International Journal of Engineering, Mathematical and Physical Sciences
ISSN: 2517-9934
Vol:4, No:1, 2010

A comparison of marginal and joint generalized quasi-likelihood estimating equations based on the Com-Poisson GLM: Application to car breakdowns data

N. Mamode Khan and V. Jowaheer

Abstract—In this paper, we apply and compare two generalized estimating equation approaches to the analysis of car breakdowns data in Mauritius. Number of breakdowns experienced by a machinery is a highly under-dispersed count random variable and its value can be attributed to the factors related to the mechanical input and output of that machinery. Analyzing such under-dispersed count observation as a function of the explanatory factors has been a challenging problem. In this paper, we aim at estimating the effects of various factors on the number of breakdowns experienced by a passenger car based on a study performed in Mauritius over a year. We remark that the number of passenger car breakdowns is highly under-dispersed. These data are therefore modelled and analyzed using Com-Poisson regression model. We use the two types of quasi-likelihood estimation approaches to estimate the parameters of the model: marginal and joint generalized quasi-likelihood estimating equation approaches. Under-dispersion parameter is estimated to be around 2.14 justifying the appropriateness of Com-Poisson distribution in modelling under-dispersed count responses recorded in this study.

Keywords—Breakdowns, Under-dispersion, Com-Poisson, Generalized Linear Model, marginal quasi-likelihood estimation, joint quasi-likelihood estimation.

I. Introduction

With an increase in the number of cars on the roads of Mauritius, the number of accidents resulting from and resulting into car breakdowns have also increased. The local road traffic branch authority [6] has reported that 34 percent of vehicles that contribute to serious accident crashes are cars. It has been remarked that such cars usually consist of defective auto-parts, defective door latches, seat belts, roofs, ignition systems and fuel systems. According to the RAC patrols report [7] in UK, many car owners do not have time to check their cars regularly, do not notice anything about the state of their cars and do not understand the specificities of their cars. In fact, one of the causes of breakdown is that many car owners do not understand the system of their cars. For example, car may undergo a breakdown if the tyres' pressure are not checked at least weekly. It is important to make sure that a tyre must have at least 1.6 mm of tread in a continuous band at least as wide as three-quarters of the central tread breadth which is important in wet or icy weather. The engine oil and the coolant levels are also important determinants for the smooth running of the vehicles. The improper engine high-tension leads may yet be another important cause of breakdown. Obviously, these factors will contribute to car breakdowns. Moreover, we remark that car breakdown causes a lot of traffic jam on the roads of Mauritius especially on the motor way and during peak hours. Breakdown of cars is thus a serious issue. In this paper, we analyze car breakdowns data that has been collected in the year 2008. The organization of the paper is as follows: In section 2, we describe the main factors leading to breakdown of cars in Mauritius in 2008. In section 3, we present the Com-Poisson regression model that will be used to analyze the data. In the last section, we provide the results and conclusions.

N.Mamode Khan is is a PhD student in the Department of Mathematics, Faculty of Science, University of Mauritius(e-mail: alimamode@yahoo.co.uk)
V. Jowaheer is is an Associate Professor in the Department of Mathematics, Faculty of Science, University of Mauritius(e-mail: vandnaj@uom.ac.mu)

II. Causes of breakdowns

Very often, the age of the car is a factor leading to major and minor breakdowns. In fact, we note that most of the cars which get frequent breakdowns range between 9 and 15 years. As a car grows older, it is more prone to breakdowns despite the fact that it might be well taken care of. The parts of the car become worn out and due to the age of the car, its spare parts are unavailable on the local market. As a result, the car users have to resort to alternatives or substitutes. Together with age, the mileage of the car can be a contributing factor in breakdowns if regular servicing and checks are not carried out. Mileage can be used as an indicator for servicing and checks, failing which, breakdowns may occur. Previous breakdowns may also be a factor leading to future and repetitive breakdowns. It has been found that a car which has been through an accident is more prone to breakdowns as the mechanism can be affected if not properly repaired or replaced,. However, this may depend on the make of the car. A recent event is the introduction of cars that use gas in Mauritius and some mechanics are encountering problems to repair the filtering part of such cars. It is also remarked that sometimes the engine system of the car and ultimately the horsepower is purposely changed but whether the new engine can be adjusted to the system of the car is questionable. In general the age, the mileage, the number of accidents that the car has made and the fuel or gas consumption of the car are the main factors that may influence car breakdowns in Mauritius. We have interviewed 15,000 randomly selected car owners and collected data on the number of breakdowns their cars have suffered during the year 2008 along with the information on the following explanatory variables: the age of the car, the mileage of the car, the number of accidents that the car has encountered, the number of times the car visit the mechanic and the fuel or gas consumption (1-petrol and 0-gas). Our objective is to assess the effect of these factors on the number of breakdowns. In fact, the mean number of breakdowns recorded during 2008 is 1.2333 while the variance is 0.3455. This indicates that the data is under-dispersed. To model under-dispersed data under a regression set-up, Jowaheer and Mamode Khan [2] have developed a Com-Poisson regression model. In the next section, we give a description of this model and provide the estimating equations to estimate the regression and under-dispersion parameter

