Analyzing the factors effecting the passenger car break downs using Com-Poisson GLM

N. Mamode Khan and V. Jowaheer

Abstract—Number of break downs 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 observations 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 break downs experienced by a passenger car based on a study performed in Mauritius over a year. We remark that the number of passenger car break downs is highly under-dispersed. These data are therefore modelled and analyzed using Com-Poisson regression model. We use quasi-likelihood estimation approach to estimate the parameters of the model. Under-dispersion parameter is estimated to be 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, 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 break downs have also increased. The local road traffic branch authority [6] has reported in 2004 that 34 percent of vehicles that contribute to serious accident crashes are cars. In our survey, we have found that such cars usually consist of defective auto-parts, defective door latches, seat belts, roofs, ignition systems and fuel systems. Many car owners do not have time to check their cars in the morning and do not notice anything about the state of their cars. According to the RAC patrols report [7] in UK, there are many car owners who do not understand the specificities of their cars. Obviously, these factors will contribute to car break downs. Moreover, we remark that breakdown cars cause 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 break downs data that has been collected in the year 2008. The organization of the paper is as follows: In section 2, we describe the 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.

II. CAUSES OF BREAKDOWNS

Very often, the age of the car is a factor leading to major and minor break downs. In fact, we note that most of the cars which get frequent break downs range between 9 and 15 years. As a car grows older, it is more prone to break downs 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 break downs if regular servicing and checks are not carried out. Mileage can be used as an indicator for servicing and checks, failing which, break downs may occur. Previous break downs may also be a factor leading to future and repetitive break downs. It has been found that a car which has been through an accident is more prone to break downs as the mechanism can be affected if not properly repaired or replaced. However, this depends 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 break downs in Mauritius. We have interviewed 1500 randomly selected car owners and collected data on the number of break downs their cars have suffered during the year 2008 along with the information on the following explanatory variables: the age of the car, the average 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 denoted (1-petrol and 0-gas). Our objective is to assess the effect of these factors on the number of break downs. In fact, the mean number of break downs 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, Guikema [1] and Jowaheer and Mamode Khan [2] developed a Com Poisson generalized linear model (GLM).
for the efficient estimation of the parameters of this model, Jowaheer and Mamode Khan [2] have developed a joint quasi-likelihood technique (JGQL). Their technique provides consistent and equally efficient estimates as the maximum likelihood approach. In this section, we use this approach to analyze the breakdown data. The Com Poisson regression model is given by:

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

where \( y_i \) is the number of breakdowns corresponding to the car of the \( i \)th individual and \( \lambda_i \) is the vector of covariates corresponding to \( y_i \). By letting \( \beta \) be the vector of regression parameters such that \( \beta_j \) is the regression effect of the \( j \)th covariate on the breakdowns, we write

\[ \lambda_i = \exp(x_i^T \beta) \]

and reformulate the equation (1) as

\[ f(y_i) = \frac{[\exp(x_i^T \beta)]^{y_i} \exp[(x_i^T \beta)\nu - 1]}{(y_i!)^\nu \exp[\nu \exp(x_i^T \beta)]} \]

From equation (4),

\[ E(Y_i) = \theta_i = \lambda_i^{1/\nu} - \nu - \frac{1}{2\nu} \]

and

\[ Var(Y_i) = \frac{\lambda_i^{1/\nu}}{\nu} \]

To estimate the parameters \( \beta \) and \( \nu \), we solve the joint quasi-likelihood equation given by

\[ \sum_{i=1}^{I} D_i^T V_i^{-1} D_i = 0, \]

where \( f_i = \{y_i, y_i^2\}^T, \mu_i = E(f_i), V_i = cov(f_i), D_i = \frac{\partial E(f_i)}{\partial \nu} \). The components of equation (7) are derived by Jowaheer and Mamode Khan [2]. For convenience, we reproduce these formulae in the appendix. The QL estimates of \( \beta \) and \( \nu \) are obtained by solving equation (7) iteratively until convergence using Newton-Raphson technique. At \( r \)th iteration,

