

The Poisson Inverse Gaussian Regression Model in the Analysis of Clustered Counts Data

M. M. Shoukri^{1,2}, M. H. Asyali², R. VanDorp³ and D. Kelton³

¹*University of Western Ontario*

²*King Faisal Specialist Hospital and Research Centre*
and ³*University of Guelph*

Abstract: We explore the possibility of modeling clustered count data using the Poisson Inverse Gaussian distribution. We develop a regression model, which relates the number of mastitis cases in a sample of dairy farms in Ontario, Canada, to various farm level covariates, to illustrate the methodology. Residual plots are constructed to explore the quality of the fit. We compare the results with a negative binomial regression model using maximum likelihood estimation, and to the generalized linear mixed regression model fitted in SAS.

Key words: Bessel functions, dispersion test, general linear mixed model, maximum likelihood estimation, residual plots.

1. Introduction

Mastitis may be defined as inflammation of the mammary gland of dairy cows, and may be categorized into subclinical and clinical mastitis. Clinical mastitis (CM) shows alterations of the milk in the form of flakes or pus, and is associated with a considerable increase in somatic cell counts (SCC). On the other hand, sub-clinical mastitis is characterized by a slight increase in the SCC and cannot be seen by eye. Identification of udders infected with sub-clinical mastitis may be performed with bacteriology and California Mastitis Test (CMT). The CMT is a simple rapid means for detecting mammary gland infection and has had wide acceptance and use by veterinarians and dairymen in routine mastitis prevention and control programs.

Two main groups of mastitis causing pathogens can be distinguished, environmental and contagious pathogens. Environmental pathogens are always present in the surroundings of the cow. Infection of the udder by these environmental pathogens is accompanied with a sudden increase in SCC. Contagious pathogens are transmitted during the milking process, and are usually spread from infected

to non-infected susceptible animals during that process. The most common organism involved in dairy mammary disease is *Staphylococcus aureus*, commonly found on the skin of human hands and the udder skin of cows. Cows with contagious mastitis are usually culled to prevent the spread of the disease to other healthy cows in the herd. Due to these losses, the disease is of considerable economic importance.

Farm management practices play an important role in controlling and reducing the incidence of the disease. Hygiene is identified as an important factor. The use of organic (straw and sawdust) versus inorganic (sand) is associated with an increase in intra-mammary infections. Tail docking is practiced to prevent the tail from hitting the udder and spread of the disease causing pathogen. Milking technique is also identified as a route to spread mastitis from cow to cow. Some farmers milk the teats by hand (post milking) after the machine milking to reduce residual milk in the quarters, and such practice might serve as an ideal growth flora for bacteria.

We intend to identify the extent of the relationship between the above risk factors and the incidence of sub-clinical mastitis. The data are available from 57 Ontario dairy farms. These herds were willing to participate in the sentinel dairy herd project, which was instigated to identify the reasons behind the increase in SCC over time. From July 1997 to February 1999, each farm was visited three to six times. At each visit, composite samples were taken from each milking cow in the herd, and bacteriology was performed at the Animal Health Laboratory, Ontario Veterinary College, at the University of Guelph, to identify the pathogen causing sub-clinical mastitis. The primary response variable was the number of cows tested positive for the disease. Table 1 shows a summary of the data by visit. Herd level covariates were: type of bedding, tail docking, and post milking. These factors were coded as 0 when the producer adopted the factor, otherwise the factor was coded as 1. The fundamental feature of the data is its hierarchical or multilevel (clustered) structure, which frequently arises in many areas of investigations such as veterinary epidemiology, agriculture, and community trials. The highest hierarchy was the herd, and the lowest hierarchy was the visit. At each visit, the number of cows cultured, and the number of cows tested positive (the response variable) were recorded.

There are two general classes of models for analyzing clustered data: cluster-specific (CS) and marginal or population averaged (PA). The PA approach specifies the marginal of the response of the j th unit in the i th cluster y_{ij} , and the Generalized Estimating Equation is the method of choice (Liang and Zeger 1986; Zeger et al. 1988). The most common CS approaches are Generalized Linear Mixed Models (GLMMS), which extend the class of generalized linear models by including random effects in the linear predictor (Breslow and Clayton 1993;

Table 1: Summary Mastitis Data collected from 57 farms at six visits to each farm.

Visit	Number of farms visited	Total number of cultured cows	Total number of cows with CM	Percentage of positive cows
1	57	2721	277	10
2	57	2907	365	12.6
3	57	2837	342	12
4	55	2585	262	10
5	48	2304	306	13.3
6	3	141	30	21.3

Schall 1991; McGilchrist 1994; Goldstein 1991 and Longford 1994). The above CS approach fits models to data using either the Taylor approximation of the log likelihood function (Goldstein 1991 and Longford 1994) or the penalized quasi-likelihood (PQL) (Breslow and Clayton's 1993). Wolfinger and O'Connell (1993) describe an algorithm for computing PQL estimates and implement it with a SAS macro called GLIMMIX. Neuhaus and Segal (1997) showed that these approaches may yield highly biased estimates of covariate effects and the variance components of the random effects distribution. Recently, Nehaus (2001) recommended that care must be taken when using approximate methods in the analysis of multilevel data. To avoid most of the problems outlined by Neuhaus (2001), we use an exact modeling procedure.

