Lifetime Data Anal (2007) 13:533–544
DOI 10.1007/s10985-007-9071-z

Estimation using penalized quasilikelihood and quasi-pseudo-likelihood in Poisson mixed models

Xihong Lin

Received: 2 October 2007 / Accepted: 14 November 2007 / Published online: 16 December 2007

Abstract We consider two estimation schemes based on penalized quasilikelihood and quasi-pseudo-likelihood in Poisson mixed models. The asymptotic bias in regression coefficients and variance components estimated by penalized quasilikelihood (PQL) is studied for small values of the variance components. We show the PQL estimators of both regression coefficients and variance components in Poisson mixed models have a smaller order of bias compared to those for binomial data. Unbiased estimating equations based on quasi-pseudo-likelihood are proposed and are shown to yield consistent estimators under some regularity conditions. The finite sample performance of these two methods is compared through a simulation study.

Keywords Asymptotic bias · Estimating equations · Generalized linear mixed models · Laplace expansion · Overdispersion · Variance components

1 Introduction

Poisson models have a lengthy history as models for analyzing count and rate data, which arise frequently in epidemiology, etiology, toxicology, and other biomedical fields. A common complication is that the observed variation of counts substantially exceeds that attributable to the Poisson distribution. This overdispersion must be taken into account to make valid statistical inferences.

A classical approach to this problem is to model the extra-Poisson variation by assuming that the means are drawn from a gamma distribution. This leads to a negative binomial model (McCullagh and Nelder 1989). Alternatively, a log-normal distribution for the means has been advocated (Hinde 1982; Morton 1987). Generalized

X. Lin (✉)
Department of Biostatistics, Harvard School of Public Health, Boston, MA 02115, USA
e-mail: xlin@hsph.harvard.edu

linear mixed models (Breslow and Clayton 1993) provide a general framework to model multiple sources of random variation for overdispersed and correlated Poisson outcomes. In view of the often intractable numerical integration required by a full likelihood analysis, Breslow and Clayton (1993) proposed a Laplace-based penalized quasilikelihood (PQL) method to estimate regression coefficients and variance components. A key feature of the PQL approach is that it can be easily implemented by iteratively fitting a linear mixed model to a modified dependent variable. Similar approaches have been proposed using alternative arguments by Stiratelli et al. (1984), Schall (1991), Liu and Pierce (1993), among others. Despite its simplicity, the performance of PQL in Poisson mixed models has not been evaluated. As an alternative to these likelihood based approaches, estimating equations using the quasilikelihood/method of moments (QL/M) have also been constructed to deal with overdispersion by Breslow (1984); Breslow (1990), Heagerty and Lele (1998), Lawless (1987), Morton (1987), Thall and Vail (1990), and others.

Although maximum likelihood estimation using numerical integration can be performed for fitting mixed models for correlated Poisson data, e.g., using SAS NLMIXED, due to its computational burden, the use of such a procedure is limited in practice to clustered/longitudinal data with a small number of random effects and a small/moderate sample size. For example, it is difficult to apply such a numerical MLE procedure to spatial Poisson count data, where the number of spatial random effects is large and requires a high-dimensional integration in calculating MLEs. It is hence desirable to study alternative computational easier estimation procedures.

In this article, we use the results of Breslow and Lin (1995)and Lin and Breslow (1996) to study the asymptotic bias in regression coefficients and variance components estimated by PQL in Poisson mixed models. We show the PQL estimators of both regression coefficients and variance components in Poisson mixed models have a smaller order of bias compared to those for binomial data. Following the work of Davidian and Carroll (1987), quasi-pseudo-likelihood (QL/PL) based estimating equations are proposed and are shown to yield consistent estimators. The performance of these two methods is compared through simulation.

The rest of the paper is organized as follows. Section 2 describes Poisson mixed models. Section 3 presents penalized quasilikelihood (PQL) estimation and studies the asymptotic bias of the PQL estimators of regression coefficients and variance components in Poisson mixed models. Section 4 proposes the quasi-pseudo-likelihood estimation procedure and studies its asymptotic properties in Poisson mixed models. Section 5 evaluates the finite sample performance of these two estimation procedures using simulation studies. Section 6 describes an extension using REML estimation, followed by discussions in Sect. 7.

