Biometrika (1997), **84**, 2, *pp.* 309–326
Printed in Great Britain

Variance component testing in generalised linear models with random effects

By XIHONG LIN

Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, U.S.A.
e-mail: xlin@sph.umich.edu

Summary

There is considerable interest in testing for overdispersion, correlation and heterogeneity across groups in biomedical studies. In this paper, we cast the problem in the framework of generalised linear models with random effects. We propose a global score test for the null hypothesis that all the variance components are zero. This test is a locally asymptotically most stringent test and is robust in the special sense that the test does not require specifying the joint distribution of the random effects. We also propose individual score tests and their approximations for testing the variance components separately. Both tests can be easily implemented using existing statistical software. We illustrate these tests with an application to the study of heterogeneity of mating success across males and females in an experiment on salamander matings, and evaluate their performance through simulation.

Some key words: Correlated data; Generalised linear mixed model; Laplace approximation; Locally asymptotically most stringent test; Overdispersion; Score test.

1. Introduction

Overdispersed and correlated data are frequently observed in many biomedical studies and may be modelled using generalised linear models with random effects (Breslow & Clayton, 1993). Various approximate and Bayesian inference procedures have been proposed to fit such models (Breslow & Lin, 1995; Zeger & Karim, 1991). However, little has been done with respect to testing in generalised linear models with random effects.

There is considerable interest in testing for overdispersion, heteroscedasticity and correlation among outcomes. These tests may be used to ascertain when ordinary generalised linear models are applicable in view of their simplicity and appeal. Another area of potential application is to test for variation over groups in epidemiological studies and clinical trials. For example, in genetic epidemiology, such tests may be used to study familial aggregation of a disease, which may be determined by genetic factors; in ecologic studies, it is of interest to investigate whether or not population disease rates, which may be influenced by environmental factors, are homogeneous in different geographical regions; in clinical trials, it is important to test for institutional variation in the effects of therapy on survival; in longitudinal studies, it may be useful to study whether or not random subject effects are present. Several authors, including Cox (1983), Dean & Lawless (1989), Dean (1992) and Smith & Heitjan (1993), used score tests to detect the presence of overdispersion in standard binomial and Poisson regression models, or more generally in generalised linear models. Gray (1995) and Liang (1987) applied similar tests for homogeneity in different groups. Jacqmin-Gadda & Commenges (1995) extended Liang's (1987) results to canonical generalised linear models with a random intercept and generalised estimating equations with a single correlation parameter.

The authors mentioned above only considered the simple special cases of overdispersion among independent outcomes, or a single source of heterogeneity. However, many problems encountered in practice involve multiple sources of variation (Breslow & Clayton, 1993; Drum & McCullagh, 1993). One example is an experiment on salamander matings (McCullagh & Nelder, 1989), where it is of interest to study whether or not there is heterogeneity of mating success among female and male salamanders. We develop in this paper a unifying theory for testing for correlation and heterogeneity in the framework of generalised linear models with random effects. We use the Laplace expansion of the integrated log-quasilikelihood to derive a global score test for the null hypothesis that all the variance components are zero. The asymptotic distribution and optimality of the proposed tests are discussed. All the overdispersion and homogeneity tests discussed above can be obtained as special cases of generalised global variance component tests. Individual score tests and their approximations are proposed for testing the variance components separately. The tests are applied to the salamander data, and their performance is evaluated through simulation.

2. The generalised linear model with random effects

Suppose that the data are composed of n observations, with responses y_i, $p \times 1$ covariate vectors x_i associated with fixed effects and $q \times 1$ covariate vectors z_i associated with random effects. Conditional on a $q \times 1$ vector of unobservable random effects b, the responses y_i are assumed to be independent with means $E(y_i|b) = \mu_i^b$ and variances $\mathrm{var}(y_i|b) = V(\mu_i^b) = \phi a_i^{-1} v(\mu_i^b)$, where ϕ is a scale parameter, a_i is a prior weight and $v(.)$ is a variance function. Suppose that the y_i follow a conditional generalised linear model

$$g(\mu_i^b) = \eta_i^b = x_i^{\mathrm{T}}\alpha + z_i^{\mathrm{T}}b, \tag{1}$$

where $g(.)$ is a monotonic differentiable link function and α is a $p \times 1$ vector of fixed effects. Let

$$\mu^b = (\mu_1^b, \ldots, \mu_n^b)^{\mathrm{T}}, \quad \eta^b = (\eta_1^b, \ldots, \eta_n^b)^{\mathrm{T}}, \quad g(\mu^b) = \{g(\mu_1^b), \ldots, g(\mu_n^b)\}^{\mathrm{T}}.$$

In matrix form, model (1) can be written

$$g(\mu^b) = \eta^b = X\alpha + Zb, \tag{2}$$

where the design matrices X and Z have rows x_i^{T} and z_i^{T}.

The model specification is completed by assuming that the random effects b are generated from some distribution F with mean zero and covariance matrix $D(\theta)$, where θ is an $m \times 1$ vector of unknown variance components. The magnitude of θ can be used to measure the degree of overdispersion and correlation. Without loss of generality, we postulate that each component of $D(\theta)$ is a linear function of θ and that $D(\theta) = 0$ if $\theta = 0$. We further assume that the third and higher-order moments of the random effects b are of order $o(\|\theta\|)$. This condition is satisfied if the random effects b have an exponential-family distribution (McCullagh & Nelder, 1989, p. 350), or a mixture of exponential-family distributions (Johnson & Kotz, 1970, p. 88). When $\theta = 0$, model (2) reduces to the ordinary generalised linear model $g(\mu_i) = \eta_i = x_i^{\mathrm{T}}\alpha$, where $\mu_i = E(y_i)$.

The integrated quasilikelihood of (α, θ) thus takes the form

$$L(\alpha, \theta) = \exp\{l(\alpha, \theta)\} = \int \exp\left\{\sum_{i=1}^{n} l_i(\alpha; b)\right\} dF(b; \theta), \tag{3}$$

where

$$l_i(\alpha; b) \propto \int_{y_i}^{\mu_i^b} \frac{a_i(y_i - u)}{\phi v(u)}\, du \tag{4}$$

defines the conditional log-quasilikelihood of α given b. Breslow & Clayton (1993) and Zeger & Karim (1991) studied more restrictive models, the generalised linear mixed models, where the distribution of the random effect b is fully specified as multivariate normal.

Model (2) has broad application in biomedical studies (Breslow & Clayton, 1993). It encompasses both nested and crossed designs. For clustered designs, we assume that the data are arranged in a series of K independent clusters (Zeger & Karim, 1991). Let $\mathcal{Y}_k$ be an $n_k \times 1$ outcome vector of the kth cluster, and let X_k and Z_k be $n_k \times p$ and $n_k \times s$ covariate matrices. Model (2) can be written

$$g(\mu_k^{b_k}) = X_k \alpha + Z_k b_k, \tag{5}$$

where $\mu_k^{b_k} = E(\mathcal{Y}_k | b_k)$ and b_k is an $s \times 1$ vector of random effects. For hierarchical and crossed designs, we consider the random effect models with m independent random effects $b = (b_1^{\mathrm{T}}, \ldots, b_m^{\mathrm{T}})^{\mathrm{T}}$, where the jth random effect b_j is a $q_j \times 1$ random vector and follows some distribution F_j with mean 0 and variance $\theta_j I_{q_j}$ (Drum & McCullagh, 1993; Lin & Breslow, 1996), and I_{q_j} here is an identity matrix of dimension q_j. Model (2) hence takes the form

$$g(\mu^b) = X\alpha + Z_1 b_1 + \ldots + Z_m b_m, \tag{6}$$

where Z_j is a design matrix associated with the jth random effect b_j and often consists of zeros and ones.

3. The global variance component tests

3·1. *A likelihood expansion*

To derive a global score test for $\theta = 0$, it is necessary to calculate $\partial l(\alpha, \theta)/\partial\theta$ and evaluate the resultant derivative at $\theta = 0$. However, a direct calculation of this derivative from (3) is often difficult, since it may involve multidimensional integrals. We hence adopt an alternative approach by taking an expansion of the integrated log-likelihood $l(\alpha, \theta)$ about $\theta = 0$ using the Laplace method.

