On the asymptotic Moments of Pearson Type Statistics based on Resampling Procedures

Herwig Friedl

Institute of Statistics, Technical University Graz, A-8010 Graz, Austria

Summary

This paper presents a discussion of overdispersed generalized linear regression models. Commonly used estimates of the unknown dispersion parameter are the Pearson statistic or the modification of Farrington (1996), who proposed a first order linear correction term to Pearson's statistic. We like to focus our attention onto a bootstrap statistic based on resampling from some observed centered Pearson residuals. Especially, if these residuals are also scaled by a constant term the respective bootstrap dispersion corresponds to the empirical variance of the Pearson residuals. Hence, this statistic includes an additional quadratic term. Assuming exponential family properties, this bootstrap dispersion estimate is shown to be asymptotically uncorrelated to the regression parameter estimates. Moreover, approximations to the first two unconditional moments are obtained, which can be easily compared to those of the Pearson statistic.

Keywords: Generalized Linear Models, Bootstrap, Pearson Residuals, Cumulants, Stochastic Expansions.

1 Introduction

Following McCullagh and Nelder (1989) we consider independent responses $y_1, \ldots, y_n$ with $E(y_i) = \mu_i$ and

$$Var(y_i) = \phi V(v_i, \mu_i) = \phi V_i, \tag{1}$$

where $\phi > 0$ denotes a possibly unknown dispersion parameter and the variance function $V > 0$ only depends on the mean and on some known weights v_i. Throughout this paper it is also demanded that the third and fourth moment exist and behave like those of an exponential family distribution. In terms of the respective cumulants, this means that we are assuming

$$\begin{aligned} \kappa_{3i} &= E(y_i - \mu_i)^3 = \phi^2 V_i V_i', & (2) \\ \kappa_{4i} &= E(y_i - \mu_i)^4 - 3Var^2(y_i) = \phi^3 \left(V_i^2 V_i'' + V_i V_i'^2 \right), & (3) \end{aligned}$$

the primes standing for differentiation with respect to μ. Moreover, we will also use the means of standardized cumulants, i.e.

$$\begin{aligned} \overline{\rho}_3^2 &= \frac{1}{n} \sum_{i=1}^{n} \frac{\kappa_{3i}^2}{Var^3(y_i)} = \frac{\phi}{n} \sum_{i=1}^{n} \frac{V_i'^2}{V_i}, \\ \overline{\rho}_4 &= \frac{1}{n} \sum_{i=1}^{n} \frac{\kappa_{4i}}{Var^2(y_i)} = \frac{\phi}{n} \sum_{i=1}^{n} \left(V_i'' + \frac{V_i'^2}{V_i} \right). \end{aligned}$$

Example 1: When modelling proportions of binomial responses y_i with denominators m_i we have $E(y_i/m_i) = \mu_i$ and $Var(y_i/m_i) = \mu_i(1-\mu_i)/v_i$ with $v_i = m_i$. So $\phi = 1$, $V_i' = (1-2\mu_i)/v_i$ and $V_i'' = -2/v_i$ resulting in

$$\overline{\rho}_3^2 = \frac{1}{n} \sum_i \frac{(1-2\mu_i)^2}{m_i \mu_i (1-\mu_i)}, \quad \overline{\rho}_4 = \frac{1}{n} \sum_i \frac{1 - 6\mu_i(1-\mu_i)}{m_i \mu_i (1-\mu_i)}.$$

Example 2: Poisson variates with $Var(y_i) = \mu_i$ ($\phi = 1$) give

$$\overline{\rho}_3^2 = \overline{\rho}_4 = \frac{1}{n} \sum_i \frac{1}{\mu_i}.$$

In a generalized linear regression model the means μ_i are related to explanatory variables $x_i = (x_{i1}, \ldots, x_{ip})^t$ by a known inverse link function h, that is

$$\mu = h(\eta),$$

where $\eta = X\beta$ is the linear predictor, $X = (x_{ij})$ is the $n \times p$ model matrix and $\beta = (\beta_1, \ldots, \beta_p)^t$ is a vector of p unknown regression parameters. We further assume that h is a third order continuously differentiable function, and write h' when differentiating h with respect to the linear predictor η.

In the context of quasi-likelihood estimation, Wedderburn (1974) has defined parameter estimates $\hat{\beta}$ as the solution of $u_\beta(\beta) = 0$, the system of p quasi-likelihood equations with

$$u_\beta(\beta) = \sum_{i=1}^{n} \frac{y_i - \mu_i}{V_i} \frac{\partial \mu_i}{\partial \beta}. \tag{4}$$

These estimating functions are unbiased in the sense $E(u_\beta(\beta)) = 0$.

