Testing and Adjusting for Departures from Nominal Dispersion in Generalized Linear Models

Author(s): Philip J. Smith and Daniel F. Heitjan

Source: *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1993, Vol. 42, No. 1 (1993), pp. 31-41

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2347407

Appl. Statist. (1993)
42, *No.* 1, *pp.* 31–41

Testing and Adjusting for Departures from Nominal Dispersion in Generalized Linear Models

By PHILIP J. SMITH†

Centers for Disease Control, Atlanta, USA

and DANIEL F. HEITJAN

Pennsylvania State University College of Medicine, Hershey, USA

[Received March 1990. Final revision August 1991]

SUMMARY

In this paper we describe a score test of the hypothesis of no departure from nominal dispersion in a generalized linear model. We also give a method for adjusting the nominal variance–covariance matrix of the estimated regression coefficients when overdispersion is suspected. This procedure is an alternative to the traditional method of adjusting each element of the variance–covariance matrix by the same factor. We illustrate our method for the one-parameter exponential family of distributions in a logistic analysis of a factorial experiment and a Poisson regression analysis of bioassay data.

Keywords: Bioassay; Information measures; Logistic regression; Poisson regression; Score test

1. Introduction

When the actual variance of a measured response y exceeds its nominal variance under an assumed probability model, y is said to be overdispersed. McCullagh and Nelder (1989), p. 125, maintain that overdispersion is a common attribute of data arising in many fields of application and advise statistical practitioners to 'assume that overdispersion is present to some extent'.

With generalized linear models, interest often centres on estimates of regression parameters in relation to their standard errors. As Cox and Snell (1989) point out, when overdispersion is present, standard errors that do not adjust for overdispersion are too small and may result in misleading inferences. To avoid this, McCullagh and Nelder recommend that the variance-covariance matrix of estimated regression coefficients be multiplied by a dispersion factor to adjust for the alleged overdispersion.

This paper presents a score test for assessing whether overdispersion is present in generalized linear models. Our test statistic measures the discrepancy between the actual and nominal information for each regression parameter, a quantity whose distribution depends on the degree of overdispersion. Because these discrepancies can vary from one coefficient to another, we recommend adjusting elements of the variance-covariance matrix of the regression coefficients differentially. This

†*Address for correspondence*: Division of Diabetes Translation, Centers for Disease Control, Mail Stop K-10, 1600 Clifton Road, Atlanta, GA 30333, USA.

 0035–9254/93/42031 $2.00

TABLE 1
Data for seeds O. aegyptiaca 75 and 73, bean and cucumber root extracts

O. aegyptiaca 73						O. aegyptiaca 75					
Bean			Cucumber			Bean			Cucumber		
y	m	y/m	y	m	y/m	y	m	y/m	y	m	y/m
10	39	0.26	5	6	0.83	8	16	0.50	3	12	0.25
23	62	0.37	53	74	0.72	10	30	0.33	22	41	0.54
23	81	0.28	55	72	0.76	8	28	0.29	15	30	0.50
26	51	0.51	32	51	0.63	23	45	0.51	32	51	0.63
17	39	0.44	46	79	0.58	0	4	0.00	3	7	0.43
			10	13	0.77						

procedure differs from the traditional practice—advocated by McCullagh and Nelder (1989)—of adjusting each element of the variance–covariance matrix by the same factor.

Consider the data of Table 1, taken from Crowder (1978), on the number of seeds germinating in a factorial experiment in which two types of seed and two root extracts, bean and cucumber, were compared. In this example, the response variable is the number $\{y_i\}$ of seeds germinating for each replicate. Because some proportions within certain cells of the design are rather disparate, these data have been used by Crowder (1978), Williams (1982) and Cox and Snell (1989) to illustrate different methods of accounting for overdispersion. We note, however, that most of the apparently aberrant proportions are associated with cases for which the sample size is small relative to the other replicates in the cell. Thus, it is not obvious that the variation observed is excessive and should be attributed to overdispersion. In Section 4.1 we use our methods to assess the strength of evidence for overdispersion in this data set.

As another example, Table 2 lists data from a titration bioassay originally discussed by McCarthy *et al.* (1958) and reanalysed by Breslow (1990a). In this experiment viral activity was assessed from pock counts at a series of dilutions of the viral medium. As can be seen, the mean and variance of counts differ radically for each dilution; thus the customary assumption of Poisson variation is questionable. In Section 4.2 we discuss the pock count data more fully and show the extent to which estimated coefficients of a Poisson regression model should be adjusted to account for overdispersion.

