THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

ROBUSTNESS OF MAXIMUM LIKELIHOOD ESTIMATES FOR MIXED POISSON REGRESSION MODELS

by

Kimberly Weems and Paul J. Smith

Institute of Statistics Mimeograph Series No. 2537

August 2002

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

Mimeo Series # 2537
Robustness of Maximum
Likelihood Estimates for
Mixed Poisson Regression Models
by
Kimberly Weems and
Paul J. Smith
August 2002

Robustness of Maximum Likelihood Estimates for Mixed Poisson Regression Models

K.S. Weems

Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA

P.J. Smith

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

Abstract

Mixed Poisson regression models, a class of generalized linear mixed models, are commonly used to analyze count data that exhibit overdispersion. Because inference for these models can be computationally difficult, simplifying distributional assumptions are often made. We consider an influence function representing effects of infinitesimal perturbations of the mixing distribution. This function enables us to compute Gâteaux derivatives of maximum likelihood estimates (MLEs) under perturbations of the mixing distribution for Poisson–gamma and Poisson–lognormal models. Provided the first two moments exist, these MLEs are robust in the sense that their Gâteaux derivatives are bounded.

Key words: Generalized linear mixed models, Robustness, Influence functions, Gâteaux derivatives, Maximum likelihood estimation

1991 MSC: 62A10, 62F35, 62J12

1 Introduction

This paper investigates estimation in mixed Poisson regression models, an important class of generalized linear mixed models (GLMMs). For these models, the conditional distribution of the response is Poisson with a random mean that depends on the random effects. The Poisson model assumes that the mean is equal to the variance; however, for count data, the variance is usually larger than the mean. Such overdispersion is often attributed to unobserved heterogeneity in the linear predictor or positive correlation that exists between responses. This overdispersion may be accounted for by the use of random effects. The Poisson mixed model as a model for overdispersed count data has been studied by several authors, including Breslow (1984), Lawless (1987), Dean (1991), Dean et al. (1989), Yanez and Wilson (1995), Van de Ven and Weber (1995), and Hougaard et al. (1997).

When estimating the parameters, the mixing distribution of the random effects is often assumed to be normal for computational convenience. Our goal is to determine the robustness of the maximum likelihood estimates (MLEs) of the fixed effects and of the variance component when the mixing distribution is slightly contaminated. Specifically, the distribution of the random effects is assumed to be an ϵ-contamination of a member of a specified parametric family. This introduces misspecification of the marginal distribution of the response. We study the effects of this misspecification by using an infinitesimal approach.

Although the issue of robustness has been considered by several authors for linear mixed models, considerably less work has been done on robustness of

Although the issue of robustness has been considered by several authors for linear mixed models, considerably less work has been done on robustness of parameter estimates in GLMMs. Neuhaus et al. (1992) examined mixing distribution misspecification in logistic mixed models. Using an approach by White (1980) for general model misspecification, they showed that the parameter estimates are typically inconsistent under misspecification; however, the magnitude of the bias in the fixed effects is small. In addition, their simulation studies suggest that valid standard error estimates of the fixed effects can be obtained under misspecification.

Gustafson (1996) used an influence function approach (Hampel et al., 1986) (Huber, 1981) to examine the robustness of maximum likelihood estimates for certain conjugate mixture models under mixing distribution misspecification. Extending Gustafson's approach to include a regression structure in the mean, we develop a related, yet slightly more complex version of the influence function, which is explained in the next section.

The paper is organized as follows. In Section 2, we introduce our model, and discuss the influence function approach used to determine an estimate's robustness. Sections 3, 4, and 5 focus on robustness of MLEs in mixed Poisson regression models. We consider in detail two popular models: Poisson–gamma and Poisson–lognormal. For the first model, we are able to calculate the influence function of the MLEs explicitly and to perform a simulation study. However, because of the intractable integrals that are involved, we obtain only asymptotic results for the second model. All proofs are given in the appendix.

