Estimating Regression Effects in Com Poisson Generalized Linear Model

Vandna Jowaheer, Naushad A. Mamode Khan

Abstract—Com Poisson distribution is capable of modeling the count responses irrespective of their mean variance relation and the parameters of this distribution when fitted to a simple cross sectional data can be efficiently estimated using maximum likelihood (ML) method. In the regression setup, however, ML estimation of the parameters of the Com Poisson based generalized linear model is computationally intensive. In this paper, we propose to use quasi-likelihood (QL) approach to estimate the effect of the covariates on the Com Poisson counts and investigate the performance of this method with respect to the ML method. QL estimates are consistent and almost as efficient as ML estimates. The simulation studies show that the efficiency loss in the estimation of all the parameters using QL approach as compared to ML approach is quite negligible, whereas QL approach is lesser involving than ML approach.

Keywords—Com Poisson, Cross-sectional, Maximum Likelihood, Quasi likelihood

I. INTRODUCTION

COUNT observations are encountered in various fields of research. Such observations may be equi, over or underdispersed. Poisson and Poisson-gamma (negative binomial) or Poisson-lognormal distributions are well accepted to model the equidispersed and overdispersed counts respectively [1,2,4,9,12,14]. However, analyzing underdispersed counts where the mean is much higher than the variance is quite a challenge. Recently, Shmueli et al. [13] proposed to use the Conway Maxwell Poisson (Com Poisson) distribution, originally developed by Conway and Maxwell [3], to model counts which may be equi, over and under dispersed. Kadane et al. [8] and Shmueli et al. [13] studied the basic properties of Com- Poisson distribution and the fitting of this distribution to over and under dispersed cross sectional count data. It has been pointed out that the estimation of the model parameters by ML method was comparatively more intensive than by weighted least squares (WLS) as well as Bayesian techniques. However, it is well known that ML estimates are more efficient as compared to WLS and Bayesian estimates.

In the regression setup, Guikema[7] developed a Com Poisson generalized linear model (GLM) and studied its application to set up a regression model for risk analysis . These authors implemented a fully Bayesian estimation

Dr V. Jowaheer is an associate Professor in the Department of Mathematics, Faculty of Science,University of Mauritius; email: vandnaj@uom.ac.mu).
N.A.Mamode Khan is a PhD student in the Department of Mathematics, Faculty of Science,University of Mauritius; email: alimamode@yahoo.co.uk).

approach based on Markov chain Monte Carlo estimation of the parameters. Lord et al[10] compared, in terms of goodness of fit statistics, the Com Poisson and negative binomial GLMs for analyzing motor vehicle crashes, where the estimates of the model parameters were based on Bayesian methods. According to these authors, Com Poisson GLM performs as well as the traditional negative binomial model when the counts are overdispersed but offers a better alternative when the counts are underdispersed. It is therefore evident that owing to its flexibility Com Poisson GLM can handle the counts subjected to either over or under dispersion. However, the issue of efficient estimation of model parameters should be taken into due consideration in order to draw valid conclusions from the model. ML estimation provides the most efficient estimates but is expected to be computationally intensive[6]. In this paper, we propose quasi-likelihood (QL) estimation technique to estimate the parameters of Com Poisson regression model. This technique provides considerably efficient estimates of all the parameters. This paper is divided into five sections. In the second section, a brief description of Com Poisson GLM is provided. The third section is devoted to ML estimation technique. In the same section, we present the QL estimation method and derive QL estimates. The next section provides a simulation study and MLEs are compared with QLEs. In section 5, we present comments on the simulation study and conclusion.