We consider a random effects Poisson model where the mixing distribution is the inverse-Gaussian. Details of the derivations, the method for incorporation of the covariates, and the procedures used to obtain estimates of the parameters are given in Section 2. In Section 3 we fit the model to the mastitis data to illustrate the potential differences between the Poisson inverse-Gaussian Model and other models such as the negative binomial and the generalized linear mixed model. A general discussion is presented in Section 4.

2. Poisson Inverse Gaussian Model (PIGM)

2.1 Over-dispersion test

Assume k clusters the i th of which has n_i observations. Define Y_{ij} to be the number of events of interest (cows with clinical mastitis) in the i th farm at visit j . Because of the hierarchical structure of the data, one should expect that the counts Y_{ij} within a particular farm to be correlated. The result of such

within-farm correlation is that the variance of counts would be much larger than the mean, a phenomenon known “over-dispersion.” This over-dispersion may be modeled as follows (Cox, 1983): first we assume that conditional on the i th cluster (farm) effect, denoted by ν_i we have

$$\begin{aligned} E(y_{ij}|\nu_i) &= \nu_i\mu_{ij} \quad \text{and} \\ \text{var}(y_{ij}|\nu_i) &= \nu_i\mu_{ij} \end{aligned}$$

Second, we assume that the random cluster effect ν_i has $E(\nu_i) = 1$, and $\text{var}(\nu_i) = \lambda < \infty$. Therefore

$$\begin{aligned} E(y_{ij}) &= \mu_{ij}, \quad \text{and} \\ \text{var}(y_{ij}) &= E[\text{var}(y_{ij}|\nu_i)] + \text{var}[E(y_{ij}|\nu_i)] = \mu_{ij}(1 + \lambda\mu_{ij}), \end{aligned}$$

regardless of distribution assigned to ν_i . Here ν_i is a random variable that summarizes the effect of our inability to model all pertinent explanatory factors and other sources of variation. The above set-up has been discussed by many authors (Breslow 1984, Thall 1988, Stukel 1993, Burnett *et al.* 1992, and Zeger *et al.* 1988). Therefore $\lambda > 0$ indicates the presence of over-dispersion, $\lambda < 0$ under-dispersion, and $\lambda = 0$ equi-dispersion. Under equi-dispersion, the Poisson model is the most commonly used model to analyze count data. It seems natural then that our attempting to construct a plausible parametric model should be preceded by a test for over-dispersion.

Following Cameron and Trivedi (1998, p.78), a test of over-dispersion can empirically be carried out by using the method of least squares to fit the regression equation:

$$D_{ij} = \frac{(y_{ij} - \bar{y}_i)^2 - y_{ij}}{\bar{y}_i} = \lambda\bar{y}_i + \epsilon_{ij}$$

where, $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$, and ϵ_{ij} is an error term. The reported t -statistic for $\tilde{\lambda}$ is asymptotically normal under the null hypothesis $H_0 : \lambda = 0$ versus $H_1 : \lambda > 0$. For the mastitis data, it was found that $\tilde{\lambda} = 0.1169$ and the standard error $SE(\tilde{\lambda}) = 0.0232$, from which $t = 0.1169/0.0232 = 5.03$ (p -value < 0.001). Therefore, based on the evidence in the data, we conclude that over-dispersion should not be ignored.

We therefore need to develop models for count data, which properly account for over-dispersion. In the next section, we develop a PIGM as a competitor to the negative binomial model (NBM).

There are several reasons for comparing the PIGM to a regression model assuming a negative binomial distribution for the counts. First, although the NBM is commonly used to model over-dispersed count data, most data analysts are not familiar with the PIGM, as it has not been fitted to clustered count

data. Secondly; there is not one single model that provides a reasonable fit to a given data set, particularly in the absence of knowledge of the biological mechanism generating the responses. Therefore, it is not uncommon for data analysts to attempt fitting several models and investigate the advantages and the disadvantages of each. For example, Stukel (1993) analyzed clinical trial data of non-melanoma skin cancer and familial polyposis, using several models for over-dispersed longitudinal count data. As well, Have and Hartzel (1995), and more recently Molenberghs and Geys (2001) discussed several modeling strategies to analyze clustered binary data in developmental toxicity studies.

2.2 Model specifications

Assume that conditional on the i th cluster (farm) the within cluster observations are conditionally independent so that their joint density is

$$P(y_{i1}, y_{i2}, \dots, y_{in_i} | \nu_i) = \prod_{j=1}^{n_i} \frac{e^{-\nu_i \mu_{ij}} (\nu_i \mu_{ij})^{y_{ij}}}{y_{ij}!},$$

$\mu_{ij} > 0, i = 1, 2, \dots, k, j = 1, 2, \dots, n_i$.