2 Methods and Model Specification

Before discussing our model, it is necessary to introduce some terminology and notation. Let Ψ be an estimating function. We define the functional $T(F)$ to be the solution (in θ) of

$$\int \Psi(x;\theta)F(dx) = 0.$$

Under general conditions, $T = T(X_1, X_2, \cdots, X_n)$ can be regarded as a functional $T(F_n)$ applied to the empirical cdf of the data $\mathbf{X} = (X_1, X_2, \cdots, X_n)$. Then $T(F_n)$ estimates $T(F)$, where F is the true distribution of the data. If the distribution of $\mathbf{X}$ is perturbed from F to $F_\epsilon = (1-\epsilon)F + \epsilon G$, then one might study $(\partial/\partial\epsilon)T(F_\epsilon)|_{\epsilon=0}$ as a measure of the sensitivity of T to departures from an assumed model distribution F. The quantity $\dot{T}(F;G) = (\partial/\partial\epsilon)T(F_\epsilon)|_{\epsilon=0}$ is the Gâteaux derivative of $T(F)$ in the direction G. We find $\dot{T}(F;G)$ by formal implicit differentiation of the equation

$$\int \Psi(x;T(F_\epsilon))[F + \epsilon(G - F)](dx) = 0, \tag{1}$$

which yields

$$\dot{T}(F;G) = \left[-\int \nabla_\theta \Psi(x;T(F))F(dx)\right]^{-1} \left[\int \Psi(x;T(F))G(dx)\right]. \tag{2}$$

It is appropriate to examine the quantity given in (2) because it gives a first-order approximation to the asymptotic bias in estimating θ that is introduced by the ϵ-contamination of F by a distribution G. That is, applying Taylor's theorem,

$$\begin{aligned} T(F_\epsilon) &= T(F) + \epsilon\, \dot{T}(F;G) + o(\epsilon) \\ &= \theta + \epsilon\, \dot{T}(F;G) + o(\epsilon). \end{aligned}$$

The following definition of robustness is used throughout this article.

Definition 1 *Let T be a functional (estimate). Suppose that the distribution function F is contaminated by an epsilon amount of a distribution G. Then T is robust against misspecification of F if*

$$\left[-\int \nabla_\theta \Psi(x;T(F))F(dx)\right]^{-1}\left[\int \Psi(x;T(F))G(dx)\right] \tag{3}$$

is bounded for all G.

We now specify our model as follows. For $i = 1, ..., n$, let Y_i denote the response variables, U_i denote the random effects, and X_i denote covariates. Suppose that the conditional distribution of $Y_i|X_i, U_i$ is Poisson with mean $U_i\mu_i$, where $\mu_i = \mu(X_i) = \exp(\beta_0 + \beta_1 X_i)$, with β_0 and β_1 as unknown regression parameters. Suppose also that U_i and X_i are independent. Note that this model specification is multiplicative on the Y_i scale.

Let F denote the nominal distribution function of the random effects U_i and f denote its corresponding density function. The mean and variance of U_i are 1 and τ, respectively. We are interested in estimating $\theta = (\beta_0, \beta_1, \tau)$ via maximum likelihood. The marginal density of Y_i is given by

$$P(Y_i = y_i; X_i, \theta) = \int_0^\infty \frac{(u\mu_i)^{y_i} \exp(-u\mu_i)}{y_i!} f(u)du, \tag{4}$$

with marginal mean and variance given by μ_i and $\mu_i(1 + \mu_i\tau)$, respectively.

By maximizing the log likelihood $l(\theta, \mathbf{Y}) = \log(L(\theta; \mathbf{Y}))$, where

$$L(\theta; \mathbf{Y}) = \prod_{i=1}^{n} P(Y_i = y_i; X_i, \theta),$$

$\mathbf{Y} = (Y_1, \cdots, Y_n)$, we can determine the MLE $\hat{\theta}$. Our model is formulated for a single regressor X_i. We note, however, that the case of multiple regressors is handled similarly.

Consider the estimating function $\Psi = \nabla_\theta l(\theta; \mathbf{Y})$. Then (3) can be written as $\int_0^\infty I^{-1}(\theta) s(\theta, u) G(du)$, where $I(\theta) = -\mathbb{E}\left[\nabla_\theta^{\otimes 2} l(\theta; \mathbf{Y})\right]$ is the Fisher information matrix and $s(\theta; u) = \mathbb{E}[\nabla_\theta l(\theta; \mathbf{Y}) | U_i = u]$ is the conditional expected score matrix. Here, all expectations are taken with respect to the nominal distribution function F. We can think of the integrand $I^{-1}(\theta) s(\theta, u)$ as an influence function for misspecification of the mixing distribution (Gustafson, 1996). By integrating this quantity, we capture the effect of a contaminating distribution G on the parameter estimate for θ. Rephrasing Definition 1, if

$$\int_0^\infty \mathbf{IF}(u; \hat{\theta}, F) G(du) = \int_0^\infty I^{-1}(\theta) s(\theta, u) G(du) \tag{5}$$

is bounded for all G, then $\hat{\theta}$ is robust. Notice that if $I(\theta)$ is well-behaved, then bounding the integral of the influence function reduces to bounding the integral of the conditional expected score matrix. We use the notation $\mathbf{IF}(u; \hat{\theta}, F)$ to represent the 3×1 matrix of influence functions for the individual parameter estimates. That is, $\mathbf{IF}(u; \hat{\theta}, F) = [IF(u; \hat{\beta}_0, F), IF(u; \hat{\beta}_1, F), IF(u; \hat{\tau}, F)]^T$.

