Fitting Longitudinal Data with Hierarchical Generalized Linear Models *

Na Xia and Zhong-Zhan Zhang†
College of Applied Sciences, Beijing University of Technology
Beijing, 100022, P.R.China

Abstract We propose a class of hierarchical generalized linear models (HGLMs) with random dispersions in this paper, and focus on the properties of the L-N estimators for the fixed effect $\boldsymbol{\beta}$ in the extended Poisson-Gamma models which are typical hierarchical generalized linear models. Under the proper assumptions on response variables and some smoothing conditions, we obtain the strong consistency and the convergence rate of the L-N estimator based on the combination of L-N and quasi-likelihood. Furthermore, we also obtain its asymptotical normality based on the combination of the central limit theorem and the law of large numbers. At last, an example is presented for the illustration of the proposed model and the inference method.

Keywords H-likelihood; L-N estimators; Poisson-Gamma models; strong consistency; random effect
2000 MR Subject Classification 62J12, 62F12

1. Introduction

The analysis of longitudinal data has been studied extensively in recent years [5]. An important model class for longitudinal data is generalized linear models with normal random effects [2]. Lee and Nelder [7, 8] introduced hierarchical generalized linear models(HGLMs) by including random components in the linear predictor with arbitrary distributions in generalized linear models(GLMs)[11]. The main idea was using the joint likelihood of response and random effects, which was called hierarchical-likelihood(H-likelihood), to substitute the marginal likelihood for inference of HGLMs [7]. They developed a hierarchical algorithm to give estimators of fixed effects and random effects through maximizing the H-likelihood given dispersion components, and an estimator of dispersion through an adjusted profile H-likelihood given fixed and random effects. This method has an advantage that integrating

*Research supported by NSFC (10371005), EYTP(VE00074) and PHR(IHLB).
†Corresponding Author, Email: zzhang@bjut.edu.cn

out of random effects is not required. The obtained estimators derived from maximizing the H-likelihood are termed as maximum H-likelihood estimators (MHLEs). The frame work has been heuristically extended to double HGLMs further [9].

However, the frame work needs a theoretical development. In fact, as that pointed out by some authors, the MHLE might be biased [7]. In this paper, we extend a typical model family in HGLMs, that is, Poisson-Gamma model to a model which allows random dispersions, and investigate theoretically the H-likelihood method for the extended Poisson-Gamma family. For convenience, with the distribution of the random effect $\boldsymbol{v}$ completely known, we refer to the MHLEs for the fixed effect $\boldsymbol{\beta}$ as L-N estimators. The main purpose is to study the strong consistency and its asymptotical normality of the L-N estimators. Under the conditions which will be given in following section, we demonstrate the strong consistency and the asymptotical normality as the number of groups $t \to \infty$ and the convergence rate of the L-N estimator for the Poisson-Gamma models (P-G models), based on the combination of L-N and quasi-likelihood. The convergence rate is $O(t^{-(\delta-1/2)}(\log\log t)^{1/2})$, where t^{δ} is a lower bound of smallest eigenvalue of information matrix. It is the optimal rate obviously in the case of $\delta = 1$, which is also the rate given by the law of the iterated logarithm.

The organization of this paper is as follows. In Section 2, we shall briefly introduce the models and the main inference methods. In Section 3, we shall give the asymptotic properties of the L-N estimator for the fixed effect $\boldsymbol{\beta}$ in P-G models. In the last section, we shall give an example and the numerical results.

2. Models and the Inference Methods

Consider a trial in which t groups are involved, and there are n_i subjects in the ith group. Denote by y_{ij} and $\boldsymbol{X}_{ij}$ respectively the response and the observation vector of covariates of the jth subject in the ith group; where $i = 1, \cdots, t,$ and $j = 1, \cdots, n_i$. Let $n = \sum_{i=1}^{t} n_i$, and $\boldsymbol{Y}_i = (y_{i1}, \cdots, y_{in_i})^{\tau}$. Suppose that the responses $y_{i1}, \cdots, y_{in_i}$ in the same group are associated to a common random component u_i, and the response vectors $\boldsymbol{Y}_1, \cdots, \boldsymbol{Y}_t$ for different groups are independently distributed. Let $v = v(u)$ be some known strictly monotonic function of u.

In HGLMs, the kernel of the conditional log likelihood for $\boldsymbol{Y} = (\boldsymbol{Y}_i^{\tau}, \cdots, \boldsymbol{Y}_t^{\tau})^{\tau}$ given the random components $\boldsymbol{v} = (\boldsymbol{v}_1^{\tau}, \cdots, \boldsymbol{v}_t^{\tau})^{\tau}$ has the GLM form

$$\sum_{i=1}^{t} \Big\{ \boldsymbol{Y}_i^{\tau} \boldsymbol{\theta}_i' - \boldsymbol{1}^{\tau} \boldsymbol{b}(\boldsymbol{\theta}_i') \Big\} / \phi.$$

where $\boldsymbol{v}_i = (v_i, \cdots, v_i)^{\tau}$ is a n_i dimension vector. The kernel of the likelihood for $\boldsymbol{v}$ is assumed having the form

$$\sum_{i=1}^{t} \Big\{ a_1(\alpha) v_i - a_2(\alpha) b(v_i) \Big\}.$$

Then the kernel of the H-likelihood becomes

$$H = \sum_{i=1}^{t} \Big\{ \boldsymbol{Y}_i^\tau \boldsymbol{\theta}_i' - \boldsymbol{1}^\tau \boldsymbol{b}(\boldsymbol{\theta}_i') \Big\} / \phi + \sum_{i=1}^{t} \Big\{ a_1(\alpha) v_i - a_2(\alpha) b(v_i) \Big\}. \tag{2.1}$$