III. Com-Poisson Regression model

Recently, Shmueli et al. [5] proposed the Conway Maxwell Poisson (Com Poisson) distribution to model counts which may be equi-, over- and under- dispersed. Kadane et al. [3] and Shmueli et al. [5] studied the basic properties of this distribution and the fitting of this distribution to over -and under - dispersed cross sectional count data. In regression set-up, Com Poisson generalized linear model (GLM) have been designed by Guikema [1] and Jowaheer and Mamode Khan [2] to study the effect of covariates on the under- and over-dispersed count responses. Guikema [1] has used Bayesian techniques whereas Jowaheer and Mamode Khan [2] have developed a joint quasi-likelihood technique (JGQL) to estimate the parameters of the model. The JGQL approach provides consistent and equally efficient estimates as the maximum likelihood approach. The Com Poisson regression model is given by:

$$f(y_i) = \frac{\lambda_i^{y_i}}{(y_i!)^{\nu}} \frac{1}{Z(\lambda_i, \nu)}, \tag{1}$$

where y_i is the number of breakdowns corresponding to the car of the i^{th} individual and X_i is the vector of covariates corresponding to y_i. By letting β be the vector of regression parameters such that β_j is the regression effect of the j^{th} covariate on the breakdowns, we write

$$\lambda_i = \exp(x_i^T \beta) \tag{2}$$

In equation (1), the parameter ν corresponds the dispersion index. More specifically, the values $\nu > 1$ correspond to equi-, over- and under- dispersion. Since equation (1) does not have closed form expressions, we use an asymptotic expression for $Z(\lambda_i, \nu)$ proposed by Shmueli et al.[5] given by

$$Z(\lambda_i, \nu) \simeq \frac{\exp(\nu \lambda_i^{\frac{1}{\nu}})}{\lambda_i^{\frac{\nu-1}{2\nu}} (2\pi)^{\frac{\nu-1}{2}} \sqrt{\nu}} \tag{3}$$

and reformulate the equation (1) as

$$f(y_i) = \frac{\exp(x_i^T \beta y_i)[\exp(x_i^T \beta (\frac{\nu-1}{2\nu}))(2\pi)^{\frac{\nu-1}{2\nu}} \sqrt{\nu}]}{(y_i!)^{\nu} [\exp(\nu \exp(\frac{x_i^T \beta}{\nu}))]} \tag{4}$$

From equation (4),

$$E(Y_i) = \theta_i = \lambda_i^{1/\nu} - \frac{\nu - 1}{2\nu} \tag{5}$$

and

$$Var(Y_i) = \frac{\lambda_i^{1/\nu}}{\nu} \tag{6}$$

IV. Estimation of the regression and under-dispersion parameters

To estimate the parameters β and ν, we consider two estimation approaches:

1) Marginal generalized quasi-likelihood estimating equation approach (MGQL)
2) Joint generalized quasi-likelihood estimating equation approach (JGQL)

A. Marginal GQL

In this section, we develop two marginal QLEs under Com-Poisson regression model. The first QLE is to estimate the vector of regression parameters β based on observations y_i while the second QLE is to estimate the dispersion index ν. The QLE to estimate β is given by

$$\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}^{-1} (y_i - \theta_i) = 0, \tag{7}$$

where

$$V_{i,\beta} = \frac{\lambda_i^{1/\nu}}{\nu}. \tag{8}$$

and

$$D_{i,\beta} = \frac{\partial \theta_i}{\partial \beta^T} = \frac{\lambda_i^{\frac{1}{\nu}}}{\nu} x_i^T \tag{9}$$

