AN APPLICATION OF THE BAYESIAN POISSON REGRESSION IN MODELLING ROOMMATE CONFLICT AMONG UNIVERSITY OF CAPE COAST STUDENTS

Acquah Joyce De-Graft, Research Fellow
Department of Peace Studies, University of Cape Coast, Cape Coast, Ghana
E-mail: joycedaftacquah@gmail.com

ABSTRACT
This paper introduces Bayesian analysis and demonstrates its application to parameter estimation of the Poisson regression via Markov Chain Monte Carlo (MCMC) algorithm using roommate conflict data. The Bayesian Poisson regression estimation is compared with the classical Poisson regression. Both the classical Poisson regression and the Bayesian Poisson regression provide similar results and suggest that the frequency of roommate conflicts decreases with family size, number of roommates one has and being in a love relationship. The results also show a reduction of standard errors associated with some coefficients obtained from the Bayesian analysis, thus bringing greater stability to the coefficients. It is concluded that Bayesian Poisson regression estimation via MCMC algorithm offers an alternative framework for modelling roommate conflict data.

KEY WORDS
Poisson regression, Posterior Distribution, Markov Chain Monte Carlo, roommate conflict, Bayesian Analysis.

The Poisson regression model has been applied extensively to model count data as demonstrated in Winkelmann (1994), Winkelmann and Zimmermann (1991;1995), Cameron and Trivedi (1986; 1996), Winkelmann (2000;2008), Jun (2018), Land et al. (1996) and Acquah (2016). For example, Acquah (2016) applied the Poisson regression model to investigate the relationship between frequency of conflicts as the dependent variable and gender, age, family size, roommates of different religion, prior experience in a boarding house, number of roommates one has, years of education, being in a love relationship, number of sibling and remittance as the covariates. The classical Poisson regression results suggest that prior experience in a boarding house and number of Sibling is associated with increased roommate conflicts whilst family size, number of roommates one has and being in a love relationship are associated with a reduction in roommate conflict. Despite it wide spread application, the assumptions of the Poisson regression model turn out to be unrealistic. More specifically, the Poisson model involves the assumption that the mean is equal to the variance and, therefore, the model cannot account for the empirical regularity that count data are over-dispersed.

A fundamental research question which remains is that, is there an alternative method that can model the roommate conflict data without the limitation of the classical Poisson regression model and will it lead to similar results and conclusion. The foregoing discussion point to the fact that there is a need to employ a flexible model that overcome the limitations of the classical Poisson model in analysing roommate conflict data.

In order to overcome the limitations inherent in the classical estimation of the Poisson regression model, this paper introduces the Bayesian Poisson Regression modelling as an alternative approach. The Bayesian Poisson regression model offers increase flexibility that can provide substantial added value in the analysis of count data. This added value arises essentially from two factors. First, the Bayesian estimation is flexible and does not require compliance with demanding assumptions as suggested in the classical Poisson regression model. Secondly, the Bayesian model provides the opportunity of introducing prior information into the analysis.

This flexibility is further enhanced by the use of the Markov Chain Monte Carlo (MCMC) based sampling methods. Development in Markov Chain Monte Carlo (MCMC)

methods has made it possible to fit various nonlinear regression models. Irrespective of these developments, few studies have employed the MCMC based approach to model the Poisson regression. As a result, very little is understood about the concept of Bayesian analysis and its application to the Poisson regression via MCMC algorithm.

The aim of the present study is to analyse the Poisson regression model estimated by the Bayesian approach in comparison to the classical estimation using roommate conflict data.

This article therefore introduces Bayesian analysis and demonstrates its application to parameter estimation of the Poisson regression via Markov Chain Monte Carlo (MCMC) algorithm. Fundamentally, this study explores the application of the Bayesian Poisson regression and compares it with the classical Poisson regression using roommate conflict data.

MATERIALS AND METHODS OF RESEARCH

The present study is interested in comparing the Bayesian Poisson regression and the classical Poisson regression models in the estimation of the relationship between the frequency of roommate conflict and its determinate. The study therefore uses the roommate conflict data used by Acquah (2016). The data was derived from 117 students in the University of Cape Coast. The data consist of frequency of conflicts as the dependent variable and gender, age, family size, roommates of different religion, prior experience in a boarding house, number of roommates one has, years of education, being in a love relationship, number of sibling and remittance as the covariate. This data which provides information on roommate conflicts among university students allows this study to investigate the determinants of roommate conflict using a Bayesian analysis.

Bayesian Poisson Regression Model. This research considers Bayesian count data modelling with a Poisson distribution for analysing frequency of roommate conflicts. Consider a random variable y that follows a Poisson distribution with parameter λ. Then its distribution is defined as follows:

$$y \sim \text{Poisson}(\lambda) \tag{1}$$

$$P(y = y / \lambda) = \frac{e^{-\lambda}\lambda^{y}}{y!}, y = 0,1,2,\dots \tag{2}$$

Where: the expectation E (y) and variance of y are equal to parameter λ.

