David Kraus

Checking proportional rates in the two-sample transformation model

*Kybernetika*, Vol. 45 (2009), No. 2, 261--278

Persistent URL: `http://dml.cz/dmlcz/140076`

KYBERNETIKA — VOLUME *45* (2009), NUMBER 2, PAGES 261 – 278

# CHECKING PROPORTIONAL RATES IN THE TWO–SAMPLE TRANSFORMATION MODEL

David Kraus

Transformation models for two samples of censored data are considered. Main examples are the proportional hazards and proportional odds model. The key assumption of these models is that the ratio of transformation rates (e. g., hazard rates or odds rates) is constant in time. A method of verification of this proportionality assumption is developed. The proposed procedure is based on the idea of Neyman's smooth test and its data-driven version. The method is suitable for detecting monotonic as well as nonmonotonic ratios of rates.

*Keywords:* Neyman's smooth test, proportional hazards, proportional odds, survival analysis, transformation model, two-sample test

*AMS Subject Classification:* 62N01, 62N03

## 1. INTRODUCTION

This paper deals with simple models for two samples (e. g., the control and treatment group) of survival data under random censorship. Various models have been proposed in the literature to describe the situation when the survival distributions in two samples differ. The aim of this paper is to develop new methods of assessment of fit for one class of these models, proportional rate models.

The most frequent model is the proportional hazards model which assumes that the ratio of the hazard rates $\alpha_1(t), \alpha_2(t)$ is constant over time, that is there exists a real constant $\beta$ such that $\alpha_2(t)/\alpha_1(t) = e^\beta$ for all $t$. The effect of treatment on the failure rate remains the same in the course of time. In some situations the effect of treatment decays for large times and hazard rates converge to each other. A popular model for this situation is the proportional odds model. Let $A_k(t) = \int_0^t \alpha_k(s)\,\mathrm{d}s$ be the cumulative hazard function and $S_k(t) = e^{-A_k(t)}$ the survival function in the $k$th sample, $k = 1, 2$. Denote $\Gamma_k(t) = (1 - S_k(t))/S_k(t)$ the odds function giving the odds of dying before time $t$ versus surviving up to $t$. The proportional odds model assumes $\Gamma_2(t)/\Gamma_1(t) = e^\beta$ for all times. A common feature of these two main examples is that they assume constancy of the ratio of some functions. It is important to check this assumption.

These two models are considered within a wider class of semiparametric linear transformation models as follows (for more details and references see, for instance, Bagdonavičius and Nikulin [5] or Martinussen and Scheike [15]). Let $S_\omega$ be a known survival function (of a continuous nonnegative variable $\omega$), let $A_\omega = -\log S_\omega$ be the corresponding cumulative hazard. Assume that there exists a continuous increasing function $G_k$ defined on the positive half-line with $G_k(0) = 0$ such that in the $k$th sample the survival function is $S_k(t) = S_\omega(G_k(t))$, and the cumulative hazard is $A_k(t) = A_\omega(G_k(t))$. The functions $G_k$ are called cumulative rates. Denote the (noncumulative) rate $g_k(t) = \mathrm{d}G_k(t)/\mathrm{d}t$ and the hazard function $\alpha_k(t) = \mathrm{d}A_k(t)/\mathrm{d}t$. These noncumulative functions are in the one-to-one relationship $\alpha_k(t) = \alpha_\omega(G_k(t))g_k(t) = q_\omega(A_k(t))g_k(t)$, where $q_\omega(t) = \alpha_\omega(A_\omega^{-1}(t))$.

It is assumed that the functions $G_1, G_2$ are proportional, i. e., there exists real $\beta$ such that $G_1(t) = G_0(t)$, $G_2(t) = e^\beta G_0(t)$ for all $t \in [0, \tau]$. The baseline cumulative rate $G_0$ is unknown and not specified parametrically. Denote by $R$ the survival time (distributed according to $S_k$ in the $k$th sample). It is easily verified that in the $k$th sample the transformed survival time $G_k(R)$ follows the distribution $S_\omega$. This implies the multiplicative model $G_0(R) = e^{-\beta z_k}\omega$, equivalently the linear model $\log G_0(R) = -\beta z_k + \log\omega$, where $z_k = 1[k = 2]$. That is, after the unknown transformation $\log G_0$ the survival times follow a location-shift model in the known error distribution of $\log\omega$.

Both main models, proportional hazards and proportional odds, fit in this framework.

The proportional hazards model is obtained for $\omega$ following the unit exponential distribution, that is $S_\omega(t) = e^{-t}$, $A_\omega(t) = t$, $\alpha_\omega(t) = q_\omega(t) = 1$. Then the cumulative rate $G_k$ is the cumulative hazard $A_k$, $g_k$ is the hazard rate $\alpha_k$, and the model for $\log G_0(R)$ is a location model in the extreme value distribution.

When $\omega$ comes from the log-logistic distribution with $S_\omega(t) = 1/(1+t)$, $A_\omega(t) = \log(1+t)$, $\alpha_\omega(t) = 1/(1+t)$ and $q_\omega(t) = e^{-t}$, we get the proportional odds model since the cumulative rate $G_k$ has the meaning of the odds function (because $G_k(t) = S_\omega^{-1}(S_k(t)) = (1 - S_k(t))/S_k(t)$). The rate $g_k(t) = \mathrm{d}G_k(t)/\mathrm{d}t$ may be called the odds-rate. The transformed time $\log G_0(R)$ has a shifted logistic distribution. In this model the hazard rates are $\alpha_k(t) = e^{\beta z_k} g_0(t)/(1 + e^{\beta z_k} G_0(t))$. Thus the hazard ratio $e^\beta(1 + G_0(t))/(1 + e^\beta G_0(t))$ converges to 1 as $t \to \infty$, and this convergence is monotonic (from above when $e^\beta > 1$, from below when $e^\beta < 1$). Therefore, this model is a popular alternative to proportional hazards when the hazards appear to approach to each other for large times.

Reliability theory provides a different view of transformation models. The function $S_\omega$ is called resource, and $G_k$ is the rate of resource usage. So in the proportional hazards model there are proportional rates of the exponential resource usage, in proportional odds the resource is log-logistic. Another example is a lognormal resource.

Linear transformation models are related to frailty models. Let $U$ be frailty variables, i. e., unobservable positive random variables with a known distribution with expectation 1 which act multiplicatively on the hazard rate. That is, the conditional hazard of observations in the $k$th sample is $\alpha_k(t|U = u) = e^{\beta z_k} g_0(t)u$. Then the marginal survival function is $S_k(t) = \mathsf{E}\, S_k(t|U) = \mathsf{E} \exp\{-e^{\beta z_k} G_0(t)U\} = L_U(e^{\beta z_k} G_0(t))$, where $L_U$ denotes the Laplace transform of the distribution of $U$. When $R$ comes from the $k$th sample, the survival function of $e^{\beta z_k} G_0(R)$ is $L_U$

(because $S_k(R)$ is uniformly distributed). Hence the Laplace transform $L_U$ of the frailty distribution equals the survival function $S_\omega$ of the error variable $\omega$ in the transformation model. Also, as the conditional (on $U = u$) proportional hazards model $e^{\beta z_k} g_0(t)u$ is the transformation model $\log G_0(R) = -\beta z_k - \log u + \log \omega_0$ with $\log \omega_0$ being extreme-value distributed, we see that $\log G_0(R)$ unconditionally follows a tranformation model with errors $\log \omega = -\log U + \log \omega_0$ (thus the error distribution is the distribution of the difference of an extreme-value variable and $\log U$, which are independent).

A model without frailties ($U = 1$ a.s.) has $L_U(t) = S_\omega(t) = e^{-t}$, thus it is a proportional hazards model. When frailties are unit exponential, $L_U(t) = S_\omega(t) = 1/(1+t)$, so the model is a proportional odds model. This agrees with the fact that the difference of two independent extreme-value variables ($\log \omega_0$ and $\log U$) is logistic. More generally, if frailties are gamma distributed with parameters $(1/v, 1/v)$ (expectation 1, variance $v$), it follows that $L_U(t) = S_\omega(t) = (1+vt)^{-1/v}$, $\alpha_\omega(t) = (1+vt)^{-1}$. This model is the proportional generalised odds model of Dabrowska and Doksum [8] (in this model $G_k$ are the generalised odds functions $v^{-1}(1-S_k(t)^v)/S_k(t)^v$, they are proportional, while the hazard rates $\alpha_k(t) = e^{\beta z_k} g_0(t)/(1 + v e^{\beta z_k} G_0(t))$ converge to each other).