TABLE 2
Bioassay data

Dilution factor	*Membrane pock counts*	*Mean*	*Variance*
1	116, 151, 171, 194, 196, 198, 208, 259	186.6	1781.1
2	71, 74, 79, 93, 94, 115, 121, 123, 135, 142	104.7	667.3
4	27, 33, 34, 44, 49, 51, 52, 59, 67, 92	50.8	360.4
8	8, 10, 15, 22, 26, 27, 30, 41, 44, 48	27.1	195.0
16	5, 6, 7, 7, 8, 9, 9, 9, 11, 20	9.2	16.8

The problem of overdispersion has attracted the attention of numerous researchers. Much early work studied overdispersion in Poisson models; examples are papers by Fisher (1950) and David and Johnson (1952). Papers by Moran (1954a, b, 1958), Armitage and Spicer (1956), Armitage (1959) and Neyman and Scott (1966) have considered overdispersion in bioassay models. Still others have considered overdispersion relative to Poisson regression and log-linear models; some papers are by Anderson (1988), Breslow (1984, 1990b), Dean and Lawless (1989) and Wilson (1989). Recently, various researchers have considered more general approaches to the overdispersion problem. One approach consists of modelling overdispersion by including additional variance parameters in the model, as advocated by Efron (1986), Albert and Peeple (1989) and Smyth (1989). A second approach is to consider the parameter of interest to be random and to test for overdispersion, as proposed by Bühler *et al.* (1965), Potthoff and Whittinghill (1966), Moran (1973), Lindsay (1980, 1982, 1983), Cox (1983) and Zelterman and Chen (1988). Other approaches are given by Moore (1986) and Barnwal and Paul (1988) in related problems. Our objective in this paper is to generalize the results of these researchers who considered only one-parameter problems to multiparameter generalized linear models; such models would be appropriate for both the seed and the pock data.

In Section 2 we describe our methods for detecting departures from nominal dispersion for the one-parameter exponential family of generalized linear models. These methods are founded on a mixture model methodology described by Lindsay (1980, 1982, 1983). In Section 3, we describe our method for adjusting the variance–covariance matrix of estimated regression coefficients when we wish to protect against departures from nominal dispersion. This method is based on adjustments described in the context of model misspecification tests by White (1982) and Royall (1986). In Section 4 we apply our methods to the seed and pock data; Section 5 concludes with a discussion.

2. Detecting Overdispersion

Let $y_1, \ldots, y_n$ be independently distributed and let y_i have probability density belonging to the one-parameter exponential family

$$f_i(y_i; \theta_i, \phi) = \exp[\{y_i\theta_i - b(\theta_i)\}/a_i(\phi) + c(y_i, \phi)], \tag{1}$$

$i = 1, \ldots, n$, where θ_i is the natural parameter, $b(\ )$ is the cumulant function, $a_i(\ )$ is a function of the scale parameter ϕ (and possibly a known prior weight w_i) and $c(y_i, \phi)$ is a normalizing constant. Then y_i has mean $\mu_i = E(y_i|\theta_i) = b'(\theta_i)$ and variance $V_i = V(y_i|\theta_i) = a_i(\phi)\, b''(\theta_i)$, where $b'(\ )$ and $b''(\ )$ denote the first and second derivatives of $b(\ )$. Assume further that the relationship between the mean μ_i of the random component (1) and the vector of p covariate factors $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^{\mathrm{T}}$ is given by a known monotonic link function $g(\ )$ having continuous first and second derivatives:

$$\eta_i = g(\mu_i) = \tilde{\boldsymbol{\beta}}_i^{\mathrm{T}}\mathbf{x}_i \tag{2}$$

$i = 1, \ldots, n$, where $\tilde{\boldsymbol{\beta}}_i$ is a vector of coefficients generated from an unknown distribution F with mean $\boldsymbol{\beta}$ and variance diag$(\boldsymbol{\tau})$, $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_p)^{\mathrm{T}}$. If $\tau_j = 0, j = 1, \ldots, p$, then equation (2) corresponds to a generalized linear model with no overdispersion. However, when $\tau_j > 0$ the variance of y is increased because of the variability in $\tilde{\boldsymbol{\beta}}_i$; y is therefore overdispersed.