Following Solomon & Cox (1992) and Breslow & Lin (1995), we expand the integrated quasilikelihood (3), for small values of the variance components, by taking a quadratic expansion $\sum_{i=1}^{n} l_i(\alpha; b)$ about the true means of the random effects ($b = 0$) before integration. A straightforward two-term Taylor expansion gives

$$\exp\left\{\sum_{i=1}^{n} l_i(\alpha; b)\right\} = \exp\left\{\sum_{i=1}^{n} l_i(\alpha; 0)\right\}\left(1 + \sum_{i=1}^{n} \frac{\partial l_i(\alpha; 0)}{\partial \eta_i} z_i^{\mathrm{T}} b \right.$$
$$\left. + \frac{1}{2} b^{\mathrm{T}}\left[\left\{\sum_{i=1}^{n} \frac{\partial l_i(\alpha; 0)}{\partial \eta_i} z_i\right\}\left\{\sum_{i=1}^{n} \frac{\partial l_i(\alpha; 0)}{\partial \eta_i} z_i^{\mathrm{T}}\right\} + \sum_{i=1}^{n} \frac{\partial^2 l_i(\alpha; 0)}{\partial \eta_i^2} z_i z_i^{\mathrm{T}}\right] b + \varepsilon\right),$$

where the residual ε contains the third and higher order terms of b. Writing the integrated quasilikelihood (3) as $L(\alpha, \theta) = E[\exp\{\sum_{i=1}^{n} l_i(\alpha; b)\}]$ and using the moment assumptions on the random effects b in § 2, we have

$$L(\alpha, \theta) = \exp\left\{\sum_{i=1}^{n} l_i(\alpha; 0)\right\}$$
$$\times \left\{1 + \frac{1}{2}\operatorname{tr}\left(\left[\left\{\sum_{i=1}^{n} \frac{\partial l_i(\alpha; 0)}{\partial \eta_i} z_i\right\}\left\{\sum_{i=1}^{n} \frac{\partial l_i(\alpha; 0)}{\partial \eta_i} z_i^{\mathrm{T}}\right\} + \sum_{i=1}^{n} \frac{\partial^2 l_i(\alpha; 0)}{\partial \eta_i^2} z_i z_i^{\mathrm{T}}\right] D(\theta)\right)\right.$$
$$\left. + o(\|\theta\|)\right\}.$$

The marginal log-quasilikelihood $l(\alpha, \theta)$ can then be written, compare Breslow & Lin (1995, eqn (8)),

$$l(\alpha, \theta) = \sum_{i=1}^{n} l_i(\alpha; 0) + \frac{1}{2}\operatorname{tr}\left[Z^{\mathrm{T}}\left\{\frac{\partial l(\alpha; 0)}{\partial \eta}\frac{\partial l(\alpha; 0)}{\partial \eta^{\mathrm{T}}} + \frac{\partial^2 l(\alpha; 0)}{\partial \eta\, \partial \eta^{\mathrm{T}}}\right\} Z D(\theta)\right] + o(\|\theta\|), \quad (7)$$

where $\partial l(\alpha; 0)/\partial \eta$ is an $n \times 1$ vector whose ith component is $\partial l_i(\alpha; 0)/\partial \eta_i$, and $\partial^2 l(\alpha; 0)/\partial \eta\, \partial \eta^{\mathrm{T}}$ is an $n \times n$ diagonal matrix with elements $\partial^2 l_i(\alpha; 0)/\partial \eta_i^2$ on the diagonal. Note that (7) mimics a Laplace expansion of $l(\alpha, \theta)$ (Solomon & Cox, 1992). Solomon & Cox (1992) and Breslow & Lin (1995) restricted their attention to the special case $m = 1$ and assumed that the random effects were normally distributed. Lin & Breslow (1996) extended these authors' results to the independent random effect model (6) with the random effects b_j assumed to be normal. Equation (7) allows a general structure of $D(\theta)$ and does not require normality of the random effects b.

3·2. *The global variance component tests*

In this section, we use the results in § 3·1 to derive a global score test for the null hypothesis $H_0: \theta = 0$ under the general model (2), which corresponds to no overdispersion and no correlation among outcomes. Let Δ and W be $n \times n$ diagonal matrices with elements

$$\delta_i = 1/g'(\mu_i), \quad w_i = [V(\mu_i)\{g'(\mu_i)\}^2]^{-1} = \{V(\mu_i)\}^{-1}\delta_i^2,$$

respectively, where $\mu_i = E(y_i)$ under H_0 and $g(\mu_i) = x_i^{\mathrm{T}}\alpha$. Let

$$W_o = -\partial^2 l(\alpha; 0)/\partial \eta\, \partial \eta^{\mathrm{T}} = \operatorname{diag}\{w_{oi}\} = \operatorname{diag}\{w_i + e_i(y_i - \mu_i)\},$$

where $e_i = \{V'(\mu_i)g'(\mu_i) + V(\mu_i)g''(\mu_i)\}/[V^2(\mu_i)\{g'(\mu_i)\}^3]$ and is zero for canonical links. Note that, under H_0, the matrix W is the generalised linear model working-weight matrix and $W = E(W_o)$. From the log-quasilikelihood expansion (7), some calculations give the efficient score $U_\theta(\hat{\alpha}_0) = \{U_{\theta_1}(\hat{\alpha}_0), \ldots, U_{\theta_m}(\hat{\alpha}_0)\}^{\mathrm{T}}$ for θ at $\theta = 0$ as follows:

$$U_{\theta_j}(\hat{\alpha}_0) = \left.\frac{\partial l(\alpha, \theta)}{\partial \theta_j}\right|_{\theta = 0, \alpha = \hat{\alpha}_0}$$
$$= \frac{1}{2}\operatorname{tr}[\{W\Delta^{-1}(y - \mu)(y - \mu)^{\mathrm{T}}\Delta^{-1}W - W_o\}Z\dot{D}_j Z^{\mathrm{T}}]$$
$$= \frac{1}{2}\{(y - \mu)^{\mathrm{T}}\Delta^{-1}WZ\dot{D}_j Z^{\mathrm{T}} W\Delta^{-1}(y - \mu) - \operatorname{tr}(W_o Z\dot{D}_j Z^{\mathrm{T}})\}, \quad (8)$$

where $\mu=(\mu_1,\ldots,\mu_n)^{\mathrm{T}}$, $\dot D_j=\partial D/\partial\theta_j|_{\theta=0}$ $(j=1,\ldots,m)$, and $\hat\alpha_0$, the maximum likelihood estimator of α under $\theta=0$, can be easily obtained by fitting the generalised linear model $g(\mu_i)=x_i^{\mathrm{T}}\alpha$ to the observations y_i. Note that the score U_θ in (8) compares the weighted actual and nominal observed covariance of y.

For the clustered model (5), the score $U_\theta(\hat\alpha_0)$ can be written as

$$U_{\theta_j}(\hat\alpha_0)=\frac{1}{2}\sum_{k=1}^{K}\{(\mathcal{Y}_k-\mu_k^{b_k})^{\mathrm{T}}\Delta_k^{-1}W_kZ_k\dot D_j^kZ_k^{\mathrm{T}}W_k\Delta_k^{-1}(\mathcal{Y}_k-\mu_k^{b_k})-\mathrm{tr}(W_{ok}Z_k\dot D_j^kZ_k^{\mathrm{T}})\},\quad (9)$$

where $g(\mu_k^{b_k})=X_k\alpha$, $\dot D_j^k$ is the kth block of the block diagonal matrix $\dot D_j$, and the matrices Δ_k, W_k and W_{ok} are diagonal matrices with elements δ_{kr}, w_{kr} and w_{okr} $(r=1,\ldots,n_k)$. Note that the subscript kr denotes the rth observation of the kth cluster. When only a random intercept is specified for each cluster, that is $m=1$, the score statistic $U_\theta(\hat\alpha_0)$ in (9) reduces to

$$U_\theta(\hat\alpha_0)=\frac{1}{2}\sum_{k=1}^{K}\left[\left\{\sum_{r=1}^{n_k}w_{kr}\delta_{kr}^{-1}(y_{kr}-\mu_{kr})\right\}^2-\sum_{r=1}^{n_k}w_{okr}\right].\quad (10)$$

Note that (10) corresponds to the overdispersion test statistic of Cox (1983) and Dean (1992), and the homogeneity test of Liang (1987) and Jacqmin-Gadda & Commenges (1995). For the hierarchical and crossed random effect model (6), the score statistic $U_\theta(\hat\alpha_0)$ in (8) takes the form

$$U_{\theta_j}(\hat\alpha_0)=\frac{1}{2}\{(y-\mu)^{\mathrm{T}}\Delta^{-1}WZ_jZ_j^{\mathrm{T}}W\Delta^{-1}(y-\mu)-\mathrm{tr}(W_oZ_jZ_j^{\mathrm{T}})\}.\quad (11)$$

To test for $H_0:\theta=0$, we construct a global score statistic as follows:

$$\chi_G^2=U_\theta(\hat\alpha_0)^{\mathrm{T}}\tilde I(\hat\alpha_0)^{-1}U_\theta(\hat\alpha_0),\quad (12)$$

where $\tilde I$ is the efficient information matrix of θ evaluated under H_0, taking the form

$$\tilde I=I_{\theta\theta}-I_{\alpha\theta}^{\mathrm{T}}I_{\alpha\alpha}^{-1}I_{\alpha\theta}.$$

Here,

$$I_{\theta\theta}=E\left(\frac{\partial l}{\partial\theta}\frac{\partial l}{\partial\theta^{\mathrm{T}}}\right),\quad I_{\alpha\theta}=E\left(\frac{\partial l}{\partial\alpha}\frac{\partial l}{\partial\theta^{\mathrm{T}}}\right),\quad I_{\alpha\alpha}=E\left(\frac{\partial l}{\partial\alpha}\frac{\partial l}{\partial\alpha^{\mathrm{T}}}\right),$$

where $l=l(\alpha,\theta)$, and the scores $\partial l/\partial\theta$, $\partial l/\partial\alpha$ and the expectations are all calculated at $\theta=0$. This score statistic χ_G^2 is closely related to White's (1982) information matrix test. Note that the scale parameter ϕ is assumed to be known in deriving (12). It is straightforward to extend (12) to the cases where ϕ needs to be estimated.

