Bartlett correction factors in logistic regression models

Lawrence H. Moulton *
University of Michigan, Ann Arbor, MI, USA

Lisa A. Weissfeld
University of Pittsburgh, Pittsburgh, PA, USA

Roy T. St. Laurent **
University of Michigan, Ann Arbor, MI, USA

Abstract: Bartlett correction factors for likelihood ratio tests of parameters in conditional and unconditional logistic regression models are calculated. The resulting tests are compared to the Wald, likelihood ratio, and score tests, and a test proposed by Moolgavkar and Venzon in *Modern Statistical Methods in Chronic Disease Epidemiology*, (Wiley, New York, 1986).

Keywords: Conditional logistic regression; Likelihood ratio test; Score test; Wald test.

1. Introduction

Epidemiologic case-control studies frequently have small sample sizes due to limited numbers of cases. It is common in such studies for logistic regression analyses to be employed, using maximum likelihood estimation and likelihood ratio, score, or Wald tests of parameter effects. However, these procedures are only asymptotically valid, and depend on the accuracy of the large sample approximations employed. The Wald test is attractive due to the ease with which it may be constructed. However, it is not invariant to parameter transformations. In addition, as Jennings (1986) illustrates it can exhibit unreliable behavior in that for a given sample size, parameter estimates farther from the null value may be judged less significant than those closer to the null value (see also Hauck and Donner, 1977).

Correspondence to: Roy St. Laurent, Ph.D., Department of Biostatistics, School of Public Health, University of Michigan, 109 South Observatory, Ann Arbor, MI 48109-2029, USA.
* Now with the Department of International Health at The Johns Hopkins University.
** Authorship is randomly ordered.

Numerous methods for drawing inferences about the parameters in the logistic regression model have been developed for use in small sample situations. Cox (1970) has described an exact approach to inference about a single parameter of the model. This approach has been extended and rendered more feasible via efficient algorithms for calculating joint and conditional distributions of the sufficient statistics (Bayer and Cox, 1979; Tritchler, 1984; Kirji, Mehta and Patel, 1987). However, the method still may take hours of computation, and current software permits only binary covariates to be included in the regression model.

Moolgavkar and Venzon (1986) have proposed a small sample correction to Wald-based confidence regions motivated by differential geometric considerations. These regions, based on transformations of the standard Wald regions, are obtained by solving a system of quasi-linear differential equations and may be inverted in the usual fashion to obtain a test of hypothesis.

Another approach to parameter inference is to improve upon the small sample performance of the likelihood ratio test. Bartlett (1937) proposed multiplying the likelihood ratio test statistic by a correction factor so that its distribution under the null hypothesis would more closely follow that of a χ^2 random variable. This approach, generalized by Lawley (1956), consists of matching the first moment of the test statistic to that of a χ^2 random variable. The effect is to bring the moments of the corrected statistic to within $O(n^{-2})$ of those of the appropriate chi-squared distribution (Lawley, 1956).

In this paper we give explicit expressions for the Bartlett correction factor in tests of location parameters in conditional and unconditional logistic regression models. In addition, we evaluate the performance of the resulting corrected likelihood ratio test statistic and compare its performance to the usual likelihood ratio, Wald, and score test statistics as well as the Moolgavkar–Venzon modified Wald procedure.

2. Calculation of Bartlett correction factors

2.1. Introduction

Let $L(\beta)$ denote the log likelihood function for the p-vector β based on n observations, and let L and $\hat{L}_p$ denote the value of the log likelihood evaluated at the true parameter vector β, and the maximum likelihood estimate $\hat{\beta}$, respectively. Lawley (1956) showed that under mild regularity conditions $2\mathrm{E}(\hat{L}_p - L) = p + \epsilon_p + O(n^{-2})$, where the correction term ϵ_p is of order n^{-1} and may be written (Cordeiro, 1983)

$$\epsilon_p = \kappa^{rs}\kappa^{tu}\left\{\tfrac{1}{4}\kappa_{rstu} - \kappa_{rst}^{(u)} + \kappa_{rt}^{(su)}\right\} - \kappa^{rs}\kappa^{tu}\kappa^{vw}$$

$$\times\left\{\tfrac{1}{6}\kappa_{rtv}\kappa_{suw} + \tfrac{1}{4}\kappa_{rtu}\kappa_{svw} - \kappa_{rtv}\kappa_{sw}^{(u)} - \kappa_{rtu}\kappa_{sw}^{(v)} + \kappa_{rt}^{(v)}\kappa_{sw}^{(u)} + \kappa_{rt}^{(u)}\kappa_{sw}^{(v)}\right\}.$$