Section 2 explains the simplified partial likelihood estimation procedure needed in subsequent considerations. In Section 3, I develop Neyman's smooth test of the proportional rates assumption, the main contribution of the paper. Section 4 reviews and extends some other testing methods. Smooth tests and other procedures are compared via simulations reported in Section 5. A real data illustration can be found in Section 6. Technical material (theorems and proofs) is deferred to Section 7 which closes the paper.

## 2. ESTIMATION PROCEDURE

Let data consist of pairs $(T_{j,i}, \delta_{j,i})$, $j = 1, 2$, $i = 1, \ldots, n_j$, where $T_{j,i} = \min(R_{j,i}, C_{j,i})$ are possibly censored survival times $R_{j,i}$ ($R_{j,i}$ are independent, with hazard function $\alpha_j$), $\delta_{j,i} = 1[T_{j,i} = R_{j,i}]$ are failure indicators, and censoring times $C_{j,i}$ are mutually independent and independent of $R_{j,i}$. The standard counting process notation is used. Set $N_{j,i}(t) = 1[T_{j,i} \le t, \delta_{j,i} = 1]$, $\bar N_j(t) = \sum_{i=1}^{n_j} N_{j,i}(t)$, $\bar N(t) = \bar N_1(t) + \bar N_2(t)$, $Y_{j,i}(t) = 1[T_{j,i} \ge t]$, $\bar Y_j(t) = \sum_{i=1}^{n_j} Y_{j,i}(t)$. Let these processes be observed on a finite interval $[0, \tau]$.

For estimation of $\beta$, I use the procedure of Bagdonavičius and Nikulin [4] based on a simplification of the partial likelihood as follows. The partial likelihood takes the form

$$\begin{aligned} C(\tau; \beta, A_1, A_2) &= \sum_{j=1}^{2} \sum_{i=1}^{n_j} \int_0^\tau \log\left( \frac{\lambda_{j,i}(t)}{\sum_{k=1}^{2} \sum_{l=1}^{n_k} \lambda_{k,l}(t)} \right) \mathrm{d}N_{j,i}(t) \\ &= \int_0^\tau \log[q_\omega(A_1(t))] \,\mathrm{d}\bar N_1(t) + \int_0^\tau \log[q_\omega(A_2(t)) e^\beta] \,\mathrm{d}\bar N_2(t) \\ &\qquad - \int_0^\tau \log[\bar Y_1(t) q_\omega(A_1(t)) + \bar Y_2(t) q_\omega(A_2(t)) e^\beta] \,\mathrm{d}\bar N(t). \end{aligned}$$

Here $A_1, A_2$ depend on $\beta$ which complicates differentiation when we want to derive a score equation. However, $A_1, A_2$ can be estimated directly without knowing $\beta$ by Nelson–Aalen estimators $\hat{A}_j(t) = \int_0^t \bar{Y}_j(s)^{-1} \mathrm{d}\bar{N}_j(s)$ computed separately in each sample. Therefore, we work with $C(\tau; \beta, \hat{A}_1, \hat{A}_2)$ instead of $C(\tau; \beta, A_1, A_2)$. Here it considerably simplifies calculations, especially when taking derivatives with respect to $\beta$. Then the score vector $\frac{\partial}{\partial\beta} C(\tau; \beta, \hat{A}_1, \hat{A}_2)$ is $U_1(\tau; \beta, \hat{A}_1, \hat{A}_2)$, where the score process equals

$$
\begin{aligned}
U_1(t; \beta, A_1, A_2) &= \bar{N}_2(t) - \int_0^t \frac{\bar{Y}_2(s) q_\omega(A_2(s)) e^\beta}{\bar{Y}_1(s) q_\omega(A_1(s)) + \bar{Y}_2(s) q_\omega(A_2(s)) e^\beta} \mathrm{d}\bar{N}(s) \\
&= \int_0^\tau \frac{\bar{Y}_1(s) q_\omega(A_1(s)) \bar{Y}_2(s) q_\omega(A_2(s))}{\bar{Y}_1(s) q_\omega(A_1(s)) + \bar{Y}_2(s) q_\omega(A_2(s)) e^\beta} \\
&\qquad \times \left( \frac{\mathrm{d}\bar{N}_2(s)}{\bar{Y}_2(s) q_\omega(A_2(s))} - e^\beta \frac{\mathrm{d}\bar{N}_1(s)}{\bar{Y}_1(s) q_\omega(A_1(s))} \right). \qquad (1)
\end{aligned}
$$

The estimator $\hat{\beta}$ of the parameter $\beta$ is defined as the maximiser of $C(\tau; \beta, \hat{A}_1, \hat{A}_2)$, that is, by concavity of $C$ as a function of $\beta$, the solution to $U_1(\tau; \beta, \hat{A}_1, \hat{A}_2) = 0$. (Here and further in the paper, the left-continuous version of the Nelson–Aalen estimator $\hat{A}_j(t-)$ is used in the integrand in $C$ and $U_1$ to preserve predictability.)

Note that for the proportional hazards model ($q_\omega \equiv 1$) this estimation procedure agrees with the usual partial likelihood method.

Having computed $\beta$, one can obtain a Breslow-type model-based estimator of $G_0$ ($= G_1$) in the form

$$
\hat{G}_0(t) = \int_0^t \frac{\mathrm{d}\bar{N}(s)}{\bar{Y}_1(s) q_\omega(\hat{A}_1(s)) + \bar{Y}_2(s) q_\omega(\hat{A}_2(s)) e^{\hat{\beta}}}.
$$

Before proceeding to main results we need to know that the simplified partial likelihood estimation procedure yields a consistent estimator of $\beta$. This is verified by Lemma 1 in Section 7.

Other estimation procedures have been developed for regression models with general covariates. Bagdonavičius and Nikulin [3] use the modified partial likelihood. Variants of this approach are reviewed in Section 8.2 of Martinussen and Scheike [15]. Murphy, Rossini and van der Vaart [16] propose the full nonparametric maximum likelihood estimation. Chen, Jin and Ying [6] develop a method based on the iterative solution of certain martingale estimating equations.

## 3. NEYMAN'S SMOOTH TEST

The general idea of Neyman's smooth test (see Rayner and Best [17]) is based on embedding the null model into a model where the departure from the null is expressed by a $d$-dimensional parameter. That is, the general alternative that the hypothesis does not hold is replaced by a $d$-dimensional alternative. In the present context the null model assumes that the ratio of rates is constant over time, i. e., $g_2(t) = e^\beta g_1(t)$. Thus the Neyman embedding is most conveniently and most naturally formulated in terms of these transformation rates $g_k(t) = \mathrm{d}G_k(t)/\mathrm{d}t$. Under the alternative model the logarithm of the time-varying rate ratio is expressed as a linear combination of some bounded basis functions $\psi_1(t), \dots, \psi_d(t)$, that is

$$g_2(t) = \exp\{\beta + \theta^\mathsf{T}\psi(t)\}g_1(t). \tag{2}$$

These functions must be linearly independent and independent of 1 (then the model is identifiable). The Neyman-type smooth test of goodness of fit of the proportional rate model is the score test of $\theta = 0$ versus $\theta \neq 0$ in (2).

In the proportional hazards model the formulation of the embedding in terms of the noncumulative rates $g_k$ is the obvious choice because $g_k$ is the hazard rate and the name of the model actually contains the word hazard. On the other hand, in the proportional odds model one can be tempted to work directly with the odds functions $G_k = \Gamma_k$. This is not a good idea because $G_k$ is an increasing (cumulative) function, thus in a model like $G_2(t) = \exp\{\beta + \theta^\mathsf{T}\psi(t)\}G_1(t)$ one would have to work with some monotonicity constraints. On the contrary, noncumulative rates may be arbitrary positive which poses no restrictions on $\beta$ and $\psi_j$ in (2).