Denote by κ_{ri} the rth cumulant of y_i. To obtain the efficient information matrix $\tilde I$, we assume that, under H_0, the third and fourth cumulants of y_i are related to the second cumulant via $\kappa_{(r+1)i}=\kappa_{2i}\,\partial\kappa_{ri}/\partial\mu_i$ $(r=2,3)$, where $\kappa_{2i}=\phi a_i^{-1}v(\mu_i)$. Note that this equality is a property of exponential-family distributions (McCullagh & Nelder, 1989, p. 350). It follows that (McCullagh & Nelder, 1989, p. 361)

$$\kappa_{3i}=(\phi a_i^{-1})^2v'(\mu_i)v(\mu_i),\quad \kappa_{4i}=(\phi a_i^{-1})^3[v''(\mu_i)v(\mu_i)+\{v'(\mu_i)\}^2]v(\mu_i).$$

Now let $A_j=Z\dot D_jZ^{\mathrm{T}}=\{a_{ii'}^j\}$ and let a^j be an $n\times1$ vector with elements a_{ii}^j $(j=1,\ldots,m)$. Let R be an $n\times n$ matrix with diagonal elements

$$r_{ii}=w_i^4\delta_i^{-4}\kappa_{4i}+2w_i^2+e_i^2\kappa_{2i}-2w_i^2\delta_i^{-2}e_i\kappa_{3i}$$

and off-diagonal elements $r_{ii'} = 2w_i w_{i'}$ $(i \neq i')$. Note that the last two terms in r_{ii} diminish for canonical links. Denote by C a diagonal matrix with $c_i = w_i^3 \delta_i^{-3} \kappa_{3i} - w_i \delta_i^{-1} e_i \kappa_{2i}$ on the diagonal. The c_i can be simplified as $w_i^3 \delta_i^{-3} \kappa_{3i}$ when canonical links are used. After some calculations, we obtain

$$I_{\theta_j \theta_k} = \frac{1}{4} J^{\mathrm{T}} (A_j \cdot R \cdot A_k) J = \frac{1}{4} \left(\sum_{i=1}^{n} a_{ii}^j a_{ii}^k r_{ii} + 2 \sum_{i<i'} a_{ii'}^j a_{ii'}^k r_{ii'} \right), \tag{13}$$

$$I_{\alpha \theta_j} = \frac{1}{2} X^{\mathrm{T}} C \alpha^j = \frac{1}{2} \sum_{i=1}^{n} a_{ii}^j c_i x_i, \tag{14}$$

$$I_{\alpha\alpha} = X^{\mathrm{T}} W X = \sum_{i=1}^{n} w_i x_i x_i^{\mathrm{T}}, \tag{15}$$

where $G \cdot H$ denotes the component-wise multiplication of conformable matrices G and H, and J denotes a vector of ones. We supply the detailed derivation of (13) in Appendix 1. The derivations of (14) and (15) are similar and are omitted.

One important feature of the proposed global variance component test statistic χ_G^2 is that a detailed specification of the distribution function $F(b; \theta)$ of the random effects is not necessary. The test is therefore robust in the special sense that it is against arbitrary mixed model alternatives where only the first two moments of the random effects are specified. The following gives the asymptotic properties of the proposed global score test. Note that 'asymptotic' in Proposition 1 refers to the number of clusters $K \to \infty$ with cluster sizes n_k bounded under the clustered model (5), and $n, q_j \to \infty$ $(j = 1, \ldots, m)$ with the number of observations at any level of any random factor bounded under the hierarchical and crossed model (6). Under these assumptions, we will simply describe the asymptotics as $n \to \infty$.

PROPOSITION 1. *Under the clustered model* (5) *and the hierarchical and crossed model* (6) *and some regularity conditions, we have the following.*

(i) *The global score statistic χ_G^2 follows a chi-squared distribution with m degrees of freedom asymptotically under $\theta = 0$.*

(ii) *The global score test based on χ_G^2 is a locally asymptotically most powerful test if $m = 1$, and is a locally asymptotically most stringent test if $m > 1$* (*Bhat & Nagnur,* 1965).

Proof. (i) For the clustered model (5), the regularity conditions and the proof are straightforward extensions of those in Moran (1971) and are hence omitted. Note that an additional regularity condition we need to assume for model (5) is that the number of observations per cluster n_k is bounded as the number of clusters $K \to \infty$. For the hierarchical and crossed model (6), the score statistic U_θ may not be written as a sum of independent random variables. Hence the regularity conditions and the proof are much more complicated and are given in Appendix 2.

(ii) For $m = 1$, see Liang (1987). For $m > 1$, see Bhat & Nagnur (1965). Note that a locally asymptotically most powerful test does not exist when $m > 1$ (Bhat & Nagnur, 1965). □

One of the regularity conditions, Condition 2, in Appendix 2 for the hierarchical and crossed model (6) is that the observations at any level of any random factor are bounded by a constant E as the sample size n and the number of levels (q_j) of each random factor b_j $(j = 1, \ldots, m)$ increase to infinity (Hartley & Rao, 1967). This assumption allows us to use the same constant $n^{-\frac{1}{2}}$ to normalise each component of the score statistic U_θ in (11) and the same constant $1/n$ to normalise each component of the information matrix I in (A2·1). For the salamander data discussed in detail in § 5, this assumption requires that the number of females and the number of males go to infinity with each animal mating with a finite number of animals of opposite sex.

Proposition 1 may not be generalised to fully crossed designs, where the number of the observations at a given level of one factor is proportional to the number of levels of another factor, as in two-way balanced random effect models with both levels going to infinity (Gumpertz & Pantula, 1992). This is because we may not be able to find a $(p+m)\times 1$ vector of constants t_n, which satisfies $t_n \to \infty$ as $n, q_j \to \infty$, in order to normalise the information matrix I in (A2·1) under H_0. In other words, there may not exist a $(p+m)\times 1$ diagonal matrix T_n with t_n on the diagonal such that $T_n^{-1/2} I T_n^{-1/2}$ converges to a positive matrix I^0 under H_0 as $n, q_j \to \infty$ (Gumpertz & Pantula, 1992). This follows from the observation that $I_{\alpha\theta}$ in (14) and $I_{\alpha\alpha}$ in (15) are sums of n nonzero terms, but the number of nonzero terms used to calculate $I_{\theta\theta}$ in (13) is of higher order in n; it is $O(n^{1+\tau})$ for some positive constant τ. Further research is needed to study the asymptotic distribution of χ^2_G when a design is fully crossed.

A correction may be made to equation (8) to adjust for small-sample bias associated with the estimation of α. If we use the approximation $E\{y_i - \mu_i(\hat{\alpha}_0)\}^2 \simeq (1-h_i)V(\mu_i)$ (Williams, 1987), where h_i is the ith diagonal element of the hat matrix $H = W^{\frac{1}{2}} X(X^{\mathrm{T}} W X)^{-1} X^{\mathrm{T}} W^{\frac{1}{2}}$, the bias-corrected version of the score statistic (8) is

$$
\begin{aligned}
U_{C\theta_j}(\hat{\alpha}_0) &= \frac{1}{2}\operatorname{tr}[\{W\Delta^{-1}(y-\mu)(y-\mu)^{\mathrm{T}}\Delta^{-1}W - \tilde{W}_o\} Z\dot{D}_j Z^{\mathrm{T}}], \\
&= \frac{1}{2}\{(y-\mu)^{\mathrm{T}}\Delta^{-1} W Z \dot{D}_j Z^{\mathrm{T}} W \Delta^{-1}(y-\mu) - \operatorname{tr}(\tilde{W}_o Z \dot{D}_j Z^{\mathrm{T}})\},
\end{aligned} \tag{16}
$$

where $j = 1, \ldots, m$ and $\tilde{W}_o = \operatorname{diag}\{(1-h_i)w_i + e_i(y_i - \mu_i)\}$. Equation (12) may then be modified as $\chi^2_C = U^{\mathrm{T}}_{C\theta}\tilde{I}^{-1}U_{C\theta}$. We examine the performance of the proposed correction by means of simulation in § 5. A similar correction was employed by Dean & Lawless (1989) and Dean (1992) for simple overdispersed Poisson and binomial models. Note that the hat matrix H also plays an important role in generalised linear models diagnostics (Williams, 1987).

4. The individual variance component tests

A disadvantage of the global variance component test (12) is that it does not indicate specific departures from the null hypothesis. For independent data, Smith & Heitjan (1993) suggested the following statistic for testing $H_0: \theta_j = 0$ in overdispersed generalised linear models:

$$
S_{\theta_j}(\hat{\alpha}_0) = U_{\theta_j}(\hat{\alpha}_0)/\{\tilde{I}_{jj}(\hat{\alpha}_0)\}^{\frac{1}{2}}, \tag{17}
$$

where $\tilde{I}_{jj}$ $(j = 1, \ldots, m)$ is the jth diagonal element of $\tilde{I}$. Under the null hypothesis that all components of θ are zero $(\theta = 0)$ and given the regularity conditions discussed in § 3·2, the statistic S_{θ_j} asymptotically has a standard normal distribution. However, this may not hold under the composite null hypothesis that $\theta_j = 0$ and the other θ_k $(k \neq j)$ are unspecified. Note that, under the independent random effect model (6), which means that the random effects are independent, S_{θ_j} is in fact identical to the score statistic for testing

$\theta_j = 0$ if one fits a misspecified random effect model including only the jth random effect b_j: $g(\mu^{b_j}) = X\alpha + Z_j b_j$.

