World Academy of Science, Engineering and Technology
International Journal of Industrial and Manufacturing Engineering
Vol:8, No:7, 2014

Risk Factors for Defective Autoparts Products Using Bayesian Method in Poisson Generalized Linear Mixed Model

Pitsanu Tongkhow, Pichet Jiraprasertwong

Abstract**—**This research investigates risk factors for defective products in autoparts factories. Under a Bayesian framework, a generalized linear mixed model (GLMM) in which the dependent variable, the number of defective products, has a Poisson distribution is adopted. Its performance is compared with the Poisson GLM under a Bayesian framework. The factors considered are production process, machines, and workers. The products coded RT50 are observed. The study found that the Poisson GLMM is more appropriate than the Poisson GLM. For the production Process factor, the highest risk of producing defective products is Process 1, for the Machine factor, the highest risk is Machine 5, and for the Worker factor, the highest risk is Worker 6.

Keywords**—**Defective autoparts products, Bayesian framework, Generalized linear mixed model (GLMM), Risk factors.

I. Introduction

AUTOPARTS industry is one of the largest industries in Thailand. An increasing global demand for automotive vehicles, Thailand has become the favorable place for autopart investment. Fig. 1 shows that there are approximately 1,700 local automotive parts suppliers in Thailand, of which about 700 are Original Equipment Manufacturers [1]. Since all major Japanese automakers have opened manufacturing sites in Thailand, many of their parts manufacturers have relocated here as well to meet their customer demands.

Producing defective products is one of major problems in an autopart factory. Moreover, It defects the production cost and customer satisfaction. This motivated us to find the risk factors in order to minimize the number of the defective products.

Generalized Linear Mixed Model (GLMM) under a Bayesian frame work is adopted, since it allows for different sources of variability in a mean response. The mean of the response contains both a function of some explanatory variables and a function of random variables called random effects. Bayesian method has recently become more attractive, mainly because of recent advances in a computational methodology [2]. A GLMM can be formulated as a hierarchical model under a Bayesian framework.

P. Tongkhow is with the Department of Industrial Engineering, Faculty of Engineering, Rajamangala University of Technology Phra Na Khon, Bangkok, Thailand (phone: 662-9132424 ext. 228; fax: 662-9132424 ext. 228; e-mail: pitsanu.t@rmutp.ac.th).

P. Jiraprasertwong is with the Department of Industrial Engineering Faculty of Engineering, Rajamangala University of Technology Phra Na Khon, Bangkok, Thailand (phone: 662-9132424 ext. 222; fax: 662-9132424 ext. 222; e-mail pichet.J2503@hotmail.com).

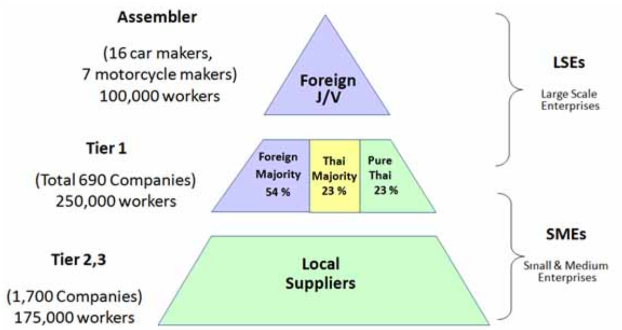

Fig. 1 Structure of Thai automotive intrustry, Source: Thai Autoparts Manufacturers Association

In this paper we apply a Poisson GLMM to the defective product data collected from an autopart factory in Pathum Thani province of Thailand during the period from January, 2012 to July, 2012. We also compare the performance of the Poisson GLMM with the Poisson GLM. The structure of this paper is as follows. In Section II, we define notation for the Poisson GLMMs, Bayesian Method and the application. The result is illustrated in Section III. We discuss and conclude the paper in Section IV and Section V, respectively.

