Thailand Statistician
July 2012; 10(2): 183-197
http://statassoc.or.th
Contributed paper

A Model for Overdispersion and Underdispersion using Latent Markov Processes

Jeeraporn Thaithanan [a]*, Sujit K. Ghosh [b], and Chinnaphong Bumrungsup [a]

[a] Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani 12120, Thailand.

[b] Department of Statistics, NC State University, Raleigh, NC 27695-8203, USA.

*Author for correspondence; e-mail: jepum@yahoo.com

Received: 11 April 2012

Accepted: 22 August 2012

Abstract

A new model for both overdispersion and underdispersion using latent Markov processes modeled a stationary processes is proposed. The parameters in this model can be estimated by the Bayesian method. The performance of the proposed method for the new model, evaluating in term of bias, MSE and coverage probability, has been explored using numerical methods based on simulated and real data.

Keywords: Bayesian Method, Markov Processes, Monte Carlo Simulation, Overdispersion, Underdispersion, Zero-Altered Distribution.

1. Introduction

Count data are commonly modeled using various parametric discrete models in practice such as Poisson and Negative Binomial distributions, etc. The Poisson distribution, which is generally used as the standard distribution for count data, is most frequently used by many practitioners, though it is restricted by the fact that the mean of this distribution is equal to its variance, known as equidispersion. However, in many practical scenarios the mean of a data set is often observed to be less than the variance of data depicting a property known as overdispersion. Kibria [1] and Yang, et al. [2] suggested that the overdispersion often occur with Poisson likelihood. If we ignored, it can lead to loss of efficiency or obtain incorrect inference. Sometimes, count data contain a large proportion of zeros, which can also result into overdispersion. For example the number of heart attacks in a given period or the number of car accidents in a given period often have excess of counts with zero. Therefore, in lieu of using Poisson model, the other count data models such as zero-inflated Poisson model could be used. However, the case of underdispersion in the data is also important and is not inevitable even though it is relatively rare in practice. This problem is not as well explored as the overdispersion problem, but it can be observed in practice [3]. Ridout and Besbeas [4] presented many underdispersed count models such as the double Poisson distribution etc.

To develop models that can account for both overdispersion and underdispersion is more important, mostly based on parametric models, Ghosh and Kim [5] were among the first ones to develop a new class of zero-altered distribution that can account for both dispersion and more importantly their models are not restricted to be parametric. It is a flexible class of semiparametric models that avoids practical limitations. It is appropriate only independent count data. However, in many situations, count data often are observed over time or in space across various locations. Such count data tend to be correlated across time and/or space, so we call these as correlated count data. The models that we discussed above are inadequate for modeling correlated count data. Many researchers proposed the models for correlated count data, for example, Dagne [6] presented a Bayesian hierarchical zero-inflated Poisson model which account for both overdispersion and excess zeros for correlated count data. Lee and Wang [7] presented a multi-level ZIP regression model for correlated count data with excess zeros. Note that most of models only account for overdispersion using correlated count data. But the problem of underdispersion for correlated count data is also not inevitable. Hence, in this article we would propose a new model that can account for both overdispersion and underdispersion using latent Markov processes (see [8-10]). The parameters in this model can be estimated via the Bayesian method which is developed as an alternative to analyze such data. It performs well and can be helpful in complex modeling situations where a frequentist method is difficult to estimate or does not exist. In addition, we also evaluate the performance of the proposed methods for this model in term of bias, MSE and coverage probability of 95% posterior intervals using simulated data scenarios.

In Section 2, we described a detail of model for both overdispersion and underdispersion using latent Markov processes. The Bayesian analysis for estimating parameters in the model is presented in Section 3. We illustrate the results based on a simulation and case study in Section 4. Finally, the conclusion is given in Section 5.

2. Model for both Overdispersion and Underdispersion using Latent Markov Processes

We propose the following models for $B_t's$ and $U_t's$ which are allowed to be correlated over time:

$$B_1 \sim \text{Bernoulli}\left(\frac{\alpha_0}{\alpha_0+\beta_0}\right), \quad 0 \le \alpha_0, \beta_0 \le 1$$

$$B_t | B_{t-1} \sim \text{Bernoulli}\left(\alpha_0(1-B_{t-1}) + (1-\beta_0)B_{t-1}\right), \quad t = 2, 3, \dots$$

$$U_1 \sim \text{Poisson}\left(\lambda_1 + \log(1 - e^{-\lambda_0} + e^{-\lambda_1})\right), \quad \lambda_0, \lambda_1 \ge 0$$

$$U_t | U_{t-1} \sim \text{Poisson}\left(\lambda_0 I(U_{t-1} = 0) + \lambda_1 I(U_{t-1} \ne 0)\right), \quad t = 2, 3, \dots$$

and finally set $X_t = B_t(1+U_t)$, for $t = 1, 2, \dots$ where X_t be a discrete valued random variable (r.v.) taking values in $I = \{0, 1, 2, \dots\}$. B_t and U_t are independent.

For the above model, we choose the parameter of the initial distribution of B_1 to be $\frac{\alpha_0}{\alpha_0+\beta_0}$ because this choice leads to a stationary process for $B_t's$. Note that the mean of B_t ($E[B_t]$) and the variance of B_t ($\text{Var}[B_t]$) do not change over time. Additionally, we choose the parameter of the initial distribution of U_1 to be $\lambda_1 + \log(1 - e^{-\lambda_0} + e^{-\lambda_1})$. This choice also leads to a stationary process for $U_t's$ because the mean of U_t ($E[U_t]$) and the variance of U_t ($\text{Var}[U_t]$) do not change over time. Hence, we have a stationary process for $X_t's$, which implies that all marginal distributions possess the same distribution which can account for both over and under dispersion. See the derivation of the mean and the variance of B_t and U_t in Appendix A.