Furthermore, the likelihood function of Poisson random variable y is as follows:

$$F(y_1, y_2, \dots, y_n / \lambda) = \prod_{i=1}^{n} \frac{1}{y_i!} \lambda^{y_i} e^{-\lambda} \infty \lambda^{\Sigma y_i} e^{-n\lambda} \tag{3}$$

The above equation is in the form of $\lambda^c e^{-d\lambda}$, and this is the gamma distribution with parameter c and d. We can therefore select gamma distribution as conjugate prior for the Poisson parameter.

Bayesian modelling has increasingly been used in regression analysis. In the Bayesian framework, there are three key components associated with parameter estimation: the prior distribution, the likelihood function, and the posterior distribution. These three components are formally combined by Bayes' rule as:

Posterior distribution = Prior distribution x likelihood function

Bayesian count data modelling starts with the following expression for the posterior distribution:

$$P(\theta / y) = \frac{P(y/\theta)P(\theta)}{P(y)} \tag{4}$$

Where: θ is the model parameter, and y is the response variable to be predicted. $P(\theta)$ and $P(\theta/y)$ are the prior and posterior probabilities of parameter θ respectively. $P(y/\theta)$ represents the likelihood function of y given θ, respectively. Additionally, $P(y)$ is calculated by the following integration:

$$P(y) = \int P(y/\theta)P(\theta)d\theta \qquad (5)$$

Using Bayesian modelling, we determined the model parameter of posterior distribution. Because we were interest in the mean of the parameters, we had to select a prior distribution to begin Bayesian modelling. In general informative priors are used for the prior distribution in Bayesian modelling. For the purpose of this study, we assume a multivariate Normal prior on β. In order to employ Bayesian computing such as Markov Chain Monte Carlo (MCMC) an informative conjugate prior is adopted to alleviate the computational burden. For Bayesian count data modelling, we constructed the Poisson regression model with Gamma distribution as prior. The Poisson regression model is defined as follows.

Therefore, we get the Bayesian count data modelling as follows:

$$y_i|x_i \sim poisson(\lambda_i) \qquad (6)$$
$$\lambda_i = \exp(x_i'\beta) + \alpha_i, \alpha_i \sim normal(0, \sigma^2) \qquad (7)$$
$$p(\beta) : Gamma\ prior\ density\ of\ \beta \qquad (8)$$

Metropolis Hastings Algorithm. Metropolis-Hasting algorithm is an iterative algorithm that generates a Markov chain and permits empirical estimation of posterior distributions. The Metropolis Hastings algorithm (MH) produces samples from a probability distribution using full joint density function. Drawing from Gill (2002) a basic MH algorithm is made up of the following steps:

1. Establish starting values S for the parameter: $\theta^{j=0} = S$. Set $j = 1$.

The starting values can be obtained via maximum likelihood estimation.

2. Draw a ´´candidate`` parameter, $\theta^c$ from a ´´proposal density,`` $\alpha(.)$.

The simulated value is considered a ´´candidate´´ because it is not automatically accepted as a draw from the distribution of interest. It must be evaluated for acceptance.

3. Compute the ratio $R = \dfrac{f(\theta^c)\alpha(\theta^{j-1} | \theta^c)}{f(\theta^{j-1})\alpha(\theta^c | \theta^{j-1})}$

4. Compare R with a $U(0,1)$ random draw u. If $R > u$, then set $\theta^j = \theta^c$. Otherwise, set $\theta^j = \theta^{j-1}$

5. Set $j = j + 1$ and return to step 2 until enough draws are obtained.

A detail discussion on the Metropolis Algorithm is presented in Gill (2014).

RESULTS AND DISCUSSION

The model specification with frequency of conflicts as the dependent variable and gender, age, family size, roommates of different religion, prior experience in a boarding house, number of roommates one has, years of education, being in a love relationship, number of sibling and remittance as the covariates was estimated for both the Bayesian and classical Poisson regression. The classical Poisson regression and the Bayesian Poisson regression results suggest that prior experience in a boarding house and number of Sibling is associated with increases in roommate conflicts whilst family size, number of roommates one has and being in a love relationship are associated with a reduction in roommate conflict.

The posterior moments in the Bayesian Poisson estimation was obtained after a burn in period of 50,000 iterations and a follow up period of 250,000, storing every 20th iteration.

Using the posterior mean as a point estimate, Table 1 compares the classical Poisson estimates with the MCMC output.