II. Methodology and Application

A. Poisson GLMMs

For $i = 1,...,m$, conditional on $\mathbf{b}_i, y_i$ are independent and assumed to have a Poisson distribution [3]:

$$y_i \mid \mathrm{b_i}, \overset{\text{iid}}{\sim} \mathrm{Pois}(\mu_i)\,, \tag{1}$$

A Poisson GLMM is expressed as:

$$\log(\mu_i) = \eta_i = \mathbf{x}_i^T \boldsymbol{\beta} + \mathbf{z}_i^T \mathbf{b}_i\,, \tag{2}$$

where $\mathrm{E}(y_i \mid \mathbf{b}_i) = \mu_i$, $\boldsymbol{\beta}$ is a p x 1 vector of fixed-effect coefficients related to covariates $\mathbf{x}_i$, $\mathbf{b}_i$ is a q x 1 vector of random-effect coefficients related to covariates $\mathbf{z}_i$, and y_i are observations. Typically, we assume

$$\mathbf{b}_i \overset{iid}{\sim} \mathrm{N}_q(\mathbf{0}, \mathbf{D})\,. \tag{3}$$

B. Bayesian Method

For a vector of data $\mathbf{y}=(y_1,...,y_m)^T$ and a vector of parameters $\boldsymbol{\theta}=(\theta_0,\theta_1,...,\theta_n)^T$, a hierarchical Bayesian model [4] is expressed as:

$$\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{y})=\frac{\mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\mathrm{f}(\mathbf{y})} \quad (4)$$

where $\mathrm{f}(\mathbf{y}\,|\,\boldsymbol{\theta})$ is a likelihood, $\pi(\boldsymbol{\theta}\,|\,\mathbf{y})$ is a posterior distribution which stands for the marginal probability density of the parameter vector $\boldsymbol{\theta}$ given the data $\mathbf{y}$, $\pi(\boldsymbol{\theta})$ is a prior distribution of $\boldsymbol{\theta}$, and $\mathrm{f}(\mathbf{y})$ is the marginal distribution of data $\mathbf{y}$. A hierarchical Bayesian model usually consists of three stages: at the first stage, a linear model is set up given the fixed effects and random effects; at the second stage, distributions of the fixed-effect coefficients and random effects are specified given the hyper parameters; at the last stage, prior distributions are given for the hyper parameters.

The Gibbs sampling MCMC available in Open BUGS is widely used for parameter estimation in a Bayesian method. The MCMC algorithms are the class of algorithms for sampling from probability distributions based on constructing a Markov chain that has the desired distribution as its equilibrium distribution [5]. A set of vectors $\boldsymbol{\theta}$ with density $\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{Y})$ in which the model parameters can be estimated is the final result of the MCMC.

Sampling from the posterior $\mathrm{p}(\boldsymbol{\theta}\,|\,\mathbf{y}), \boldsymbol{\theta}=(\theta_0,\theta_1,...,\theta_n)$ the Gibbs sampler requires a random starting point of parameters of interest, $\boldsymbol{\theta}^{(0)}=\left(\theta_1^{(0)},\theta_2^{(0)},...,\theta_n^{(0)}\right)$

The steps of Gibbs sampling are

1) Generate $\theta_1^{(1)}$ from $\pi(\theta_1\,|\,\theta_2^{(0)},...,\theta_n^{(0)},\mathbf{y})$.
2) Generate $\theta_2^{(1)}$ from $\pi(\theta_2\,|\,\theta_1^{(1)},\theta_3^{(0)},...,\theta_n^{(0)},\mathbf{y})$.
   Use updated value of $\theta_1^{(1)}$.
3) Generate $\theta_3^{(1)}$ from $\pi(\theta_3\,|\,\theta_1^{(1)},\theta_2^{(1)},...,\theta_n^{(0)},\mathbf{y})$.
   Use updated value of $\theta_1^{(1)}$ and $\theta_2^{(1)}$.
4) Generate $\theta_4^{(1)},...,\theta_n^{(1)}$ similarly to step 1 to 3
5) Generate $\boldsymbol{\theta}^{(2)}$ using $\boldsymbol{\theta}^{(1)}$ as a starting point and continually using the most updated values.
6) Repeat until we get M samples, with each sample being a vector of $\boldsymbol{\theta}^{(1)},\boldsymbol{\theta}^{(2)},...,\boldsymbol{\theta}^{(M)}$, where M is the number of samples .
7) Monte Carlo integration on those draws to the quantity of interest can be done. For example, the mean of θ_0 results from:

$$\mathrm{E}(\theta_0)=\frac{1}{M}\sum_{i=1}^{M}\theta_0^{(i)} \quad (5)$$

C. Application to Autoparts Defective Products

Data: The defective products data were collected from an autopart factory in Pathum Thani province of Thailand during the period from January, 2012 to July, 2012 (189 times). The products coded RT50 are observed. The considered factors related to producing defective products are 4 production Processes, 7 Machines, and 11 Workers.

Model: For each product, let y_i, $i=1,...,189$ denote the number of defective product observed at time i The proposed model is:

$$\log(\mu_i) = \log(Total_i) + \beta_0 + \sum_{j=1}^{3}\beta_j\, process_j + \sum_{j=4}^{9}\beta_k\, machine_{j-2} + \sum_{j=10}^{19}\beta_j\, worker_{j-8} + b_i\ , \quad (6)$$

where the "process", "machine", and "worker" are dummy variables. The "total" is the total products at each time which is treated as an offset. Note that we use only the random intercept, i.e. $\mathbf{z}_i^T\mathbf{b}_i=b_i$. We assume N(0,1.0E06) for $\beta_0,...,\beta_{19}$. We also assume $\mathrm{N}(0,\tau_b^2)$ for b_i and IG(0.1,0.001) for τ_b^2 [6].

The proposed method is compared with the Poison GLM [7], [8] without the random intercept:

$$\log(\mu_i) = \log(Total_i) + \beta_0 + \sum_{j=1}^{3}\beta_j\, process_j + \sum_{j=4}^{9}\beta_k\, machine_{j-2} + \sum_{j=10}^{19}\beta_j\, worker_{j-8}\ . \quad (7)$$

The deviance information criterion (DIC) is used for model comparison. The model with smaller DIC is better than the model with larger DIC [9]. The Gibbs sampling MCMC is run in Open BUGS.

III. Result

The Gibbs sampling MCMC is converged when being run for 25,000 iterations and the initial 5,000 iterations are discarded as burn-in. The history plots of some parameters, as an example, in Fig. 2 show no trends, and snake around the mean. As well as their Kenel density plots in Fig. 3 do not show a multi-modal curve. The summary of the parameter estimate from the Poisson GLMM are presented in Table I, compared with the one from the Poisson GLM in Table II. The DIC of the proposed model is 1,127 which is smaller the one from the Poisson GLM (2,898), indicating that the Poisson GLMM is more appropriate than the Poisson GLM.