Letting $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^{\mathrm{T}}$ and using the overdispersed model (2), the contribution to the likelihood for the ith observation is

$$\exp l_i(y_i; \boldsymbol{\beta}, \boldsymbol{\tau}) = p_i(y_i; \boldsymbol{\beta}, \boldsymbol{\tau}) = \int f_i(y_i; \tilde{\boldsymbol{\beta}}_i)\, \mathrm{d}F(\tilde{\boldsymbol{\beta}}_i; \boldsymbol{\beta}, \boldsymbol{\tau}),$$

and the null hypothesis of no overdispersion is

$$H_0: \boldsymbol{\tau} = \mathbf{0}. \tag{3}$$

Using L'Hôpital's rule and assuming regularity conditions that permit interchanging the orders of integration and differentiation (Cramér, 1946), the score statistics for hypothesis (3) are

$$\begin{aligned} U_{j.} &= \sum_{i=1}^{n} \frac{\partial}{\partial \tau_j} \ln p_i(y_i; \boldsymbol{\beta}, 0) \\ &= \frac{1}{2} \sum_{i=1}^{n} \left(\left[\frac{\partial \{\ln f_i(y_i; \boldsymbol{\beta})\}}{\partial \beta_j} \right]^2 - \left[- \frac{\partial^2 \{\ln f_i(y_i; \boldsymbol{\beta})\}}{\partial \beta_j^2} \right] \right), \end{aligned} \tag{4}$$

$j = 1, \ldots, p$, where $\boldsymbol{\beta}$ is evaluated at its maximum likelihood estimate under the null hypothesis. Liang (1987) describes the derivation of a similar score statistic in the context of case–control studies. Note that equation (4) is the difference between actual and nominal observed information for β_j. In the exponential family model of equations (1) and (2),

$$U_{j.} = \frac{1}{2} \sum_{i=1}^{n} \left[\frac{(y_i - \mu_i)^2 x_{ij}^2}{V_i^2 g'(\mu_i)^2} - \frac{x_{ij}^2}{V_i g'(\mu_i)^2} - \frac{(y_i - \mu_i)\{V_i' g'(\mu_i) + V_i g''(\mu_i)\} x_{ij}^2}{V_i^2 g'(\mu_i)^3} \right], \tag{5}$$

where $V_i' = a_i(\phi)\, b'''(\theta_i)/b''(\theta_i)$ and $b'''(\theta_i)$ denotes the third derivative of the cumulant function.

To test the null hypothesis (3) we propose the score statistic

$$\chi^2 = U^{\mathrm{T}} C_{\tau}^{-1} U \tag{6}$$

where

$$C_{\tau} = \mathscr{I}_{\tau\tau} - \mathscr{I}_{\beta\tau}^{\mathrm{T}} \mathscr{I}_{\beta\beta}^{-1} \mathscr{I}_{\beta\tau} \tag{7}$$

is the covariance matrix of U corrected for estimation of $\boldsymbol{\beta}$. Here,

$$\mathscr{I}_{\tau\tau} = \sum_i E\left\{ \left(\frac{\partial l_i}{\partial \boldsymbol{\tau}} \frac{\partial l_i}{\partial \boldsymbol{\tau}^{\mathrm{T}}} \right) \right\}, \qquad \mathscr{I}_{\beta\tau} = \sum_i E\left\{ \left(\frac{\partial l_i}{\partial \boldsymbol{\beta}} \frac{\partial l_i}{\partial \boldsymbol{\tau}^{\mathrm{T}}} \right) \right\},$$

$$\mathscr{I}_{\beta\beta} = \sum_i E\left\{ \left(\frac{\partial l_i}{\partial \boldsymbol{\beta}} \frac{\partial l_i}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \right) \right\}$$

where the scores $\partial l_i/\partial \boldsymbol{\tau}$ and $\partial l_i/\partial \boldsymbol{\beta}$ and their expectations are calculated at $\boldsymbol{\tau} = 0$ with $\boldsymbol{\beta}$ replaced by its maximum likelihood estimate under hypothesis (3). Appendix A gives explicit expressions for elements of matrices in equation (7). When H_0 is true, the asymptotic distribution of statistic (6) is the χ^2-distribution with p degrees of freedom (see Rao (1973), pp. 415–420). When statistic (6) is large relative to its degrees of freedom, a departure from nominal dispersion is indicated.

