Variance Estimation in Generalized Linear Models using the Bootstrap

H. Friedl

Abstract

This paper provides an incomplete survey of some approaches to the application of nonparametric bootstrap techniques in generalized linear models. Without exception all of the mentioned resampling schemes are based on first-order approximations. Therefore, they are principally useful to motivate alternative variance estimators for $\hat{\beta}$ but cannot treat aspects of its bias. It is shown that the definitions of the resulting bootstrap replications $\hat{\beta}^*$ differ only with respect to the regarded residuals or the used matrix of weights.

1 Introduction

Consider a generalized linear model (GLM) where the components y_i of $y = (y_1, \ldots, y_n)'$ are independent random variates with corresponding densities or probability mass functions from the exponential dispersion family parameterized by (θ_i, ϕ_i). This equals the notation in McCullagh and Nelder (1989) and implies $\mathrm{E}(y_i) = \mu_i$ and $\mathrm{Var}(y_i) = \phi_i V(\mu_i)$. From an applied point of view it is often appropriate to further assume $\phi_i = \phi a_i$, where ϕ denotes a possibly unknown but constant dispersion parameter and the a_i's are known constants. This special variance structure is obtained by modelling arithmetic means of n_i independently distributed variates ($a_i = 1/n_i$) all with equal variance ϕ, or by considering means of n_i homogeneous Bernoulli variates where we have to choose again $a_i = 1/n_i$ but fix $\phi = 1$, amongst other examples.

Using canonical link functions $g(\mu_i) = \theta_i = \eta_i = x_i'\beta$, where $X = (x_1, \ldots, x_n)'$ is the matrix of known explanatory variables and $\beta = (\beta_1, \ldots, \beta_p)'$ is a vector of unknown parameters, the maximum likelihood estimate (MLE) $\hat{\beta}$ is defined as solution of the score equations

$$u_j(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} \frac{1}{a_i} x_{ij}(y_i - \mu_i(\beta)) = 0, \qquad j = 1, \ldots, p, \tag{1}$$

and is not affected by the specific value of ϕ. In the case of noncanonical link models the derivative $\partial\theta_i/\partial\eta_i$ is appended additionally to the score function.

Because the scores only depend on the mean and the variance, Wedderburn (1974) suggested to use the weaker second moment assumptions rather than assuming an exponential family. This results in the well known quasi-score function, in which $V(\mu_i)^{-1}\partial\mu_i/\partial\eta_i$ plays the same role as $\partial\theta_i/\partial\eta_i$ in the score function for noncanonical links. Therefore, postulating $V(\mu_i) = \partial\mu_i/\partial\eta_i$ gives a quasi-canonical link function and the maximum quasi-likelihood estimate of β is again defined by (1). Thus the properties of this estimate correspond to those of the MLE.

Obviously the system of estimating equations (1) is highly nonlinear in β. Therefore, we use the first-order Newton approximation

$$u(\hat{\beta}) = 0 = u(\beta) + (\partial u(\beta)/\partial\beta')(\hat{\beta} - \beta) \tag{2}$$

leading to a weighted least squares representation

$$\hat{\beta} = (X'W(\mu)X)^{-1}X'W(\mu)z(\mu), \tag{3}$$

in which $W(\mu) = \mathrm{diag}\,(w(\mu_i))$ with $w(\mu_i) = V(\mu_i)/a_i$ and

$$z_i(\mu_i) = \eta_i + \frac{y_i - \mu_i}{V(\mu_i)}, \qquad i = 1, \ldots, n, \tag{4}$$

are standardized adjusted variates. Since $\mathrm{E}(z_i) = \eta_i = x_i'\beta$ and $\mathrm{Var}(z_i) = \phi a_i / V(\mu_i) = \phi / w(\mu_i)$, a linear heteroscedastic model is defined by (4) and $\hat\beta$ in (3) is the weighted least squares estimate where the weights are given by $W(\mu)$. But the z_i's are unknown because we do not know the μ_i's. Therefore, an iterative weighted least squares procedure is applied.