The functions $\psi_j(t)$ are typically some standard basis functions on $[0, 1]$ in transformed time, i. e., of the form $\psi_j(t) = \phi_j(P(t)/P(\tau))$, where $\phi_j(u), u \in [0, 1]$, are, for instance, Legendre polynomials of order $1, \dots, d$, or cosines $\sqrt{2}\cos(j\pi u)$. The time-transformation $P(t)$ is a nondecreasing nonnegative continuous function with $P(0) = 0$, thus $P(t)/P(\tau)$ maps $[0, \tau]$ on $[0, 1]$. Although $P(t)$ can be any function with these properties, it should be related to the underlying distribution as its purpose is to make the course of time in some sense uniform and hence better exploit the flexibility of the shape of $\phi_j$, see [12] for a discussion. Therefore, in practice $P(t)$ is replaced by a data-dependent estimator $\hat{P}(t)$. Here I use $\hat{P}^*(t) = 1 - \exp\{-\hat{A}^*(t)\}$, where $\hat{A}^*(t)$ is the Nelson–Aalen estimator computed from the pooled sample. The quantity $\hat{A}^*(t)$ consistently estimates $A^*(t) = \int_0^t \bar{y}_1(s)/\bar{y}(s)\,\mathrm{d}A_1(s) + \int_0^t \bar{y}_2(s)/\bar{y}(s)\,\mathrm{d}A_2(s)$, where $\bar{y}_j(t) = a_j S_j(t)(1 - C_j(t))$ denotes the uniform limit in probability of $n^{-1}\bar{Y}_j(t)$, $C_j(t)$ is the distribution function of censoring times and $a_j \in (0, 1)$ is the limit of $n_j/n$ (see Section 7 for details). If the censoring distribution is the same in both samples ($C_1(t) = C_2(t)$), the limit of $\hat{P}^*(t)$ is the distribution function corresponding to the mixture of survival distributions $S_1, S_2$ with weights $a_1, a_2$, i. e., $P^*(t) = a_1(1 - S_1(t)) + a_2(1 - S_2(t))$. Thus $P^*(t)$ is the distribution of a 'typical' observation.

Now let us finally derive the score test of significance of $\theta$. If $C(\tau; \beta, \theta, \hat{A}_1, \hat{A}_2)$ denotes the simplified partial likelihood in the extended model (2), the score vector for inference about $\theta$ is $U_2(\tau; \beta, \theta, \hat{A}_1, \hat{A}_2) = \frac{\partial}{\partial\theta}C(\tau; \beta, \theta, \hat{A}_1, \hat{A}_2)$. The score test of significance of $\theta$ employs the score vector $U_2(\tau; \hat{\beta}, 0, \hat{A}_1, \hat{A}_2)$, denoted $U_2(\tau; \hat{\beta}, \hat{A}_1, \hat{A}_2)$ for short. Notice that

$$U_2(\tau; \beta, A_1, A_2) = \int_0^\tau \psi(t) U_1(\,\mathrm{d}t; \beta, A_1, A_2).$$

In Section 7, I show that the score $n^{-1/2}U_2(\tau; \hat{\beta}, \hat{A}_1, \hat{A}_2)$ is asymptotically (with $n \to \infty$) normal with mean zero and variance matrix consistently estimated by $n^{-1}\hat{\Xi}$

given by (8). Consequently, the distribution of the quadratic test statistic

$$T_d = U_2(\tau; \hat{\beta}, \hat{A}_1, \hat{A}_2)^{\mathsf{T}} \hat{\Xi}^{-1} U_2(\tau; \hat{\beta}, \hat{A}_1, \hat{A}_2)$$

is approximately $\chi^2$ with $d$ degrees of freedom. Large values of $T_d$ lead to rejection of the hypothesis.

I consider a data-driven version of Neyman's smooth test. The problem of choosing the suitable number of basis functions is addressed by the approach proposed by Ledwina in [13] based on a modification of Schwarz's selection rule. The number of basis functions is the maximiser of penalised score statistics, i. e., $S = \arg\max_{k=1,\dots,d}\{T_k - k \log n\}$. The data-driven test statistic is $T_S$. Under the null hypothesis, the selector $S$ converges in probability to 1, and thus $T_S$ is asymptotically $\chi^2$-distributed with one degree of freedom. This approximation is known to be inaccurate (seriously anticonservative). Therefore, a more accurate two-term approximation provided in [12, eq. (12)] is used.

## 4. OTHER TESTS

### 4.1. Komogorov–Smirnov-type test

A simple test can be based on the simplified partial likelihood score process $U_1(t; \hat{\beta}, \hat{A}_1, \hat{A}_2)$, $t \in [0, \tau]$. When the fit of the proportional rate model is good, this process fluctuates around zero. When the model is not valid, the score process is expected to be far from zero. This may be measured by the Kolmogorov–Smirnov-type statistic $\sup_{t\in[0,\tau]} |U_1(t; \hat{\beta}, \hat{A}_1, \hat{A}_2)|$. Wei [21] used this test for the two-sample proportional hazards model ($A_\omega(t) = t$), Bagdonavičius and Nikulin [4] extended it to general two-sample transformation models.

Bagdonavičius and Nikulin [4] proved that under the proportional rate model the score process is asymptotically Gaussian with mean zero. In the present paper this convergence is proved (Lemma 3 in Section 7) as an intermediate result for the proof of the asymptotic distribution of the Neyman test statistic. The process is of the bridge type (equal to zero at times 0 and $\tau$). In the special case of the proportional hazards model the score process converges to the Brownian bridge. In general, however, its limiting covariance structure is complicated and one has to resort to simulations. The standard resampling technique of Lin, Wei and Ying [14] (see also Martinussen and Scheike [15]) can be used as the martingale representation is available, see eqs. (6) and (7). We obtain simulated paths of the test process by replacing unobservable martingale increments $dM_{k,i}(t)$ at failure times by randomly generated independent standard normal variables.

### 4.2. Gill–Schumacher-type test

Gill and Schumacher [11] proposed a simple procedure for verifying proportionality of hazard functions in two samples. The idea is to compare two weighted estimators of the hazard ratio. Here I use their idea and extend this approach to the general transformation setting.

Consider a weight function $K(t), t \in [0,\tau]$, which is a nonnegative predictable process. Assume that $n^{-1}K(t)$ converges in probability to some deterministic function $k(t)$, uniformly in $t \in [0,\tau]$. Then the proportionality parameter $\eta = e^{\beta} = g_2(t)/g_1(t)$ may be estimated by

$$\hat{\eta} = \frac{\int_0^\tau K(t)\,\mathrm{d}\hat{G}_2(t)}{\int_0^\tau K(t)\,\mathrm{d}\hat{G}_1(t)}.$$

The variable $\hat{\eta}$ converges to $\{\int_0^\tau k(t)g_2(t)\,\mathrm{d}t\}/\{\int_0^\tau k(t)g_1(t)\,\mathrm{d}t\} = \eta$. Now consider weights $K_1, K_2$ with the same properties as $K$. Denote $\hat{\eta}_j = \hat{\rho}_{j2}/\hat{\rho}_{j1}$, $\hat{\rho}_{jk} = \int_0^\tau n^{-1}K_j(t)\,\mathrm{d}\hat{G}_k(t)$, $j = 1,2$, $k = 1,2$. Under the null hypothesis both $\hat{\eta}_1$ and $\hat{\eta}_2$ consistently estimate $\eta$, hence their difference $\hat{\eta}_2 - \hat{\eta}_1$ will fluctuate around zero. On the other hand, when the rate ratio $g_2(t)/g_1(t)$ is nonconstant and $K_1$ and $K_2$ emphasize time periods with different values of $g_2(t)/g_1(t)$, the difference $\hat{\eta}_2 - \hat{\eta}_1$ will be far from zero. Following [11], rewrite

$$\hat{\eta}_2 - \hat{\eta}_1 = \frac{\hat{\rho}_{22}\hat{\rho}_{11} - \hat{\rho}_{21}\hat{\rho}_{12}}{\hat{\rho}_{21}\hat{\rho}_{11}},$$

and use $\hat{\rho}_{22}\hat{\rho}_{11} - \hat{\rho}_{21}\hat{\rho}_{12}$ as the test statistic. In Section 7, this statistic is shown to be asymptotically zero-mean normal, a variance estimator is provided, and a consistency result is given (the test is consistent against monotonic rate ratios provided the limit of $K_2(t)/K_1(t)$ is monotonic).

