J. Pharm. Biomed. Sci., 1(1) 19- 22, 2010

ISSN 2090 – 424X
Journal of Pharmaceutical and Biomedical Sciences
www.textroad.com

Analyzing the Computational Complexities of the Quasi-Likelihood Estimation Approach: Application to Breastfeeding Data

*Naushad Mamode Khan[1], [2]Cheika Jahangeer and [3]Maleika Heenaye-Mamode Khan

[1]Department of Economics and Statistics, Faculty of Social Studies and Humanities, University of Mauritius, Africa
[2] Ministry of Education and Human resources, Mauritius, Africa
[3]Department of Computer Science and Engineering, University of Mauritius, Africa

ABSTRACT: **Problem statement**: The quasi-likelihood estimation technique is a popular approach to estimate parameters in regression models. It is based on the correct specification of the mean and variance components. It has been shown to provide equally efficient estimates as the maximum likelihood approach in various studies. However, its computational complexities have not yet been explored so far.
Approach: In this paper, we use the quasi-likelihood approach to analyze a breastfeeding data based on the Generalized Poisson distribution and calculate its computational cost.
Results: We conclude that the quasi-likelihood estimation technique computes the unknown parameters at fairly low cost and involves few numbers of flop counts.
Key words: Quasi-Likelihood Equations (QLE), computational complexity, Generalized Poisson Regression model (GPR)

INTRODUCTION

Breast milk is the most ideal food for an infant and it should meet most of the nutritional requirements if sufficiently supplied. Exclusive breastfeeding is recommended for the first six months of life followed by nutritionally adequate and safe complementary foods with continued breastfeeding up to two years of age or beyond. Breastfeeding has a lot of health benefits in the prevention of acute and chronic diseases and it helps tremendously by lowering infant mortality rate. Mauritius as all the other developing nation is also affected by a reduction in the incidence of exclusive breastfeeding which can be accounted by factors such as maternal age, employment, length of maternity leave, place of antenatal treatment, information obtained on breastfeeding, type of delivery and place of delivery.
In this study, the Generalized Poisson Regression model (GPR) have been used (Famoye and Singh, 2006; Famoye, 1993) to analyze the practices of exclusive breastfeeding in Mauritius which were based on a random subset of data collected from a survey on breastfeeding in Mauritius over the period 2006-2008. The organization of the paper is as follows: In the next sections, we describe the factors influencing exclusive breastfeeding in Mauritius since 2006 and review the GPR model. Estimation of parameters is done using quasi-likelihood estimation technique and eventually we analyze the exclusive breastfeeding data and present the results. We will also calculate the computational complexity of the estimation method.

MATERIALS AND METHODS

Maternal age is one of the factors that can adversely affect breastfeeding rates among mothers. Employment, maternity leave and the length of maternity leave are also very influential on the mother's choice of infant feeding practice and on exclusive breastfeeding. Although working mothers may have a great awareness of the benefits of breastfeeding, many of them are rather reluctant to practice exclusive breastfeeding as compared to the unemployed mothers.

Working outside the home and being a full-time worker is related to a reduction in breastfeeding. One of the most important reasons for mothers to stop breastfeeding at 6 months or earlier was"returning to work" (Integrated care for mother and child, 2004). Information on breastfeeding can influence a mother's choice of feeding practice. Other authors have stipulated that health education could improve the present status on infant feeding practices (Singhania *et al.*, 1990). The lack of appropriate information on

*Corresponding author: Naushad Mamode Khan, Department of Economics and Statistics, Faculty of Social Studies and Humanities, University of Mauritius, Africa.

breastfeeding can sometimes acts as a barrier to its practice though women are strongly determined to breastfeed. Moreover, continuous support using a nutrition education 'communication mix' is prone to be more effective to result in positive behavior change towards infant feeding practices (Sethi *et al*., 2003).

Sufficient support needs to be given to breastfeeding mothers to further enhance breastfeeding practice beyond the first month and also imparting proper knowledge to mothers and grandmothers is very important to establish good infant feeding practices (Barton, 2001). As for the type of delivery, it may have a negative effect on breastfeeding initiation where caesarian section seems to be a big barrier for the rightly timed initiation of breastfeeding. Caesarian section is becoming an increasingly common practice in the private hospitals among the upper and middle income groups and this seems to be an obstacle to successful breastfeeding (Reddy, 1995).The site of antenatal treatment and the place of delivery can also have an impact on proper breastfeeding practice among the Mauritian mothers. There are two types of hospital set-up in Mauritius namely the public hospitals and the private hospitals where there is a difference in the services provided at the antenatal care, prenatal and postnatal care.