Referring to (3), it follows that at least to a first-order of approximation

$$\mathrm{E}(\hat\beta) = \beta, \quad \mathrm{Cov}(\hat\beta) = \phi(X'W(\mu)X)^{-1}.$$

In case of constant variance and identity link models both properties are exact. However, for the whole class of GLM's these properties hold asymptotically only, as $n \to \infty$. To remove the $O(n^{-1})$ term of the bias we either use an estimate of it like in Copas (1988) or we modify the likelihood function by Jeffreys prior as in Firth (1992, 1993). Hence, both techniques are shrinkage corrections. The covariance structure is typically estimated by calculating $\mathrm{Cov}(\hat\beta)$ at $\mu = \hat\mu$. If the dispersion parameter ϕ is unknown, appropriate estimates are given by the mean residual deviance or by

$$\hat\phi = \frac{1}{n-p} \sum_{i=1}^{n} \frac{(y_i - \hat\mu_i)^2}{a_i V(\hat\mu_i)} = \frac{1}{n-p} X^2, \tag{5}$$

where X^2 is the generalized Pearson statistic.

In the following, two bootstrap approaches will be considered in order to motivate alternative estimates. Both techniques are based on bootstrap methods appropriate for linear models. To construct bootstrap estimators for the bias and the variance-covariance matrix of the least squares estimate $\hat\beta = (X'X)^{-1}X'y$ in heteroscedastic linear models like $y_i = x_i'\beta + \epsilon_i$, where we further assume that the ϵ_i's are independent and all from an unknown distribution function F with zero means and variances σ_i^2, nonparametric bootstrap procedures are mainly based on the exchangeability of the standardized error terms ϵ_i/σ_i. Because of the additivity of the errors in the linear model, the distributions of the responses y_i are completely specified by F and differ only in the first two moments. Hence, resampling r_i^* from the empirical distribution function $\hat F$ of the observed centered and standardized residuals

$$r_i = \frac{e_i - \bar e}{\sqrt{\frac{1}{n}\sum_k (e_k - \bar e)^2}}, \qquad i = 1, \ldots, n,$$

where $e_i = y_i - x_i'\hat\beta$ and $\bar e = n^{-1}\sum_k e_k$, reflects all of the assumptions and resulting properties of the linear model with respect to the bootstrap distribution *, which gives probability mass $1/n$ at every observed r_i. Thus, the moments of r_i^* are given by $\mathrm{E}^*(r_i^*) = 0$ and $\mathrm{Var}^*(r_i^*) = 1$ for all $i = 1, \ldots, n$. The bootstrap sample $y^* = (y_1^*, \ldots, y_n^*)'$ is defined to be

$$y_i^* = x_i'\hat\beta + \hat\sigma_i r_i^*. \tag{6}$$

Wu (1986) used the estimate $\hat\sigma_i = |e_i|/\sqrt{1-h_i}$, where $h_i = x_i'(X'X)^{-1}x_i$ is the i-th diagonal element of the hat matrix. With (6), the bootstrap value corresponding to $\hat\beta$ is

$$\hat\beta^* = (X'X)^{-1}X'y^*$$

and we get $\mathrm{E}^*(\hat\beta^*) = \hat\beta$ and $\mathrm{Cov}^*(\hat\beta^*) = (X'X)^{-1}X'\hat\Sigma X(X'X)^{-1}$, where $\hat\Sigma = \mathrm{diag}(\hat\sigma_i^2)$.

In homoscedastic linear intercept models with $\sigma_i^2 = \sigma^2$, $i = 1, \ldots, n$, resampling from e_i gives $\mathrm{Cov}^*(\hat\beta^*) = \hat\sigma_e^2(X'X)^{-1}$, with $\hat\sigma_e^2 = n^{-1}\sum_k e_k^2$. The bias in the estimation of σ^2 can be simply avoided by resampling from the adjusted residuals $e_i/\sqrt{1-p/n}$.