Table 1 – Classical Poisson and Posterior Moments

Variable	Classical Poisson		Posterior	
	Mean	Std. Error	Mean	Std. Error
Intercept	-1.5394	1.0387	-1.5667	1.0480
GEN	0.1672	0.1706	0.1594	0.1724
AGE	0.0271	0.0287	0.0253	0.0284
FSIZE	-0.2633 **	0.0994	-0.2618	0.0990
DIFR	0.0385	0.1763	0.0336	0.1766
BOD	0.5919 **	0.2036	0.6117	0.2040
NRM	-0.1199 *	0.0481	-0.1229	0.0488
EDU	0.1079	0.0685	0.1104	0.0691
LOR	-0.3813 *	0.1482	-0.3859	0.1491
NOS	0.3335 ***	0.0970	0.3323	0.0968
REM	0.0002	0.0003	0.0002	0.0002

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*
NOTE: GEN= gender, AGE= age, FSIZE= family size, DIFR= Roommates of different religion, BOD= Prior experience in a boarding house, NRM= Number of roommates one has, EDU= years of education, LOR= Being in a love relationship, NOS= Number of Sibling and REM= Remittance.

It should be emphasized that a negative sign of a parameter indicates that high values of the variables tends to decrease the frequency of conflict. A positive sign implies that high values of the variables will increase the frequency of conflict. In effect the frequency of roommate conflict decreases with Family size, Number of roommates one has and Being in a love relationship.

The estimated means and standard errors appear quite close with minimum difference between the classical Poisson estimate and MCMC output or posterior summary. Noticeably, the results show a reduction of standard errors associated with the coefficients of age, family size and number of sibling obtained from the Bayesian analysis, thus bringing greater stability to theses coefficients. It's interesting that the results agree so closely, considering fundamentally how different these estimation procedures are.

Table 2 – Posterior Distribution Summaries of parameters from MCMC Poisson regression

Variables	Posterior Means	Standard Error	Quantiles of Posterior Distributions			
			2.5%	25%	75%	97.5%
Intercept	-1.5667	1.0480	-3.6551	-2.257e+00	-0.8746	0.4983
GEN	0.1594	0.1724	-0.1795	4.086e-02	0.2755	0.5011
AGE	0.0253	0.0284	-0.0306	6.571e-03	0.0443	0.0809
FSIZE	-0.2618	0.0990	-0.4491	-3.303e-01	-0.1941	-0.0666
DIFR	0.0336	0.1766	-0.3239	-8.368e-02	0.1522	0.3761
BOD	0.6117	0.2040	0.2255	4.688e-01	0.7471	1.029
NRM	-0.1229	0.0488	-0.2205	-1.556e-01	-0.0887	-0.0303
EDU	0.1104	0.0691	-0.0280	6.426e-02	0.1571	0.2428
LOR	-0.3859	0.1491	-0.6785	-4.840e-01	-0.284	-0.0976
NOS	0.3323	0.0968	0.1413	2.671e-01	0.3989	0.5206
REM	0.0002	0.0002	-0.0003	5.178e-05	0.0004	0.0007

NOTE: GEN= gender, AGE= age, FSIZE= family size, DIFR= Roommates of different religion, BOD= Prior experience in a boarding house, NRM= Number of roommates one has, EDU= years of education, LOR= Being in a love relationship, NOS= Number of Sibling and REM= Remittance.

Bayesian Poisson regression results reveals a positive relationship between frequency of roommate conflicts and the regression covariates (prior experience in a boarding house and number of Sibling) whilst a negative relationship is revealed between frequency of roommate conflicts and the regression covariates (family size, number of roommates one has and being in a love relationship).

Elements such as the quantiles of the parameter posterior distributions and the posterior probability of positive or negative values for each input parameter are of primary interest.

The estimated posterior mean of the effect of prior experience in a boarding house is 0.611 with (0.22, 1.03) at a 95% credible interval. However, the parameter for prior experience in a boarding house is a distribution and further conclusions can be derived. For example, there is a small chance (2.5%) to have an estimate of the very low value of 0.22 or the high value of 1.03 but is more likely that the estimate is 0.468 to 0.747. These observations lead to the conclusion that the prior experience in a boarding house has a positive effect on the frequency of roommate conflicts. This observation agrees with the results of the Poisson regression analysis, where we found that prior experience in a boarding house has a positive and significant effect on the frequency of roommate conflicts.

The posterior distributions of the number of Sibling and its corresponding quantiles given in Table 2 indicates that this parameter is mostly around 0.33 with a 2.5% probability taking a low value of 0.1413 or a high value of 0.5206. Graphically, most of the mass of the posterior distributions of number of Sibling variables are in the positive as illustrated in the plots of their posterior distributions in appendix I. These observations lead to the conclusion that number of Sibling has a positive effect on the frequency of roommate conflicts. This observation agrees with the results of the Poisson regression analysis, where we found that the number of Sibling has a positive and significant effect on the frequency of roommate conflicts.