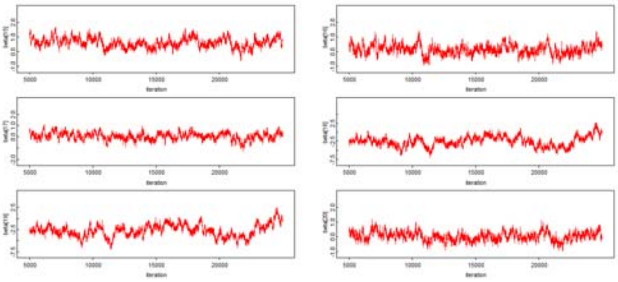

Fig. 2 History plots indicating MCMC convergence

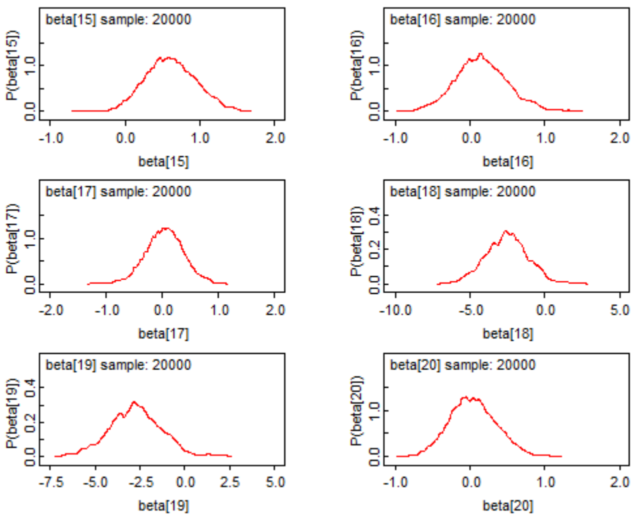

Fig. 3 Kernel density plots indicating MCMC convergence

The parameter estimate from the proposed model in Table I indicates that, for the production process factor, the rate of producing defective products in production Process 2 is 1.0840 times of the rate in Process 1, the rate in Process 2 is 18.02% more than the rate in Process 1, but the rate in Process 4 is 92.87% less than the rate in Process 1. For the Machine factor, the rates of producing defective products from Machine 2 and 5 are 6.97% and 76.65% more than the rate from Machine 1, respectively, and the rates from Machine 3, 4, 6 , 7 are 43.21%, 32.4%, 16.95%, 28.87% more than the rate from Machine 1, respectively. For the Worker factor, the rates of producing defective products from Worker 2, 4, 5, 6, 7,8,11 are 28.43%, 27.10%, 30.40%, 84.67%, 16.80%, 4.17%, 3.99% more than the rate from Worker1, respectively, and the rates from Worker 3, 9, 10 are 89.89%, 91.98%, 93.54% less than the rate form Worker 1.

IV. Discussion

A hierarchical Bayesian method via MCMC algorithms allows more complicated models that classical methods based on maximum likelihood are unable to estimate. It use DIC to compare models with different methods including hierarchical models, where classical model fit statistics cannot compare different methods or hierarchical models. A GLMM is more appropriate than the GLM because it can account for the subject-specific random effects which naturally happen to individuals. The highest risk factors of producing defective products in each type of factors are Process 1, Machine 5, and Worker 6; therefore, those factors should be closely studied in to solve the problems. The Poisson GLMM can be applied to other kinds of problems in which the observations are count data.

TABLE I
PARAMETER ESTIMATE FROM THE POISSON GLMM WITH DIC = 1,127

Factor	Mean	Standard Deviation	MC Error	95% Credible Interval		Relative Rate
Intercept	-4.277	0.729	0.061	-5.875	-3.098	.
Process1	Ref.	.	.	.	.	.
Process2	0.734	0.458	0.038	0.000	1.747	2.084
Process3	0.166	0.491	0.040	-0.654	1.301	1.180
Process4	-2.641	0.421	0.028	-3.420	-1.753	0.071
Machine1	Ref.	.	.	.	.	.
Machine2	0.067	0.697	0.047	-1.355	1.402	1.070
Machine3	-0.566	0.626	0.047	-1.683	0.716	0.568
Machine4	-0.392	0.457	0.037	-1.276	0.414	0.676
Machine5	0.569	1.050	0.081	-1.506	2.800	1.767
Machine6	-0.186	0.507	0.039	-1.157	0.797	0.831
Machine7	-0.341	0.444	0.036	-1.231	0.464	0.711
Worker1	Ref.	.	.	.	.	.
Worker2	0.250	0.378	0.023	-0.471	0.986	1.284
Worker3	-2.292	1.310	0.103	-4.942	0.319	0.101
Worker4	0.240	0.306	0.020	-0.382	0.851	1.271
Worker5	0.265	0.351	0.023	-0.498	0.900	1.304
Worker6	0.613	0.336	0.024	-0.014	1.281	1.847
Worker7	0.155	0.347	0.023	-0.510	0.877	1.168
Worker8	0.041	0.331	0.022	-0.622	0.697	1.042
Worker9	-2.523	1.415	0.111	-5.267	0.227	0.080
Worker10	-2.739	1.414	0.111	-5.536	-0.020	0.065
Worker11	0.039	0.307	0.021	-0.538	0.660	1.040