Following the generalized linear model approach, we relate the parameters μ_{ij} to the covariates $x_i \in R^p$ through the log-link function so that

$$\log \mu_{ij} = \beta_0 + \beta_1 \text{num}_{ij} + \beta_2 \text{bedorg}_{ij} + \beta_3 \text{postmilk}_{ij} + \beta_4 \text{tail}_{ij},$$

where $\beta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4)^T$ is the vector of parameters, num_{ij} is the number of animals cultured from i th farm at visit j , and the rest of the covariates are as explained in Section 1.

The specification of the joint distribution of $Y_i = (Y_{i1}, Y_{i2}, \dots, Y_{in_i})^T$ is completed by assigning a distribution for ν_i . If we assign gamma distribution for ν_i , the resulting marginal density of Y_i is the negative binomial (see, Collings and Margolin 1985 and McCullagh and Nelder 1989). Alternatively, we assume that ν_i has an inverse-Gaussian distribution, parameterized to have $E(\nu_i) = 1$ and $\text{var}(\nu_i) = \lambda > 0$ so that the pdf of ν_i is given by

$$g(\nu_i) = (2\pi\lambda)^{-1/2} \nu_i^{-3/2} e^{-(\nu_i-1)^2/2\lambda\nu_i}.$$

The evaluation of the (marginalized) likelihood for the i th farm involves integrating out ν_i and is equal to

$$L_i = L(Y_i | \beta, \lambda) = \prod_{j=1}^{n_i} L_{ij}, \quad (2.1)$$

where

$$L_{ij} = \left\{ \frac{(\mu_{ij})^{y_{ij}}}{y_{ij}!} \right\} (2\pi\lambda)^{-1/2} e^{\lambda^{-1}} \int_0^\infty x^{y_i-3/2} e^{-x(\mu_i+1/2\lambda)-1/(2\lambda x)} dx,$$

where $\mu_i = \sum_{j=1}^{n_i} \mu_{ij}$.

From Gardshteyn and Ryzhik (1980, p. 970) the integral in (2.1) is of the form

$$\int_0^\infty e^{ax^h-bx^{-h}} x^{s-1} dx = (2/h)(b/a)^{s/2h} K_{s/h}(2\sqrt{ab}) \quad (2.2)$$

where $K_j(\cdot)$ is the modified Bessel function of the second kind. Therefore L_i can be written as

$$L_i = L(Y_i|\beta, \lambda) = \prod_{j=1}^{n_i} \left\{ \frac{(\mu_{ij})^{y_{ij}}}{y_{ij}!} \right\} \left(\frac{2}{\pi\lambda} \right)^{1/2} e^{1/\lambda} (1 + 2\lambda\mu_i)^{-s_i/2} K_{s_i}(z_i) \quad (2.3)$$

where $s_i = y_i - 1/2$, $z_i = (1 + 2\lambda\mu_i)^{1/2}/\lambda$, and $y_i = \sum_{j=1}^{n_i} y_{ij}$.

Define,

$$p(y) = c(\lambda)(1 + 2\lambda\mu)^{-(y-1/2)/2} K_{y-1/2}(\sqrt{1 + 2\lambda\mu}/\lambda).$$

Then,

$$p(y+1) = c(\lambda)(1 + 2\lambda\mu)^{-(2y+1)/4} K_{y+1/2}(\sqrt{1 + 2\lambda\mu}/\lambda).$$

where $c(\lambda) = (2/\pi\lambda)^{1/2} e^{1/\lambda}$.

The Bessel functions has the following important properties:

$$\begin{aligned} K_{1/2}(a) &= K_{-1/2}(a) = (\pi/2a)^{1/2} e^{-a} \\ K_{3/2}(a) &= (1 + 1/a)K_{1/2}(a). \end{aligned} \quad (2.4)$$

Therefore,

$$\frac{K_{y+1/2}(a)}{K_{y-1/2}(a)} = (1 + 2\lambda\mu)^{1/2} M(y) \quad (2.5)$$

where

$$M(y) = (y+1) \frac{p(y+1)}{p(y)}, \quad (2.6)$$

from which it follows that

$$M(0) = (1 + 2\lambda\mu)^{-1/2}.$$

Moreover,

$$p(0) = \exp \left[\frac{1}{\lambda} (1 - (1 + 2\lambda\mu)^{1/2}) \right].$$

The importance of equations (2.5) and (2.6) will be apparent when we derive the estimating equations of the model parameters in Appendix I. We will show that the partial derivatives of the log-likelihood function with respect to the model parameters do not depend on the Bessel functions, rather they depend on $M(y)$ which can be recursively computed.