For testing proportional hazards Gill and Schumacher [11] discussed several choices of the weight functions and recommended the logrank weight $\bar{Y}_1(t)\bar{Y}_2(t)/(\bar{Y}_1(t) + \bar{Y}_2(t))$ and the Prentice–Wilcoxon weight $\hat{S}^*(t-)\bar{Y}_1(t)\bar{Y}_2(t)/(\bar{Y}_1(t) + \bar{Y}_2(t))$, where $\hat{S}^*(t-)$ is the left-continuous Kaplan–Meier estimator computed from the combined sample. In transformation models analogs of these weights are

$$\frac{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))}{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-)) + \bar{Y}_2(t)q_\omega(\hat{A}_2(t-))}, \quad \hat{S}^*(t-)\frac{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))}{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-)) + \bar{Y}_2(t)q_\omega(\hat{A}_2(t-))}. \tag{3}$$

Note that a test related to that of Gill and Schumacher [11] was proposed by Sengupta, Bhattacharjee and Rajeev [18]. They focused on alternatives where the cumulative hazard ratio $A_2(t)/A_1(t)$ is monotonic, which is a slightly broader class of alternatives than alternatives with monotonic $\alpha_2(t)/\alpha_1(t)$. Dauxois and Kirmani [9] applied their idea to testing proportional odds against monotonic $\Gamma_2(t)/\Gamma_1(t)$. These tests are based on statistics of the same form $\hat{\rho}_{22}\hat{\rho}_{11} - \hat{\rho}_{21}\hat{\rho}_{12}$ but with $\hat{\rho}_{jk}$ defined as $\int_0^\tau K_j(t)\hat{G}_k(t)\,\mathrm{d}t$ instead of $\int_0^\tau K_j(t)\,\mathrm{d}\hat{G}_k(t)$. Unlike the Gill–Schumacher-type tests, these tests are not rank tests as they depend on the actual spaces between event times due to the Lebesgue integration.

## 5. SIMULATION STUDY

I carried out a simulation study of the behaviour of three tests of proportionality: the data-driven smooth test ($T_S$), the Kolmogorov–Smirnov (KS) test and the Gill–

**Table 1.** Scenarios for the simulation study.

| | | $\alpha_1(t)$ | $\alpha_2(t)$ | 25 % cens. $(a,b)$ | 45 % cens. $(a,b)$ |
|---|---|---|---|---|---|
| A | (Prop. hazards) | 0.5 | 2 | $(0,5)$ | $(0,2)$ |
| B | (Prop. odds) | $e^{-1}(1+e^{-1}t)^{-1}$ | $e^{1.5}(1+e^{1.5}t)^{-1}$ | $(2,5)$ | $(0,3)$ |
| C | (Monot. ratios) | 1 | $\frac{5}{3}t^{2/3}$ | $(0,3.8)$ | $(0,2)$ |
| D | (Nonmon. rat.) | 0.75 | $\frac{3}{2}(t-1)^2$ | $(0,5.5)$ | $(0,3)$ |

**Table 2.** Estimated rejection probabilities on the nominal level 5 % for configurations A to D with 25 % and 45 % censoring proportions. Sample sizes $n_1 = n_2 = 50$. Figures based on 10 000 simulation repetitions (standard deviation 0.005).

| | A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|---|
| | 25 % | 45 % | 25 % | 45 % | 25 % | 45 % | 25 % | 45 % |
| | Hypothesis: proportional hazards | | | | | | | |
| $T_S$ | 0.056 | 0.048 | 0.322 | 0.169 | 0.577 | 0.417 | 0.926 | 0.692 |
| KS | 0.052 | 0.050 | 0.528 | 0.322 | 0.595 | 0.487 | 0.584 | 0.493 |
| GS | 0.042 | 0.034 | 0.249 | 0.076 | 0.707 | 0.548 | 0.122 | 0.172 |
| | Hypothesis: proportional odds | | | | | | | |
| $T_S$ | 0.414 | 0.220 | 0.052 | 0.049 | 0.507 | 0.314 | 0.926 | 0.683 |
| KS | 0.218 | 0.114 | 0.043 | 0.044 | 0.490 | 0.358 | 0.304 | 0.505 |
| GS | 0.221 | 0.153 | 0.021 | 0.024 | 0.525 | 0.388 | 0.058 | 0.293 |

Schumacher (GS) test. I repeatedly (10 000 times) generated two samples (each of size 50) of survival times under four scenarios. These include proportional hazards, proportional odds, and monotonic and nonmonotonic ratios of hazard rates and odds rates. Survival times were censored by independent variables distributed uniformly on intervals $(a, b)$ adjusted to produce approximately 25 % and 45 % censored observations. Parameters of the simulation design are summarised in Table 1. On the level of 5 %, tests of both proportional hazards and proportional odds were performed. The Kolmogorov–Smirnov test was performed with 1000 resampled processes. The data-driven smooth test was used with $d = 5$, with the Legendre polynomial basis, with the two-term approximation of the distribution of the test statistic. The GS test used the weights (3), the statistic was compared to asymptotic critical points.

Table 2 reports estimates of rejection probabilities. It is seen that Neyman's test and the Kolmogorov–Smirnov test preserve the level very well (see scenario A for proportional hazards and B for proportional odds). The Gill–Schumacher test tends to be slightly conservative, mainly under proportional odds. Under alternatives with monotonic ratios of hazard and/or odds rates (A–C), the overall performance seems to be comparable for all three tests. In the nonmonotonic situation D, it is no surprise that the Gill–Schumacher-type test does not do well as it is designed to be sensitive against monotonic alternatives. The main message of the results

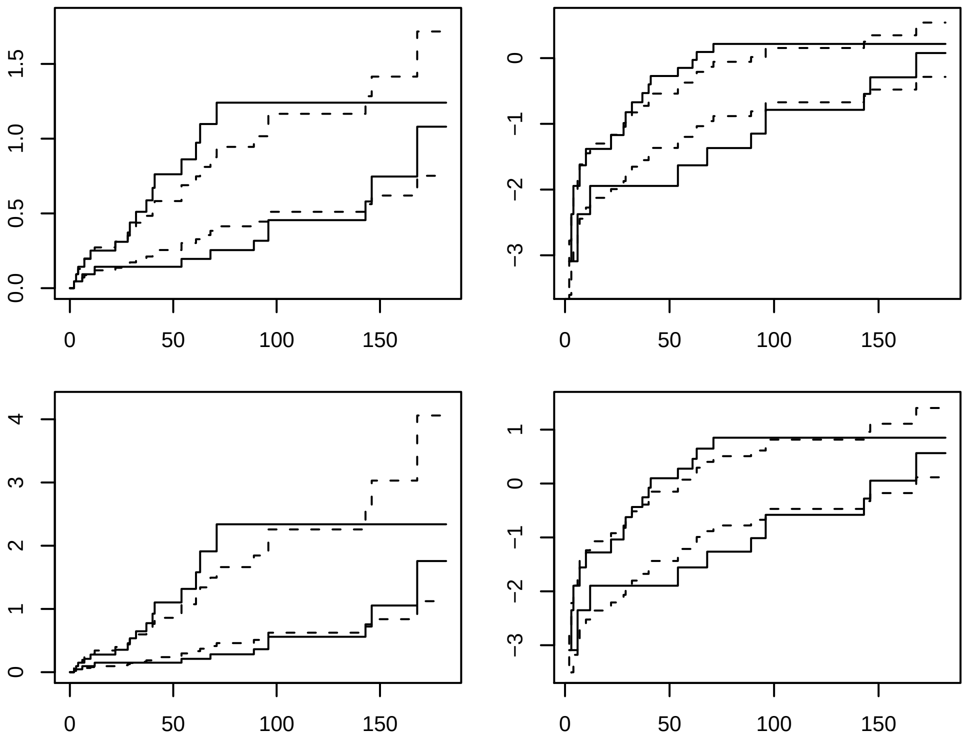


**Fig.** Estimated cumulative hazards and odds functions for the chronic active hepatitis data. Upper row: cumulative hazards (left panel) and log-cumulative hazards (right). Lower row: odds (left) and log-odds functions (right). In each plot: solid curves are estimates computed separately for the treatment group (lower curves) and control group (upper curves), dashed lines show corresponding model-based estimates. Time from the beginning of the trial is in months.

concerning the power of the proposed smooth test is that this test maintains stable power for a variety of departures from proportionality. I performed simulations for other combinations of distributions, and never met a situation where the smooth test dramatically lost compared to the other methods.