Here all of the indices run from 1 to p and the usual summation convention is applied. The κ's are defined in terms of $L(\beta)$ as

$$\kappa_{rs} = \mathrm{E}\left\{\frac{\partial^2 L(\beta)}{\partial\beta_r\partial\beta_s}\right\}, \quad \kappa_{rst} = \mathrm{E}\left\{\frac{\partial^3 L(\beta)}{\partial\beta_r\partial\beta_s\partial\beta_t}\right\}, \quad \kappa_{rs}^{(t)} = \frac{\partial\kappa_{rs}}{\partial\beta_t}, \quad \kappa_{rs}^{(tu)} = \frac{\partial^2\kappa_{rs}}{\partial\beta_t\partial\beta_u},$$

and so on, where β_r is the rth component of β. In this notation, the information matrix $A = \{-\kappa_{rs}\}$ and its inverse is denoted $A^{-1} = \{-\kappa^{rs}\}$.

Suppose that $\beta^{\mathrm{T}} = (\beta_{1,1}, \ldots, \beta_{1,q}, \beta_{2,1}, \ldots, \beta_{2,p-q})$ and under the null hypothesis $\beta_{2,1} = \cdots = \beta_{2,p-q} = 0$. Let $\hat{L}_q$ be the maximized log likelihood under the null hypothesis, and ϵ_q be the correction term calculated under the null model. Then dividing the usual likelihood ratio test statistic $2(\hat{L}_p - \hat{L}_q)$ by the Bartlett factor $1 + (\epsilon_p - \epsilon_q)/(p - q)$ will improve the approximation of its distribution by a χ^2_{p-q} random variable under the null model. Generally the correction factor will be a function of β and should be estimated from the null model via maximum likelihood.

2.2. Unconditional logistic regression

The unconditional logistic regression model may be written

$$\mathrm{logit}\, \mathrm{P}(y_i = 1 \mid x_i) = x_i^{\mathrm{T}}\beta, \qquad i = 1, \ldots, n, \tag{1}$$

where x_i is the p-vector of covariates for the ith observation, the y_i are independent Bernoulli random variables, and β is a p-dimensional parameter vector. The log likelihood is then

$$L(\beta) = \sum_{i=1}^{n} \left[y_i \log\{\theta_i/(1 - \theta_i)\} + \log(1 - \theta_i) \right], \tag{2}$$

with

$$\theta_i = \exp(x_i^{\mathrm{T}}\beta)/\{1 + \exp(x_i^{\mathrm{T}}\beta)\}.$$

Cordeiro (1983, equation 8), gives an expression for ϵ_p applicable to any generalized linear model. Applying his result to (2), we have

$$\epsilon_p = \tfrac{1}{4}\mathrm{tr}(HZ_d^2) + \tfrac{1}{12}1_n^{\mathrm{T}}F(2Z^{(3)} + 3Z_d ZZ_d)F1_n, \tag{3}$$

where the $n \times n$ matrices F, H, Z, and Z_d are given by $F = V(I_n - 2\theta)$, $H = V(6V - I_n)$, $Z = X(X^{\mathrm{T}}VX)^{-1}X^{\mathrm{T}}$, and Z_d is a diagonal matrix composed of the diagonal elements of Z. Here $V = \mathrm{diag}\{\theta_i(1 - \theta_i)\}$ and $\theta = \mathrm{diag}\{\theta_i\}$ are $n \times n$ matrices and 1_n is an n-vector of 1's.