Furthermore, the correlation between B_t and B_{t-1} is then given by

$$\text{Corr}[B_t, B_{t-1}] = 1 - (\alpha_0 + \beta_0), \qquad -1 \le 1 - (\alpha_0 + \beta_0) \le 1 \quad \forall \alpha_0, \beta_0 \in [0,1]$$

and the correlation between U_t and U_{t-1} is given by

$$\text{Corr}[U_t, U_{t-1}] = \frac{\text{cov}[U_t, U_{t-1}]}{\sqrt{\text{Var}[U_t]\text{Var}[U_{t-1}]}}$$

where

$$\text{cov}[U_t, U_{t-1}] = \left[\frac{(\lambda_1 - \lambda_0)e^{-\lambda_1}}{1 - e^{-\lambda_0} + e^{-\lambda_1}}\right] E[U_t]$$

$$E[U_t] = \frac{\lambda_0 e^{-\lambda_0} + \lambda_1(1 - e^{-\lambda_0})}{1 - e^{-\lambda_0} + e^{-\lambda_1}}, \quad \forall t$$

$$\mathrm{Var}[U_t] = \lambda_1 + (\lambda_0 - \lambda_1)\left[\frac{e^{-\lambda_1}}{1 - e^{-\lambda_0} + e^{-\lambda_1}}\right]\left[1 + (\lambda_0 - \lambda_1)\frac{1 - e^{-\lambda_0}}{1 - e^{-\lambda_0} + e^{-\lambda_1}}\right], \ \ \forall t.$$

Parameter estimation in the proposed model is discussed in the next section.

3. Bayesian Analysis

In Bayesian analysis, we start with the prior distribution that contains all information about the parameter values before observing the data. When the prior distribution is updated, we obtain the posterior distribution. Actually, the posterior distribution is proportional to the product of the likelihood function and the prior distribution of the parameters.

Here, we propose the prior distribution for the parameter $\theta = (\alpha_0, \beta_0, \lambda_0, \lambda_1)$ as follows:

$$\alpha_0 \sim \text{Beta}(a_1, b_1) \ \text{ and } \ \beta_0 \sim \text{Beta}(a_2, b_2)$$
$$\lambda_0 \sim \text{Gamma}(c_1, d_1) \ \text{ and } \ \lambda_1 \sim \text{Gamma}(c_2, d_2).$$

Consider the likelihood function of (α_0, β_0), we have

$$L(\alpha_0, \beta_0 | B_1, \dots, B_n) \approx \prod_{t=2}^{n}\left[(1-\beta_0)^{B_{t-1}B_t}\,(\alpha_0)^{(1-B_{t-1})B_t}\right]\left[(\beta_0)^{B_{t-1}(1-B_t)}\,(1-\alpha_0)^{(1-B_{t-1})(1-B_t)}\right] \tag{3.1}$$

and the likelihood function of (λ_0, λ_1) is given by

$$L(\lambda_0, \lambda_1 | U_1, \dots, U_n) \approx \prod_{t=2}^{n}\left(\frac{e^{-\lambda_0}{\lambda_0}^{U_t}}{U_t!}\right)^{I(U_{t-1}=0)}\left(\frac{e^{-\lambda_1}{\lambda_1}^{U_t}}{U_t!}\right)^{I(U_{t-1}\neq 0)}. \tag{3.2}$$

Then, the posterior distribution of (α_0, β_0) is given by

$$\alpha_0 | B_1, \dots, B_n \sim \text{Beta}\left(\sum_{t=2}^{n} B_t - \sum_{t=2}^{n} B_{t-1}B_t + a_1, n - \sum_{t=2}^{n} B_t - \sum_{t=2}^{n} B_{t-1} - \sum_{t=2}^{n} B_{t-1}B_t + b_1\right)$$

$$\beta_0 | B_1, \dots, B_n \sim \text{Beta}\left(\sum_{t=2}^{n} B_{t-1} - \sum_{t=2}^{n} B_{t-1}B_t + a_2, \sum_{t=2}^{n} B_{t-1}B_t + b_2\right)$$

and the posterior distribution of (λ_0, λ_1) is given by

$$\lambda_0 | U_1, \dots, U_n \sim \text{Gamma}\left(\sum_{t=2}^{n} I(U_{t-1}=0)U_t + c_1, \sum_{t=2}^{n} I(U_{t-1}=0) + d_1,\right)$$

$$\lambda_1 | U_1, \dots, U_n \sim \text{Gamma}\left(\sum_{t=2}^{n}\left(1 - \sum_{t=2}^{n} I(U_{t-1}=0)\right)U_t + c_2, n - 1 - \sum_{t=2}^{n} I(U_{t-1}=0) + d_2\right)$$

The function $I(U_{t-1}=0)$ is defined as

$$I(U_{t-1}=0)=\begin{cases}1 & \text{if } U_{t-1}=0,\\ 0 & \text{if } U_{t-1}\neq 0.\end{cases}$$

Here, we use the Bayesian software package known as WinBUGS [11] using a Markov Chain Monte Carlo (MCMC) method to compute the posterior mean as the Bayes estimators of $(\alpha_0, \beta_0, \lambda_0, \lambda_1)$ in the proposed model given in Section 2.