## 6. ILLUSTRATION

A real example is taken from Collett [7, Appendix D.1]. The data concern survival times of patients with chronic active hepatitis. There were 44 patients, 22 of them (randomly selected) received a drug (11 died, 11 were censored), the remaining 22 were in the control group (16 deaths, 6 survivors). Figure displays estimates of cumulative hazards, odds functions, and their logarithms (i. e., complementary log-

**Table 3.** Results of tests of fit for the chronic active hepatitis data.

| | Proportional hazards | | Proportional odds | |
|---|---|---|---|---|
| | Statistic | $p$-value | Statistic | $p$-value |
| $T_1$ | 2.09 | 0.148 | 0.45 | 0.502 |
| $T_2$ | 7.71 | 0.021 | 5.36 | 0.069 |
| $T_3$ | 7.85 | 0.049 | 5.67 | 0.129 |
| $T_4$ | 8.30 | 0.081 | 6.09 | 0.192 |
| $T_5$ | 9.36 | 0.096 | 7.29 | 0.200 |
| $T_S$ | 7.71 | 0.005 | 5.36 | 0.061 |
| KS | 3.00 | 0.052 | 2.53 | 0.169 |
| GS | 1.23 | 0.220 | 0.46 | 0.648 |

log and logit transform of the Kaplan–Meier estimate). These estimates obtained separately from each sample are plotted by solid lines. If the proportional assumption holds, the vertical distance between log-curves should be approximately constant. Estimates based on proportional rate models are plotted by dashed lines. Results of goodness-of-fit tests are summarised in Table 3.

The partial likelihood estimate in the proportional hazards model is $\hat{\beta} = 0.826$ ($e^{\hat{\beta}} = 2.28$). The data-driven smooth test (with maximum dimension $d = 5$) rejects the hypothesis of proportional hazards. Schwarz's selection rule selects two basis functions (the linear and quadratic Legendre polynomial), which corresponds to the fact that the hazard ratio appears to be nonmonotonic. The quadratic function contributes to the description of the hazard ratio most; using more than two basis functions does not increase the statistic much. The Gill–Schumacher test does not reject the hypothesis of proportionality (the logrank and Prentice–Wilcoxon weighted estimates of $\eta = e^{\beta}$ are 2.37 and 2.73, respectively) as this test is focused against alternatives with monotonic ratios.

If we are interested in the proportional odds model, the simplified partial likelihood procedure gives the estimate $\hat{\beta} = 1.29$ ($e^{\hat{\beta}} = 3.63$), and two weighted estimates used in the Gill–Schumacher test are $\hat{\eta}_1 = 3.74$, $\hat{\eta}_2 = 3.93$ (with weights (3)). Plots of (log-)odds functions indicate a similar type of departure from proportionality as plots of cumulative hazards; results, however, do not lead to rejection on the 5 % level.

## 7. ASYMPTOTIC RESULTS

### 7.1. Assumptions

It is assumed that $n^{-1}\bar{Y}_j(t)$, $j = 1, 2$, converge in probability uniformly in $t \in [0, \tau]$ to some functions $\bar{y}_j(t)$ bounded away from zero. This is satisfied if $n_j/n \to a_j \in (0, 1)$, $S_j(\tau) > 0$ and $1 - C_j(\tau) > 0$ ($C_j$ is the distribution function of censoring variables) because then by the Glivenko–Cantelli theorem $\bar{y}_j(t) = a_j S_j(t)(1 - C_j(t))$.

Further assume that $q_\omega(t) = \alpha_\omega(A_\omega^{-1}(t))$ is continuously differentiable on $[0, \tau]$. For both main examples this is satisfied ($q_\omega(t) = 1$ for proportional hazards, $q_\omega(t) =$

$e^{-t}$ for proportional odds). Denote the derivative $\dot{q}_\omega(t)$.

### 7.2. Consistency of the estimation procedure

**Lemma 1.** (Convergence of $\hat{\beta}$) Assume that the rate ratio $g_2(t)/g_1(t)$ is $e^{\beta_0(t)}$, i. e., it may or may not be constant. Then the estimator $\hat{\beta}$ defined as the solution to $U_1(\tau;\beta,\hat{A}_1,\hat{A}_2)=0$ converges in probability to the solution $\bar{\beta}$ to the limiting estimating equation

$$\int_0^\tau \frac{\bar{y}_1(t)q_\omega(A_1(t))\bar{y}_2(t)q_\omega(A_2(t))}{\bar{y}_1(t)q_\omega(A_1(t))+\bar{y}_2(t)q_\omega(A_2(t))e^{\bar{\beta}}}(e^{\beta_0(t)}-e^{\bar{\beta}})\,g_0(t)\,\mathrm{d}t=0. \tag{4}$$

Specifically, if the proportional rate model holds ($\beta_0$ is constant), $\hat{\beta}$ consistently estimates $\beta_0$.

P r o o f. The proof is analogous to that for the Cox model (see Theorem 8.3.1 of Fleming and Harrington [10] or Theorem VII.2.1 of Andersen, Borgan, Gill and Keiding [1], and also Struthers and Kalbfleisch [19]). The maximiser of $C(\tau;\beta,\hat{A}_1,\hat{A}_2)$ is the same as the maximiser of

$$\begin{aligned}
& n^{-1}(C(\tau;\beta,\hat{A}_1,\hat{A}_2)-C(\tau;\bar{\beta},\hat{A}_1,\hat{A}_2))\\
= \ & n^{-1}(\beta-\bar{\beta})\bar{N}_2(\tau)\\
& \quad -n^{-1}\int_0^\tau \log\Big(\frac{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))+\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))e^{\beta}}{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))+\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))e^{\bar{\beta}}}\Big)\mathrm{d}\bar{N}(t)\\
= \ & n^{-1}(\beta-\bar{\beta})\bar{\Lambda}_2(\tau)\\
& \quad -n^{-1}\int_0^\tau \log\Big(\frac{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))+\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))e^{\beta}}{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))+\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))e^{\bar{\beta}}}\Big)\mathrm{d}\bar{\Lambda}(t)\\
& \quad +n^{-1}(\beta-\bar{\beta})\bar{M}_2(\tau)\\
& \quad -n^{-1}\int_0^\tau \log\Big(\frac{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))+\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))e^{\beta}}{\bar{Y}_1(t)q_\omega(\hat{A}_1(t-))+\bar{Y}_2(t)q_\omega(\hat{A}_2(t-))e^{\bar{\beta}}}\Big)\mathrm{d}\bar{M}(t).
\end{aligned}$$

Here the last two terms converge to zero by Lenglart's inequality. Hence, by the uniform consistency of Nelson–Aalen estimators, $n^{-1}(C(\tau;\beta,\hat{A}_1,\hat{A}_2)-C(\tau;\bar{\beta},\hat{A}_1,\hat{A}_2))$ converges in probability to

$$\begin{aligned}
(\beta-\bar{\beta})\int_0^\tau & \bar{y}_2(t)q_\omega(A_2(t))e^{\beta_0(t)}g_0(t)\,\mathrm{d}t\\
& -\int_0^\tau \log\Big(\frac{\bar{y}_1(t)q_\omega(A_1(t))+\bar{y}_2(t)q_\omega(A_2(t))e^{\beta}}{\bar{y}_1(t)q_\omega(A_1(t))+\bar{y}_2(t)q_\omega(A_2(t))e^{\bar{\beta}}}\Big)\\
& \qquad \times[\bar{y}_1(t)q_\omega(A_1(t))+\bar{y}_2(t)q_\omega(A_2(t))e^{\beta_0(t)}]g_0(t)\,\mathrm{d}t.
\end{aligned} \tag{5}$$

Then, in the light of concavity of $n^{-1}(C(\tau;\beta,\hat{A}_1,\hat{A}_2)-C(\tau;\bar{\beta},\hat{A}_1,\hat{A}_2))$, Lemma 8.3.1 in [10] (see also Appendix II in [2]) yields that the maximiser of $n^{-1}(C(\tau;\beta,\hat{A}_1,\hat{A}_2)-C(\tau;\bar{\beta},\hat{A}_1,\hat{A}_2))$ converges in probability to the maximiser of (5), which is by concavity the solution to (4). $\square$

### 7.3. Asymptotics for Neyman's test

**Lemma 2.** The process $n^{-1/2}U_1(\cdot;\beta_0,\hat{A}_1,\hat{A}_2)$ is asymptotically distributed as the process

$$V_1(t) = \int_0^t l_{12}(s)\,\mathrm{d}V_{12}(s) - \int_0^t l_{11}(s)\,\mathrm{d}V_{11}(s) - \int_0^t V_{12}(s)\,\mathrm{d}h_{12}(s) + \int_0^t V_{11}(s)\,\mathrm{d}h_{11}(s),$$

where $V_{1j}$ are independent zero-mean continuous Gaussian martingales with variance functions $\int_0^t \bar{y}_j(s)^{-1}\mathrm{d}A_j(s)$, and the functions $h_{1j}$ and $l_{1j}$ are uniform limits in probability of $n^{-1}H_{1j}$ and $n^{-1}L_{1j}$ defined below in the proof.

