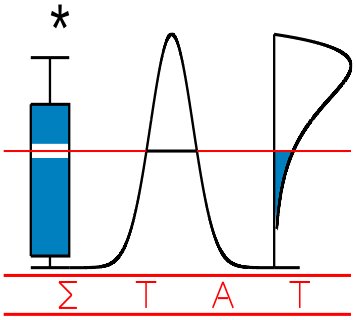
T E C H N I C A L

R E P O R T

0436

ONE-SIDED TESTS IN SHARED FRAILTY MODELS

NGUTI, R., CLAESKENS, G. and P. JANSSEN



I A P S T A T I S T I C S

N E T W O R K

INTERUNIVERSITY ATTRACTION POLE

http://www.stat.ucl.ac.be/IAP

One-sided tests in shared frailty models

Rosemary Nguti[12], Gerda Claeskens[3] and Paul Janssen[2]

[1] *University of Nairobi and International Livestock Research Institute*

[3] *Université catholique de Louvain*

[2] *Limburgs Universitair Centrum*

Abstract

Tests for the presence of heterogeneity in frailty models use an alternative hypothesis in which the heterogeneity parameter is subject to an inequality constraint. As a result the classical likelihood ratio asymptotic chi-squared distribution theory is no longer valid. Our main result states the limiting distribution of the likelihood ratio and score statistic for the one-sided testing problem. The resulting distribution is a mixture of chi-squared distributed random variables. The results are shown to hold when the data might be subject to right censoring and when covariate information is present. A data example on a diabetic retinopathy study illustrates the tests.

Key words: Inference under inequality constraints, Frailty models, Likelihood ratio test, Mixture of χ^2-distributions, Score test, Survival data.

1 Introduction

Consider a set of n subjects, clusters or groups. For cluster i ($i = 1, \ldots, n$) we observe a vector of n_i variables T_{ij}, together with a censoring indicator δ_{ij} which takes on the value 1 if the event of interest is observed and $\delta_{ij} = 0$ otherwise, as well as covariates x_{ij}. The data take the form $(T_{i1}, \delta_{i1}, x_{i1})$, ..., $(T_{in_i}, \delta_{in_i}, x_{in_i})$. Denoting the true event times by Y_{ij}, $T_{ij} = \min(Y_{ij}, C_{ij})$ where C_{ij} is the censoring time. In a shared frailty model the within cluster dependence is described by the independent frailty variables $Z_1, \ldots, Z_n$ representing

[1]Department of Mathematics, University of Nairobi and Biometrics Unit, International Livestock Research Institute, Nairobi, Kenya.

[2]Center for Statistics, Limburgs Universitair Centrum, B-3590 Diepenbeek, Belgium.

[3]Institute of Statistics, Voie du Roman Pays 20, B-1348 Louvain-la-Neuve, Belgium.

E-mail: claeskens@stat.ucl.ac.be. Corresponding author.

unobserved common risk factors. This leads to a "proportional" hazards model with random effects (Clayton, 1978, Clayton & Cuzick, 1985). We want to test homogeneity (no association between the survival outcomes) versus heterogeneity (there is positive association). We assume that, conditional on the frailty variables Z_i, the Y_{ij} $(j = 1, \ldots, n_i)$ are independent with, for $Z_i = z$, a Weibull$(z\lambda_j, \gamma)$ distribution with $\lambda_j > 0$ and $\gamma > 0$. This distribution is a standard model for parametric survival data. In a first case we take Z_i having a gamma distribution with mean 1 and variance θ. For more information on this gamma frailty model and examples of its use, see Hougaard (2000, Chapter 7) and Therneau and Grambsch (2000, Chapter 9). In the second case, in Section 4, we discuss the situation of a positive stable distribution for the frailty variables.

At time t, we denote for subject j within cluster i the conditional hazard function $h(t|z) = z\mu_j(t)$ where $\mu_j(t) = \lambda_j \gamma t^{\gamma-1} \exp(\beta x_{ij})$ and the corresponding cumulative function $M_j(t) = \lambda_j t^{\gamma} \exp(\beta x_{ij})$. Let D_i denote the total number of events for cluster i, $D_i = \sum_{j=1}^{n_i} \delta_{ij}$. From a modelling point of view it should be understood that all $\lambda_j = \lambda$ when the cluster sizes are not all the same, while in the case of equal cluster sizes, for example data on pairs, we might use different parameters λ_j. The likelihood for these data is given by

$$\prod_{i=1}^{n} \left[(-1)^{D_i} \prod_{j=1}^{n_i} \{\mu_j(T_{ij})\}^{\delta_{ij}} L^{(D_i)}\{\sum_{j=1}^{n_i} M_j(T_{ij})\} \right]$$

where L is the Laplace transform of the distribution of the frailty Z. For the gamma frailty $L(s) = (1 + \theta s)^{-1/\theta}$. This leads to the following form of the likelihood

$$\prod_{i=1}^{n} \left[\frac{\theta^{D_i} \Gamma(1/\theta + D_i)}{\Gamma(1/\theta)} \left[\prod_{j=1}^{n_i} \{\lambda_j \gamma T_{ij}^{\gamma-1} \exp(\beta x_{ij})\}^{\delta_{ij}} \right] \left\{ 1 + \theta \sum_{j=1}^{n_i} \lambda_j T_{ij}^{\gamma} \exp(\beta x_{ij}) \right\}^{-1/\theta - D_i} \right]. \tag{1}$$

Often a fixed number of observations $n_i = q$ is recorded. Twin studies provide typical examples of bivariate event data. Another example are the times to failure for several similar human organs like time to blindness of the right and the left eye as in studies on diabetic retinopathy.

The relevant testing problem for the null hypothesis of no association between the survival

outcomes in a cluster is

$$H_0 : \theta = 0 \quad \text{versus} \quad H_a : \theta > 0. \tag{2}$$

In this situation a one-sided testing problem is natural. Matters are non-standard for a twofold reason: (i) the heterogeneity parameter is on the boundary of its parameter space under H_0 and (ii) the alternative model contains an inequality constraint. The boundary problem can be circumvented via a (perhaps artificial) embedding of the natural parameter space in a slightly bigger space which includes negative values for the association parameter θ (see Hougaard, 2000, Section 7.3.3). Adjustment for one-sided testing, however, remains to be dealt with; also in an enlarged parameter space techniques for testing under inequality constraints are needed. Commenges & Anderson (1995) and Murphy & van der Vaart (1997) only consider two-sided tests. We obtain the asymptotic distribution of the likelihood ratio and score statistic for the *one-sided* testing problem.

Even though one-sided testing has a long history, going back to Chernoff (1954), its asymptotic distributional behaviour has been studied only recently. For more references on this issue, see Self & Liang (1987) for independent and identically distributed data. Silvapulle & Silvapulle (1995) study one-sided score tests, see also Verbeke & Molenberghs (2003) in the context of mixed linear models. Vu & Zhou (1997) derive general theoretical results; more information is contained in the overview paper by Sen & Silvapulle (2002).