The following assumptions are made throughout this article.

A1. Let $\mathbb{F} = \{F | F \text{ is a cdf on } (0, \infty), \int uF(du) = 1, \text{ and } \int u^2 F(du) = 1 + \tau\}$. To ensure identifiability of the parameters, we let the nominal distribution $F \in \mathbb{F}$ and, likewise, the contaminating distribution $G \in \mathbb{F}$.

A2. $n^{-1}I \longrightarrow I^*$ as $n \longrightarrow \infty$, where I^* is positive definite.

A3. The covariates $X_1, X_2, \cdots, X_n$ are an i.i.d. sample from a nondegenerate distribution whose support is a compact region of $\mathbb{R}^2$.

A4. The interchange of expectation and differentiation of the log likelihood and its derivatives is permitted.

Remark. We note that if $\int_0^\infty uG(du) = \infty$, then we have no robustness. Therefore, if $\mathbf{IF}(u; \hat{\theta}, F)$ has a linear component, the statistic is not fully robust, but it is robust against contamination satisfying assumption A1.

In the following sections, we use an influence function approach to determine the effect of G on the MLEs for the fixed effect parameters β_0 and β_1 and the random effect parameter τ. We consider the Poisson–gamma and Poisson–lognormal models. We note that the subscript i is often omitted to simplify notation.

3 Poisson–gamma Model

In this section, we examine a Poisson–gamma model, where the nominal mixing distribution F is gamma with shape parameter $1/\tau$ and scale parameter τ. Thus, we have

$$f(u) = \frac{1}{\tau^{1/\tau}\Gamma(1/\tau)} u^{1/\tau - 1} \exp(-u/\tau).$$

It is well–known that this mixture results in a negative–binomial distribution for Y. Let $\rho = \mu(1 + \mu\tau)^{-1}$. For the Poisson–gamma model, the influence function for $\hat{\theta}$ is given by

$$\mathbf{IF}(u;\hat{\theta},F) = I^{-1}(\theta)s(\theta;u)$$
$$= \begin{pmatrix} \mathbb{E}_X(\rho) & \mathbb{E}_X(X\rho) & 0 \\ \mathbb{E}_X(X\rho) & \mathbb{E}_X(X^2\rho) & 0 \\ 0 & 0 & i_{33} \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{E}_X[\rho(u-1)] \\ \mathbb{E}_X[X\rho(u-1)] \\ s_{31} \end{pmatrix} \quad (6)$$

where

$$i_{33} = \tau^{-4}\mathbb{E}_X \int_0^\infty \left\{\mathbb{E}_{Y|X,U}[\psi'(\tau^{-1}) - \psi'(Y+\tau^{-1})|X, U=u] - \tau^2\rho\right\} F(du),$$

$$s_{31} = -\tau^{-2}\mathbb{E}_X \left\{\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1}) - \psi(\tau^{-1})|X, U=u] - \tau\rho(u-1) - \log(1+\mu\tau)\right\},$$

and ψ is the digamma function. We consider the family of integrals $\int_0^\infty \mathbf{IF}(u;\hat{\theta},F)G(du)$.

Our first result follows from elementary matrix algebra and assumption A1.

Proposition 1 *We have*

$$\int_0^\infty IF(u;\hat{\beta}_0,F)G(du) = \int_0^\infty IF(u;\hat{\beta}_1,F)G(du) = 0.$$

Proposition 1 gives a very strong robustness result for $\hat{\beta}_0$ and $\hat{\beta}_1$. We conclude that contamination of the mixing distribution has practically no effect on the regression parameter estimates.