Cox and Hinkley (1974) point out that composite tests like statistic (6) have the disadvantage that if clear evidence of discrepancy with H_0 is obtained the test itself gives no indication of the nature of the departure. Thus, letting $C_\tau = \{C_{jk}\}$, to investigate more specific departures from nominal dispersion we may use the statistics

$$T_{\tau_j} = U_{j.}/C_{jj}^{1/2}, \qquad j = 1, \ldots, p.$$

From equation (5), an approximation to the mean of $U_{j.}$ for given $\boldsymbol{\tau}$ is

$$E[U_{j.} \mid \boldsymbol{\tau}] \doteq \frac{1}{2} \sum_{i=1}^{n} x_{ij}^2 \left(\sum_{k=1}^{p} x_{ik}^2 \tau_k \right) \Big/ V_i^2 g'(\mu_i)^4 \geqslant 0. \tag{8}$$

Thus, under the null hypothesis the T_{τ_j}-statistics have zero mean, unit variance and are approximately normally distributed, whereas under overdispersion they have positive mean. Tests should therefore be one sided, rejecting for large values of T_{τ_j}. Jones *et al.* (1989) have shown that tests of this form used to assess heterogeneity of odds ratios have good power. If all p tests are applied, and q_{obs} denotes the value of the smallest observed p-value, then pq_{obs} is an upper bound for the significance level of the overall procedure.

3. Compensating for Overdispersion

Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^{\mathrm{T}}$ denote the design matrix and $W = \operatorname{diag}\{w_i\}$, where $w_i^{-1} = V_i g'(\mu_i)^2$. When a departure from overdispersion is indicated, it is common to adjust the nominal variance-covariance matrix of $\hat{\boldsymbol{\beta}}$ by multiplying by a dispersion factor $\tilde{\sigma}^2$:

$$V(\hat{\boldsymbol{\beta}}) = \tilde{\sigma}^2 (\mathbf{X}^{\mathrm{T}} W \mathbf{X})^{-1}. \tag{9}$$

A possible dispersion factor is obtained by estimating the average amount by which the actual observed information exceeds the nominal:

$$\begin{aligned} \tilde{\sigma}^2 &= (n-p)^{-1} \sum_{i=1}^{n} \left[\frac{\partial \{\ln f_i(y_i; \boldsymbol{\beta})\}}{\partial \beta_j} \right]^2 \Big/ \left[-\frac{\partial^2 \{\ln f_i(y_i; \boldsymbol{\beta})\}}{\partial \beta_j^2} \right] \\ &= (n-p)^{-1} \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{V_i^2 g'(\mu_i)^2} \Big/ \left[\frac{1}{V_i g'(\mu_i)^2} + \frac{(y_i - \mu_i)\{V_i' g'(\mu_i) + V_i g''(\mu_i)\}}{V_i^2 g'(\mu_i)^3} \right]. \end{aligned} \tag{10}$$

This expression is useful when non-canonical link functions are used. For canonical links, dispersion factor (10) reduces to

$$\tilde{\sigma}^2 = \sum_{i=1}^{n} \frac{(y_i - \mu_i)^2}{V_i} \Big/ (n-p) \tag{11}$$

and is identical with the factor recommended by Payne (1987), McCullagh and Nelder (1989) and Aitkin *et al.* (1989). Adjustment factor (11) multiplies the estimated variances and covariances of all regression coefficients by the same amount. In a rough way, this scaling attempts to replace nominal variances with observed variances.

Alternatively, each element of the covariance matrix can be adjusted differentially. In particular, let

$$\tilde{\mathscr{I}}_{\beta\beta} = \sum_{i=1}^{n} \left[\frac{\partial\{\ln f_i(y_i;\theta_i)\}}{\partial\beta} \frac{\partial\{\ln f_i(y_i;\theta_i)\}}{\partial\beta^{\mathrm{T}}} \right]$$

$$= \sum_{i=1}^{n} \{(y_i-\mu_i)^2/V_i^2 g'(\mu_i)^2\}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}$$

where the expectation is with respect to the distribution of y_i under hypothesis (3). A consistent estimate of the variance–covariance matrix of estimated regression coefficients is given by

$$V(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^{\mathrm{T}}W\mathbf{X})^{-1}\tilde{\mathscr{I}}_{\beta\beta}(\mathbf{X}^{\mathrm{T}}W\mathbf{X})^{-1}. \tag{12}$$

This matrix, proposed in a more general context by White (1982) and Royall (1986), is robust in that it adjusts for departures from the nominal model (such as overdispersion) if they are present. Using equation (12) amounts to premultiplying the nominal covariance matrix $(\mathbf{X}^{\mathrm{T}}W\mathbf{X})^{-1}$ by $(\mathbf{X}^{\mathrm{T}}W\mathbf{X})^{-1}\tilde{\mathscr{I}}_{\beta\beta}$.