2 The Poisson mixed model

Denote by (y_i, x_i, z_i) $(i = 1, \ldots, n)$ the observed data, where each y is a count or a rate, each x is a vector of explanatory variables associated with fixed effects, and each z is a vector of explanatory variables associated with random effects. Conditional on c unobservable random effects $b = (b_1^T, \ldots, b_c^T)^T$, the observations y_i are assumed to be independent with means μ_i^b and variances $\phi a_i^{-1}\mu_i^b$, where ϕ is a scale parameter and a_i is a weight, and to follow a log-linear model

$$\log \mu^b = X\alpha + Z_1 b_1 + \cdots + Z_c b_c, \tag{1}$$

where $\mu^b = (\mu_1^b, \ldots, \mu_n^b)^T$, $X = (x_1^T, \ldots, x_n^T)^T$, and $Z_j = (z_{1j}^T, \ldots, z_{nj}^T)^T$ is the covariate matrix associated with the jth random effect b_j of length $q_j \times 1$ ($j = 1, \ldots, c$). The elements of the matrices Z_j are typically 0 and 1. Model (1) encompasses both nested and crossed designs. Its wide biological and biomedical applications have been demonstrated by Morton (1987), Thall and Vail (1990), and Breslow and Clayton (1993). For example, Morton (1987) presented data on trap catches of insects, where the numbers of Australian bush flies caught by each of 10 traps were counted every hour during an 8-h-period each day over nine consecutive days. The aim was to study how the catch rate was affected by various environmental variables, including temperature, windspeed, radiation and humidity. The author considered a two-level mixed model with nested trap and period random effects.

Denote $\theta = (\theta_1, \ldots, \theta_c)^T$ and $D(\theta) = \text{diag}(\theta_1 I_{q_1}, \ldots, \theta_c I_{q_c})$, where I_{q_j} is a $q_j \times q_j$ identity matrix. Assuming that the random effects b_j are independent and are distributed as $N(0, \theta_j I_{q_j})$ the integrated likelihood of (α, θ) is

$$L(\alpha, \theta) = \exp\{\ell(\alpha, \theta)\} \propto |D|^{-\frac{1}{2}} \int \exp\left\{\sum_{i=1}^n \ell_i(\alpha; b) - \frac{1}{2} b^T D(\theta) b\right\} db, \tag{2}$$

where, apart from a constant,

$$\ell_i(\alpha; b) = \exp\left\{\frac{a_i}{\phi}\left(y_i \ln \mu_i^b - \mu_i^b\right)\right\}$$

defines the conditional Poisson loglikelihood.

Although full likelihood inference is complicated by no apparent simplification of the multiple integrals, the marginal mean μ and covariance matrix V of the outcome vector $y = (y_1, \ldots, y_n)^T$ can be easily obtained as follows:

$$\begin{aligned} \mu_i &= E(y_i) = \exp\left\{x_i^T\alpha + \frac{1}{2} z_i^T D(\theta) z_i\right\} \\ V_{ii} &= \text{var}(y_i) = \phi a_i^{-1}\mu_i + \mu_i^2\left[\exp\left\{z_i^T D(\theta) z_i\right\} - 1\right] \\ V_{ii'} &= \text{cov}(y_i, y_{i'}) = \mu_i \mu_{i'}\left[\exp\left\{z_i^T D(\theta) z_{i'}\right\} - 1\right] \quad (i \neq i'), \end{aligned} \tag{3}$$

where $z_i = (z_{i1}^T, \ldots, z_{ic}^T)^T$. It follows that the random effects only result in an offset, $z_i^T D(\theta) z_i / 2$, in the marginal mean (Zeger et al. 1988).

3 Penalized quasilikelihood

3.1 Estimation under penalized quasilikelihood

Applying the Laplace approximation to the integrated likelihood (2) and postulating that the modified generalized linear model (GLM) weights vary slowly with α, Breslow and Clayton (1993) estimated (α, b) for fixed θ by jointly maximizing Green's (1987) penalized quasilikelihood (PQL)

$$\sum_{i=1}^{n} \ell_i(\alpha; b) - \frac{1}{2} b^T D^{-1} b.$$

Denoting by $\tilde{b}$ the solution to $Z^T r_b - D^{-1} b = 0$, where r_b is a $n \times 1$ vector of residuals $a_i(y_i - \mu_i^b)/\phi$, the estimating equations of the PQL estimator $\hat{\alpha}_P$ of α are

$$X^T \tilde{r} = 0, \tag{4}$$

where $\tilde{r}$ represents r_b evaluated at $\tilde{b}$. Equivalently, $\hat{\alpha}_P$ iteratively solves

$$(X^T \tilde{V}^{-1} X)\alpha = X^T \tilde{V}^{-1} Y, \tag{5}$$

where $\tilde{V} = \tilde{W}^{-1} + ZDZ^T = \tilde{W}^{-1} + \sum_{j=1}^{c} \theta_j Z_j Z_j^T$, $\tilde{W} = \mathrm{diag}(a_i \mu_i^{\tilde{b}}/\phi)$, and Y is the working vector with components $Y_i = x_i^T \alpha + z_i^T \tilde{b} + a_i(y_i - \mu_i^{\tilde{b}})/\left(\phi \mu_i^{\tilde{b}}\right)$.