2.3. Conditional logistic regression

Case-control epidemiologic studies often employ matching or stratification in order to increase efficiency and reduce bias. In a regression framework, the preferred analysis of highly stratified data involves conditioning on stratum

membership (Breslow et al., 1978). The resulting conditional likelihood function is then maximized to yield conditional maximum likelihood estimates.

Consider a case-control study with K matched sets (or strata), where each stratum consists of one case and R controls, $n = K(R+1)$, $j = 0$ will denote the case in each stratum, so that $y_{0k} = 1$, and $y_{jk} = 0$ for $j \neq 0$. Here $i = 1, \ldots, K$ and $j = 0, \ldots, R$. We consider the logistic model

$$\text{logit } P(y_{jk} = 1 \mid x_{jk}) = \alpha_k + x_{jk}^T \beta,$$

where α_k depends on the values of the matching variables. Conditioning on the $1:R$ structure of the stratum leads to elimination of α_k resulting in the conditional log likelihood function

$$L(\beta) = \sum_{k=1}^{K} \sum_{j=0}^{R} y_{jk} \log(\mu_{jk}) = \sum_{k=1}^{K} \log(\mu_{0k}). \tag{4}$$

Here μ_{jk} depends upon the parameter vector β through the pair of transformations $\pi_{jk} = \exp(x_{jk}^T \beta)$ and $\mu_{jk} = \pi_{jk}/\Sigma \pi_{mk}$ (notation of Pregibon, 1984). Here x_{jk} is the p-vector of covariates for the jth individual in the kth matched set and π_{jk} is the corresponding odds ratio. The model depends upon β only through the odds ratio, therefore $x_{jk}^T \beta$ need not include an unknown constant, as it is not estimable.

The conditional logistic regression model is not a generalized linear model, hence the expressions for the correction term ϵ_p developed by Cordeiro (1983) cannot be applied to (4).

Since $\mu_{0k} + \cdots + \mu_{Rk} = 1$,

$$\tilde{x}_{jk} = x_{jk} - \sum_{m=0}^{R} \mu_{mk} x_{mk},$$

can be thought of as 'centered' x's. The Bartlett adjustment for the conditional model is a function of the covariates only through the $\tilde{x}_{jk}$'s. For example, the $p \times p$ information matrix A, may be written $A = \{-\kappa_{rs}\} = \Sigma A_k$, where $A_k = \Sigma \mu_{jk} \tilde{x}_{jk} \tilde{x}_{jk}^T = \tilde{X}_k^T W_k \tilde{X}_k$. Here the $(R+1) \times p$ matrix of centered covariates is $\tilde{X}_k = (\tilde{x}_{0k}, \ldots, \tilde{x}_{Rk})^T$, and W_k is an $(R+1) \times (R+1)$ diagonal matrix with μ_{jk} on the diagonal, $j = 0, \ldots, R$. It follows that the information matrix can be written $A = \tilde{X}^T W \tilde{X}$, where W is block diagonal with blocks $W_1, \ldots, W_K$, and $\tilde{X}$ is the $K(R+1) \times p$ matrix of centered covariates combined over matched sets.

The correction term for the one-to-R matched design is

$$\epsilon_p = \tfrac{1}{6} I_n^T W P^{(3)} W I_n - \tfrac{1}{4} I_n^T P_d^T W (I - PW) P_d I_n + \tfrac{1}{2} I_{R+1}^T U I_{R+1} + \tfrac{1}{4} \sum_{k=1}^{K} T_k^2, \tag{5}$$

where $T_k = \text{tr}(A_k A^{-1})$, and the $(R+1) \times (R+1)$ matrix $U = \Sigma W_k (\tilde{X}_k A_k^{-1} \tilde{X}_k^T)^{(2)} W_k$. In addition, $P = \tilde{X}(\tilde{X}^T W \tilde{X})^{-1} \tilde{X}^T$ and $P_d = \text{diag}(p_{ii})$, where p_{ii} is the ith diagonal element of P, $i = 1, \ldots, K(R+1)$. $P^{(3)}$ is the matrix whose elements are the cube of the elements of P. A derivation of this result can be found in the Appendix. Note that our calculations are carried out with respect

to the number of strata, K. The moments of the resulting likelihood ratio statistics thus are corrected to order $O(K^{-2})$ (Lawley, 1956).

It is informative to write (5) as

$$\epsilon_p = \tfrac{1}{4}\bar{\rho}_{13}^2 + \tfrac{1}{6}\bar{\rho}_{23}^2 - \tfrac{1}{4}\bar{\rho}_4,$$

where $\bar{\rho}_{13}^2 = 1_n^T P_d^T WPWP_d 1_n$, $\bar{\rho}_{23}^2 = 1_n^T WP^{(3)} W1_n$, and $\bar{\rho}_4 = 1_n^T P_d^T WP_d 1_n - 2(1_{R+1}^T U1_{R+1}) - \Sigma T_k^2$. The terms $\bar{\rho}_{13}^2$, $\bar{\rho}_{23}^2$ measure the squared skewness, and $\bar{\rho}_4$ measures the kurtosis of the score vector (McCullagh and Cox, 1986).

3. Numerical comparisons

3.1. Study design

To assess the agreement between the nominal and actual level of the adjusted likelihood ratio test statistic in unconditional and conditional logistic regression, and to explore the power of the test, a numerical study was conducted comparing the score, likelihood ratio, and Wald tests to the Bartlett corrected likelihood ratio test when testing a single element of β to be equal to zero. In addition, these results were compared to the confidence interval based test of Moolgavkar and Venzon (1986).

For each of several sample sizes n, B replicates of the basic experiment were performed. For each replicate, an n-vector x of independent pseudo-random Uniform $(0, 1)$ deviates was generated. For $n \leq 10$, the exact level and power were calculated based on a complete enumeration of the possible 2^n response vectors y, each weighted by its likelihood. Computing restrictions required simulation of the y vectors from the logistic model when $n > 10$. These values were generated from another set of Uniform $(0, 1)$ deviates. The multiplicative-congruential uniform pseudo-random generator of the GAUSS™ system was employed. For each sample size, we report the observed level and power as the mean rejection rate of the null hypothesis over the B replicates.

3.2. Unconditional logistic regression results

Here we used the model $x_i^T \beta = \beta_0 + \beta_1 x_i$, and compared the level and power of the five methods for testing the null hypothesis $\beta_1 = 0$ at nominal level $\alpha = 0.05$.

For this model with $p = 2$, the Moolgavkar–Venzon limits for β require the simultaneous solution of the following equations:

$$\frac{d\beta_1(t)}{dt} = \text{sg} Z_{\alpha/2} \left[\sum_{i=1}^{n} x_i^2 \theta_i(1-\theta_i) - \left\{ \sum_{i=1}^{n} x_i \theta_i(1-\theta_i) \right\}^2 \Big/ \sum_{i=1}^{n} \theta_i(1-\theta_i) \right]^{-1/2},$$

$$\frac{d\beta_0(t)}{dt} = -\left(\frac{\partial^2 L(\beta)}{\partial \beta_0^2} \right)^{-1} \left(\frac{\partial^2 L(\beta)}{\partial \beta_0 \partial \beta_1} \right) \frac{d\beta_1(t)}{dt},$$

Table 1
Comparison of average rejection rates for five tests, at nominal level $\alpha = 0.05$, of $H_0 : \beta_1 = 0$ in an unconditional logistic regression model. n = total number of observations. Columns 3–7, average rejection rate $\times 100$. Column 8, maximum standard error across the five averages $\times 100$.

β_1	n	Wald	Moolgavkar–Venzon	Score	Likelihood ratio	Bartlett corrected	Maximum std. error
0	10	0	9.00	5.68	9.01	6.35	0.27
	20	3.35	7.22	5.58	6.42	5.28	0.25
	40	4.56	6.15	5.31	5.73	5.14	0.15
1	10	0	10.96	7.42	11.76	8.44	0.28
	20	4.94	12.72	9.73	11.22	9.52	1.05
	40	12.52	16.22	14.31	15.37	14.21	0.70
2	10	0	13.97	11.01	18.15	13.35	0.79
	20	6.09	24.71	15.99	19.19	16.33	3.56
	40	30.24	39.45	34.77	37.41	35.25	1.45

with $\beta(0) = \hat{\beta}$, $L(\beta)$ and θ_i defined in (1) and (2). Here $Z_{\alpha/2}$ is the $(1 - \alpha/2)$th percentage point of the standard normal distribution and sg $= -1, +1$ for the lower and upper confidence limits respectively. These equations are integrated from $t = 0$ to 1. The solution to these equations can be obtained numerically using the Adams–Gear algorithm (Moolgavkar and Venzon, 1986).

Results are presented for $n = 10, 20, 40$, with $\beta_0 = 0$ for all runs. In Table 1 we give the estimates of level, and power for $\beta_1 = 1$, 2, for each of the five methods. Table 1 also includes the maximum standard error of the mean over the five tests, calculated from the B replicates. For $n = 10$, $B = 25$ replicates were used, while for $n = 20$ and 40, we took $B = 5$ replicates of 5000 simulations each.

We see that the poor performance of the Wald statistic was markedly improved by the Moolgavkar–Venzon correction. The score test and the Bartlett corrected likelihood ratio test both outperform the likelihood ratio test in the level $(\beta_1 = 0)$ runs, the latter being too liberal. Even though the Bartlett correction factor is designed to improve performance in small samples, at the smallest sample size $(n = 10)$ it performed slightly worse than the score test. However, for $n = 20$ and 40, rejection rates for the Bartlett corrected likelihood ratio test were closest to the nominal level.

The likelihood ratio test and Moolgavkar–Venzon test are both rather liberal, which translates into greater power, as seen in the $\beta_1 = 1$ and $\beta_1 = 2$ runs, where their rejection rates are all greater than those of the score and Bartlett corrected likelihood ratio test. For these runs, the power for the score and Bartlett corrected likelihood ratio test are very close.

Considering the relative ease with which the score test may be calculated, and the similarity between its behavior and that of the Bartlett corrected likelihood ratio test, based on these results it would appear that the score test is the preferred method of testing in this situation.

Table 2
Comparison of average rejection rates for five tests, at nominal level $\alpha = 0.05$, of $H_0 : \beta_1 = 0$ in a conditional logistic regression model. $R = 1$, $K = $ number of strata. Columns 3–7, average rejection rate $\times 100$. Column 8, maximum standard error across the five averages $\times 100$.

β_1	K	Wald	Moolgavkar–Venzon	Score	Likelihood ratio	Bartlett corrected	Maximum std. error
0	5	0	1.00	1.25	10.75	7.50	0.68
	6	0	4.25	3.00	9.00	6.63	0.64
	7	0	6.44	3.81	8.19	6.19	0.41
	8	0	7.63	3.75	6.91	5.44	0.38
	9	0	8.36	4.06	6.94	5.31	0.32
	10	0	8.25	4.17	6.71	5.29	0.24
	20	2.12	6.97	4.77	5.87	5.13	0.51
	40	3.89	5.77	4.82	5.29	4.93	0.14
1	5	0	0.96	1.78	13.22	9.43	0.81
	6	0	5.21	4.33	11.84	9.01	0.77
	7	0	8.13	5.87	11.48	8.95	0.63
	8	0	10.23	6.34	10.67	8.70	0.60
	9	0	12.13	7.08	11.19	8.92	0.44
	10	0	12.82	7.63	11.34	9.31	0.47
	20	6.34	16.23	12.38	14.54	13.14	1.77
	40	21.76	27.17	24.71	25.99	25.04	1.44
2	5	0	1.16	3.18	19.57	14.45	1.40
	6	0	7.07	7.96	19.25	15.31	1.40
	7	0	12.38	11.51	20.16	16.37	1.37
	8	0	17.60	13.58	20.66	17.54	1.44
	9	0	21.69	15.51	22.44	18.72	1.19
	10	0	24.20	17.31	23.64	20.27	1.30
	20	20.85	40.18	33.69	37.27	34.95	6.13
	40	65.06	70.97	68.38	69.86	68.82	3.20

3.3. Conditional logistic regression results

Here the Moolgavkar–Venzon confidence interval is obtained by solving two related differential equations. For $p = 1$, the interval endpoints are found by integrating

$$\frac{d\beta(t)}{dt} = \mathrm{sg} Z_{\alpha/2} \left[\sum_{k=1}^{K} \sum_{j=1}^{R} (x_{jk} - x_{0k})^2 \pi_{jk}(t) \pi_{0k}(t) \left\{ \sum_{m=0}^{R} \pi_{mk}(t) \right\}^{-2} \right]^{-1/2},$$

from $t = 0$ to 1, with $\beta(0) = \hat{\beta}$ and $\pi_{jk}(t) = \exp\{x_{jk}^{\mathrm{T}} \beta(t)\}$. Numerical solutions are obtained using the Adams–Gear algorithm.

For the conditional model, the performance of the five methods was compared in the one-to-one matched case-control ($R = 1$) model with one covariate via a test of the null hypothesis $\beta_1 = 0$ at nominal level $\alpha = 0.05$.

Results are presented for $K = 5(1)10$, 20, 40 matched sets. Table 2 gives the simulation estimates of level and power (for $\beta_1 = 1$, 2) for each of the five

methods and the maximum standard error of the mean, across the five tests. For $K \le 10$, $B = 25$ replicates were used, while for $K = 20$ and 40, $B = 5$ replicates were run of 5000 simulations each.

In general, the simulation results for the conditional model parallels those of the unconditional model. For small sample sizes, the score and Wald tests are generally too conservative, while the likelihood ratio test and Bartlett corrected likelihood ratio test are too liberal. The Moolgavkar–Venzon procedure performs rather erratically (moving from conservative to liberal as K increases) and poorly overall when compared to most other methods. The rejection rates for the Bartlett corrected likelihood ratio test, with the exception of the $K = 6$ run, are closest to the nominal level.

In the power runs, $\beta_1 = 1$, 2, the Bartlett corrected likelihood ratio test clearly performed better than the score test. The Moolgavkar–Venzon test and likelihood ratio test exhibited the highest power, but this performance should be discounted by their liberal results in the level runs.

4. Discussion

For the conditional analysis of one-to-one matched data with a single covariate ($p = 1$), the Bartlett correction for the null hypothesis $\beta_1 = 0$ has a particularly simple form. Let $z_{jk} = x_{jk} - \bar{x}_{.k}$, the stratum mean centered covariate, and let s_k^2 be the variance of x_{jk} within the kth stratum. Then $\bar{\rho}_{13}^2 = \bar{\rho}_{23}^2 = \hat{\gamma}_1^2 / K$ and $\bar{\rho}_4 = (\hat{\gamma}_2 - 3\hat{C}^2)/K$, where $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are the standardized skewness and kurtosis of z_{jk} ($j = 0, \ldots, R$, $k = 1, \ldots, K$), respectively, and $\hat{C}$ is the coefficient of variation of s_k^2 ($k = 1, \ldots, K$). Thus $\epsilon_1 = (5\hat{\gamma}_1^2 - 3\hat{\gamma}_2 + 9\hat{C}^2)/(12K)$.

In the case of one-to-one matching ($R = 1$), $\hat{\gamma}_1 = 0$ and ϵ_1 may be simplified to

$$\epsilon_1 = \frac{1}{2} \frac{\sum (x_{0k} - x_{1k})^4}{\left\{ \sum (x_{0k} - x_{1k})^2 \right\}^2},$$

where sums in the numerator and denominator run over $k = 1, \ldots, K$. This expression may also be written in terms of deviations $x_{jk} - \bar{x}_{.k}$. In either form, the correction is essentially a measure of the kurtosis of the distribution of the covariates.