To obtain the maximum likelihood estimators of the parameters we construct the total (marginalized) likelihood L , which is the product of k terms of equation (2.3).

$$L = \prod_{i=1}^k L_i.$$

The log-likelihood function $\ell = \log(L)$ is given by:

$$\begin{aligned} \ell = & \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} x_{ij}^T \beta + \frac{k}{\lambda} - (k/2) \log \lambda \\ & - \sum_{i=1}^k \left(\frac{2y_{i.} - 1}{4} \right) \log(1 + 2\lambda\mu_{i.}) + \sum_{i=1}^k \log K_{s_i}(z_i) \end{aligned} \quad (2.7)$$

In Appendix I, we provide the estimating equations for β and λ .

2.3 Estimating unknown parameters

As in the nonlinear mixed models, the maximum likelihood equations cannot be explicitly solved. It is natural to use the Newton-Raphson algorithm or any of its variants to estimate the parameter vector $\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \lambda)^T$. However, like other iterative techniques this method depends on the validity of a quadratic approximation to the surface of ℓ , and for many nonlinear mixed-effects models, straightforward approximations do not provide consistent estimates regardless of the number of clusters, or the number of observations per cluster (see, Demidenko 1997). Therefore, we needed an optimization algorithm that could efficiently handle this complicated log-likelihood function. After some exploration, we have chosen to use “fminunc” (unconstrained function minimization) routine of the Optimization Toolbox of the MATLABTM Software (The MathWorks, Inc., Natick, MA) to locate the minimum of the negative log-likelihood function. The routine implements a subspace trust region method which is based on the interior-reflective Newton method described in Coleman and Li (1994, 1996). Each iteration in this “large-scale optimization” algorithm involves the approximate solution of a large linear system using the method of reconditioned conjugate gradients.

The routine requires the user to provide an initial estimate for the parameter vector θ . Since PIGM reduces to Poisson regression model when $\lambda = 0$, we

obtained the initial estimate by fitting a Poisson regression model to mastitis data using SAS (Genmod procedure). The result of the optimization, i.e., log-likelihood maximization, was:

$$\hat{\theta} = (3.212, 0.130, -0.034, -0.816, -0.798, 0.392)^T$$

To ensure that the global maximum was located, we perturbed the initial parameter estimates and verified that this solution was indeed the optimum. The optimization routine also produced the Hessian matrix from which we calculated Fisher's information matrix, $I(\theta)$. The asymptotic standard errors of the estimated parameters are the square roots of the elements on the diagonal of $I(\theta)$.

3. Analysis of Mastitis Data

Clinical mastitis (CM) is one of the endemic diseases and conditions of dairy cattle in many countries. Such disease causes significant losses to the dairy industry both in terms of the reduction in output levels and wastage of resources incurred, in addition to the costs of disease prevention and treatment. Various studies, although followed different methods of assessments, have estimated direct costs associated with CM, but not the wider effects of disease, such as impact on human health, animal welfare and livestock markets. In the UK, Bennett *et al.* (1992) demonstrated that mastitis treatment costs are substantial, owing to the large number of cows requiring treatment and the relatively high costs of treatment (which includes milk withdrawal following antibiotic use). Therefore, effective control and prevention are functions of our ability to identify potential farm level management practices that affect the distribution of the disease.

In Table 2, we provide a summary fitting of the data using the PIGM, NBM and the generalized linear mixed model (GLMM) with log-link (Schall 1991, McGilchrist 1994). The estimating equations of the two models are given in Appendices I & II. The MATLAB codes for fitting the PIGM and the NBM are available from the authors. The GLIMMIX is a SAS macro (1996) that fits the GLMM. It uses the pseudo likelihood method for estimating parameters of the GLMM as proposed by Wolfinger and O'Connell (1993). In summary, the pseudo likelihood approach fits a linearized pseudo variable (i.e., a transformation of Y_{ij} onto to the linear scale using the weighted normal mixed model.) The name pseudo likelihood was used because the likelihood function maximized at each iteration, is that of the linearized (pseudo) variable and not that of the original data (Breslow and Clayton, 1993).

For the GLIMMIX, the residual log-likelihood, rather than the log-likelihood, is provided by SAS, and is not included in the summary. Note that, the only non-categorical covariate included in the model was the "number cultured." This covariate z_{ij} was standardized so that it is modeled as $z_{ij}^* = (z_{ij} - \bar{z}_i)/s_i$ where z_{ij}

is the number of animals cultured in the i th farm at the j th visit, \bar{z}_i is the average number of animals cultured in the i th farm, and s_i is the standard deviation. As can be seen, the PIGM and NBM models are equivalent as they comprise the same significant covariates and the values of their maximized log-likelihood functions are almost the same. However, there is a difference in the sign of the “bedding” covariate. This covariate has been coded in such a way that the code “1” is given to farms that used sands, and code “0” for farms that used organic material. Therefore a positive sign for the regression coefficient of bedding means that, for a randomly selected farm, the use of sand should increase the mean number of diseased animals (on the log-scale) on that farm, as was shown by the NBM and the GLIMMIX. This is in contrast to what the PIGM fit gives, which shows that the use of sand as a bedding material, should decrease the number mastitis cases. This is in support of what is widely known among veterinarians.