Estimation of θ under PQL proceeds by approximating Y by a multivariate normal distribution with covariance matrix $\tilde{V}$, which is a function of θ. Since we are interested in asymptotic bias, we restrict our attention to the maximum likelihood (ML) estimating equations, rather than the restricted maximum likelihood (REML) version. The estimating equations of θ are

$$-\frac{1}{2}\mathrm{tr}\left(\tilde{V}^{-1}\frac{\partial \tilde{V}}{\partial \theta_j}\right) + \frac{1}{2}(Y - X\hat{\alpha}_P)^T \tilde{V}^{-1} \frac{\partial \tilde{V}}{\partial \theta_j} \tilde{V}^{-1}(Y - X\hat{\alpha}_P) = 0. \tag{6}$$

Note that the dependence of $\tilde{W}$ on θ is ignored in calculating $\partial \tilde{V}/\partial \theta_j$ and $\partial \tilde{V}/\partial \theta_j$ is thus equal to $Z_j Z_j^T$. The joint solution to (5) and (6), is defined as the PQL estimator and is denoted by $(\hat{\alpha}_P, \hat{\theta}_P)$. Note that the scale parameter ϕ can be estimated together with θ (Breslow and Clayton 1993).

3.2 Asymptotic bias in PQL estimators

Without loss of generality, we assume the following regularity conditions throughout this section in our asymptotic analysis:

(I) The design matrix X contains an intercept.
(II) The matrices $Z_j (j = 1, \ldots, c)$ are standard design matrices with only zeros and ones and exactly one 1 in each row and at least one 1 in each column. The ith row of Z_j thus serves to indicate which level of the jth random effect b_j enters into the equation for the ith data point.

Conditional (I) is often satisfied in practical regression problems. Conditional (II) is applicable to several longitudinal data settings, such as longitudinal data with random intercepts and time-point specific random effects. They are also applicable to hierarchical data, and spatial data, such as the adjacent neighborhood correlation structure considered in Breslow and Clayton (1993). Condition (II) does not include longitudinal data with random linear slopes. Note however, that general asymptotic bias results of PQL of Lin and Breslow (1996) do not require condition (II), including general Poisson mixed models. Only the simplified asymptotic bias results below for the Poisson data assume condition (II).

Following Solomon and Cox (1992), Breslow and Lin (1995) and Lin and Breslow (1996) proposed a quadratic approximation to the marginal loglikelihood $\ell(\alpha, \theta)$ in (2) about $\theta = 0$ for small values of θ. Relating the PQL estimating equations to the true score equations of α and θ, they showed that the asymptotic biases in regression coefficients and variance components estimated by PQL were typically of order $\|\theta\|$ for small values of θ, except for Gaussian outcomes. In this section, we investigate in detail the bias of the PQL estimators for Poisson outcomes.

Using the quadratic approximation of $\ell(\alpha, \theta)$ about $\theta = 0$ (Lin and Breslow 1996, Eq. 8) and taking a linear expansion of the score $\partial\ell/\partial\alpha$ evaluated at the true MLE $\alpha = \hat{\alpha}$ about $\alpha = \hat{\alpha}_P$, we find the PQL estimator $\hat{\alpha}_P$ differs from $\hat{\alpha}$ by

$$\hat{\alpha}_P - \hat{\alpha} = \frac{1}{2}(X^T W_0 X)^{-1} X^T W_1 Z^{(2)} J\theta + o(\|\theta\|), \tag{7}$$

where

$$\begin{aligned}
W_0 &= \operatorname{diag}(W_{0i}) = \operatorname{diag}(a_i \mu_i^0/\phi) \\
W_1 &= \operatorname{diag}\left(\frac{\partial W_{0i}}{\partial \eta_i^0}\right) = W_0 \\
\log(\mu_i^0) &= \eta_i^0 = x_i^T \alpha \\
J &= \operatorname{diag}(\mathbf{1}_{q_1}, \ldots, \mathbf{1}_{q_c}).
\end{aligned}$$

Here $\mathbf{1}_{q_j}$ is a $q_j \times 1$ vector of ones and $Z^{(2)} = \left\{z_{kl}^2\right\}$. A detailed derivation of (7) was given in Breslow and Lin (1995) and Lin and Breslow (1996).

