Biometrika (2001), **88**, 3, *pp*. 739–751

Printed in Great Britain

Order-restricted score tests for homogeneity in generalised linear and nonlinear mixed models

BY DANIEL B. HALL

Department of Statistics, University of Georgia, Athens, Georgia 30602, U.S.A.

dhall@stat.uga.edu

AND JENS T. PRÆSTGAARD

A. C. Nielsen, 150 N. Martingale Road, Schaumburg, Illinois 60173-2076, U.S.A.

jens.praestgaard@acnielsen.com

SUMMARY

Lin (1997), building on work by Solomon & Cox (1992) and Lin & Breslow (1996), introduced a score test for homogeneity in the generalised linear mixed model. In this paper we propose an improvement to Lin's test that exploits the fact that covariance parameters associated with random effects in the model are constrained so as to form a positive semidefinite covariance matrix. Therefore, an omnibus score test will have power against alternatives that are known never to occur. Our improvement works by concentrating the score test into directions given by the positive semidefiniteness constraint. We apply the same sort of modification to the bias-corrected version of her score test in the generalised linear mixed model context and, in a nonlinear mixed model context, to a score test analogous to hers. We show how to obtain the limiting distribution of our proposed test statistic. We also argue, via simulations and local alternative power calculations, that the restricted test has better power than the unrestricted one.

Some key words: Lagrange multiplier test; Nonlinear regression; Random effect; Order-restricted inference; Tangent cone; Variance component.

1. INTRODUCTION

Both generalised linear models and normal error nonlinear regression models have been extended by incorporating random effects into the expectation function of the model. Although these generalised linear mixed models and nonlinear mixed models have been enthusiastically received by both practitioners and researchers, substantial theoretical and practical challenges remain. A natural question is whether or not the inclusion of random effects and the accompanying, often-cumbersome mixed model methodology is necessary for any particular dataset. Lin (1997) has addressed this issue by proposing a score test for homogeneity in the generalised linear mixed model. One of the nice features of her test is that it does not require the fitting of the mixed model, because its approximations to the score function and information matrix are derived under the null hypothesis of homogeneity. As a consequence, the analyst can directly check the significance of the hypothesised covariance structure against an ordinary fixed-effect null model. If no significant improvement is obtained by incorporating random effects, for example to model the within-unit correlation, then there is no need for the analyst to invoke the full mixed model fitting machinery.

In both generalised linear mixed models and nonlinear mixed models, the data vector y is assumed to have density $\exp\{l(y; b, \alpha)\}$ conditional on b, an unobservable random vector. Here, α is an unknown vector of regression parameters, which typically are of primary interest. The random regression parameters b are assumed to follow some zero mean kernel distribution $F(b; \theta)$ whose variance we denote by $\mathrm{var}_\theta(b) = \Sigma(\theta)$, for θ varying in a parameter space Θ satisfying $\Sigma(0) = 0$. When we refer to the 'test for homogeneity' we mean the test of $H_0 : \Sigma(\theta) = 0$ against the alternative hypothesis specified for Σ.

Lin (1997) develops a score test for H_0 based on $T = S' I_e^{-1} S$, where I_e denotes the efficient information matrix of θ under H_0 and S denotes the derivative with respect to θ of the loglikelihood function evaluated at the point $(\alpha, \theta) = (\hat{\alpha}, 0)$, $\hat{\alpha}$ being the maximiser of $l(y; 0, \alpha)$. This test is an omnibus test, in the sense that it has power in all directions from $\theta = 0$. We argue that the power of the test can be improved by 'concentrating' it in those directions that are given a priori by the constraint that $\Sigma(\theta)$ be a valid covariance matrix. By simulations and theoretical considerations, it can be argued that a more powerful score test is based on $\tilde{T} = \tilde{S}' I_e^{-1} \tilde{S}$, where $\tilde{S}$ denotes the projection of S with respect to the weight matrix I_e^{-1} on to the set $\mathscr{M} = \{I_e t : t \in \mathscr{T}_C\}$. Here $\mathscr{T}_C$ denotes the tangent cone of the model, which is roughly the cone generated by the derivatives at $\theta = 0$ of all differentiable paths $\Sigma(\theta)$ through $\Sigma(0) = 0$.

Projected score tests have been considered previously by Gourieroux et al. (1982), Rogers (1986) and Silvapulle & Silvapulle (1995). This work is in the context of order-restricted inference about a vector of location parameters. In such problems, the constrained parameter set is a polyhedral cone in finite samples, whereas the positive semidefiniteness constraint only produces a polyhedral cone in the limit. Stram & Lee (1994) describe the asymptotic distribution of the likelihood ratio statistic for testing for zero variance components in the linear mixed model. These authors identify certain pairs of hypotheses $H_0 : \theta \in \Theta_0$, $H_1 : \theta \in \Theta_1$, $\Theta_0 \subset \Theta_1$ so that the likelihood ratio test of H_0 versus H_1 is asymptotically free of the true value θ_0. In each case the limiting distribution is more concentrated at 0 than is the standard chi-squared distribution with $\dim(\theta_1) - \dim(\theta_0)$ degrees of freedom. Their results would be valid also for nonlinear mixed models, but, for the overall test of homogeneity that uses $H_0 : \theta = 0$, the likelihood ratio test would not have a distribution that is free of θ_0 when $\dim(\theta) > 1$.

2. ORDER-RESTRICTED SCORE TESTS FOR HOMOGENEITY

For an $n \times 1$ data vector y following either a generalised linear or nonlinear mixed model as described in § 1, the likelihood function is

$$L(\alpha, \theta) = \int \exp\{l(y; b, \alpha)\}\, dF(b; \theta). \tag{1}$$

Following Lin (1997), we assume that $E_\theta(\|b\|^r) = o(\|\theta\|)$ as $\|\theta\| \to 0$ for all $r > 2$. Our set-up differs slightly from Lin's in that we specify a true density for the conditional distribution of y given b, and we allow the resulting model to be a nonlinear mixed model or a generalised linear mixed model. We expect that extension to the quasilikelihood case should be straightforward.

Lin's score test requires calculation of the score and information under the null hypothesis $\theta = 0$. We can do this without having to evaluate the integral (1), because the likelihood function admits the Laplace expansion (Solomon & Cox, 1992)

$$L(\alpha, \theta) = \exp\{l(y; 0, \alpha)\}\left\{1 + \frac{1}{2}\phi(\alpha, \theta) + o(\|\theta\|)\right\}, \tag{2}$$

where

$$\phi(\alpha, \theta) \equiv \sum_{i,j} \sigma_{ij}\left\{\frac{\partial^2}{\partial b_i\, \partial b_j} l(y; b, \alpha) + \frac{\partial l(y; b, \alpha)}{\partial b_i}\frac{\partial l(y; b, \alpha)}{\partial b_j}\right\}.$$

Here, σ_{ij} is the (i, j)th element of $\Sigma(\theta)$, the sum is over all components of b, and all derivatives with respect to components of b are evaluated at $b = 0$. Let $\gamma = (\alpha, \theta)$. The efficient score $S \equiv \partial \log L/\partial\gamma|_{\{\gamma=(\hat{\alpha},0)\}}$ can now be evaluated by differentiating the expansion (2). Also, the information I can be evaluated by differentiating twice in (2) and taking the expectation at $\gamma = (\hat{\alpha}, 0)$. If we partition the information matrix as

$$I = \begin{pmatrix} I_{\alpha\alpha} & I_{\alpha\theta} \\ I_{\theta\alpha} & I_{\theta\theta} \end{pmatrix},$$

and define the efficient information as $I_e = I_{\theta\theta} - I_{\theta\alpha} I_{\alpha\alpha}^{-1} I_{\alpha\theta}$, then Lin's test statistic is $T \equiv S' I_e^{-1} S$. Lin shows that, under some regularity conditions, T has a limiting chi-squared distribution with $\dim(\theta)$ degrees of freedom.

In the generalised linear mixed model context, formulae for S, $I_{\theta\theta}$, $I_{\alpha\theta}$ and $I_{\alpha\alpha}$ are given in equations (8), (13), (14) and (15), respectively, of Lin (1997). In the nonlinear mixed model context, we use the formulation of Wolfinger & Lin (1997). These authors give an approximate loglikelihood, and corresponding score vector and information matrix, which are based on a quadratic Taylor series expansion of $\exp\{l(y; b, \alpha)\}$ about $b = 0$ in (1). These authors ignore a term in $\partial^2 l(y; b, \alpha)/\partial b\, \partial b'|_{b=0}$ which involves the second derivative of the nonlinear regression function, $f(X, \alpha, Z, b)$, with respect to b. In addition, in the computation of the score and information matrix, they ignore the dependence of $\partial f/\partial b'|_{b=0}$ on α and θ, essentially treating these parameters orthogonally. In our experience, these approximations have resulted in only minor differences in the test statistics as compared to the results obtained when the omitted terms are retained. Therefore, for computational simplicity we utilise the Wolfinger & Lin formulae.

We propose a modification of Lin's score test in which we first project S on to $\mathscr{M} = I_e \mathscr{T}_C = \{I_e t : t \in \mathscr{T}_C\}$ before forming the test statistic. The projection is with respect to I_e^{-1}. Here, the tangent cone $\mathscr{T}_C$ of the model is defined as the convex closure of the set of limits of the form $\lim_{t\to 0}(\theta_t - 0)/t$, where $\{\theta_t, t \geqslant 0\}$ is a differentiable path in the parameter space for θ. For more on the concept of a tangent cone, see Geyer (1994), who follows Self & Liang (1987) and Chernoff (1954). The projected score statistic $\tilde{S}$ is the minimiser of $(S - \tilde{S})' I_e^{-1}(S - \tilde{S})$ over the set $\mathscr{M}$. Since the components of θ are restricted to form a positive semidefinite covariance matrix, $\mathscr{T}_C$ will not be a vector space of dimension $\dim(\theta)$. We propose to use the test statistic $\tilde{T} \equiv \tilde{S}' I_e^{-1} \tilde{S}$ to test the null hypothesis of homogeneity.

Just as for T, we obtain cut-off values for $\tilde{T}$ from its limiting distribution, which we describe in Theorem 1.

Theorem 1. *Let n be a measure of sample size so that, as $n \to \infty$, $n^{-1} I_e$ tends in probability to J_e, a positive definite, symmetric matrix, with Cholesky decomposition $J_e = DD'$. Let $U_1, \ldots, U_{\dim(\theta)}$ be independent, standard normal random variables, and let $\tilde{U}$ denote the orthogonal projection of $U = (U_1, \ldots, U_{\dim(\theta)})'$ on to the set $D'\mathscr{T}_C = \{D't : t \in \mathscr{T}_C\}$. Then, under some regularity conditions, it holds that* $\mathrm{pr}(\tilde{T} \leqslant x) \to \mathrm{pr}(\tilde{U}'\tilde{U} \leqslant x)$ *as $n \to \infty$.*

The distribution in Theorem 1 is a mixture of chi-squared distributions, all of which have degrees of freedom less than or equal to $\dim(\theta)$. The mixing weights depend on the parameterisation of $F(.;\,\theta)$ through the tangent cone $\mathscr{T}_C$ and on the limit J_e of the efficient information matrix through its Cholesky factor D. In practice, the limiting distribution must be approximated by substituting $n^{-1}I_e$ for J_e in all calculations; in § 3 we show how this is done.

Based on the simulation results presented in § 4, we are confident that the restricted test $\tilde{T}$ has better power than the unrestricted test T. We now give a local asymptotic power argument to the same effect. Let q_α and $\tilde{q}_\alpha$ denote the asymptotic cut-off values for a size $1-\alpha$ test of homogeneity based on, respectively, T and $\tilde{T}$. Assume that the nth stage parameter θ_n is not exactly zero, but satisfies $n^{\frac{1}{2}}\theta_n \to \delta$, where then by necessity $\delta \in \mathscr{T}_C$. Define $\mu = D'\delta$, and let $\mathrm{proj}(U+\mu|\mathscr{C})$ denote the orthogonal projection of $U+\mu$ on to the set $\mathscr{C} = D'\mathscr{T}_C$.

THEOREM 2. *Under suitable regularity conditions*, $\mathrm{pr}(T > q_\alpha) \to \mathrm{pr}\{\|U+\mu\|^2 > q_\alpha\}$ *while* $\mathrm{pr}(\tilde{T} > \tilde{q}_a) \to \mathrm{pr}\{\|\mathrm{proj}(U+\mu|\mathscr{C})\|^2 > \tilde{q}_\alpha\}$.

To put Theorem 2 in context, let $V \sim N(\mu, 1)$ and consider testing $H_0: \mu = 0$ versus either $H_1: \mu \in \mathscr{C}$ or $H_2: \mu \neq 0$, based on V. Then stating that

$$\mathrm{pr}\{\|U+\mu\|^2 > q_\alpha\} \leqslant \mathrm{pr}\{\|\mathrm{proj}(U+\mu|\mathscr{C})\|^2 > \tilde{q}_\alpha\}, \tag{3}$$

for all α and $\mu \in \mathscr{C}$, is equivalent to stating that the likelihood ratio test for H_1 is uniformly more powerful than the likelihood ratio test for H_2 if $\mu \in \mathscr{C}$. This result has been conjectured for general convex cones $\mathscr{C}$ and is believed to be true, but, as far as we are aware, a proof is only available in special cases. In the salamander data of § 3·1, the set $\mathscr{C}$ has a particular form, so that the conjecture is known to hold by Theorem 1 of Tsai (1992). This theorem is the most extensive such result of which we are aware.

Contrary to Lin's claim, in the cases where (3) holds, her test is not locally asymptotically most stringent, as defined in equation (1.3) of Bhat & Nagnur (1965). This is because the remark of these authors before their equation (3·6) is no longer valid when the tangent, Bhat & Nagnur's τ and our δ, is restricted to lie in a tangent cone $\mathscr{T}_C$ different from all of $\mathscr{R}^{\dim(\delta)}$, but such that $\mathscr{C}$ satisfies (3). Arguably, Bhat & Nagnur are not clear about the set in which their potential tangents are allowed to vary. In the text between their equations (1.1) and (1.2) they allow for an arbitrary set Γ, which is our $\mathscr{T}_C$; however, in deriving their equation (3.6) they assume that Γ is all of Euclidian space, $\mathscr{R}^{\dim(\delta)}$.

Lin (1997) proposes a modification of her score test statistic T that adjusts for the small-sample bias associated with the estimation of α. We will denote this modified test statistic by T^*. It arises by bias-correcting the score statistic S using the approximation $E\{y_i - \mu_i(\hat{\alpha})\} \simeq (1-h_i)\,\mathrm{var}(y_i)$, where h_i is the ith diagonal element of the generalised linear model 'hat' matrix (Williams, 1987). This $T^* = (S^*)'I_e^{-1}S^*$ can also be concentrated in directions given by the positive definiteness constraint, but the appropriate projection matrix is no longer I_e^{-1}. Instead we project S^* on to $\mathscr{T}_C$ with respect to $(I_e^*)^{-1}$, where I_e^* is a bias-corrected version of I_e. This I_e^* can be computed using the formulae given by Lin (1997) in equations (13), (14) and (15) of her paper, with modification in which $(1-h_i)w_i$ is used in place of w_i. The projected version of this bias-corrected test statistic is thus $\tilde{T}^* = (\tilde{S}^*)'(I_e^*)^{-1}\tilde{S}^*$, where $\tilde{S}^*$ is the minimiser of $(S^* - \tilde{S}^*)'(I_e^*)^{-1}(S^* - \tilde{S}^*)$ over the cone $\mathscr{T}_C$. We examine the performance of T^* and $\tilde{T}^*$ via simulation in § 4. We also examine an alternative bias-corrected version of T, $T^{**} \equiv S^*(I_e^*)^{-1}S^*$, where a bias correction is applied to the score statistic and information matrix. Note that T^{**}, rather than T^{*}, is more appropriately the non-projected version of $\tilde{T}^{*}$.

3. Examples

3·1. *Salamander data*

We first consider the famous salamander data of McCullagh & Nelder (1989, § 14.5). These data were also used by Lin (1997) who considered the model

$$\operatorname{logit}\{E(y_{ij} \mid b_i^{(f)}, b_j^{(m)})\} = x_{ij}'\alpha + b_i^{(f)} + b_j^{(m)} \quad (i = 1, \ldots, n_f, j = 1, \ldots, n_m), \tag{4}$$

where n_f and n_m are the numbers of female and male salamanders, respectively, involved in the experiment. We recall that Y_{ij} denotes the outcome of the mating of female i with male j; the covariate vector is $X_{ij} = (1, \textsc{wff}_i, \textsc{wsm}_j, \textsc{wsfm}_{ij})'$ including indicators for female and male population, 1 if whiteside and 0 otherwise, and an interaction term; and $b_i^{(f)}$, $b_j^{(m)}$ are independent subject-specific random effects so that $b_i^{(f)} \sim (0, \sigma_f^2)$, $b_i^{(m)} \sim (0, \sigma_m^2)$. We consider with Lin the score test for the hypothesis $H_0 : \theta = (\sigma_f^2, \sigma_m^2)' = 0$.

In this example, the 2×2 matrix $\Sigma(\theta)$ has a diagonal structure, with elements σ_f^2 and σ_m^2. The only constraint on θ is that both of its components be nonnegative, so the tangent cone becomes $\mathscr{T}_{\mathrm{C}} = [0, \infty) \times [0, \infty)$. The information matrix I can be calculated by the expansion (2), and it turns out to have off-diagonal blocks $I_{\alpha\theta} = I_{\theta\alpha}' = 0$. The efficient information matrix is therefore particularly simple: $I_e = I_{\theta\theta}$. Furthermore, $I_{\theta\theta}$ itself turns out to be diagonal, so the set on to which we must project is $\mathscr{M} = I_e \mathscr{T}_{\mathrm{C}} = \mathscr{T}_{\mathrm{C}}$. The projected score function $\tilde{S}$ is now the Euclidian projection of $S = (S_1, S_2)$ on to $[0, \infty)^2$; that is, $(\tilde{S}_1, \tilde{S}_2)' = (S_1 \vee 0, S_2 \vee 0)'$, the componentwise positive part of the score function. Denote the diagonal elements of I_e by i_{11}, i_{22}. Then the projected test statistic is

$$\tilde{T} = (S_1 \vee 0)^2 / i_{11} + (S_2 \vee 0)^2 / i_{22},$$

and, since $S_1 / i_{11}^{1/2}$ and $S_2 / i_{22}^{1/2}$ are asymptotically standard normal, we can ascertain directly that

$$\operatorname{pr}(\tilde{T} \leqslant x) \to \tfrac{1}{2} \operatorname{pr}\{\chi^2(1) \leqslant x\} + \tfrac{1}{4} \operatorname{pr}\{\chi^2(2) \leqslant x\} + \tfrac{1}{4}.$$

Table 1 contains test statistics for the three repetitions of the salamander experiment separately, with $n_f = n_m = 20$ in each case, as well as for the pooled data with $n_f = n_m = 60$. As a result of the large amount of heterogeneity exhibited in these data, both components of S are positive. Hence, projecting the score vector has no effect; that is, $\tilde{S} = S$. The same is true of the bias-corrected score vector S^{*}. However, the reference distributions for $\tilde{T}$

Table 1. *Test statistics for salamander data. All p-values are less than* 0·005

Test statistic	Summer	Autumn 1	Autumn 2	Pooled
T	17·68	11·33	16·92	40·99
T^{*}	18·98	12·40	18·05	42·20
T^{**}	19·67	12·90	18·87	42·67
$\tilde{T}$	17·68	11·33	16·92	40·99
$\tilde{T}^{*}$	19·67	12·90	18·87	42·67

and $\tilde{T}^*$ are shifted to the left compared with $\chi^2(2)$, so that these tests yield more significant results than their non-projected counterparts, T and T^{**}.

3·2. *A two-period cross-over study*

Next we consider data from a two-period cross-over study presented and analysed by Layard & Arveson (1978). In this study two drugs, A and B, for the treatment of nausea following chemotherapy were compared in 10 patients who received the drugs in the order (A, B) and 10 patients who received the drugs in the order (B, A). The response measured was the number of nausea episodes experienced during the first two hours following chemotherapy. We illustrate the tests discussed in this paper based on the model $Y_{ij}|b_{i1}, b_{i2} \sim \text{Po}\{\exp(X_{ij}'\alpha + b_{1i} + b_{2i}\,\text{TRT}_{ij})\}$. Here the covariate vector is $X_{ij} = (1, \text{TRT}_{ij}, \text{TP}_j, \text{DS}_i)'$, where TRT_{ij} is an indicator for drug A, the standard drug, TP_j is an indicator for time period two, and DS_i is an indicator that subject i received drug sequence (A, B) rather than (B, A). The subject-level random effects, $b_i = (b_{1i}, b_{2i})'$, are assumed to have mean 0 and 2×2 variance-covariance matrix $\Sigma(\theta) = (\sigma_{jk})$ $(j, k = 1, 2)$, where we take $\theta = (\sigma_{11}, \sigma_{12}, \sigma_{22})'$.

To yield a valid covariance matrix, the components of θ must satisfy $\sigma_{11} \geqslant 0$, $\sigma_{22} \geqslant 0$, and $|\sigma_{12}| \leqslant (\sigma_{11}\sigma_{12})^{\frac{1}{2}}$. Since the partial derivatives from the right satisfy

$$\left.\frac{\partial\{(\sigma_{11}\sigma_{22})^{\frac{1}{2}}\}}{\partial\sigma_{11}}\right|_{(\sigma_{11}=0)} = \infty \quad (\sigma_{22} > 0), \quad \left.\frac{\partial\{(\sigma_{11}\sigma_{22})^{\frac{1}{2}}\}}{\partial\sigma_{22}}\right|_{(\sigma_{22}=0)} = \infty \quad (\sigma_{11} > 0),$$

the constraint on σ_{12} vanishes in the limit, and the tangent cone becomes $\mathscr{T}_C = \{\theta : \sigma_{11} \geqslant 0, \sigma_{22} \geqslant 0\}$. The efficient score S, the information I and the efficient information I_e can all be determined through straightforward calculations as in Lin (1997). For the restricted score test we must project S on to the set $\mathscr{M} = I_e\mathscr{T}_C$; that is, we must find $\tilde{S} = \arg\min_{\tilde{s} \in I_e\mathscr{T}_C} (S - \tilde{s})' I_e^{-1} (S - \tilde{s})$. If we let $EE' = I_e$ be the Cholesky decomposition, then

$$\tilde{S} = \arg \min_{\tilde{s} \in EE'\mathscr{T}_C} (E^{-1}S - E^{-1}\tilde{s})'(E^{-1}S - E^{-1}\tilde{s})$$

so that $\tilde{S} = E\tilde{L}$, where $\tilde{L}$ is the Euclidian projection of $L = E^{-1}S$ on to $E'\mathscr{T}_C$. With this change of variables, the restricted score test is $\tilde{T} = \tilde{L}'\tilde{L}$.

According to Theorem 1, the limiting distribution of $\tilde{T}$ involves the Cholesky factor D. For approximating this distribution, we replace this matrix by its empirical counterpart E, and we obtain the cut-off values for $\tilde{T}$ from the estimated asymptotic distribution. This distribution is that of $\tilde{U}'\tilde{U}$, where $U = (U_1, U_2, U_3)'$ has independent $N(0, 1)$ components, and $\tilde{U}$ can be calculated as the Euclidian projection of U on to the set $E'\mathscr{T}_C$.

We now describe how to determine this projection. Let e_1, e_2 and e_3 be the columns of E'. Then the set on to which we project is

$$\mathscr{C} = \{x_1e_1 + x_2e_2 + x_3e_3 : x_1, x_3 \geqslant 0\}.$$

This set is a polyhedral cone in $\mathscr{R}^3$, so $\tilde{U}$ can be determined by techniques from Robertson et al. (1988, § 2.7). From these authors' results, $\tilde{U}$ can be written as

$$\tilde{U} = \sum_{F \in \mathscr{F}} \text{proj}\{U\,|\,\text{af}(F)\} 1_{\{U \in D_F\}},$$

where $1_{\{A\}}$ is the indicator of the event A, the sum is over the collection $\mathscr{F}$ of faces of $\mathscr{C}$, $\text{af}(F)$ denotes the affine span of face F, and D_F is the pre-image of F, that is, the part of $\mathscr{R}^3$ for which the projection on to $\mathscr{C}$ falls on F. Furthermore, the distribution of $\tilde{U}'\tilde{U}$ is given by

$$\mathrm{pr}(\tilde{U}'\tilde{U} \leqslant x) = \sum_{F \in \mathscr{F}} \mathrm{pr}[\chi^2\{\dim(F)\} \leqslant x]\, \mathrm{pr}(U \in D_F). \qquad (5)$$

The challenging part in determining (5) is to determine the level probabilities $\mathrm{pr}(U \in D_F)$. For the cone $\mathscr{C}$, they can be found as follows. Let $L_{ijk} = \mathrm{span}(e_i, e_j, e_k)$, $L_{ij} = \mathrm{span}(e_i, e_j)$, $L_i = \mathrm{span}(e_i)$, and let A_{123}, A_{12}, A_{23}, A_2 form the disjoint partition of $\mathscr{R}^3$ portrayed in Fig. 1. In Fig. 1, in which we look down on L_{13}, it is apparent that the cone $\mathscr{C}$ has four faces with, respectively, affine spans and pre-images (L_{123}, A_{123}), (L_{12}, A_{12}), (L_{23}, A_{23}) and (L_2, A_2). It follows from (5) that

$$\begin{aligned}\mathrm{pr}(\tilde{T} \leqslant x) = {} & \mathrm{pr}(U \in A_2)\, \mathrm{pr}\{\chi^2(1) \leqslant x\} + \mathrm{pr}(U \in A_{123})\, \mathrm{pr}\{\chi^2(3) \leqslant x\} \\ & + \{\mathrm{pr}(U \in A_{12}) + \mathrm{pr}(U \in A_{23})\}\, \mathrm{pr}\{\chi^2(2) \leqslant x\} + o(1),\end{aligned}$$

where χ^2 denotes a random variable with a chi-squared distribution with v degrees of freedom. Since $U \sim N(0, I_3)$, we can go to polar coordinates and evaluate these level probabilities as

$$\delta \equiv \mathrm{pr}(U \in A_{123}) = (2\pi)^{-1} |\arccos\{e_1' e_3/(\|e_1\|\, \|e_3\|)\}|,$$

$$\mathrm{pr}(U \in A_{12} \cup A_{23}) = \tfrac{1}{2}, \quad \mathrm{pr}(U \in A_2) = \tfrac{1}{2} - \delta.$$

Hence, the limiting law of $\tilde{T}$ is given by

$$\mathrm{pr}\,(\tilde{T} \leqslant x) = (\tfrac{1}{2} - \delta)\, \mathrm{pr}\{\chi^2(1) \leqslant x\} + \tfrac{1}{2}\, \mathrm{pr}\{\chi^2(2) \leqslant x\} + \delta\, \mathrm{pr}\{\chi^2(3) \leqslant x\} + o(1).$$

The coefficient δ takes values between 0 and $\frac{1}{2}$, so the most conservative null distribution occurs when $\delta = \frac{1}{2}$; it is an equal mixture of $\chi^2(2)$ and $\chi^2(3)$.

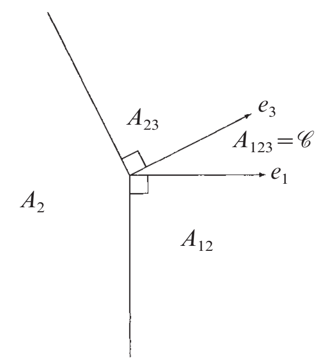

Fig. 1. The cone $\mathscr{C}$ viewed looking down on $L_{13} = \mathrm{span}(e_1, e_3)$.

In Table 2 we report test statistics and estimated values of δ for the cross-over data. In this example, the results based on all tests are qualitatively the same; however, the evidence for subject-level heterogeneity is appreciably weaker based on $\tilde{T}$ and $\tilde{T}^*$ as compared with their non-projected counterparts. Based on the relatively large p-values for all tests, we would fail to reject the null hypothesis of subject-level homogeneity. This result is consistent with the outcome of a maximum likelihood fit of the alternative hypothesis generalised linear mixed model which yields a maximum likelihood estimate of $\hat{\theta} = 0$. Based on the test results, we discarded the random effects and fitted the Poisson log-linear model. This model fitted adequately, yielding a deviance of 28·26 on 36 residual degrees of freedom. The p-value for a likelihood ratio test for treatment effect from this model is $p=0{\cdot}023$ so there is fairly strong evidence of the superiority of drug B over drug A.

Table 2. *Test statistics, estimates of δ and critical values based on the cross-over data*

Test statistic	Observed value	$\hat{\delta}$	0·05-level critical value	p-value
T	3·82	—	7·81	0·282
T^*	2·36	—	7·81	0·501
T^{**}	2·64	—	7·81	0·451
$\tilde{T}$	1·11	0·121	3·56	0·491
$\tilde{T}^*$	0·689	0·119	3·55	0·614

3·3. *The general case*

Suppose $\Sigma(\theta)$ is an unstructured $m \times m$ covariance matrix parameterised by the $m(m+1)/2$-dimensional vector θ that contains the distinct elements of Σ. If the elements of θ are ordered so that the m diagonal elements of Σ appear first, the tangent cone becomes $\mathscr{T}_C=[0,\infty)^m \times(-\infty,\infty)^{m(m-1)/2}$. The Cholesky factor $E=(e_1|\ldots|e_{\dim(\theta)})$ satisfying $EE'=I_e$ will be a $\dim(\theta)$-dimensional triangular matrix, and for calculating the projected test statistic from the score S we must project $E^{-1}S$ on to the cone

$$\mathscr{C}=E'\mathscr{T}_C=\left\{\sum_{i=1}^{\dim(\theta)} x_i e_i : x_1,\ldots,x_m \geqslant 0\right\}.$$

The limiting distribution is that of $\tilde{U}'\tilde{U}$, where $\tilde{U}$ is the projection of a $\dim(\theta)$-dimensional vector with independent, standard normal elements on to $\mathscr{C}$.

The restricted test poses two computational challenges above Lin's work: obtaining the Cholesky decomposition of I_e and computing the projection on to $\mathscr{C}$. Software for finding Cholesky decompositions is readily available. Algorithms for determining the projection on to polyhedral cones such as $\mathscr{C}$ are given by Dykstra (1983); however, as far as we know, these algorithms are not available as part of common statistical software packages. As is common in order-restricted inference, we have here a situation for which increased efficiency is available, but at a cost in terms of increased theoretical and computational complexity.

4. SIMULATIONS

4·1. *A model for the salamander mating example*

For comparability, we performed simulations based on the salamander data that were identical in design to the simulations presented by Lin (1997). That is, we took $n_f=n_m=60$ as in the pooled data and generated 360 binary observations according to model (4), where the random effects $b_i^{(f)}$ and $b_j^{(m)}$ were generated from the distribution

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

with mean 0 and variance $\theta=\sigma_f^2=\sigma_m^2=\pi(1-\pi)\gamma^2+\tau^2$. We consider the same four cases of F as does Lin.

Case 1. Homogeneity: $\pi=\gamma=\tau^2=\theta=0$.

Case 2. Normal: $\pi = \gamma = 0$, $\tau^2 = \theta = \{0{\cdot}25, 0{\cdot}50, 0{\cdot}75, 1{\cdot}00\}$.

Case 3. Unimodal normal mixture: $\pi = 0{\cdot}25$,

$$(\gamma, \theta) = \{(0{\cdot}80, 0{\cdot}25), (1{\cdot}10, 0{\cdot}50), (1{\cdot}40, 0{\cdot}75), (1{\cdot}70, 1{\cdot}00)\}, \quad \tau^2 = \theta - \pi(1-\pi)\gamma^2.$$

Case 4. Bimodal mixture: $\pi = 0{\cdot}25$,

$$(\gamma, \theta) = \{(1{\cdot}10, 0{\cdot}25), (1{\cdot}40, 0{\cdot}50), (1{\cdot}70, 0{\cdot}75), (2{\cdot}00, 1{\cdot}00)\}, \quad \tau^2 = \theta - \pi(1-\pi)\gamma^2.$$

Here, 3000 datasets were generated according to this model for each of the 13 choices of F corresponding to Cases 1–4. In each case, the fixed effects were set equal to their values based on fitting model (4) to the original data with restricted maximum likelihood: $\alpha = (1{\cdot}18, -0{\cdot}32, -2{\cdot}84, 3{\cdot}35)'$. Results appear in Table 3.

Table 3. *Estimated test size, $\theta = 0$, and power, $\theta \neq 0$, based on simulated data from model* (4) *for the salamander data*

Random effects distribution	θ	T	T^*	T^{**}	$\tilde{T}$	$\tilde{T}^*$
		Rejection rate				
Normal	0·00	0·051	0·049	0·051	0·050	0·058
	0·25	0·289	0·310	0·313	0·401	0·435
	0·50	0·692	0·717	0·720	0·788	0·813
	0·75	0·899	0·910	0·913	0·946	0·955
	1·00	0·970	0·975	0·975	0·984	0·987
Unimodal mixture	0·25	0·317	0·340	0·345	0·436	0·463
	0·50	0·697	0·721	0·725	0·798	0·822
	0·75	0·903	0·915	0·916	0·949	0·955
	1·00	0·974	0·978	0·978	0·989	0·990
Bimodal mixture	0·25	0·331	0·356	0·359	0·452	0·486
	0·50	0·725	0·750	0·757	0·823	0·841
	0·75	0·924	0·932	0·933	0·958	0·964
	1·00	0·981	0·985	0·985	0·992	0·994

4·2. *A log-linear model for the cross-over data with random intercept and slope*

We assume that, conditional on bivariate normal, subject-specific random effects, $b_i = (b_{1i}, b_{2i})'$, the response for subject i in period j is Poisson with conditional mean that satisfies the log-linear model of § 3·2. The value of α in this model was taken to be $\alpha = (-0{\cdot}70, 0{\cdot}93, 0{\cdot}46, 0{\cdot}053)'$. From this model, we randomly generated datasets for each of four specifications of θ and three specifications of K, the number of subjects, and examined the rejection rates of the five test statistics described in §§ 2 and 3. Results appear in Table 4. The power estimates, for $\theta \neq 0$, were all obtained based on 3000 Monte Carlo replicates, but to obtain more precise estimates of size we utilised 5000 replicates at $\theta = 0$.

From Table 4 it is apparent that the power is greater for $\tilde{T}$ than for T and greater for $\tilde{T}^*$ than for either T^* or T^{**}. A referee has raised the concern that these findings may result from the fact that the projected tests have larger actual sizes than their non-projected counterparts, rather than because of any inherent power superiority of these tests. To address this concern we examined the empirical power of the non-projected tests when computed based on significance levels set equal to the observed sizes of $\tilde{T}$ and $\tilde{T}^*$. That is, we 'levelled the playing field' by comparing the power of T, T^* and T^{**} with that of

Table 4. *Estimated test size, $\theta = 0$, and power, $\theta \neq 0$, based on simulated data from the log-linear mixed model for the cross-over data*

K	θ'	Rejection rate T	T^*	T^{**}	$\tilde{T}$	$\tilde{T}^*$
20	(0, 0, 0)	0·031	0·038	0·050	0·037	0·067
	(0·59, −0·12, 0·29)/4	0·103	0·127	0·156	0·120	0·186
	(0·59, −0·12, 0·29)/2	0·244	0·274	0·315	0·270	0·362
	(0·59, −0·12, 0·29)	0·570	0·603	0·645	0·604	0·692
40	(0, 0, 0)	0·042	0·048	0·054	0·055	0·075
	(0·59, −0·12, 0·29)/4	0·145	0·171	0·190	0·190	0·245
	(0·59, −0·12, 0·29)/2	0·413	0·456	0·482	0·470	0·543
	(0·59, −0·12, 0·29)	0·840	0·864	0·873	0·875	0·902
80	(0, 0, 0)	0·043	0·048	0·051	0·055	0·071
	(0·59, −0·12, 0·29)/4	0·240	0·270	0·281	0·303	0·349
	(0·59, −0·12, 0·29)/2	0·646	0·672	0·686	0·709	0·754
	(0·59, −0·12, 0·29)	0·976	0·980	0·981	0·985	0·989

Table 5. *Estimated power based on simulated data from the log-linear mixed model for the cross-over data. Here we examine the empirical power of T, T^* and T^{**} when the nominal size is set equal to observed size of $\tilde{T}$ or $\tilde{T}^*$*

K	θ'	Rejection rate T	T^*	T^{**}	$\tilde{T}$	$\tilde{T}^*$	Nominal size
20	(0·59, −0·12, 0·29)/4	0·093	0·112	0·138	0·120	—	0·037
		0·118	0·143	0·172	—	0·186	0·067
	(0·59, −0·12, 0·29)/2	0·227	0·258	0·297	0·270	—	0·037
		0·260	0·298	0·343	—	0·362	0·067
	(0·59, −0·12, 0·29)	0·550	0·589	0·623	0·604	—	0·037
		0·589	0·627	0·666	—	0·692	0·067
40	(0·59, −0·12, 0·29)/4	0·152	0·178	0·198	0·190	—	0·055
		0·175	0·208	0·231	—	0·245	0·075
	(0·59, −0·12, 0·29)/2	0·426	0·465	0·490	0·470	—	0·055
		0·459	0·498	0·522	—	0·543	0·075
	(0·59, −0·12, 0·29)	0·845	0·868	0·878	0·875	—	0·055
		0·860	0·882	0·889	—	0·902	0·075
80	(0·59, −0·12, 0·29)/4	0·251	0·280	0·291	0·303	–	0·055
		0·281	0·309	0·320	—	0·349	0·071
	(0·59, −0·12, 0·29)/2	0·654	0·684	0·698	0·709	—	0·055
		0·679	0·710	0·717	—	0·754	0·071
	(0·59, −0·12, 0·29)	0·978	0·981	0·982	0·985	—	0·055
		0·980	0·984	0·985	—	0·989	0·071

$\tilde{T}$ when the three non-projected tests were performed at a level equal to the observed size of $\tilde{T}$, for example 0·037 for $K = 20$; and by comparing the power of T, T^* and T^{**} with that of $\tilde{T}^*$ when the non-projected tests were performed at level equal to the observed size of $\tilde{T}^*$, for example 0·067 for $K = 20$. These results appear in Table 5. Note that, in each case, the empirical power of $\tilde{T}$ still exceeds that of Lin's T and T^*, and the empirical power of $\tilde{T}^*$ exceeds that of all three non-projected tests.

5. Discussion

The simulation results presented in § 4 clearly indicate that projecting Lin's score tests for homogeneity on to the tangent cone of the model results in power increases that can be substantial. Power gains were greatest for the projected and bias-corrected statistic $\tilde{T}^*$, although this test did exhibit somewhat inflated size. With respect to both size and power, $\tilde{T}$, the projected version of Lin's original uncorrected statistic, performed best among all of the test statistics considered here. Our simulations reproduce Lin's (1997) result, showing a slight advantage for the bias-corrected T^* over T in terms of size and power. However, the results were even better for T^{**}, leading us to conclude that, if a bias correction is to be employed, it should be applied to both the score vector and information matrix. In simulations not reported here, we compared the non-bias-corrected statistics T and $\tilde{T}$ in a nonlinear mixed model context. Results were quite similar to those reported in § 4. In the nonlinear mixed model context, $\tilde{T}$ produced substantial gains in power and also had much closer to nominal size. Development of bias-corrected tests in the nonlinear mixed model setting is a worthwhile area for further research.

Acknowledgement

The work of the first author was supported by a University of Georgia Faculty Research Grant. We thank Michael Perlman for helpful discussions, Xihong Lin for making available her Gauss code for computing her test statistics, and an anonymous referee for comments that led to improvements in the paper.

Appendix

Proof of Theorem 2

We now prove Theorem 2, noting that a proof of Theorem 1 is contained herein. We write $\gamma = (\alpha, \theta)$ and recall that the true parameter is $\hat{\gamma}_{n0} \equiv (\alpha_0', \theta_n')'$, where n is some measure of sample size, and we assume that $n^{1/2}\theta_n \to \delta \in \mathscr{T}_C$. Let $Z \sim N(0, J)$ be of dimension $\dim(\alpha) + \dim(\theta)$ with positive definite variance-covariance matrix J. Let $\hat{\gamma}_n = (\hat{\alpha}_n', 0')'$, where $\hat{\alpha}_n$ is the maximum likelihood estimator of α under the assumption $\theta = 0$. We partition $S(\gamma)$, $I(\gamma)$, Z and J as follows:

$$S(\gamma) = \begin{Bmatrix} S_\alpha(\gamma) \\ S_\theta(\gamma) \end{Bmatrix}, \quad I(\gamma) = \begin{Bmatrix} I_{\alpha\alpha}(\gamma) & I_{\alpha\theta}(\gamma) \\ I_{\theta\alpha}(\gamma) & I_{\theta\theta}(\gamma) \end{Bmatrix}, \quad Z = \begin{pmatrix} Z_\alpha \\ Z_\theta \end{pmatrix}, \quad J = \begin{pmatrix} J_{\alpha\alpha} & J_{\alpha\theta} \\ J_{\theta\alpha} & J_{\theta\theta} \end{pmatrix}.$$

We should really index S and I by n, but this would make the notation unnecessarily cumbersome. With this notation, $\hat{\gamma}_n$ solves $S_\alpha(\hat{\gamma}_n) = 0$, and Lin's test statistic is

$$T = S_\theta(\hat{\gamma}_n)' I_e^{-1} S_\theta(\hat{\gamma}_n),$$

where $I_e = \{I_{\theta\theta}(\hat{\gamma}_n) - I_{\theta\alpha}(\hat{\gamma}_n) I_{\alpha\alpha}(\hat{\gamma}_n)^{-1} I_{\alpha\theta}(\hat{\gamma}_n)\}$. Our projected test statistic is

$$\tilde{T} = \mathrm{proj}\{S_\theta(\hat{\gamma}_n) | \mathscr{M}, I_e^{-1}\}' I_e^{-1} \, \mathrm{proj}\{S_\theta(\hat{\gamma}_n) | \mathscr{M}, I_e^{-1}\},$$

where $\mathrm{proj}(.|\mathscr{M}, I_e^{-1})$ denotes the projection operator on to $\mathscr{M}$ with respect to I_e^{-1}.

We assume that the following conditions hold.

Condition 1. We have that $n^{-1/2} S(\hat{\gamma}_{n0}) = Z + o_p(1)$, where $Z \sim N(0, J)$.

Condition 2. We have that $\text{plim}_n\, n^{-1}(\partial/\partial\gamma)S(\hat{\gamma}_{n0}) = \text{plim}_n\, n^{-1}I(\hat{\gamma}_{n0}) = J$.

Condition 3. We have that $\hat{\alpha}_n = \alpha_0 + O_p(n^{-1/2})$.

Condition 4. For some neighbourhood ω of $(\alpha_0, 0)$ it holds that

$$\sup_{\gamma\in\omega}\max_{i,j,k} \frac{\partial^3 \log L(\gamma)}{\partial\gamma_i\,\partial\gamma_j\,\partial\gamma_k} = O_p(n^{-1}).$$

Under Conditions 1–4 it holds that

$$0 = n^{-1/2}S_\alpha(\hat{\gamma}_n)$$
$$= n^{-1/2}S_\alpha(\hat{\gamma}_{n0}) - n^{-1}\left\{\left(\frac{\partial}{\partial\alpha}\right)S_\alpha(\hat{\gamma}_{n0}) : \left(\frac{\partial}{\partial\theta}\right)S_\alpha(\hat{\gamma}_{n0})\right\}\begin{Bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha_0) \\ n^{1/2}(0-\theta_n)\end{Bmatrix} + o_p(1), \quad (A1)$$

$$n^{-1/2}S_\theta(\hat{\gamma}_n) = n^{-1/2}S_\theta(\hat{\gamma}_{n0}) - n^{-1}\left\{\left(\frac{\partial}{\partial\alpha}\right)S_\theta(\hat{\gamma}_{n0}) : \left(\frac{\partial}{\partial\theta}\right)S_\theta(\hat{\gamma}_{n0})\right\}\begin{Bmatrix} n^{1/2}(\hat{\alpha}_n - \alpha_0) \\ n^{1/2}(0-\theta_n)\end{Bmatrix} + o_p(1). \quad (A2)$$

By solving (A1) for $n^{1/2}(\hat{\alpha}_n - \alpha)$ and using Condition 2, we find that

$$n^{1/2}(\hat{\alpha}_n - \alpha_0) = J_{\alpha\alpha}^{-1}Z_\alpha + J_{\alpha\alpha}^{-1}J_{\alpha\theta}\delta + o_p(1). \quad (A3)$$

Then, inserting (A3) into (A2) shows that

$$n^{-1/2}S_\theta(\hat{\gamma}_n) = Z_\theta + J_{\theta\theta}\delta - J_{\theta\alpha}(J_{\alpha\alpha}^{-1}Z_\alpha + J_{\alpha\alpha}^{-1}J_{\alpha\theta}\delta) + o_p(1) \equiv W + o_p(1),$$

where we define

$$W = Z_\theta - J_{\theta\alpha}J_{\alpha\alpha}^{-1}Z_\alpha + (J_{\theta\theta} - J_{\theta\alpha}J_{\alpha\alpha}^{-1}J_{\alpha\theta})\delta, \quad J_e = (J_{\theta\theta} - J_{\theta\alpha}J_{\alpha\alpha}^{-1}J_{\alpha\theta}),$$

so we have by Condition 2 that $n^{-1}I_e = J_e + o_p(1)$. Hence, the unrestricted test statistic satisfies $T = W'J_e^{-1}W + o_p(1)$. It further holds that

$$\tilde{T} = \text{proj}\{S_\theta(\hat{\gamma}_n)\,|\,\mathscr{M}, I_e^{-1}\}'I_e^{-1}\,\text{proj}\{S_\theta(\hat{\gamma}_n)\,|\,\mathscr{M}, I_e^{-1}\}$$
$$= \text{proj}\{n^{-1/2}S_\theta(\hat{\gamma}_n)\,|\,nI_e\mathscr{T}_C, nI_e^{-1}\}'nI_e^{-1}\,\text{proj}\{n^{-1/2}S_\theta(\hat{\gamma}_n)\,|\,nI_e\mathscr{T}_C, nI_e^{-1}\}$$
$$= \text{proj}(W\,|\,J_e\mathscr{T}_C, J_e^{-1})'J_e^{-1}\,\text{proj}(W\,|\,J_e\mathscr{T}_C, J_e^{-1}) + o_p(1). \quad (A4)$$

Now, $W \sim N(J_e\delta, J_e)$, so we can write it as $W = D(U + D'\delta)$, where $DD' = J_e$. Then

$$W'J_e^{-1}W = \{D(U+D'\delta)\}'(DD')^{-1}D(U+D'\delta) = (U+\mu)'(U+\mu),$$

where $\mu = D'\delta$. This proves the first claim of Theorem 2. As for $\tilde{T}$, we have that

$$\text{proj}(W\,|\,J_e\mathscr{T}_C, J_e^{-1}) = \arg\min_{k\in J_e\mathscr{T}_C}(W-k)'J_e^{-1}(W-k)$$
$$= \arg\min_{k\in J_e\mathscr{T}_C}(U + D'\delta - D^{-1}k)'(U + D'\delta - D^{-1}k)$$
$$= D\arg\min_{l\in D'\mathscr{T}_C}(U + D'\delta - l)'(U - D'\delta - l)$$
$$= D\,\text{proj}(U+\mu\,|\,D'\mathscr{T}_C, 1).$$

If we use $DD' = J_e$ and (A4), it follows that

$$\tilde{T} = \text{proj}(U+\mu\,|\,D'\mathscr{T}_C, 1)'\;\text{proj}(U+\mu\,|\,D'\mathscr{T}_C, 1) + o_p(1).$$

REFERENCES

BHAT, B. R. & NAGNUR, B. N. (1965). Locally asymptotically most stringent tests and Lagrangian multiplier tests of linear hypotheses. *Biometrika* **52**, 459–68.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.* **25**, 573–8.

DYKSTRA, R. D. (1983). An algorithm for restricted least squares regression. *J. Am. Statist. Assoc.* **78**, 837–42.

GEYER, C. J. (1994). On the asymptotics of constrained *M*-estimation. *Ann. Statist.* **22**, 1993–2010.

GOURIEROUX, C., HOLLY, A. & MONFORT, A. (1982). Likelihood ratio test, Wald test, and Kühn-Tücker test in linear models with inequality constraints on the regression parameters. *Econometrica* **50**, 63–80.

LAYARD, M. W. J. & ARVESON, J. N. (1978). Analysis of Poisson data in crossover experimental designs. *Biometrika* **34**, 421–8.

LIN, X. (1997). Variance component testing in generalised linear models with random effects. *Biometrika* **84**, 309–26.

LIN, X. & BRESLOW, N. E. (1996). Bias correction in generalised 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.

ROBERTSON, T., WRIGHT, F. T. & DYKSTRA, R. L. (1988). *Order Restricted Statistical Inference*. New York: Wiley.

ROGERS, A. (1986). Modified Lagrange multiplier tests for problems with one-sided alternatives. *J. Economet.* **31**, 341–61.

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

SILVAPULLE, M. J. & SILVAPULLE, P. (1995). A score test against one-sided alternatives. *J. Am. Statist. Assoc.* **90**, 342–9.

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

STRAM, D. O. & LEE, J. W. (1994). Variance components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171–7.

TSAI, M. (1992). On the power superiority of likelihood ratio tests for restricted alternatives. *J. Mult. Anal.* **42**, 102–9.

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

WOLFINGER, R. D. & LIN, X. (1997). Two Taylor-series approximation methods for nonlinear mixed models. *Comp. Statist. Data Anal.* **25**, 465–90.

[*Received May* 1999. *Revised February* 2001]