If in addition, x_{jk} is a binary covariate, then $\epsilon_1 = (2m)^{-1}$ where m is the number of strata in which $x_{0k} \ne x_{1k}$ ($0 \le m \le K$), so that the Bartlett factor becomes $1 + (2m)^{-1}$. Here the likelihood ratio test statistic may be written $2m\{\log 2 + r \log r + (1 - r) \log(1 - r)\}$, where r is the proportion of the m strata for which $x_{0k} \ne x_{1k}$ and $x_{0k} = 1$ (so that $1 - r$ is the proportion for which $x_{1k} = 1$). Performing calculations similar to those of Frydenberg and Jensen (1989) for this likelihood, we found the order of the correction in this case to be less than $O(K^{-1})$. This agrees with their findings that in the case of fully discrete data, the Bartlett correction offers little, if any, improvement. More-

over, for up to at least $K = 50$ strata, the tail-area probability of the Bartlett corrected likelihood ratio test is the same as for that of the score (McNemar) test at size $\alpha = 0.05$. This stands in contrast to our results for the continuous covariate case (Table 2), where the Bartlett corrected likelihood ratio test performs better than the score test.

The results of our runs for both the conditional and unconditional models indicate that the computational burden of obtaining the Moolgavkar–Venzon interval may not be worth its modest performance when used for testing. Instead of trying to improve the Wald test, better results are had by improving the likelihood ratio test via the relatively simple Bartlett correction.

Many factors may conspire to render even a data set with a large number of observations one in which standard large-sample asymptotics do not provide sufficiently accurate approximations. These include extreme response probabilities, many estimated parameters with respect to the number of observations, sparse corners of the covariate space, and link functions that yield relatively flat likelihood functions. In such situations, it may be prudent to use a Bartlett corrected likelihood ratio test. For the logit link case we have studied here, the score test is also to be preferred over the Wald and likelihood ratio test.

Appendix: Derivation of (5)

From the definitions of μ_{jk} and $\tilde{x}_{jk}$ following (4), these identities are easily established:

$$\sum_{j=0}^{R} \mu_{jk} \tilde{x}_{jk} = 0, \tag{A.1}$$

$$\frac{\partial}{\partial \beta} \mu_{jk} = \mu_{jk} \tilde{x}_{jk}, \tag{A.2}$$

$$\frac{\partial}{\partial \beta^{\mathrm{T}}} \tilde{x}_{jk} = - \sum_{m=0}^{R} \mu_{mk} x_{mk} \tilde{x}_{mk}^{\mathrm{T}} = - \sum_{m=0}^{R} \mu_{mk} \tilde{x}_{mk} \tilde{x}_{mk}^{\mathrm{T}}. \tag{A.3}$$

Taking second partial derivatives of the log likelihood in (4), and making use of the identities above, gives

$$\frac{\partial^2 L(\beta)}{\partial \beta_r \partial \beta_s} = - \sum_{k=1}^{K} \sum_{j=0}^{R} \mu_{jk} \tilde{x}_{jkr} \tilde{x}_{jks}^{\mathrm{T}}, \tag{A.4}$$

where $\tilde{x}_{jkr}$ is the rth element of the p-vector $\tilde{x}_{jk}$. As the elements of the second partial derivative matrix in (A.4) are non-stochastic, it follows that the right-hand side of (A.4) is $-\kappa_{rs}$. Hence

$$A = \sum_{k=1}^{K} \sum_{j=0}^{R} \mu_{jk} \tilde{x}_{jk} \tilde{x}_{jk}^{\mathrm{T}},$$

which also may be written $A = \tilde{X}^{\mathrm{T}} W \tilde{X}$.