Although we found statistically significant relationship between CM and management practices, those management practices are not necessarily causally related to the CM on the farm. Clearly “postmilk” which is directly related to udder health, should be either stopped or preceded by proper preparation (e.g., through hand washing and use of disinfectants). Moreover, the risk of transmitting the pathogens may be substantially reduced if “tail-docking” is practiced. Although previous studies (Barkema *et al.* 1999) demonstrated strong association between types of bedding and the distribution of CM, such an association was absent in our data, as can be seen from Table 2. The justification is that in our data, only one farm did not use organic bedding, and that particular farm was visited three times only. Therefore, this lack of association may be attributed to insufficient data.

To test the adequacy of the models, we provide two sets of residual plots. As was advocated by Davison and Gigli (1989) we plot Pearson’s residuals r_{ij} ,

$$r_{ij} = \frac{y_{ij} - \hat{\mu}_{ij}}{\left[\hat{\mu}_{ij}(1 + \hat{\lambda}\hat{\mu}_{ij})\right]^{1/2}} \quad (3.1)$$

where $\hat{\mu}_{ij} = \exp(x_{ij}^T \hat{\beta})$, against the normal quantiles. Deviation of the plot of r_{ij} against the normal quantiles from a straight line, indicates that the residuals are not normally distributed. In Figure 1, we notice, for both models, a slight deviation from a straight line at the upper tails of the distribution as would be expected for count data.

We also provided plots of the “deviance residuals” for over-dispersed count data, defined as

$$d_{ij} = 2\text{sign}(y_{ij} - \mu_{ij}) \left[y_{ij} \log(y_{ij}/\hat{\mu}_{ij}) - (y_{ij} + \hat{\lambda}^{-1}) \log \left(\frac{y_{ij} + \hat{\lambda}^{-1}}{\hat{\mu}_{ij} + \hat{\lambda}^{-1}} \right) \right] \quad (3.2)$$

Table 2: Summary of the PIGM, NBM and GLIMMIX (SAS) analysis of the mastitis data.

Model	Term	Estimate	St. error	<i>p</i> -value
PIGM	intercept	3.212	0.820	0.000
	number-cultured	0.130	0.028	0.000
	bedding	-0.034	0.659	0.958
	postmilk	-0.816	0.397	0.040
	tailed	-0.798	0.267	0.002
	λ (dispersion parameter) log-likelihood (-1484)	0.392	0.093	0.000
NBM	intercept	3.195	0.468	0.000
	number-cultured	0.131	0.055	0.017
	bedding	0.120	0.374	0.748
	postmilk	-1.055	0.234	0.000
	tailed	-0.692	0.145	0.000
	α (dispersion parameter) log-likelihood (-1451)	0.447	0.054	0.000
GLIMMIX (SAS)	intercept	2.614	0.522	0.000
	number-cultured	0.128	0.001	0.000
	bedding	0.119	0.661	0.857
	postmilk	-1.068	0.523	0.041
	tailed	-0.676	0.424	0.111
	σ^2 (dispersion parameter)	0.402	—	—

(McCullagh and Nelder 1989, p.39 and Cameron and Trivedi 1998, p.142). Figure 2 shows the plots of d_{ij} versus the observed counts y_{ij} for both models. As can be seen, both models provide adequate fit to the data.

4. Discussion

During the last fifteen years, hierarchical statistical modeling of epidemiological data has been an area of intense research. Depending on the objectives of the investigation, an appropriate model is selected. If interest is focused on the efficient estimation of the parameters of the regression function on the expense of the other parameters that characterize the components of dispersion, then the Generalized Estimation Equation approach of Liang and Zeger (1986) would be an appropriate tool, and the regression coefficients would have a population

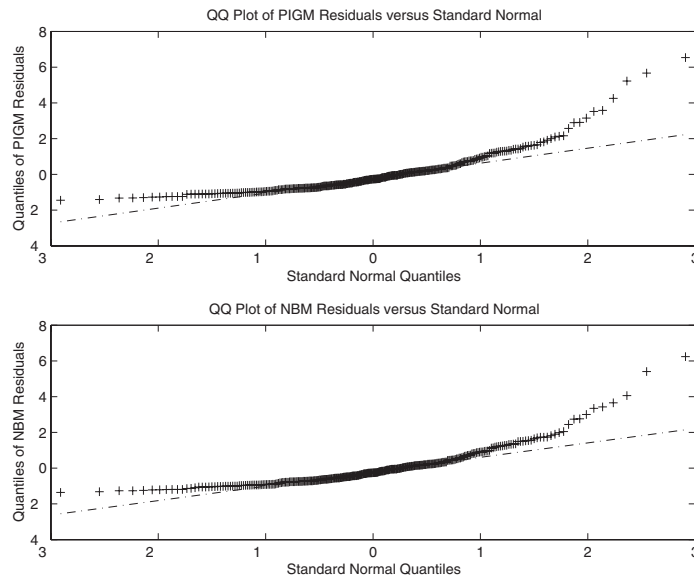


Figure 1. Upper panel: Quantiles of the Pearson's residuals of the PIGM against the normal quantiles. Lower panel: Quantiles of the Pearson's residuals of the NBM against the normal quantiles.