The specification of errors in GLM's is generally different from linear models. Therefore, the empirical distribution function of some residuals cannot describe the distribution of the y_i's. Furthermore, there are possibly substantial differences in higher order moments of the y_i's so that the residuals are not generally exchangeable. Hence, a bootstrap technique which is based on resampling from somehow standardized residuals seems to be inappropriate for these models. It is shown in section 2 that the method of residual resampling yields bootstrap replications, which can be regarded at least as linear approximation to the MLE $\hat\beta$. The score resampling technique described in section 3 results in a more robust estimator for the covariance matrix of the score vector which is used to estimate $\mathrm{Cov}(\hat\beta)$. This method is based on sampling from artificially generated pseudo random numbers instead of resampling observed residuals. Moreover, it will be shown that both approaches cannot provide a reduction of the bias.

An application of the proposed resampling techniques to the analysis of binary responses is presented in section 4.

2 Resampling from Residuals

In this section we confine our attention onto approximation (3). This implies the application of bootstrap techniques appropriate for linear heteroscedastic models with additive errors of the special form $(y_i - \mu_i)/V(\mu_i)$ all with zero mean. However, the error variances generally differ so we cannot assume exchangeability. In terms of theoretical Pearson residuals, (4) can be rewritten as

$$z_i(\mu_i) = \eta_i + \sqrt{a_i/V(\mu_i)}\frac{y_i - \mu_i}{\sqrt{a_i V(\mu_i)}} = \eta_i + w(\mu_i)^{-1/2}\epsilon(\mu_i), \qquad i = 1, \ldots, n,$$

where $\epsilon(\mu_i) = (y_i - \mu_i)/\sqrt{a_i V(\mu_i)}$. These errors have constant variance ϕ and differ only with respect to higher order moments. Therefore, they are rather exchangeable which is required to apply nonparametric residual resampling techniques.

Let $e_1^*, \ldots, e_n^*$ denote a random sample from the centered empirical distribution function based on $e_1 - \overline{e}, \ldots, e_n - \overline{e}$ with $\overline{e} = n^{-1}\sum_k e_k$, where

$$e_i = \epsilon(\hat\mu_i) = \frac{y_i - \hat\mu_i}{\sqrt{a_i V(\hat\mu_i)}} \tag{7}$$

is the i-th observed Pearson residuals. Hence, $\mathrm{E}^*(e_i^*) = 0$ and

$$\mathrm{Var}^*(e_i^*) = \frac{1}{n}\sum_{k=1}^{n} e_k^2 - \overline{e}^2 = \hat\phi_e$$

with E^* and Var^* referring to the bootstrap distribution. Note that the variance of the residuals under the bootstrap model is smaller than the estimator (5).

Following the considerations above, the resampled pseudo observations are defined by

$$z_i^* = \hat\eta_i + w(\hat\mu_i)^{-1/2} e_i^*,$$

for which $\mathrm{E}^*(z_i^*) = \hat\eta_i$ and $\mathrm{Var}^*(z_i^*) = \hat\phi_e / w(\hat\mu_i)$.

Moulton and Zeger (1991) suggest to use another set of residuals. Regarding canonical link functions, the first-order approximation $\hat\eta \approx \eta + V(\mu)^{-1}(\hat\mu - \mu)$ yields $y - \hat\mu \approx (I - H(\mu))(y - \mu)$, where $H(\mu) = W(\mu)^{1/2}X(X'W(\mu)X)^{-1}X'W(\mu)^{1/2}$ is the generalized version of the hat matrix. Therefore, the covariance matrix of the ordinary residual vector $y - \hat\mu$ can be approximated by $\mathrm{Cov}(y)(I - H(\mu))$, so that $(1 - h(\hat\mu_i))\phi a_i V(\hat\mu_i)$ is a natural estimate of the variance of the i-th residual under the assumed variance model. Hence, they sampled from observed standardized Pearson residuals

$$r_i = \frac{e_i}{\sqrt{1 - h(\hat\mu_i)}} = \frac{y_i - \hat\mu_i}{\sqrt{1 - h(\hat\mu_i)}\sqrt{a_i V(\hat\mu_i)}}, \tag{8}$$

all of them reflecting equal variance ϕ. For a detailed discussion of a very general definition of residuals see Cox and Snell (1968).