II. THE COMPOISSON REGRESSION MODEL

Let y_i be the ith count response ($y_i = 0,1,2,\ldots; i = 1,2,\ldots n$) and x_i be the p dimensional vector of covariates corresponding to y_i. Let β be the p dimensional vector of regression parameters such that β_j ($j = 1,\ldots,p$) is the regression effect of the jth covariate on the responses. The Com Poisson regression model is given by:

$$P(Y = y_i) = \frac{\lambda_i^{y_i}}{(y_i!)^{\upsilon}} \frac{1}{Z(\lambda_i,\upsilon)}$$

where

$$Z(\lambda_i,\upsilon) = \sum_{j=0}^{\infty} \frac{\lambda_i^{j}}{(j!)^{\upsilon}} \qquad , \lambda_i > 0, \upsilon > 0 \qquad (1)$$

and

$$\ln(\lambda_i) = x_i^T \beta \qquad (2)$$

In equation (1), υ is the dispersion parameter which conducts the modeling of equi, under and overdispersed count data. More specifically, the values $\upsilon = 1$, $\upsilon < 1$ and $\upsilon > 1$ correspond to equi, over and under(dispersion. Since equation (1) does not have closed form expressions for its moments, Shmueli et al. [13] derived an asymptotic expression for $Z(\lambda_i, \upsilon)$ given by

$$\frac{\exp(\upsilon \lambda_i^{\frac{1}{\upsilon}})}{\lambda_i^{\frac{\upsilon-1}{2\upsilon}} (2\pi)^{\frac{\upsilon-1}{2}} \sqrt{\upsilon}} \qquad (3)$$

Now, we reformulate the equation (1) to include regression effects as follows and use this expression to calculate the moments

$$p(y_i) = \frac{\exp(x_i^T y_i \beta)[\exp(x_i^T \beta(\frac{\upsilon-1}{2\upsilon}))(2\pi)^{\frac{\upsilon-1}{2}} \sqrt{\upsilon}]}{(y_i!)^{\upsilon}(\exp(\upsilon \exp(x_i^T \beta / \upsilon)))} \qquad (4)$$

To estimate the parameters β and υ, we investigate the maximum likelihood approach and the quasi likelihood techniques. In the next section, we develop the estimating equations for each method and use the iterative method to obtain estimates of the parameters β and υ.

III. ESTIMATION TECHNIQUES

A. Maximum Likelihood approach

The loglikelihood function based on model (4) can be written as

$$\ell(\beta,\upsilon) = \sum_{i=1}^{n} y_i(x_i^T \beta) - \upsilon \sum_{i=1}^{n} \ln(y_i!) + \frac{\upsilon-1}{2\upsilon} \sum_{i=1}^{n} x_i^T \beta - \upsilon \sum_{i=1}^{n} \exp(x_i^T \beta / \upsilon) + n\frac{\upsilon-1}{2} \ln(2\pi) +$$
$$\frac{n}{2} \ln(\upsilon) \qquad (5)$$

The derivates of the loglikelihood function with respect to β and υ are obtained as

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^{n} y_i x_{ij} - x_{ij} \exp(x_i^T \beta / \upsilon) + \frac{1}{2}(1 - \frac{1}{\upsilon})x_{ij} \qquad (6)$$

$$\frac{\partial \ell}{\partial \upsilon} = \frac{n}{2\upsilon} + \frac{n}{2} \ln(2\pi) + \sum_{i=1}^{n} x_i^T \beta \exp(x_i^T \beta / \upsilon) / \upsilon + \frac{x_i^T \beta}{2\upsilon} - \exp(\frac{x_i^T \beta}{\upsilon}) - \ln(y_i!) \qquad (7)$$

$$\frac{\partial^2 \ell}{\partial \beta_j^2} = \sum_{i=1}^{n} [-x^2_{ij} \exp(x_i^T \beta / \upsilon)] / \upsilon \qquad (8)$$

$$\frac{\partial^2 \ell}{\partial \beta_j \partial \upsilon} = \sum_{i=1}^{n} [-x^2_{ij} \exp(x_i^T \beta / \upsilon)] / \upsilon^2 + \frac{x_{ij}}{2\upsilon^2}] \qquad (9)$$

$$\frac{\partial^2 \ell}{\partial \upsilon^2} = \frac{-n}{2\upsilon^2} + \sum_{i=1}^{n} [-(x^T_i \beta)^2 \exp(x_i^T \beta / \upsilon) / \upsilon^3 (\frac{x^T_i \beta}{\upsilon^3}] \qquad (10)$$

The maximum likelihood estimates of the regression parameters and the dispersion parameter υ are derived by the Newton Raphson technique and are given in equation (11).