4. Simulation and Case Study

In this section, we present the performance of the Bayesian method for this model which can account for both overdispersion and underdispersion using latent Markov processes. We use a simulation study to present the performance of the proposed method. We consider the posterior summary statistics such as posterior mean, median, standard deviation and 95% posterior intervals for each parameter estimate. This provides the Monte Carlo bias (median of posterior median - true value), the Monte Carlo mean squared error (MSE) and the coverage probability of 95% posterior intervals as the proportion of times the true value of parameter was included in the 95% posterior intervals to measure the performance of estimates. We also analyze a real data set shown in Figure 4 using the method in Section 3.

4.1 A Simulation Study

A simulation study is carried out to study the performance of the Bayesian method. We generate data sets B_t and U_t of size n = 50, n = 100 and n = 300 from the following model

$$B_1 \sim \text{Bernoulli}\left(\frac{\alpha_0}{\alpha_0+\beta_0}\right), \quad 0 \le \alpha_0, \beta_0 \le 1$$

$$B_t|B_{t-1} \sim \text{Bernoulli}\left(\alpha_0(1-B_{t-1})+(1-\beta_0)B_{t-1}\right), \quad t=2,3,\dots$$

$$U_1 \sim \text{Poisson}\left(\lambda_1+\log(1-e^{-\lambda_0}+e^{-\lambda_1})\right), \quad \lambda_0, \lambda_1 \ge 0$$

$$U_t|U_{t-1} \sim \text{Poisson}\left(\lambda_0 I(U_{t-1}=0)+\lambda_1 I(U_{t-1}\neq 0)\right), \quad t=2,3,\dots$$

and finally set $X_t = B_t(1+U_t)$, for $t=1,2,\dots$

We choose the true value $\alpha_0 = \beta_0$ =(0.25, 0.50, 0.75) and $\lambda_0 = \lambda_1 =$ (1,2). In total, we have 36 different combinations with each of sample size n. We replicate the data generation each of sample size n for N = 500 Monte Carlo (MC) simulation runs. We use WinBUGS to generate 10,000 additional iterations obtaining the posterior estimates following a burn-in of 5,000 iterations with the prior distribution as

$$\alpha_0 \sim \text{Beta}(1,1)$$

$$\beta_0 \sim \text{Beta}(1,1)$$

$$\lambda_0 \sim \text{Gamma}(0.1, 0.1)$$

$$\lambda_1 \sim \text{Gamma}(0.1, 0.1).$$

Some results are summarized numerically in Table 1-2. The posterior summary of the Bayes estimators of $(\alpha_0, \beta_0, \lambda_0, \lambda_1)$, consisting of the posterior mean, sd (standard deviation), median, 95% posterior intervals, bias, MSE and coverage probability of 95% posterior intervals, are presented in Table 1-2. In addition, we illustrate the performance of parameters by Box Plots for each of the sample size n=50 (first row), n=100 (second row) and n=300 (third row) as shown in Figures 1. Furthermore, we also present a graphical summary of absolute bias and MSE for all cases in Figures 2-3, respectively, for evaluating the performance of the proposed Bayesian methods.

In Figures 2-3, it is clear that the Bayesian method provides asymptotically unbiased estimators of α_0 and β_0 nearly for all parameter combinations. Notice that not only the magnitude of the biases is very small but also has small MSE for all sample sizes. The biases and MSE of α_0 and β_0 tend to decrease when the sample sizes increase. This implies that it looks close to the true value. In addition, the performance of the proposed method, as measured by the coverage probability of 95% posterior intervals, is good. It is close to the desired value of 0.95 in each case.

In the case of the Bayes estimator of λ_0, the Bayesian method still provides an asymptotically unbiased estimator for all combinations of $\beta_0 =$ 0.25 such as the combinations of (α_0 =0.75, β_0 =0.25, λ_0 =1, λ_1 =2), etc. It is clear that the biases and MSE are very small for all sample sizes. In particular, the MSE of λ_0 appears small and significantly decreases when the sample size is large. Meanwhile, all combinations of β_0 =0.50 also have small biases and the MSE is small and tend to decrease when increasing the sample sizes. Furthermore, the coverage probability of 95% posterior intervals of λ_0 is close to the nominal 95% level in each case.

In the case of the Bayes estimator of λ_1, the Bayesian method still provides an asymptotically unbiased estimator for all combinations of $\alpha_0 =$ (0.50,0.75) and ($\lambda_0 \le \lambda_1$) such as the combinations of (α_0 =0.75, β_0 =0.25, λ_0 =1, λ_1 =2), etc. It is clear that not only the magnitude of the biases are very small but also has small MSE for all sample sizes. In addition, the MSE of λ_1 tend to decrease when increasing the sample sizes. When we consider the coverage probability of 95% posterior intervals of λ_1, note that it is close to 0.95. It means that the estimator of λ_1 is close to the true value.