A survey was carried out in 2008 to 2010 with mothers having an infant between the age of 6 and 24 months old using a questionnaire. A random subset of this data consisting of 2500 mothers was chosen. The average practice of exclusive breastfeeding during the first six months of the baby's life is approximately 305 while the variance 106 was noted. This indicates that the data is under-dispersed. To model such data under a regression set-up, the GPR model is used as the following (Famoye and Singh (2006) and Jahangeer *et al*., (2009))

Generalized Poisson Regression model (GPR):

Let y_i be a count response and X_i be a p-dimensional vector of covariates for subject i(i = 1, . . . , I). Let β be the vector of regression parameters such that β_j (j = 1, 2, . . . , p) is the regression effect of the jth covariate on the incidence of exclusive breastfeeding among Mauritian mothers. The density function of y_i is given by:

$$f(y_i,\theta_i,\alpha)=\left(\frac{\theta_i}{1+\alpha\theta_i}\right)^{y_i}\frac{(1+\alpha y_i)^{y_i-1}}{y_{i!}}\exp\left[\frac{-\theta_i(1+\alpha y_i)}{1+\alpha_{\theta_i}}\right] \qquad (1)$$

y_i = 0,1,2,3….., where $\theta_i=\exp(X_i^T\beta)$ The mean of y_i is given by θ_i and the variance of y_i is given by $\theta_i(1+\alpha\theta_i)^2$ where $\alpha < 0$ represents count data with under-dispersion. Following Jahangeer *et al*. (2009), we developed two quasi-likelihood equations (QLE): 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 based on the squared observations. The QLE to estimate β is given by:

$$\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}^{-1}(y_i-\theta_i)=0 \qquad (2)$$

where, $V_{i,\beta}^{-1}=\theta_i(1+a\theta_i)^2.D_{i,\beta}=\frac{\partial\theta_i}{\partial\beta^T}=\theta_i X_i^T$ is a p × 1 matrix. The QLE to estimate is given by:

$$\sum_{i=1}^{I} D_{i,\alpha}^T V_{i,\alpha}^{-1}(y_i^2-\eta_i)=0 \qquad (3)$$

where, $\eta_i=\theta_i(1+a\theta_i)^2+\theta_i^2$ and $D_{i,a}=2\theta_i^2(1+a\theta_i)$. $V_{i,a}$ is the variance of Y_i^2 and is calculated using:

$$V_{i,\alpha}=E(Y_i^4)-E(Y_i^2)^2 \qquad (4)$$

Where:

$$E(Y_i^4)=\frac{3\theta_i^2}{(1-\alpha\theta_i)^6}+\theta_i\left(\frac{15}{(1-\alpha\theta_i)^2}-\frac{20}{(1-\alpha\theta_i)}+6\right)\frac{1}{(1-\alpha\theta_i)^5} \qquad (5)$$

following (Famoye, 1993); Johnson *et al*. (1993). The Newton-Raphson technique is then applied to the two estimating equations. The iterative r^{th} equations are given as follows: At the r^{th} iteration:

$$(\hat{\beta}_{r+1})=(\hat{\beta}_r)+\left[\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}^{-1} D_{i,\beta}\right]_r^{-1}\left[\sum_{i=1}^{I} D_{i,\beta}^T V_{i,\beta}^{-1}(y_i-\theta_i)\right]_r \qquad (6)$$

$$(\hat{\alpha}_{r+1})=(\hat{\alpha}_r)+\left[\sum_{i=1}^{I} D_{i,\alpha}^T V_{i,\alpha}^{-1} D_{i,a}\right]_r^{-1}\left[\sum_{i=1}^{I} D_{i,\alpha}^T V_{i,\alpha}^{-1}(y_i^2-\eta_i)\right]_r \qquad (7)$$

where:

$\hat{\beta}_r$ and $\hat{\alpha}_r$ = The values of $\hat{\beta}_r$ and $\hat{\alpha}_r$ at the r^{th} iteration

$[.]_r$ = The value of the expression at the r^{th} iteration

The estimates of $\hat{\beta}_r$ and $\hat{\alpha}_r$ are consistent and efficient following Jahangeer *et al.* ,2009. The algorithm to estimate the parameters is also described in Jahangeer *et al.* ,2009 .

RESULTS

The covariates are the intercept term, age of the mothers, length of maternity leave, place of antenatal treatment, information on infant feeding practices, type of delivery and place of delivery. The results are presented in the table below. They have been obtained by taking small initial values of the regression parameters. The entry in brackets represents the standard errors of each estimate.