Resampling from the centered set $\{r_i - \overline{r}\}$ leads to $\mathrm{E}^*(r_i^*) = 0$ and $\mathrm{Var}^*(r_i^*) = n^{-1}\sum_k r_k^2 - \overline{r}^2 = \hat\phi_r$. Referring to the resampled

$$z_i^* = \hat\eta_i + w(\hat\mu_i)^{-1/2} r_i^*$$

we have $\mathrm{E}^*(z_i^*) = \hat\eta_i$ and $\mathrm{Var}^*(z_i^*) = \hat\phi_r / w(\hat\mu_i)$.

No matter from what type of residuals we are resampling, the resulting bootstrap replication of $\hat{\beta}$ is given by

$$\hat{\beta}^* = (X'W(\hat{\mu})X)^{-1}X'W(\hat{\mu})z^*, \tag{9}$$

where the vector z^* is either based on the Pearson residuals (7) or on their standardized version defined in (8). Moreover, both possibilities yield $\mathrm{E}^*(\hat{\beta}^*) = \hat{\beta}$, for which reason the residual resampling approach is not helpful to describe the bias of $\hat{\beta}$. Of course, this is generally caused by the fact that only a first-order approximation has been considered. But with this technique a dispersion factor is incorporated into the variance term. Depending on the residuals we have resampled from, we get $\mathrm{Cov}^*(\hat{\beta}^*) = \hat{\phi}_.(X'W(\hat{\mu})X)^{-1}$, where $\hat{\phi}_.$ denotes either $\hat{\phi}_e$ or $\hat{\phi}_r$. Therefore, the only modification in using $\mathrm{Cov}^*(\hat{\beta}^*)$ to estimate $\mathrm{Cov}(\hat{\beta})$ is given by the different dispersion estimator. Especially, resampling from the unstandardized Pearson residuals (7) results in the dispersion estimate $\hat{\phi}_e = n^{-1}X^2 - \overline{e}^2$. Because $\overline{e}^2$ should be relatively small compared to the sum of squared residuals, this estimator is roughly of the same magnitude as $n^{-1}X^2$. In case of $\overline{e}^2 = 0$ we have $\hat{\phi}_e = (n-p)n^{-1}\hat{\phi}$ where $\hat{\phi} = X^2/(n-p)$ is again the 'moment-based estimator' for ϕ. Like in constant-variance, linear models, resampling from the observed Pearson residuals scaled by $\sqrt{1-p/n}$ avoids this bias for moderate values of p. If the differences in the $h(\hat{\mu}_i)$'s are neglectable, the scaling factors $1-h(\hat{\mu}_i)$ can be replaced by their mean value $1-p/n$. A comparative study concerning the behaviour of resampling techniques based on different scaling methods in the logistic regression case is given in Friedl and Tilg (1994).

In order to approximate the distribution of deviance differences, Hinde (1992) also resampled from the standardized residuals defined by (8), however, in order to construct replications of the original responses y_i, namely

$$y_i^* = \hat{\mu}_i + \sqrt{a_i V(\hat{\mu}_i)(1-h(\hat{\mu}_i))}\,r_i^*. \tag{10}$$