4.2 Monthly Claim Counts

We apply the proposed model based on the Bayesian method with the claim counts data which is observed over time and tend to be correlated across time. All claims are collected due to a burn related injury [12]. The data is shown in Figure 4, which consists of monthly claim counts of workers in the heavy manufacturing industry from the Richmond claims center between January 1987 and December 1994. Empirical mean and variance of the data are given by 8.6042 and 11.3575, respectively, indicating that the data is overdispersed. We use WinBUGS to compute the posterior mean as the Bayes estimators of $(\alpha_0, \beta_0, \lambda_0, \lambda_1)$ with the prior distribution as $\alpha_0 \sim$ Beta$(1,1)$, $\beta_0 \sim$ Beta$(1,1)$, $\lambda_0 \sim$ Gamma$(0.1,0.1)$ and $\lambda_1 \sim$ Gamma$(0.1,0.1)$. The result is presented in Table 3. First, we consider the MC error, it gives us an idea about how good the parameters estimation are. Note that the MC error in Table 3 gives small values for all parameters. The posterior mean and 95% intervals of α_0 and β_0 are 0.5103, [0.0376, 0.9735] and 0.0098, [0.0003, 0.0365], respectively. In addition, the posterior mean and 95% intervals of λ_0 and λ_1 are 1.7780, [0.2042, 5.0900] and 7.6550, [7.1110, 8.2170], respectively.

Next, we present the predicted values of X_t compared with the observed values (X_t) to show how good is the prediction as shown in Figure 5. The bias of predicted values is shown in Figure 6. Notice that most of the bias of predicted biases is small and only 7 of the 95% intervals do not contain zero out of 96 cases. So, we can conclude that the proportion was included in the 95% intervals expected as ≈ 93%. It is close to 0.95. Hence, we could use this model to apply for such data based on the Bayesian method.

Table 1. Posterior Summary for (α_0 =0.50, β_0 =0.25, λ_0 =1, λ_1 =2).

n	Parameter	Mean	SD	Median	95% Posterior Interval	Bias	MSE	Coverage Prob.
50	α_0	0.5140	0.1130	0.5149	(0.2931,0.7291)	0.0099	0.0131	0.9640
	β_0	0.2674	0.0728	0.2628	(0.1388,0.4212)	0.0099	0.0058	0.9360
	λ_0	1.1917	0.6737	1.0457	(0.3596,2.8801)	-0.0292	0.3099	0.9520
	λ_1	1.9983	0.2603	1.9868	(1.5216,2.5393)	-0.0065	0.0703	0.9380
100	α_0	0.5030	0.0826	0.5032	(0.3417,0.6632)	0.0042	0.0070	0.9500
	β_0	0.2594	0.0524	0.2570	(0.1640,0.3681)	0.0020	0.0029	0.9400
	λ_0	1.1287	0.3925	1.0788	(0.5133,2.0287)	0.0125	0.1583	0.9500
	λ_1	2.0140	0.1848	2.0083	(1.6685,2.3916)	0.0095	0.0355	0.9380
300	α_0	0.5009	0.0492	0.5015	(0.4046,0.5970)	0.0015	0.0024	0.9540
	β_0	0.2545	0.0306	0.2515	(0.1970,0.3166)	0.0015	0.0011	0.9460
	λ_0	1.0857	0.2068	1.0545	(0.7190,1.5267)	0.0545	0.0477	0.9360
	λ_1	2.0144	0.1062	2.0160	(1.8117,2.2276)	0.0160	0.0104	0.9540

Table 2. Posterior Summary for (α_0 =0.75, β_0 =0.25, λ_0 =1, λ_1 =2).

n	Parameter	Mean	SD	Median	95% Posterior Interval	Bias	MSE	Coverage Prob.
50	α_0	0.7167	0.1122	0.7276	(0.4727,0.9006)	-0.0248	0.0124	0.9520
	β_0	0.2638	0.0687	0.2597	(0.1418,0.4088)	0.0032	0.0048	0.9520
	λ_0	1.1528	0.5969	1.0341	(0.3615,2.6228)	-0.0400	0.2489	0.9740
	λ_1	1.9767	0.2444	1.9667	(1.5270,2.4830)	-0.0362	0.0601	0.9580
100	α_0	0.7390	0.0817	0.7452	(0.5636,0.8795)	-0.0005	0.0066	0.9640
	β_0	0.2600	0.0495	0.2579	(0.1695,0.3626)	0.0050	0.0027	0.9520
	λ_0	1.1054	0.3532	1.0661	(0.5317,1.9007)	0.0145	0.1068	0.9620
	λ_1	1.9878	0.1734	1.9826	(1.6627,2.3417)	-0.0285	0.0327	0.9280
300	α_0	0.7440	0.0493	0.7465	(0.6418,0.8340)	-0.0035	0.0024	0.9540
	β_0	0.2514	0.0287	0.2508	(0.1973,0.3095)	0.0008	0.0009	0.9480
	λ_0	1.0874	0.1943	1.0555	(0.7405,1.4996)	0.0555	0.0475	0.9200
	λ_1	1.9950	0.1000	1.9965	(1.8037,2.1954)	-0.0035	0.0103	0.9540

(a) (α_0 =0.50, β_0 =0.25, λ_0 =1, λ_1 =2) (b) (α_0 =0.75, β_0 =0.25, λ_0 =1, λ_1 =2)