V. Conclusion

Risk factors for defective products in autoparts factories are investigated using a Poisson GLMM under a Bayesian framework. The performance of the proposed model is compared with the Poisson GLM. The factors, production Process, Machines, and Workers are considered. The products coded RT50 are observed. We found that the Poisson GLMM has a better performance than the Poisson GLM. The highest risk factors of producing defective products in each type of factors are Process 1, Machine 5, and Worker 6.

TABLE II
PARAMETER ESTIMATE FROM THE POISSON GLM WITH DIC = 2,898

Factor	Mean	Standard Deviation	MC Error	95% Credible Interval	
Intercept	-3.680	0.210	0.016	-4.103	-3.255
Process1	Ref.	.	.	.	.
Process2	-0.205	0.191	0.014	-0.590	0.165
Process3	-0.762	0.194	0.014	-1.148	-0.379
Process4	-2.991	0.156	0.009	-3.305	-2.689
Machine1	Ref.	.	.	.	.
Machine2	0.522	0.129	0.006	0.268	0.771
Machine3	-0.485	0.200	0.014	-0.890	-0.099
Machine4	0.107	0.099	0.005	-0.087	0.296
Machine5	0.588	0.318	0.014	-0.010	1.219
Machine6	-0.142	0.109	0.006	-0.356	0.069
Machine7	0.057	0.098	0.005	-0.133	0.246
Worker1	Ref.	.	.	.	.
Worker2	0.293	0.078	0.002	0.142	0.446
Worker3	-2.853	0.369	0.021	-3.585	-2.161
Worker4	0.168	0.065	0.002	0.040	0.294
Worker5	0.517	0.071	0.002	0.381	0.657
Worker6	0.455	0.066	0.002	0.327	0.586
Worker7	0.182	0.069	0.002	0.049	0.321
Worker8	0.081	0.067	0.002	-0.049	0.214
Worker9	-2.949	0.403	0.022	-3.744	-2.182
Worker10	-2.935	0.398	0.022	-3.716	-2.176
Worker11	0.197	0.066	0.002	0.069	0.327

ACKNOWLEDGMENT

Authors gratefully acknowledge the Rajamangala University of Technology Phra Nakhon and the faculty of Engineering for their technical and financial support.

REFERENCES

[1] BOI. (2014, Jan 16). “Thailand: the detroit of asia”. Available: http://www.boi.go.th/index.php?page=opp_automotive.

[2] B. L. Bazán , *Bayesian Modeling User Manual WinBUGS* , Department of Sciences: Pontifical Catholic University of Peru, 2011.

[3] P. McCullagh, J. Nelder. *Generalized Linear Models*, 2nd ed., Boca Raton: Chapman and Hall/CRC, 2008.

[4] P. Congdon, *Bayesian Statistical Modelling*, 2nd ed., John Wiley & Sons: New York, 2006.

[5] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, 2nd ed., Springer: New York, 2004.

[6] A. Thomas, N. Best, D. Lunn, R. Arnold, and D. Spiegelhalter, “GeoBUGS User Manual”, London: MRC Biostatistics Unit, Institute of PublicHealth, 2007.

[7] P. McCullagh, and J. Nelder, *Generalized Linear Models*, 2nd ed. Boca Raton: Chapman and Hall/CRC, 1989.

[8] CH. Adithep, B. Wasan, and T. Pitsanu, “Risk Factors for Defective Products in Autoparts Factory Using Generalized Linear Model (GLM),” *Nation Conference 9th*, Kamphaengsaen campus, Kasetsart university, Thailand, 2012, pp. 238-246.

[9] WinBUGS. (2014, Jan 9). “The BUGS Project: DIC. Cambridge: MRC Biostatistics Unit”. Available: URL: http://dvorak.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml.