RESAMPLING METHODS IN GENERALIZED LINEAR MODELS USEFUL IN ENVIRONMETRICS

HERWIG FRIEDL AND ERNST STADLOBER
Institute of Statistics, Technical University Graz,
Lessingstr. 27, A-8010 Graz, Austria

Abstract

Generalized linear models are important tools for analyzing relationships between binary, count or continuous response variables and predictors with fixed effects. In this paper we present a survey on bootstrap methods based on (extended) quasi-likelihood assumptions. We discuss two approaches: one-step residual resampling and score resampling to estimate the variability of functions in the linear parameters of the model and an iterative procedure which allows to define replicates of the dependent variate. With the latter we are able to estimate nonlinear parameters in the variance function and to compare non-nested models. The power of these resampling schemes is illustrated by air sampler data concentrating on the number of bacteria colonies observed at outdoor sites in the area of Graz.

1 INTRODUCTION

Generalized linear models (GLM's) are often appropriate to describe the dependence of data from some explanatory variables. Especially in the environmental sciences the analysis of observations representing discrete variates like counts or frequencies is of growing importance. An appropriate model enables to gain new insight into the stochastic and the systematic component of the data generating mechanism. Moreover, parameters and functions thereof can be easily studied which help to describe the behaviour of the data effected by the explanatory variables.

McCullagh and Nelder[1] introduced GLM's where independent responses y_i, $i = 1, \ldots, n$, are assumed to have a distribution from the exponential family. One of the key ideas in GLM's is to relate the expected values $E(y_i) = \mu_i$ onto linear predictors $\eta_i = x_i^t\beta$ with vectors of covariates $x_i = (x_{i1}, \ldots, x_{ip})^t$ and common parameters $\beta = (\beta_1, \ldots, \beta_p)^t$ by some known inverse link function, $\mu_i = h(\eta_i)$. Because of the exponential family, $Var(y_i) = \sigma^2 V(\mu_i)$ holds, where the variance function $V(\cdot)$ depends only on the mean, and σ^2 is a possibly unknown dispersion parameter. Let $l(y, \mu)$ be the log-likelihood of the sample under the considered distribution. Then for fixed σ^2 the maximum likelihood (ML) estimate $\hat{\beta}$ is derived as solution of the system of p score functions

$$u_\beta(y,\mu) = \frac{\partial l(y,\mu)}{\partial\beta} = \sum_{i=1}^n u_\beta(y_i,\mu_i) = \sum_{i=1}^n \frac{\partial l(y_i,\mu_i)}{\partial\mu_i}\frac{\partial\mu_i}{\partial\beta} = \sum_{i=1}^n \frac{y_i-\mu_i}{\sigma^2 V(\mu_i)}\frac{\partial\mu_i}{\partial\beta}, \tag{1}$$

and is independent of σ^2. Goodness-of-fit can be evaluated by means of the log-likelihood ratio statistic

$$\frac{1}{\sigma^2}D(y,\mu) = -2\Big(l(y,\mu) - l(y,y)\Big), \tag{2}$$

usually referred to as scaled deviance. Maximizing the likelihood is therefore equivalent to minimizing the deviance.

Section 2 deals with a more general framework. GLM's with suitable mean-variance relations for the responses lead to quasi-likelihoods. Models based on different variance functions can be compared by the extended quasi-likelihood or the pseudo-likelihood approach. Several bootstrap procedures are discussed in Section 3. Residual resampling and score resampling yield both one-step parameter replications resulting in alternative estimates of the variance-covariance matrix $Var(\hat{\beta})$. However, an iterative scheme also allows for generating replications of the responses y_i. In Section 4 all those methods are applied to bacteria counts observed during a period of one year in different outdoor sites. Finally, some discussions and conclusions are presented in Section 5.

2 EXTENDED LIKELIHOOD CONCEPTS

2.1 Quasi-Likelihood

For the analysis of counts or frequencies the assumption of Poisson or binomial distributions often seems to be too inflexible, which is mainly caused by the restrictive variance models involved. On the other hand, the only distributional information in the definition of the ML estimate $\hat{\mu}$ are the first two moments of y. Hence, Wedderburn[2] widened the scope of GLM's by assuming only a suitable mean-variance relation instead of specifying the full distribution of the responses. This enables to model overdispersion as well as to consider distributions outside the exponential family. Negative-binomial or beta-binomial variation are examples which are often used in practise. Moreover, variance functions without explicitely specifying a distribution can also be handled within this context. For arbitrary variance models of the form $Var(y) = \sigma^2 V(\mu)$ with σ^2 variable, the so called quasi-score function $(y - \mu)/\sigma^2 V(\mu)$ has important properties in common with a log-likelihood derivative concerning μ. Analogously to the log-likelihood function $l(y, \mu)$ the log-quasi likelihood is defined as

$$q(y, \mu) = \sum_{i=1}^{n} q(y_i, \mu_i) = \sum_{i=1}^{n} \int^{\mu_i} \frac{y_i - t}{\sigma^2 V(t)} dt \tag{3}$$

plus any function depending on y and σ^2 only. If $q(y, \mu)$ exists then $\exp(q(y, \mu))$ should behave like a likelihood for the mean μ. A detailed discussion of quasi-distributions is given in Nelder and Lee[3]. The maximum quasi-likelihood (QL) estimate $\hat{\beta}$ again solves (1) and is not affected by σ^2. To estimate the dispersion parameter, Wedderburn[2] suggested to use the bias-corrected mean Pearson statistic

$$\hat{\sigma}^2_X = \frac{1}{n-p} X^2(y, \hat{\mu}) = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)} \tag{4}$$

and X^2 is another suitable measure of discrepancy of a fit. By analogy to (2), the quasi-deviance of the sample is

$$D(y, \mu) = -2\sigma^2 \Big(q(y, \mu) - q(y, y) \Big) = 2 \sum_{i=1}^{n} \int_{\mu_i}^{y_i} \frac{y_i - t}{V(t)} dt, \tag{5}$$

which is strictly positive except at $y = \mu$.

2.2 Extended Quasi-Likelihood

The QL function (3) can be used to compare different link functions or linear predictors. To examine different variance models on the same sample Nelder and Pregibon[4] introduced the extended quasi-likelihood (EQL) function. Such an analysis is necessary since the variance function determines the units of measurement for D and X^2 and distinguishing between these measures across different variance functions is impossible. If the variance function is not known exactly one may embed it into a parameterized family like the power family $Var(y) = \sigma^2 V_\alpha(\mu) = \sigma^2\mu^\alpha$ where σ^2 and α are both possibly unknown. For a fixed value of α, the EQL function is

$$q_\alpha^+(y,\mu) = q_\alpha(y,\mu) - \frac{1}{2}\sum_{i=1}^{n} \log(2\pi\sigma^2 V_\alpha(y_i)) \tag{6}$$

and the corresponding scaled extended quasi-deviance takes simply the form

$$D_\alpha^+(y,\mu) = \frac{1}{\sigma^2} D_\alpha(y,\mu) + \sum_{i=1}^{n} \log(2\pi\sigma^2 V_\alpha(y_i)). \tag{7}$$

D_α is the quasi-deviance (5) and q_α is the QL function (3) with V replaced by V_α. Therefore, maximizing (6) results in the QL estimate $\hat{\beta}$ and the dispersion estimate $D_\alpha(y,\hat{\mu})/n$. To adjust for finite samples the estimate

$$\hat{\sigma}_D^2 = D_\alpha(y,\hat{\mu})/(n-p) \tag{8}$$

should be used in practise. In contrast to the QL function, the EQL function also behaves like a likelihood with respect to σ^2 because

$$E\left(\frac{\partial q_\alpha^+}{\partial\sigma^2}\right) = 0, \qquad Var\left(\frac{\partial q_\alpha^+}{\partial\sigma^2}\right) = -E\left(\frac{\partial^2 q_\alpha^+}{\partial(\sigma^2)^2}\right) \approx \frac{n}{2}\frac{1}{\sigma^4}.$$

Approximate $(1-\gamma)\%$ profile QL intervals for the parameter α can be obtained in the usual way as

$$\{\alpha | 2(q_{max}^+ - q_\alpha^+) \le \chi^2_{1,1-\gamma}\},$$

where $q_{max}^+ = \max_\alpha q_\alpha^+(y,\hat{\mu})$ and $\chi^2_{1,1-\gamma}$ is the $(1-\gamma)$-quantile of the χ^2 distribution with one degree of freedom.

2.3 Pseudo-Likelihood

Given an estimate $\hat{\beta}$, which is not necessarily the ML estimator defined through (1), Carroll and Ruppert[5] maximized a normal likelihood in α and σ^2, setting the variance of y_i to $\sigma^2 V_\alpha(\hat{\mu}_i)$. The corresponding scaled pseudo-deviance function may be written as

$$D_\alpha^P(y,\hat{\mu}) = \frac{1}{\sigma^2} X_\alpha^2(y,\hat{\mu}) + \sum_{i=1}^{n} \log(2\pi\sigma^2 V_\alpha(\hat{\mu}_i)), \tag{9}$$

with Pearson statistic $X_\alpha^2(y,\hat{\mu}) = \sum_i (y_i - \hat{\mu}_i)^2 / V_\alpha(\hat{\mu}_i)$, where the means $\hat{\mu}_i$ are considered to be fixed and are evaluated in $\hat{\beta}$. (9) differs from (7) by having $\hat{X}_\alpha^2$ instead of D_α and $V_\alpha(\hat{\mu}_i)$ instead of $V_\alpha(y_i)$. The corresponding dispersion estimate is $X_\alpha^2(y,\hat{\mu})/n$. So Wedderburn's suggestion (4) can be interpreted as a pseudo-likelihood (PL) estimate for given α, which additionally takes into account the loss in degrees of freedom resulting from estimating β.

The comparative study in Nelder and Lee[3] is based on simulations and examples. It demonstrates that PL estimates are often outperformed by the (extended) QL estimators in terms of the mean-squared error.

3 BOOTSTRAP TECHNIQUES

3.1 Covariance-matrix of the parameter estimates

Our particular interest is now to estimate the variance-covariance matrix of $\hat{\beta}$ or of any other function in those parameters. A first order expansion of the score vector $u_\beta(y,\hat{\mu})$ in (1) around the true parameter β gives

$$0 \approx u_\beta(y,\mu) + E_p\left(\frac{\partial u_\beta(y,\mu)}{\partial \beta^t}\right)(\hat{\beta}-\beta) = u_\beta(y,\mu) - Var_p\left(u_\beta(y,\mu)\right)(\hat{\beta}-\beta),$$

where the score-derivative is replaced by its expected value $E_p(\cdot)$ under the considered model with p parameters. This is also known as Fisher's scoring technique for estimating β. Let $X = (x_1,\dots,x_n)^t$, $V = \mathrm{diag}(V(\mu_i))$ and $W = \mathrm{diag}(h_i'^2/V(\mu_i))$ with $h_i' = \partial h_i/\partial\eta_i$. Then Fisher's information matrix is $Var_p(u_\beta(y,\mu)) = X^tWX/\sigma^2$. With this the approximation above yields

$$\begin{aligned} Var(\hat{\beta}) &\approx Var_p^{-1}\left(u_\beta(y,\mu)\right) Var(u_\beta(y,\mu)) Var_p^{-1}\left(u_\beta(y,\mu)\right) \\ &= \sigma^4 (X^tWX)^{-1} Var(u_\beta(y,\mu))(X^tWX)^{-1}. \end{aligned} \tag{10}$$

If the assumed variance $Var(y) = \sigma^2 V(\mu)$ holds then the information sandwich (10) simplifies to the well known asymptotic result

$$Var(\hat{\beta}) \approx \sigma^2 (X^tWX)^{-1}, \tag{11}$$

usually estimated in replacing the unknown μ by $\hat{\mu}$. In case of unknown σ^2 we substitute (4) or (8).

The ML estimate $\hat{\beta}$ can also be analysed by the expansion mentioned above. If the true parameter β is known then

$$\hat{\beta}^1 = (X^tW(\beta)X)^{-1}X^tW(\beta)z(\beta) \tag{12}$$

is a one-step approximation of $\hat{\beta}$ with vector of pseudo-observations

$$z(\beta) = X\beta + W^{-1/2}(\beta)V^{-1/2}(\beta)(y-\mu(\beta)) = X\beta + W^{-1/2}(\beta)\,\epsilon(\beta). \tag{13}$$

Actually, β is unknown and (12) is used iteratively by starting with a proper value β^0 and iterating until convergence. Note that the error terms $\epsilon_i(\beta)$ are independent with $E(\epsilon(\beta)) = 0$ and $Var(\epsilon(\beta)) = \sigma^2 I_n$. So (13) defines a heteroscedastic linear model for $z(\beta)$ with $E(z(\beta)) = X\beta$ and $Var(z(\beta)) = \sigma^2 W^{-1}(\beta)$.

3.2 Residual Resampling

Simonoff and Tsai[6] and also Moulton and Zeger[7] introduced a bootstrap procedure which yields another variance estimation technique. They adapted the well known mechanism from ordinary linear homoscedastic models simply onto model (13). Resampling e_i^* (with replacement) from centered residuals $e_i = \epsilon_i(\hat{\beta}) - \bar{\epsilon}(\hat{\beta})$, with $\bar{\epsilon}(\beta) = n^{-1}\sum_j \epsilon_j(\beta)$, generates independent bootstrap replicates of the pseudo-observations

$$z^* = X\hat{\beta} + W^{-1/2}(\hat{\beta})\,e^* \tag{14}$$

with $E^*(z^*) = X\hat{\beta}$ and $Var^*(z^*) = \tilde{\sigma}^2 W^{-1}(\hat{\beta})$. These moments correspond to the bootstrap distribution which puts mass $1/n$ at each observed e_i. Following (12), the bootstrap values of the parameters are

$$\hat{\beta}^* = (X^t W(\hat{\beta})X)^{-1} X^t W(\hat{\beta}) z^* \tag{15}$$

with bootstrap moments $E^*(\hat{\beta}^*) = \hat{\beta}$ and $Var^*(\hat{\beta}^*) = \tilde{\sigma}^2 (X^t W(\hat{\beta})X)^{-1}$. The latter corresponds to the estimated version of (11) having a modified dispersion $\tilde{\sigma}^2 = \sum_i e_i^2/n$ instead of using $\hat{\sigma}_X^2$ or $\hat{\sigma}_D^2$. Note that (15) is a one-step procedure only.

Simonoff and Tsai[6] sampled from the centered residuals e_i, whereas Moulton and Zeger[7] scaled these error terms by $(1-\hat{H}_{ii})^{-1/2}$, where $\hat{H}_{ii}$ is the i-th diagonal element of the generalized hat matrix $H(\hat{\beta}) = W^{1/2}(\hat{\beta})X(X^t W(\hat{\beta})X)^{-1} X^t W^{1/2}(\hat{\beta})$. It is shown in Friedl[8], that scaling the residuals e_i simply by $(1-p/n)^{-1/2}$ is also appropriate and that the moments of the resulting dispersion estimate $\tilde{\sigma}^2$ are comparable with those of the mean Pearson statistic (4). Therefore, resampling in our case study is done from scaled and centered residuals

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)(1-p/n)}} - \frac{1}{n}\sum_{j=1}^{n} \frac{y_j - \hat{\mu}_j}{\sqrt{V(\hat{\mu}_j)(1-p/n)}} \tag{16}$$

yielding a bootstrap dispersion

$$\tilde{\sigma}^2 = \frac{1}{n-p}\left(\sum_{i=1}^{n} \epsilon_i^2(\hat{\beta}) - n\bar{\epsilon}^2(\hat{\beta})\right). \tag{17}$$

The asymptotic behaviour of this estimate is studied in Friedl[8].

3.3 Score Resampling

A second approach is directly based on the variance sandwich (10), where the approximative variance-covariance matrix of $\hat{\beta}$ is given as product of the inverse Fisher information matrix, which is part of the Fisher scoring method defining $\hat{\beta}$, and the true variance-covariance matrix of the score. Therefore, Lele[9] has used a resampling technique to construct an estimate for $Var(u_\beta(y,\hat{\mu}))$ which should result in a useful estimate of $Var(\hat{\beta})$.

The score vector (1) is the sum of n independent components $u_\beta(y_i,\mu_i)$ each of zero mean but unequal variances

$$Var(u_\beta(y_i,\mu_i)) = \frac{Var(y_i)}{\sigma^4 V^2(\mu_i)} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta^t},$$

the derivative evaluated at the true parameter β. Since the second moments depend on covariates, they are different and we have to resample under a non i.i.d. model. This individual moment structure is reflected by use of Wu's[10] weighted bootstrap

$$u_\beta^*(y_i,\hat{\mu}_i) = u_\beta(y_i,\hat{\mu}_i)t_i^*,$$

which is also referred to as 'wild' or 'external' bootstrap. Here, the t_i^* are drawn from some distribution F_t with zero mean and unit variance. Hence, with respect to F_t we have $E^*(u_\beta^*(y_i,\hat{\mu}_i)) = 0$ and

$$Var^*(u_\beta^*(y_i,\hat{\mu}_i)) = \frac{(y_i - \hat{\mu}_i)^2}{\sigma^4 V^2(\hat{\mu}_i)} \frac{\partial \mu_i}{\partial \beta} \frac{\partial \mu_i}{\partial \beta^t},$$

with derivative at the ML estimate $\hat{\beta}$. Thus the weighted bootstrap yields a plug-in estimate where the variance of y_i is estimated through $(y_i - \hat{\mu}_i)^2$. To reflect the independence of the components $u_\beta(y_i, \mu_i)$ we have to resample the t_i^*'s independently. Therefore, the bootstrap variance of the total bootstrap score $u_\beta^*(y, \hat{\mu}) = \sum_i u_\beta^*(y_i, \hat{\mu}_i)$ simply adds up resulting in

$$Var^*(u_\beta^*(y, \hat{\mu})) = \frac{1}{\sigma^4} X^t S^2(\hat{\beta}) W(\hat{\mu}) X$$

with $S^2(\hat{\beta}) = \text{diag}((y_i - \hat{\mu}_i)^2 / V(\hat{\mu}_i))$. Substituting this expression into the variance formula (10) yields a variance sandwich of the form

$$Var^*(\hat{\beta}^*) = (X^t W(\hat{\beta}) X)^{-1} X^t S^2(\hat{\beta}) W(\hat{\beta}) X (X^t W(\hat{\beta}) X)^{-1}. \tag{18}$$

Friedl[11] has shown that the corresponding parameter values can be written again as one-step estimate

$$\hat{\beta}^* = (X^t W(\hat{\beta}) X)^{-1} X^t W(\hat{\beta}) z^*, \tag{19}$$

with resampled pseudo-observations

$$z^* = X\hat{\beta} + T^* W^{-1/2}(\hat{\beta}) \epsilon(\hat{\beta}),$$

where $T^* = \text{diag}(t_i^*)$. Obviously, we are using here all observed residuals from the ML fit, each multiplied by an individual realisation t_i^*. So the bootstrap moments are $E^*(z^*) = X\hat{\beta}$ and $Var^*(z^*) = S^2(\hat{\beta}) W^{-1}(\hat{\beta})$. Hence, the elements of S^2 estimate the dispersions of the responses. Consequently, $\hat{\beta}^*$ has again mean $E^*(\hat{\beta}^*) = \hat{\beta}$ but variance (18).

If all observed residuals are equal then $S^2(\hat{\beta}) = \hat{\sigma}^{2*} I_n$ holds and each response y_i has bootstrap dispersion $\hat{\sigma}^{2*}$. For that case the expression above simplifies to a direct estimation of the asymptotic result (11). Moreover, the mean Pearson statistic (4) can be interpreted as the arithmetic mean of all individual dispersion estimates based on a bootstrap technique.

In their simulation study Friedl and Tilg[12] found that this technique is not appropriate for overdispersed binomial responses in logistic regression models with moderate sample sizes. This may be caused by the very special structure of $S^2(\hat{\beta})$ in this case, where each element is used as an individual dispersion estimate based only on a sample of size one.

3.4 Resampled Responses

The estimation procedures for the variance-covariance matrix of $\hat{\beta}$ avoided resampling of y values, but for other applications bootstrap quantities y_i^* are explicitely needed. They are defined as

$$y_i^* = \hat{\mu}_i + \sqrt{V(\hat{\mu}_i)} e_i^*, \tag{20}$$

which corresponds to Efron's[13] bootstrap in heteroscedastic linear regression models. As in the residual resampling approach we are sampling from the centered and scaled residuals e_i given in (16). Centering is necessary to reflect the considered mean model $E^*(y_i^*) = \hat{\mu}_i$ whereas $Var^*(y_i^*) = \tilde{\sigma}^2 V(\hat{\mu}_i)$. Moreover, as shown in Friedl[8] this kind of scaling implies $E(\tilde{\sigma}^2) \approx E(X^2(y, \hat{\mu}))$ with X^2 from (4).

Bootstrap replicates $\hat{\beta}^*$ are obtained by applying the iterative weighted least squares algorithm (12) onto the resampled data (y_i^*, x_i). However, with this technique a restricted response range can be lost by artificially assuming linearity of the errors in (20). E.g. in case of discrete distributions, where only non-negative integers are observable, the range restriction is guaranteed by the adaptation

$$y_i^* = \max\{\hat{\mu}_i + \sqrt{V(\hat{\mu}_i)} e_i^*, 0\}. \tag{21}$$

If necessary, integer values can be constructed by rounding these values to the nearest integer.

4 CASE STUDY: AIR SAMPLER DATA

To utilize the methods and techniques described in Sections 2 and 3 we analyse data from environmental microbiology. During a period of one year every two weeks airborne microorganisms were monitored at seven outdoor sample sites in and around Graz, Austria. For this an Andersen six-stage viable microbial particle sizing air sampler was used which ran four minutes drawing particles at a rate of one cubic feet (28.3 liters) per minute. Then the six Petri dishes were removed, each containing an agar medium appropriate for the microorganisms which may be encountered. After incubation the total number of colonies formed was counted. Motivation of this investigation was to determine the levels of aerogen bacteria and molds in the air of Greater Graz. Thus, the data will definitely describe the number of bacteria observed within 113.2 liters of air.

The study was designed (i) to evaluate the effects of measuring at different locations and (ii) to analyse the seasonal influence described by the concurrent temperature in degree Celsius. A sampling plan was developed carefully to monitor, control and model air quality with respect to bacteria load for seven typical regions of a medsize town represented by the following locations:

1. village zone, nearby big farms with liquid manure pits and dung-hills,
2. grassland and arable land, without buildings,
3. suburban area with one-family houses and small farms,
4. busy crossing, nearby a slaughter-house,
5. public park on top of the Schloßberg in the center of Graz,
6. living area with apartment buildings and gardens,
7. like 6 but additionally with compost arrangements.

Our statistical analysis is based on a sample of $n = 178$ elements and has been carried out in GLIM4[14]. Four measurements are missing because of bad weather conditions.

The summary statistics in Table 1 are based on the site specific samples each consisting of about 26 observations. Means and medians show plausible results, which has been expected at those locations. Moreover, standard deviations and interquartile ranges seem to be strongly related to their respective mean values. The series of boxplots in Figure 1 give a more detailed insight into the data and reflect the behaviour just discussed.

For an appropriate model we assume that the mean counts at different sites are proportional and that the factors of proportionality are constant over the year. We further suppose that the same seasonal effect is present for each site and that the number of colonies uniformly changes with temperature. Because no zero bacteria count was observed, a log-link will be used to retain the multiplicative mean structure, based only on main effects.

site	1	2	3	4	5	6	7
mean	43.6	24.3	25.7	20.6	15.8	10.8	12.7
s.d.	25.5	17.0	23.6	12.4	11.5	5.9	10.5
med.	36.0	17.5	15.0	16.0	13.0	8.5	11.0
iqr	28.0	28.0	23.0	18.0	10.0	9.0	9.0

Table 1: Estimates for the mean-variance relation

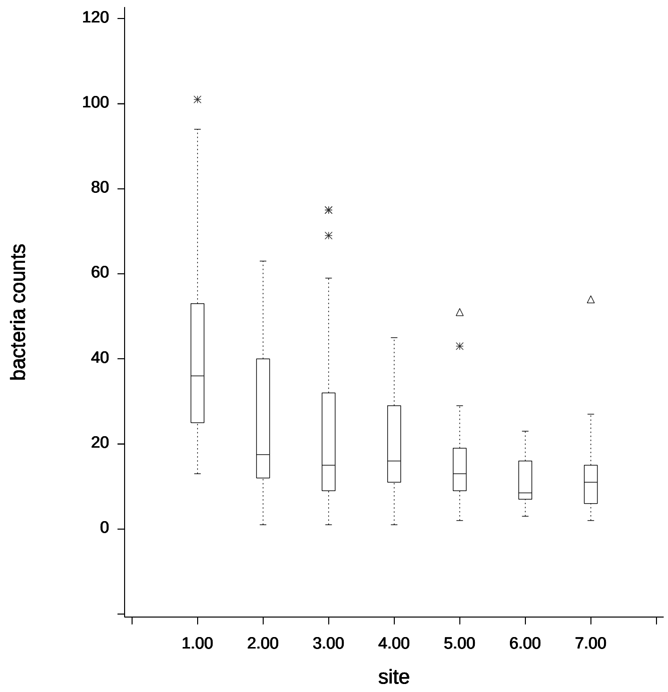

Figure 1: Boxplots of bacteria counts observed at seven different sites

4.1 Loglinear Models

The first step is to study the loglinear model of the form

$$\eta = \log(\mu) = \beta_1 + \beta_{\mathrm{site}_j} + \beta_8 \text{ temp} \qquad j = 2, \ldots, 7$$

for which the counts y_i are assumed to be Poisson distributed with $V(\mu_i) = \mu_i$, $\sigma^2 = 1$. The resulting deviance (2) of 1586.6 is about ten times as large as the residual degrees of freedom 170. This suggests to consider also a model assuming overdispersed Poisson variation in the data. Hence, $\sigma^2 > 1$ is treated to be unknown and estimated by the mean deviance (8) with $\alpha = 1$ as $\hat{\sigma}^2_D = 9.33$. The mean Pearson statistic (4) as well as the bootstrap dispersion (17) result in $\hat{\sigma}^2_X = \tilde{\sigma}^2 = 10.11$, caused by a fairly small residual mean. Estimates of both models are summarized in Table 2, columns 1, 2 and 3, 4, respectively.

Because of the large overdispersion, Poisson variation seems to be inappropriate. Moreover, from Table 1 we observe that the site specific median values are nearly of the same magnitude as the respective interquartile ranges implying rather a quadratic variance structure (gamma model) than a linear one (Poisson model). An adequate log-link model with quadratic variance results in a deviance of 88.2 and a scale parameter of about 0.50 for the estimates (4) or (17) and 0.52 in case of using (8). Table 2 (columns 5, 6 and 7) shows that the parameter estimates are similar those of the overdispersed Poisson model, but with slightly larger standard errors.

The assumption of a standard negative-binomial model for GLM's accords to the use of a variance model of the form $Var(y_i) = \mu_i + \mu_i^2/\theta$. Hereby, the unknown parameter θ is assumed to be fixed as μ varies. As in the Poisson model it is also assumed that there is no extra parameter to describe dispersion ($\sigma^2 = 1$). The ML estimates for β and θ are calculated as discussed in Nelder and Lee[3]. The values of $\hat{\beta}$, given in Table 2 column 8, are very similar those assuming linear or quadratic variance. Like in case of the gamma model, the estimated standard errors of the site effects in column 9 are all nearly of the same size but a bit smaller those under the

$Var(y) = \mu$		$Var(y) = \sigma^2\mu$		$Var(y) = \sigma^2\mu^2$			$Var(y) = \mu + \mu^2/\theta$		
estim.	s.e.	s.e. (4)	s.e. (8)	estim.	s.e. (4)	s.e. (8)	estim.	s.e.	parameter
3.499	.039	.123	.118	3.469	.157	.159	3.474	.142	1
-0.586	.050	.159	.153	-0.618	.198	.202	-0.614	.180	site(2)
-0.537	.050	.158	.152	-0.595	.200	.204	-0.587	.182	site(3)
-0.785	.054	.170	.164	-0.826	.201	.204	-0.819	.183	site(4)
-1.053	.059	.187	.180	-1.035	.201	.204	-1.040	.185	site(5)
-1.432	.067	.213	.205	-1.424	.199	.202	-1.428	.186	site(6)
-1.263	.063	.200	.192	-1.252	.199	.202	-1.256	.185	site(7)
0.020	.002	.005	.005	0.023	.005	.005	0.023	.005	temp

Table 2: Loglinear models assuming Poisson, overdispersed Poisson, gamma and negative-binomial variability for bacteria counts

quadratic variance model. The ML estimate of θ is 2.621 with standard error 0.311, which is another indicator against the inappropriate Poisson model ($\theta \to \infty$).

Figure 2 compares the (estimated) variance models from above with the site specific mean-variance relations where the overdispersed Poisson and gamma models are plotted under both dispersion estimates. In the Poisson case the deviance based estimate is smaller than the mean Pearson statistic but vice versa for the gamma case. Both, quadratic and negative-binomial variability deliver good fits of the sample estimates. However, in the following we will concentrate on the purely quadratic variance model and look at it as a member of the power variance family $Var(y_i) = \sigma^2\mu_i^\alpha$.

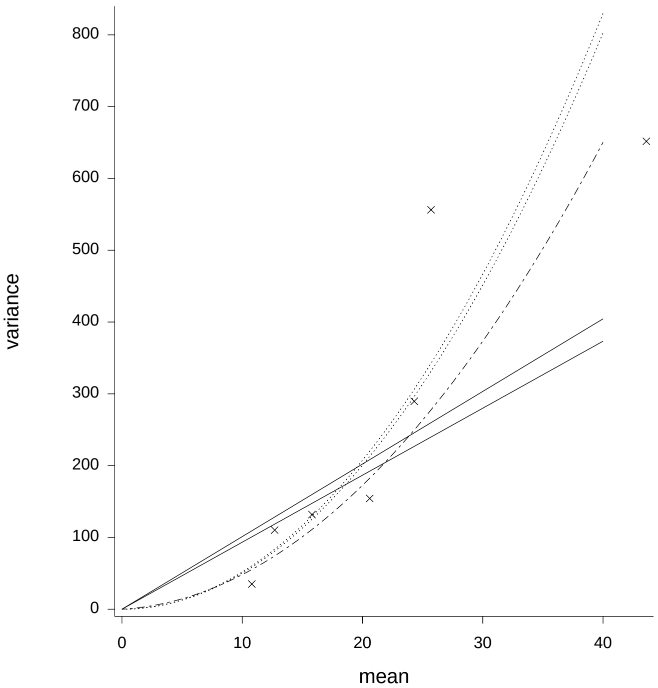

Figure 2: Variance models considered: overdispersed Poisson (solid), quadratic variance (dotted), negative-binomial (dashed), sample estimates ($\times$)

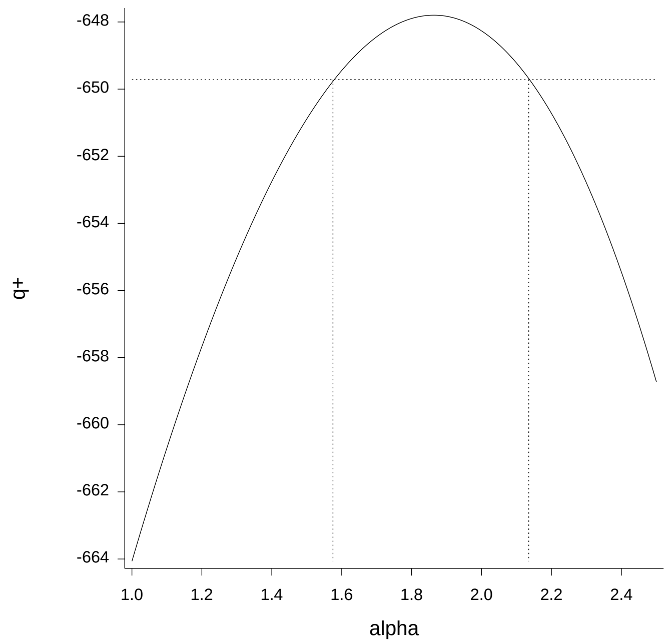

Figure 3: Extended quasi likelihood function with 95% profile quasi-likelihood interval for the power variance family $V_\alpha(\mu) = \mu^\alpha$

4.2 Checking the variance function

With the methods of Section 2 we can assess both variance structures, the overdispersed Poisson and the gamma. Given α in the power variance family, we maximize the EQL function (6) in its finite-sample adjusted form

$$q_\alpha^+(y,\hat{\mu}) = -\frac{1}{2}(n-p) - \frac{1}{2}\frac{n-p}{n}\left(n\log\left(2\pi\frac{D_\alpha(y,\hat{\mu})}{n-p}\right) + \alpha\sum_{i=1}^{n}\log(y_i)\right) \tag{22}$$

by the maximum QL estimate $\hat{\beta}$ where σ^2 is replaced by its EQL estimate (8). Figure 3 displays $q_\alpha^+(y,\hat{\mu})$ for a range of α values $(1.0(0.005)2.5)$ and also the 95% profile QL interval $(1.54, 2.10)$, where the overdispersed Poisson variance with $\alpha = 1$ is not included. However, the profile QL estimate $\hat{\alpha} = 1.84$ maximizes (22) which is close to the gamma variance with $\alpha = 2$.

It is also possible to obtain respective bootstrap results by applying Monte-Carlo techniques. Since the EQL function (22) depends on the minimized deviance as well as on the observation vector y we have to use replications of y_i as defined in (20). Let $\hat{\alpha}$, $\hat{\beta}$ and $\hat{\sigma}^2$ denote the parameter values which maximize $q_\alpha^+(y,\mu)$. In a first stage we resample under this estimated EQL model using replicates $y_i^* = \hat{\mu}_i + \hat{\mu}_i^{\hat{\alpha}/2} e_i^*$ where the bootstrap residuals e_i^* are resampled from (16) with $V(\hat{\mu}_i)$ replaced by $V_{\hat{\alpha}}(\hat{\mu}_i) = \hat{\mu}_i^{\hat{\alpha}}$. But sometimes, if this procedure results in negative replications y_i^* a correction according to (21) is indicated. In the second stage the bootstrap data (y_i^*, x_i) are fitted by EQL assuming a variance model of the form $\sigma^2\mu^\alpha$ with σ^2 and α both unknown. This gives a bootstrap value $\hat{\alpha}^*$ through the profile QL technique.

These two stages are repeated $B = 1000$-times and result in an arithmetic mean of 1.69 (estimated standard deviation of 0.11), which can be used to approximate the expected value of $\hat{\alpha}$. Furthermore, a 95% bootstrap confidence interval based on the respective percentiles

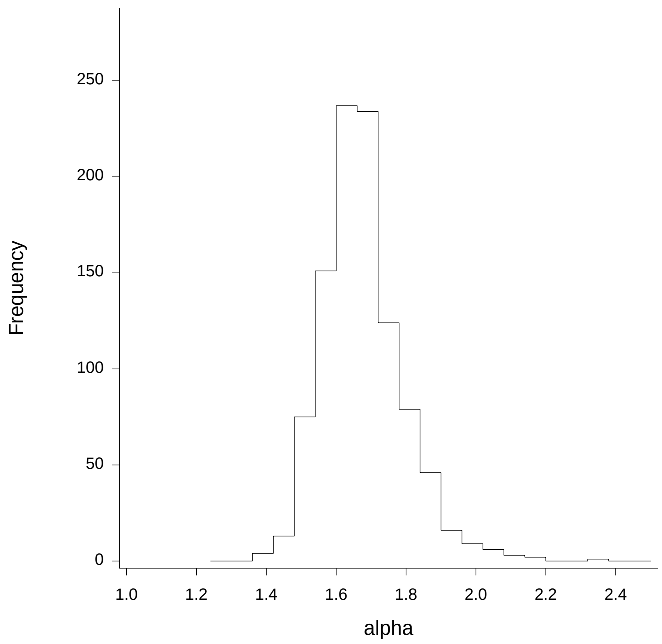

Figure 4: Bootstrap histogram of 1000 replications $\hat{\alpha}^*$ for the power variance family $V_\alpha(\mu) = \mu^\alpha$

of the bootstrap histogram in Figure 4 is $(1.50, 1.96)$, which is shifted to the left and a bit smaller compared with the respective profile QL interval $(1.54, 2.10)$. This percentile interval is of correct coverage probability if the bootstrap distribution of $\hat{\alpha}^*$ has median $\hat{\alpha}$. Otherwise, a bias-corrected form of this interval should be used. The Monte-Carlo estimation of this so called BC-interval (without acceleration) gives $(1.74, 2.36)$ and covers also the quadratic variance function. But we have to be careful by interpreting the right bound, because it already represents the maximum of all $\hat{\alpha}^*$ realisations. Finally, another bootstrap confidence interval is constructed by the hybrid method as $(\hat{\alpha} - h^*_{0.975}/\sqrt{n}, \hat{\alpha} - h^*_{0.025}/\sqrt{n})$ where h^*_γ denotes the γ-quantile of the histogram of $\sqrt{n}(\hat{\alpha}^* - \hat{\alpha})$. The result $(1.72, 2.18)$ covers the quadratic variance function with $\alpha = 2$, is shifted to the right and a bit smaller compared to the QL interval. A discussion of bootstrap confidence intervals is given in Shao and Tu[15].

4.3 Comparing non-nested models

In the sequel we restrict our considerations to models with a quadratic variance function as suggested in the previous discussion. This is equivalent to the assumption that the coefficient of variation is constant for all observations, i.e. $Var(y_i) = \sigma^2\mu_i^2$ with $E(y_i) = \mu_i$. For a small value of σ^2, the corresponding variance-stabilizing transformation is $\log(y_i)$, which has approximative moments $E(\log y_i) \approx \log \mu_i - \sigma^2/2$ and $Var(\log y_i) \approx \sigma^2$. Therefore, it will be interesting whether a lognormal distribution for y_i is also supported by the data. Asymptotic efficiencies of both approaches are compared in Firth[16]. For finite sample sizes the bootstrap will possibly help to distinguish between the two non-nested models:

G: $E(y_i) = \exp(\eta_i)$, $Var(y_i) = \sigma^2\mu_i^2$,

L: $E(\log y_i) = \eta_i$, $Var(\log y_i) = \sigma^2$.

gamma		lognormal		
estim.	s.e.	estim.	s.e.	parameter
3.469	.159	3.260	.172	1
-0.618	.202	-0.707	.218	site(2)
-0.595	.204	-0.895	.221	site(3)
-0.826	.204	-0.871	.221	site(4)
-1.035	.204	-1.124	.221	site(5)
-1.424	.202	-1.429	.219	site(6)
-1.252	.202	-1.371	.218	site(7)
0.023	.005	0.027	.006	temp

Table 3: Results for the gamma/log-link and lognormal/identity-link model

Parameter estimates and their standard errors are given in Table 3. The observed deviance for the gamma model is 88.18 whereas for the lognormal model it is 103.26 giving an observed deviance difference of -15.08. The estimated dispersion parameters based on (8) are 0.519 and 0.607, respectively. To compare both intercept parameters we have to adjust the intercept of the lognormal model by adding $\hat{\sigma}^2/2 = 0.304$.

Hinde[17] chooses between non-nested models using resampled deviance differences of both models. To simulate such differences we have to assume model G as well as model L to be the correct one. Therefore, bootstrapping is based on the following two steps:

1. Fit the data (y_i, x_i) under model G and denote the estimates by $\hat{\mu}_{i|G}$ and the residuals by $\epsilon_{i|G} = (y_i - \hat{\mu}_{i|G})/\hat{\mu}_{i|G}$.
   Given this estimated gamma model simulate bootstrap responses $y^*_{i|G} = \hat{\mu}_{i|G} + \hat{\mu}_{i|G}\epsilon^*_{i|G}$ and fit again model G based on $(y^*_{i|G}, x_i)$ which results in the deviance $D^*_{G|G}$. Use the same resampled responses to fit also model L for the data $(\log y^*_{i|G}, x_i)$ and get $D^*_{L|G}$. Calculate the deviance differences $d^*_G = D^*_{G|G} - D^*_{L|G}$.

2. Fit $(\log y_i, x_i)$ under model L to get the estimate $\hat{\mu}_{i|L}$ for the mean of $\log y_i$ and corresponding residuals $\epsilon_{i|L} = \log y_i - \hat{\mu}_{i|L}$.
   Now assume that the lognormal model holds and simulate responses $\log y^*_{i|L} = \hat{\mu}_{i|L} + \epsilon^*_{i|L}$. Fitting model G for $(y^*_{i|L}, x_i)$ and model L for $(\log y^*_{i|L}, x_i)$ yields deviances $D^*_{G|L}$ and $D^*_{L|L}$, respectively. Define deviance differences $d^*_L = D^*_{G|L} - D^*_{L|L}$.

Histograms of 1000 replicates of d^*_G and d^*_L are shown in Figure 5 by solid and dotted lines. The Monte-Carlo mean (standard deviation) is -24.07 (11.63) under model G and -13.58 (8.80) in case of model L. Hence, the observed difference of -15.08 is near to the mean difference under the lognormal model. Moreover, given model G 85.3% of the replicates are smaller than the observed difference whereas under model L this percentage is 36.2%.

This shows no clear support but gives a slight indication that the data are also consistent with the simpler linear regression model. However, to demonstrate how the bootstrap works within the class of GLM's we will further on assume quadratic variability of the responses.

4.4 Estimating standard errors

Up to now, we only used techniques based on resampled responses y^*. Resampling from residuals or from score components can be also applied easily under the quadratic variance model with log-link function. Because of the very special structure $h'(\eta)^2 = V(\mu)$ the matrix of weights simplifies to $W(\hat{\beta}) = I_n$.

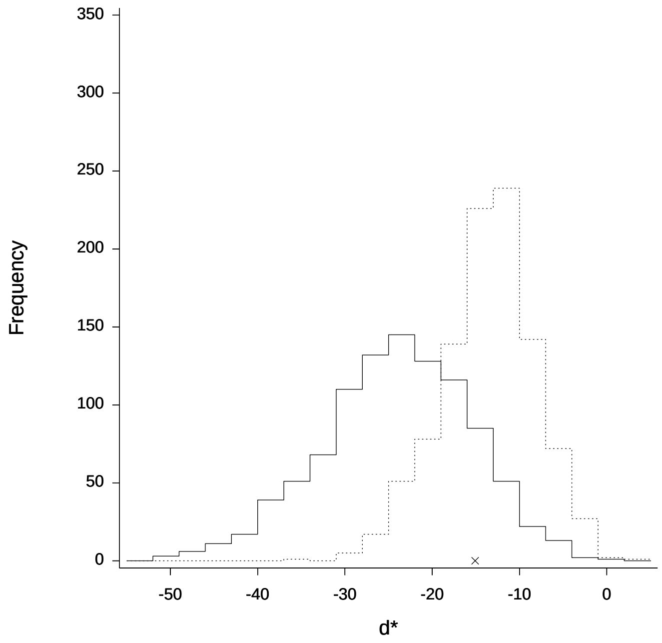

Figure 5: Bootstrap histograms of deviance differences given the estimated gamma (solid lines) and lognormal (dotted lines) model

Thus, for the residual resampling approach we have to consider the replicates $z^* = X\hat{\beta} + e^*$, where the e_i^*'s are drawn from (16). Bootstrap values of the parameters are easily calculated from (15) giving

$$\hat{\beta}^* = \hat{\beta} + (X^tX)^{-1}X^te^*.$$

Boxplots of 1000 parameter replicates $\hat{\beta}_1^*, \hat{\beta}_1^* + \hat{\beta}_j^*, j = 2, \ldots, 7$, characterizing the total site effects of every location are compared in Figure 6. All of them show rather symmetrical distributions and reflect equal variability. Respective means and standard deviations of the replicates $\hat{\beta}_j^*$ can be found in Table 4, columns 1 and 2. These Monte-Carlo results confirm the theoretical bootstrap moments $E^*(\hat{\beta}^*) = \hat{\beta}$ and $Var^*(\hat{\beta}^*) = \tilde{\sigma}^2(X^tX)^{-1}$, where the square roots of the diagonal elements of the latter are given in the third column.

To study the score resampling technique, we generate a set of 1000 bootstrap values according to (19), namely

$$\hat{\beta}^* = \hat{\beta} + (X^tX)^{-1}X^tT^*\epsilon(\hat{\beta}),$$

where the t_i^*'s are drawn from a standard normal distribution. As stated before, bootstrap moments are again $E^*(\hat{\beta}^*) = \hat{\beta}$ but $Var^*(\hat{\beta}^*) = (X^tX)^{-1}X^tS^2(\hat{\beta})X(X^tX)^{-1}$. The resulting mean values, given in Table 4, are very close the corresponding likelihood results. This indicates once more, that none of these techniques can be used to describe or to reduce the bias of the likelihood estimator.

Monte-Carlo and theoretical standard deviations are nearly the same, but both differ from the residual resampling outcome in the following two directions: Firstly, they are generally smaller and, secondly, they vary much more than the respective estimates from the residual resampling approach. Especially the estimates according to sites 3, 5 and 7 are conspicuously larger than those of 2, 4 and 6. This is mainly caused by the use of $S^2(\hat{\beta})$, the matrix of squared residuals. From Figure 1 it is obvious that there are some sites with extremely large observations

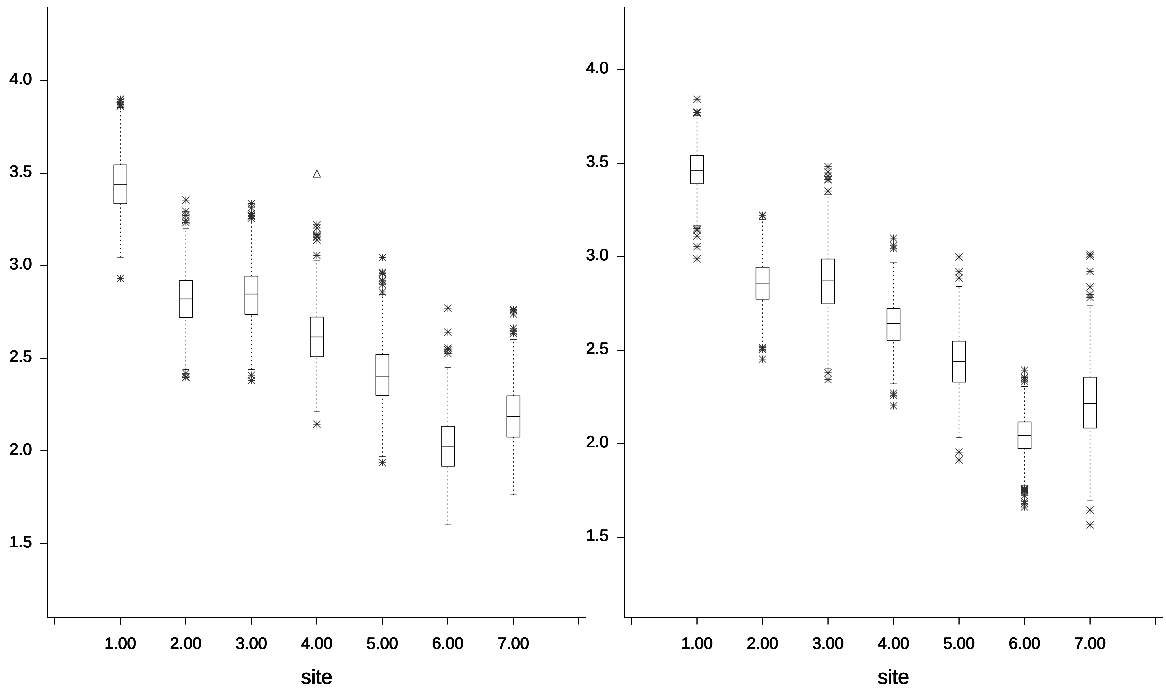

Figure 6: Boxplots of 1000 replicates of the site effects based on resampling from residuals (left) and from score components (right)

yielding huge residuals. For example, the most striking value at site 7 gives a squared residual larger than 14, whereas all others are smaller than five. Therefore, the standard deviation of a special site parameter estimated by score resampling just reflects the goodness-of-fit observed at this location and illustrates the behaviour of the respective group of data with regard to the assumed and estimated model. In contrast, the estimated deviations from residual resampling are determined in terms of the fixed design matrix X multiplied by a single mean measure of goodness-of-fit based on all observed data.

Residual Resampling			Score Resampling			
mean(MC)	s.d.(MC)	s.d.(th)	mean(MC)	s.d.(MC)	s.d.(th)	parameter
3.445	.154	.157	3.465	.116	.119	1
-0.621	.189	.198	-0.608	.163	.164	site(2)
-0.599	.195	.200	-0.596	.205	.206	site(3)
-0.825	.196	.201	-0.825	.159	.156	site(4)
-1.032	.202	.201	-1.026	.180	.181	site(5)
-1.415	.197	.199	-1.423	.144	.149	site(6)
-1.252	.205	.199	-1.249	.212	.210	site(7)
0.023	.005	.005	0.023	.004	.004	temp

Table 4: Estimates from the respective histograms of 1000 parameter replicates (MC) compared with theoretical bootstrap results (th) given a quadratic variance model

5 CONCLUSIONS

Our paper provides an incomplete survey on likelihood concepts and resampling methods in GLM's, where the design is considered to be fixed. Moreover, it is assumed that the observations are independent and that the mean model is specified correctly with respect to the link function and the linear predictor. Hence, neither variable selection procedures nor observation vector resampling are discussed, the latter being only appropriate to study the correlation structure between the set of explanatory variables and the response variate.

However, the main advantage of the QL approach is to replace the full distributional specification about the random component by a suitable variance assumption. It is used in the case study to consider loglinear models for bacteria counts with site as factor and temperature as variable, assuming (overdispersed) Poisson, gamma and negative-binomial variability. Resampling from residuals automatically corrects for dispersion and the bootstrap dispersion parameter has asymptotically similar properties as the mean Pearson statistic. The score resampling approach is also robust against mis-specified variance structures. However, the results of our case study show no substantial differences between the QL and bootstrap variance estimators.

An extension of the QL function allows for the comparison of different variance functions embedded into a parameterized family like the power family. The choice of quadratic variability for our data is supported by a profile QL interval and by a variety of bootstrap intervals.

In a Monte-Carlo study we have also compared two non-nested models by using deviance differences. To discriminate between these models the observed difference is related to the simulated sets of resampled deviance differences under each of the estimated models. The simulation indicates that both models are consistent to the data, however, with a slight preference for the lognormal model.

Acknowledgments

The authors are grateful for the supply of the data. The corresponding project was carried out by the Information Centre of Environmental Protection, Graz, (project no. 01/95) in co-operation with the Institute of Hygiene of the University of Graz.

References

[1] McCullagh, P. and Nelder, J.A. *Generalized Linear Models (2nd edn)*, Chapman and Hall, London, 1989.

[2] Wedderburn, R.W.M. 'Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method', *Biometrika*, **61**, 439–447 (1974).

[3] Nelder, J.A. and Lee, Y. 'Likelihood, quasi-likelihood and pseudolikelihood: Some comparisons', *Journal of the Royal Statistical Society, Series B*, **54**, 273–284, (1992).

[4] Nelder, J.A. and Pregibon, D. 'An extended quasi-likelihood function', *Biometrika*, **74**, 221–232, (1987).

[5] Carroll, R.J. and Ruppert, D. 'Robust estimation in heteroscedastic linear models', *Annals of Statistics*, **10**, 429–441 (1982).

[6] Simonoff, J.S. and Tsai, C.-L. 'Jackknifing and bootstrapping quasi-likelihood estimators', *Journal of Statistical Computation and Simulation*, **30**, 213–232, (1988).

[7] Moulton, L.H. and Zeger, S.L. 'Bootstrapping generalized linear models', *Computational Statistics & Data Analysis*, **11**, 53–62, (1991).

[8] Friedl, H. 'On the asymptotic moments of Pearson type statistics based on resampling procedures', submitted, (1996).

[9] Lele, S. 'Resampling using estimating equations', *In: Estimating Functions (ed. by V.P. Godambe)*, Oxford University Press, Oxford, 295–304, (1992).

[10] Wu, C.F.J. 'Jackknife, bootstrap and other resampling methods in regression analysis (with discussion)', *Annals of Statistics*, **14**, 1261–1350, (1986).

[11] Friedl, H. 'Variance estimation in generalized linear models using the bootstrap', *In: Symposia Gaussiana, Conf. B (ed. by Mammitzsch and Schneeweiß)*, De Gruyter, Berlin, 217–226, (1995).

[12] Friedl, H. and Tilg, N. 'Variance estimates in logistic regression using the bootstrap', *Communications in Statistics – Theory and Methods*, **24**, 473–486, (1994).

[13] Efron, B. 'Bootstrap methods: Another look at the jackknife', *Annals of Statistics*, **7**, 1–26, (1979).

[14] Francis, B., Green, M. and Payne, C. (eds.) *GLIM4 – The Statistical System for Generalized Linear Interactive Modelling*, Clarendon Press, Oxford, 1993.

[15] Shao, J. and Tu, D. *The Jackknife and Bootstrap*, Springer, New York, 1995.

[16] Firth, D. 'Multiplicative errors: log-normal or gamma?', *Journal of the Royal Statistical Society, Series B*, **50**, 266–268, (1988).

[17] Hinde, J. 'Choosing between non-nested models: a simulation approach', *In: Advances in GLIM and Statistical Modelling (ed. by Fahrmeir, L. et al.)*, Lecture Notes in Statistics, **78**, Springer, New York, 119–124, (1992).