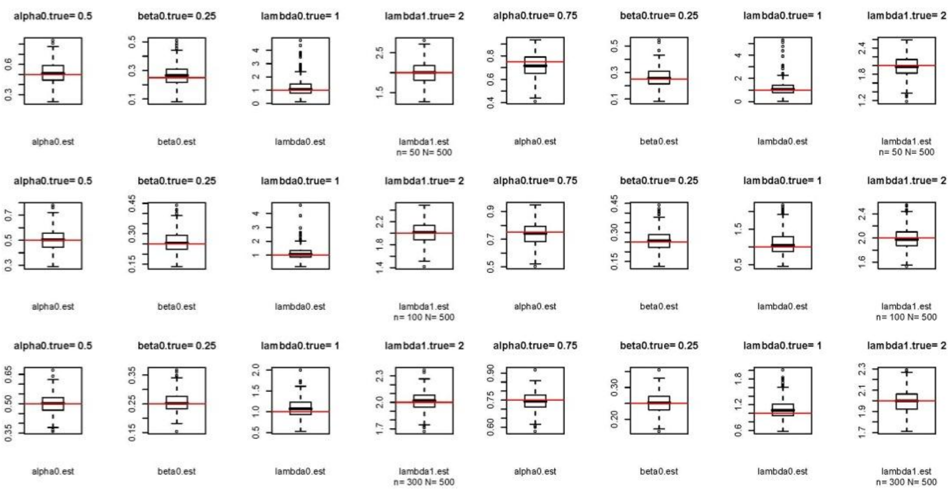

Figure 1. Performance of the Bayesian method.

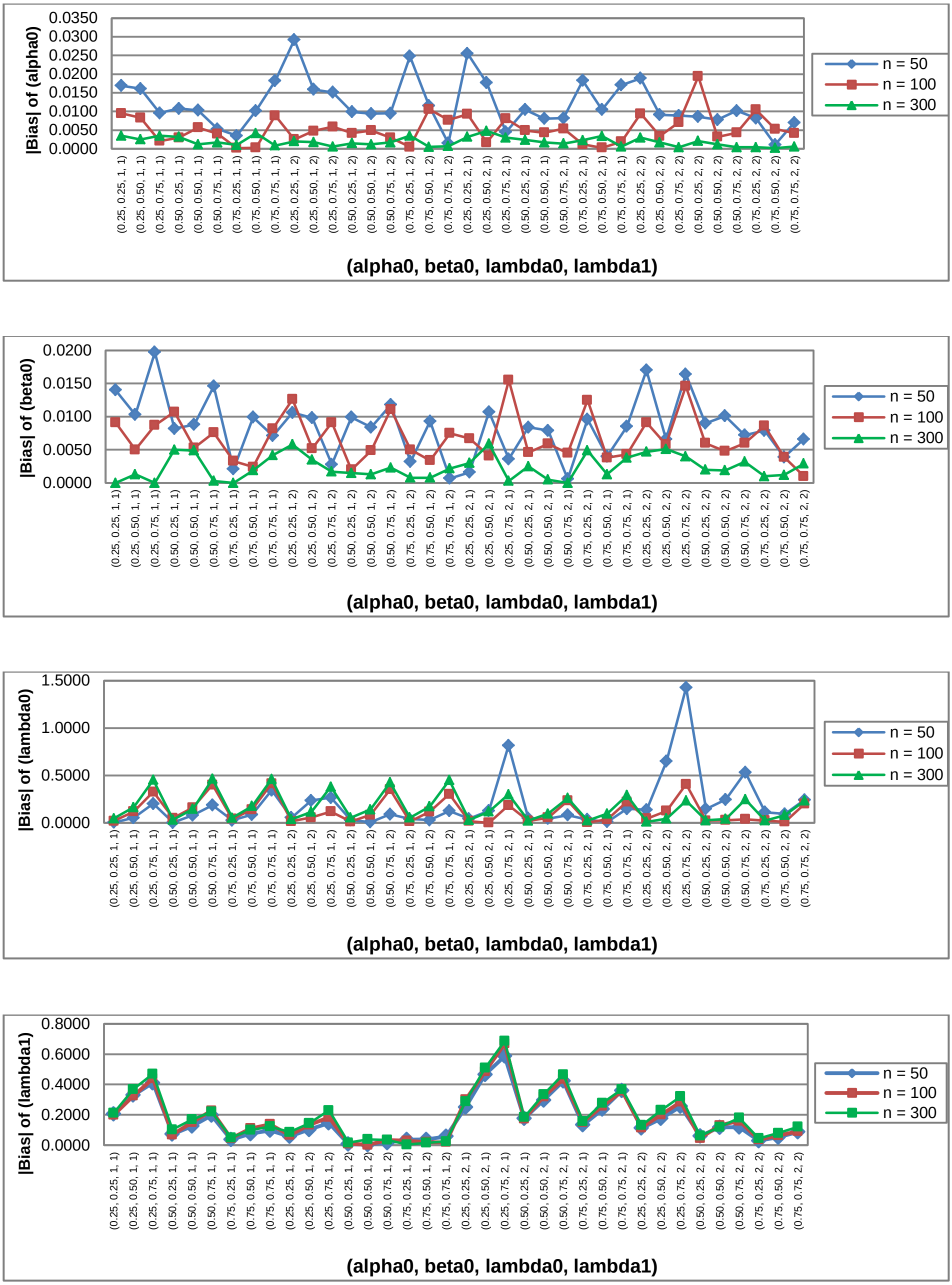

Figure 2. Absolute Bias of the Bayes Estimators of $(\alpha_0, \beta_0, \lambda_0, \lambda_1)$.