In this section, we study the score statistic for testing the composite null hypothesis $H_0: \theta_j = 0$. For simplicity, we restrict our attention to the independent random effect model (6) and hence the alternative hypothesis is one-sided, $H_a: \theta_j > 0$. We further assume that the random effects b_j $(j = 1, \ldots, m)$ are normally distributed, which allows us to use the existing maximum likelihood method or its approximations (Breslow & Lin, 1995; Zeger & Karim, 1991) to estimate the unknown parameters α and $\theta_{-j} = (\theta_1, \ldots, \theta_{j-1}, \theta_{j+1}, \ldots, \theta_m)^{\mathrm{T}}$ under H_0.

Rewrite $l(\alpha, \theta)$ as

$$l(\alpha, \theta) = \ln \int \exp\{l(y;\, b_j) + l(b_j)\}\, db_j, \tag{18}$$

where

$$l(y;\, b_j) = \ln \int \exp\left\{\sum_{i=1}^{n} l_i(y;\, b) + l(b_{-j})\right\} db_{-j},$$

$b_{-j}^{\mathrm{T}} = (b_1^{\mathrm{T}}, \ldots, b_{j-1}^{\mathrm{T}}, b_{j+1}^{\mathrm{T}}, \ldots, b_m^{\mathrm{T}})$, $l_i(y;\, b)$ was given in equation (4), and $l(b_j)$ and $l(b_{-j})$ are the normal log-likelihood functions for b_j and b_{-j}. Taking a Laplace expansion of (18) about $\theta_j = 0$, similar to that used in § 3·1, one can show that the efficient score for testing $H_0: \theta_j = 0$ is

$$\begin{aligned}\mathscr{U}_{\theta_j}(\hat\alpha, \hat\theta_{-j}) &= \left.\frac{\partial l(\alpha, \theta)}{\partial \theta_j}\right|_{\theta_j = 0, \hat\alpha, \hat\theta_{-j}} \\ &= \frac{1}{2} E[\{(y - \mu^{b_{-j}})^{\mathrm{T}} (\Delta^{b_{-j}})^{-1} W^{b_{-j}} Z_j Z_j^{\mathrm{T}} W^{b_{-j}} (\Delta^{b_{-j}})^{-1} (y - \mu^{b_{-j}}) \\ &\quad - \operatorname{tr}(Z_j^{\mathrm{T}} W_o^{b_{-j}} Z_j)\} \mid y],\end{aligned} \tag{19}$$

where $\mu^{b_{-j}} = E(y \mid b_{-j})$ and the expectations are taken under the null hypothesis $\theta_j = 0$. Note that $(\hat\alpha, \hat\theta_{-j})$ is the maximum likelihood estimator of (α, θ_{-j}) under the null hypothesis $\theta_j = 0$ and can be obtained by fitting the null generalised linear mixed model

$$g(\mu^{b_{-j}}) = X\alpha + \sum_{k \neq j} Z_k b_k. \tag{20}$$

The terms $W^{b_{-j}}$, $W_o^{b_{-j}}$, $\Delta^{b_{-j}}$ in equation (19) take the same form as W, W_o, Δ in § 3·2 except that $\mu^{b_{-j}}$ is used in place of μ^b.

Let $\gamma^{\mathrm{T}} = (\alpha^{\mathrm{T}}, \theta_{-j}^{\mathrm{T}})$ and let $\hat\gamma$ be the maximum likelihood estimator of γ under the null model (20). The score test for $\theta_j = 0$ then takes the form

$$\mathscr{S}_{\theta_j}(\hat\gamma) = \mathscr{U}_{\theta_j}(\hat\gamma) / \{\tilde{\mathscr{I}}_{jj}(\hat\gamma)\}^{\frac{1}{2}}, \tag{21}$$

where $\tilde{\mathscr{I}}_{jj}$ is the efficient information for θ_j and is given by

$$\tilde{\mathscr{I}}_{jj} = \mathscr{I}_{\theta_j \theta_j} - \mathscr{I}_{\gamma\theta_j}^{\mathrm{T}} \mathscr{I}_{\gamma\gamma}^{-1} \mathscr{I}_{\gamma\theta_j}. \tag{22}$$

Here,

$$\mathscr{I}_{\theta_j\theta_j} = E\left(\frac{\partial l}{\partial \theta_j}\frac{\partial l}{\partial \theta_j}\right), \quad \mathscr{I}_{\gamma\theta_j} = E\left(\frac{\partial l}{\partial \gamma}\frac{\partial l}{\partial \theta_j}\right), \quad \mathscr{I}_{\gamma\gamma} = E\left(\frac{\partial l}{\partial \gamma}\frac{\partial l}{\partial \gamma^{\mathrm{T}}}\right),$$

where $l = l(\alpha, \theta)$, and the scores $\partial l/\partial\theta_j$, $\partial l/\partial\gamma$ and the expectation are all calculated under $\theta_j = 0$. Under the regularity conditions discussed in § 3·2, using arguments similar to those used in § 3·2, one can show that $\mathscr{S}_{\theta_j}$ follows an asymptotically standard normal distribution under the null hypothesis $\theta_j = 0$, and is a locally asymptotically most powerful test. Note that, under the independent random effect model (6), the test based on $\mathscr{S}_{\theta_j}$ should be one-sided.

Except in the case of linear mixed models, $\mathscr{U}_{\theta_j}$ and $\tilde{\mathscr{I}}_{jj}$ often do not have closed form expressions and may involve multiple integrations. We therefore used the Laplace method to approximate $\mathscr{U}_{\theta_j}$ and $\tilde{\mathscr{I}}_{jj}$. We assume in the subsequent discussion that the null hypothesis $\theta_j = 0$ holds and denote $l_i(y; b)$ by $l(y; b_{-j})$.

Denote by $\hat{b}_{-j}$ the maximum point of $\sum l_i(y_i; b_{-j}) + l(b_{-j})$, where

$$l_i(y_i; b_{-j}) \propto \int \frac{a_i(y_i - u)}{\phi v(u)}\, du,$$

with the integral over $(y_i, \mu_i^{b_{-j}})$.

Let $Y = X\alpha + Z_{-j}\hat{b}_{-j} + (\Delta^{\hat{b}_{-j}})^{-1}(y - \mu^{\hat{b}_{-j}})$ be the working vector under the null generalised linear mixed model (20) (Breslow & Clayton, 1993), where

$$Z_{-j} = (Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_m).$$

Some calculations using the Laplace method show that $\mathscr{U}_{\theta_j}$ in (19) can be approximated by

$$\mathscr{U}^*_{\theta_j}(\hat{\alpha}, \hat{\theta}_{-j}) = \frac{1}{2}\{(Y - X\alpha)^{\mathrm{T}} V_j^{-1} Z_j Z_j^{\mathrm{T}} V_j^{-1}(Y - X\alpha) - \operatorname{tr}(Z_j^{\mathrm{T}} V_j^{-1} Z_j)\}\,|_{\hat{\alpha}, \hat{\theta}_{-j}}, \tag{23}$$

where $V_j = (W^{b_{-j}})^{-1} + \sum_{k \neq j} \theta_k Z_k Z_k^{\mathrm{T}}$; see Appendix 3 for details.

Equation (23) is in fact identical to the penalised quasilikelihood estimating equation for θ_j evaluated at $\theta_j = 0$ (Breslow & Clayton, 1993, eqn (14)). It is also equivalent to the score function of θ_j evaluated at $\theta_j = 0$ when one assumes that the working vector Y follows a linear mixed model $Y = X\alpha + Z_1 b_1 + \ldots + Z_m b_m + \varepsilon$, where ε follows a normal distribution with mean 0 and covariance $(W^b)^{-1}$. The efficient information $\tilde{\mathscr{I}}_{jj} = E[\{\mathscr{U}_{\theta_j}(\hat{\alpha}, \hat{\theta}_{-j})\}^2]$ can then be approximated by $\tilde{\mathscr{I}}^*_{jj} = E[\{\mathscr{U}^*_{\theta_j}(\hat{\alpha}, \hat{\theta}_{-j})\}^2]$ with

$$\tilde{\mathscr{I}}^*_{jj} = \mathscr{I}^*_{\theta_j\theta_j} - (\mathscr{I}^*_{\theta_{-j}\theta_j})^{\mathrm{T}}(\mathscr{I}^*_{\theta_{-j}\theta_{-j}})^{-1}\mathscr{I}^*_{\theta_{-j}\theta_j}, \tag{24}$$

where, if we define

$$\mathscr{I}^*_{\theta_k\theta_{k'}} = \frac{1}{2}\operatorname{tr}(Z_{k'}^{\mathrm{T}} V_j^{-1} Z_k Z_k^{\mathrm{T}} V_j^{-1} Z_{k'}) \quad (k, k' = 1, \ldots, m),$$

and $\mathscr{I}^*_{\theta_{-j}\theta_j}$ is an $(m-1)\times 1$ vector with elements $\mathscr{I}^*_{\theta_k\theta_j}$ $(k \neq j)$, then $\mathscr{I}^*_{\theta_{-j}\theta_{-j}}$ is an $(m-1)\times(m-1)$ matrix with elements $\mathscr{I}^*_{\theta_k\theta_{k'}}$ $(k, k' \neq j)$. Note that $\mathscr{I}^*_{\theta_k\theta_{k'}}$ resembles the information matrix equation (15) of Breslow & Clayton (1993).