Therefore, the bootstrap variance $\mathrm{Var}^*(y_i^*) = \phi_r^* a_i V(\hat{\mu}_i)(1-h(\hat{\mu}_i))$ is implied by this resampling strategy, which seems to underestimate the variance fixed by the model. Applying (10) to discrete probability models like Poisson, binomial or negative binomial results in replications, which are possibly negative or noninteger. Because of those difficulties, Firth et al. (1991) used bootstrap replications

$$y_i^* = [\max(0, \hat{\mu}_i + \sqrt{a_i V(\hat{\mu}_i)}e_i^*)]$$

to guarantee nonnegative integer responses for Poisson models. We mention that (9) defines a one-step procedure without any iteration, noting that the bootstrap works with $\hat{\mu}$ instead of μ in (3). To describe some distributional properties of the MLE $\hat{\beta}$ or a function thereof by a nonparametric bootstrap procedure, we do not need to define replications y^* explicitelly.

3 An Application of Wu's Bootstrap

Local linearisation of the score function $u(\hat{\beta})$ immediately gives

$$\mathrm{Cov}(\hat{\beta}) \approx F^{-1}(\beta)\mathrm{Cov}(u(\beta))F^{-1}(\beta),$$

where F is the $p \times p$ Fisher information matrix. For canonical link models this matrix does not include random components and is simply $F(\beta) = -\partial u(\beta)/\partial\beta' = \phi^{-1}X'W(\beta)X$. To the regarded level of approximation, an improved estimation of the score covariance should result in a better estimate of the covariance of $\hat{\beta}$ (Lele, 1992).

Now we interpret each component $u_j(\beta)$, $j = 1, \ldots, p$, of the scorevector (1) as an arithmetic mean of n independent but not identically distributed variates $u_{ij}(\beta)$, $i = 1, \ldots, n$, each with zero mean and variance $x_{ij}^2\mathrm{Var}(y_i)/(\phi^2 a_i^2)$. A general discussion on bootstrapping under non i.i.d. models is given in Liu (1988). Wu (1986) has introduced a weighted bootstrap procedure yielding bootstrap scores

$$u_{ij}^* = \overline{u}_j(\hat{\beta}) + \left(u_{ij}(\hat{\beta}) - \overline{u}_j(\hat{\beta})\right) t_i^* = u_{ij}(\hat{\beta})t_i^*, \tag{11}$$

because $\overline{u}_j(\hat{\beta}) = n^{-1}\sum_i u_{ij}(\hat{\beta}) = 0$. The t_i^*'s are drawn independently from some distribution with zero mean and unit variance. Hence we have $\mathrm{E}^*(u_{ij}^*) = 0$ and $\mathrm{Var}^*(u_{ij}^*) = u_{ij}(\hat{\beta})^2 = x_{ij}^2\left(y_i - \hat{\mu}_i\right)^2/(\phi^2 a_i^2)$, where E^* and Var^* denote mean and variance with respect to the distribution of $t_1^*, \ldots, t_n^*$. Therefore, the variance of y_i is estimated by $(y_i - \hat{\mu}_i)^2$ which is obviously more robust against misspecifications of the variance model $V(\mu_i)$. This sampling procedure is also referred to as external or wild bootstrap.

Let $u_j^* = \sum_i u_{ij}(\hat{\beta}) t_i^*$ denote the j-th component of the bootstrap score vector $u^* = (u_1^*, \ldots, u_p^*)'$. Then the resulting bootstrap replication of $\hat{\beta}$ is simply given by replacing the score u by u^* in (2) giving $\hat{\beta}^* = \hat{\beta} + F(\hat{\beta})^{-1} u^*$, where F corresponds to the derivative of u. Equivalently, this can be also written as

$$\hat{\beta}^* = \hat{\beta} + (X'W(\hat{\mu})X)^{-1} X' S(\hat{\mu})^{1/2} t^*, \tag{12}$$

where $S(\mu)^{1/2} = \mathrm{diag}(|y_i - \mu_i| / a_i)$. In terms of bootstrap pseudo observations

$$z_i^* = \hat{\eta}_i + w(\hat{\mu}_i)^{-1/2} e_i t_i^*,$$

this can also be written in the usual weighted least squares representation. Relating to the variances of the original pseudo observations z_i defined in (4), $\mathrm{Var}(z_i) = \phi / w(\mu_i)$, it is interesting to note that

$$\mathrm{Var}^*(z_i^*) = \hat{\phi}_i / w(\hat{\mu}_i)$$

is induced by that sampling mechanism, where the i-th dispersion estimate is

$$\hat{\phi}_i = \frac{(y_i - \hat{\mu}_i)^2}{a_i V(\hat{\mu}_i)}, \qquad i = 1, \ldots, n.$$