$$\begin{pmatrix} \beta_{r+1} \\ \upsilon_{r+1} \end{pmatrix} = \begin{pmatrix} \beta_r \\ \upsilon_r \end{pmatrix} + \begin{pmatrix} \frac{\partial^2 \ell}{\partial \beta^2} & \frac{\partial^2 \ell}{\partial \beta \partial \upsilon} \\ \frac{\partial^2 \ell}{\partial \beta \partial \upsilon} & \frac{\partial^2 \ell}{\partial \upsilon^2} \end{pmatrix}^{-1} \begin{pmatrix} \frac{\partial \ell}{\partial \beta} \\ \frac{\partial \ell}{\partial \upsilon} \end{pmatrix} \qquad (11)$$

where the expressions within the brackets is evaluated at $\beta = \beta_r$ and $\upsilon = \upsilon_r$. For an initial vector of the regression parameter $\beta = \beta_0$ and dispersion parameter $\upsilon = \upsilon_0$, we iterate the equation (11) until convergence. The estimators are consistent and under mild regularity conditions, for $n \to \infty$,

$n^{\frac{1}{2}}[(\beta^{ML}, \upsilon^{Ml}) - (\beta,\upsilon)]^T$ has an asymptotic normal distribution.

B. Quasi likelihood approach

In this subsection, we derive the estimates of the regression parameters and dispersion parameter discussed by Wedderburn [15] and Firth et al.[5]. This method requires only the first two moments of the ComPoisson distribution but since the distribution does not have closed form expressions for its moments in terms of the parameters λ and υ, following Shmueli et al. [13], we use equation (4) to obtain the $E(Y_i)$. Hence,

$$E(Y_i) = \frac{\lambda_i^{\frac{1}{\upsilon}}}{\upsilon} - \frac{\upsilon - 1}{2\upsilon} \qquad (12)$$

and

$$Var(Y_i) = \frac{\lambda_i^{\frac{1}{\upsilon}}}{\upsilon} \qquad (13)$$

By letting

$$\theta_i = E(Y_i) \qquad (14)$$

we can write

$$Var(Y_i) = \frac{\theta_i}{\upsilon} + \frac{\upsilon - 1}{2\upsilon^2} \qquad (15)$$

Other moments can be obtained by using the recursive formula

$$E(Y_i^{r+1}) = \lambda_i \frac{d}{d\lambda} E(Y^r) + E(Y)E(Y^r) \qquad (16)$$

Note that both ϑ_i and $Var(Y_i)$ are functions of β and υ. To estimate β and υ, we solve the joint quasi(likelihood equation given by

$$\sum_{i=1}^{n} D_i^T V_i^{-1} (f_i - \mu_i) = 0 \qquad (17)$$

where $\qquad f_i = (y_i, y_i^2), \mu_i = (\theta_i, m_i)^T \qquad$ where

$m_i = E(Y_i^2) = \theta_i^2 + \dfrac{\lambda_i^{\frac{1}{\upsilon}}}{\upsilon}$. V_i is the covariance matrix of the score vector f_i and D_i is the $2 \times (p+1)$ matrix of derivatives given by

$$D_i = \begin{pmatrix} \dfrac{\partial \theta_i}{\partial \beta^T} & \dfrac{\partial \theta_i}{\partial \upsilon} \\ \dfrac{\partial m_i}{\partial \beta^T} & \dfrac{\partial m_i}{\partial \upsilon} \end{pmatrix} \qquad (18)$$

Where,

$$\frac{\partial \theta_i}{\partial \beta^T} = \frac{\lambda_i^{\frac{1}{\upsilon}} x_i^T}{\upsilon} \qquad (19)$$

$$\frac{\partial \theta_i}{\partial \upsilon} = \frac{\upsilon - 1}{2\upsilon^2} - \frac{1}{2\upsilon} - \frac{\lambda^{\frac{1}{\upsilon}} x_i^T \beta}{\upsilon^2} \qquad (20)$$

$$\frac{\partial m_i}{\partial \beta^T} = \frac{x_i^T (2\lambda_i^{\frac{1}{\upsilon}} + 2\upsilon\lambda_i^{\frac{2}{\upsilon}} - \upsilon\lambda_i^{\frac{1}{\upsilon}})}{\upsilon^2} \qquad (21)$$

$$\frac{\partial m_i}{\partial \upsilon} = \frac{[2\lambda_i^{\frac{1}{\upsilon}}\upsilon\ln(\lambda_i) + \upsilon - 1 - 4\lambda_i^{\frac{2}{\upsilon}}\ln(\lambda_i)\upsilon - 4\lambda_i^{\frac{1}{\upsilon}}\upsilon - 4\lambda_i^{\frac{1}{\upsilon}}\ln(\lambda_i)]}{2\upsilon^3} \qquad (22)$$