P r o o f. By the martingale central limit theorem, the process $n^{1/2}(\hat{A}_1 - A_1, \hat{A}_2 - A_2)$ converges in distribution to $(V_{11}, V_{12})$, which is a standard result on the Nelson–Aalen estimator. Rewrite $U_1(t;\beta_0,\hat{A}_1,\hat{A}_2)$ completely in terms of $\hat{A}_j$ as follows

$$\begin{aligned} U_1(t;\beta_0,\hat{A}_1,\hat{A}_2) = & \int_0^\tau \frac{\bar{Y}_1(s)q_\omega(\hat{A}_1(s))}{\bar{Y}_1(s)q_\omega(\hat{A}_1(s)) + \bar{Y}_2(s)q_\omega(\hat{A}_2(s))e^{\beta_0}}\bar{Y}_2(s)\,\mathrm{d}\hat{A}_2(s) \\ & - \int_0^\tau \frac{\bar{Y}_2(s)q_\omega(\hat{A}_2(s))e^{\beta_0}}{\bar{Y}_1(s)q_\omega(\hat{A}_1(s)) + \bar{Y}_2(s)q_\omega(\hat{A}_2(s))e^{\beta_0}}\bar{Y}_1(s)\,\mathrm{d}\hat{A}_1(s). \end{aligned}$$

When in this expression $\hat{A}_j$ are replaced by $A_j$, the result is zero. Thus the asymptotic distribution of $U_1(t;\beta_0,\hat{A}_1,\hat{A}_2)$ (minus zero) can be inferred from that of $\hat{A}_j - A_j$ with the help of the functional delta method. Using the chain rule and a lemma on differentiation of integration (Proposition II.8.6 in [1] or Lemma 3.9.17 in [20]), we obtain that $n^{-1/2}U_1(t;\beta_0,\hat{A}_1,\hat{A}_2)$ is asymptotically equivalent to

$$\begin{aligned} & \int_0^t n^{-1}L_{12}(s)n^{1/2}(d\hat{A}_2(s) - \mathrm{d}A_2(s)) - \int_0^t n^{-1}L_{11}(s)n^{1/2}(d\hat{A}_1(s) - \mathrm{d}A_1(s)) \\ & - \int_0^t n^{1/2}(\hat{A}_2(s) - A_2(s))n^{-1}\mathrm{d}H_{12}(s) + \int_0^t n^{1/2}(\hat{A}_1(s) - A_1(s))n^{-1}\mathrm{d}H_{11}(s), \end{aligned} \tag{6}$$

where

$$L_{11}(t) = \frac{\bar{Y}_1(t)\bar{Y}_2(t)q_\omega(A_2(t))e^{\beta_0}}{\bar{Y}_1(t)q_\omega(A_1(t)) + \bar{Y}_2(t)q_\omega(A_2(t))e^{\beta_0}},$$

$$L_{12}(t) = \frac{\bar{Y}_1(t)q_\omega(A_1(t))\bar{Y}_2(t)}{\bar{Y}_1(t)q_\omega(A_1(t)) + \bar{Y}_2(t)q_\omega(A_2(t))e^{\beta_0}},$$

$$H_{11}(t) = \int_0^t \frac{\bar{Y}_1(s)\dot{q}_\omega(A_1(s))\bar{Y}_2(s)q_\omega(A_2(s))e^{\beta^0}}{[\bar{Y}_1(s)q_\omega(A_1(s)) + \bar{Y}_2(s)q_\omega(A_2(s))e^{\beta^0}]^2}[\bar{Y}_1(s)\mathrm{d}A_1(s) + \bar{Y}_2(s)\,\mathrm{d}A_2(s)],$$

$$H_{12}(t) = \int_0^t \frac{\bar{Y}_1(s)q_\omega(A_1(s))\bar{Y}_2(s)\dot{q}_\omega(A_2(s))e^{\beta_0}}{[\bar{Y}_1(s)q_\omega(A_1(s)) + \bar{Y}_2(s)q_\omega(A_2(s))e^{\beta_0}]^2}[\bar{Y}_1(s)\,\mathrm{d}A_1(s) + \bar{Y}_2(s)\,\mathrm{d}A_2(s)].$$

□

**Lemma 3.** The process $n^{-1/2}U_1(\cdot;\hat\beta,\hat A_1,\hat A_2)$ is asymptotically distributed as the process $V_1(t)-d_1(t;\beta_0,A_1,A_2)d_1(\tau;\beta_0,A_1,A_2)^{-1}V_1(\tau)$, where $V_1$ is the process of Lemma 2 and the function $d_1(t;\beta,A_1,A_2)$ is the uniform limit in probability of $n^{-1}D_1(t;\beta,\hat A_1,\hat A_2)$ defined below.

P r o o f. The proof follows by Lemma 2 after a straightforward use of Taylor's expansion which gives

$$\begin{aligned} n^{-1/2}U_1(t;\hat\beta,\hat A_1,\hat A_2) &= n^{-1/2}U_1(t;\beta_0,\hat A_1,\hat A_2)-n^{-1}D_1(t;\beta_t^*,\hat A_1,\hat A_2)n^{1/2}(\hat\beta-\beta_0)\\ &= n^{-1/2}U_1(t;\beta_0,\hat A_1,\hat A_2)-\{n^{-1}D_1(t;\beta_t^*,\hat A_1,\hat A_2)\}\{n^{-1}D_1(\tau;\beta_\tau^*,\hat A_1,\hat A_2)\}^{-1}\\ &\qquad\times n^{-1/2}U_1(t;\beta_0,\hat A_1,\hat A_2), \qquad (7)\end{aligned}$$

where $\beta_t^*$ lies on the line segment between $\beta_0$ and $\hat\beta$, and

$$\begin{aligned} D_1(t;\beta,\hat A_1,\hat A_2) &= -\frac{\partial}{\partial\beta}U_1(t;\beta,\hat A_1,\hat A_2)\\ &= \int_0^t \frac{\bar Y_1(s)q_\omega(\hat A_1(s))\bar Y_2(s)q_\omega(\hat A_2(s))e^\beta}{[\bar Y_1(s)q_\omega(\hat A_1(s))+\bar Y_2(s)q_\omega(\hat A_2(s))e^\beta]^2}\,\mathrm{d}\bar N(s).\end{aligned}$$

□

**Theorem 1.** (Asymptotic distribution of the score) The score vector $n^{-1/2}U_2(\tau;\hat\beta,\hat A_1,\hat A_2)$ converges in distribution to a mean zero Gaussian vector with variance matrix that is consistently estimated by $n^{-1}\hat\Xi$ given below in (8).

P r o o f. By Taylor's expansion about $\beta_0$,

$$\begin{aligned} n^{-1/2}U_2(\tau;\hat\beta,\hat A_1,\hat A_2) &= n^{-1/2}U_2(\tau;\beta_0,\hat A_1,\hat A_2)\\ &\quad - n^{-1}D_2(\tau;\beta^{**},\hat A_1,\hat A_2)\{n^{-1}D_1(\tau;\beta^*,\hat A_1,\hat A_2)\}^{-1}n^{-1/2}U_1(\tau;\beta_0,\hat A_1,\hat A_2),\end{aligned}$$

where $D_2(\tau;\beta,\hat A_1,\hat A_2) = -\frac{\partial}{\partial\beta}U_2(\tau;\beta,\hat A_1,\hat A_2)$, and $\beta^*$ and $\beta^{**}$ are on the line segment between $\beta_0$ and $\hat\beta$ (to be technically precise, note that each component of $D_2$ has its own $\beta^{**}$, all between $\beta_0$ and $\hat\beta$).