The estimated posterior mean of the effect of family size is 0.261 with (-0.44, -0.06) at a 95% credible interval. However, the parameter for family size is a distribution and further conclusions can be derived. For example, there is a small chance (2.5%) to have an estimate of the very low value of - 0.44 or the high value of -0.06 but is more likely that the estimate is between -0.33 to -0.19. These observations lead to the conclusion that the family size has a negative effect on the frequency of roommate conflicts. This observation agrees with the results of the Poisson regression analysis, where we found that the family size has a negative and significant effect on the frequency of roommate conflicts.

The posterior distributions of the number of roommates one has and its corresponding quantiles given in Table 2 indicates that this parameter is mostly around -0.123 with a 2.5% probability taking a low value of -0.221 or a high value of -0.03. Graphically, most of the mass of the posterior distributions of number of roommates one has variables are in the negative as illustrated in the plots of their posterior distributions in appendix I. These observations lead to the conclusion that the number of roommates one has a negative effect on the frequency of roommate conflicts. This observation agrees with the results of the Poisson regression analysis, where we found that one's number of roommates has a negative and significant effect on the frequency of roommate conflicts.

The posterior distributions of being in a love relationship and its corresponding quantiles given in Table 1 indicates that the parameters is mostly around -0.386 with a 2.5% probability taking a low value of – 0.679 and a high value of -0.098. Graphically, the mass of the posterior distributions of being in a love relationship variable are distributed in the negative as illustrated in the plots of its posterior distributions in appendix I. These observations lead to the conclusion that being in a love relationship has a negative effect on the frequency of roommate conflicts. This observation agrees with the results of the Poisson regression analysis, where we found that being in a love relationship has a negative and significant effect on the frequency of roommate conflicts.

CONCLUSION

The classical estimation of the Poisson regression model has some important limitations which can be resolved with possible alternative methods. The goal of this study was therefore to introduce Bayesian analysis as an alternative approach and demonstrate its application to parameter estimation of the Poisson regression model in a comparative analysis with the Classical Poisson regression estimation. This study finds that the Bayesian

Markov Chain Monte Carlo algorithm offers an alternative framework for estimating the Poisson regression model.
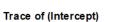
Both the classical Poisson regression and the Bayesian Poisson regression results suggest a positive relationship between frequency of roommate conflicts and the regression covariates (prior experience in a boarding house and number of Sibling) whilst a negative relationship is revealed between frequency of roommate conflicts and the regression covariates (family size, number of roommates one has and being in a love relationship). Furthermore, a comparison of the classical and Bayesian approach to modelling the Poisson regression reveals lower standard errors of some estimated coefficients in the Bayesian approach for the Poisson regression model. Thus the Bayesian Poisson regression is more stable. Notably, the alternative methods lead to similar conclusions. Fundamentally, this study has demonstrated the application of the Bayesian MCMC algorithm to Poisson regression estimation within the context of roommate conflict data.

REFERENCES

1. Acquah, J. G. (2016). Understanding Roommate Conflict among University of Cape Coast Students: A Poisson Regression Approach. Journal of Social and Development Sciences,, 7(1), 73-81.
2. Cameron, A. C., & Trivedi, P. K. (1986). Econometric models based on count data. Comparisons and applications of some estimators and tests. Journal of applied econometrics, 1(1), 29-35.
3. Cameron, A. C., & Trivedi, P. K. (1996). 12 Count data models for financial data. Elsevier, 14(1), 363-391.
4. Gill, J. (2014). Bayesian methods: A social and behavioral sciences approach (3 ed.). Boca Raton, United State of America: CRC press.
5. Jun, S. (2018). Bayesian count data modeling for finding technological sustainability. Sustainability, 10(9), 1-14.
6. Land, K. C., McCall, P. L., & Nagin, D. S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models: With empirical applications to criminal careers data. Sociological Methods & Research, 24(4), 387-442.
7. Winkelmann, R. (1994). "Application to Labor Mobility." Count Data Models. Berlin, Heidelberg: Springer.
8. Winkelmann, R. (2000). Econometric analysis of count data. (3 ed.). Berlin, Heidelberg: Springer .
9. Winkelmann, R. (2008). Econometric analysis of count data. (5 ed.). Berlin, Heidelberg: Springer .
10. Winkelmann, R., & Zimmermann, K. F. (1991). A new approach for modeling economic count data. Economics Letters, 37(2), 139-143.
11. Winkelmann, R., & Zimmermann, K. F. (1995). Recent developments in count data modelling: theory and application. Journal of economic survey, 9(1), 1-24.

APPENDIX I