The covariance matrix of f_i can be expressed as

$$V_i = \begin{pmatrix} var(Y_i) & cov(Y_i, Y_i^2) \\ & var(Y_i^2) \end{pmatrix} \qquad (23)$$

where the elements in V_i are derived iteratively from the equation (16)

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

$$Cov(Y_i, Y_i^2) = E(Y_i^3) - E(Y_i)E(Y_i^2)$$

$$= \frac{2\lambda_i^{\frac{1}{\upsilon}} + 2\upsilon\lambda_i^{\frac{2}{\upsilon}} - \upsilon\lambda_i^{\frac{1}{\upsilon}}}{\upsilon^2} \qquad (24)$$

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

$$= \frac{\lambda_i^{\frac{1}{\upsilon}}\upsilon^2 + 4\lambda_i^{\frac{3}{\upsilon}}\upsilon^2 + 10\lambda_i^{\frac{2}{\upsilon}}\upsilon - 4\lambda_i^{\frac{1}{\upsilon}}\upsilon + 4\lambda_i^{\frac{1}{\upsilon}} - 4\lambda_i^{\frac{2}{\upsilon}}\upsilon^2}{\upsilon^3} \qquad (25)$$

Now the Newton Raphson technique can be used to obtain the quasi-likelihood estimates of β and υ. Hence the QL estimates are given by

$$\begin{pmatrix} \beta_{r+1} \\ \upsilon_{r=1} \end{pmatrix} = \begin{pmatrix} \beta_r \\ \upsilon_r \end{pmatrix} + [\sum_{i=1}^{n} D_i^T V_i^{-1} D_i]^{-1} [\sum_{i=1}^{n} D_i^T V_i^{-1} (f_i - \mu_i)] \qquad (26)$$

For an initial values of $\beta = \beta_0$ and dispersion parameter $\upsilon = \upsilon_0$, we iterate the equation (26) until convergence. The estimators are consistent and under mild regularity conditions, for $n \to \infty,$, it may be shown that $n^{\frac{1}{2}}[(\beta^{QL}, \upsilon^{QL}) - (\beta, \upsilon)]^T$ has an asymptotic normal distribution.

IV. SIMULATION STUDY

In the simulation study, we first generate Com Poisson counts by using the relationship described by Shmueli et al. [13]

$$P(Y_i + 1) = P(Y_i)\lambda_i(y_i + 1)^{-\upsilon} \qquad (27)$$

where $\qquad P(Y_i = 0) = \dfrac{\lambda_i^{\frac{\upsilon - 1}{2\upsilon}}(2\pi)^{\frac{\upsilon - 1}{2}}\sqrt{\upsilon}}{\exp(\upsilon\lambda_i^{\frac{1}{\upsilon}})} \qquad (28)$

Minka et al. [11] calculated the sum of these probabilities starting from $P(Y_i = 0)$ until the sum exceeds the value of a simulate Uniform (0,1) variable. Using this method of simulation, it is remarked that for small values of υ, the Com(Poisson counts take different values between 0 and 10 while for very large values of υ, the counts are repeated. Hence, we consider $n = 20, 60, 100$ and 500 and different values of υ for small, medium and large sample sizes. Thus for the above choices of n, the simulation results will exhibit small and large sample performances of the methodology. For any $i = 1 \ldots n$, the first covariate is chosen as

$$x_{i1} = \begin{cases} 1 & (i=1,\ldots\ldots,n/4), \\ 0 & (i=(n/4)+1,\ldots,3n/4) \\ 1 & (i=(3nl/4)+1,\ldots,n) \end{cases}$$

and for the second covariate, we generate n standard normal values. For convenience, the true values of the regression parameters are assumed $\beta_0 = \beta_1 = 1$. For each structure, 1,000 simulations were run for each of the following values of $\upsilon =0.5,0.75,0.85,1,1.5,2$. The results are presented in the following table