where ϕ is the dispersion parameter. Given $\boldsymbol{v}$, the link function is given by $\boldsymbol{\mu}_i' = \boldsymbol{E}_\beta(\boldsymbol{Y}_i|\boldsymbol{v}) = \boldsymbol{h}(\boldsymbol{X}_i^\tau \boldsymbol{\beta} + \boldsymbol{v}_i) = \dot{\boldsymbol{b}}(\boldsymbol{\theta}_i')$, where $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \cdots, \boldsymbol{X}_{in_i})$; and the conditional covariance matrix is $\boldsymbol{COV_\beta}(\boldsymbol{Y}_i|\boldsymbol{v}) = \ddot{\boldsymbol{b}}(\boldsymbol{\theta}_i')\phi$, where $\dot{\boldsymbol{b}}(\boldsymbol{\theta}_i') = \left(\frac{\partial b}{\partial \theta_{i1}'}, \cdots, \frac{\partial b}{\partial \theta_{in_i}'} \right)^\tau$, $\ddot{\boldsymbol{b}}(\boldsymbol{\theta}_i')$ denotes the $n_i \times n_i$ second order partial derivative matrix for $\boldsymbol{b}(\boldsymbol{\theta}_i')$ with respect to $\boldsymbol{\theta}_i'$. $\boldsymbol{h} : \mathbb{R}^{n_i} \to \mathbb{R}^{n_i}$ is a bijective and twice continuously differentiable function. Lee and Nelder employ

$$\frac{\partial H}{\partial \boldsymbol{\beta}} = 0, \quad \frac{\partial H}{\partial \boldsymbol{v}} = 0 \tag{2.2}$$

as estimating equations of $\boldsymbol{\beta}$ and $\boldsymbol{v}$ [7, 8].

Now suppose that each y_{ij} is disturbed by an individual random effect η_{ij}, u_i and η_{ij} are all unobserved, and given u_i, $(y_{i1}, \eta_{i1}), \cdots, (y_{in_i}, \eta_{in_i})$ are conditionally independent. Consider multiplicative Poisson-Gamma models, in which $y_{ij}|(u_i, \eta_{ij})$ is distributed as over-dispersed Poisson(μ_{ij}''), $\eta_{ij}|u_i$ is distributed as Gamma(ρ_{ij},ρ_{ij}), and u_i is distributed as Gamma(α,α), a Gamma distribution with mean 1 and a known variance parameter α; where $\rho_{ij} = \rho_{ij}(\boldsymbol{X}_{ij}^\tau \boldsymbol{\beta}, u_i)$ is a known function of $\boldsymbol{X}_{ij}^\tau \boldsymbol{\beta}$ and u_i, probably except a unknown parameter.

Some notations are listed below.

$$\begin{array}{lll}
\boldsymbol{\mu}_i'' = (\mu_{i1}'', \cdots, \mu_{in_i}'')^\tau, & \boldsymbol{\mu}_i' = (\mu_{i1}', \cdots, \mu_{in_i}')^\tau, & \\
\boldsymbol{\mu}_i = (\mu_{i1}, \cdots, \mu_{in_i})^\tau, & \boldsymbol{\mu}' = (\boldsymbol{\mu}_1'^\tau, \cdots, \boldsymbol{\mu}_t'^\tau); & \\
\boldsymbol{U} = \alpha \text{ diag}(u_1, \cdots, u_t), & \boldsymbol{X}_{ij} = (x_{1ij}, \cdots, x_{pij})^\tau, & \boldsymbol{X} = (\boldsymbol{X}_1, \cdots, \boldsymbol{X}_t); \\
\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\tau, & \boldsymbol{\beta}_0 = (\beta_{01}, \cdots, \beta_{0p})^\tau, & \boldsymbol{\theta}_i' = (\theta_{i1}', \cdots, \theta_{in_i}')^\tau; \\
\boldsymbol{b}(\boldsymbol{\theta}_i') = \Big(b(\theta_{i1}'), \cdots, b(\theta_{in_i}') \Big)^\tau, & \boldsymbol{\eta}_i = (\eta_{i1}, \cdots, \eta_{in_i})^\tau, & \boldsymbol{Z} = (Z_1, \cdots, Z_t); \\
\boldsymbol{S}_i = \text{diag}(\mu_{i1}, \mu_{i2}, \cdots, \mu_{in_i}), & \boldsymbol{\mathcal{U}}_i = \text{diag}(\mu_{i1}', \mu_{i2}', \cdots, \mu_{in_i}'), & \boldsymbol{\mathcal{U}} = \text{diag}(\boldsymbol{\mathcal{U}}_1, \cdots, \boldsymbol{\mathcal{U}}_t); \\
\boldsymbol{w} = \mathbf{log}\, \boldsymbol{\mu}' + \boldsymbol{\mathcal{U}}^{-1}(\boldsymbol{Y} - \boldsymbol{\mu}'), & \boldsymbol{R} = \boldsymbol{U}\boldsymbol{v} + (\alpha \boldsymbol{1} - \alpha \boldsymbol{e}^{\boldsymbol{v}}), & \boldsymbol{1} = (1, \cdots, 1)^\tau,
\end{array}$$

where $\boldsymbol{Z}_i$ is the $t \times n_i$ matrix and its the ith row elements are 1, others are 0. For univariate function $b(x)$, use $\boldsymbol{b}(\boldsymbol{x})$ to denote column vector $(b(x_1), b(x_2), \cdots, b(x_p))^\tau$ for a p column vector $\boldsymbol{x} = (x_1, \cdots, x_p)^\tau$.