4. Applications

4.1. *Seed Germination Data*

Consider again Crowder's data in Table 1. To test for a departure from nominal dispersion we use GLIM (Payne, 1987) to fit a logistic model with parameters representing the factors of the 2×2 experiment. Table 3 gives results of the analysis. Like Crowder (1978), we find that the most important significant main effect is extracts and that the interaction between extracts and seeds is significant. Thus, we test departures from nominal dispersion by using the full interaction model.

For binomial data as in Table 1, let π_i denote the success probability and let y_i denote the number of germinated seeds out of m_i for the ith observation. Then the log-likelihood for observation i may be written in exponential form as

$$\ln f_i(y_i;\pi_i) = y_i \ln\left(\frac{\pi_i}{1-\pi_i}\right) + m_i \ln(1-\pi_i).$$

In the notation of Section 2, $\theta_i = \ln\{\pi_i/(1-\pi_i)\}$ and the cumulant function is $b(\theta_i) = \ln(1+\exp\theta_i)$. For logistic regression, the mean $\mu_i = m_i\pi_i$ is linked to the linear predictor by the logit function

TABLE 3
Analysis of Crowder's seed data

Model	*Degrees of freedom*	*Deviance*
1	20	98.72
Seeds	19	96.18
Extracts	19	42.75
Seeds + extracts	18	39.69
Interaction	17	33.28

$$\eta_i = g(\mu_i) = \ln\left(\frac{\mu_i}{m_i - \mu_i}\right) = \ln\left(\frac{\pi_i}{1-\pi_i}\right), \qquad i = 1, \ldots, n.$$

In this case,

$$U_{j.} = \frac{1}{2}\sum_{i=1}^{n} \{(y_i - m_i\pi_i)^2 - m_i\pi_i(1-\pi_i)\}x_{ij}^2.$$

Table 4 lists the test statistics T_{τ_j} of the four parameters along with parameter estimates and nominal and adjusted standard errors. Although all the T_{τ_j}-statistics are positive—suggesting a degree of overdispersion—one-sided tests suggest that the departures from nominal binomial dispersion are not appreciable. The χ^2-statistic (6) for assessing departures from nominal dispersion is $\chi^2 = 3.96$ with 4 degrees of freedom, yielding a p-value of 0.41. Thus, there is little evidence against null hypothesis (3) of no overdispersion when the full model is fitted. Finally, Table 4 shows that using equation (12) to adjust the standard errors of the regression coefficients yields only modest increases over the nominal standard errors.

Although this result may appear to contradict the popular conviction that these data are overdispersed, neither Crowder (1978), Williams (1982) nor Cox and Snell (1989) formally tested the hypothesis of no overdispersion. Our analyses agree with theirs in that we too find some evidence of overdispersion; we differ in assessing the significance of the overdispersion and finding that it is not great.

4.2. *Pock Count Data*

Letting $\{y_i\}$ denote the individual pock counts observed at the dilution factors $\{x_i'\}$, $j = 1, \ldots, n_k$, Fig. 1 gives a plot of $\ln x_i'$ *versus* $\ln y_i$. This figure strongly suggests a model of the form

$$\ln \mu_i = \tilde{\beta}_{i0} + \tilde{\beta}_{i1}x_{i2} \qquad (13)$$

where $x_{i2} = \ln x_i'$ and μ_i is the mean of y_i for its corresponding dilution factor. Assuming that the pock counts are distributed as Poisson variates at each dilution level, the corresponding generalized linear model has link function $g(\ ) = \ln(\ )$, natural parameter $\theta = \ln \mu_i$, cumulant function $b(\theta) = \exp\theta$ and $a_i(\phi) = 1$. Letting $x_{i1} = 1$, the score statistics (5) for testing departures from nominal dispersion are

TABLE 4
Statistics for the seed data

Parameter	$U_{j.}$	$C_{jj}^{1/2}$	T_{τ_j}	*Maximum likelihood estimate of* β_j	*Nominal standard error*	*Adjusted standard error*
Intercept	62.82	33.96	1.85	−0.558	0.126	0.176
Seeds	8.38	16.52	0.51	0.146	0.223	0.287
Extracts	25.22	24.89	1.01	1.318	0.177	0.242
Interaction	0.78	12.74	0.06	−0.778	0.306	0.374

