sum_{k=1}^{m}\frac{\partial}{\partial\varphi_k}\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left(\widehat{p_3}^{(r)}\varphi_k\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_z}\right)=0\\
\frac{\partial\mathbf{Q}}{\partial p_1} &= \sum_{z=1}^{n}\sum_{k=1}^{m}\frac{\partial}{\partial p_1}\widehat{\nu}_{z1}^{(r+1)}I_{\{i\}}(y_z)\ln\left(\widehat{p_1}^{(r)}\right)\\
&\quad+\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left((1-\widehat{p_1}^{(r)}-\widehat{p_2}^{(r)})\varphi_k\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_i}\right)=0\\
\frac{\partial\mathbf{Q}}{\partial p_2} &= \sum_{z=1}^{n}\sum_{k=1}^{m}\frac{\partial}{\partial p_2}\widehat{\nu}_{z2}^{(r+1)}I_{\{j\}}(y_z)\ln\left(\widehat{p_2}^{(r)}\right)\\
&\quad+\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left((1-\widehat{p_1}^{(r)}-\widehat{p_2}^{(r)})\varphi_k\frac{(\widehat{\alpha}_k^{(r)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{\widehat{\alpha}_k^{(r)}}\left(\frac{\widehat{d}_z^{(r)}}{\widehat{d}_z^{(r)}+\widehat{\alpha}_k^{(r)}}\right)^{y_i}\right)=0
\end{aligned}
$$

The above equation can not solve explicitly, so

$$
\widehat{\alpha}_k^{(r+1)}=\widehat{\alpha}_k^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial\alpha_k^2}\right)\frac{-\partial\mathbf{Q}}{\partial\alpha_k^2};\qquad \widehat{\mathbf{B}}_z^{(r+1)}=\widehat{\mathbf{B}}_z^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial\mathbf{B}_z^2}\right)\frac{-\partial\mathbf{Q}}{\partial\mathbf{B}_z^2};
$$

$$
\widehat{\varphi}_k^{(r+1)}=\widehat{\varphi}_k^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial\varphi_k^2}\right)\frac{-\partial\mathbf{Q}}{\partial\varphi_k^2};\qquad \widehat{p_1}^{(r+1)}=\widehat{p_1}^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial p_1^2}\right)\frac{-\partial\mathbf{Q}}{\partial p_1^2};
$$

$$
\widehat{p_2}^{(r+1)}=\widehat{p_2}^{(r)}+E^{-1}\left(\frac{-\partial^2\mathbf{Q}}{\partial p_2^2}\right)\frac{-\partial\mathbf{Q}}{\partial p_2^2}.
$$

Then

$$\begin{aligned}
\iota_c^{(r+1)} &= \sum_{z=1}^{n}\widehat{\nu}_{z1}^{(r+1)}I_{\{i\}}(y_z)\ln\left(\widehat{p_1}^{(r)}\right)+\sum_{z=1}^{n}\widehat{\nu}_{z2}^{(r+1)}I_{\{j\}}(y_z)\ln\left(\widehat{p_2}^{(r)}\right)\\
&+\sum_{z=1}^{n}\sum_{k=1}^{m}\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left(\widehat{p_3}^{(r+1)}\varphi_k\frac{(\widehat{\alpha}_k^{(r+1)}+y_z-1)!}{y_z!(\widehat{\alpha}_k^{(r+1)}-1)!}\left(\frac{\widehat{\alpha}_k^{(r+1)}}{\widehat{d}_z^{(r+1)}+\widehat{\alpha}_k^{(r+1)}}\right)^{\widehat{\alpha}_k^{(r+1)}}\right)\\
&+\sum_{z=1}^{n}\sum_{k=1}^{m}\widehat{\nu}_{zk}^{(r+1)}I_{A^*}(y_z)\ln\left(\frac{\widehat{d}_z^{(r+1)}}{\widehat{d}_z^{(r+1)}+\widehat{\alpha}_k^{(r+1)}}\right)^{y_z}.
\end{aligned}$$