is a $p \times 1$ matrix. The QLE to estimate ν is given by

$$\sum_{i=1}^{I} D_{i,\alpha}^T V_{i,\nu}^{-1} (y_i^2 - \eta_i) = 0, \tag{10}$$

where

$$\eta_i = E(Y_i^2) = \frac{\lambda_i^{1/\nu}}{\nu} + [\lambda_i^{1/\nu} - \frac{\nu - 1}{2\nu}]^2 \tag{11}$$

and

$$D_{i,\nu} = \frac{1}{2\nu^3}[2\lambda_i^{\frac{1}{\nu}} \nu \ln(\lambda_i) + \nu - 1 - 4\lambda_i^{\frac{2}{\nu}} \ln(\lambda_i)\nu]$$
$$+ \frac{1}{2\nu^3}[-4\lambda_i^{\frac{1}{\nu}} \nu - 4\lambda_i^{\frac{1}{\nu}} \ln(\lambda_i)] \tag{12}$$

.
$V_{i,\nu}$ is the variance of Y_i^2 and is calculated using

$$V_{i,\nu} = E(Y_i^4) - E(Y_i^2)^2 \tag{13}$$

where the moments are derived iteratively from the moment generating function of y_i which is given by

$$E[Y_i^{r+1}] = \lambda \frac{d}{d\lambda} E[Y^r] + E[Y]E[Y^r] \tag{14}$$

and hence,

$$E(Y_i^4) = \frac{\lambda_i^{\frac{1}{\nu}} \nu^2 + 4\lambda_i^{\frac{3}{\nu}} \nu^2 + 10\lambda_i^{\frac{2}{\nu}} \nu - 4\lambda_i^{\frac{1}{\nu}} \nu + 4\lambda_i^{\frac{1}{\nu}} - 4\lambda_i^{\frac{2}{\nu}} \nu^2}{\nu^3}$$
$$+ [\frac{\lambda_i^{1/\nu}}{\nu} [\lambda_i^{1/\nu} - \frac{\nu - 1}{2\nu}]^2]^2 \tag{15}$$

following equation (14). The Newton-Raphson technique is then applied to the two estimating equations. The iterative equations are given as follows: At the rth iteration,

$$\left(\hat{\beta}_{r+1}\right) = \left(\hat{\beta}_r\right) + [\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}{}^{-1} D_{i,\beta}]_r^{-1} [\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}{}^{-1}(y_i - \theta_i)] \quad (16)$$

$$\left(\hat{\nu}_{r+1}\right) = \left(\hat{\nu}_r\right) + [\sum_{i=1}^{I} D_{i,\nu}^T V_{i,\nu}{}^{-1} D_{i,\nu}]_r^{-1} [\sum_{i=1}^{I} D_{i,\nu}^T V_{i,\nu}{}^{-1}(y_i^2 - \eta_i)] \quad (17)$$

where $\hat{\beta}_r$ and $\hat{\nu}_r$ are the values of $\hat{\beta}$ and $\hat{\nu}_r$ at the r^{th} iteration. $[.]_r$ is the value of the expression at the r^{th} iteration. The estimators are consistent and under mild regularity conditions, for $I \to \infty$, it may be shown that $I^{\frac{1}{2}}((\hat{\beta}) - (\beta))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I[\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}{}^{-1} D_{i,\beta}]^{-1} [\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}{}^{-1}(y_i - \theta_i)(y_i - \theta_i)^T V_{i,\beta}{}^{-1} D_{i,\beta}][\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}{}^{-1} D_{i,\beta}]^{-1}$ and $I^{\frac{1}{2}}((\hat{\nu}) - (\nu))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I[\sum_{i=1}^{I} D_{i,\nu}^T V_{i,\nu}{}^{-1} D_{i,\nu}]^{-1} [\sum_{i=1}^{I} D_{i,\nu}^T V_{i,\nu}{}^{-1}(y_i^2 - \eta_i)(y_i^2 - \eta_i)^T V_{i,\nu}{}^{-1} D_{i,\nu}][\sum_{i=1}^{I} D_{i,\nu}^T V_{i,\nu}{}^{-1} D_{i,\nu}]^{-1}$ The algorithm to estimate the parameters works as follows: For an initial estimate of β and ν, we iterate equation (16) until convergence, then use the updated β to update ν in equation (17). We then replace the updated β and ν in equation (16) and iterate until convergence. Having obtained the new β, we replace in equation (17) to obtain a new ν and the cycle continues until both values converge.