For canonical link models $h'_i = V_i$ is applicable. A numerical solution of (4) is then obtained by applying the iterative weighted least squares procedure

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

It is further assumed that Fisher's scoring technique is used for non-canonical link models. This allows us in both situations to write the adjusted dependent variates as

$$z_i = \eta_i + (y_i - \mu_i)/h'_i \tag{5}$$

and the diagonal matrix $W = diag(w_i)$ contains the iterative weights

$$w_i = h_i'^2 / V_i.$$

It is well known (and it will be also shown later) that at least asymptotically $\hat{\beta}$ is unbiased with variance-covariance matrix $\phi(X^t W X)^{-1}$. Estimated standard errors may therefore be calculated as square roots of diagonal elements from the matrix

$$\widehat{Var}(\hat{\beta}) = \phi(X^t \hat{W} X)^{-1}. \tag{6}$$

Estimates for the unknown dispersion parameter ϕ are sometimes based on the deviance or on the generalized Pearson statistic. A critical discussion of the asymptotic behaviour of both statistics is given in Firth (1991). Approximations to the unconditional moments of the Pearson statistic and the extended version of Farrington (1996) are quoted in the following section.

However, an alternative strategy to construct dispersion estimates is to use the bootstrap technique. Like in linear regression models, where we assume responses with constant variance, the bootstrap procedure is mainly based on resampling randomly, with replacement, from the observed Pearson residuals after the likelihood fit (Simonoff and Tsai (1988), Moulton and Zeger (1990), Friedl (1995)). An outline of that technique within generalized linear models is given in the third section of this paper. No matter which estimate we are discussing, we have to consider the system of estimating equations (4) for β together with a further equation for ϕ. It is shown in section 4, that the estimating function for ϕ is unbiased, if we resample from Pearson residuals additionally scaled by $(1-1/n)^{1/2}$. Finally, the asymptotic properties of the corresponding bootstrap dispersion estimate are obtained in section 5.

2 Dispersion Estimates

The constant dispersion parameter ϕ in (1) can be interpreted in a variety of different ways. Used in an ordinary linear regression model for responses with $Var(y_i) = \sigma^2$, ϕ just describes the constant variance σ^2. For measurements with multiplicative errors, ϕ reflects the constant coefficient of variation together with the quadratic variance structure $V_i = \mu_i^2$.

Extra-variation is often concerned with omitted variables in the model inducing unobserved herterogeniety. One way to model this situation is to assume a GLM for the conditional distribution $y_i|z_i$ of the responses y_i given an additional random effect z_i in the linear predictor $\eta_i + z_i$. The negative-binomial and the beta-binomial distribution are both well known examples for this kind of models. But with the variance structure ϕV_i we are only able to cover simple overdispersion relative to our assumed variance model. In the following example it is shown how the formulation of the negative-binomial distribution has to be changed to lead to a constant overdispersion model.

Example 3: Suppose that $y_i|\nu_i \sim Pois(\nu_i)$ and $\nu_i \sim \Gamma(\theta_i, \lambda_i)$. Then the marginal distribution of y_i is negative-binomial with moments $E(y_i) = \theta_i/\lambda_i = \mu_i$ and $Var(y_i) = \theta_i/\lambda_i + \theta_i/\lambda_i^2$. For fixed $\theta_i = \theta$ and varying λ_i we have a linear exponential family model with $Var(y_i) = \mu_i + \mu_i^2/\theta$. If we assume that $\lambda_i = \lambda$ is fixed as θ_i varies in the Γ-mixing distribution, then

$$Var(y_i) = \mu_i + \mu_i/\lambda = \Big(1 + \frac{1}{\lambda}\Big)\mu_i = \phi\mu_i.$$

This is still a negative-binomial regression model but is no longer in the linear exponential family.

If ϕ is unknown, the mean Pearson statistic

$$\hat{\phi} = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{V}_i} = \frac{1}{n}X^2(y, \hat{\mu}) \tag{7}$$

can be a favourable estimate of it. For small sample sizes it is better to replace $1/n$ by $1/(n-p)$, a direct consequence of result (10) below. In exponential family regression models with canonical links approximate conditional and unconditional moments of Pearson's statistic $X^2(y, \hat{\mu})$ may be found in McCullagh (1985).