To account for the loss of degrees of freedom from estimating α and to correct for small sample bias in $\mathscr{U}^*_{\theta_j}$, we may use the restricted maximum likelihood version of the approximate efficient score,

$$\mathscr{U}^*_{R\theta_j}(\hat{\alpha}, \hat{\theta}_{-j}) = \frac{1}{2}\{(Y - X\hat{\alpha})^{\mathrm{T}} V_j^{-1} Z_j Z_j^{\mathrm{T}} V_j^{-1}(Y - X\hat{\alpha}) - \operatorname{tr}(Z_j^{\mathrm{T}} P_j Z_j)\}, \tag{25}$$

where $P_j = V_j^{-1} - V_j^{-1}X(X^{\mathrm{T}} V_j^{-1} X)^{-1} X^{\mathrm{T}} V_j^{-1}$. The restricted maximum likelihood efficient information of θ_j is

$$\tilde{\mathscr{I}}^*_{Rjj} = \mathscr{I}^*_{R\theta_j\theta_j} - (\mathscr{I}^*_{R\theta_{-j}\theta_j})^{\mathrm{T}}(\mathscr{I}^*_{R\theta_{-j}\theta_{-j}})^{-1}(\mathscr{I}^*_{R\theta_{-j}\theta_j}), \quad (26)$$

where, if we define

$$\mathscr{I}^*_{R\theta_k\theta_{k'}} = \frac{1}{2}\,\mathrm{tr}(Z_{k'}^{\mathrm{T}} P_j Z_k Z_k^{\mathrm{T}} P_j Z_{k'}) \quad (k, k' = 1, \ldots, m),$$

$\mathscr{I}^*_{R\theta_{-j}\theta_j}$ and $\mathscr{I}^*_{R\theta_{-j}\theta_{-j}}$ are defined in the same way as $\mathscr{I}^*_{\theta_{-j}\theta_j}$ and $\mathscr{I}^*_{\theta_{-j}\theta_{-j}}$ except that $\mathscr{I}^*_{R\theta_k\theta_{k'}}$ is used in place of $\mathscr{I}^*_{\theta_k\theta_{k'}}$. The score statistic $\mathscr{S}_{\theta_j}$ for testing $\theta_j = 0$ can then be approximated by

$$\mathscr{S}^*_{\theta_j} = \mathscr{U}^*_{R\theta_j}/(\tilde{\mathscr{I}}^*_{Rjj})^{\frac{1}{2}} \quad (27)$$

evaluated at $(\hat{\alpha}, \hat{\theta}_{-j})$. We estimate $\hat{\alpha}$ and $\hat{\theta}_{-j}$ using the penalised quasilikelihood of Breslow & Clayton (1993). An advantage of using the penalised quasilikelihood estimators of (α, θ_{-j}) is that the working vector Y and the working weight matrix $W^{b_{-j}}$ used to calculate $\mathscr{S}^*_{\theta_j}$ are by-products of fitting the null model (20) using penalised quasilikelihood. As a consequence, the approximate score statistic $\mathscr{S}^*_{\theta_j}$ can be easily calculated using the existing SAS macro GLIMMIX (Wolfinger, 1995), which implements penalised quasilikelihood.

It is worth noting that, for linear mixed models, the approximation $\mathscr{S}^*_{\theta_j}$ is the exact score statistic for testing $\theta_j = 0$. For classical random effect analysis of variance models, such as the two-way random effect analysis of variance model, one can show that $\mathscr{U}^*_{R\theta_j}$ is closely related to the F-statistic for the test of $\theta_j = 0$. Specifically, apart from a constant factor, $\mathscr{U}^*_{R\theta_j}$ is the numerator of the F-statistic minus its mean. We evaluate the performance of the approximate individual score test $\mathscr{S}^*_{\theta_j}$ in § 6.

5. Application to the salamander mating data

McCullagh & Nelder (1989, § 14.5) presented an interesting dataset on salamander matings. The experiment involved two populations of salamanders, rough butt (RB) and whiteside (WS). Ten males and ten females from each population were mated in a crossed design, with six matings for each salamander. This resulted in 120 correlated binary observations. The experiment was repeated three times during the summer and autumn of 1986. The same 40 salamanders were used in the summer and the first autumn experiments. A new set of 40 salamanders was utilised in the second autumn study. The data for the three experiments are given in Tables 14.4–14.6 in McCullagh & Nelder (1989). The questions of interest are whether or not the four mating probabilities across the two salamander populations are the same and whether or not there is heterogeneity across males and females. We focus here on the second question. For a detailed discussion of the first question, see Breslow & Clayton (1993) and Drum & McCullagh (1993).

We analysed the data separately for each repetition and jointly for the three experiments combined. Denote by y_{ij} the binary outcome for the ith female and jth male. As in Drum & McCullagh (1993), the conditional logistic model for each experiment is

$$\mathrm{logit}\{E(y_{ij} | b_i^f, b_j^m)\} = x_{ij}^{\mathrm{T}}\alpha + b_i^f + b_j^m \quad (i, j = 1, \ldots, 20), \quad (28)$$

where b_i^f and b_j^m are random effects from the female and male individuals in the pair and are assumed to be independent variates with means 0 and variances σ_f^2 and σ_m^2, respectively. The covariate vector x_{ij} is set to be $(1, \mathrm{ws}_i^f, \mathrm{ws}_j^m, \mathrm{ws}_{ij}^{fm})^{\mathrm{T}}$, where ws_i^f is the indicator for whiteside female ($0 = $ RB, $1 = $ WS), ws_j^m is the indicator for whiteside male ($0 = $ RB, $1 = $ WS), and ws_{ij}^{fm} is their interaction.

To estimate the fixed effects and variance components jointly, several authors considered the logistic-normal model, where the random effects were assumed to be normally distributed. See Lin & Breslow (1996) for a survey.

To test for overall heterogeneity across males and females for each experiment, we set the null hypothesis $\sigma_f^2 = \sigma_m^2 = 0$. We here relax the normality assumption on the random effects for testing this global null hypothesis. Note that the regularity condition, Condition 2 in Appendix 2, was approximated for this dataset, since the design was partially crossed and each animal was only allowed to mate with six animals of the opposite sex. We then applied the approximate individual variance component tests discussed in § 4 to test for $\sigma_f^2 = 0$ and $\sigma_m^2 = 0$ separately. In the individual tests, we assumed that random effects were normally distributed and we estimated the unknown parameters under the null hypothesis using penalised quasilikelihood. Table 1 lists the global test statistics and the individual test statistics for each of the three experiments and the pooled data. In the analysis of the pooled data, we assumed that a new set of animals had been used in the first autumn experiment (Drum & McCullagh, 1993), and that the conditional logistic model (28) held for $i, j = 1, \ldots, 60$.

For the purpose of illustration, the first row of Table 1 gives the restricted maximum likelihood estimates of the variance components by Drum & McCullagh (1993), who assumed the random effects were normally distributed. The bias-corrected global score statistics were slightly larger than the uncorrected quantities. In the summer experiment, the individual variance component tests suggested that there was significant evidence of heterogeneity among females ($p = 0{\cdot}00$), but little evidence of heterogeneity among males ($p = 0{\cdot}41$). In the first autumn experiment, there was strong evidence of heterogeneity among both males ($p = 0{\cdot}03$) and females ($p = 0{\cdot}00$). In the second autumn experiment, significant heterogeneity was found among male salamanders ($p = 0{\cdot}00$), but there was little evidence of heterogeneity among females ($p = 0{\cdot}19$). For the pooled data, the large values of the test statistics allowed us to reject strongly the null hypothesis of homogeneity among males ($p = 0{\cdot}00$) and females ($p = 0{\cdot}00$).

These results agreed with the restricted maximum likelihood estimates of the variance components and the pseudo log-likelihood plots of Drum & McCullagh (1993). It should be noted that the Drum & McCullagh pseudo-log-likelihood plots are not the profile log-likelihood plots of the variance components, since these authors constructed their restricted log-likelihoods of θ by assuming that the binary y_{ij} were normally distributed. The simulation results of Lin & Breslow (1996) show that their restricted maximum likelihood

Table 1. *Test statistics for variance components in the salamander experiment*

	Summer		Fall 1		Fall 2		Pooled	
	σ_f^2	σ_m^2	σ_f^2	σ_m^2	σ_f^2	σ_m^2	σ_f^2	σ_m^2
REML	1·68	0·34	2·46	1·44	0·69	2·40	1·67	1·50
$\mathscr{S}^*_{\theta_j}$	2·64	0·22	3·21	1·92	0·89	4·10	3·57	3·38
(p-value)	0·00	0·41	0·00	0·03	0·19	0·00	0·00	0·00
χ^2_G (d.f. = 2)	17·68		11·33		16·92		40·99	
(p-value)	0·00		0·00		0·00		0·00	
χ^2_C (d.f. = 2)	18·98		12·40		18·05		42·21	
(p-value)	0·00		0·00		0·00		0·00	

REML, restricted maximum likelihood (Drum & McCullagh, 1993, Table 4).

estimates are almost unbiased, but are often associated with large variances. Hence the results based on their pseudo-log-likelihood plots may provide less evidence against $\sigma_f^2 = 0$ and $\sigma_m^2 = 0$ compared to the proposed score tests.

6. A simulation study

A simulation study was performed to assess the size and power of the proposed global score statistic χ_G^2, its bias-corrected counterpart χ_C^2 and the individual score statistic $\mathscr{S}_{\theta_j}^*$. The design considered in the simulation study was identical to that used in the actual salamander experiment, with female and male random effects generated from normal or mixed normal distributions with different values for the variance components. We assumed that the summer experiment was repeated three times and 10 different females and 10 different males were used each time.