From (A.4) it also follows that $\kappa_{rs}^{(t)} = \kappa_{rst}$, $\kappa_{rs}^{(tu)} = \kappa_{rstu}$, and so on. Therefore the expression for the Bartlett correction may be written

$$\epsilon_p = \tfrac{1}{4}\kappa^{rs}\kappa^{tu}\kappa_{rstu} - \kappa^{rs}\kappa^{tu}\kappa^{vw}\{\tfrac{1}{6}\kappa_{rtv}\kappa_{suw} + \tfrac{1}{4}\kappa_{rtu}\kappa_{svw}\}. \tag{A.5}$$

Taking partial derivatives of κ_{rs} from (A.4), and employing the identities (A.1)–(A.3), the expression for a mixed third partial derivative of $L(\beta)$ is written

$$\kappa_{rst} = \frac{\partial^3 L(\beta)}{\partial\beta_r \partial\beta_s \partial\beta_t} + \sum_{k=1}^{K}\sum_{j=0}^{R} \mu_{jk}\tilde{x}_{jkr}\tilde{x}_{jks}\tilde{x}_{jkt}. \tag{A.6}$$

The expression for a mixed fourth partial derivative of $L(\beta)$ is obtained from κ_{rst} in a similar fashion, yielding

$$\kappa_{rstu} = \frac{\partial^4 L(\beta)}{\partial\beta_r \partial\beta_s \partial\beta_t \partial\beta_u}$$

$$= \sum_{k=1}^{K}\sum_{j=0}^{R} \mu_{jk}\tilde{x}_{jkr}\tilde{x}_{jks}\tilde{x}_{jkt}\tilde{x}_{jku}$$

$$- \sum_{k=1}^{K}(A_{krs}A_{ktu} + A_{krt}A_{ksu} + A_{kru}A_{kst}), \tag{A.7}$$

where $A_{krs} = \sum_{j=0}^{R}\mu_{jk}\tilde{x}_{jkr}\tilde{x}_{jks}$.

Substituting the expressions (A.4), (A.6) and (A.7) into (A.5) and employing a fair amount of matrix algebra inspired by McCullagh (1987, page 218), and equation (11) of Cordeiro (1983), we obtain the matrix form for the Bartlett correction given in (5).

References

Bartlett, M.S., Properties of sufficiency and statistical tests, *Proceedings of the Royal Statistical Society A*, **160** (1937) 268–282.

Bayer, L., and C. Cox, Algorithm AS142: Exact tests of significance in binary regression models, *Appl. Statist.*, **28** (1979) 319–324.

Breslow, N.E., N.E. Day, K.T. Halvorsen, R.L. Prentice and C. Sabai, Estimation of multiple relative risk functions in matched case-control studies, *Amer. J. Epidemiol.*, **108** (1978) 299–307.

Cordeiro, G.M., Improved likelihood ratio statistics for generalized linear models, *J. Roy. Statist. Soc. B*, **45** (1983) 404–413.

Cox, D.R., *The Analysis of Binary Data* (Methuen, London, 1970).

Frydenberg, M., and J.L. Jensen, Is the 'improved likelihood ratio statistic' really improved in the discrete case?, *Biometrika*, **76** (1989) 655–661.

Hauck, W.W., and A. Donner, Wald's test as applied to hypotheses in logit analysis, *J. Am. Statist. Assoc.*, **72** (1977) 851–853.

Jennings, D.E., Judging inference adequacy in logistic regression, *J. Am. Statist. Assoc.*, **81** (1986) 471–476.

Kirji, K.F., C.R. Mehta and N.R. Patel, Computing distributions for exact logistic regression, *J. Amer. Statist. Assoc.*, **82** (1987) 1110–1117.

Lawley, D.N., A general method for approximating to the distribution of likelihood ratio criteria, *Biometrika*, **43** (1956) 295–303.

McCullagh, P., *Tensor Methods in Statistics* (Chapman and Hall, London, 1987).

McCullagh, P., and D.R. Cox, Invariants and likelihood ratio statistics, *Ann. Statist.*, **14** (1986) 1419–1430.

Moolgavkar, S.H., and D.J. Venzon, Confidence regions for case-control and survival studies with general relative risk functions, in: S.H. Moolgavkar and R.L. Prentice (Eds.), *Modern Statistical Methods in Chronic Disease Epidemiology* (Wiley, New York, 1986).

Pregibon, D., Data analytic methods for matched case-control studies, *Biometrics*, **40** (1984) 639–651.

Tritchler, D., An algorithm for exact logistic regression, *J. Amer. Statist. Assoc.*, **79** (1984) 709–711.