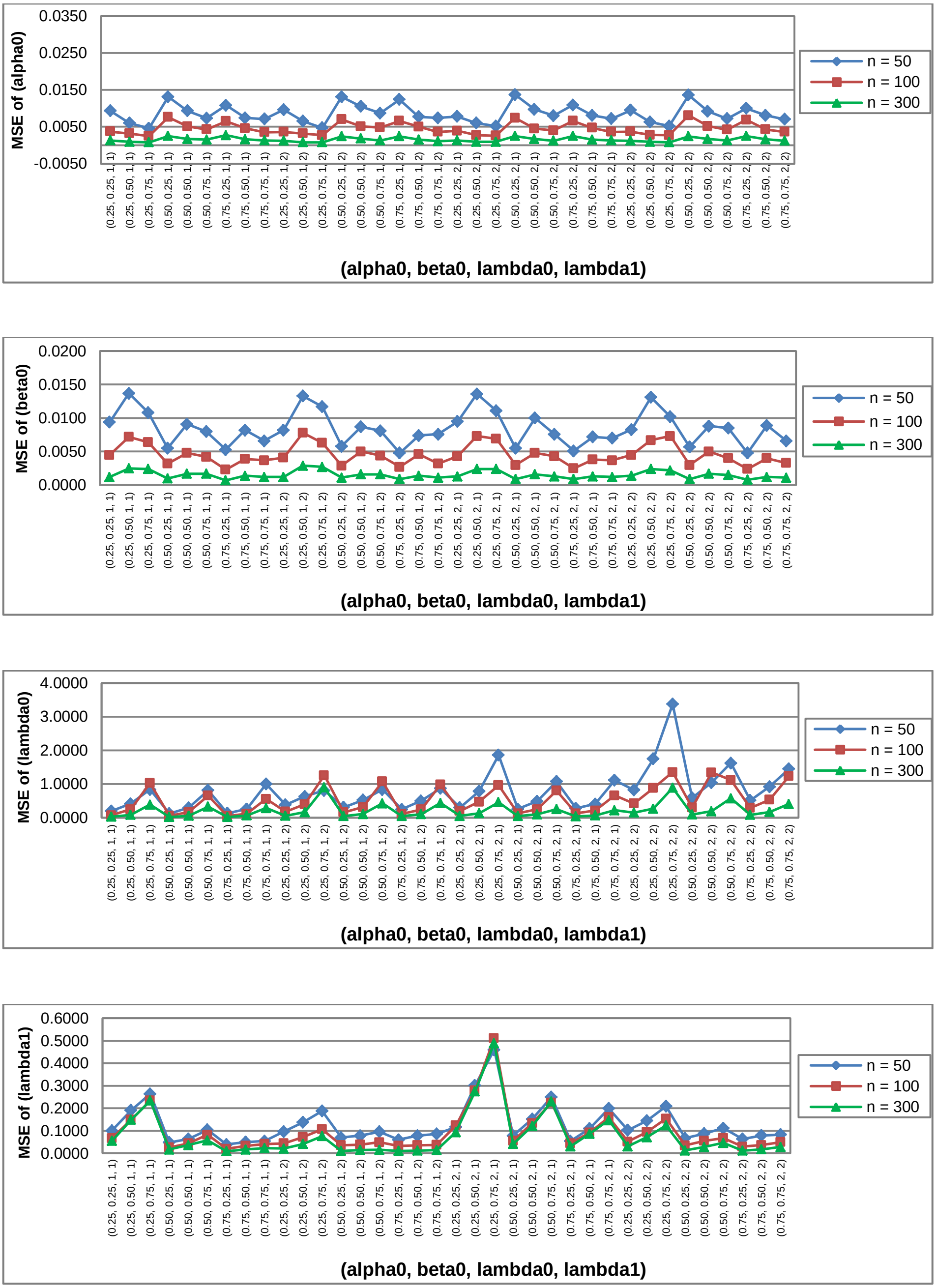

Figure 3. MSE of the Bayes Estimators of $(\alpha_0, \beta_0, \lambda_0, \lambda_1)$.

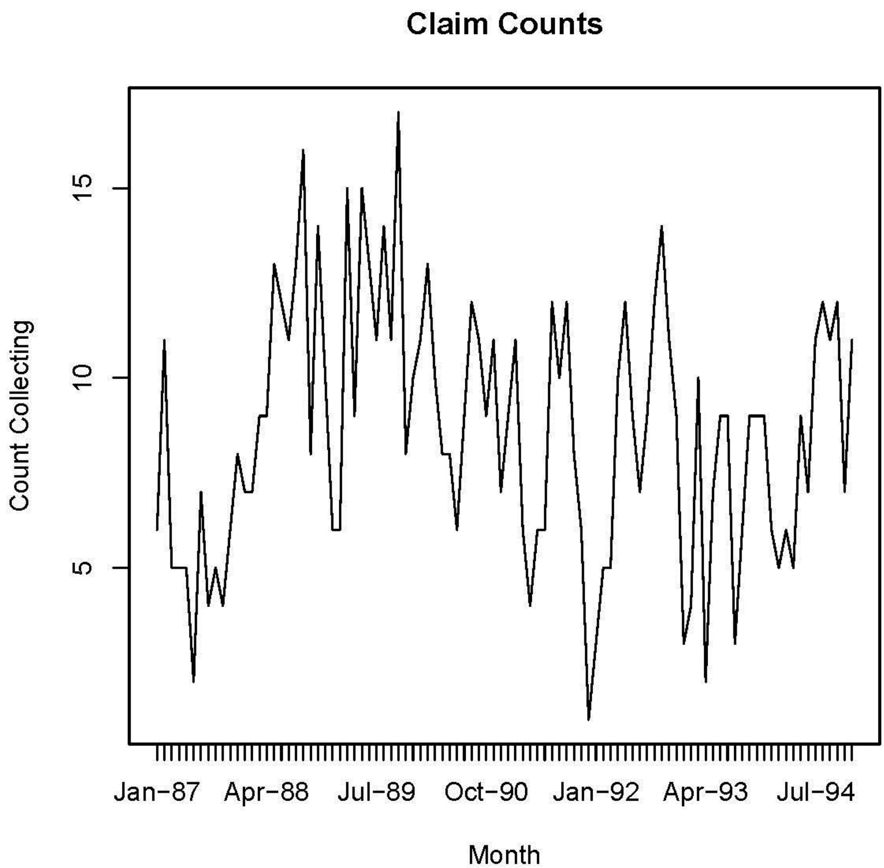

Figure 4. The Claim Counts Data.