By adding a first order component to the mean Pearson statistic (7), Farrington (1996) introduced a more general family of dispersion estimates, namely

$$\hat{\phi}_a = \frac{1}{n}\sum_{i=1}^{n}\frac{(y_i - \hat{\mu}_i)^2}{\hat{V}_i} + \frac{1}{n}\sum_{i=1}^{n}\hat{a}_i(y_i - \hat{\mu}_i), \tag{8}$$

where the a_i are functions of μ_i, but do not depend on ϕ.

For an exponential family model with any admissible link function

$$Cov(\hat{\phi}_a, \hat{\beta}) = O(n^{-2}) \tag{9}$$

applies to the class of statistics (8). Moreover, the first two unconditional moments are

$$
\begin{aligned}
E(\hat{\phi}_a) &= \phi\Big(1-\frac{p}{n}\Big) + \frac{\phi}{2n}\sum_{i,j}\left(a_i + \phi\frac{V_i'}{V_i}\right)\frac{h_j''}{V_j}h_i'h_j'Q_{ij}Q_{jj} \\
&\quad -\frac{\phi}{2n}\sum_{i=1}^{n}\left\{\phi\frac{V_i''}{V_i}h_i'^2 + \left(a_i+\phi\frac{V_i'}{V_i}\right)h_i''\right\}Q_{ii} + O(n^{-2}) \quad (10)\\
Var(\hat{\phi}_a) &= \frac{\phi^2}{n}\left(\overline{\rho}_4 - \overline{\rho}_3^2 + 2\right) + \frac{\phi}{n^2}\sum_{i=1}^{n}V_i\left(a_i+\phi\frac{V_i'}{V_i}\right)^2 \\
&\quad -\frac{\phi}{n^2}\sum_{i,j}\left(a_i+\phi\frac{V_i'}{V_i}\right)\left(a_j+\phi\frac{V_j'}{V_j}\right)h_i'h_j'Q_{ij} + O(n^{-2}), \quad (11)
\end{aligned}
$$

where the $n \times n$ matrix $Q = X(X^tWX)^{-1}X^t$ denotes the variance matrix of $\hat{\eta} = X\hat{\beta}$. These general results hold also for non-canonical links. For $\phi = 1$ known, the special choice $a_i = -V_i'/V_i$, proposed in Farrington (1996), eliminates the dependence of $E(\hat{\phi}_a)$ from the first order bias of $\hat{\beta}$, minimizes the variance (11) and results in much simpler expressions for the moments (10) and (11). If ϕ is treated to be unknown then none of these properties generally continue to hold.

Expressions for the mean and the variance of the mean Pearson statistic $\hat{\phi}$ based on GLM's with general link functions h can be easily derived from above by setting $a_i = 0$. If we additionally substitute the definition of $\overline{\rho}_3^2$ for the exponential family case to simplify the variance term then

$$
\begin{aligned}
E(\hat{\phi}) &= \phi\Big(1-\frac{p}{n}\Big) + \frac{\phi^2}{2n}\sum_{i,j}\frac{V_i'h_j''}{V_iV_j}h_i'h_j'Q_{ij}Q_{jj} \quad (12)\\
&\quad -\frac{\phi^2}{2n}\sum_{i=1}^{n}\frac{V_i''h_i'^2 + V_i'h_i''}{V_i}Q_{ii} + O(n^{-2}) \\
Var(\hat{\phi}) &= \frac{\phi^2}{n}\left(\overline{\rho}_4 + 2\right) - \frac{\phi^3}{n^2}\sum_{i,j}\frac{V_i'V_j'}{V_iV_j}h_i'h_j'Q_{ij} + O(n^{-2}). \quad (13)
\end{aligned}
$$

The respective moments of the Pearson statistic (7) obtained by McCullagh (1985) are available by setting again $a_i = 0$ but using the canonical link property $h_i' = V_i$, resulting also in $h_i'' = V_iV_i'$.

Example 4: Modelling binary data, the estimate $\hat{\phi}_a$ based on the optimal value $a_i = -V_i'/V_i = -(1-2\mu_i)/\mu_i(1-\mu_i)$ is completely noninformative in the sense that each observation contributes exactly 1 to the value of $n\phi_a$. It is shown in Osius and Rojek (1992) that the Pearson statistic can then be interpreted as a score test statistic with respect to an enlarged model. The same also happens in the k-sample problem with $X^2(y,\mu)$ defined in (7). The respective model can be written as $y_i \sim b(1,\mu_i)$ with $\log(\mu_i/(1-\mu_i)) =$

$\beta_1 + \beta_j[x_i = j]$, $j = 2, \ldots, k$, where x denotes a k-level factor and $[e]$ is the indicator function for event e.