Under conditions (I) and (II), we have $Z^{(2)} J = ZJ = \mathbf{1}_n \mathbf{1}_c^T$ and $(X^T W_0 X)^{-1} X^T W_0 Z^{(2)} J = \Delta$, where Δ is a $n \times c$ matrix whose elements in the first row are 1 and whose remaining elements are 0. It follows that

$$\hat{\alpha}_P - \hat{\alpha} = \begin{pmatrix} \frac{1}{2}\sum_{j=1}^{c}\theta_j \\ 0 \\ \vdots \\ 0 \end{pmatrix} + o(\|\theta\|). \tag{8}$$

The significance of this result is that the asymptotic bias of the PQL regression coefficient estimator is of order $o(\|\theta\|)$ for small values of the variance components, except for the intercept.

In order to study the asymptotic bias of the PQL variance component estimator for small values of θ, one may take a linear expansion of the PQL estimating equations for θ in (6) and the true profile score $\partial\ell(\hat{\alpha}(\theta), \theta)/\partial\theta$ about $\theta = 0$ separately and then relate $\hat{\theta}_P$ and $\hat{\theta}$. Some calculations yield

$$\hat{\theta}_P - \hat{\theta} = (C_P^{-1}C - I)\hat{\theta} + o(\|\theta\|), \tag{9}$$

where

$$\begin{aligned} C &= \frac{1}{2}J^T(Z^TW_0Z)^{(2)}J + \frac{1}{4}J^TZ^{(2)^T}W_2Z^{(2)}J \\ &\quad -\frac{1}{4}J^TZ^{(2)^T}W_1X(X^TW_0X)^{-1}X^TW_1Z^{(2)}J \\ C_P &= \frac{1}{2}J^T(Z^TW_0Z)^{(2)}J \\ W_2 &= \text{diag}\left(\frac{\partial^2W_{0i}}{\partial\eta_i^{0\,2}}\right) = W_0. \end{aligned} \tag{10}$$

See Breslow and Lin (1995) and Lin and Breslow (1996) for more details.

Under conditions (I) and (II), Eq. 9 can be greatly simplified. Specifically, the identities $W_0X(X^TW_0X)^{-1}X^TW_0X = W_0X$, $W_0 = W_1 = W_2$ and $Z^{(2)}J = \mathbf{1}_n\mathbf{1}_c^T$ give

$$W_1X(X^TW_0X)^{-1}X^TW_1Z^{(2)}J = W_0\mathbf{1_n1_c^T} = W_2Z^{(2)}J.$$

It follows that the last two terms in (10) coincide and $C_P = C$. The bias of the PQL estimator $\hat{\theta}_P$ under the Poisson mixed models becomes

$$\hat{\theta}_P - \hat{\theta} = o(\|\theta\|). \tag{11}$$

A direct consequence of Eqs. 8 and 11 is that the bias correction procedure proposed by Breslow and Lin (1995) and Lin and Breslow (1996) is not necessary for Poisson outcomes. Our previous work shows that PQL seriously breaks down for binary outcomes and the bias of the PQL estimators is of order $\|\theta\|$. The results reported here indicate that even though the PQL estimators may still remain biased for Poisson outcomes, the order of the bias is smaller compared to that for binary outcomes ($o(\|\theta\|)$ vs. $\|\theta\|$). It follows that the performance of PQL is better for Poisson outcomes compared to binary data, even in problems involving small Poisson means, where the normal theory based Laplace approximation of the marginal likelihood may be expected to be less accurate. We will further examine these results through simulation in Sect. 5.

4 Quasi-pseudo-likelihood

4.1 Estimation under quasi-pseudo-likelihood

Since both the marginal means and covariances of the observations y_i have closed form expressions, we propose a quasi-pseudo-likelihood (QL/PL) approach (Davidian and Carroll 1987) to construct consistent estimators of (α, θ) under the Poisson loglinear mixed model (1). An advantage of this method is that numerical integration required by a full likelihood analysis can be avoided. If the variance components θ were known, estimation of α would proceed by solving the quasilikelihood score equations (McCullagh and Nelder 1989)

$$\begin{aligned} U_\alpha(\alpha, \theta) &= \frac{\partial \mu^T}{\partial \alpha} V^{-1}(y-\mu) \\ &= X^T \mathrm{diag}(\mu) V^{-1}(y-\mu) = 0, \end{aligned} \tag{12}$$

where μ and V, which are functions of α and θ, represent the marginal mean and covariance matrix of y and were given in Sect. 2.