The influence function for $\hat{\tau}$ is a complicated expression which involves the digamma function ψ. From (6) we see that

$$IF(u;\hat{\tau},F) = \frac{s_{31}}{i_{33}},$$

which can be written explicitly as follows:

$$\frac{-\tau^2\mathbb{E}_X\left\{\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1})-\psi(\tau^{-1})|X,U=u]-\tau\rho(u-1)-\log(1+\mu\tau)\right\}}{\mathbb{E}_X\int_0^\infty\left\{\mathbb{E}_{Y|X,U}[\psi'(\tau^{-1})-\psi'(Y+\tau^{-1})|X,U=u]-\tau^2\rho\right\}F(du)}. \tag{8}$$

The following lemma by Lawless (1987), concerning ψ and ψ', will be used to analyze this expression.

Lemma 1 *(Lawless, 1987) Let Z be a random variable with possible values $0,1,2,\cdots$, and let $\mathbb{E}(Z)<\infty$. Then*

$$\mathbb{E}[\psi(Z+\tau^{-1})-\psi(\tau^{-1})]=\tau\sum_{j=0}^{\infty}\frac{1}{(1+\tau j)}\Pr(Z>j),$$

and

$$\mathbb{E}[\psi'(\tau^{-1})-\psi'(Z+\tau^{-1})]=\tau^2\sum_{j=0}^{\infty}\frac{1}{(1+\tau j)^2}\Pr(Z>j).$$

It follows from Lemma 1 that the integrand in the denominator of (8) is positive. Observe also that the denominator is not a function of u. Therefore, we focus our attention on the numerator, or more specifically, the quantity

$$\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1})-\psi(\tau^{-1})|X,U=u]. \tag{9}$$

Using Lemma 1, we now compute an upper bound for (9). We refer the reader to the appendix for its proof.

Lemma 2 *We have*

$$\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1})-\psi(\tau^{-1})|X,U=u]\leq\tau(u\mu-1+\exp(-u\mu)).$$

The next lemma gives a lower bound for (9), which comes immediately from the fact that ψ is increasing on $\mathbb{R}^+$. See Abramowitz and Stegun (1972) for specifics.

Lemma 3 *We have*

$$\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1})-\psi(\tau^{-1})|X,U=u]\geq 0.$$

We have the following theorem.

Proposition 2 *Let $\mu=\mu(X)$. The integral $\int_0^\infty IF(u;\hat{\tau},F)G(du)$ is bounded by the quantities*

$$\frac{\tau^2\mathbb{E}_X\log(1+\mu\tau)}{\mathbb{E}_X\int_0^\infty\left\{\mathbb{E}_{Y|X,U}[\psi'(\tau^{-1})-\psi'(Y+\tau^{-1})|X,U=u]-\tau^2\rho\right\}F(du)}$$

and

$$\frac{-\tau^2\mathbb{E}_X\left\{\tau(\mu-1+\exp(-\mu))-\log(1+\mu\tau)\right\}}{\mathbb{E}_X\int_0^\infty\left\{\mathbb{E}_{Y|X,U}[\psi'(\tau^{-1})-\psi'(Y+\tau^{-1})|X,U=u]-\tau^2\rho\right\}F(du)}.$$

We omit the proof to this result. However, it follows directly from Lemmas 2 and 3, by replacing the expression $\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1})-\psi(\tau^{-1})|X,U=u]$ in (8) with its appropriate bounds and then integrating with respect to G.

Propositions 1 and 2 show that the MLEs of a Poisson–gamma model are robust against mixing distribution misspecification.

4 A Simulation Study

In a small simulation study, we explored how well the theoretical results for a Poisson–gamma model describe the true performance of MLEs when the gamma mixing distribution is contaminated.

For $i=1,\cdots,1000$, we generated covariates X_i that follow a standard normal distribution and formed the regression structure $\mu_i=\beta_0+X_i\beta_1$, where $\beta_0=2$

and $\beta_1 = 2$. Next, we let F denote the (nominal) gamma mixing distribution and G denote the (contaminating) lognormal distribution. We generated random effects $U_i \sim (1-\epsilon)F + \epsilon G$ with a mean of 1 and a variance of $\tau = 1$ for $\epsilon = 0, 0.01, 0.05, 0.10, 1$. Then, Poisson responses Y_i were generated, conditionally on U_i and X_i, with a mean of $U_i\mu_i$.

Finally, using the same X_is, we simulated 1000 sets of Monte Carlo samples, estimating β_0, β_1, and τ via maximum likelihood. The MLEs were computed using the function `glm.nb` (Venables and Ripley, 1999) in Splus.

Table 1

Estimated parameter values under an assumed Poisson–gamma model. The true mixing distribution is $(1-\epsilon)F + \epsilon G$, where F is gamma and G is lognormal. The true parameter values are $(\beta_0,\beta_1,\tau) = (2,2,1)$. For a sample of size $n = 1000$, the average standard error of the estimates is 0.04.