3 The Residual Resampling Approach

To study the behaviour (especially the variability) of functions of the estimated parameters $\hat{\beta}$, Moulton and Zeger (1991) introduced a one-step bootstrap technique. Like in ordinary linear regression models their bootstrap procedure is based on the exchangability of certain residuals.

Let us write the additive model for the adjusted dependent variates (5) as

$$z_i = \eta_i + \epsilon_i \sqrt{V_i}/h_i' = \eta_i + \epsilon_i w_i^{-1/2},$$

where

$$\epsilon_i = \frac{y_i - \mu_i}{\sqrt{V_i}}$$

denotes the theoretical Pearson residual. If the considered model holds then we have simply $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \phi$. For that reason the quantities ϵ_i are exchangeable at least up to their second moments.

Simonoff and Tsai (1988) resampled from the empirical distribution function F^* of the centered observed Pearson residuals

$$\hat{\epsilon}_i - \frac{1}{n}\sum_{k=1}^{n} \hat{\epsilon}_k. \tag{14}$$

Note that for linear intercept models with constant variance the second term in this expression vanishes. The bootstrap distribution F^*, which corresponds to that resampling process, is constructed by putting mass $1/n$ at each observed residual (14). So all these resampled residuals have zero mean and equal variance with respect to F^*. However, this is not true with respect to the distribution of the y_i's.

To remove the dependency of the variance from the model matrix X, Moulton and Zeger (1991) investigated the behaviour of the more nearly exchangeable observed standardized Pearson residuals

$$\frac{\hat{\epsilon}_i}{\sqrt{1 - \hat{H}_{ii}}},$$

where $\hat{H}_{ii}$ denotes the i-th diagonal element of the generalized hat matrix $\hat{H} = \hat{W}^{1/2} X (X^t \hat{W} X)^{-1} X^t \hat{W}^{1/2} = \hat{W}^{1/2} \hat{Q} \hat{W}^{1/2}$ and $\hat{W}$ is evaluated in $\hat{\beta}$.

To avoid the computation of the scaling factors $1 - \hat{H}_{ii}$ they may be replaced by their mean value $1 - p/n$. The simulation study in Friedl and Tilg (1994) showed that the behaviour of the residuals scaled by $(1 - \hat{H}_{ii})^{1/2}$ or by $(1 - p/n)^{1/2}$ is quite similar at least in logistic regression models of overdispersed binomial variates.

The one-step bootstrap procedure results in the coefficient

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

where z^* is the vector of $z_i^* = \hat{\eta}_i + \hat{w}_i^{-1/2} e_i^*$ and e_i^* is sampled under F^* from the observed set $\{e_1, \ldots, e_n\}$. To relate the following discussion onto all types of residuals mentioned before we define quantities in the general form

$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{b_i \hat{V}_i}}. \tag{15}$$

Therefore, choosing $b_i = 1$ results in resampling from the observed Pearson residuals without extra scaling and $b_i = 1 - \hat{H}_{ii}$ or $b_i = 1 - p/n$ yields their standardized versions.

No matter from what type of residuals we like to resample from, the bootstrap variance-covariance matrix is

$$Var^*(\hat{\beta}^*) = \tilde{\phi}(X^t \hat{W} X)^{-1}.$$

This is similar to (6) except the unknown dispersion parameter ϕ is estimated by

$$\tilde{\phi} = \frac{1}{n} \sum_{i=1}^{n} e_i^2 - \left(\frac{1}{n} \sum_{i=1}^{n} e_i \right)^2 = \frac{1}{n} \sum_{i=1}^{n} (e_i - \overline{e})^2 . \tag{16}$$

In the unstandardized case, this bootstrap dispersion can be considered as a shrinkage modification of the mean Pearson statistic (7).