To study the performance of the global variance component tests, χ_G^2 and χ_C^2, we generated 360 binary observations from model (28) by assuming that the random effects b_i^f and b_j^m $(i, j = 1, \ldots, 60)$ were independent random variates from distributions

$$F_f = F_m = F = \pi N\{-(1-\pi)\beta, \tau^2\} + (1-\pi)N\{\pi\beta, \tau^2\}$$

(Butler & Louis, 1992), and had means 0 and variances $\sigma_f^2 = \sigma_m^2 = \sigma^2 = \pi(1-\pi)\beta^2 + \tau^2$ (Johnson & Kotz, 1970, p. 89). We considered the following four cases of F.

Case 1. For the null, $\beta = \tau^2 = \sigma^2 = 0$.

Case 2. For the normal, $\beta = 0$, $\tau^2 = \sigma^2 = 0{\cdot}25, 0{\cdot}50, 0{\cdot}75, 1{\cdot}00$.

Case 3. For the unimodal normal mixture, $\pi = 0{\cdot}25$, $\beta = 0{\cdot}80, 1{\cdot}10, 1{\cdot}40, 1{\cdot}70$, $\tau^2 = \sigma^2 - \pi(1-\pi)\beta^2$, where $\sigma^2 = 0{\cdot}25, 0{\cdot}50, 0{\cdot}75, 1{\cdot}00$.

Case 4. For the bimodal normal mixture, $\pi = 0{\cdot}25$, $\beta = 1{\cdot}10, 1{\cdot}40, 1{\cdot}70, 2{\cdot}00$, $\tau^2 = \sigma^2 - \pi(1-\pi)\beta^2$, where $\sigma^2 = 0{\cdot}25, 0{\cdot}50, 0{\cdot}75, 1{\cdot}00$.

Note that the selected value of $\pi = 0{\cdot}25$ in Cases (3) and (4) produces moderately skewed distributions. The values of β in Case 3 were chosen in such a way that the means of the mixtures have almost the widest separations that yield unimodal distributions. We set the true values of the fixed effects to be $\alpha = (1{\cdot}06, -3{\cdot}05, -0{\cdot}72, 3{\cdot}77)^T$, which correspond to the restricted maximum likelihood estimates (Drum & McCullagh, 1993; Lin & Breslow, 1996) fitted to the actual data.

To study the performance of the individual variance component tests based on $\mathscr{S}_{\sigma_f^2}^*$ and $\mathscr{S}_{\sigma_m^2}^*$, we assumed that the random effects b_i^f and b_j^m $(i, j = 1, \ldots, 60)$ were normally distributed. As a result of the balanced design, we only need to study the size and power of $\mathscr{S}_{\sigma_f^2}^*$ for different values of σ_f^2 and a fixed value of σ_m^2. For simplicity, we fixed σ_m^2 to be $0{\cdot}5$ and chose σ_f^2 from $0{\cdot}00, 0{\cdot}25, 0{\cdot}50, 0{\cdot}75, 1{\cdot}00$. Breslow & Lin (1995) and Lin & Breslow (1996) show that point estimation using the normal-based Laplace approximation in generalised linear mixed models may not perform well when the data are binary and that its performance improves as the binomial denominator increases. We suspect that the same may be true when the Laplace approximation is applied to the score statistic $\mathscr{U}_{\theta_j}$. We therefore replicated the simulation by generating 360 binomial observations y_{ij} from model (28) whose denominators were $N = 1, 2, 4, 8$.

Table 2 presents empirical sizes and powers for the global test statistic χ_G^2 and its bias-corrected version χ_C^2 for binary data $(N = 1)$ based on 2000 simulations. The nominal sizes

Table 2. *Empirical sizes and powers of global variance component tests observed in* 2000 *simulations*

Distribution of (b_f, b_m)		(σ_f^2, σ_m^2) (0·00, 0·00)	(0·25, 0·25)	(0·50, 0·50)	(0·75, 0·75)	(1·00, 1·00)
Normal	χ_G^2	0·053	0·284	0·675	0·893	0·969
	χ_C^2	0·051	0·306	0·698	0·903	0·973
Unimodal normal mixture	χ_G^2	0·053	0·308	0·706	0·913	0·979
	χ_C^2	0·051	0·335	0·727	0·922	0·983
Bimodal normal mixture	χ_G^2	0·053	0·321	0·729	0·922	0·981
	χ_C^2	0·051	0·340	0·749	0·935	0·984

of the tests were set to be 0·05. The results for testing $\sigma_f^2 = \sigma_m^2 = 0$ showed that the actual sizes of the tests were very close to 0·05. The bias-corrected test statistic χ_C^2 slightly outperformed the uncorrected quantity χ_G^2 in terms of size and power. Greater powers were associated with the distributions which deviated further from normality. As the variance components increased, the powers of the tests approached 1 quickly.

Table 3 shows the empirical sizes and powers for the individual test statistics $\mathscr{S}^*_{\sigma_f^2}$ based on 2000 simulations. When the data were binary ($N = 1$), the size of the Laplace-based approximate score statistic $\mathscr{S}^*_{\sigma_f^2}$ was 0·026 at the 0·05 level and hence the test was too conservative. To see if the unsatisfactory performance of $\mathscr{S}^*_{\sigma_f^2}$ could be improved by using less biased estimates of α and σ_m^2 under the null hypothesis $\sigma_f^2 = 0$, we used the corrected penalised quasilikelihood estimates of α and σ_m^2 proposed by Lin & Breslow (1996). The improvement was minimal and hence the results are not shown. This suggests that the Laplace approximation to the score statistic (19) does not work well when the data are highly sparse, as in the binary case. As the binomial denominator increased, the performance of the approximation quickly improved in terms of size and power. As the binomial denominator became $N = 8$, the approximate score test $\mathscr{S}^*_{\sigma_f^2}$ did an excellent job of achieving the nominal size at the 0·05 level. These results agree with the theoretical and numerical results of Breslow & Clayton (1993) and Breslow & Lin (1995) on the performance of the point estimators using the Laplace approximation.

Table 3. *Empirical sizes and powers of individual variance component test* $\mathscr{S}^*_{\sigma_f^2}$ *given* $\sigma_m^2 = 0{\cdot}50$, *observed in* 2000 *simulations*; *N*, *binomial denominator*

N	$\sigma_f^2 = 0{\cdot}000$	$\sigma_f^2 = 0{\cdot}250$	$\sigma_f^2 = 0{\cdot}500$	$\sigma_f^2 = 0{\cdot}750$	$\sigma_f^2 = 1{\cdot}000$
1	0·026	0·201	0·504	0·756	0·892
2	0·029	0·610	0·940	0·997	1·000
4	0·038	0·951	1·000	1·000	1·000
8	0·049	1·000	1·000	1·000	1·000

7. Discussion

We have proposed simple global variance component tests, which are locally asymptotically most stringent tests and are robust in the sense that no assumption about the parametric form of the random effects is made. The global tests have closed form expressions and only require fitting conventional generalised linear models. Our simulation results demonstrate that the test statistics perform very well when the number of levels of each random effect (q_j), which represents the amount of available information on estimating its corresponding variance component, is moderate or large. When the number of levels of each random effect is small, e.g. less than 15, we suspect that the critical values of the global test statistics, which are based on large sample theory, may be less accurate (Dean, 1992). Small sample approximations, such as the saddlepoint approximation or the Edgeworth approximation (McCullagh & Nelder, 1989), may be employed to create better critical values.

The individual variance component tests often do not have closed form expressions. The performance of the Laplace-based approximate individual tests is unsatisfactory when the data are binary. As the binomial denominator increases, the Laplace approximation quickly improves, as therefore does the approximate individual score test. Following Breslow & Lin (1995) and Lin & Breslow (1996), further research is needed to study the asymptotic bias of this approximation. It is also of potential interest to evaluate the efficient score statistic (19) using Monte Carlo simulation methods such as importance sampling and Metropolis sampling.

An alternative approach to testing for global overdispersion and heterogeneity is to use the Wald test or the likelihood ratio test estimating θ by specifying a form for F, such as Gaussian. However, estimation of α and θ is often greatly hampered by intractable numerical integrations (Breslow & Clayton, 1993). Since the null hypothesis $\theta=0$ places θ on the boundary of the parameter space, the resultant Wald statistic and likelihood ratio statistic follow a mixed chi-squared distribution, rather than a chi-squared distribution, under some regularity conditions (Self & Liang, 1987). Similar arguments apply to the individual variance components.

The variance component tests discussed herein are designed to be effective against one type of overdispersion and heterogeneity, for which the random effects exert their influence on the same scale as the fixed effects. Other tests may be more effective when the extra variation enters in a different fashion. For example, in beta-binomial and Poisson-gamma models, the random effects and fixed effects act multiplicatively. The forms of the score tests for overdispersion and heterogeneity may then be different (Dean, 1992). Detailed studies of the properties of the tests proposed in this paper when applied to such scenarios would be a worthy subject for further research.

The tests proposed in this paper can be extended to test for a subset of the variance components equal to zero. They can also be extended to test for variance components in nonlinear mixed models (Wolfinger, 1993) using a similar Laplace expansion approach, and to test for correlation parameters in generalised estimating equations (Liang & Zeger, 1986) using second-order estimating equations (Jacqmin-Gadda & Commenges, 1995).