Therefore, the dispersion estimate (5) can be interpreted as the bias corrected mean value of n individual dispersion estimates.

Finally we have $\mathrm{E}^*(\hat{\beta}^*) = \hat{\beta}$ whereas the covariance matrix of $\hat{\beta}^*$ is

$$\mathrm{Cov}^*(\hat{\beta}^*) = (X'W(\hat{\mu})X)^{-1} X' S(\hat{\mu}) X (X'W(\hat{\mu})X)^{-1}. \tag{13}$$

If $a_i V(\hat{\mu}_i) = (y_i - \hat{\mu}_i)^2$ for all i, then S equals W and (13) reduces to the Fisher information F evaluated in $\mu = \hat{\mu}$, which is the usual estimator for $\mathrm{Cov}(\hat{\beta})$ assuming no extra dispersion.

Using the same notation as in (12), the residual resampling estimate (9) is simply

$$\hat{\beta}^* = \hat{\beta} + (X'W(\hat{\mu})X)^{-1} X' W(\hat{\mu})^{1/2} r^*$$

or with e^* instead of r^*. This shows the analogy between the previous mentioned residual resampling technique and this kind of score resampling, where different types of residuals and variance estimates are suggested. Of course, (12) should not be used to approximate the whole distribution of $\hat{\beta}$, because up to this point only moment assumptions has been made on the sample $t_1^*, \ldots, t_n^*$. Strengthening the assumptions by supposing $t_i^* \overset{iid}{\sim} N(0,1)$ will result in $\hat{\beta}^* \sim N(\hat{\beta}, \mathrm{Cov}^*(\hat{\beta}^*))$, similar to the limiting distributional behaviour of the MLE $\hat{\beta}$.

A further interesting definition of a bootstrap vector $\hat{\beta}^*$ should be mentioned here. Consider the bootstrap score as a function in β, e.g.

$$u_j^*(\beta) = \frac{1}{\phi} \sum_{i=1}^{n} \frac{1}{a_i} x_{ij} (y_i - \mu_i(\beta)) t_i^*, \qquad j = 1, \ldots, p.$$

The behaviour of the bootstrap vector defined by $u^*(\hat{\beta}^*) = 0$ presumably would also approximate the respective behaviour of $\hat{\beta}$ (Lele, 1992). A first-order expansion immediately yields

$$\hat{\beta}^* = (X'W(\beta)^{1/2} T^* W(\beta)^{1/2} X)^{-1} X' W(\beta)^{1/2} T^* W(\beta)^{1/2} z(\beta), \tag{14}$$

with the sampling matrix $T^* = \mathrm{diag}(t_1^*, \ldots, t_n^*)$ and z defined in (4). The observation vector resampling technique in Moulton and Zeger (1991) gives a very similar definition. Because of resampling from observation vectors (y_i, x_i') the distribution of $t_1^*, \ldots, t_n^*$ corresponds to the multinomial$(n; 1/n, \ldots, 1/n)$. Evaluating the right side of (14) in $\hat{\beta}$ their 'closed form' representation of the covariance matrix of the one-step bootstrap vector $\hat{\beta}^*$ equals (13).