4 Unbiased Estimating Equations

As mentioned before the estimating functions defining the components of $\hat{\beta}$ are unbiased. To study the behaviour of the bootstrap dispersion an additional estimating equation is required. The residuals e_i in (16) differ only in their standardisation by $b_i > 0$. So we write the dispersion estimate in the general form

$$\frac{1}{n} \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{b_i \hat{V}_i} - \left(\frac{1}{n} \sum_{i=1}^{n} \frac{y_i - \hat{\mu}_i}{\sqrt{b_i \hat{V}_i}} \right)^2 .$$

The corresponding estimating function given in (17) below is unbiased, if

$$\sum_{i=1}^{n} \frac{1}{b_i} = \frac{n^2}{n-1}.$$

In the simple case $b_i = b$ for all i, this condition for unbiasedness is fullfilled if $b = 1 - 1/n$. This special choice yields

$$u_\phi(\beta, \phi) = \sum_{i=1}^{n} \left(\frac{(y_i - \mu_i)^2}{V_i} - \left(1 - \frac{1}{n}\right)\phi \right) - \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{y_i - \mu_i}{\sqrt{V_i}} \right)^2 , \tag{17}$$

and the corresponding bootstrap dispersion is given by

$$\tilde{\phi} = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{V}_i}} - \frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{\mu}_i}{\sqrt{\hat{V}_i}}\right)^2. \tag{18}$$

Hence, resampling from the residuals (15) with $b_i = 1-1/n$ yields a bootstrap dispersion $\tilde{\phi}$, which is the empirical variance of the ordinary Pearson residuals and may be used as an estimate for the unknown dispersion ϕ. Therefore, the statistic (18) together with the estimating function (17) will be discussed from now on in more detail.

Example 5: As stated in the previous example 4, neither the mean Pearson statistic $\hat{\phi}$ nor Farington's proposed estimate $\hat{\phi}_a$ with $a_i = -V_i'/V_i$ can be used for binary data in a k-sample model. Because the mean of the Pearson residuals is zero in this special situation, the estimate $\tilde{\phi}$ is also equal to the constant value $X^2(y,\hat{\mu})/(n-1) = n/(n-1)$.

For an arbitrary value ϕ we have $E(u_\phi(\beta,\phi)) = 0$ and

$$Var(u_\phi(\beta,\phi)) = \phi^2\Big(1-\frac{1}{n}\Big)^2\sum_{i=1}^{n}\left(\frac{E(y_i-\mu_i)^4}{\phi^2 V_i^2} - \frac{n-3}{n-1}\right).$$

This expression simplifies by using standardized cumulants instead of moments. Then the above variance expression becomes

$$Var(u_\phi(\beta,\phi)) = \phi^2(n-1)\left(\Big(1-\frac{1}{n}\Big)\overline{\rho}_4 + 2\right). \tag{19}$$

If κ_{4i} is bounded and $Var(y_i)$ is bounded away from zero, then $\overline{\rho}_4 = O(1)$ and $Var(\tilde{\phi}) = O(n)$ implying that $u_\phi(\beta,\phi) = O_p(n^{1/2})$. Both conditions are usually fullfilled if the elements of X realises on a bounded domain. A counterexample can be constructed by considering poisson-distributed responses as in example 2 with $\mu_i = 1/i$ resulting in $\overline{\rho}_4 = (n+1)/2$.

We may obtain more accurate estimates for small or moderate n by replacing $1-1/n$ with $1-p/n$, since its expected value will be closer to zero when evaluated at $\hat{\beta}$. This is equivalent to scaling by the mean value of the diagonal elements of the generalized hat matrix.

5 Asymptotic Expansion

To study the asymptotic behaviour of the dispersion estimate (18), we will use similar arguments as Farrington (1996). Let $\theta = (\beta_1,\ldots,\beta_p,\phi)^t$ denote the vector of all true parameters in our model. Similarily, let $u = (u_1,\ldots,u_p,u_{p+1})^t$ be the system of all $p+1$ estimating functions $(u_\beta^t, u_\phi)^t$ and define the respective estimates by $u(\hat{\theta}) = 0$. Then a quadratic Taylor expansion of $n^{-1/2}u(\hat{\theta})$ around θ yields

$$\hat{\theta} - \theta = n^{-1/2}Z_1 + n^{-1}Z_2 + O(n^{-3/2}),$$

where

$$\begin{aligned} Z_1 &= -n^{-1/2}\Delta^{-1}u, \\ Z_2 &= -n^{-1/2}\Delta^{-1}EZ_1 - \Delta^{-1}Z_1^t\nu Z_1/2 \end{aligned}$$

are both $O_p(1)$. Here the $(p+1)\times(p+1)$ matrices $\Delta=(\delta_{rs})$, $E=(e_{rs})$ and $\nu_r=(\nu_{r,st})$ have entries

$$\begin{aligned} \delta_{rs} &= E(\partial u_r/\partial\theta_s)/n, \\ e_{rs} &= \partial u_r/\partial\theta_s - n\delta_{rs}, \\ \nu_{r,st} &= E(\partial^2 u_r/\partial\theta_s\partial\theta_t)/n \end{aligned}$$

with $1\le r,s,t\le p+1$. So the first two moments of $\hat\theta$ are given by

$$\begin{aligned} E(\hat\theta-\theta) &= n^{-1}E(Z_2)+O(n^{-2}), && (20) \\ Var(\hat\theta) &= n^{-1}E(Z_1Z_1^t)+O(n^{-2}). && (21) \end{aligned}$$