In this paper we focus on the survival context and show rigorously when and under which conditions the general results of Vu and Zhou (1997) are applicable. A careful analysis is required, especially in regression models with covariate information. In addition, we allow for right censored observations.

Section 2 states the asymptotic distribution of the likelihood ratio statistic for this testing problem, the definition and distribution of the score statistic is dealt with in Section 3. Section 4 presents an extension to the positive stable frailty model. We apply our findings to the diabetic retinopathy study (Huster et al. 1989) in Section 5, and a discussion on possible extensions can be found in Section 6. Section 7 collects all proofs.

2 One-sided likelihood ratio tests

To test for within subject (or cluster) dependence we consider the testing problem (2). For the further discussion it is convenient to work with the following transformed Weibull parameters: $\eta_j = -\log \lambda_j$ $(j = 1, \ldots, q)$ and $\alpha = -\log \gamma$. This transformation yields the full real line as parameter space of the transformed parameters, which simplifies notational aspects and allows to focus better on the heterogeneity parameter. Further we use $\nu = (\eta_1, \ldots, \eta_q, \alpha, \beta)$ as shorthand notation for the set of nuisance parameters and let $\tau = (\theta, \nu)$. In terms of the transformed parameter vector τ the parameter space is $\Theta = [0, \infty) \times \mathbb{R}^{q+2}$ and the testing problem can be written as

$$H_0 : \tau \in \Theta_0 = \{0\} \times \mathbb{R}^{q+2} \text{ against } H_a : \tau \in \Theta_1 = (0, \infty) \times \mathbb{R}^{q+2}$$

with corresponding likelihood ratio statistic

$$\mathcal{L}_n = 2 \left\{ \sup_{\tau \in \Theta} L_n(\tau) - \sup_{\tau \in \Theta_0} L_n(\tau) \right\}$$

where $L_n(\tau)$ is the log likelihood function.

In obtaining the asymptotic distribution of the likelihood ratio statistic we distinguish, for transparency of the proof, three situations. First we focus on the frailty model without the extra complication of censoring and covariates. Without loss of generality we will assume that each cluster contains two observations. Bjarnason and Hougaard (2000) use this model to study the Fisher information matrix. The idea behind the simplification is to fully understand the statistical properties for a simple, though practically relevant, model. In a second situation we study complete data including covariates. Third, we deal with the effect of censoring.

2.1 Complete survival data

For bivariate data (T_{i1}, T_{i2}), $i = 1, \ldots, n$ with $\lambda_1 = \lambda_2 = \lambda$, no censoring and no covariates, it follows from (1) that the log likelihood in terms of the original parametrisation equals

$$\begin{aligned} L_n(\theta, \lambda, \gamma) &= 2n \log(\lambda\gamma) + n \log(1 + \theta) + \sum_{i=1}^{n} [(\gamma - 1)(\log T_{i1} + \log T_{i2}) \\ &\quad -(2 + 1/\theta) \log\{1 + \theta\lambda(T_{i1}^{\gamma} + T_{i2}^{\gamma})\}]. \end{aligned}$$

THEOREM 1. *The likelihood ratio statistic $\mathcal{L}_n$ for testing the one-sided heterogeneity hypotheses $H_0 : (\theta, \eta, \alpha) \in \{0\} \times \mathbb{R}^2$ against $H_a : (\theta, \eta, \alpha) \in (0, \infty) \times \mathbb{R}^2$ in the shared gamma frailty model with Weibull baseline hazard has an asymptotic null distribution which is an equal mixture of a point mass at zero and a chi-square distribution with one degree of freedom, that is, $\mathcal{L}_n \to_d \frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$ as $n \to \infty$.*

We give the proof of this theorem in Section 7. The structure of the proof makes clear how to deal with boundary parameters as well as with the one-sided aspect of the testing problem. The main part of the proof consists of showing how for parametrically modelled survival data the results of Vu and Zhou (1997) apply. For independent and identically distributed data we could also have applied the results of Self and Liang (1987). The main difficulty lies in working with the matrix of negative second derivatives of the log likelihood, and its limit, the Fisher information matrix. We obtain explicit expressions for these matrices, allowing for a detailed proof. This may be seen as an addition to earlier results obtained by simulation only.

This result has an immediate impact on how to determine (asymptotic) critical values and p-values for likelihood ratio tests for heterogeneity. Further it is well known that one-sided testing as compared to two-sided testing is a less conservative strategy in rejecting the null hypothesis of independence.

2.2 Complete survival data with covariate information

It is possible to generalize Theorem 1 to the situation of non-identically distributed observations. This allows the distribution of the lifetimes T_{ij} to depend on covariate information. The log likelihood from (1) now reads

$$\begin{aligned} L_n(\theta, \lambda, \gamma) &= 2n\log(\lambda\gamma) + n\log(1+\theta) + \sum_{i=1}^{n}\{\beta(x_{i1}+x_{i2}) + (\gamma-1)(\log T_{i1} + \log T_{i2})\} \\ &\quad -(2+1/\theta)\sum_{i=1}^{n}\log\{1+\theta\lambda(T_{i1}^{\gamma}\exp(\beta x_{i1}) + T_{i2}^{\gamma}\exp(\beta x_{i1}))\}]. \end{aligned}$$

We here let β be univariate, extension to more than one covariate is immediate. Let $F_n(\tau)$ be the matrix of the negative of the second derivatives of $L_n(\tau)$ and define $G_n(\nu) = E[F_n(0, \nu)]$. Unlike the covariate-free case where the Fisher information matrix G_n only depends on the

parameter η and its determinant is independent of any parameter (see the proof of Theorem 1 in Section 7) in the presence of covariates there is additional dependence of the matrix G_n on the coefficients β and η. Restrictions on the allowable range of β and η values are necessary for the asymptotic theory to go through. These restrictions depend on the type of covariates. It is not uncommon that restrictions apply, for example in joint modelling of outcomes via copulas (Nelsen 1999) the parameter space is restricted as well.