The variables $n^{-1}D_1(\tau;\beta^*,\hat A_1,\hat A_2), n^{-1}D_2(\tau;\beta^{**},\hat A_1,\hat A_2)$ converge in probability to $d_1(\tau;\beta_0,A_1,A_2), d_2(\tau;\beta_0,A_1,A_2)$, respectively, where $d_1$ is explained in Lemma 3 and $d_2(\tau;\beta,A_1,A_2) = \int_0^\tau \psi(t)d_{11}(\mathrm{d}t;\beta,A_1,A_2)$. The $(1+d)$-dimensional vector $n^{-1/2}(U_1(\tau;\beta_0,\hat A_1,\hat A_2),U_2(\tau;\beta_0,\hat A_1,\hat A_2)^\mathsf{T})^\mathsf{T}$ jointly converges weakly to the zero-mean Gaussian vector $(V_1(\tau),V_2(\tau)^\mathsf{T})^\mathsf{T}$ with $V_1$ given in Lemma 3 and $V_2(\tau) = \int_0^\tau \psi(t)V_1(t)$.

Let us derive suitable variance estimators. Integrating by parts (or using Fubini's theorem) in (6) yields that $n^{-1/2}U_1(\tau;\beta_0,\hat A_1,\hat A_2)$ has the same asymptotic distribution as

$$\int_0^\tau n^{-1}[L_{12}(t) + H_{12}(t))]n^{1/2}(\mathrm{d}\hat{A}_2(s) - \mathrm{d}A_2(s)) \\ - n^{-1}H_{12}(\tau)n^{1/2}(\hat{A}_2(\tau) - A_2(\tau)) \\ - \int_0^\tau n^{-1}[L_{11}(t) + H_{11}(t)]n^{1/2}(\mathrm{d}\hat{A}_1(s) - \mathrm{d}A_1(s)) \\ + n^{-1}H_{11}(\tau)n^{1/2}(\hat{A}_1(\tau) - A_1(\tau)).$$

For $n^{-1/2}U_2(\tau;\beta_0,\hat{A}_1,\hat{A}_2)$ we get an analogous expression with $L_{2j}(t) = \psi(t)L_{1j}(t)$ instead of $L_{1j}(t)$ and $H_{2j}(t) = \int_0^t \psi(s)\,\mathrm{d}H_{1j}(s)$ instead of $H_{1j}(t)$. Thus the asymptotic variance $\Sigma_{11}$ of $n^{-1/2}U_1(\tau;\beta_0,\hat{A}_1,\hat{A}_2)$, covariance vector $\Sigma_{21}$ of $n^{-1/2}U_2(\tau;\beta_0,\hat{A}_1,\hat{A}_2)$, $n^{-1/2}U_1(\tau;\beta_0,\hat{A}_1,\hat{A}_2)$, and variance matrix $\Sigma_{22}$ of $n^{-1/2}U_2(\tau;\beta_0,\hat{A}_1,\hat{A}_2)$ can be consistently estimated by

$$n^{-1}\hat{\Sigma}_{kk'} \\ = \sum_{j=1}^{2}\int_0^\tau n^{-1}[\hat{L}_{kj}(t) + \hat{H}_{kj}(t) - \hat{H}_{kj}(\tau)][\hat{L}_{k'j}(t) + \hat{H}_{k'j}(t) - \hat{H}_{k'j}(\tau)]^\intercal \frac{\mathrm{d}\hat{A}_j(t)}{\bar{Y}_j(t)},$$

$k,k' = 1,2$. Here $\hat{L}_{kj}$ and $\hat{H}_{kj}$ are defined like $L_{kj}$ and $H_{kj}$ with $\beta_0, A_1, A_2$ replaced by $\hat{\beta}, \hat{A}_1, \hat{A}_2$ (with left-continuous Nelson–Aalen estimators in integrands because of predictability). The very final conclusion is that the asymptotic variance of $n^{-1/2}U_2(\tau;\hat{\beta},\hat{A}_1,\hat{A}_2)$ is estimated by

$$n^{-1}\hat{\Xi} = n^{-1}\hat{\Sigma}_{22} - n^{-1}\hat{D}_2\hat{D}_1^{-1}\hat{\Sigma}_{21}^\intercal - n^{-1}\hat{\Sigma}_{21}\hat{D}_1^{-1}\hat{D}_2^\intercal + n^{-1}\hat{D}_2\hat{D}_1^{-1}\hat{\Sigma}_{11}\hat{D}_1^{-1}\hat{D}_2^\intercal \tag{8}$$

(having set $\hat{D}_k = D_k(\tau;\hat{\beta},\hat{A}_1,\hat{A}_2)$). □

**Theorem 2.** (Consistency of Neyman's test) Assume that the true rate ratio is time-varying of the form $e^{\beta_0(t)}$. Let $\bar{\beta}$ be as in Lemma 1. Suppose that the basis functions satisfy the condition

$$\int_0^\tau \psi(t)\frac{\bar{y}_1(t)q_\omega(A_1(t))\bar{y}_2(t)q_\omega(A_2(t))}{\bar{y}_1(t)q_\omega(A_1(t)) + \bar{y}_2(t)q_\omega(A_2(t))e^{\bar{\beta}}}(e^{\beta_0(t)} - e^{\bar{\beta}})\,g_0(t)\,\mathrm{d}t \neq 0 \tag{9}$$

(at least one component differs from zero). Then the rejection probability of Neyman's test approaches 1 as $n \to \infty$.

P r o o f. In view of the definition $U_2(\tau;\hat{\beta},\hat{A}_1,\hat{A}_2) = \int_0^\tau \psi(t)U_1(\,\mathrm{d}t;\hat{\beta},\hat{A}_1,\hat{A}_2)$ and Lemma 1, $n^{-1}U_2(\tau;\hat{\beta},\hat{A}_1,\hat{A}_2)$ converges in probability to the left-hand side of (9). The variance matrix estimator $n^{-1}\hat{\Xi}$ converges to some finite matrix. Thus $n^{-1}$ times the score test statistic converges in probability to a nonzero number. □

The consistency condition means, loosely speaking, that the choice of the basis functions is not 'completely wrong'. More precisely, the left-hand side of (9) is the limiting estimating equation for the parameters $\theta = (\theta_1, \ldots, \theta_d)^\mathsf{T}$ in the smooth model (2) evaluated at $\theta = 0$. The inequality (9) means that $\theta = 0$ does not solve the estimating equation, that is, the basis function contribute to the description of the true time-varying rate ratio. In other words, the test is consistent against alternatives whose projection on the smooth model (2) does not fall to the null model.

### 7.4. Asymptotics for the Gill–Schumacher test

Assume that $n^{-1}K_j(t)$, $j = 1, 2$, converge in probability uniformly in $t \in [0, \tau]$ to some functions $k_j(t)$ bounded away from zero. For instance, the logrank-type and Prentice–Wilcoxon-type weights (3) satisfy this condition by the convergence of $n^{-1}\bar{Y}_j(t)$ and the Kaplan–Meier estimator.

**Theorem 3.** (Asymptotic distribution of the GS statistic) Under the null hypothesis of proportionality of $g_1, g_2$, the test statistic $n^{1/2}(\hat{\rho}_{22}\hat{\rho}_{11} - \hat{\rho}_{21}\hat{\rho}_{12})$ is asymptotically normal with mean zero and variance given by (11) below, which is consistently estimated by (12).

P r o o f. Denote $\rho_{jk} = \int_0^\tau k_j(t)\,\mathrm{d}G_k(t)$ and rewrite

$$\hat{\rho}_{22}\hat{\rho}_{11} - \hat{\rho}_{21}\hat{\rho}_{12} = (\hat{\rho}_{22} - \rho_{22})\hat{\rho}_{11} + (\hat{\rho}_{11} - \rho_{11})\rho_{22} - (\hat{\rho}_{21} - \rho_{21})\hat{\rho}_{12} \\ - (\hat{\rho}_{12} - \rho_{12})\rho_{21} + \rho_{22}\rho_{11} - \rho_{21}\rho_{12}. \quad (10)$$

Under the hypothesis it is $\rho_{j2} = \eta\rho_{j1}$, hence the last two terms together are zero. Further, $\hat{\rho}_{jk}$ converges in probability to $\rho_{jk}$. It remains to explore the weak convergence of $n^{1/2}(\hat{\rho}_{jk} - \rho_{jk})$ jointly for $j = 1, 2$, $k = 1, 2$.