Writing

$$\Delta = \begin{pmatrix} \Delta_{\beta\beta} & 0 \\ \Delta_{\phi\beta} & \Delta_{\phi\phi} \end{pmatrix}$$

with elements

$$\begin{aligned} \Delta_{\beta\beta,rs} &= -\frac{1}{n}\sum_{i=1}^n \frac{1}{V_i}\frac{\partial\mu_i}{\partial\beta_r}\frac{\partial\mu_i}{\partial\beta_s}, \\ \Delta_{\phi\beta,r} &= -\frac{\phi}{n}\Big(1-\frac{1}{n}\Big)\sum_{i=1}^n\frac{V_i'}{V_i}\frac{\partial\mu_i}{\partial\beta_r}, \qquad \Delta_{\phi\phi} = -\Big(1-\frac{1}{n}\Big), \end{aligned}$$

it follows that the inverse matrix is

$$\Delta^{-1} = \begin{pmatrix} \Delta_{\beta\beta}^{-1} & 0 \\ (1-\frac{1}{n})^{-1}\Delta_{\phi\beta}\Delta_{\beta\beta}^{-1} & -(1-\frac{1}{n})^{-1} \end{pmatrix}.$$

Moreover, the variance matrix of the β-score is

$$E(u_\beta u_\beta^t) = -\phi n\Delta_{\beta\beta}$$

and for the covariance of u_ϕ and u_β we get

$$E(u_\phi u_\beta^t) = \Big(1-\frac{1}{n}\Big)\sum_{i=1}^n \frac{\kappa_{3i}}{V_i^2}\frac{\partial\mu_i}{\partial\beta^t}.$$

Evaluation of the lower left part of the matrix $E(Z_1Z_1^t)$ gives

$$Cov(\tilde\phi,\hat\beta) = \frac{1}{n^2}\Big(1-\frac{1}{n}\Big)^{-1}\Big(-\phi n\Delta_{\phi\beta}\Delta_{\beta\beta}^{-1} - E(u_\phi u_\beta^t)\Delta_{\beta\beta}^{-1}\Big)+O(n^{-2}).$$

The above expression becomes exactly zero, if the exponential family property (2) holds. Then

$$E(u_\phi u_\beta^t) = -\phi n \Delta_{\phi\beta}$$

and the covariance is

$$Cov(\tilde{\phi}, \hat{\beta}) = O(n^{-2}). \qquad (22)$$

Hence, for responses y_i which are members of the exponential family, the parameter estimate $\hat{\beta}$ and the dispersion estimate $\tilde{\phi}$ are asymptotically uncorrelated. As stated before in (9) the same result also applies to the general Farrington family $\hat{\phi}_a$, which includes the Pearson statistic $\hat{\phi}$ for the choice $a_i = 0$.

In the next steps we evaluate the remaining parts of $E(Z_1 Z_1^t)$. The upper left terms describe

$$Var(\hat{\beta}) = -\frac{\phi}{n}\Delta_{\beta\beta}^{-1} + O(n^{-2}) = \phi(X^tWX)^{-1} + O(n^{-2}),$$

which was already used in the variance estimate (6). Furthermore,

$$Var(\tilde{\phi}) = \frac{1}{n^2}\Big(1 - \frac{1}{n}\Big)^{-2}\Big(E(u_\phi^2) + \phi n \Delta_{\phi\beta}\Delta_{\beta\beta}^{-1}\Delta_{\phi\beta}^t\Big) + O(n^{-2})$$

can be derived from the lower right corner of $E(Z_1 Z_1^t)$ in (21). In terms of $Q = X(X^tWX)^{-1}X^t$ and with result (19), we finally get

$$Var(\tilde{\phi}) = \frac{\phi^2}{n}\left(\overline{\rho}_4 + \frac{n}{n-1}2\right) - \frac{\phi^3}{n^2}\sum_{i,j}\frac{V_i'}{V_i}\frac{V_j'}{V_j}h_i'h_j'Q_{ij} + O(n^{-2}). \qquad (23)$$

Comparing this result with the variance (13) of Faringtons's statistic with $a_i = 0$ the only difference is the factor $n/(n-1) > 1$ above.