We estimate θ by generalizing the pseudolikelihood approach of Davidian and Carroll (1987). Given $\hat{\alpha}_Q(\theta)$, the solution to (12), the pseudo-loglikelihood of θ is defined as the normal loglikelihood $\ell_N(\hat{\alpha}_Q, \theta)$, where

$$\ell_N(\alpha, \theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}(y-\mu)^T V^{-1}(y-\mu). \tag{13}$$

Estimating equations for θ are obtained by treating $\hat{\alpha}_Q$ fixed and differentiating ℓ_N with respect to θ, ignoring the dependence of μ on θ, as follows:

$$U_\theta(\alpha, \theta_j) = -\frac{1}{2}\mathrm{tr}\left(V^{-1}\frac{\partial V}{\partial \theta_j}\right) + \frac{1}{2}(y-\mu)^T V^{-1}\frac{\partial V}{\partial \theta_j} V^{-1}(y-\mu) = 0, \tag{14}$$

where

$$\frac{\partial V_{ii}}{\partial \theta_j} = \left[\frac{1}{2}\phi a_i^{-1}\mu_i + \mu_i^2\left\{2\exp(z_i^T D z_i) - 1\right\}\right] z_{ij}^T z_{ij}$$

$$\frac{\partial V_{ii'}}{\partial \theta_j} = \frac{1}{2}\mu_i \mu_{i'}(z_{ij} + z_{i'j})^T(z_{ij} + z_{i'j})\exp(z_i^T D z_{i'}) - \frac{1}{2}\mu_i \mu_{i'}(z_{ij}^T z_{ij} + z_{i'j}^T z_{i'j}).$$

Note that $\ell_N(\hat{\alpha}_Q(\theta), \theta)$ is not the corresponding objective function and only serves to motivate the estimating equations (14). The quasi-pseudo-likelihood (QL/PL) estimator $(\hat{\alpha}_Q, \hat{\theta}_Q)$ is defined as the joint solution to (12) and (14). The scale parameter ϕ can be estimated jointly with θ by replacing $\partial V/\partial\theta_j$ with $\partial V/\partial\phi$ in (14). Note that Eqs. 6 and 14 have similar structures. They differ by the fact that under PQL, the pseudolikelihood estimating equations for θ are constructed for the modified GLM working vector Y.

An advantage of the QL/PL method is that the QL/PL estimating equations for both regression coefficients and variance components are strictly unbiased (Davidian and Carroll 1987) and hence the QL/PL estimators are asymptotically consistent. It can be easily seen that the estimating Eq. 14 correspond to the Prentice and Zhao (1991) second-order generalized estimating equations with a Gaussian working matrix. Compared to the PQL estimators, QL/PL estimators are consistent but are likely to have larger variances. We will compare the finite sample performance of these estimators using simulation.

4.2 Asymptotic properties of QL/PL estimators

Before proving strong consistency and asymptotic normality for $\hat{\gamma}_Q = (\hat{\alpha}_Q^T, \hat{\theta}_Q^T)^T$, we first discuss the features of the mixed model (1). As indicated by Miller (1977), under the mixed model (1), the numbers of levels of the random effects may increase at different rates and different variance component estimators may thus require different normalization constants. For example, in a two-way balanced design, the number of levels of the row random effect b_1 and the number of levels of the column random effect b_2 may increase to infinity at different rates. Asymptotic results thus must be modified to take this possibility into account. Another difficulty is that the standard laws of large numbers that require independent random vectors are not applicable, since the observations are not necessarily independent under the mixed model (1). In order to overcome these difficulties, we here adopt the assumptions on the design sequence given by Miller (1977), as follows:

(A.1) The matrix X has full rank p and $n > p + c$.

(A.2) The matrices $Z_j Z_j^T$ $(j = 1, \dots, c)$ are linearly independent, that is $\sum_{j=1}^{c} a_j Z_j Z_j^T = 0$ implies $a_j = 0$ for all $j = 1, \dots, c$.

(A.3) If each q_j $(j = 1, \dots, c)$ can be considered as a function of n, then $\lim_{n\to\infty} q_j = \infty$.