To explain this issue we take the same modelling situation as in the previous section, yet with covariates added. Denote $M_{kl} = \sum_{i=1}^{n} \sum_{j=1}^{2} x_{ij}^k \exp(l\beta x_{ij})$ and $N_{kl} = \sum_{i=1}^{n} \exp\{\beta(kx_{i1} + lx_{i2})\}$. With this notation the symmetric Fisher information matrix $G_n = G_n(\eta, \beta)$ has components: $G_{1,1} = n + 4(M_{0,3} - M_{0,2} + N_{2,1} + N_{1,2} - N_{1,1})$, $G_{2,2} = M_{0,1}$, $G_{3,3} = -2n(\psi(1) + \eta) + (\psi(2) + \eta)M_{0,1} + \{(\psi(2) + \eta)^2 + \zeta(2,2)\}M_{0,1}$, $G_{4,4} = M_{2,1}$, $G_{1,2} = 2N_{1,1} + 2M_{0,2} - 2M_{0,1}$, $G_{1,3} = -2(\psi(2)+\eta)M_{0,1}+2(\psi(3)+\eta)M_{0,2}+2(\psi(2)+\eta)N_{1,1}$, $G_{1,4} = 2M_{1,1}-2M_{1,2}-\sum_{i=1}^{n}(x_{i1}+ x_{i2})\exp\{\beta(x_{i1} + x_{i2})\}$, $G_{2,3} = (\psi(2) + \eta)M_{0,1}$, $G_{2,4} = -M_{1,1}$ and $G_{3,4} = -(\psi(2) + \eta)M_{1,1}$. Here ψ denotes the digamma function and $\zeta(2,2) = \int_0^\infty te^{-2t}/(1 - e^{-t})dt$. One of the key issues is that the matrix G_n is positive definite. Without this assumption none of the results mentioned in this paper is guaranteed to hold. We consider three applications.

Example 1. If $\beta = 0$, this is the situation of the previous section. No restrictions on the parameter space apply. In this case $\det(G_n) = n^3(2\pi^2 - 8)$, independent of η, and the matrix G_n is positive definite.

Example 2. Suppose n is even and for a binary covariate, half of the observations $x_{i1} = 1$ with the corresponding $x_{i2} = 0$. This situation occurs for example for a treatment of the eyes when only one eye is treated, either left or right, and the other serves as a control. A specification of the matrix G_n gives that the entries of $\tilde{G} = G_n/n$ are given by $\tilde{G}_{1,1} = 1+4\exp(3\beta)$, $\tilde{G}_{2,2} = 1 + \exp(\beta)$, $\tilde{G}_{3,3} = -2(\psi(1) + \eta) + \{\psi(2) + \eta + (\psi(2) + \eta)^2 + \zeta(2,2)\}\{1 + \exp(\beta)\}$, $\tilde{G}_{4,4} = \exp(\beta)$, $\tilde{G}_{1,2} = 2\exp(2\beta)$, $\tilde{G}_{1,3} = 2(\psi(3) + \eta)\exp(2\beta) + 1$, $\tilde{G}_{1,4} = \exp(\beta) - 2\exp(2\beta)$, $\tilde{G}_{2,3} = (\psi(2)+\eta)\{1+\exp(\beta)\}$, $\tilde{G}_{2,4} = -\exp(\beta)$ and $\tilde{G}_{3,4} = -(\psi(2)+\eta)\exp(\beta)$. Direct calculation shows that the determinant of $\tilde{G}$ can take on both positive and negative values. Requiring the matrix $G_n(\eta, \beta)$ to be positive definite implies a restriction on β and the intensity η.

Figure 1 gives the allowable area of covariate combinations. Taking e.g. $\eta = 2$ we need to require that a positive $\beta < 2.1674$ which corresponds to a risk difference between treatment and non-treatment of size 8.736. When both η and β are negative, no further restrictions apply.

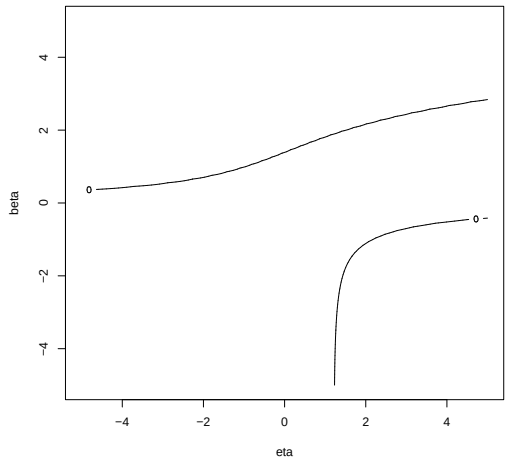


Figure 1. The allowable range of values of the intensity η (horizontally) and regression coefficient β (vertically) for survival data with a binary covariate, is located in between the curves of zero value for det(G_n).

Example 3. Let the covariate data be generated from distributions F_j such that $x_{ij} = F_j^{-1}(u_i)$ where $u_i = (i - 1/2)/n$. It is now easy to compute limits for $n \to \infty$ for the entries of the matrix G_n. For example, $n^{-1}M_{k,l} \to \sum_{j=1}^{2} E[X_j^k \exp(l\beta X_j)]$ where $X_j \sim F_j$. Also in this case, restrictions on the regression coefficient and the intensity are needed for the matrix G_n to be positive definite.

THEOREM 2. *Assume that, for n tending to infinity, the limiting Fisher information matrix G is positive definite and that $|\beta x|$ is finite uniformly over the coefficient β and the covariate space of x. Then the limiting null distribution of the likelihood ratio statistic $\mathcal{L}_n$ for testing the one-sided heterogeneity hypothesis in the shared gamma frailty model with Weibull baseline hazard is $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$.*

The proof of Theorem 2 (see Section 7) relies on Theorem 2.2 in Vu and Zhou (1997). This is a situation where Self and Liang's (1987) results do not apply since the data are not identically distributed. As explained in the examples above, it is crucial to check the assumptions of the theorem for the particular dataset at hand. Depending on the covariate

design, restrictions to the parameter space might apply. "Guessing" asymptotic results might for correlated survival data lead to situations where the theoretical results are not applicable. By our knowledge, such a rigorous verification for the class of survival data is nowhere available.

2.3 Censored survival data

For right censored bivariate data the general (log)likelihood (1) reduces to

$$\begin{aligned} L_n(\theta) &= -\frac{1}{\theta}\sum_{\{i:D_i=0\}}\log\{1+\theta(T_{i1}+T_{i2})\} - (\frac{1}{\theta}+1)\sum_{\{i:D_i=1\}}\log\{1+\theta(T_{i1}+T_{i2})\} \\ &\quad -(\frac{1}{\theta}+2)\sum_{\{i:D_i=2\}}\log\{1+\theta(T_{i1}+T_{i2})\} + N_2\log(1+\theta) \end{aligned}$$

where $N_2 = \#\{i : D_i = 2\}$ and for simplicity we choose $\lambda_1 = \lambda_2 = \gamma = 1$ and $\beta = 0$ (no covariates). Since the heterogeneity parameter is the only remaining parameter, the likelihood ratio test for

$$H_0 : \theta = 0 \text{ versus } H_a : \theta > 0 \tag{3}$$

is $\mathcal{L}_n = 2\{\sup_{\theta\geq 0} L_n(\theta) - L_n(0)\}$. This log likelihood is, with $\lambda(t) \equiv 1$ and $\tau = \infty$, a particular example of the likelihood and the log likelihood expression given on pages 1476 and 1490 of Murphy and van der Vaart (1997). Indeed an equivalent way to describe the information contained in (T_{ij}, δ_{ij}) $(i = 1, \ldots, n,\ j = 1, 2)$ is through the counting process

$$N_i(t) = \sum_{j=1}^{2} N_{ij}(t) = \sum_{j=1}^{2} I\{T_{ij} \leq t, \delta_{ij} = 1\}$$

and the risk process

$$Y_i(t) = \sum_{j=1}^{2} Y_{ij}(t) = \sum_{j=1}^{2} I\{T_{ij} \geq t\}.$$

It then follows that

$$\frac{[\{1+\theta N_i(t-)\}Y_i(t)]^{\Delta N_i(t)}}{(1+\theta\int_0^\infty Y_i(t)dt)^{1/\theta+N(\infty)}} = \begin{cases} \{1+\theta(T_{i1}+T_{i2})\}^{-1/\theta-D_i} & \text{for } D_i = 0 \text{ or } 1 \\ (1+\theta)\{1+\theta(T_{i1}+T_{i2})\}^{-1/\theta-D_i} & \text{for } D_i = 2 \end{cases}.$$