	Parameter Estimates		
ϵ	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\tau}$
0	2.00	2.00	1.00
0.01	2.00	2.00	1.00
0.05	2.00	2.00	0.84
0.10	2.00	2.00	0.74
1	2.00	2.00	0.67

Observing Table 1, we found that for all amounts of contamination, the regression parameter estimates are unbiased. However, as ϵ increased, the bias in

τ increased significantly. This behavior is not surprising since the theoretical robustness results obtained for the regression parameters are much stronger than the one obtained for τ.

It is important to note the $\epsilon = 1$ case. This situation often occurs in practice. That is, the data follow a Poisson–lognormal distribution; however, the estimates are computed under an assumed Poisson–gamma distribution. Even under complete mixing distribution misspecification, we still obtained unbiased estimates for the regression parameters. The variance component estimate, however, showed significant bias.

We also compared Fisher's information and the Monte-Carlo sample estimates for the variance of the MLEs. For $\epsilon \leq 0.10$, we observed reasonably good agreement between these variance estimates. When $\epsilon = 1$, Fisher's information tended to underestimate the variance of the MLEs, implying that inference based on the MLEs under complete mixing distribution misspecification may be unreliable.

5 Poisson–lognormal Model

Let $\sigma^2 = \log(1 + \tau)$. The lognormal density is given by

$$f(u) = \frac{1}{u\sigma(2\pi)^{1/2}} \exp\left(-\frac{(\log u + \sigma^2/2)^2}{2\sigma^2}\right).$$

Thus, the Poisson–lognormal probabilities are given by

$$P_y = \frac{\mu^y}{y!\sqrt{2\pi\sigma^2}} \int_0^\infty u^{y-1} \exp\left(-\frac{(\log u + \sigma^2/2)^2}{2\sigma^2}\right) \exp(-u\mu) du. \qquad (10)$$

Unlike the Poisson–gamma distribution, the Poisson–lognormal marginal probabilities cannot be written in a closed form. However, Shaban (1988) gives a general method to compute the first derivatives of the log-likelihood function. We use these results to compute the conditional expected score matrix $s(\theta; u)$ whose entries are given below:

$$s(\beta_0; u) = \mathbb{E}_X \mathbb{E}_{Y|X,U} \left\{ P_Y^{-1}(YP_Y - (Y+1)P_{Y+1}) \Big| X, U = u \right\},$$
$$s(\beta_1; u) = \mathbb{E}_X \mathbb{E}_{Y|X,U} \left\{ XP_Y^{-1}(YP_Y - (Y+1)P_{Y+1}) \Big| X, U = u \right\}$$

and

$$s(\tau; u) = (1+\tau)^{-1} \mathbb{E}_X \mathbb{E}_{Y|X,U} \{ P_Y^{-1}[(Y^2 - Y/2)P_Y - (Y+1)(2Y+1/2)P_{Y+1} + (Y+1)(Y+2)P_{Y+2}] | X, U = u \}.$$

Since $\mathbf{IF}(u; \hat{\theta}, F)$ is a linear combination of the terms of $s(\theta; u)$, we focus on bounding the integrals of these terms. However, notice that the terms in the score matrix $s(\theta; u)$ involve the ratios P_{y+1}/P_y and P_{y+2}/P_y. We first determine the behavior of these ratios before analyzing $\mathbf{IF}(u; \hat{\theta}, F)$.

Let $w = \log(u)$ and $v = y - 1/2$. We rewrite (10) as

$$P_y = \frac{\mu^y}{y!} \exp\{\sigma^2 y(y-1)/2\} H(y), \tag{11}$$

where

$$H(y) = (2\pi\sigma^2)^{-1/2} \int_{-\infty}^{\infty} \exp\left\{ -\frac{1}{2\sigma^2}(\sigma^2 v - w)^2 - \mu e^w \right\} dw. \tag{12}$$

It follows that

$$\frac{P_{y+1}}{P_y} = \frac{\mu}{y+1} \exp\{\sigma^2 y\} \frac{H(y+1)}{H(y)} \tag{13}$$

and

$$\frac{P_{y+2}}{P_y} = \frac{\mu^2}{(y+1)(y+2)} \exp\{\sigma^2(2y+1)\} \frac{H(y+2)}{H(y)}. \tag{14}$$

To gain insight into the behavior of the Poisson–lognormal probability ratios, we study the behavior of $H(y+\rho)/H(y)$, for ρ fixed. We use several results of de

Bruijn (1953), who studied the asymptotic properties of a function similar to H using saddlepoint techniques. From his methods, we show that asymptotically the ratios $H(y+\rho)/H(y)$ behave like $\exp\{-\rho\sigma^2 y\}$.