average interpretation. On the other hand, if in addition to the regression coefficients, the dispersion parameter is of interest, the cluster specific model is more appropriate. We would like to emphasize that had the main objectives of the study been the estimation of the disease prevalence, a binomial regression with over-dispersion parameter would be appropriate. However the main objective of the present project was the identification and estimation of the effects of the herd level covariates that influence the variability of the disease. Due to the fact that the number of cultured cows was large (excluding the sixth visit), and that the percentage of positive cows did not change significantly over time, we assumed, conditional on the herd, that the number of mastitis cases can be approximately modeled by the Poisson distribution, whose mean is a function of the herd level covariates, and an unobservable random effect.

The inverse-Gaussian distribution may be an attractive choice as a mixing distribution for models of count data that exhibit over dispersion caused by the hierarchical structure. When suitably parameterized it can be used as an alternative to the widely used gamma distribution. Although the use of the negative binomial, which results from mixing the Poisson with the gamma is long standing, the Poisson inverse-Gaussian can be used to analyze data with this structure

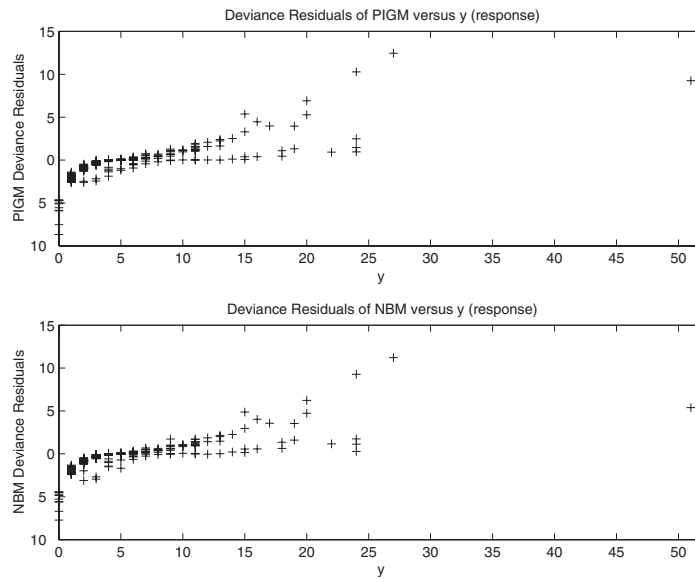


Figure 2. Upper panel: Deviance residuals of the PIGM versus the observed counts y_{ij} . Lower panel: Deviance residuals of the NBM versus the observed counts y_{ij} .

(see; Wilmot 1987). Moreover, we were able to fit the PIGM quite easily.

The GLIMMIX, although uses a general approximation to the integral in equation (2.1), to our surprise produced somewhat similar results to those of the PIGM and NBM. Evidence from the literature suggested that the GLIMMIX fitting procedure of Schall, McGilchrist and Breslow and Clayton are biased, and a full mixture model should be employed to analyze clustered data. We would like to point out that the difference in the sign of one of the coefficients (bedding) is not unexpected and has been reported in similar context by Lesaffre and Spiessens (2001) and Neuhaus and Segal (1997). Although the two models give close fits, common knowledge regarding the effect of the “bedding” factor, supports the selection of the PIGM over the NBM.

References

- Barkema, H. W., Schukken, Y .H., Lam, T. J., Beiboer, M. L., Benedictus, G. and Brand, A. (1999). Management practices associated with the incidence rate of clinical mastitis *Journal of Dairy Science*, **82**, 1643-1654.
- Benett, R. M., Christiansen, K., and Clifton-Hadley, R. S. (1999). Estimating the costs associated with endemic diseases of dairy cattle, *Journal of Dairy Research* **66**,

455-459.