Table 1: Recurrence rates from cervical carcinoma

	Tumour Extension								
	0			1			2		
	Mitotic Activity								
L	0	1	2	0	1	2	0	1	2
0	9/55	7/48	5/21	6/21	2/20	10/17	0/2	2/5	2/7
1	3/8	2/8	2/5	1/4	3/6	2/2	6/6	4/5	6/8
2	0/4	6/7	3/5	0/1	4/4	1/2	0/2	5/6	4/4
3	1/1	4/4	8/8	1/1	3/3	1/1	2/3	3/3	5/6

4 Example

As an illustration of the behaviour of the proposed variance estimates, a binomial logistic model is considered for data on the recurrence of cervical carcinoma within five years after the first operative treatment. Each of the $n = 313$ patients underwent radical abdominal hysterectomy. This study aimed to identify factors known at the time of primary treatment that contribute to the risk of recurrent cancer. The ratios y_i/n_i, where the y_i-s denote the numbers of patients with recurrence out of totals n_i, are outlined in Table 1. The factors used to model tumour prognosis are the tumour extension E (0-cervix only, 1-border zone involved, 2-parametria involved), the mitotic activity M (0-low, 1-medium, 2-high) and the number of lymph node stations involved L (0-none, 1-one, 2-two, 3-more than two involved).

The estimates $\hat{\beta}$ under the binomial distribution model $y_i/n_i \sim Bin(n_i, \pi_i)/n_i$ are given in Table 2. This model implies $\mathrm{E}(y_i/n_i) = \pi_i$ and $\mathrm{Var}(y_i/n_i) = \phi_i V(\pi_i)$, where $\phi_i = \phi a_i$ with $a_i = 1/n_i$, fixed $\phi = 1$ and $V(\pi_i) = \pi_i(1 - \pi_i)$.

Because the observed value of the mean Pearson statistic is 44.86 on only 32 degrees of freedom, there is some evidence of overdispersion relative to the binomial variance. Assuming the constant dispersion model $\mathrm{Var}(y_i) = \phi a_i V(\pi_i)$ with unknown ϕ results again in the MLE $\hat{\beta}$. As dispersion estimate we get $\hat{\phi} = 1.402$. We further investigated the variance model with $\phi_i = (1 + (n_i - 1)\tau^2)/n_i$, where τ^2 can be interpreted as the correlation coefficient between patients with a specific (L,E,M)-diagnosis group independent of its size. However, the estimation procedure proposed in Williams (1982) resulted in the relatively small value $\hat{\tau}^2 = 0.08$. This seems to be no evidence that the observations are correlated.

Table 2: Maximum likelihood estimation and theoretical bootstrap results

		Estimated Standard Errors of $\hat{\beta}$					
	$\hat{\beta}$	$\phi = 1$	$\hat{\phi} = 1.402$	$\hat{\phi}_e = 1.241$	$\hat{\phi}_r = 1.416$	$\hat{\phi}_{r'} = 1.396$	Wu
1	-1.918	0.258	0.305	0.288	0.307	0.305	0.292
L	1.036	0.156	0.185	0.174	0.187	0.184	0.144
E	0.530	0.180	0.214	0.201	0.214	0.213	0.176
M	0.440	0.178	0.210	0.198	0.212	0.210	0.162

The residual resampling techniques of section 2 yield $\hat{\phi}_e = 1.241$ or $\hat{\phi}_r = 1.416$ according to the type of residuals (e_i or r_i) we are resampling from. Replacing the h_i-s by their mean value p/n in the definition of r_i results in $\hat{\phi}_{r'} = 1.396$ which is very similar to $\hat{\phi}_r$. In the Monte-Carlo study of Friedl and Tilg (1994) it is shown that the behaviour of $\hat{\phi}_r$ and $\hat{\phi}_{r'}$ is quite similar to that of $\hat{\phi}$ and both outperform $\hat{\phi}_e$ in case of binomial and beta-binomial distributed variates. Note that a dispersion estimate is automatically incorporated into the variance by the bootstrap although we assumed binomial variability. Table 2 shows the estimated standard errors of $\hat{\beta}$ which are proportional to those under the binomial model. The factor

Table 3: Monte-Carlo results based on 1000 bootstrap replications

	Estimated Standard Errors of $\hat{\beta}$			
	e	r	r'	Wu
1	0.285	0.259	0.305	0.294
L	0.174	0.173	0.187	0.149
E	0.202	0.196	0.221	0.179
M	0.199	0.187	0.214	0.164

of proportionality is the respective dispersion estimator.