Theorem 1 *Let ρ be fixed. As $y \longrightarrow \infty$,*

$$\log \frac{H(y+\rho)}{H(y)} = -\rho\left\{\sigma^2 y - \log y + \frac{\log y}{\sigma^2 y}\right\} + O\left(\frac{1}{y}\right). \tag{15}$$

Corollary 1 *There exist constants y_0 and C that depend only on μ and σ^2, such that if $y > y_0$,*

$$\frac{H(y+\rho)}{H(y)} \leq C \exp\{-\rho\sigma^2 y\}. \tag{16}$$

Briefly, we sketch the proof of Theorem 1. Using a power series expansion, an asymptotic representation of H is given in terms of its saddlepoint. Next, we obtain a first–order Taylor series approximation for the saddlepoint, and we substitute this approximation into the H expansion. Finally, an asymptotic representation of the ratios of H functions is obtained. For more details of de bruijn's results and the proof to Theorem 1, see the appendix.

We now state our main result for this section.

Theorem 2 *The integral of the conditional expected score matrix that has entries given by*

$$\int_0^\infty \mathbb{E}_X \mathbb{E}_{Y|X,U}\left\{P_Y^{-1}(YP_Y - (Y+1)P_{Y+1})\Big| X, U = u\right\} G(du), \tag{17}$$

$$\int_0^\infty \mathbb{E}_X \mathbb{E}_{Y|X,U}\left\{XP_Y^{-1}(YP_Y - (Y+1)P_{Y+1})\Big| X, U = u\right\} G(du), \tag{18}$$

and

$$
\begin{aligned}
(1+\tau)^{-1}\int_0^\infty \mathbb{E}_X\mathbb{E}_{Y|X,U}\Big\{&(Y^2-Y/2)-(Y+1)(2Y+1/2)\frac{P_{Y+1}}{P_Y}\\
&+\,(Y+1)(Y+2)\frac{P_{Y+2}}{P_Y}\Big|\,X,U=u\Big\}\,G(du), \qquad (19)
\end{aligned}
$$

is uniformly bounded over $G \in \mathbb{F}$.

The proof of this result, given in the appendix, follows mainly from the relationships established in equations (13) and (14), along with Corollary 1.

It follows from Theorem 2 that $\int_0^\infty \mathbf{IF}(u;\hat{\theta},F)G(du)$ is also uniformly bounded in G. Therefore, we conclude that the MLE $\hat{\theta}$ for the Poisson–lognormal model is robust against mixing distribution misspecification.

6 Conclusions

We have focused on the effects of mixing distribution misspecification on MLEs in mixed Poisson regression models. Extending the influence function approach of Hampel et al. (1986) to the unobservable random effects, we computed bounds for the Gâteaux derivatives of regression parameter estimates and the variance component estimate. For the Poisson–gamma and the Poisson–lognormal models, the MLEs are robust against small perturbations of the mixing distribution, provided that the first two moments exist. These results were obtained by using properties of the digamma function and saddlepoint approximations.

In a limited simulation study, the robustness of the regression estimates was verified for misspecified Poisson–gamma models. However, the robustness of the variance component estimate was verified only for small perturbations from the nominal model. We conclude that, in practice, the modeling assumptions do not have a substantial effect on regression parameter estimates. We suggest considering an alternative estimation method for the variance component if the modeling assumptions may be incorrect.

7 Acknowledgments

The work of K.S. Weems was partially supported by a fellowship from the National Physical Science Consortium. She would like to thank Professor D. Boos for his helpful comments and suggestions.

A Proof for Poisson-gamma Model

Proof of Lemma 2 By using Lemma 1 and conditioning on X and U, we have that

$$\begin{aligned}
&\mathbb{E}_{Y|X,U}[\psi(Y+\tau^{-1}) - \psi(\tau^{-1})|X, U=u] \\
&\quad = \tau \sum_{j=0}^{\infty} \frac{1}{(1+\tau j)} \Pr(Y > j|X, U=u) \\
&\quad \leq \tau \sum_{k=1}^{\infty} \Pr(Y = k|X, U=u) \int_0^{k-1} \frac{dj}{(1+\tau j)} \\
&\quad = \tau u\mu \sum_{k=1}^{\infty} \Pr(Y = k-1|X, U=u) - \tau \sum_{k=1}^{\infty} \Pr(Y=k|X,U=u) \\
&\quad = \tau(u\mu - 1 + \exp(-u\mu)).
\end{aligned}$$

B Proofs for Poisson-lognormal Model

De bruijn used a series of Taylor expansions and substitutions to prove several asymptotic results concerning the saddlepoint of the H function given in (12). We present, without proof, the ones that are very important to this paper. All of these results can be found in de Bruijn (1953).