- Breslow, N. E. (1984). Extra-Poisson variation in the log-linear models. *Applied Statistics* **33**, 38-44.
- Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *J. Amer. Statist. Assoc.*, **88**, 9-25.
- Burnett, R. T., Shedden, J. and Krewski, D. (1991). Nonlinear regression models for correlated count data. *Environmetrics* **3**, 211-222.
- Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge University Press.
- Coleman, T. F. and Li, Y. (1994). On the convergence of reflective Newton method for large-scale nonlinear minimization subject to bounds, *Mathematical Programming* **67**, No. 2, 189-224.
- Coleman, T. F. and Li, Y. (1996). An interior trust-region approach for nonlinear minimization subject to bounds. *SIAM Journal on Optimization* **6**, 418-445.
- Collings, B. J. and Margolin, B. H. (1985). Testing goodness of fit for the Poisson assumption when observations are not identically distributed. *J. Amer. Statist. Assoc.* **74**, 411-418.
- Cox, D. R. and HINKLEY, D. (1974). *Theoretical Statistics*. Chapman Hall.
- Cox, D. R. (1983). Some remarks on over-dispersion. *Biometrika* **70**, 269-274.
- Dean, C. B. (1992). Testing for over-dispersion in Poisson and binomial regression models. *J. Amer. Statist. Assoc.*, **87** 451-457.
- Demidenko, E. (1997). Asymptotic properties of nonlinear mixed-effects models. In *Modelling Longitudinal and Spatially Correlated Data* (Edited "Lecture Notes in Statistics", **122**, 49-62). Springer.
- Goldstein, H. (1991). Nonlinear multi-level models with an application to discrete response data. *Biometrika* **7**, 45-51.
- Gradshteyn, I. S. and Ryzhik, I. M. (1980). *Tables of Integrals*. Series and Products: Corrected and Enlarged Edition, Academic Press.
- Have, T. and Hartzel, T. (1995). Comparison of two approaches to analyzing correlated binary data in developmental toxicity studies. *Teratology* **52**, 267-276.
- Lesaffre, E. and Spiessens, B. (2001). On the effect of the number of quadrature points in a logistic random-effects model: An example. *Applied Statistics* **50**, part 3, 325-335.
- Liang, K. Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-122.
- Longford, N. T. (1994). Logistic regression with random coefficients. *Computational Statistics and Data Analysis* **17**, 1-15.

- Mc-Cullagh, P. and Nelder, J. A. (1989). *Generalized linear models*, 2nd ed. Chapman and Hall.
- Mc-Gilchrist, C. A. (1994). Estimation in generalized mixed models. *J. R. Statist. Soc., Series-B* **58**, 61-69.
- Molenberghs, G. and Geys, H. (2001). Multivariate clustered data analysis in developmental toxicity studies. *Statistica Neerlandica* **55**, 319-345.
- Moran, P. A. P. (1970). On asymptotically optimal tests of composite hypotheses. *Biometrika* **57**, 45-55.
- Neuhaus, J. M. and Segal, M. R. (1997). An assessment of approximate maximum likelihood estimators in generalized linear mixed models. In *Modelling Longitudinal and Spatially Correlated Data* (Edited "Lecture Notes in Statistics", **122**, 11-22). Springer.
- Neuhaus, J. M. (2001). Assessing change with longitudinal and clustered binary data, *Annual Review of Public Health*, **22**, 115-128.
- SAS Institute Inc. (1996). *SAS/STAT Software: Changes and Enhancements through Release 6.11*. Cary, NC, USA.
- Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika* **78**, 719-727.
- Stukel, T. A. (1993). Comparison of methods for the analysis of longitudinal interval count data. *Statistics in Medicine* **12**, 1339-1351.
- Thall, P. F. and Vail, S. C. (1990). Some covariance models for longitudinal count data with over-dispersion. *Biometrics* **46**, 657-671.
- Wilmot, G. E. (1987). The Poisson-Inverse Gaussian distribution as an Alternative to the negative binomial. *Scandinavian Actuarial Journal* 113-127.
- Wolfinger, R. and O'Connell, M. (1993). Generalized linear mixed models: A pseudo-likelihood approach. *J. Statist. Comput. Simul.* **48**, 233-243.
- Zeger, S. L., Liang, K. Y. and Albert, P. A. (1988). Models for longitudinal data: A generalized estimating equation approach. *Biometrics* **44**, 1049-1060.

Appendix I: Estimating Equations for PIGM

The log-likelihood function of the PIGM is given by

$$\begin{aligned} \ell = & \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij} \sum_{r=0}^p x_{ijr} \beta_r + (k/\lambda) - (k/2) \log \lambda \\ & - \sum_{i=1}^k \left(\frac{2y_{i\cdot} - 1}{4} \right) \log(1 + 2\lambda\mu_{i\cdot}) + \sum_{i=1}^k K_{s_i}(z_i) \end{aligned}$$

where $y_{i\cdot} = \sum_{j=1}^{n_i} y_{ij}$, $\mu_{i\cdot} = \sum_{j=1}^{n_i} \mu_{ij}$, $\mu_{ij} = \exp(\sum_{r=0}^p x_{ijr}\beta_r)$, $s_i = y_{i\cdot} - 1/2$ and $z_i = (1 + 2\lambda\mu_{i\cdot})^{1/2}/\lambda$. Then,

$$\frac{\partial \ell}{\partial \beta_r} = \sum_{i=1}^k \left[\sum_{j=1}^{n_i} y_{ij} x_{ijr} - \frac{\lambda \sum_{j=1}^{n_i} x_{ijr} \mu_{ij}}{1 + 2\lambda\mu_{i\cdot}} \left(\frac{2y_{i\cdot} - 1}{2} \right) + \frac{\partial \log K_{s_i}(z_i)}{\partial \beta_r} \right] \quad \text{and}$$

$$\frac{\partial \ell}{\partial \lambda} = -\frac{k}{\lambda^2} - \frac{k}{2\lambda} - \sum_{i=1}^k \frac{\mu_{i\cdot}}{1 + 2\lambda\mu_{i\cdot}} \left(\frac{2y_{i\cdot} - 1}{2} \right) + \sum_{i=1}^k \frac{\partial \log K_{s_i}(z_i)}{\partial \lambda}.$$