A lot of calculus needs to be done to calculate the mean (20), because Z_2 also depends on the matrices E and ν_r. But these matrices are of rather simple structure, namely

$$E = \begin{pmatrix} E_{\beta\beta} & 0 \\ E_{\phi\beta} & 0 \end{pmatrix}$$

with elements

$$\begin{aligned} E_{\beta\beta,rs} &= \sum_{i=1}^{n}\frac{y_i - \mu_i}{V_i}\left(\frac{h_i''}{h_i'^2} - \frac{V_i'}{V_i}\right)\frac{\partial\mu_i}{\partial\beta_r}\frac{\partial\mu_i}{\partial\beta_s}, \\ E_{\phi\beta,r} &= \sum_{i=1}^{n}\frac{-(y_i-\mu_i)(2V_i + (y_i-\mu_i)V_i')}{V_i}\frac{\partial\mu_i}{\partial\beta_r} + \phi\Big(1 - \frac{1}{n}\Big)\sum_{i=1}^{n}\frac{V_i'}{V_i}\frac{\partial\mu_i}{\partial\beta_r} \\ &\quad + \frac{1}{n}\left(\sum_{i=1}^{n}\frac{y_i-\mu_i}{\sqrt{V_i}}\right)\left(\sum_{i=1}^{n}\frac{2V_i + (y_i-\mu_i)V_i'}{V_i^{3/2}}\frac{\partial\mu_i}{\partial\beta_r}\right) \end{aligned}$$

and the matrix of expected second derivatives ν is either

$$\begin{pmatrix} \nu_{\beta_r} & 0 \\ 0 & 0 \end{pmatrix} \qquad \text{or} \qquad \begin{pmatrix} \nu_{\phi} & 0 \\ 0 & 0 \end{pmatrix}$$

with entries

$$\begin{aligned}
\nu_{\beta_r,st} &= \frac{1}{n}\sum_{i=1}^{n}\left\{2\frac{V_i'}{V_i^2} - \frac{3}{V_i}\frac{h_i''}{h_i'^2}\right\}\frac{\partial\mu_i}{\partial\beta_r}\frac{\partial\mu_i}{\partial\beta_s}\frac{\partial\mu_i}{\partial\beta_t}, \\
\nu_{\phi,st} &= \frac{1}{n}\Big(1-\frac{1}{n}\Big)\sum_{i=1}^{n}\left\{\frac{2V_i - \phi V_i V_i'' + 2\phi V_i'^2}{V_i^2} - \phi\frac{V_i'}{V_i}\frac{h_i''}{h_i'^2}\right\}\frac{\partial\mu_i}{\partial\beta_s}\frac{\partial\mu_i}{\partial\beta_t} \\
&\quad -\frac{2}{n^2}\sum_{i\neq j}\frac{1}{\sqrt{V_iV_j}}\frac{\partial\mu_i}{\partial\beta_s}\frac{\partial\mu_i}{\partial\beta_t}.
\end{aligned}$$

After computing some expectations of matrix products and with $trace(H) = p$ we finally have

$$\begin{aligned}
E(\tilde{\phi}) = \phi\Big(1-\frac{p}{n}\Big) &+ \frac{\phi^2}{2n}\left\{\sum_{i,j}\frac{V_i'}{V_i}\frac{h_j''}{V_j}h_i'h_j'Q_{ij}Q_{jj} - \sum_{i=1}^{n}\frac{V_i''h_i'^2 + V_i'h_i''}{V_i}Q_{ii}\right\} \\
&+ \frac{\phi}{n^2}\Big(1-\frac{1}{n}\Big)^{-1}\sum_{i\neq j}\frac{h_i'h_j'}{\sqrt{V_iV_j}}Q_{ij} + O(n^{-2}). \qquad (24)
\end{aligned}$$