This shows that $L_n(\theta)$ can be written in the format of the log likelihood expression in Murphy and van der Vaart (1997).

THEOREM 3. *Assume that, for n tending to infinity, the limiting Fisher information G is positive. The likelihood ratio statistic $\mathcal{L}_n$ for testing the one-sided heterogeneity hypothesis (3) in the shared gamma frailty model with constant baseline hazard has an asymptotic null distribution* $\frac{1}{2}\chi^2_0 + \frac{1}{2}\chi^2_1$.

The proof is stated in Section 7. Verifying the crucial assumption on the positiveness of the Fisher information corresponds to obtaining first, second, and third moments of $T_{i1}+T_{i2}$ (see equation (11) in the proof). An explicit calculation of these moments requires information on the censoring proportion and the joint distribution of (C_{i1}, C_{i2}). In contrast to the previous theorems, the results of Vu and Zhou (1997) are not applicable to censored data.

3 One-sided score tests

If parameters constrained under the null hypothesis belong to the interior of the parameter space, it is well known that likelihood ratio, Wald and score statistics have asymptotically the same distribution under the null hypothesis. Under inequality constraints in the alternative hypothesis this is no longer true. A score statistic is not uniquely defined, see Silvapulle and Silvapulle (1995). Robertson, Wright and Dykstra (1988, pp. 320–321) propose a score statistic which has the disadvantage of requiring estimation of model parameters both under the null and alternative hypothesis. Silvapulle and Silvapulle (1995) propose a different score-type statistic which only requires estimation under the null hypothesis. Under mild regularity conditions, they obtain that under the null hypothesis asymptotically the score statistic follows the same mixture distribution as the likelihood ratio statistic.

We first state the general form of the score statistic to test the heterogeneity hypothesis. Partition the Fisher information matrix $G(\nu)$, evaluated at $\theta = 0$, such that the upper left block corresponds to the parameter θ constrained to zero under the null hypothesis and the lower right block is defined by the nuisance parameters ν. Specifically,

$$G(\nu) = \begin{pmatrix} G_{00}(\nu) & G_{01}(\nu) \\ G_{01}^T(\nu) & G_{11}(\nu) \end{pmatrix}.$$

Further, define $G^{00}(\nu) = \{G^{-1}(\nu)\}_{00} = \{G_{00}(\nu) - G_{01}(\nu)G_{11}^{-1}(\nu)G_{01}^T(\nu)\}^{-1}$, let $\hat{\nu}$ be the

maximum likelihood estimator of the nuisance parameters under the null hypothesis and let $S_{n,\theta}(0,\hat{\nu})$ denote the first component of the score vector consisting of the partial derivative of the log likelihood with respect to θ evaluated at $(0,\hat{\nu})$. The other score components are zero by definition of maximum likelihood estimation. We now reformulate Theorem 1 of Silvapulle and Silvapulle (1995) for the one-sided hypothesis of heterogeneity.

THEOREM 4. [Silvapulle and Silvapulle, 1995] *Assume that there exists non-singular matrices $V(\tau)$ and $W(\tau)$ such that as $n \to \infty$, $n^{-1/2}S_n(\tau) \to_d N(0, V(\tau))$ and for any $a > 0$, $\sup\left[n^{-1/2}\{S_n(\tau + n^{-1/2}h) - S_n(\tau)\} + W(\tau)h : \|h\| \leq a\right] = o_P(1)$. Then the score statistic $\mathcal{S}_n = Q^t(A_{00})^{-1}Q - \inf\{(Q-b)^t(A_{00})^{-1}(Q-b) : b \geq 0\} = \mathcal{L}_n + o_P(1)$, where $Q = n^{-1/2}W^{00}S_n(0,\hat{\nu})$ with $W^{00} = (W_{00} - W_{01}W_{11}^{-1}W_{10})^{-1}$ and $A = \{W^t(\tau)V(\tau)^{-1}W(\tau)\}^{-1}$.*

With $V(\tau) = W(\tau)$ equal to the Fisher information matrix G, $A_{00} = G^{00}(\nu)$. Since A_{00} is one-dimensional, the score statistic with parameters estimated under the null hypothesis simplifies to

$$\begin{aligned}\mathcal{S}_n &= n^{-1}G^{00}(0,\hat{\nu})S_{n,\theta}^2(0,\hat{\nu}) - n^{-1}G^{00}(0,\hat{\nu})\inf[\{S_{n,\theta}(0,\hat{\nu}) - b\}^2 : b \geq 0] \\ &= n^{-1}G^{00}(0,\hat{\nu})\max\{0, S_{n,\theta}(0,\hat{\nu})\}^2.\end{aligned}$$

This structure of the score test invites to appreciate the asymptotic distribution result; either $\mathcal{S}_n(0,\hat{\nu}) = 0$ which occurs with probability 1/2 or $\mathcal{S}_n = n^{-1}G^{00}(0,\hat{\nu})S_{n,\theta}^2(0,\hat{\nu})$, which has asymptotically a chi-square distribution with one degree of freedom.

COROLLARY 1. *(i) For a shared gamma frailty model with exponential baseline hazard a score statistic for testing the heterogeneity hypothesis (2) is given by*

$$\mathcal{S}_n = \frac{1}{3n^2}\max\{0, S_{n,\theta}(0,\hat{\eta})\}^2.$$

(ii) For a Weibull distribution as the baseline hazard function a score statistic for testing the heterogeneity hypothesis (2) is given by

$$\mathcal{S}_n = \frac{\pi^2}{3n^2(\pi^2-4)}\max\{0, S_{n,\theta}(0,\hat{\nu})\}^2.$$

For both models the corresponding score statistic has, under the null hypothesis, asymptotic distribution $\frac{1}{2}\chi_0^2 + \frac{1}{2}\chi_1^2$.

More details on the calculations are given in Section 7.

This general recipe also applies for the situation with added covariates. Here we substitute an estimator of the Fisher information matrix G that is consistent under the null hypothesis. The data analysis in Section 5 employs the empirical Fisher information matrix with maximum likelihood estimators of (η, β) inserted.