Acknowledgement

The author gratefully acknowledges the Editor, Associate Editor and referees for their valuable comments that substantially improved the paper. The author thanks Professors Roderick Little and Edsel Pena for helpful discussion. This work was supported in part by the University of Michigan Rackham Grant and OVPAMA's Faculty Career Development Award Fund.

Appendix 1

Derivation of (13)

In what follows, we assume that the expectations are taken under $\theta=0$. Write $A_j = Z \cdot \dot{D}_j \cdot Z^{\mathrm{T}} = \{a_{ii'}^j\}$, where $Z \cdot \dot{D}_j$ defines the element-by-element multiplication of Z and $\dot{D}_j$.

From (8), the (j, k)th component $(j, k = 1, \ldots, m)$ of $I_{\theta\theta}$ is

$$\begin{aligned} I_{\theta_j\theta_k} &= E(U_{\theta_j} U_{\theta_k}) \\ &= \frac{1}{4} E\left(\left[\sum_{i=1}^{n} a_{ii}^j \{w_i^2 \delta_i^{-2}(y_i-\mu_i)^2 - w_i - e_i(y_i-\mu_i)\} + 2\sum_{i<i'} a_{ii'}^j w_i \delta_i^{-1} w_{i'} \delta_{i'}^{-1}(y_i-\mu_i)(y_{i'}-\mu_{i'})\right]\right. \\ &\quad \left.\times \left[\sum_{i=1}^{n} a_{ii}^k \{w_i^2 \delta_i^{-2}(y_i-\mu_i)^2 - w_i - e_i(y_i-\mu_i)\} + 2\sum_{i<i'} a_{ii'}^k w_i \delta_i^{-1} w_{i'} \delta_{i'}^{-1}(y_i-\mu_i)(y_{i'}-\mu_{i'})\right]\right) \\ &= \frac{1}{4}\left[\sum_{i=1}^{n} a_{ii}^j a_{ii}^k E\{w_i^2 \delta_i^{-2}(y_i-\mu_i)^2 - w_i - e_i(y_i-\mu_i)\}^2\right. \\ &\quad \left. + 4\sum_{i<i'} a_{ii'}^j a_{ii'}^k w_i^2 \delta_i^{-2} w_{i'}^2 \delta_{i'}^{-2} E\{(y_i-\mu_i)^2 (y_{i'}-\mu_{i'})^2\}\right]. \end{aligned}$$

If we use the identities $E(y_i-\mu_i)^4 = \kappa_{4i} + 3\kappa_{2i}^2$ (McCullagh & Nelder, 1989, p. 361) and $\kappa_{2i} = V(\mu_i)$, some calculations give

$$I_{\theta_j\theta_k} = \frac{1}{4}\left(\sum_{i=1}^{n} a_{ii}^j a_{ii}^k r_{ii} + 2\sum_{i<i'} a_{ii'}^j a_{ii'}^k r_{ii'}\right),$$

where $r_{ii} = w_i^4 \delta_i^{-4} \kappa_{4i} + 2w_i^2 + e_i^2 \kappa_{2i} - 2w_i^2 \delta_i^{-2} e_i \kappa_{3i}$ and $r_{ii'} = 2w_i w_{i'}$. Equation (13) follows immediately.

Appendix 2

Proof of the asymptotic distribution of χ_G^2

Here we study the asymptotic distribution of χ_G^2 under $H_0: \theta = 0$, i.e. independence, for the hierarchical and crossed model (6). Let

$$U = \begin{pmatrix} U_\alpha \\ U_\theta \end{pmatrix}, \quad I = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\theta} \\ I_{\alpha\theta}^{\mathrm{T}} & I_{\theta\theta} \end{pmatrix}, \quad I^0 = \begin{pmatrix} I_{\alpha\alpha}^0 & I_{\alpha\theta}^0 \\ I_{\alpha\theta}^{0\mathrm{T}} & I_{\theta\theta}^0 \end{pmatrix}, \tag{A2·1}$$

where

$$U_\alpha = X^{\mathrm{T}} W \Delta^{-1}(y-\mu) = \sum_{i=1}^{n} U_{\alpha,i} = \sum_{i=1}^{n} x_i w_i \delta_i^{-1}(y_i-\mu_i)$$

is the score vector for α and U_θ is given in (11). We first prove the asymptotic normality of $n^{-\frac{1}{2}}U(\alpha_0)$, where α_0 is the true value of α. Since the designs may be crossed, we may not be able to write the score vector $n^{-\frac{1}{2}}U(\alpha_0)$ as a sum of independent random vectors, so that the standard asymptotic theory for independent random variables may not be applied here.

We assume the following regularity conditions under independence $(\theta = 0)$, where the first two assumptions are similar to those given in Hartley & Rao (1967).

Condition 1. Each of the design matrices Z_j has only zeros and ones, with exactly one 1 in each row and at least one 1 in each column.

Condition 2. As $n, q_j \to \infty$ $(j = 1, \ldots, m)$, the number of observations at any level of any random factor is bounded by a constant E for all n. It follows that $q_j = O(n)$ $(j = 1, \ldots, m)$.

Condition 3. The log-quasilikelihood of α has the usual asymptotic properties, including consistency of $\hat{\alpha}_0$ and the linear expansion (White, 1982)

$$n^{\frac{1}{2}}(\hat{\alpha}_0 - \alpha_0) = \{nI_{\alpha\alpha}^{-1}(\alpha_0)\}\{n^{-\frac{1}{2}}U_\alpha(\alpha_0)\} + o_p(1).$$

Condition 4. There exists a positive definite matrix I^0, which has a partition as given in (A2·1), such that

$$\lim_{n, q_j \to \infty} n^{-1} I = I^0.$$

Note that this assumption is reasonable if Conditions 1 and 2 hold.

Condition 5. The sequences $\{w_i \delta_i^{-1}(y_i - \mu_i)\}$, $\{x_i\}$ and $\{w_{oi}\}$ are uniformly bounded for all $i = 1, \ldots, n$.

Condition 6. For any given $m \times 1$ constant vector λ_2, let

$$\Omega_{\lambda_2} = \sum_{j=1}^{m} \frac{1}{2} \lambda_{2j} \Delta^{-1} W Z_j Z_j^{\mathrm{T}} W \Delta^{-1}.$$

Then $y_{\lambda_2}^* = \Omega_{\lambda_2}(y - \mu) = (y_{\lambda_2,1}^*, \ldots, y_{\lambda_2,n}^*)^{\mathrm{T}}$ forms an M-dependent sequence for some constant M.

This assumption is reasonable given Conditions 1 and 2. It ensures that the experiment and the design matrices Z_j are constructed in such a way that the matrix Ω_{λ_2} is sufficiently sparse.

Proof of Proposition 1. For any given $(p+m) \times 1$ constant vector $\lambda = (\lambda_1^{\mathrm{T}}, \lambda_2^{\mathrm{T}})^{\mathrm{T}}$, where λ_1 is a $p \times 1$ vector and λ_2 is an $m \times 1$ vector, we have

$$\lambda^{\mathrm{T}} U = \lambda_1^{\mathrm{T}} U_\alpha + \lambda_2^{\mathrm{T}} U_\theta = \left(\sum_{i=1}^{n} \lambda_1^{\mathrm{T}} U_{\alpha,i} \right) + \left\{ (y-\mu)^{\mathrm{T}} \Omega_{\lambda_2} (y - \mu) - \mathrm{tr}\left(\sum_{j=1}^{m} \frac{1}{2} \lambda_{2j} Z_j Z_j^{\mathrm{T}} W_o \right) \right\}, \tag{A2·2}$$

where $\Omega_{\lambda_2} = \sum_{j=1}^{m} \frac{1}{2} \lambda_{2j} \Delta^{-1} W Z_j Z_j^{\mathrm{T}} W \Delta^{-1}$.

Let $y_{\lambda_2}^* = \Omega_{\lambda_2}(y - \mu) = (y_{\lambda_2,1}^*, \ldots, y_{\lambda_2,n}^*)^{\mathrm{T}}$. It follows from Conditions 1 and 2 that

$$\Gamma_{\lambda_2} = \mathrm{tr}\left(\sum_{j=1}^{m} \frac{1}{2} \lambda_{2j} Z_j Z_j^{\mathrm{T}} W_o \right) = \sum_{i=1}^{n} \Gamma_{\lambda_2,i},$$

where $\Gamma_{\lambda_2,i} = \frac{1}{2}(\sum_{j=1}^{m} \lambda_{2j}) w_{oi}$, and each $y_{\lambda_2,i}^*$ is a weighted sum of at most mE terms of the $(y_{i'} - \mu_{i'})$ $(i' = 1, \ldots, n)$. As a consequence, (A2·2) can be written as

$$\lambda^{\mathrm{T}} U = \sum_{i=1}^{n} U_{\lambda,i} = \sum_{i=1}^{n} \{\lambda_1^{\mathrm{T}} U_{\alpha,i} + (y_i - \mu_i) y_{\lambda_2,i}^* - \Gamma_{\lambda_2,i}\}.$$

From Condition 5, the sequences

$$\{x_i^{\mathrm{T}} w_i \delta_i^{-1}(y_i - \mu_i)\}, \quad \{w_i \delta_i^{-1} w_j \delta_j^{-1}(y_i - \mu_i)(y_j - \mu_j)\} \quad (i, j = 1, \ldots, n)$$

are uniformly bounded. Since the w_{oi} are also uniformly bounded under Condition 5, we have that the $U_{\lambda,i}$ $(i = 1, \ldots, n)$ are uniformly bounded for any given λ.