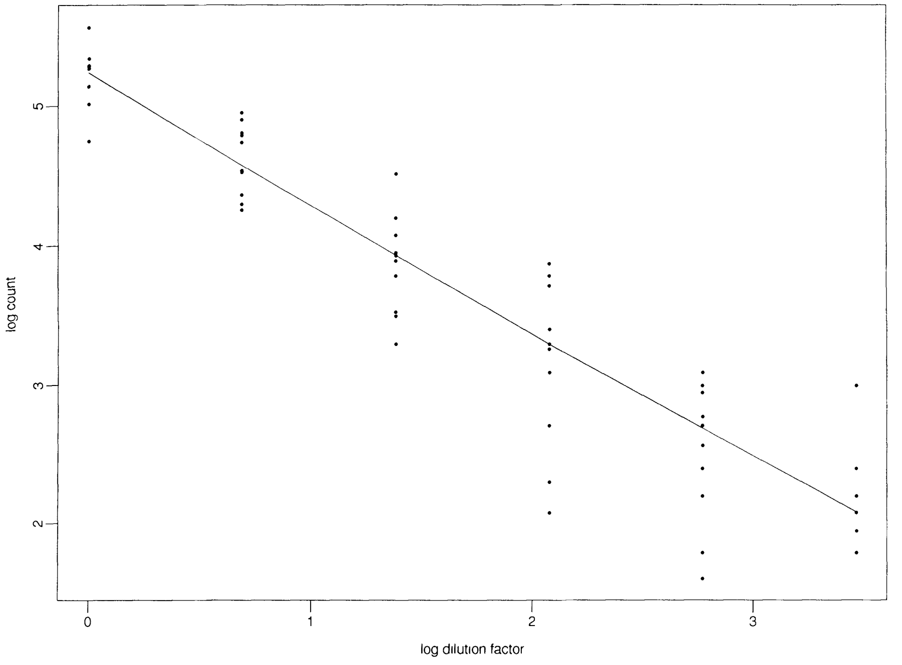

Fig. 1. Armitage's pock count data

$$U_{j.} = \frac{1}{2} \sum_{i=1}^{n} \{(y_i - \mu_i)^2 - \mu_i\} x_{ij}^2, \qquad j = 1, 2.$$

Table 5 lists the test statistics T_{τ_j} of the slope and intercept in model (13) along with parameter estimates and nominal and adjusted standard errors. Both standardized scores T_{τ_1} and T_{τ_2} are large and positive, indicating overdispersion. Score statistic (6) for the null hypothesis of no overdispersion is $\chi^2 = 597.7$ with 2 degrees of freedom, corresponding to a p-value of 0 in single precision. Thus our tests confirm the presence of overdispersion that Table 5 suggests. Table 5 also illustrates the effect of overdispersion on variability estimates. Use of equation (12) adjusts the nominal standard errors upwards by factors of 2–3.

5. Discussion

Crowder (1978) and Williams (1982) proposed methods to accommodate overdispersion specifically in binomial models. In comparison, our methods apply to distributions of exponential family type, encompassing a much broader variety of models including proportional hazards models, probit and logistic models, multiple regression and analysis of variance, Poisson regression and log-linear models.

In the binomial case, we note that the methods of Crowder (1978) and Williams (1982) are intended only to adjust for overdispersion, although Crowder does outline an overdispersion test. Neither researcher considers the problem of underdispersion. Our model does not include underdispersion alternatives either, but our test statistic could be used to detect underdispersion. This is because the score statistic $U_{j.}$ is the

TABLE 5
Statistics for the pock count data

Parameter	$U_{j.}$	$C_{jj}^{1/2}$	T_{τ_j}	*Maximum likelihood estimate of* β_j	*Nominal standard error*	*Adjusted standard error*
Intercept	10369.2	460.7	22.5	5.242	0.022	0.061
Slope	8337.7	496.0	16.8	−0.919	0.019	0.043

difference between the actual and nominal observed information for β_j, and therefore negative values of $U_{j.}$ suggest underdispersion. Power functions for underdispersion tests would need to involve stochastic models for underdispersion; these might include models of negative correlation among units or of less than nominal variability.

Acknowledgements

Professor Smith's work was supported by grants HL 43880, HL 33407 and HL 40916, and Professor Heitjan's work was supported by grant HL 33407 from the US National Institutes of Health. The authors thank Kinley Larntz, Bruce Lindsay and two referees for many valuable suggestions.