4 The positive stable frailty model

Two major classes of frailty models are currently in use. The shared gamma frailty model we discussed above and one where the frailty variables follow a stable distribution. An extension to the stable frailty model requires techniques to deal with an infinite valued Fisher information matrix.

In the positive stable frailty model, the joint survival function for a bivariate observation is $S(t_1, t_2) = \exp\{-\lambda^{\tilde\theta}(t_1^\gamma + t_2^\gamma)^{\tilde\theta}\}$ where $0 < \tilde\theta \leq 1$. The corresponding log likelihood for the data reads as follows: $\tilde{\mathcal{L}}_n(\tilde\theta, \lambda, \gamma) =$

$$\sum_{i=1}^{n}[-\lambda^{\tilde\theta}(T_{i1}^\gamma + T_{i2}^\gamma)^{\tilde\theta} + \log(\lambda^{\tilde\theta}\tilde\theta(T_{i1}^\gamma + T_{i2}^\gamma)^{\tilde\theta-2}\gamma^2 T_{i1}^{\gamma-1}T_{i2}^{\gamma-1}) + \log\{\lambda^{\tilde\theta}(T_{i1}^\gamma + T_{i2}^\gamma)^{\tilde\theta} - (\tilde\theta - 1)\}]$$

Independence corresponds here to $\tilde\theta = 1$ and the relevant testing problem for heterogeneity is $H_0 : \tilde\theta = 1$ versus $H_a : \tilde\theta \in (0, 1)$. When λ and γ are known, for example $\lambda = \gamma = 1$, the score value at $\tilde\theta = 1$,

$$\tilde S_n(1) = \frac{d\tilde{\mathcal{L}}_n(1)}{d\tilde\theta} = \sum_{i=1}^{n}[1 + \{2 - (T_{i1} + T_{i2})\log(T_{i1} + T_{i2}) - (T_{i1} + T_{i2})^{-1}].$$

Although this function is different from the score function considered in Tawn (1988), the essence of the problem is the same: the infinite variance of the term $(T_{i1} + T_{i2})^{-1}$. Following arguments as in Tawn (1988), in particular using Feller (1971), we derive that

$$2^{1/2}(n\log n)^{-1/2}\tilde S_n(1) \to_d N(0, 1).$$

Hence $2(n\log n)^{-1}\max\{0, \tilde S_n(1)\}^2 \to_d \chi_0^2/2 + \chi_1^2/2$. This also is the main ingredient to derive the same asymptotic distribution for the likelihood ratio test statistic. For the situation where nuisance parameters (such as λ and γ) are present, methods as in Section 8 of Tawn (1988) can be used to derive the asymptotic distributions of test statistics for heterogeneity.

5 Data example

We use the Diabetic Retinopathy Study (Huster et al. 1989) to test for heterogeneity by considering time to blindness in each eye of 197 patients with diabetic Retinopathy. One eye of each patient is randomly selected for treatment and the other eye is observed without treatment. The data are bivariate right censored data with a treatment indicator as covariate.

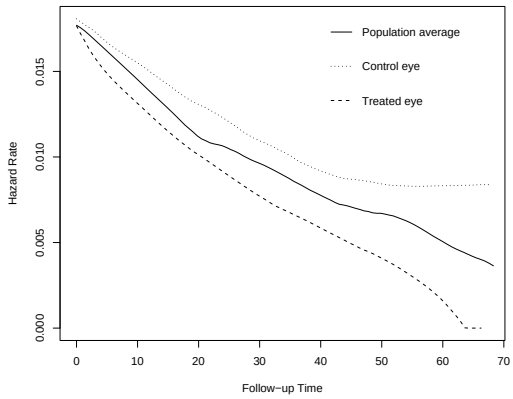


Figure 2. Nonparametric hazard function estimates for data from the Diabetic Retinopathy study.

The monotone decreasing trend of the nonparametric hazard function estimates (Figure 2) motivates that the Weibull baseline is an acceptable parametric way to describe the data. The Weibull parameters are λ and γ. We obtain the following results:

Table 1. Diabetic Retinopathy Study: parameter estimates (standard errors).

	Parameter			
	θ	λ	γ	β
Full model:	0.712 (0.145)	0.011 (0.190)	0.888 (0.006)	0.382 (0.046)
Null model:	—	0.015 (0.126)	0.799 (0.005)	0.280 (0.027)

The log likelihood values for null and full model are -846.499 and -841.272 respectively. This gives the observed value 10.454 of the likelihood ratio statistic. The corresponding p-value for the one-sided likelihood ratio test equals 0.0006.

The presence of heterogeneity is confirmed by the construction of a profile likelihood based confidence interval for θ. For a given value of θ we maximize $\mathcal{L}_n(\theta, \lambda, \beta)$ with respect to λ and β: for a given θ we denote these maximizers as $\lambda(\theta)$ and $\beta(\theta)$. We then obtain

the profile likelihood $\mathcal{L}_n(\theta, \lambda(\theta), \beta(\theta))$ and follow the method similar as explained in Morgan (1992) to obtain the profile likelihood based confidence interval. From Figure 3 it is clear that the confidence interval for θ does not contain zero, hence heterogeneity is present in the data; 95% confidence limits are (0.32, 1.20).

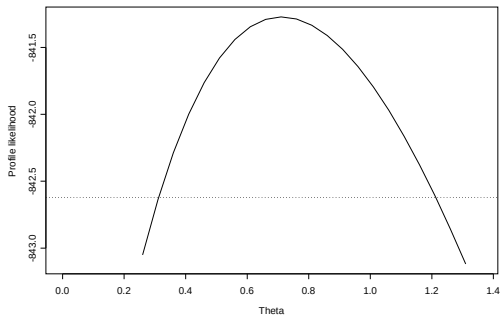


Figure 3. Profile likelihood $\mathcal{L}_n(\theta, \lambda(\theta), \beta(\theta))$ for data from the Diabetic Retinopathy study.

6 Discussion

In Section 2.3 we discussed one-sided heterogeneity tests in frailty models for censored data. For clarity we restricted attention to the situation where the heterogeneity parameter is the only unknown parameter. The likelihood expressions given in Murphy (1995) and Murphy and van der Vaart (1997) will be useful to extend our Theorem 3 to more complex parametric and semi-parametric frailty models. The related study of score tests for censored data is a further interesting theme. The discussion will be slightly more complicated than the one given in Section 3 since the censored data score vector does not have mean zero (see for example Theorem 2 in Murphy, 1995). One way to define a score statistic would be to start with the centered score vector by subtracting its mean value and then proceed similarly as in Section 3. If not the empirical mean but the population mean is used, this approach requires information on the censoring mechanism. See also Commenges and Anderson (1995) for a related two-sided score test.