Intercept	**1.2732**	**(0.3541)**
Age	**-5.6184**	**(0.0281)**
Length of maternity leave	**0.9541**	**(0.0345)**
Place of antenatal treatment	**-8.7641**	**(0.3456)**
Information	**10.1912**	**(0.1011)**
Type of delivery	**-2.5520**	**(0.1421)**
Place of delivery	**-1.3115**	**(0.1276)**
$\hat{\alpha}$	**-1.5611**	**(0.3091)**

Complexity analysis of the QLE approach:

To measure the computational complexities of the QLE algorithms, we focus on the two estimating equations 6 and 7. Basically, both equations require a gradient matrix, a covariance matrix and the score vector. For the estimation of the vector of regression parameters β, the number of flop counts involved are as follows:

- The inverse of $V_{i,\beta}$ requires O(1) number of operations
- This inverse is then multiplied with the gradient matrix $D_{i,\beta}$ twice. This operation involves O(p) number of flop counts
- The component $D^T_{i,\beta}V_{i,\beta}^{-1}$ is then multiplied with (y_i-(θ_i)) and this requires only O(p) operations

These above operations are iterated r times until convergence and are summed at each step I times. Thus, the total complexity may be given by the formula O(Irp). Similarly, the total number of flop counts involved in estimating the dispersion parameter is O(Ir) since it is a single parameter. Hence, the total number of flop counts is O(Irp) + O(Ir).

DISCUSSION

The negative value of the age factor indicates that age can adversely affect the practice of exclusive breastfeeding which was very pertinent among young mothers of less than 18 years old. The positive estimate of the length of maternity leave shows that as the number of days of maternity leave increases, mothers are more likely to adopt a better infant feeding practice and it also increases the incidence of exclusive breastfeeding. The estimated value concerning the place of antenatal treatment reflects the current situation of the private and public health care in Mauritius. More information on proper infant feeding practices is being dispersed in the public health institution as compared to the private health sector, thus justifying the negative sign. Similarly, the estimate for the place of delivery is negative, thus showing a disparity at the level of the private and public health institutions where only the latter have adopted the Baby Friendly Hospital Initiative (BFHI). As for the regression estimate corresponding of the type of delivery, it indicates that mothers undergoing caesarian are less likely to practice exclusive breastfeeding. The information parameter estimate justifies that mothers who have been well informed are more likely to adopt appropriate practice infant feeding practices and exclusive breastfeeding for the recommended time. We also find out the number of flop counts to estimate the two sets of parameters turn around O(Irp) + O(Ir) which are not huge. This implies that in the longitudinal case, the number of flop counts will considerably increase as the covariance structure will be in matrix form.

CONCLUSION

Overall, the quasi-likelihood estimation technique is a computationally efficient technique in analyzing cross-sectional data. Besides, quasi-likelihood estimation based on the Generalized Poisson model yields reliable and consistent regression estimates. However, as reported by Jahangeer *et al.* (2009), the estimate of the dispersion parameter does not quite reflect the true situation when compared with the mean-variance ratio of the exclusive breastfeeding data. To overcome this problem, an alternative under-dispersed discrete model may be used.

REFERENCES

Barton, S.B., 2001. Infant feeding practices of how low-income rural mothers. Am. J. Mater. Child Nurs., 26: 93-97.

Famoye, F., 1993. Restricted generalized poison regression model. Commun. Stat., 22: 1335-1354.

Famoye, F. and K.P. Singh, 2006. Zero-inflated generalized poison regression model with an application to domestic violence data. J. Data Sci., 4: 117-130.

Integrated Care for Mother and Child, 2004. Baby friendly and beyond, issue One, USA., pp: 1-6.

Jahangeer, C, N. Mamode Khan and M. Heenaye Mamode Khan, 2009. Analyzing the factors influencing exclusive breastfeeding using the Generalized Poisson Regression model, 3:6.

Johnson, N.L., S. Kotz and A.W. Kemp, 1993. Univariate Discrete Distributions. Wiley, New York

Reddy, S., 1995. Breastfeeding practices, problems and prospects. J. Family Welfare, 41: 43- 51.

Sethi, V., S. Kashyap and V. Seth, 2003. Effect of Nutrition education of mothers on infant feeding practices. Indian J. Pediatr., 70: 463-466.

Singhania, R.S. Kabra and A. Bansal, 1990. Infant feeing practices in educated mothers from upper socio-economic status. Indian Pediat., 27: 591-593.