The only difference between this result and the mean value of the mean Pearson statistic with general link function given in formula (10) is the extra term

$$\sum_{i\neq j}\frac{h_i'h_j'}{\sqrt{V_iV_j}}Q_{ij} = \sum_{i\neq j}H_{ij}, \qquad (25)$$

which is the sum over all entries of the general hat matrix without diagonal elements. In linear regression models, where $h(\eta) = \eta$ and $V_i = 1$, W is the $n\times n$ identity matrix. If X additionally includes the $n\times 1$ vector $1_n = (1,\ldots,1)^t$ (as in intercept models) then the sum over all entries in H equals n, because H is a projection matrix onto the space generated by the columns of X which especially results in $H1_n = 1_n$. Thus the elements of any row of H add up to one. Therefore, for this classical model the above sum (25) is equal to $n-p$ and expression (24) simplifies to the well known $X^2(y,\hat{\mu})$ result

$$E(\tilde{\phi}) = E\Big(\frac{1}{n-1}X^2(y,\hat{\mu})\Big) = \phi\frac{n-p}{n-1} + O(n^{-2}).$$

More generally, for generalized linear intercept models with $W = I_n$ as above, (25) equals $n-p$ resulting in

$$E(\tilde{\phi}) = \phi\frac{n-p}{n-1} + \frac{\phi^2}{2n}\tau + O(n^{-2}),$$

where τ denotes the difference within the parenthesis of (24). Equivalently, this class includes all models where $h_i'^2 = V_i$ holds. An important member is the squared binomial variance $V_i = \mu_i^2(1-\mu_i)^2$ for percentages when used together with the logit link function. In the following example a model for data with constant coefficient of variation is discussed in more detail.

Example 6: Gamma variability together with the log link function gives $V_i^{1/2} = \mu_i = h_i' = h_i''$. Hence, $\tau = -2p$ and

$$E(\tilde{\phi}) = \phi\frac{n-p}{n-1} - \phi^2\frac{p}{n} + O(n^{-2}).$$

However, for any generalized linear regression model only the upper bound $\sum_{i,j}|H_{ij}| \leq np^{1/2}$ holds by applying the Cauchy-Schwarz inequality and the third term in (24) is of order n^{-1}.

Example 7: It is interesting to note that in case of canonical link models $\sum_j H_{ij}V_j^{1/2} = V_i^{1/2}$ holds for every i. Hence, $\tau = 0$ results for loglinear Poisson models but the sum (25) is not longer equal to $n-p$.

6 Discussion

It is shown that the empirical variance $\tilde{\phi}$ of the Pearson residuals has a nice bootstrap motivation. In this context, $\tilde{\phi}$ can be seen as the dispersion parameter of the resampled observed centered Pearson residuals with respect to their empirical distribution. This estimate represents also a natural extension to the sum of squared residuals in ordinary linear intercept models, where the residual mean is zero. Real data applications often showed that the magnitude of the amount of the residual mean $\bar{e}$ in GLM's is considerably small.

Given the likelihood fit, the realisation of this statistic is very easy to calculate. Like the mean Pearson statistic, $\tilde{\phi}$ is asymptotically independent of $\hat{\beta}$. The asymptotic variance is very similar that of the mean Pearson statistic. Some considered examples also motivate to use its bias corrected version $\tilde{\phi}(n-1)/(n-p)$. To correct for further terms in the bias expression the elements of the generalized hat matrix are needed. This can be computationally expensive and is therefore not suggested.

Acknowledgment

The author is grateful to G. Kauermann for helpful discussions. He is also thankful to both referees for many valuable suggestions which led to an improvement in the presentation.

References

[1] Farrington, C.P. (1996), On assessing goodness of fit of generalized linear models to sparse data, *Journal of the Royal Statistical Society* **58**(2), 349–360.

[2] Firth, D. (1991), Generalized linear models, in: *Statistical Theory and Modelling*, Eds.: Hinkley, D.V., Reid, N., Snell, E.J., Chapman and Hall, London, 55–82.

[3] Friedl, H. (1995), Variance estimation in generalized linear models using the bootstrap, in: *Symposia Gaussiana, Conf. B*, Eds.: Mammitzsch, Schneeweiß, Walter de Gruyter, Berlin, 217–226.

[4] Friedl, H., Tilg, N. (1994), Variance estimates in logistic regression using the bootstrap, *Communications in Statistics - Theory and Methods* **24**(2), 473–486.

[5] McCullagh, P. (1985), On the asymptotic distribution of Pearson's statistic in linear exponential-family models, *International Statistical Review* **53**(1), 61–67.

[6] Moulton, L.H., Zeger, S.L. (1991), Bootstrapping generalized linear models, *Computational Statistics & Data Analysis* **11**, 53–63.

[7] Osius, G., Rojek, D. (1992), Normal goodness-of-fit tests for multinomial models with large degrees of freedom, *Journal of the American Statistical Association* **87**, 1145–1152.

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

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