(A.4) If Z_j is labeled so that $\text{rank}(Z_j) = q_j$ is decreasing, that is $q_j \geq q_k$ if $j < k$, then $\lim_{n\to\infty} q_k/q_j$ exists for $j < k$. The limit may be zero. This assumption requires that the design sequence is expanded in an orderly way.

(A.5) Define sequences τ_j, $j = 1, \dots, c$, depending on n, as follows:

$$\tau_j = \text{rank}[Z_{j_s}, \dots, Z_c] - \text{rank}[Z_{j_s}, \dots, Z_{j-1}, Z_{j+1}, \dots, Z_c],$$

where Z_{j_s} is the matrix of the largest rank such that $q_{j_s}/q_j = O(1)$. The τ_j represent the dimension of the part of the column space of Z_j which is orthogonal to that of Z_k for $k > j_s$, and can be regarded as the degree of freedom for the random effect b_j. Suppose that $\lim_{n\to\infty} \tau_j/q_j$ exists and is positive for each $j = 1, \ldots, c$. This assumption ensures that the jth random effect is not asymptotically confounded with the others.

In the trap data set of Sect. 2, there are $q_1 = 720$ levels of the trap effect and $q_2 = 72$ levels of the period effect. For a more detailed discussion of these conditions and illustrative examples, see Gumpertz and Pantula (1992). Denote $M_n = \text{diag}(\sqrt{n}, \ldots, \sqrt{n}, \sqrt{\tau_1}, \ldots, \sqrt{\tau_c})$, $U_n(\gamma) = (U_\alpha^T, U_\theta^T)^T$, and the true value $\gamma_0 = (\alpha_0^T, \theta_0^T)^T$. The following theorem states the strong consistency of the QL/PL estimator $\hat{\gamma}_Q$, assuming ϕ is known (often equals to one). Theorem 1 can be easily extended to incorporate the QL/PL estimator of ϕ, when ϕ is unknown, by slightly modifying the assumptions (A.1)–(A.5) and conditions of Theorem 1 (Gumpertz and Pantula 1992).

Theorem 1 *Suppose that assumptions* (*A.1*)–(*A.5*) *are satisfied for the true value* γ_0. *Define the* $(p + c) \times (p + c)$ *Hessian matrix as*

$$H_n(\gamma) = -\frac{\partial U_n(\gamma)}{\partial \gamma^T}.$$

Suppose that there exists a constant $\delta_n \downarrow 0$ *and that* n, ν_j *and* $\sqrt{n}\delta_n, \sqrt{\nu_j}\delta_n (j = 1, \ldots, c)$ *tend to infinity, such that*

$$\begin{aligned} \|\delta_n M_n^{-1} U_n(\gamma_0)\| &\stackrel{a.s.}{\to} 0 \\ M_n^{-1} H_n(\gamma_0) M_n^{-1} &\stackrel{a.s.}{\to} H, \end{aligned}$$

where $\stackrel{a.s.}{\to}$ *denotes the convergence almost surely,* $\|A\| = (\sum_{ij} A_{ij}^2)^{1/2}$ *for any* A *and* H *is a positive definite matrix, and*

$$\lim_{n\to\infty} \sup \delta_n \left[\text{tr}(M_n^{-1} M_n^{-1})\right]^{-1/2} \|M_n^{-1}(H_n(\gamma^*) - H_n(\gamma_0))M_n^{-1}\| \le M < \infty \tag{15}$$

almost surely, where γ^* *lies on a line segment connecting* $\hat{\gamma}_Q$ *and* γ_0. *Then the equations* $U_n(\gamma)$ *have a solution* $\hat{\gamma}_Q$ *such that* $\hat{\gamma}_Q \stackrel{a.s.}{\to} \gamma_0$.

Proof Denote by $Q_n(\gamma) = -U_n(\gamma)^T U_n(\gamma)/2$ an objective function. Theorem 3.3.1 of Gumpertz and Pantula (1992) implies that, with probability approaching 1, there exists a solution $\hat{\gamma}_Q$ of the equations $U_n(\gamma) = 0$ in a small neighborhood of γ_0 such that $Q_n(\gamma)$ attains a minimum. Further application of Corollary 1 of Gumpertz and Pantula (1992) yields $\hat{\gamma}_Q \stackrel{a.s.}{\to} \gamma_0$. □

Corollary 1 *In addition to the conditions of Theorem* 1, *assume that*

$$M_n^{-1} U_n(\gamma_0) \stackrel{d}{\to} N(0, \Sigma),$$

where $\xrightarrow{d}$ *denotes the convergence in distribution and* Σ *is a positive definite matrix. Then*

$$M_n(\hat{\gamma}_Q - \gamma_0) \xrightarrow{d} N(0, H^{-1}\Sigma H^{-1})$$

Proof See Corollary 2 of Gumpertz and Pantula (1992).