Appendix A: Derivation of C_{jk}

Letting l_i denote the log-likelihood for observation i under the null hypothesis of no departure from nominal dispersion,

$$\begin{aligned} U_{j.} &= \sum_i \frac{\partial l_i}{\partial \tau_j} \\ &= \frac{1}{2} \sum_i \left(\frac{\partial l_i}{\partial \beta_j}\right)^2 + \sum_i \frac{\partial^2 l_i}{\partial \beta_j^2}. \end{aligned}$$

Thus, the (j, k)th element of $\mathscr{I}_{\tau\tau}$ is

$$\begin{aligned} \mathscr{I}_{\tau_j\tau_k} = E\Bigg\{ & \sum_i \left(\frac{\partial l_i}{\partial \beta_j}\right)^2 \sum_i \left(\frac{\partial l_i}{\partial \beta_k}\right)^2 + \sum_i \left(\frac{\partial l_i}{\partial \beta_j}\right)^2 \sum_i \left(\frac{\partial^2 l_i}{\partial \beta_k^2}\right) \\ & + \sum_i \left(\frac{\partial^2 l_i}{\partial \beta_j^2}\right) \sum_i \left(\frac{\partial l_i}{\partial \beta_k}\right)^2 + \sum_i \left(\frac{\partial^2 l_i}{\partial \beta_j^2}\right) \sum_i \left(\frac{\partial^2 l_i}{\partial \beta_k^2}\right)\Bigg\} \\ = \frac{1}{4}\Bigg(& \sum_{i=1}^n \psi_i^{(4)} v_{ij}^2 v_{ik}^2 + 2 \sum_{h<i}^n \psi_i^{(2)} \psi_h^{(2)} v_{ij}^2 v_{hk}^2 + \sum_{i=1}^n \psi_i^{(2)} v_{ij}^2 \sum_{i=1}^n c_{ik} + \sum_{i=1}^n \psi_i^{(3)} v_{ij}^2 d_{ik} \\ & + \sum_{i=1}^n \psi_i^{(2)} v_{ik}^2 \sum_{i=1}^n c_{ij} + \sum_{i=1}^n \psi_i^{(3)} v_{ik}^2 d_{ij} + \sum_{i=1}^n c_{ij} \sum_{i=1}^n c_{ik} + \sum_{i=1}^n \psi_i^{(2)} d_{ij} d_{ik} \Bigg). \end{aligned}$$

Here, $\psi_i^{(2)} = \kappa_{2,i}$, $\psi_i^{(3)} = \kappa_{3,i}$ and $\psi_i^{(4)} = \kappa_{4,i} + 3\kappa_{2,i}^2$ are the second, third and fourth central moments of y_i and $\kappa_{r,i} = a_i(\phi)^{r-1} \partial^r b(\theta_i)/\partial \theta_i^r$ is the rth cumulant of y_i. Also,

$$v_{ij} = x_{ij}/V_i g'(\mu_i),$$
$$c_{ij} = -x_{ij}^2/V_i g'(\mu_i)^2,$$

and

$$d_{ij} = -\{V_i' g'(\mu_i) + V_i g''(\mu_i)\}x_{ij}^2/V_i^2 g'(\mu_i)^3.$$

A similar derivation yields the (j, k)th element of $\mathscr{I}_{\beta\tau}$:

$$\mathscr{I}_{\beta_j\tau_\kappa} = \frac{1}{2}\left(\sum_{i=1}^{n} \psi_i^{(3)} v_{ik}^2 v_{ij} + \sum_{i=1}^{n} \psi_i^{(2)} d_{ik} v_{ij}\right).$$

References

Aitkin, M., Anderson, D., Francis, B. and Hinde, J. (1989) *Statistical Modeling in GLIM*. New York: Oxford Science.

Albert, J. H. and Peeple, P. A. (1989) A Bayesian approach to some over-dispersion models. *Can. J. Statist.*, **17**, 333–344.

Anderson, D. (1988) Some models for over-dispersed binomial data. *Aust. J. Statist.*, **30**, 125–148.

Armitage, P. (1959) Host variability in dilution experiments. *Biometrics*, **15**, 1–9.

Armitage, P. and Spicer, C. C. (1956) The detection of variation in host susceptibility in dilution counting experiments. *J. Hyg. Camb.*, **54**, 401–414.

Barnwal, R. K. and Paul, S. R. (1988) Analysis of one-way layout of count data with negative binomial variation. *Biometrika*, **75**, 215–222.

Breslow, N. E. (1984) Extra-Poisson variation in log-linear models. *Appl. Statist.*, **33**, 38–44.

——— (1990a) Further studies in the variability of pock counts. *Statist. Med.*, **9**, 615–626.