The recurrence relations

$$\begin{aligned} K_{s+1}(z) &= K_{s-1}(z) + \frac{2s}{z} K_s(z) \\ z \frac{\partial K_s(z)}{\partial z} &= -z K_{s+1}(z) + s K_s(z) \end{aligned} \quad (\text{A.1})$$

(see; Gradshteyn and Ryzhik 1980, p. 970) together with equations (2.5) and (2.6) can be used to show that

$$\frac{\partial \log K_{y-1/2}(z)}{\partial \lambda} = \frac{M(y)(1 + \lambda\mu)}{\lambda^2} - \frac{(1 + \lambda\mu)(y - 1/2)}{\lambda(1 + 2\lambda\mu)} \quad (\text{A.2})$$

where $z = (1 + 2\lambda\mu)^{1/2}/\lambda$. (Note that subscripts have been dropped for convenience.) Therefore,

$$\frac{\partial \ell}{\partial \lambda} = \sum_{i=1}^k [-\lambda^{-2} - y_{i\cdot} \lambda^{-1} + M(y_{i\cdot})(1 + \lambda\mu_{i\cdot}) \lambda^{-2}] \quad (\text{A.3})$$

Again, using the recurrence relations (A.1), and equations (2.5) and (2.6) we can show that

$$\frac{\partial \log K_{y-1/2}(z)}{\partial \beta_r} = \left[-M(y)(1 + 2\lambda\mu)^{1/2} + \frac{\lambda(y - 1/2)}{(1 + 2\lambda\mu)^{1/2}} \right] \frac{\mu x_r}{(1 + 2\lambda\mu)^{1/2}} \quad (\text{A.4})$$

Therefore,

$$\frac{\partial \ell}{\partial \beta_r} = \sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij} - M(y_{i\cdot}) \mu_{ij}] x_{ijr}. \quad (\text{A.5})$$

Clearly, the Bessel functions do not appear in the estimating equations and $M(y_{i\cdot})$ can be recursively computed when solving

$$\frac{\partial \ell}{\partial \lambda} = 0, \quad \frac{\partial \ell}{\partial \beta_r} = 0, \quad r = 0, 1, \dots, p$$

Appendix II: Estimating Equations for NBM

The log-likelihood function of the NBM is given by

$$\ell_{NB} = \sum_{i=1}^k \sum_{j=1}^{n_i} \left[y_{ij} \sum_{r=0}^p x_{ijr} \beta_r - (y_{ij} + \lambda^{-1}) \log(1 + \lambda \mu_{ij}) + \sum_{t=1}^{y_{ij}} \log(1 + \lambda(t-1)) \right].$$

Then,

$$\begin{aligned} \frac{\partial \ell_{NB}}{\partial \beta_r} &= \sum_{i=1}^k \sum_{j=1}^{n_i} \frac{x_{ijr}}{1 + \lambda \mu_{ij}} (y_{ij} - \mu_{ij}) \quad \text{and} \\ \frac{\partial \ell_{NB}}{\partial \lambda} &= \sum_{i=1}^k \sum_{j=1}^{n_i} \left[\frac{\log(1 + \lambda \mu_{ij})}{\lambda^2} - \frac{1 + \lambda y_{ij}}{\lambda(1 + \lambda \mu_{ij})} \mu_{ij} + \sum_{t=1}^{y_{ij}} \frac{t-1}{1 + \lambda(t-1)} \right]. \end{aligned}$$

Received September 3, 2002; accepted November 21, 2002.

M. M. Shoukri
 Department of Biostatistics, Epidemiology and Scientific Computing
 King Faisal Specialist Hospital and Research Center
 MBC-03 / P. O. Box 3354
 Riyadh, KSA 11211
 shoukri@kfshrc.edu.sa

M. H. Asyali
 Department of Biostatistics, Epidemiology and Scientific Computing
 King Faisal Specialist Hospital and Research Center
 MBC-03 / P. O. Box 3354
 Riyadh, KSA 11211
 asyali@kfshrc.edu.sa

R. VanDorp
 Department of Animal and Poultry Sciences
 University of Guelph
 Guelph, Ontario, N1G 2W1
 Canada

D. Kelton
 Department of Population Health
 University of Guelph
 Guelph, Ontario, N1G 2W1
 Canada