It can also easily be shown that $E(U_{\lambda,i}) = 0$ and that $\{U_{\lambda,i}\}$ is an M-dependent sequence under Condition 6. An application of Theorem 7.3.1 of Chung (1974) gives

$$n^{-\frac{1}{2}} \lambda^{\mathrm{T}} U(\alpha_0) \to N\{0, \lambda^{\mathrm{T}} I^0(\alpha_0) \lambda\} \tag{A2·3}$$

in distribution as $n, q_j \to \infty$. Note that we here have also used Condition 4 in the proof of (A2·3). Using the Cramer–Wald device, we have that $n^{-\frac{1}{2}} U(\alpha_0) \to N\{0, I^0(\alpha_0)\}$ in distribution.

We next take a linear expansion of the efficient score $U_\theta(\hat{\alpha}_0)$ about $\alpha = \alpha_0$ and use Condition 3. This gives

$$n^{-\frac{1}{2}} U_\theta(\hat{\alpha}_0) = n^{-\frac{1}{2}} U_\theta(\alpha_0) - \{n^{-1} I_{\alpha\theta}(\alpha_0)\}^{\mathrm{T}} \{n I_{\alpha\alpha}^{-1}(\alpha_0)\} \{n^{-\frac{1}{2}} U_\alpha(\alpha_0)\} + o_p(1).$$

It follows that $n^{-\frac{1}{2}} U_\theta(\hat{\alpha}_0) \to N\{0, \tilde{I}^0(\alpha_0)\}$ in distribution, where $\tilde{I}^0 = I_{\theta\theta}^0 - I_{\alpha\theta}^{0\mathrm{T}} (I_{\alpha\alpha}^0)^{-1} I_{\alpha\theta}^0$. Using the consistency of $\hat{\alpha}_0$ and Slutsky's theorem, we have that

$$\chi_G^2 = \{n^{-\frac{1}{2}} U_\theta(\hat{\alpha}_0)\}^{\mathrm{T}} \{n^{-1} \tilde{I}^{-1}(\hat{\alpha}_0)\} \{n^{-\frac{1}{2}} U_\theta(\hat{\alpha}_0)\}$$

converges in distribution to a chi-squared distribution with m degrees of freedom as $n, q_j \to \infty$.

Appendix 3

Derivation of equation (23)

In what follows, we assume that the null hypothesis $\theta_j = 0$ holds. If we let

$$\mathscr{U}_{\theta_j}^{b_{-j}} = \frac{1}{2}\{(y-\mu^{b_{-j}})^{\mathrm{T}}(\Delta^{b_{-j}})^{-1}W^{b_{-j}}Z_jZ_j^{\mathrm{T}}W^{b_{-j}}(\Delta^{b_{-j}})^{-1}(y-\mu^{b_{-j}}) - \mathrm{tr}(Z_j^{\mathrm{T}}W_o^{b_{-j}}Z_j)\},$$

we can write $\mathscr{U}_{\theta_j}$ in equation (19) as

$$\mathscr{U}_{\theta_j} = \frac{\int \mathscr{U}_{\theta_j}^{b_{-j}} \exp\{l(y;\, b_{-j}) + l(b_{-j})\}\, db_{-j}}{\int \exp\{l(y;\, b_{-j}) + l(b_{-j})\}\, db_{-j}}. \tag{A3·1}$$

Let $\hat{b}_{-j}$ be the maximum point of $h(y;\, b_{-j}) = l(y;\, b_{-j}) + l(b_{-j})$. Application of the Laplace approximation to the numerator and denominator of (A3·1) separately gives, apart from a constant,

$$\mathscr{U}_{\theta_j} \simeq |-h''(y;\, \hat{b}_{-j})|^{\frac{1}{2}} \int \mathscr{U}_{\theta_j}^{b_{-j}} \exp\left[-\frac{1}{2}(b_{-j}-\hat{b}_{-j})^{\mathrm{T}}\{-h''(y;\, \hat{b}_{-j})\}(b_{-j}-\hat{b}_{-j})\right] db_{-j}, \tag{A3·2}$$

where

$$-h''(y;\, \hat{b}_{-j}) \simeq \Lambda_{-j} = D_{-j}^{-1} + Z_{-j}^{\mathrm{T}} W(\mu^{\hat{b}_{-j}})Z_{-j}$$

and D_{-j} is a block diagonal matrix with the kth diagonal block $\theta_k I_{q_k}$ $(k = 1, \ldots, m,\, k \neq j)$. Denoting by

$$Y = X\alpha + Z_{-j}b_{-j} + (\Delta^{b_{-j}})^{-1}(y - \mu^{b_{-j}})$$

the working vector under the null generalised linear mixed model (20), we can write equation (A3·2) as

$$\begin{aligned}\mathscr{U}_{\theta_j} \simeq \frac{1}{2}E\{&(Y - X\alpha - Z_{-j}b_{-j})^{\mathrm{T}} W(\mu^{b_{-j}})Z_jZ_j^{\mathrm{T}}W(\mu^{b_{-j}})(Y - X\alpha - Z_{-j}b_{-j}) \\ &- \mathrm{tr}(Z_j^{\mathrm{T}}W_o^{b_{-j}}Z_j)\},\end{aligned} \tag{A3·3}$$

where the expectation is taken with respect to $b_{-j} \sim N(\hat{b}_{-j}, \Lambda_{-j}^{-1})$. If we approximate Y and $W(\mu^{b_{-j}})$ by their values at $\hat{b}_{-j}$, equation (A3·3) then becomes

$$\mathscr{U}_{\theta_j} \simeq \frac{1}{2}\{(Y - X\alpha)^{\mathrm{T}} V_j^{-1}Z_jZ_j^{\mathrm{T}}V_j^{-1}(Y - X\alpha) - \mathrm{tr}(Z_j^{\mathrm{T}}V_j^{-1}Z_j)\}, \tag{A3·4}$$

which is equation (23).

References

Bhat, B. R. & Nagnur, B. N. (1965). Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypothesis. *Biometrika* **53**, 459–68.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalised linear mixed models. *J. Am. Statist. Assoc.* **88**, 9–25.

Breslow, N. E. & Lin, X. (1995). Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* **82**, 81–91.

Butler, S. M. & Louis, T. A. (1992). Random effects models with non-parametric priors. *Statist. Med.* **11**, 1981–2000.

Chung, K. L. (1974). *A Course in Probability Theory*, 2nd ed. New York: Academic Press.

Cox, D. R. (1983). Some remarks on overdispersion. *Biometrika* **70**, 269–74.

Dean, C. (1992). Testing for overdispersion in Poisson and binomial regression models. *J. Am. Statist. Assoc.* **87**, 451–7.

Dean, C. & Lawless, J. F. (1989). Tests for detecting overdispersion in Poisson regression models. *J. Am. Statist. Assoc.* **84**, 467–72.

DRUM, M. & MCCULLAGH, P. (1993). REML estimation with exact covariance in the logistics mixed model. *Biometrics* **49**, 677–89.

GRAY, R. J. (1995). Tests for variation over groups in survival data. *J. Am. Statist. Assoc.* **90**, 198–203.

GUMPERTZ, M. L. & PANTULA, S. G. (1992). Nonlinear regression with variance components. *J. Am. Statist. Assoc.* **87**, 201–9.

HARTLEY, H. O. & RAO, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika* **54**, 93–108.

JACQMIN-GADDA, H. & COMMENGES, D. (1995). Tests of homogeneity for generalized linear models. *J. Am. Statist. Assoc.* **90**, 1237–46.

JOHNSON, N. L. & KOTZ, S. (1970). *Distributions in Statistics, Continuous Univariate Distributions—1.* Boston: Houghton-Mifflin.

LIANG, K. Y. (1987). A locally most powerfully test for homogeneity with many strata. *Biometrika* **74**, 259–64.

LIANG, K. Y. & ZEGER, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

LIN, X. & BRESLOW, N. E. (1996). Bias correction in generalized linear mixed models with multiple components of dispersion. *J. Am. Statist. Assoc.* **91**, 1007–16.

MCCULLAGH, P. & NELDER, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.

MORAN, P. A. (1971). Maximum likelihood estimation in non-standard conditions. *Proc. Camb. Phil. Soc.* **70**, 441–50.

SELF, S. & LIANG, K. Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Statist. Assoc.* **82**, 605–10.

SMITH, P. J. & HEITJAN, D. F. (1993). Testing and adjusting for departures from nominal dispersion in generalized linear models. *Appl. Statist.* **41**, 31–41.

SOLOMON, P. J. & COX, D. R. (1992). Nonlinear components of variance models. *Biometrika* **79**, 1–11.

WHITE, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica* **50**, 1–25.

WILLIAMS, D. A. (1987). Generalized linear model diagnostics using the deviance and single case deletions. *Appl. Statist.* **36**, 181–91.

WOLFINGER, R. (1993). Laplace's approximation for nonlinear mixed models. *Biometrika* **80**, 791–5.

WOLFINGER, R. (1995). *The GLIMMIX SAS Macro.* Cary NC: SAS Institute Inc.

ZEGER, S. L. & KARIM, M. R. (1991). Generalized linear models with random effects: A Gibbs sampling approach. *J. Am. Statist. Assoc.* **86**, 79–86.

[*Received December* 1995. *Revised December* 1996]