It follows from Theorem 1 and Corollary 1 that the QL/PL estimator $(\hat{\alpha}_Q, \hat{\theta}_Q)$ is strongly consistent and asymptotically normal. □

5 Simulation study

The intention of this section is to compare the finite sample performance of PQL and QL/PL estimators through a simulation study. Each data set used in the simulation was composed of $I = 1000$ clusters of size $n_i = 5$. Conditional on the random effects b_i that were distributed as $N(0, \theta)$ with $\theta = 0.5$ or 1, the Poisson counts y_{ij} were generated within each cluster under the following conditional log-linear model:

$$\log(E(y_{ij}|b_i)) = \alpha_0 + \alpha_1 t_j + \alpha_2 x_i + \alpha_3 x_i t_j + b_i,$$

where $t_j = (j-3)/2$ for $j = 1, \ldots, 5$ and $x_i = 1$ for half the sample and 0 for the reminder. We set $\alpha = (0.0, 1.0, -0.5, 0.5)^T$. The choice of this particular parameter configuration was motivated by the desire to generate data with small to moderate counts. The marginal means ranged from 0.2 to 3.5 when $\theta = 0.5$. We expected that under this scenario, the normal theory based Laplace approximation might be less accurate and hence the performance of PQL might be relatively less satisfactory compared to the cases where the Poisson means were large. For each choice of θ, 1,000 datasets of 1,000 clusters each were generated and analyzed via the PQL and QL/PL procedures.

Results of the simulation are summarized in Table 1, where PQL and QL/PL estimates were obtained for each simulated data set. The estimated standard errors were similar to the simulated standard errors and are therefore omitted from Table 1. When the dispersion was small ($\theta = 0.5$), PQL yielded a slightly negatively biased estimate of the variance component, while QL/PL gave an almost unbiased estimate. With a larger variance component ($\theta = 1.0$), the bias of the PQL estimate was larger, but was still reasonable from a practical viewpoint. The regression coefficient estimates were virtually unbiased under both methods, except that the bias of the PQL intercept estimate was more substantial. The magnitude of this bias was somehow much less than the value $\theta/2$ predicted by (8). Although the bias of the QL/PL estimates was relatively small, they were associated with larger variances. Nelder and Lee (1992) reported similar results in the context of overdispersed generalized linear models, where the PL estimators were found to be more variable than the ML estimators. The mean squares errors of the PQL and QL/PL estimates were comparable, given the selected sample size and the values of the variance components.

Table 1 Sampling statistics for the PQL and QL/PL estimators

Parameter	True value	Bias		Simulated SE		MSE	
		PQL	QL/PL	PQL	QL/PL	PQL	QL/PL
Simulation 1							
θ	0.500	−0.049	−0.004	0.027	0.048	0.003	0.002
α_0	0.000	0.065	0.001	0.038	0.044	0.006	0.002
α_1	1.000	0.000	0.000	0.026	0.026	0.001	0.001
α_2	−0.500	0.011	0.000	0.057	0.062	0.003	0.003
α_3	0.500	0.000	0.000	0.045	0.045	0.002	0.002
Simulation 2							
θ	1.000	−0.118	−0.021	0.052	0.122	0.017	0.015
α_0	0.000	0.083	0.008	0.047	0.072	0.009	0.005
α_1	1.000	0.001	0.001	0.023	0.023	0.001	0.001
α_2	−0.500	0.011	0.002	0.066	0.084	0.005	0.007
α_3	0.500	0.001	0.001	0.038	0.038	0.001	0.001

6 Extension

For small sample problems, restricted maximum likelihood (REML) method can be used to construct a REML type estimating equation of θ by correcting the loss of degrees of freedom from estimating α. The REML estimating equations for θ under PQL were given by Breslow and Clayton (1993, Eq. 15). The corresponding REML estimator of θ under QL/PL can be approximated by replacing (14) with

$$U_{R\theta}(\alpha, \theta_j) = -\frac{1}{2}\mathrm{tr}\left(P\frac{\partial V}{\partial \theta_j}\right) + \frac{1}{2}(y-\mu)^T V^{-1}\frac{\partial V}{\partial \theta_j}V^{-1}(y-\mu) = 0,$$

where $P = V^{-1} - V^{-1}X^*(X^{*T}V^{-1}X^*)^{-1}X^{*T}V^{-1}$ and $X^* = \partial\mu/\partial\alpha^T$ (Davidian and Carroll, 1988).