We first introduce some notation. Let $v = y - 1/2$. Let $\lambda = \mu\sigma^2 e^{\zeta} = \sigma^2 v - \zeta$, where ζ is the saddlepoint of (12).

Lemma 4 *We have*

$$\log H(y) = -\frac{\lambda^2 + 2\lambda}{2\sigma^2} - \frac{1}{2}\log\lambda, \tag{B.1}$$

uniformly for $|\lambda| \longrightarrow \infty, |\arg\ \lambda| < 3\pi/4 - \delta,$ *where* $\delta > 0$.

To understand how (B.1) depends explicitly on y, it is necessary to examine λ.

Lemma 5 *Let* $\alpha = \mu\sigma^2$ *and* $v' = v/\mu$. *As* $y \longrightarrow \infty$

$$\lambda = \alpha v' - \log v' + \frac{\log v'}{\alpha v'} + \frac{\log^2 v'}{2\alpha^2 v'^2} - \frac{\log v'}{\alpha^2 v'^2} + O\left(\frac{\log^3 v'}{v'^3}\right), \tag{B.2}$$

$$\lambda^2 = \alpha v'^2 - 2\alpha v' \log v' + 2\log v' + \log^2 v' - \frac{\log^2 v'}{\alpha v'} - 2\frac{\log v'}{\alpha v'} + O\left(\frac{\log^3 v'}{v'^2}\right), \tag{B.3}$$

and

$$\log\lambda = \log v' + \log\alpha - \frac{\log v'}{\alpha v'} + O\left(\frac{\log^3 v'}{v'^2}\right). \tag{B.4}$$

We use de Bruijn's results to prove the two theorems from Section 5.

Proof of Theorem 1 Combining the results from Lemmas 4 and 5 gives

$$\log H(y) = -\mu\left\{\frac{\sigma^2 v^2}{2\mu} - \frac{v}{\mu}\log\left(\frac{v}{\mu}\right) + \frac{v}{\mu} + \frac{\log^2(v/\mu)}{2\mu\sigma^2} - \frac{\log^2(v/\mu)}{2\mu\sigma^4 v}\right\}$$
$$- \frac{1}{2}\left\{\log\left(\frac{v}{\mu}\right) + \log(\mu\sigma^2) - \frac{\log(v/\mu)}{\sigma^2 v}\right\} + O\left(\frac{\log^3 v}{v^2}\right). \quad \text{(B.5)}$$

Next, for ρ fixed, consider $\log H(y+\rho)$, which contains terms involving powers of $\log[(v+\rho)/\mu] = \log(v/\mu) + \rho/v + o(1/v)$. Then, after tedious calculations,

$$\log\frac{H(y+\rho)}{H(y)} = -\mu\rho\left\{\frac{\sigma^2 v}{\mu} - \frac{1}{\mu}\log\left(\frac{v}{\mu}\right) + \frac{\log(v/\mu)}{\mu\sigma^2 v} + \frac{1}{2\mu v}\right\}$$
$$- \mu\rho^2\left\{\frac{\sigma^2}{2\mu} + \frac{1}{\mu v}\right\} + o\left(\frac{1}{v}\right). \quad \text{(B.6)}$$

Recall that $v = y - 1/2$, and consider $\log(y - 1/2) = \log y - 1/2y + o(1/y)$. Then (B.6) becomes

$$\log\frac{H(y+\rho)}{H(y)} = -\rho\left\{\sigma^2 y - \log y + \frac{\log y}{\sigma^2 y} + \left(1 - \frac{\log\mu}{\sigma^2}\right)\frac{1}{y} + o\left(\frac{1}{y^2}\right)\right.$$
$$\left. - \frac{\sigma^2}{2} + \log\mu\right\} + \rho^2\left\{\frac{1}{y} + \frac{\sigma^2}{2}\right\} + o\left(\frac{1}{y}\right)$$
$$= -\rho\left\{\sigma^2 y - \log y + \frac{\log y}{\sigma^2 y}\right\} + O\left(\frac{1}{y}\right)$$

as claimed.