Table 3. WinBUGS Output.

node	mean	sd	MC error	2.5%	median	97.5%	start	sample
α_0	0.5103	0.2820	0.0028	0.0376	0.5119	0.9735	5001	10000
β_0	0.0098	0.0099	0.0002	0.0003	0.0067	0.0365	5001	10000
λ_0	1.7780	1.2800	0.0192	0.2042	1.4960	5.0900	5001	10000
λ_1	7.6550	0.2855	0.0032	7.1110	7.6480	8.2170	5001	10000

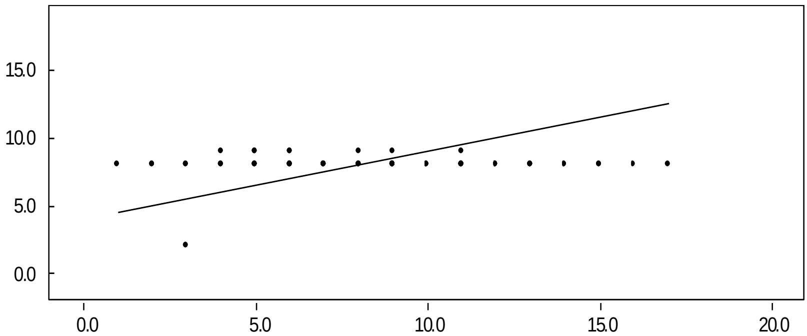

Figure 5. Predicted vs Observed Values.

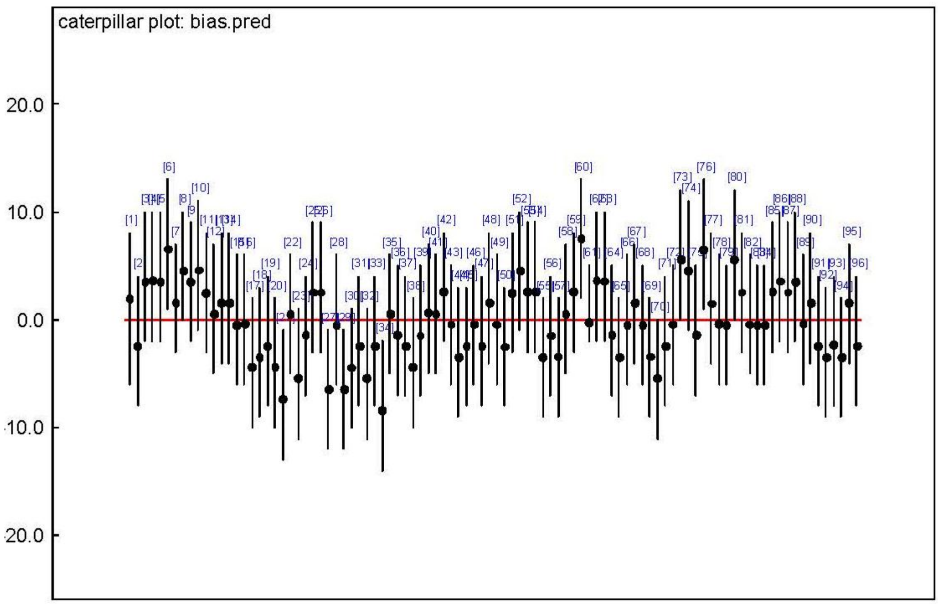

Figure 6. Bias of Predicted Values.

5. Conclusion

In this article, a new model that can account for both overdispersion and underdispersion using latent Markov processes is proposed. The parameters in this model can be estimated via the Bayesian method. From a simulation study, it is clear that the proposed method performs well and provides asymptotically unbiased estimator for the parameters combinations of ($\alpha_0 \geq \beta_0 = 0.25$) and ($\lambda_0 \leq \lambda_1$) used. Notice that the

MSE appears small and significantly decrease when increasing the sample sizes. The coverage probability of 95% posterior intervals of $(\alpha_0, \beta_0, \lambda_0, \lambda_1)$ are close to 0.95.

Acknowledgements

We would like to thank the Department of Mathematics and Statistics, Thammasat University for partial support of this research.

References

[1] Kibria, B.M.G., Applications of some discrete regression models for count data, *Pakistan Journal of Statistics and Operation Research,* 2006; 2(1): 1-16.

[2] Yang, Z., Hardin, J.W., and Addy, C.L., Testing overdispersion in the zero-inflated Poisson moel, *Journal of Statistical Planning and Inference., 2009*; 139: 3340-3353.

[3] Famoye, F., Restricted Generalized Poisson Regression Model, *Communications in Statistics - Theory and Methods*, 1993; 22(5): 1335-1354.

[4] Ridout, M. S. and Besbeas, P., An empirical model for underdispersed count data. *Statistical Modelling*, 2004; *4*: 77-89.

[5] Ghosh, S.K., and Kim, H., Semiparametric inference based on a class of zero-altered distributions, *Statistical Methodology,* 2007; 4(3), 371-383.