\[ (\beta_{r+1}, \nu_{r+1}) = (\beta_r, \nu_r) + [\sum_{i=1}^{I} D_i^T V_i^{-1} D_i]^{-1} [\sum_{i=1}^{I} D_i^T V_i^{-1} (f_i - \mu_i)], \]

where \( \beta_r \) is the value of \( \beta \) at the \( r \)th iteration. \([\cdot]\), 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 \( I^2 ((\beta, \nu) - (\beta, \nu))^{T} \) has an asymptotic normal distribution with mean 0 and covariance matrix \( I^2 \sum_{i=1}^{I} D_i^T V_i^{-1} D_i \). The components of equation (7) are derived by Jowaheer and Mamode Khan [2]. For convenience, we reproduce these formulae in the appendix. The QL estimates of \( \beta \) and \( \nu \) are obtained by solving equation (7) iteratively until convergence using Newton-Raphson technique. At \( r \)th iteration,

\[ (\beta_{r+1}, \nu_{r+1}) = (\beta_r, \nu_r) + [\sum_{i=1}^{I} D_i^T V_i^{-1} D_i]^{-1} [\sum_{i=1}^{I} D_i^T V_i^{-1} (f_i - \mu_i)], \]

where \( \beta_r \) is the value of \( \beta \) at the \( r \)th iteration. \([\cdot]\), 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 \( I^2 ((\beta, \nu) - (\beta, \nu))^{T} \) has an asymptotic normal distribution with mean 0 and covariance matrix \( I^2 \sum_{i=1}^{I} D_i^T V_i^{-1} D_i \).
APPENDIX A

JOINT GQL COMPONENTS

\[ D_i = \begin{pmatrix} \frac{\partial \theta_i}{\partial \beta^T} & \frac{\partial \theta_i}{\partial \nu} \\ \frac{\partial m_i}{\partial \beta^T} & \frac{\partial m_i}{\partial \nu} \end{pmatrix} \]

where

\[ \frac{\partial \theta_i}{\partial \beta^T} = \frac{\lambda_i^2}{\nu} x_i^T \]

\[ \frac{\partial \theta_i}{\partial \nu} = \frac{1}{2} \frac{1}{2v} - \frac{1}{2v} \frac{\lambda_i^2}{\nu^2} \frac{x_i^T \beta}{\nu^2} \]

\[ \frac{\partial m_i}{\partial \beta^T} = x_i^T \left( \frac{2\lambda_i^2 + 2\nu \lambda_i^2 - \nu \lambda_i^2}{\nu^2} \right) \]

\[ \frac{\partial m_i}{\partial \nu} = \frac{1}{2\nu^2} \left[ 2\lambda_i^2 \nu \log(\lambda_i) + \nu - 4\lambda_i^2 \log(\lambda_i)\nu - 4\lambda_i^2 \nu - 4\lambda_i^2 \log(\lambda_i) \right] \]

The covariance matrix of \( f_i \) is expressed as

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

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

\[ E[Y_i^{r+1}] = \lambda_i \frac{d}{d\lambda} E[Y_i^r] + E[Y_i] E[Y_i^r] \]

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

\[ \text{cov}(Y_i, Y_i^2) = E(Y_i^3) - E(Y_i) E(Y_i^2) \]

\[ = \frac{2\lambda_i^2 + 2\nu \lambda_i^2 - \nu \lambda_i^2}{\nu^2} \]

\[ \text{var}(Y_i^2) = E(Y_i^4) - E(Y_i^2)^2 \]

\[ = \frac{\lambda_i^2 \nu^2 + 4\lambda_i^2 \nu^2 + 10\lambda_i^2 \nu + 4\lambda_i^2 \nu + 4\lambda_i^2 \nu - 4\lambda_i^2 \nu^2}{\nu^3} \]

ACKNOWLEDGMENT

The authors would like to thank the Road Traffic Branch of Mauritius, the mechanics working at the Bala Garage and Tapsee motors for helping in the collection of data.

REFERENCES