Estimation under the $\{ij\}$-inflated negative binomial mixture model

$$\begin{aligned}
P(Y=y|p) &= p_1I_{\{i\}}(y)+p_2I_{\{j\}}(y)+p_3\frac{(N+y-1)!}{y!(N-1)!}(1-p)^yp^N \qquad \lambda=1-p\\
P(Y_{zl}=y_{zl}|\boldsymbol{\theta}) &= \int_0^1 P(Y_{zl}=y_{zl}|\boldsymbol{\theta},u_z)f_{U_z}(u_z)du_z\\
&= p_1I_{\{i\}}(y_{zl})+p_2I_{\{j\}}(y_{zl})+I_{A^*}(y_z)\sum_{k=1}^{m}p_3\varphi_k\frac{(N+y_{zl}-1)!\Gamma(\alpha_k+\alpha_k)}{y_{zl}!(N-1)!\Gamma(\alpha_k)\Gamma(\alpha_k)}\\
&\times\int_0^1 u_z^{y_{zl}+\alpha_k-1}(1-u_ze^{\mathbf{X}_z\mathbf{B}_z})^N(1-u_z)^{\alpha_k-1}(e^{\mathbf{X}_z\mathbf{B}_z})^Ndu_z\\
&= p_1I_{\{i\}}(y_{zl})+p_2I_{\{j\}}(y_{zl})+I_{A^*}(y_z)\sum_{k=1}^{m}p_3\varphi_k\frac{(N+y_{zl}-1)!\Gamma(\alpha_k+\alpha_k)}{y_{zl}!(N-1)!\Gamma(\alpha_k)\Gamma(\alpha_k)}(e^{\mathbf{X}_z\mathbf{B}_z})^N\\
&\times\sum_{d=1}^{N}N(-1)^{N-d}e^{(N-d)\mathbf{X}_z\mathbf{B}_z}\int_0^1 u_z^{y_{zl}+\alpha_k-1}u_z^{N-d}(1-u_z)^{\alpha_k-1}du_z\\
&= p_1I_{\{i\}}(y_{zl})+p_2I_{\{j\}}(y_{zl})+I_{A^*}(y_z)\sum_{k=1}^{m}\sum_{d=1}^{N}N(-1)^Ne^{(2N-d)\mathbf{X}_z\mathbf{B}_z}p_3\varphi_k\\
&\times\frac{\Gamma(N+y_{zl})\Gamma(2\alpha_k)\Gamma(\alpha_k+N-d+y_{zl})}{\Gamma(y_{zl}-1)\Gamma(N)\Gamma(\alpha_k)\Gamma(2\alpha_k+N-d+y_{zl})}
\end{aligned}$$

$$\begin{aligned}
\iota_c(\underline{\theta}|y_z,\nu_z,\mathbf{X}_z) &= \ln\Pi_{z=1}^{n}(p_1I_{\{i\}}(y_{zl})+p_2I_{\{j\}}(y_{zl})+I_{A^*}(y_z)\sum_{k=1}^{m}\sum_{d=1}^{N}N(-1)^Ne^{(2N-d)\mathbf{X}_z\mathbf{B}_z}p_3\varphi_k\\
&\times\frac{\Gamma(N+y_{zl})\Gamma(2\alpha_k)\Gamma(\alpha_k+N-d+y_{zl})}{\Gamma(y_{zl}-1)\Gamma(N)\Gamma(\alpha_k)\Gamma(2\alpha_k+N-d+y_{zl})})\\
&= \sum_{z=1}^{n}\nu_{z1}I_{\{i\}}(y_z)\ln\ p_1+\sum_{z=1}^{n}\nu_{z2}I_{\{j\}}(y_z)\ln\ p_2\\
&+\sum_{z=1}^{n}\nu_{zk}I_{A^*}(y_z)\ln\sum_{k=1}^{m}\sum_{d=1}^{N}N(-1)^Ne^{(2N-d)\mathbf{X}_z\mathbf{B}_z}p_3\varphi_k\\
&\times\frac{\Gamma(N+y_{zl})\Gamma(2\alpha_k)\Gamma(\alpha_k+N-d+y_{zl})}{\Gamma(y_{zl}-1)\Gamma(N)\Gamma(\alpha_k)\Gamma(2\alpha_k+N-d+y_{zl})})
\end{aligned}$$

where $A^*=\{0,1,\cdots\}\setminus\{i,j\}$, $\boldsymbol{\theta}=(\mathbf{B},\boldsymbol{\alpha},\boldsymbol{\varphi},p_1,p_2)$ and $I_\cdot(\cdot)$ stands for indicator function.

E-step:

$$\begin{aligned}
\widehat{\theta}^{(r)} &= (\widehat{\alpha}^{(r)}, \widehat{\mathbf{B}}^{(r)}, \widehat{\varphi}^{(r)}, \widehat{p_1}^{(r)}, \widehat{p_2}^{(r)}) \\
\widehat{\nu}_{zk}^{(r+1)} &= E\left[\nu_{zk}|y_z, \mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right] = 1 \times P\left(\nu_{zk} = 1|y_z, \mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right) + 0 \times P\left(\nu_{zk} = 0|y_z, \mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right) \\
&= \frac{f\left(y_z|\nu_{zn} = 1, \mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right) P\left(\nu_{zn} = 1|\mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right)}{f\left(y_z|\mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right)}
\end{aligned}$$