TABLE I SIMULATION RESULTS

n	υ	β_0		β_1		υ	
		MLE	QLE	MLE	QLE	MLE	QLE
20	0.5	1.0221	1.0256	0.9998	0.9989	0.5006	0.4997
		(0.0728)	(-0.0733)	(0.071)	(-0.0716)	(0.0812)	(-0.0818)
	0.75	0.9978	0.9993	1.0981	0.9997	0.7499	0.7502
		(0.0822)	(0.0826)	(0.0699)	(0.0705)	(0.0792)	(0.0798)
	1	1.0254	1.0012	1.0025	1.0015	0.9989	0.9992
		(0.0755)	(0.0758)	(0.0736)	(0.0745)	(0.0667)	(0.0672)
	1.5	1.001	1.001	0.9999	0.9999	1.4989	1.501
		(0.0832)	(0.0839)	(0.0897)	(0.0901)	(0.0756)	(0.076)
	2	1.0022	0.9999	1.0005	0.9999	1.9998	1.9999
		(0.0812)	(0.082)	(0.0881)	(0.0894)	(0.0692)	(0.0699)
60	0.5	1.0201	1.0341	0.9945	0.9981	0.5012	0.4981
		(0.0622)	(0.0631)	(0.0642)	(0.0651)	(0.0751)	(0.0756)
	0.75	1.0654	0.9986	1.0671	0.9992	0.7481	0.7485
		(0.0718)	(0.0745)	(0.0666)	(0.0698)	(0.0778)	(0.0795)
	1	1.011	1.0101	1.0003	0.9998	0.9895	0.9899
		(0.0713)	(0.0721)	(0.0725)	(0.073)	(0.0613)	(0.0628)
	1.5	1.0012	1.0008	0.9996	0.9999	1.4561	1.5012
		(0.0824)	(0.0836)	(0.0888)	(0.0891)	(0.0723)	(0.0734)
	2	1.0002	0.9991	1.0003	0.9999	1.9995	1.9999
		(0.0797)	(0.0801)	(0.0815)	(0.0818)	(0.0688)	(0.0692)
100	0.5	0.9871	1.0561	0.9999	0.9982	0.4981	0.551
		(0.0536)	(0.054)	(0.0401)	(0.0403)	(0.051)	(0.0512)
	0.75	0.9993	1.0012	0.9999	1.0052	0.7493	0.751
		(0.0522)	(0.0524)	(0.0436)	(0.044)	(0.0601)	(0.0607)
	1	0.9999	1.0001	0.9999	0.9999	1.0012	0.9996
		(0.0689)	(0.0695)	(0.0711)	(0.0718)	(0.0589)	(0.0592)
	1.5	0.9999	0.9989	0.9999	1.0021	1.5101	1.4993
		(0.0718)	(0.0721)	(0.0801)	(0.0811)	(0.0677)	(0.068)
	2	0.9999	1.0001	0.9999	0.9999	2.001	1.9996
		(0.0791)	(0.0793)	(0.0818)	(0.082)	(0.0681)	(0.068)
500	0.5	0.9997	0.9999	1.0076	1.0043	0.5032	0.501
		(0.0321)	(0.032)	(0.0385)	(0.039)	(0.0482)	(0.049)
	0.75	0.9999	0.9996	1.0001	0.9999	0.7498	0.7499
		(0.0333)	(0.0335)	(0.0288)	(0.0291)	(0.0487)	(0.0491)
	1	1.0052	0.9999	1.0025	0.9998	0.9992	0.9995
		(0.0467)	(0.0431)	(0.0481)	(0.0485)	(0.0492)	(0.0497)
	1.5	1.0003	0.9999	0.9997	0.9998	1.4989	1.501
		(0.0667)	(0.0668)	(0.0592)	(0.0598)	(0.0515)	(0.0518)
	2	1.0001	0.9999	1.0008	0.9998	1.9992	1.9995
		(0.0617)	(0.0619)	(0.0566)	(0.057)	(0.0498)	(0.0501)