B. Joint GQL

In this section, we solve the joint quasi-likelihood equation given by

$$\sum_{i=1}^{I} D_i^T V_i^{-1}(f_i - \mu_i) = 0, \quad (18)$$

where $f_i = (y_i, y_i^2)^T, \mu_i = E(f_i), V_i = cov(f_i), D_i = \frac{\partial E(f_i)}{\partial(\beta^T,\nu)}$. The components of equation (18) are derived by Jowaheer and Mamode Khan [2]. For convenience, we reproduce these formulae in the appendix. The QL estimates of β and ν are obtained by solving equation (18) iteratively until convergence using Newton-Raphson technique. At rth iteration,

$$\begin{pmatrix} \hat{\beta}_{r+1} \\ \hat{\nu}_{r+1} \end{pmatrix} = \begin{pmatrix} \hat{\beta}_r \\ \hat{\nu}_r \end{pmatrix} + [\sum_{i=1}^{I} D_i^T V_i^{-1} D_i]_r^{-1} [\sum_{i=1}^{I} D_i^T V_i^{-1}(f_i - \mu_i)]_r \quad (19)$$

where $\hat{\beta}_r$ is the value of $\hat{\beta}$ at the r^{th} iteration. $[.]_r$ is the value of the expression at the r^{th} iteration. The estimators are consistent and under mild regularity conditions, for $I \to \infty$, it may be shown that $I^{\frac{1}{2}}((\hat{\beta}, \hat{\nu}) - (\beta, \nu))^T$ has an asymptotic normal distribution with mean 0 and covariance matrix $I[\sum_{i=1}^{I} D_i^T V_i^{-1} D_i]^{-1} [\sum_{i=1}^{I} D_i^T V_i^{-1}(f_i - \mu_i)(f_i - \mu_i)^T V_i^{-1} D_i][\sum_{i=1}^{I} D_i^T V_i^{-1} D_i]^{-1}$ The algorithm works in the same way as in section IV A.

V. RESULTS

The application of Com-Poisson GLM discussed in the above section provides the following estimates

TABLE I
ESTIMATES OF THE REGRESSION AND UNDER-DISPERSION PARAMETERS BASED UNDER JGQL AND MGQL FOR THE CAR BREAKDOWN DATA YEAR 2008

Estimates	JGQL		MGQL		ER
Intercept	0.4732	(0.3425)	0.4692	(0.3487)	98
Age	2.2284	(0.0342)	2.2394	(0.0352)	97.1
Mileage	0.3541	(0.0222)	0.3301	(0.0228)	97.3
Accident	0.2666	(0.0886)	0.2592	(0.0893)	99.2
Visit	-0.2512	(0.0623)	-0.2489	(0.0625)	99
Consumption	-0.2230	(0.1344)	-0.2197	(0.1350)	99
$\hat{\nu}$	2.4101	(0.1220)	2.4221	(0.1227)	99

VI. CONCLUSION

These results are obtained by taking small initial values of the regression and under-dispersion parameters. The entry in brackets represent the standard errors of each estimate. ER represents the efficiency ratio percentage of the JGQL approach with respect to the MGQL approach. The age, mileage and accident covariates are all positive. This is reasonable because as age and mileage increase, we expect the cars to be frequently getting breakdowns. We note that age is the covariate that contributes to car breakdowns to a big extent. Moreover, the positive sign in the accident covariate indicates that as the number of accidents that the car encounters increases, the number of breakdowns of the car will also increase. This is obvious because upon meeting an accident, the mechanical parts of the car are prone to damage either in short term or long term. Further, if the car owner attends his regular car servicing or have its car regularly checked, the number of breakdowns will eventually decrease. The estimate of the consumption of the fuel by vehicle is negative meaning that the number of breakdowns is lesser among cars using petrol than gas. This may be because, in Mauritius, the local mechanics are not quite familiar with the recent development in car technology. Therefore, those people who have preferred gas as an alternative to petrol should have to change the filtering part of their cars and in many cases, the mechanics have not done these changes correctly. The estimate of the under-dispersion parameter affirms that the data is under-dispersed and dispersion parameter cannot be ignored. We note that there is no huge difference in the estimates of the regression and under-dispersion parameters between MGQL and JGQL but JGQL provides slightly more efficient estimates than MGQL,i.e, the standard errors are relative lower. Moreover, the efficiency ratio of the two methods are close to 100 percent. Nevertheless, both techniques can be used for estimation under the Com-Poisson GLM.