It holds that

- $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{e}^{\boldsymbol{x}}$, $\boldsymbol{b}(\boldsymbol{\theta}_i') = \boldsymbol{e}^{\boldsymbol{\theta}_i'}$, $b(v_i) = e^{v_i} = u_i$.
- $\mu_{ij}'' = \mu_{ij} u_i \eta_{ij} = E(y_{ij}|u, \eta)$
- $\mu_{ij}' = \mu_{ij} u_i = E(y_{ij}|\boldsymbol{u})$; $\mu_{ij} = e^{\boldsymbol{X}_{ij}^{\boldsymbol{\tau}} \boldsymbol{\beta}}$;
- $\boldsymbol{COV}(\boldsymbol{Y}_i|\boldsymbol{v}) = \boldsymbol{\phi}(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{\tau}} \boldsymbol{\beta}) \ddot{\boldsymbol{b}}(\boldsymbol{\theta}_i') = \boldsymbol{\phi}(\boldsymbol{X}_{\boldsymbol{i}}^{\boldsymbol{\tau}} \boldsymbol{\beta}) \boldsymbol{\mathcal{U}}_i$;

- $E(u_i) = 1; \quad a_1(\alpha) = a_2(\alpha) = \alpha.$
- $E(\eta_{ij}) = 1; \quad var(\eta_{ij}) < \infty.$

where $\boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta}) = \mathrm{diag}(\phi(\boldsymbol{X}_{i1}^\tau \boldsymbol{\beta}), \cdots, \phi(\boldsymbol{X}_{in_i}^\tau \boldsymbol{\beta}))$ is the dispersion matrix, $\phi(\boldsymbol{X}_{ij}^\tau \boldsymbol{\beta}) = \frac{u_i e^{\boldsymbol{X}_{ij}^\tau \boldsymbol{\beta}}}{\rho_{ij}} + 1$ is the dispersion of y_{ij}. Similar to equation (2.2), the estimating equations are

$$\sum_{i=1}^{t} \boldsymbol{X_i} \boldsymbol{\phi^{-1}}(\boldsymbol{X_i^\tau \beta})(\boldsymbol{Y_i} - \boldsymbol{e}^{\boldsymbol{X_i^\tau \beta}} u_i) = 0, \tag{2.3}$$

$$\{(\boldsymbol{Y_i} - \boldsymbol{\mu_i} u_i)^{\boldsymbol{\tau}} \boldsymbol{\phi^{-1}}(\boldsymbol{X_i^\tau \beta})\mathbf{1}\} + \alpha - \alpha u_i = 0. \tag{2.4}$$

These equations are extensions of the likelihood estimating equations. From (2.4), we have

$$\hat{u}_i = \frac{\boldsymbol{Y}_i^\tau \boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta})\mathbf{1} + \alpha}{\boldsymbol{\mu}_i^\tau \boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta})\mathbf{1} + \alpha}. \tag{2.5}$$

Denote $\boldsymbol{\mathcal{E}}_i = \boldsymbol{Y}_i - \boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} u_i$. Substituting u_i with $\hat{u}_i$ in (2.3), we obtain the equation

$$\sum_{i=1}^{t} \boldsymbol{X}_i \boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta}) \Big(\boldsymbol{\mathcal{E}}_i - e^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} \mathbf{1}^\tau \boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta}) \boldsymbol{\mathcal{E}}_i R_i(\boldsymbol{\beta}) + \boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} (u_i - 1)\alpha R_i(\boldsymbol{\beta}) \Big) = 0 \tag{2.6}$$

where $R_i(\boldsymbol{\beta}) = 1/[(\boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})^\tau \boldsymbol{\phi}^{-1}(\boldsymbol{X}_i^\tau \boldsymbol{\beta})\mathbf{1} + \alpha]$.

Equation (2.6) is an unbiased estimating equation with respect to $\boldsymbol{\beta}$. From the point view of generalized estimating equations[], we may replace the first $\boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta})$ on the left hand side by $\boldsymbol{\Lambda}_i(\boldsymbol{\beta}) = diag(\boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})(I - R_i(\boldsymbol{\beta})\boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta})\mathbf{1}(\boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}})^\tau)V_i(\boldsymbol{\beta})$, where $\boldsymbol{V}_i(\boldsymbol{\beta})$ is a given positive definite matrix, $i = 1, \cdots, t$. This leads to

$$\boldsymbol{L}_t(\boldsymbol{\beta}) \equiv \sum_{i=1}^{t} \boldsymbol{X}_i \boldsymbol{\Lambda}_i(\boldsymbol{\beta}) \Big(\boldsymbol{\mathcal{E}}_i - \boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} \mathbf{1}^\tau \boldsymbol{\phi}(\boldsymbol{X}_i^\tau \boldsymbol{\beta}) \boldsymbol{\mathcal{E}}_i R_i(\boldsymbol{\beta}) + \boldsymbol{e}^{\boldsymbol{X}_i^\tau \boldsymbol{\beta}} (u_i - 1)\alpha R_i(\boldsymbol{\beta}) \Big) = 0 \tag{2.7}$$

(2.7) is more flexible in the mean that the estimator obtained may still be consistent without the correct assumptions for the conditional covariance matrix of $\boldsymbol{Y}_i$. We focus the estimator of $\boldsymbol{\beta}$ obtained from (2.7), and call it the L-N estimator. An estimator (or prediction value) of u_i may be obtained by inserting the estimator obtained into (2.5).

3. Properties

Let $\hat{\boldsymbol{\beta}}_t$ be the solution of (2.6), and $\boldsymbol{\beta}_0$ is the true value of $\boldsymbol{\beta}$. Suppose that $\boldsymbol{\beta}_0 \in \mathbb{B}^p$, where $\mathbb{B}^p$ is a bounded subset of p-dimensional Euclidean space. Denote

$$\begin{aligned} \boldsymbol{Q}_t(\boldsymbol{\beta}) &= \tfrac{1}{t} \boldsymbol{L}_t(\boldsymbol{\beta}) \\ &= \tfrac{1}{t} \sum_{i=1}^{t} \boldsymbol{X}_i \boldsymbol{\Lambda}_i(\boldsymbol{\beta}) \Big(\boldsymbol{\mathcal{E}}_i - \boldsymbol{e}^{\boldsymbol{X_i^\tau \beta}} \mathbf{1}^\tau \boldsymbol{\phi^{-1}}(\boldsymbol{X_i^\tau \beta}) \boldsymbol{\mathcal{E}}_i R_i(\boldsymbol{\beta}) + \boldsymbol{e}^{\boldsymbol{X_i^\tau \beta}} (u_i - 1)\alpha R_i(\boldsymbol{\beta}) \Big), \\ \boldsymbol{W}_t(\boldsymbol{\beta}) &= COV(\sqrt{t} \boldsymbol{Q}_t(\boldsymbol{\beta})). \end{aligned}$$

$\boldsymbol{W}_t(\boldsymbol{\beta})$ is given as follows.

(i) When $\boldsymbol{\phi}(\boldsymbol{X}_i^\tau\boldsymbol{\beta})$ is a fixed matrix, there is

$$\begin{aligned}\boldsymbol{W}_t(\boldsymbol{\beta}) = \ & \frac{1}{t}\sum_{i=1}^{t}\boldsymbol{X}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\Big[(\boldsymbol{I}-\boldsymbol{A}_i)^\tau\boldsymbol{\phi}(\boldsymbol{X}_i^\tau\boldsymbol{\beta})\boldsymbol{S}_i(\boldsymbol{I}-\boldsymbol{A}_i)\\ & +\boldsymbol{e}^{\boldsymbol{X}_i^\tau\boldsymbol{\beta}}(\boldsymbol{e}^{\boldsymbol{X}_i^\tau\boldsymbol{\beta}})^\tau\alpha R_i^2(\boldsymbol{\beta})\Big]\boldsymbol{X}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta});\end{aligned}$$

(ii) When $\boldsymbol{\phi}(\boldsymbol{X}_i^\tau\boldsymbol{\beta})$ is a random matrix, there is

$$\begin{aligned}\boldsymbol{W}_t(\boldsymbol{\beta}) = \ & \frac{1}{t}\sum_{i=1}^{t}\boldsymbol{X}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta})\Big[\boldsymbol{E}\Big((\boldsymbol{I}-\boldsymbol{A}_i)^\tau\boldsymbol{\phi}(\boldsymbol{X}_i^\tau\boldsymbol{\beta})\boldsymbol{\mathcal{U}}_i(\boldsymbol{I}-\boldsymbol{A}_i)\Big)\\ & +\boldsymbol{e}^{\boldsymbol{X}_i^\tau\boldsymbol{\beta}}(\boldsymbol{e}^{\boldsymbol{X}_i^\tau\boldsymbol{\beta}})^\tau\alpha^2E\Big((u_i-1)^2R_i^2(\boldsymbol{\beta})\Big)\Big]\boldsymbol{X}_i\boldsymbol{\Lambda}_i(\boldsymbol{\beta}).\end{aligned}$$

where $\boldsymbol{A}_i=\boldsymbol{\phi}^{-1}(\boldsymbol{X}_i^\tau\boldsymbol{\beta})1(\boldsymbol{e}^{\boldsymbol{X}_i^\tau\boldsymbol{\beta}})^\tau R_i(\boldsymbol{\beta})$.

In order to obtain the strong consistency and the asymptotical normality of the L-N estimator, we give some additional assumptions which are stated as follows.

A1. $\{\boldsymbol{X}_i, i\geq 1\}$ *is bounded,* $\lambda_t\geq ct^\delta$ *for sufficiently large* t *and* $\delta\in(3/4,1]$, *where* λ_t *is the smallest eigenvalue of the symmetric matrix* $\sum_{i=1}^{t}\boldsymbol{X}_i\boldsymbol{X}_i^\tau$;

A2. $E\Big[\boldsymbol{COV}_{\boldsymbol{\beta}_0}(\boldsymbol{Y}_i|u_i)\Big]\geq cI, i=1,2,\cdots,\ \sup_{i\geq 1}E_{\boldsymbol{\beta}_0}\|\boldsymbol{Y}_i\|^{\bar{p}}<\infty, \bar{p}=8/3$;

A3. $\phi(\cdot)>c>0$ *is twice continuously differentiable. Moreover,* $\phi(\cdot)$ *and its first and second order partial derivative are bounded in arbitrary bounded subset;*

A4. *For all* i, $\boldsymbol{V}_i(\boldsymbol{\beta})>cI$ *for all* $\boldsymbol{\beta}\in\mathbb{B}^p$, *where* $c>0$ *is a constant independent of* i; *and all elements of* $\boldsymbol{V}_i(\boldsymbol{\beta})$ *have continuous second order partial derivatives; moreover, the elements of* $\boldsymbol{V}_i(\boldsymbol{\beta})$, *their first and second order partial derivatives are bounded in arbitrary subset of* $\mathbb{B}^p$.

A5. $\lim_{t\to\infty}\boldsymbol{W}_t(\boldsymbol{\beta}_0)=\boldsymbol{W}(\boldsymbol{\beta}_0)$; *where* $\boldsymbol{W}(\boldsymbol{\beta}_0)$ *is a positive matrix;*

A6. *For all* $\boldsymbol{\beta}\in\mathbb{B}^p$, $\lim_{t\to\infty}\boldsymbol{F_t}(\boldsymbol{\beta})=\boldsymbol{F}(\boldsymbol{\beta})$, *where* $\boldsymbol{F_t}(\boldsymbol{\beta})=\boldsymbol{E_\beta}(-\frac{\partial\boldsymbol{Q_t}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\tau})$, denote $-\frac{\partial\boldsymbol{Q_t}(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}^\tau}=\dot{\boldsymbol{Q}}_t(\boldsymbol{\beta})$;

A7. $\{n_i: i=1,\cdots,t\}$ *is bounded.*

The main result is as following.

Theorem 1 *Suppose Assumptions A1∼A4 hold, then there exists an estimator* $\hat{\boldsymbol{\beta}}_t$ *of* $\boldsymbol{\beta}_0$ *such that*

$$P\Big(\boldsymbol{L}_t(\hat{\boldsymbol{\beta}}_t)=0, \textit{for all sufficiently large } t\Big)=1$$

$$\hat{\boldsymbol{\beta}}_t-\boldsymbol{\beta}_0=O(t^{-(\delta-1/2)}(\log\log t)^{1/2})\quad a.s..$$

In the important special case of $\delta = 1$, the convergence rate is

$$\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0 = O(t^{-1/2}(\log\log t)^{1/2})$$

which is the same as the rate what the law of the iterated logarithm determined for partial sums of independent identically distributed random variables.

Lemma 1 *If $\boldsymbol{F}_t(\boldsymbol{\beta})$, $\dot{\boldsymbol{Q}}_t(\boldsymbol{\beta})$ satisfy the above assumptions, then for $\boldsymbol{\beta}_0 \in \mathbb{B}^p$, we have*

$$\|\dot{\boldsymbol{Q}}_t(\boldsymbol{\beta}_0) - \boldsymbol{F}_t(\boldsymbol{\beta}_0)\| \stackrel{a.s.}{\rightarrow} 0 \quad (t \rightarrow \infty).$$

Lemma 2 *Suppose Assumptions A1∼A7 hold, then*

$$\sqrt{t}\ \boldsymbol{Q}_t(\boldsymbol{\beta}_0) \stackrel{L}{\rightarrow} N(\mathbf{0}, \boldsymbol{W}(\boldsymbol{\beta}_0)).$$

Theorem 2 *The conditions are stated as that in Lemma 2, then*

$$\sqrt{t}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0) \stackrel{L}{\rightarrow} N(\mathbf{0}, \boldsymbol{F}^{-1}(\boldsymbol{\beta}_0)\boldsymbol{W}(\boldsymbol{\beta}_0)\boldsymbol{F}^{-1}(\boldsymbol{\beta}_0)).$$

Remark For the proofs of the theorems refer to [16]. Some conditions above could be weaken technically. For example, the asymptotic normality needs only the consistency, and the condition on the smallest eigenvalue in A1 could be given similar to [3].

4. Numerical Results

4.1 Algorithm

We develop an algorithm motivated by Fisher scoring method. It includes two steps given as following, and is shown very efficient computationally.

(1) Iterative weighted least square for estimating $\boldsymbol{\beta}$ and $\boldsymbol{v}$

$$\begin{pmatrix} \boldsymbol{XWX}^{\tau} & \boldsymbol{XWZ}^{\tau} \\ \boldsymbol{ZWX}^{\tau} & \boldsymbol{ZWZ}^{\tau} + \boldsymbol{U} \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta} + \boldsymbol{\delta\beta} \\ \boldsymbol{v} + \boldsymbol{\delta v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{XWw} \\ \boldsymbol{ZWw} + \boldsymbol{R} \end{pmatrix} \tag{4.1}$$

where $\boldsymbol{W} = \text{diag}(\phi^{-1}(\boldsymbol{X}_{\boldsymbol{11}}^{\boldsymbol{\tau}}\boldsymbol{\beta})\mu'_{11}, \cdots, \phi^{-1}(\boldsymbol{X}_{\boldsymbol{tn_t}}^{\boldsymbol{\tau}}\boldsymbol{\beta})\mu'_{tn_t})$, $\boldsymbol{\delta\beta}$ and $\boldsymbol{\delta v}$ are the adjustment of $\boldsymbol{\beta}$ and $\boldsymbol{v}$ respectively; $\boldsymbol{w}$ is the adjusted dependent variable.

(2) Estimate the variance parameter α

$$\sum_{i=1}^{t}\{v_i + \log\alpha + 1 - u_i - \psi(\alpha)\} - \frac{1}{2}tr\left\{\boldsymbol{K}diag(u_1, \cdots, u_t)\right\} = 0 \tag{4.2}$$

where $\psi(\alpha)$ is digamma function, $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}$ and $\boldsymbol{v} = \hat{\boldsymbol{v}}$ are re-evaluated in each iteration in (4.1), and $\boldsymbol{K}$ is determined from

$$\begin{pmatrix} \boldsymbol{XWX}^{\tau} & \boldsymbol{XWZ}^{\tau} \\ \boldsymbol{ZWX}^{\tau} & \boldsymbol{ZWZ}^{\tau} + \boldsymbol{U} \end{pmatrix}^{-1} = \begin{pmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{C} & \boldsymbol{K} \end{pmatrix}.$$

4.2 An example

As an illustration of the new model, we present the analysis of data arising from a clinical trial of 59 epileptics carried out by Leppik et al(1985)[14]. Patients suffering from simple or complex partial seizures were randomized to receive either the antiepileptic drug progabide(Trt=1) or a placebo(Trt=0), as an adjuvant to standard chemotherapy. Baseline data available at entry into the trial included the number of epileptic seizures recorded in the preceding 8-week period and age in years. The logarithm of $\frac{1}{4}$ the number of baseline seizures(B) and the logarithm of age(A) were treated as covariates in the analysis. A multivariate response variable consisted of the counts of seizures during the 2-weeks before each of four clinic visits(Visit, coded Visit=(-3,-1,1,3)/10). Preliminary analysis indicated that the counts were substantially lower during the fourth visit and a binary variable(V_4=1 for fouth visit, 0 otherwise) was constructed to model such effects. There are several authors have analyzed these data by using various models, i.e. Thall and Vail(1990)[15], Breslow and Clayton(1993)[2], Lee and Nelder(1996)[7], Diggle(2002)[5] and so on. We use the above Poisson-Gamma models to analyze these data, for which $t = 59, n_i = 4, p = 7, n = 236$. Our main arithmetic are stated as follows.

The model described in Section 2 might include a random dispersion. For the purpose of comparison, we present the results of fitting to the epileptic data corresponding to the following two models.

Model 1: $\rho_{ij} = u_i e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}, \quad \phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2.$

Model 2: $\rho_{ij} = e^{\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}}, \quad \phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i.$

Model 1 is a conventional model with a constant dispersion, and Model 2 has random dispersions $1 + u_i$, which allows the dispersions varying with the common group random effects.

Because some observations had been identified as outliers [15], we first give out the Q-Q plots of the standardized residuals $\tilde{r}_{ij1} = (\hat{r}_{ij1} - \bar{r}_1)/\hat{\sigma}_1$ and the standardized conditional residuals $\tilde{r}_{ij2} = (\hat{r}_{ij2} - \bar{r}_2)/\hat{\sigma}_2$ of the complete data for Model 1 and Model 2 respectively in Figure 1 and Figure 2, where $\hat{r}_{ij1} = y_{ij} - e^{\boldsymbol{X}_{ij}^{\tau}\hat{\boldsymbol{\beta}}}, \hat{r}_{ij2} = y_{ij} - e^{\boldsymbol{X}_{ij}^{\tau}\hat{\boldsymbol{\beta}}}\hat{u}_i, \bar{r}_h = \frac{1}{236}\sum_{i,j}\hat{r}_{ijh}, \hat{\sigma}_h = \sqrt{\frac{1}{235}\sum_{i,j}(\hat{r}_{ijh} - \hat{r}_{ijh})^2}, h = 1, 2$. In each figure, the above plot is of the standardized residuals, and the below one is of the standardized conditional residuals. Both figures show that the 3rd response of patient 227 (control group) and the 1st response of patient 207 (treatment group) have unusual residuals. Boxplots and other residula plots show the same. Furthermore, the conditional residual plots show the 3rd response of patient 227, corresponding to the highest points in the plots, is more outlying. A careful observation to the data digs out that the response 76 of patient 227 is about 3 times of his other responses (18, 27 and 25 respectively), while the 1st response of patient 207 is not so high compared with his other response, although his all four response are much more higher than those of other patients apparently. In fact, the conditional residual plots do not display evidence for the 1st response of patient 207 being an outlier.

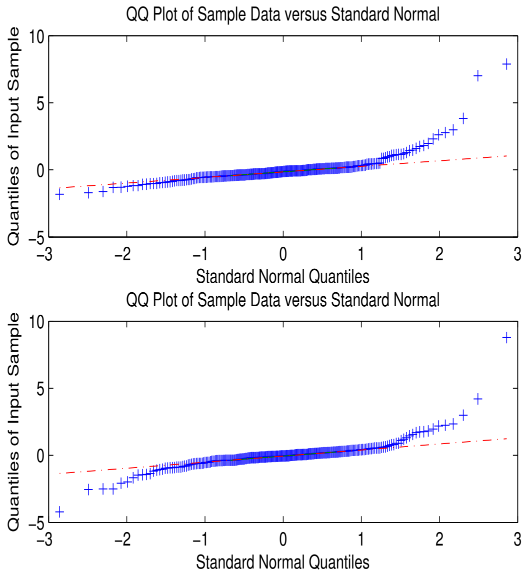

Figure 1: $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2$, complete data.

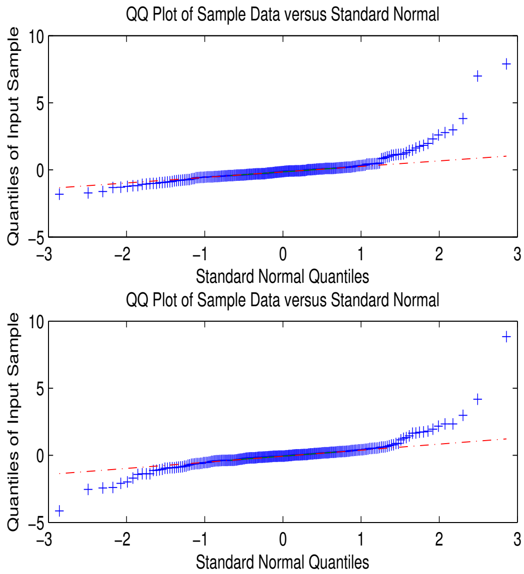

Figure 2: $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$, complete data.

For the purpose of comparison with respect to conventional analysis, we delete the data of patient 227 and 207, although this might not be necessary. Then the figures become Figure 3 and Figure 4. No more unusual observations are found in the plots.

Secondly, we give the estimates and the asymptotic standard deviations of the fixed effect βs and their p values of Wald tests for the significance of the fixed effects. Table 1 and Table 2 are the results of Model 1 and Model 2 respectively. The results for complete data in the two tables are followed by the results after deletion of the two patients given in parentheses. The two models give very different explanations for the data. The conventional model shows that only the baseline observation is significant at level 0.05, while the model with random dispersion tell us that not only the treatment, age and baseline have significant influence for the therapy, but also there is an interaction between the treatment and baseline. This is also find by Lee and Nelder with a much more complicated model [9]. In addition, Model 2 gives smaller asymptotic standard deviations for the estimates of the fixed effects, see also [15, 2].

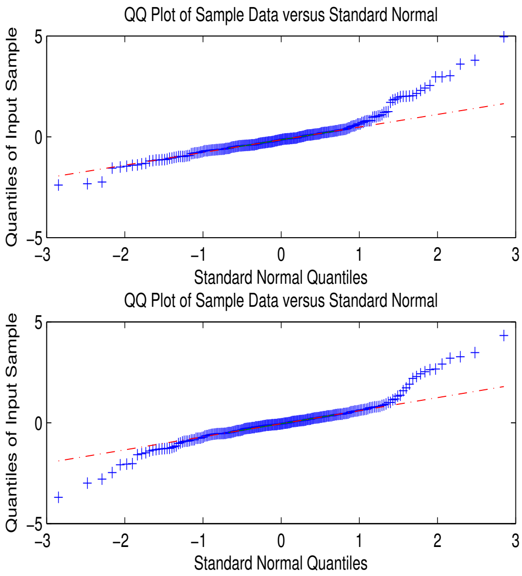

Figure 3: $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2$,
227 and 207 are deleted.

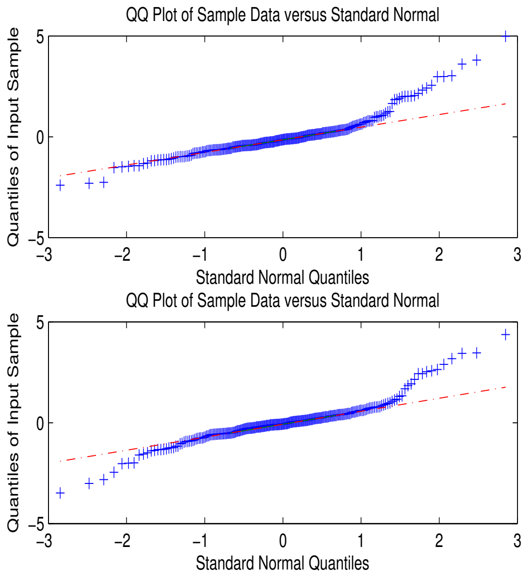

Figure 4: $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$,
227 and 207 are deleted.

This might because Model 2 accounts for heteroscedasticity, dependence among repeated observations and overdispersion simultaneously.

Table 1 $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 2$,
summaries of analyses for the eplieptics data

Parameter	$\hat{\boldsymbol{\beta}}_t$	Asymptotic standard error	p value
$\boldsymbol{\beta}_0$	-1.2512	1.5482	0.4190
	(-1.1737)	(1.5307)	(0.4433)
$\boldsymbol{\beta}_T$	-0.8802	0.5372	0.1013
	(-0.7268)	(0.5555)	(0.1907)
$\boldsymbol{\beta}_A$	0.4899	0.4525	0.2790
	(0.4787)	(0.4469)	(0.2841)
$\boldsymbol{\beta}_B$	0.8743	0.1736	4.7735×10^{-7}
	(0.7975)	(0.1747)	(4.9617×10^{-6})
$\boldsymbol{\beta}_V$	-0.1481	0.2904	0.6099
	(-0.3747)	(0.2946)	(0.2034)
$\boldsymbol{\beta}_{T*B}$	0.3383	0.2733	0.2158
	(0.2706)	(0.2916)	(0.3536)
$\boldsymbol{\beta}_{V_4}$	-0.1015	0.1547	0.5116
	(0.0302)	(0.1571)	(0.8476)

Table 2 $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$,
summaries of analyses for the eplieptics data

Parameter	$\hat{\boldsymbol{\beta}}_t$	Asymptotic standard error	p value
$\boldsymbol{\beta}_0$	-1.2913	0.7094	0.0687
	(-1.1983)	(0.4631)	(0.0097)
$\boldsymbol{\beta}_T$	-0.8958	0.1190	5.1292×10^{-14}
	(-0.7388)	(0.1606)	(4.2067×10^{-6})
$\boldsymbol{\beta}_A$	0.4888	0.0486	0
	(0.4758)	(0.0583)	(4.4409×10^{-16})
$\boldsymbol{\beta}_B$	0.8807	0.0294	0
	(0.8046)	(0.0285)	(0)
$\boldsymbol{\beta}_V$	-0.2913	5.0888	0.9544
	(-0.4372)	(2.8956)	(0.8800)
$\boldsymbol{\beta}_{T*B}$	0.3419	0.0410	0
	(0.2717)	(0.0636)	(1.9626×10^{-5})
$\boldsymbol{\beta}_{V_4}$	-0.0069	2.7286	0.9980
	(0.0866)	(1.5422)	(0.9553)

We have also compared many other choices of dispersion. The results obtained show the models with random dispersions are better than those with fixed dispersions, and the significance of the covariates is the same.

At last, we give the plots of standardized conditional residuals for the complete data and the data with patients 227 and 207 deleted respectively in Figure 5 and Figure 6 for model 2 on the log baseline counts at four occasion. The figures show the dispersion structure is needed.

It might be possible to introduce an unknown parameter in the dispersion function $\rho(\cdot,\cdot)$. We shall present further results on this in future papers.

References

[1] Bennet, G., Probability inequalities for the sum of independent random variables. *J. Am. Statist. Ass.*, Vol. 57(1962), 33-45.

[2] Breslow, N. E. and Clayton, D. G., Approximate inference in generalized linear mixed models. *J. Am. Statist. Ass.*, Vol.88(1993), 9-25.

[3] Chen, K., Hu, I. and Ying, Z., Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *The Annals of Statistics*, Vol.27(1999), No.4, 1155-1163.

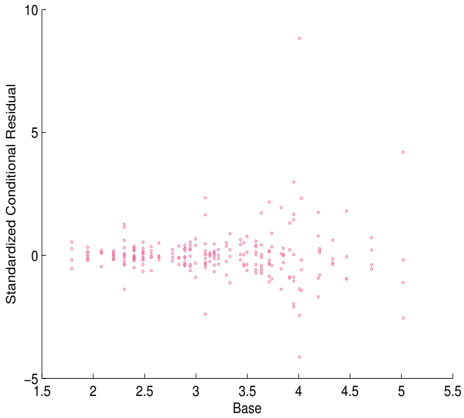

Figure 5: The complete data and $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$.

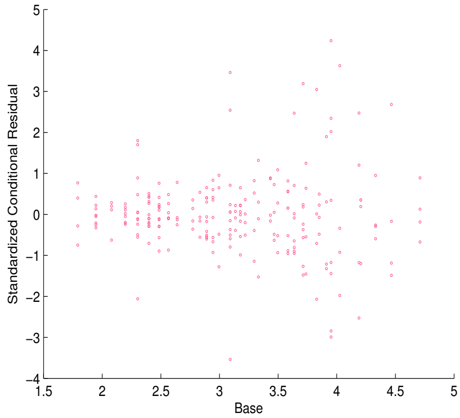

Figure 6: After deletion the two patients and $\phi(\boldsymbol{X}_{ij}^{\tau}\boldsymbol{\beta}) = 1 + u_i$.

[4] Chen, X. R. and Yue, L., The strong consistency and convergence rate of maximum quasi-likelihood estimator in GLMs. *Science in China Series A*, Vol.34(2004), No.2, 203-214 (in Chinese).

[5] Diggle, P. J., Analysis of Longitudinal Data. Oxford University Press, 2002.

[6] Fahrmeir, L. and Kaufmann, H., Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *The Annals of Statistics*, Vol.13(1985), No.1, 342-368.

[7] Lee, Y. and Nelder, J. A., Hierarchical generalized linear models (with discussion). *J. R. Statist. Soc. B*, Vol. 58(1996), No.4, 619-678.

[8] Lee, Y., Nelder, J. A. Hierarchical generalized linear models : a synthesis of generalized linear models,random-effect models and structured dispersions. *Biometrika*, 88(2001),No.4, 987-1006.

[9] Lee, Y., Nelder, J. A. Double hierarchical generalized linear models (with discussions). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, Vol.55 (2006), No.2, 139-185.

[10] Lu, C. R., Lin, Z. Y. Foundations of Probability limit Theory. Higher Education Press, Beijing (in Chinese), 1989.

[11] McCullagh, P., Nelder, J. A. Generalized Linear Models. Chapman and Hall, London, 1989.

[12] Petrov, V. V. Sums of Independent Random Variables. Springer-Verlag, Berlin, New York, 1975

[13] Stout, W. F. Almost Sure Convergence. Academic Press, New York, 1974.

[14] Leppik, I. E.,et al. A double-blind crossover evaluation of progabide in partial seizures. Neurology, 1985, 35, 285.

[15] Thall, P. F. and Vail, S.C. Some covariance models for longitudinal data with overdispersion. *Biometrics*, 46(1990), 657-671.

[16] Xia, N. and Zhang, Z.Z. Convergence Rate of the L-N Estimator in Poisson-Gamma Models. *Acta Mathematicae Applicatae Sinica*, to appear.