7 Discussion

We discuss in this article two estimation schemes, penalized quasilikelihood and quasi-pseudo-likelihood in the Poisson mixed models. Although the PQL estimators may still be biased, the order of their bias is $o(\|\theta\|)$ except for the intercept, and is smaller than that for binomial outcomes. The QL/PL estimators are shown to be strongly consistent and asymptotically normal. Although limited in scope, the results of our study suggest that the PQL estimators are more biased and less variable than the QL/PL estimators. From a practical point of view, inference on regression coefficients and variance components under PQL is satisfactory even in problems involving small Poisson means.

Our mixed model setting in (1) resembles those of Morton (1987) and Thall and Vail (1990), where clustered and nested designs were considered, quasilikelihood was used to estimate the regression coefficients, and the method of moments was adopted to estimate the variance components. Note that the asymptotic results on the QL/PL estimators reported here differ from those of Thall and Vail (1990) where independent clusters were assumed.

The QL/PL method discussed herein can be readily extended to the Probit-normal models, where marginal means and covariances of the binary/binomial observations have closed form expressions. Following similar arguments to Theorem 1 and Corollary 1, the resultant QL/PL estimators are strongly consistent and asymptotically normal. By contrast, the PQL estimators are seriously bias for problems involving correlated binary outcomes (Breslow and Clayton 1993; Breslow and Lin 1995).

Acknowledgements This article is dedicated to Professor Norman Breslow. The author would like to thank Professor Breslow for his supervision of her dissertation on corrected PQL methods in GLMMs, and for his inspiration and serving as a role model in her career. This work is supported by grants from the US National Cancer Institute (R37 CA76404 and R13 CA124365).

References

Breslow NE (1984) Extra-Poisson variation in log-linear models. Appl Statist 33:38–44
Breslow NE (1990) Tests of hypotheses in overdispersed Poisson regression and other quasi-likelihood models. J Amer Statist Assoc 85:565–571
Breslow NE, Clayton DG (1993) Approximate inference in generalized linear mixed models. J Amer Statist Assoc 88:9–25
Breslow NE, Lin X (1995) Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika 82:81–91
Davidian M, Carroll RJ (1987) Variance function estimation. J Amer Statist Assoc 82:1079–1091
Green PJ (1987) Penalized likelihood for general semi-parametric regression models. Int Statist Rev 55:245–259
Gumpertz ML, Pantula SG (1992) Nonlinear regression with variance components. J Amer Statist Assoc 87:201–209
Heagerty PJ, Lele SR (1998) A composite likelihood approach to binary spatial data. J Amer Statist Assoc 93:1099–1111
Hinde J (1982) Compound Poisson regression models. In: Gilchrist Berline R (ed) GLIM 82: Proceedings of the International Conference on Generalised Linear Models. Springer-Verlag, pp 109–121
Lawless JF (1987) Negative binomial and mixed Poisson regression. Can J Statist 15:209–225
Lin X, Breslow NE (1996) Bias correction in generalized linear mixed models with multiple components of dispersion. J Amer Statist Assoc 91:1007–1016
Liu Q, Pierce DA (1993) Heterogeneity in Mantel-Haenszel-type models. Biometrika 80:543–556
McCullagh P, Nelder JA (1989) Generalized linear models, 2nd ed. Chapman and Hall, London
Miller JJ (1977) Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance. Ann Statist 5:746–762
Morton R (1987) A generalized linear model with nested strata of extra-Poisson variation. Biometrika 74:247–257
Nelder JA, Lee Y (1992) Likelihood, quasi-likelihood and pseudolikelihood: some comparisons. J R Statist Soc B 50:266–268
Prentice RL, Zhao LP (1991) Estimating equations for parameters in means and covariances of multivariate discrete and continuous responses. Biometrics 47:825–839
Schall R (1991) Estimation in generalized linear models with random effects. Biometrika 40:917–927
Solomon PJ, Cox DR (1992) Nonlinear components of variance models. Biometrika 79:1–11
Stiratelli R, Laird N, Ware J (1984) Random effect models for serial observations with binary response. Biometrics 40:961–971
Thall PF, Vail SC (1990) Some covariance models for longitudinal count data with overdispersion. Biometrics 46:657–671
Zeger SL, Liang KY, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. Biometrics 44:1049–1060