Appendix A
Components of Joint GQL equation (18)

$$D_i = \begin{pmatrix} \partial\theta_i/\partial\beta^T & \partial\theta_i/\partial\nu \\ \partial m_i/\partial\beta^T & \partial m_i/\partial\nu \end{pmatrix}$$

where

$$\partial\theta_i/\partial\beta^T = \frac{\lambda_i^{\frac{1}{\nu}}}{\nu} x_i^T \tag{20}$$

$$\partial\theta_i/\partial\nu = \frac{1}{2}\frac{\nu-1}{2\nu^2} - \frac{1}{2\nu} - \frac{\lambda_i^{\frac{1}{\nu}} x_i^T \beta}{\nu^2} \tag{21}$$

$$\partial m_i/\partial\beta^T = x_i^T (\frac{2\lambda_i^{\frac{1}{\nu}} + 2\nu\lambda_i^{\frac{2}{\nu}} - \nu\lambda_i^{\frac{1}{\nu}}}{\nu^2}) \tag{22}$$

$$\partial m_i/\partial\nu = \frac{1}{2\nu^3}[2\lambda_i^{\frac{1}{\nu}}\nu\ln(\lambda_i) + \nu - 1 - 4\lambda_i^{\frac{2}{\nu}}\ln(\lambda_i)\nu - 4\lambda_i^{\frac{1}{\nu}}\nu$$

$$-4\lambda_i^{\frac{1}{\nu}}\ln(\lambda_i)]. \tag{23}$$

The covariance matrix of f_i is expressed as

$$V_i = \begin{pmatrix} var(Y_i) & cov(Y_i, Y_i^2) \\ & var(Y_i^2) \end{pmatrix}$$

The elements in V_i are derived iteratively from the moment generating function of y_{it} which is given by

$$E[Y_i^{r+1}] = \lambda\frac{d}{d\lambda}E[Y^r] + E[Y]E[Y^r] \tag{24}$$

By deriving the moments for y_i^2, y_i^3 and y_i^4, we obtain

$$cov(Y_i, Y_i^2) = E(Y_i^3) - E(Y_i)E(Y_i^2) \tag{25}$$

$$= \frac{2\lambda_i^{\frac{1}{\nu}} + 2\nu\lambda_i^{\frac{2}{\nu}} - \nu\lambda_i^{\frac{1}{\nu}}}{\nu^2}$$

$$Var(Y_i^2) = E(Y_i^4) - E(Y_i^2)^2$$

$$= \frac{\lambda_i^{\frac{1}{\nu}}\nu^2 + 4\lambda_i^{\frac{3}{\nu}}\nu^2 + 10\lambda_i^{\frac{2}{\nu}}\nu - 4\lambda_i^{\frac{1}{\nu}}\nu + 4\lambda_i^{\frac{1}{\nu}} - 4\lambda_i^{\frac{2}{\nu}}\nu^2}{\nu^3} \tag{26}$$

Acknowledgment

The authors would like to thank the Road Traffic Branch of Mauritius, the mechanics working at the Bala Garage and Tapsee motors for providing the relevant information and Mrs Maleika-Heenaye Mamode Khan for helping in the collection of data.

References

[1] Guikema, S. " Formulating informative data -based priors for failure probability estimation in reliability analysis", Reliability Engineering and System Safety, Vol 92, 490-502,2007.

[2] Jowaheer,V and Mamode Khan,N. " Estimating Regression Effects in Com-Poisson Generalized Linear Model", International Journal of Computational and Mathematical Sciences, Vol 3:4, 2009

[3] Kadane,J , Shmueli,G, Minka,G, Borle,T and Boatwright,P. " Conjugate analysis of the Conway Maxwell Poisson distribution", Bayesian analysis, Vol 1, 363-374,2006.

[4] Lord,D, Guikema,S and Geedipally,S. "Application of the Conway-Maxwell-Poisson Generalized Linear Model for Analyzing Motor Vehicle Crashes". Accident Analysis and Prevention, Vol. 40, 1123-1134,2008

[5] Shmueli,G, Minka,T, Borle,J and Boatwright,P. " A useful distribution for fitting discrete data, Journal of Royal Statistical Society, 2005.

[6] Road accidents in Mauritius: statistics and analysis, Ministry of Public Infrastructure, Land Transport and Shipping Traffic management and Road safety unit, July 2006.

[7] RAC patrol report: motoring organisation, United Kingdom, 2008