A first step to study testing in semiparametric shared gamma frailty models, has been made by Vu and Knuiman (2002). Simulation methods are used to obtain statistical properties of the estimators of frailty variances. A relevant further step would be a mathematical

treatment of the semiparametric model, to counterpart and confirm the results of their simulation study, hereby extending our results for parametric models.

Wald-type test statistics for testing hypothesis (2) may be employed as well. Robertson, Wright and Dykstra (1988) construct a Wald statistic for the situation where the alternative hypothesis is described by inequalities. Their test statistic requires estimation of model parameters under both the null and alternative hypothesis. Sen and Silvapulle (2002) state a Wald test statistic as a difference of the minimum of two quadratic forms which has under the null hypothesis the same asymptotic distribution as the score and likelihood ratio statistic. For more details, see the recent review paper by Sen and Silvapulle (2002).

An interesting theme for further work is to study the distributional behaviour of the tests for heterogeneity under local alternatives converging to the null hypothesis at the rate $n^{-1/2}$. As in two-sided testing problems, it is expected that the test statistics will have the same power characteristics under these local circumstances.

A further relevant issue is to provide information on good finite sample approximations of the mixing properties, i.e., can we improve the asymptotic 50:50 mixture of the χ^2_0 and the χ^2_1 by finding mixing proportions that depend on the information of the sample size? In a setting of regression spline mixed models, Claeskens (2004) calculates finite sample approximations to the mixing probabilities. In the frailty models currently under consideration the situation is more complex by the presence of nuisance parameters under the null hypothesis. Bootstrapping the distribution of the test statistic can provide another alternative to the asymptotic distribution.

7 Proofs

7.1 Proof of Theorem 1

The proof follows Vu and Zhou (1997). It is straightforward to verify that L_n, components of the score vector $S_n(\tau)$ and entries of the matrix of negative second derivatives $F_n(\tau)$ are continuous and finite on a neighborhood of the true parameter value $(0, \nu_0)$. For the derivatives with respect to θ boundedness follows by an expansion of the logarithmic function

in the derivatives. In the full likelihood model $E[S_n(0,\nu)] = 0$. For the shared gamma frailty model with Weibull baseline hazard

$$G_n(\nu) = nG(\eta) = n \begin{pmatrix} 5 & 2 & 2(2-\gamma_e+\eta) \\ 2 & 2 & 2(1-\gamma_e+\eta) \\ 2(2-\gamma_e+\eta) & 2(1-\gamma_e+\eta) & \pi^2/3 + 2(1-\gamma_e+\eta)^2 \end{pmatrix} \tag{4}$$

with γ_e the Euler constant. Note that $\det\{G(\eta)\} = 2\pi^2 - 8 > 0$. The matrix $G(\eta)$ is symmetric and positive definite and therefore has for every fixed value of η, three positive real eigenvalues. From this it follows that $\liminf_{n\to\infty} \beta_{\min}\{G_n(\nu)\}/\beta_{\max}\{G_n(\nu)\} > 0$, where $\beta_{\min}(A)$ and $\beta_{\max}(A)$ denote the smallest and the largest eigenvalue of a symmetric positive definite matrix A. The convexity of the parameter space $\Theta_1 = (0,\infty) \times \mathbb{R} \times \mathbb{R}$ implies Chernoff regularity (Geyer, 1994). The arguments above are sufficient to conclude that there exists a closed cone C_{Θ_1} with vertex at $(0,\nu_0)$ such that $C_{\Theta_1} \subseteq \Theta_1$ and for every closed neighbourhood N of $(0,\nu_0)$, $C_{\Theta_1} \cap N = \Theta_1 \cap N$, see Vu & Zhou (1997). The approximating cones for Θ_0 and Θ_1 are $C_{\Theta_0} \equiv \Theta_0$ and $C_{\Theta_1} \equiv \Theta$. The transformed cones, used to obtain the asymptotic distribution of the likelihood ratio test, are (for $j = 0, 1$): $\widetilde{C}_{n,\Theta_j} = \left\{(\widetilde{\theta},\widetilde{\eta},\widetilde{\alpha}) = G_n^{T/2}(\nu)(\theta,\eta,\alpha)^T \text{ with } (\theta,\eta,\alpha) \in C_{\Theta_j}\right\}$. The matrices $G_n^{1/2}(\nu)$ and $G_n^{T/2}(\nu)$ are the left and the corresponding right Cholesky square root of $G_n(\nu)$. A direct calculation shows that $G_n^{T/2}(\nu) = n^{1/2}G^{T/2}(\eta)$ with

$$G^{T/2}(\eta) = \begin{bmatrix} 5^{1/2} & 2/5^{1/2} & 2(2-\gamma_e+\eta)/5^{1/2} \\ 0 & (6/5)^{1/2} & 2^{1/2}(1-3\gamma_e+3\eta)/(15)^{1/2} \\ 0 & 0 & \{(\pi^2-4)/3\}^{1/2} \end{bmatrix}$$

We therefore have

$$\begin{aligned} \widetilde{C}_{n,\Theta_0} &= \left\{(\widetilde{\theta},\widetilde{\eta},\widetilde{\alpha}) : \widetilde{\theta} - (2/3)^{1/2}\widetilde{\eta} - 20^{1/2}\{3(\pi^2-4)\}^{-1/2}\widetilde{\alpha} = 0\right\} \equiv \widetilde{C}_{\Theta_0} \\ \widetilde{C}_{n,\Theta_1} &= \left\{(\widetilde{\theta},\widetilde{\eta},\widetilde{\alpha}) : \widetilde{\theta} - (2/3)^{1/2}\widetilde{\eta} - 20^{1/2}\{3(\pi^2-4)\}^{-1/2}\widetilde{\alpha} \geq 0\right\} \equiv \widetilde{C}_{\Theta_1}, \end{aligned}$$

not dependent on the sample size n.

We now show that

$$\sup_{\tau\in N_n(A)} \|G_n^{-1/2}(\nu_0)F_n(\tau)G_n^{-T/2}(\nu_0) - I_3\|_1 = o_P(1) \tag{5}$$

with P shorthand notation for P_{τ_0} where $\tau_0 = (0, \nu_0) \in \Theta_0$. In the above $\|W\|_1$ denotes the sum of the absolute values of the elements of a matrix W, and for $(0, \nu_0)$ the true parameter value,

$$N_n(A) = \left\{\tau = (\theta, \eta, \alpha) : (\theta, \eta - \eta_0, \alpha - \alpha_0) G_n(\nu_0) (\theta, \eta - \eta_0, \alpha - \alpha_0)^T \leq A^2, \tau \in \Theta\right\}.$$

Write

$$\begin{aligned} & G_n^{-1/2}(\nu_0) F_n(\tau) G_n^{-T/2}(\nu_0) - I_3 \\ & \quad = n^{-1} G^{-1/2}(\eta_0) \{F_n(0, \nu_0) - G_n(\nu_0)\} G^{-T/2}(\eta_0) \\ & \qquad + n^{-1} G^{-1/2}(\eta_0) \{F_n(\tau) - F_n(0, \nu_0)\} G^{-T/2}(\eta_0). \end{aligned} \tag{6}$$