M-step:

$$\begin{aligned}
\mathbf{Q} &= E\left[\iota_c|y_z, \mathbf{X}_z, \widehat{\boldsymbol{\theta}}^{(r)}\right] \\
&= \sum_{z=1}^{n} \widehat{\nu}_{z1}^{(r+1)} I_{\{i\}}(y_z) \ln\ \widehat{p_1}^{(r)} + \sum_{z=1}^{n} \widehat{\nu}_{z2}^{(r+1)} I_{\{j\}}(y_z) \ln\ \widehat{p_2}^{(r)} \\
&\quad + \sum_{z=1}^{n} \widehat{\nu}_{zk}^{(r+1)} I_{A^*}(y_z) \ln \sum_{k=1}^{m} \sum_{d=1}^{N} N(-1)^N e^{(2N-d)\mathbf{X}_z \widehat{\mathbf{B}}_z^{(r)}} \widehat{p_3}^{(r)} \widehat{\varphi}_k^{(r)} \\
&\quad \times \frac{\Gamma(N + y_{zl})\Gamma(2\widehat{\alpha}_k^{(r)})\Gamma(\widehat{\alpha}_k^{(r)} + N - d + y_{zl})}{\Gamma(y_{zl} - 1)\Gamma(N)\Gamma(\widehat{\alpha}_k^{(r)})\Gamma(2\widehat{\alpha}_k^{(r)} + N - d + y_{zl})})
\end{aligned}$$

where $A^* = \{0, 1, \cdots\} \setminus \{i, j\}$, $I.(\cdot)$ stands for indicator function and

$$\widehat{\theta}^{(r+1)} = (\widehat{\mathbf{B}}^{(r+1)}, \widehat{\alpha}^{(r+1)}, \widehat{\varphi}^{(r+1)}, \widehat{p_1}^{(r+1)}, \widehat{p_2}^{(r+1)})$$

have been arrived by solving the following equation

$$\frac{\partial \mathbf{Q}}{\partial \alpha_k} = 0; \qquad \frac{\partial \mathbf{Q}}{\partial \mathbf{B}_z} = 0; \qquad \frac{\partial \mathbf{Q}}{\partial \varphi_k} = 0; \qquad \frac{\partial \mathbf{Q}}{\partial p_1} = 0; \qquad \frac{\partial \mathbf{Q}}{\partial p_2} = 0$$

the above equation can not solve explicitly, so

$$\widehat{\alpha}_k^{(r+1)} = \widehat{\alpha}_k^{(r)} + E^{-1}\left(\frac{-\partial^2 \mathbf{Q}}{\partial \alpha_k^2}\right) \frac{-\partial \mathbf{Q}}{\partial \alpha_k^2}; \qquad \widehat{\mathbf{B}}_z^{(r+1)} = \widehat{\mathbf{B}}_z^{(r)} + E^{-1}\left(\frac{-\partial^2 \mathbf{Q}}{\partial \mathbf{B}_z^2}\right) \frac{-\partial \mathbf{Q}}{\partial \mathbf{B}_z^2};$$

$$\widehat{\varphi}_k^{(r+1)} = \widehat{\varphi}_k^{(r)} + E^{-1}\left(\frac{-\partial^2 \mathbf{Q}}{\partial \varphi_k^2}\right) \frac{-\partial \mathbf{Q}}{\partial \varphi_k^2}; \qquad \widehat{p_1}^{(r+1)} = \widehat{p_1}^{(r)} + E^{-1}\left(\frac{-\partial^2 \mathbf{Q}}{\partial p_1^2}\right) \frac{-\partial \mathbf{Q}}{\partial p_1^2};$$

$$\widehat{p_2}^{(r+1)} = \widehat{p_2}^{(r)} + E^{-1}\left(\frac{-\partial^2 \mathbf{Q}}{\partial p_2^2}\right) \frac{-\partial \mathbf{Q}}{\partial p_2^2}.$$

Then

$$\begin{aligned}
\iota_c^{(r+1)} &= \sum_{z=1}^{n} \widehat{\nu}_{z1}^{(r+1)} I_{\{i\}}(y_z) \ln\ \widehat{p_1}^{(r+1)} + \sum_{z=1}^{n} \widehat{\nu}_{z2}^{(r+1)} I_{\{j\}}(y_z) \ln\ \widehat{p_2}^{(r+1)} \\
&\quad + \sum_{z=1}^{n} \widehat{\nu}_{zk}^{(r+1)} I_{A^*}(y_z) \ln \sum_{k=1}^{m} \sum_{d=1}^{N} N(-1)^N e^{(2N-d)\mathbf{X}_z \widehat{\mathbf{B}}_z^{(r+1)}} \widehat{p_3}^{(r+1)} \widehat{\varphi}_k^{(r+1)} \\
&\quad \times \frac{\Gamma(N + y_{zl})\Gamma(2\widehat{\alpha}_k^{(r+1)})\Gamma(\widehat{\alpha}_k^{(r+1)} + N - d + y_{zl})}{\Gamma(y_{zl} - 1)\Gamma(N)\Gamma(\widehat{\alpha}_k^{(r+1)})\Gamma(2\widehat{\alpha}_k^{(r+1)} + N - d + y_{zl})}),
\end{aligned}$$

where $A^* = \{0, 1, \cdots\} \setminus \{i, j\}$ and $I.(\cdot)$ stands for indicator function.