Whereas all the residual resampling approaches yield estimates of a global overdispersion factor, Wu's weighted bootstrap rather describes individual overdispersion terms through the ratio of the one-sample variance estimate $(y_i - n_i\hat{\pi}_i)^2$ to the estimated variance under the binomial model given by $n_i\hat{\pi}_i(1-\hat{\pi}_i)$. These ratios range from 0.00012 up to 6.742 with arithmetic mean 1.246.

In order to compare the bootstrap variance estimates of Table 2 with Monte-Carlo results we generate $B = 1000$ replications $\hat{\beta}^*$ of the maximum likelihood estimate $\hat{\beta}$ using all of the different resampling techniques. Consistent with the findings from the previous sections the results in Table 3 show no substantial differences between the theoretical and the Monte-Carlo bootstrap results. Wu's weighted bootstrap is based on generating t_i variates from the $N(0,1)$ distribution.

5 Concluding Remarks

Restriction on canonical link models is only caused by the information matrix F which would be random otherwise. If the definition of $\hat{\beta}$ is based on the expected information, the generalization onto the whole class of GLM's including noncanonical link models is straightforward. However, defining $\hat{\beta}$ in terms of the observed information, the corresponding bootstrap information should also reflect the behaviour of some y^*. This is automatically realized by the observation vector resampling technique of Moulton and Zeger (1991). But their approach is rather appropriate for correlation models. We have avoided to define y^* explicitely because of difficulties raised by discrete distributions.

None of the bootstrap methods is able to give a bias reduction. Naturally, this also affects the quality of the variance estimation. Combining resampling strategies with the likelihood modification of Firth (1992) possibly improves the variance estimation. However, some of the bootstrap estimates are already robust against slight misspecifications of the assumed variance model like overdispersion.

References

[1] Cox, D.R., Snell, E.J., A general definition of residuals (with discussion). J. R. Statist. Soc. B 30 (1968), 248-275.

[2] Copas, J.B., Binary regression models for contaminated data. J. R. Statist. Soc. B 50 (1988), 225-265.

[3] Firth, D., Goslup, J., Hinkley, D.V., Model checking with nonparametric curves. Biometrika 78 (1991), 245-252.

[4] Firth, D., Bias reduction, the Jeffreys prior and GLIM. In: Lecture Notes in Statistics 78 (ed. by L. Fahrmeir et.al), 91-100. Springer Verlag, New York-Berlin-Heidelberg 1992.

[5] Firth, D., Bias reduction of maximum likelihood estimates. Biometrika 80 (1993), 27-38.

[6] Friedl, H., Tilg, N., Variance estimation in logistic regression using the bootstrap. Submitted to: Communications in Statistics (1994).

[7] Hinde, J.P., Choosing between non-nested models; a simulation approach. In: Lecture Notes in Statistics 78 (ed. by L. Fahrmeir et.al), 119-124. Springer Verlag, New York-Berlin-Heidelberg 1992.

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

[9] Liu, R.Y., Bootstrap procedures under some non-i.i.d. models. Ann. Statist. 16 (1988), 1696-1708.

[10] McCullagh, P., Nelder, J.A., Generalized linear models, 2nd edition. Chapman and Hall, London-New York 1989.

[11] Moulton, L.H., Zeger, S.L., Bootstrapping generalized linear models. Comp. Statist. & Data Analysis 11 (1991), 53-63.

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

[13] Williams, D.A., Extra-binomial variation in logistic linear models. Appl. Statist. 31 (1982), 144-148.

[14] Wu, C.F.J., Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). Ann. Statist. 14 (1986), 1261-1350.