For matrices W_1 and W_2, $\|W_1 W_2\|_1 \leq \|W_1\|_1 \|W_2\|_1$. Since $\|G^{-1/2}(\eta_0)\|_1 = \|G^{-T/2}(\eta_0)\|_1 \leq C(\eta_0)$, with $0 < C(\eta_0) < \infty$, (5) follows since $\|F_n(0, \nu_0)/n - G(\eta_0)\|_1 = o_P(1)$, through a componentwise application of the law of large numbers, and since $\sup_{\tau \in N_n(A)} \|\{F_n(\tau) - F_n(0, \nu_0)\}/n\|_1 = o_P(1)$. This also is obtained componentwise; we show the proof for the $[2,2]$-entry: $n^{-1}\{F_n(\tau) - F_n(0, \nu_0)\}_{[2,2]} = n^{-1} \sum_{i=1}^n H_{22}(T_i, \tau)$ where, with $U_i = T_{i1}^{e^{-\alpha}} + T_{i2}^{e^{-\alpha}}$,

$$H_{22}(T_i, \tau) = -\theta e^{-2\eta}(1 + 2\theta) \frac{U_i^2}{(1 + \theta e^{-\eta} U_i)^2} + e^{-\eta}(1 + 2\theta) \frac{U_i}{1 + \theta e^{-\eta} U_i} - e^{-\eta_0} U_i.$$

Note that $H_{22}(T_i, \tau_0) \equiv 0$. There exists a fixed positive integer n_0 such that for all $n \geq n_0$

$$\sup_{\tau \in N_n(A)} e^{-\alpha} \leq K \equiv 2(e^{-\alpha_0} + 1)$$

and, for some constant $D > 0$, $|H_{22}(T_i, \tau)| < D(T_{i1}^{2K} + T_{i2}^{2K})$. With $\mu(\tau) = E_{\tau_0} H(T_i, \tau)$ we have by the dominated convergence theorem that $\lim_{\tau \to \tau_0} \mu(\tau) = \mu(\tau_0) \equiv 0$. Now the result follows since

$$\sup_{\tau \in N_n(A)} \left| \frac{1}{n} \sum_{i=1}^n H_{22}(T_i, \tau) \right| \leq \sup_{\tau \in N_n(A)} \left| \frac{1}{n} \sum_{i=1}^n H_{22}(T_i, \tau) - \mu(\tau) \right| + \sup_{\tau \in N_n(A)} |\mu(\tau)| = o_P(1). \tag{7}$$

An application of Theorem 16(a) in Ferguson (1996), p. 108 implies indeed that the first term in the right-hand side of (7) is $o_P(1)$ (uniform law of large numbers); elementary analysis implies that the second term in the right-hand side of (7) is $o(1)$.

Similar proofs hold for all the other entries of $\{F_n(\tau) - F_n(0, \nu_0)\}/n$.

Since (T_{i1}, T_{i2}), $i = 1, \ldots, n$, are independent and identically distributed vectors, an application of classical multivariate central limit theory gives that $G_n^{-1/2} S_n(0, \nu_0) \rightarrow_d Z = (Z_1, Z_2, Z_3) \sim N(0, I_3)$.

Since by the above arguments the Vu and Zhou (1997) conditions (A1)-(A3) and (B1)-(B5) are valid, an application of their Theorem 2.2 gives that the asymptotic null distribution of $\mathcal{L}_n$, the likelihood ratio statistic, is the same as the distribution of

$$\inf_{\widetilde{\tau} \in \widetilde{C}_{\Theta_0}} |Z - \widetilde{\tau}|^2 - \inf_{\widetilde{\tau} \in \widetilde{C}_{\Theta_1}} |Z - \widetilde{\tau}|^2 \tag{8}$$

where $\widetilde{\tau} = (\widetilde{\theta}, \widetilde{\eta}, \widetilde{\alpha})$.

From the definitions of $\widetilde{C}_{\Theta_0}$ and $\widetilde{C}_{\Theta_1}$ we have

$$\inf_{\widetilde{\tau} \in \widetilde{C}_{\Theta_0}} |Z - \widetilde{\tau}|^2 = (Z_1 + aZ_2 + bZ_3)^2/(1 + a^2 + b^2) \tag{9}$$

with $a = -(2/3)^{1/2}$ and $b = -20^{1/2}\{3(\pi^2 - 4)\}^{-1/2}$. This implies that the random variable in (9) has a χ_1^2 distribution. We further have

$$\inf_{\widetilde{\tau} \in \widetilde{C}_{\Theta_1}} |Z - \widetilde{\tau}|^2 = \begin{cases} 0 & Z \in \widetilde{C}_{\Theta_1} \\ (Z_1 + aZ_2 + bZ_3)^2/(1 + a^2 + b^2) & Z \notin \widetilde{C}_{\Theta_1} \end{cases} \tag{10}$$

Moreover we have $P(Z \in \widetilde{C}_{\Theta_1}) = 0.5$. This, together with (8) – (10) implies that the asymptotic distribution of the likelihood ratio test is $0.5\chi_0^2 + 0.5\chi_1^2$.

7.2 Proof of Theorem 2

This follows along the lines of the proof of Theorem 1. We indicate the main differences. Write $G_n(\eta, \beta) = n(n^{-1}G_n(\eta, \beta))$. The conditions on x and β assure that $G(\eta, \beta) = \lim_{n\to\infty} n^{-1}G_n(\eta, \beta)$ exists, and that the symmetric matrix $G(\eta, \beta)$ is positive definite. Its Cholesky decomposition leads to cones $\tilde{C}_{n,\Theta_j}$ $(j = 0, 1)$ of which the limiting cones are defined by using the matrix $G(\eta, \beta)$. The remaining part of the proof holds under the boundedness assumption on β and on the covariate x.