Proof of Theorem 2

Here, we prove only the boundedness of the integral for β_0. The proofs for the β_1 and τ terms are similar.

Let $\mu = \mu(X)$. Using (13), consider the expectation

$$\begin{aligned}
&\mathbb{E}_X \mathbb{E}_{Y|X,U} \left\{ P_Y^{-1}(YP_Y - (Y+1)P_{Y+1}) \middle| X, U = u \right\} \\
&= \mathbb{E}_X \left\{ u\mu - \sum_{y=0}^{y_0} \mu \exp(\sigma^2 y) \frac{H(y+1)}{H(y)} p(y|u) \right. \\
&\qquad \left. - \sum_{y=y_0+1}^{\infty} \mu \exp(\sigma^2 y) \frac{H(y+1)}{H(y)} p(y|u) \right\}, \qquad \text{(B.7)}
\end{aligned}$$

where $p(y|u)$ is the Poisson mass function with mean $u\mu$. Consider the integrals (with respect to G) of the three terms in braces individually.

It is clear from Assumption A1 that $\int_0^\infty u\mu G(du) = \mu$. For the second term, we have

$$\begin{aligned}
&\sum_{y=0}^{y_0} \mu \exp(\sigma^2 y) \frac{H(y+1)}{H(y)} \frac{\mu^y}{y!} \int_0^\infty u^y \exp(-u\mu) G(du) \\
&\qquad \leq \sum_{y=0}^{y_0} \mu \exp(y(\sigma^2 - 1)) \frac{H(y+1)}{H(y)} \frac{y^y}{y!},
\end{aligned}$$

where we have used simple calculus to bound $u^y \exp(-u\mu)$ by $(y/\mu)^y \exp(-y)$. Using Corollary 1, we find an upper bound for the integral of the third term that depends only on μ and σ. Substituting for the three terms in (B.7) gives the desired uniform bound.

References

Abramowitz, M. and I. Stegun (1972). *Handbook of Mathematical Functions.* Dover, New York.

Breslow, N.E. (1984). Extra Poisson variation in log-linear models. *Applied Statistics.* 33, 38–44.

de Bruijn, N.G. (1953). The difference–differential equation $F'(x) = e^{\alpha x+\beta}F(x-1)$, I, II. *Indagationes Mathematicae.* 15, 449–458, 459–464.

Dean, C.B. (1991). Estimating equations for mixed Poisson models. In: V.P. Godambe, Ed., *Estimating Functions,* Clarendon, Oxford, 35–46.

Dean, C.B., J.F. Lawless, and G.E. Willmot (1989). A mixed Poisson–inverse Gaussian regression model. *The Canadian Journal of Statistics.* 17, 171–181.

Gustafson, P. (1996). The effect of mixing-distribution misspecification in conjugate mixture models. *The Canadian Journal of Statistics.* 24, 307–318.

Hampel F.R., E.M. Ronchetti, P.J. Rousseeuw and W.A. Stahel (1986). *Robust Statistics: The Approach Based on Influence Functions.* Wiley, New York.

Hougaard, P., M.-L. Lee, and G.A. Whitmore (1997). Analysis of overdispersed count data by mixtures of Poisson variables and Poisson processes. *Biometrics.* 53, 1225–1238.

Huber, P.J. (1981). *Robust Statistics.* Wiley, New York.

Lawless, J.F. (1987). Negative binomial and mixed Poisson regression. *The Canadian Journal of Statistics.* 15, 209–225.

Neuhaus, J.M., W.W. Hauck, and J.D. Kalbfleisch (1992). The effects of mixture distribution misspecification when fitting mixed-effects logistic models. *Biometrika.* 79, 755–762.

Shaban, S.A. (1988). Poisson-Lognormal Distributions. In: E. Crow and K. Shimizu, Eds., *Lognormal Distributions: Theory and Applications,* Marcel Dekker, New York, 195–210.

Van de Ven, R. and N.C. Weber (1995). Log-linear models for mean and dispersion in mixed Poisson regression models.. *Australian Journal of Statistics.* 37(2), 205–216.

Venables, W.N. and B.D. Ripley (1999). *Modern Applied Statistics with S-PLUS.* Springer, New York.

Yanez III, N.D. and J. Wilson (1995). Comparison of quasi-likelihood models for overdispersion. *Australian Journal of Statistics.* 37, 217–231.

White, H. (1980). Nonlinear regression on cross-sectional data. *Econometrica.* 48, 721–746.