[6] Dagne, G. A., Hierarchical Bayesian Analysis of Correlated Zero-inflated Count Data, *Biometrical Journal*, 2004; 46: 653-663.

[7] Lee, A.H., and Wang, K., Multi-level zero-inflated Poisson Regression modeling of correlated count data with excess zero, *Statistical Methods in Medical Research,* 2006; 15: 47-61.

[8] Rykov, V.V., Balakrishnan, N. and Nikulin, M.S., *Mathematical and Statistical Models and Methods in Reliability*, Springer, Boston, 2010.

[9] Vermunt, J.K., Latent Markov Model, In: M.S. Lewis-Beck, A. Bryman, and T.F. Liao (Eds.), *The Sage Encyclopedia of Social Sciences Research Methods,* 2004; 553-554.

[10] Medhi, J., *Stochastic Processes*, New Age Science Limited, 2010.

[11] Spiegelhalter, D. J., Thomas, A., Best, N. G. and Lunn, D., *WinBUGS Version 1.4 User Manual*, 2001.

[12] Freeland, R.K., *Statistical analysis of discrete time series with applications to the analysis of workers compensation claims data*, PhD Thesis, University of British Columbia, Canada, 1998.

Appendix A

The properties of B_t :

$$B_1 \sim \text{Bernoulli}\left(\frac{\alpha_0}{\alpha_0+\beta_0}\right), \quad 0 \leq \alpha_0, \beta_0 \leq 1$$

$$B_t | B_{t-1} \sim \text{Bernoulli}\,(\alpha_0(1-B_{t-1}) + (1-\beta_0)B_{t-1}), \quad t = 2, 3, \ldots$$

- The mean of $\boldsymbol{B_t}$:

$$E[B_t] = E[E(B_t|B_{t-1})]$$

$$= \alpha_0 + (1-\alpha_0-\beta_0)\mu_{t-1}, \; E[B_{t-1}] = \mu_{t-1}$$

$$t = 2, \; \mu_2 = \alpha_0 + (1-\alpha_0-\beta_0)\mu_1, \; E[B_1] = \mu_1$$

$$\vdots$$

$$t = t, \; \mu_t = \frac{\alpha_0}{\alpha_0+\beta_0} + (1-\alpha_0-\beta_0)^{t-1}\left[\mu_1 - \frac{\alpha_0}{\alpha_0+\beta_0}\right]$$

Hence,

$$E[B_t] = \frac{\alpha_0}{\alpha_0+\beta_0}, \quad \forall t$$

- The variance of $\boldsymbol{B_t}$:

$$\text{Var}[B_t] = E\left[{B_t}^2\right] - \{E[B_t]\}^2$$

$$E\left[{B_t}^2\right] = E[B_t] = \frac{\alpha_0}{\alpha_0+\beta_0}$$

Hence,

$$\text{Var}[B_t] = \frac{\alpha_0\beta_0}{(\alpha_0+\beta_0)^2}, \quad \forall t$$

The properties of U_t :

$$U_1 \sim \text{Poisson}\left(\lambda_1 + \log(1 - e^{-\lambda_0} + e^{-\lambda_1})\right), \quad \lambda_0, \lambda_1 \geq 0$$

$$U_t | U_{t-1} \sim \text{Poisson}\left(\lambda_0 I(U_{t-1} = 0) + \lambda_1 I(U_{t-1} \neq 0)\right), \quad t = 2, 3, \ldots$$

- The mean of $\boldsymbol{U_t}$:

$$E[U_t] = E[E(U_t|U_{t-1})]$$

$$= \lambda_0 P_{t-1} + \lambda_1(1 - P_{t-1})$$

$$P_t = P[U_t = 0] = E[I(U_t = 0)]$$

$$= e^{-\lambda_0}P_{t-1} + e^{-\lambda_1}(1 - P_{t-1})$$

$$= b + (a-b)P_{t-1}, \; a = e^{-\lambda_0}, \; b = e^{-\lambda_1}$$

$$t = 2, \quad P_2 = b + (a-b)P_1$$

$$\vdots$$

$$t = t, \quad P_t = \frac{b}{1-(a-b)} + (a-b)^{t-1}\left[P_1 - \frac{b}{1-(a-b)}\right]$$

where, $P_1 = P[U_1 = 0] = \dfrac{b}{1-(a-b)}$

Hence, $$P_t = \frac{e^{-\lambda_1}}{1 - e^{-\lambda_0} + e^{-\lambda_1}}, \quad \forall t$$

Therefore, $$E[U_t] = \frac{\lambda_0 e^{-\lambda_0} + \lambda_1(1 - e^{-\lambda_0})}{1 - e^{-\lambda_0} + e^{-\lambda_1}}, \quad \forall t$$

- The variance of $\boldsymbol{U_t}$ **:**

$$\text{Var}[U_t] = E\left[{U_t}^2\right] - \{E[U_t]\}^2$$

$$E\left[{U_t}^2\right] = \lambda_0 P_{t-1} + \lambda_1(1 - P_{t-1}) + {\lambda_0}^2 P_{t-1} + {\lambda_1}^2(1 - P_{t-1})$$

Then,

$$\text{Var}[U_t] = \lambda_1 + (\lambda_0 - \lambda_1)\left[\frac{e^{-\lambda_1}}{1 - e^{-\lambda_0} + e^{-\lambda_1}}\right]\left[1 + (\lambda_0 - \lambda_1)\frac{1 - e^{-\lambda_1}}{1 - e^{-\lambda_0} + e^{-\lambda_1}}\right], \quad \forall t$$