——— (1990b) Tests of hypotheses in over-dispersed Poisson regression and other quasi-likelihood models. *J. Am. Statist. Ass.*, **85**, 565–571.

Bühler, W., Fein, H., Goldsmith, D., Neyman, J. and Puri, P. S. (1965) Locally optimal tests for homogeneity with respect to very rare events. *Proc. Natn. Acad. Sci. USA*, 673–680.

Cox, D. R. (1983) Some remarks on overdispersion. *Biometrika*, **70**, 269–274.

Cox, D. R. and Hinkley, D. V. (1974) *Theoretical Statistics*. New York: Chapman and Hall.

Cox, D. R. and Snell, E. J. (1989) *Analysis of Binary Data*, 2nd edn. New York: Chapman and Hall.

Cramér, H. (1946) *Mathematical Methods of Statistics*. Princeton: Princeton University Press.

Crowder, M. J. (1978) Beta-binomial anova for proportions. *Appl. Statist.*, **27**, 34–37.

David, F. N. and Johnson, N. L. (1952) The truncated Poisson. *Biometrics*, **8**, 275–285.

Dean, C. and Lawless, J. F. (1989) Tests for detecting over-dispersion in Poisson regression models. *J. Am. Statist. Ass.*, **84**, 467–472.

Efron, B. (1986) Double exponential families and their use in generalized linear models. *J. Am. Statist. Ass.*, **81**, 709–721.

Fisher, R. A. (1950) The significance of deviations from expectation in a Poisson series. *Biometrics*, **6**, 17–24.

Jones, M. P., O'Gorman, T. W., Lemke, J. H. and Woolson, R. F. (1989) A Monte Carlo investigation of homogeneity tests of the odds ratio under various sample size configurations. *Biometrics*, **45**, 171–181.

Liang, K. Y. (1987) A locally most powerful test for homogeneity with many strata. *Biometrika*, **74**, 259–264.

Lindsay, B. G. (1980) Nuisance parameters, mixture models, and the efficiency of partial likelihood estimators. *Phil. Trans. R. Soc. Lond.*, **296**, 639–665.

——— (1982) Conditional score function: some optimality results. *Biometrika*, **69**, 503–512.

——— (1983) Efficiency of the conditional score in a mixture setting. *Ann. Statist.*, **11**, 486–497.

McCarthy, K., Downie, A. W. and Armitage, P. (1958) The antibody response in man following infection with viruses of the pox group: I, An evaluation of the pock counting method for measuring neutralizing antibody. *J. Hyg. Camb.*, **56**, 84–100.

McCullagh, P. and Nelder, J. A. (1989) *Generalized Linear Models*, 2nd edn. New York: Chapman and Hall.

Moore, D. F. (1986) Asymptotic properties of moment estimators for overdispersed counts and proportions. *Biometrika*, **73**, 583–588.
Moran, P. A. P. (1954a) The dilution assay of viruses I. *J. Hyg. Camb.*, **52**, 189–193.
——— (1954b) The dilution assay of viruses II. *J. Hyg. Camb.*, **52**, 444–446.
——— (1958) Another test for heterogeneity of host resistance in dilution assays. *J. Hyg. Camb.*, **56**, 319–322.
——— (1973) Asymptotic properties of homogeneity test. *Biometrika*, **60**, 79–85.
Neyman, J. and Scott, E. (1966) On the use of $C(\alpha)$ optimal tests of composite hypotheses. *Bull. Int. Inst. Statist.*, **41**, 477–497.
Payne, C. D. (1987) *The GLIM System Release 3.77 Manual*. Oxford: Numerical Algorithms Group.
Potthoff, R. F. and Whittinghill, M. (1966) Testing for homogeneity: II, The Poisson distribution. *Biometrika*, **53**, 183–190.
Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, pp. 415–420. New York: Wiley.
Royall, R. R. (1986) Model robust confidence intervals using maximum likelihood estimators. *Int. Statist. Rev.*, **54**, 221–226.
Smyth, G. K. (1989) Generalized linear models with varying dispersion. *J. R. Statist. Soc.* B, **51**, 47–60.
White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica*, **50**, 1–25.
Williams, D. A. (1982) Extra-binomial variation in logistic models. *Appl. Statist.*, **31**, 144–148.
Wilson, J. R. (1989) Chi-square tests for overdispersion with multiparameter estimates. *Appl. Statist.*, **38**, 441–453.
Zelterman, D. and Chen, C. F. (1988) Homogeneity tests against central-mixture alternatives. *J. Am. Statist. Ass.*, **83**, 179–182.