V. CONCLUSION

With small starting positive values for β and υ, we obtained the estimates of β and υ under maximum likelihood and quasi-likelihood approaches using equations equation (11) and equation (26) respectively. It is remarked that ML procedure is computationally very intensive as compared to QL method. For $n = 60$; $\upsilon = 0.5$ and $n = 100$; $\upsilon = 0.5$ the ML iterative process does not converge in 45 percent and 18 percent of the simulations respectively whereas the QL approach fails in only 15 percent and 10 percent of the simulations. As we increase the value of υ to 0.75 and 0.85, we note that there is a slight increase in the number of nonconvergent simulations in the ML approach under $n = 60$ and $n = 100$. However, for $n = 500$, the nonconvergence problems did not arise for both methods which is expected, as the consistency of estimators depend on large I. For values of υ greater than 1, we found out that for $n = 60$ and $n = 100$, the ML iterative process does not converge in 55 percent and 36 percent of the simulations respectively for the case with $\upsilon = 1.5$ whereas the QL approach fails only in 20 percent and 12 percent of simulations respectively. Hence, ML procedure is far more difficult to implement as compared to QL procedure, especially for underdispersed data ($\upsilon > 1$) and for small as well as medium sized samples ($n \le 100$). In terms of efficiency as well, the performance of QL method is quite satisfactory. QL approach yields almost as efficient estimates as provided by the ML. Even for small samples, the efficiency loss for the QL estimates is not more than 1 %. Thus, it is evident that QL technique is a better option for estimating the parameters of Com Poisson regression models.

REFERENCES

[1] G. Bates and J. Neymann, " Contribution to the theory of accident processes", *Publications in Statistics* Vol 1, 215-253, 1952, University of California.
[2] N. Breslow and X. Lin , " Bias correction in generalized linear mixed models", *Biometrics* Vol 82, 81-91, 1995.
[3] R. Conway and W. Maxwell, " A queuing model with state dependent service rates", *Journal of Industrial Engineering*, Vol 12, 132-136, 1962.
[4] C. Dean and R. Balshaw, " Efficiency lost by analyzing counts rather than even times in Poisson and overdispersed Poisson regression models", *Journal of American Statistical Association*, Vol 92, 1387-1398, 1997.
[5] D. Firth and L. Harris, " Quasi-likelihood for multiplicative random effects", *Biometrics*, Vol 78, 545-555, 1991.
[6] S. Guikema, " Formulating informative, data(based priors for failure probability estimation in reliability analysis", *Reliability Engineering and System Safety*, Vol 92, 490-502, 2007.
[7] S. Guikema and J. Gofelt, " A flexible count data regression model for risk analysis", *Risk analysis*, Vol 28, 213-223, 2008.
[8] J. Kadane, G. Shmueli, G. Minka, T. Borle and P. Boatwright, " Conjugate analysis of the Conway Maxwell Poisson distribution", *Bayesian analysis*, Vol 1, 363-374, 2006.
[9] J. Klein, " Semi parametric estimation of random effects using Cox model based on the EM algorithm", *Biometrics* , Vol 48, 795-806, 1992.
[10] D. Lord , S. Washington and J. Ivan, " Poisson, Poisson(Gamma and Zero inflated regression models of motor vehicle crashes: Balancing statistical fit and theory", *Accident analysis and Prevention*, Vol 37, 35-46, 2005.

[11] T. Minka, G. Shmueli, J. Kadane, J. Borle and P. Boatwright," Computing with the Com Poisson distribution", Technical report CMU statistics department, http:// www.stat.cmu.edu/tr/tr776/tr776.html, 2003.

[12] D. Moore, " Asymptotic properties of moment estimators for overdispersed counts and proportions", *Biometrics* Vol 73, 583-588, 1986.

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

[14] B. Sutradhar and K. Das, " A higher order approximation to likelihood inference in the Poisson mixed model", *Statist, Prob, Lett*, Vol 52, 59-67, 2001.

[15] R. Wedderburn, " Quasi-likelihood functions, generalized linear models and the Gauss Newton method, *Biometrics*, Vol 61, 439-447, 1974.