7.3 Proof of Theorem 3

The limiting Fisher information G is calculated as $\lim_{n\to\infty} E[-(d^2/d\theta^2)L_n(0)/n]$ where

$$\frac{d^2}{d\theta^2} L_n(0) = -\frac{2}{3} \sum_{i=1}^{n} U_i^3 + \sum_{i:D_i=1} U_i^2 + 2 \sum_{i:D_i=2} U_i^2 - N_2, \tag{11}$$

with $U_i = T_{i1} + T_{i2}$. Since $G > 0$, the log likelihood $L_n(\theta)$ is concave in a closed neighborhood $\mathcal{N}$ of zero. Define $\hat{\theta}_n = \arg\max_{\theta \in \mathcal{N}} L_n(\theta)$. For $\hat{\theta}_n$ the following properties are immediate from the more general Theorem 2 in Murphy (1994) and Theorem 1 in Murphy (1995): under the null hypothesis the likelihood estimator $\hat{\theta}_n$ strongly converges to zero and the distribution of $\sqrt{n}\hat{\theta}_n$ tends to a normal limit with zero mean and bounded limit variance $G^{-1} > 0$. This implies that $\lim_{n\to\infty} P(\hat{\theta}_n \leq 0) = 0.5$. Take $x \geq 0$. Under the null hypothesis

$$P(\mathcal{L}_n \leq x) = P(\mathcal{L}_n \leq x | \hat{\theta}_n > 0) P(\hat{\theta}_n > 0) + P(\mathcal{L}_n \leq x | \hat{\theta}_n \leq 0) P(\hat{\theta}_n \leq 0).$$

The condition on the second derivative implies the existence of a neighbourhood $[-\varepsilon, \varepsilon]$, $\epsilon > 0$, where $L_n(\theta)$ is concave. Therefore, for n sufficiently large, $\hat{\theta}_n \leq 0$ implies $\mathcal{L}_n = 0$ and therefore under the null hypothesis $P(\mathcal{L}_n \leq x | \hat{\theta}_n \leq 0) = 1$.

For $\hat{\theta}_n > 0$, a Taylor expansion yields

$$\mathcal{L}_n = 2(L_n(\hat{\theta}_n) - L_n(0)) = -\hat{\theta}_n^2 G_n(\tilde{\theta}_n)$$

with $\tilde{\theta}_n$ an intermediate point in $(0, \hat{\theta}_n)$.

Since $-n^{-1} G_n(\tilde{\theta}_n) \to G$ in probability as $n \to \infty$ it follows that the variable $Z_n = n^{1/2}\hat{\theta}_n \{-n^{-1} G_n(\tilde{\theta}_n)\}^{1/2} \to_d Z$ with $Z \sim N(0,1)$ and

$$P(\mathcal{L}_n \leq x | \hat{\theta}_n > 0) P(\hat{\theta}_n > 0) = P(0 < Z_n \leq x^{1/2}) \to \frac{1}{2} P(Z^2 \leq x).$$

Hence, under the null hypothesis,

$$P(\mathcal{L}_n \leq x) \to \frac{1}{2} P(Z^2 \leq x) + \frac{1}{2}.$$

7.4 Proof of Corollary 1

For the case of a shared gamma frailty model with an exponential baseline hazard there is only the nuisance parameter η (or $\lambda = \exp(-\eta)$). For this special case, with Fisher information matrix

$$G_n = n \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix},$$

we have that $G_n^{00} = (3n)^{-1}$, not dependent on any nuisance parameters, and hence we obtain the following score statistic

$$\mathcal{S}_n = \frac{1}{3n^2}\{S_{n,\theta}(0,\hat{\eta})\}^2 - 3n \inf_{b\geq 0}\left\{\left(\frac{1}{3n^{3/2}}S_{n,\theta}(0,\hat{\eta}) - b\right)^2\right\} = \frac{1}{3n^2}\max\{0, S_{n,\theta}(0,\hat{\eta})\}^2$$

For the Weibull baseline hazard the nuisance parameter is $\nu = (\eta, \alpha)$ and the Fisher information matrix is $G_n(\nu)$ as given in (4), from which it is deduced that $G_n^{00} = \pi^2/\{3n(\pi^2-4)\}$. Hence the resulting score statistic is obtained as given in Theorem 2.

Acknowledgements

Nguti acknowledges partial support from the Belgian Technological Cooperation (BTC). Janssen is partly supported by the Ministry of the Flemish Community (Project BIL00/28, International Scientific and Technological Cooperation). The Belgian Federal Science Policy Office provided partial support for Janssen (Interuniversity Attraction Pole Programme P5/24) and for Claeskens. Part of the research of Claeskens has been performed at Texas A&M university.

References

Bjarnason, H. & Hougaard, P. (2000). Fisher information for two gamma frailty bivariate Weibull models. *Lifetime Data Anal.* **6**, 59–71.

Chernoff, H. (1954). On the distribution of the likelihood ratio. *Ann. Math. Statist.*, **25**, 573–578.

Claeskens, G. (2004). Restricted likelihood ratio lack of fit tests using mixed spline models. *J. Roy Statist. Assoc. (Ser. B)*, too appear.

Clayton, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika* **65**, 141–151.

Clayton, D.G. & Cuzick, J. (1985). Multivariate generalizations of the proportional hazards model (with discussion). *J. R. Statist. Soc. A* **148**, 82–117.

Commenges, D. & Andersen P.K. (1995). Score test of homogeneity for survival data. *Lifetime data Anal.* **1**, 145–156.

Feller, W. (1971). *An introduction to probability theory and its applications.* 2nd edition. Wiley, New York.

Ferguson, T. (1996). *A course in large sample theory.* Chapman and Hall, London.

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

Hougaard, P. (2000). *Analysis of multivariate survival data.* Springer, New York.

Huster, W.J., Brookmeyer, R. & Self, S.G. (1989). Modelling paired survival data with covariates. *Biometrics*, **45**, 145-156.

Morgan, B.J.T. (1992). *Analysis of quantal response data*, Chapman and Hall, London.

Murphy, S.A. (1994). Consistency in a proportional hazards model incorporating a random effect. *Ann. Statist.* **22**, 712–731.

Murphy, S.A. (1995). Asymptotic theory for the frailty model. *Ann. Statist.* **23**, 182–198.

Murphy, S.A. & van der Vaart, A.W. (1997). Semiparametric likelihood ratio inference. *Ann. Statist.* **25**, 1471–1509.

Nelsen, R.G. (1999). *An introduction to copulas.* Lecture notes in Statistics, 139, Springer, New York.

Robertson, T., Wright, F.T. & Dykstra, R.L. (1988). *Order-restricted statistical inference.* Wiley, New York.

Self, S.G. & Liang, K.-Y. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non-standard conditions. *J. Amer. Statist. Assoc.* **82**, 605–610.

Sen, P.K. & Silvapulle, M.J. (2002). An appraisal of some aspects of statistical inference under inequality constraints. *J. Statist. Planning Inference* **107**, 3–43.

Silvapulle, M.J. & Silvapulle, P. (1995). A score test against one-sided alternatives. *J. Amer. Statist. Assoc.* **90**, 342–349.

Tawn, J.A. (1988). Bivariate extreme value theory: models and estimation. *Biometrika* **75**, 397–415.

Therneau, T.M. & Grambsch, P.M. (2000). Modeling survival data: extending the Cox model. Springer, New York.

Verbeke, G. & Molenberghs, G. (2003). The use of score tests for inference on variance components. *Biometrics* **59**, 254–262.

Vu, H.T.V. & Knuiman, M.W. (2002). Estimation in semiparametric marginal shared gamma frailty models. *Austr. N.-Zeal. J. Statist.*, **44**, 489–501.

Vu, H.T.V. & Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions. *Ann. Statist.* **25**, 897–916.