Recall that $G_k(t) = A_\omega^{-1}(A_k(t))$ and $\hat{G}_k(t) = A_\omega^{-1}(\hat{A}_k(t))$. By the functional delta method $n^{1/2}(\hat{G}_k(\cdot) - G_k(\cdot))$ is asymptotically equivalent to

$$\frac{1}{q_\omega(A_k(\cdot))} n^{1/2}(\hat{A}_k(\cdot) - A_k(\cdot)).$$

Thus, the asymptotic distribution of $n^{1/2}(\hat{\rho}_{jk} - \rho_{jk})$ is the same as the asymptotic distribution of

$$\int_0^\tau n^{-1}K_j(t)\mathrm{d}\left(\frac{1}{q_\omega(A_k(t))} n^{1/2}(\hat{A}_k(t) - A_k(t))\right) \\ = \int_0^\tau n^{1/2}(\hat{A}_k(t) - A_k(t))n^{-1}\mathrm{d}B_{jk}(t) + \int_0^\tau \frac{n^{-1}K_j(t)}{q_\omega(A_k(t))} n^{1/2}(\mathrm{d}\hat{A}_k(t) - \mathrm{d}A_k(t)),$$

where $\mathrm{d}B_{jk}(t) = K_j(t)\,\mathrm{d}B_k(t)$ and $\mathrm{d}B_k(t) = \mathrm{d}(1/q_\omega(A_k(t))) = -\dot{q}_\omega(A_k(t))/ /q_\omega(A_k(t))\,\mathrm{d}G_k(t)$. This asymptotic distributional equivalence holds jointly for $j = 1, 2$, $k = 1, 2$. Integrating by parts we arrive at

$$\int_0^\tau n^{-1}R_{jk}(t)n^{1/2}(d\hat{A}_k(t) - \mathrm{d}A_k(t)) + n^{-1}B_{jk}(\tau)n^{1/2}(\hat{A}_k(\tau) - A_k(\tau)),$$

where

$$R_{jk}(t) = \frac{K_j(t)}{q_\omega(A_k(t))} - B_{jk}(t).$$

Denote by $b_{jk}$ and $r_{jk}$ the limits of $n^{-1}B_{jk}$ and $n^{-1}R_{jk}$, respectively. Then by the martingale central limit theorem $n^{1/2}(\hat\rho_{jk} - \rho_{jk})$, $j = 1, 2$, $k = 1, 2$, converge to zero mean jointly normal variables. The asymptotic covariance of $n^{1/2}(\hat\rho_{jk} - \rho_{jk})$ and $n^{1/2}(\hat\rho_{j'k} - \rho_{j'k})$ is

$$\int_0^\tau (r_{jk}(t) + b_{jk}(\tau))(r_{j'k}(t) + b_{j'k}(\tau)) \frac{\mathrm{d}A_k(t)}{\bar y_k(t)},$$

while the asymptotic covariance of $n^{1/2}(\hat\rho_{jk} - \rho_{jk})$ and $n^{1/2}(\hat\rho_{j'k'} - \rho_{j'k'})$ is zero for $k \neq k'$.

Therefore, using the fact $\rho_{j2} = \eta\rho_{j1}$, it follows that the asymptotic variance of the statistic $n^{1/2}(\hat\rho_{22}\hat\rho_{11} - \hat\rho_{21}\hat\rho_{12})$ is

$$\int_0^\tau [(r_{11}(t) + b_{11}(\tau))\rho_{21} - (r_{21}(t) + b_{21}(\tau))\rho_{11}]^2 \eta^2 \frac{\mathrm{d}A_1(t)}{\bar y_1(t)} \\ + \int_0^\tau [(r_{12}(t) + b_{12}(\tau))\rho_{11} - (r_{22}(t) + b_{22}(\tau))\rho_{21}]^2 \frac{\mathrm{d}A_2(t)}{\bar y_2(t)}.$$

Finally, as $dG_2(t) = \eta dG_1(t)$ and $dG_k(t) = dA_k(t)/q_\omega(A_k(t))$, we arrive at

$$\int_0^\tau [(r_{11}(t) + b_{11}(\tau))\rho_{21} - (r_{21}(t) + b_{21}(\tau))\rho_{11}]^2 \eta \frac{q_\omega(A_1(t))\,\mathrm{d}A_2(t)}{q_\omega(A_2(t))\bar y_1(t)} \\ + \int_0^\tau [(r_{12}(t) + b_{12}(\tau))\rho_{11} - (r_{22}(t) + b_{22}(\tau))\rho_{21}]^2 \eta \frac{q_\omega(A_2(t))\,\mathrm{d}A_1(t)}{q_\omega(A_1(t))\bar y_2(t)}, \tag{11}$$

which may be consistently estimated by

$$\int_0^\tau [n^{-1}(\hat R_{11}(t) + \hat B_{11}(\tau))\hat\rho_{21} - n^{-1}(\hat R_{21}(t) + \hat B_{21}(\tau))\hat\rho_{11}]^2 \hat\eta_0 \frac{q_\omega(\hat A_1(t-))\,\mathrm{d}\hat A_2(t)}{q_\omega(\hat A_2(t-))\bar Y_1(t)/n} \\ + \int_0^\tau [n^{-1}(\hat R_{12}(t) + \hat B_{12}(\tau))\hat\rho_{11} - n^{-1}(\hat R_{22}(t) + \hat B_{22}(\tau))\hat\rho_{21}]^2 \hat\eta_0 \frac{q_\omega(\hat A_2(t-))\,\mathrm{d}\hat A_1(t)}{q_\omega(\hat A_1(t-))\bar Y_2(t)/n}. \tag{12}$$

Here $\hat R_{jk}(t)$ and $\hat B_{jk}(t)$ are defined like $R_{jk}(t)$ and $B_{jk}(t)$ but with unknown quantities replaced by their estimators (the Nelson–Aalen estimator is used in the left-continuous version where necessary to preserve predictability). The estimator $\hat\eta_0$ is defined as $(\hat\eta_1 + \hat\eta_2)/2$ to preserve some sort of symmetry (but any other consistent estimator of $\eta$ may be used as well). □

Note that the variance estimator (12) is always positive. A different estimator of a form similar to that of Gill and Schumacher [11, eq. (4)] can be derived. However, such an estimator may be negative (and my experience is that it sometimes really happens).

**Theorem 4.** (Consistency of the Gill–Schumacher test) The Gill–Schumacher-type test is consistent against alternatives satisfying $\rho_{22}\rho_{11} - \rho_{21}\rho_{12} \neq 0$. This particularly holds for alternatives with monotonic $g_2(t)/g_1(t)$ whenever $k_2(t)/k_1(t)$ is monotonic.

Proof. The variance estimator (12) converges to some finite nonzero quantity even under the alternative, and, thus, the first assertion follows from (10). The proof of the rest is the same as the proof in [11, p. 293] with hazard rates replaced by transformation rates $g_k$. $\square$

## ACKNOWLEDGEMENT

The work was supported by the Czech Science Foundation through Grant 201/05/H007, the Grant Agency of the Academy of Sciences of the Czech Republic through Grant IAA101120604 and the Ministry of Education, Youth and Sport of the Czech Republic through the project 1M06047. Computations were carried out in METACentrum (Czech academic supercomputer network).

(Received November 7, 2007.)

## REFERENCES

[1] P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding: Statistical Models Based on Counting Processes. Springer, New York 1993.

[2] P. K. Andersen and R. D. Gill: Cox's regression model for counting processes: a large sample study. Ann. Statist. *10* (1982), 1100–1120.

[3] V. Bagdonavičius and M. Nikulin: Generalized proportional hazards model based on modified partial likelihood. Lifetime Data Anal. *5* (1999), 329–350.

[4] V. Bagdonavičius and N. Nikulin: On goodness-of-fit for the linear transformation and frailty models. Statist. Probab. Lett. *47* (2000), 177–188.

[5] V. Bagdonavičius and M. Nikulin: Accelerated life models. Modeling and Statistical Analysis. Chapman & Hall/CRC, Boca Raton 2001.

[6] K. Chen, Z. Jin, and Z. Ying: Semiparametric analysis of transformation models with censored data. Biometrika *89* (2002), 659–668.

[7] D. Collett: Modelling Survival Data in Medical Research. Chapman & Hall/CRC, Boca Raton 2003.

[8] D. M. Dabrowska and K. A. Doksum: Estimation and testing in a two-sample generalized odds-rate model. J. Amer. Statist. Assoc. *83* (19í8), 744–749.

[9] J.-Y. Dauxois and S. N. U. A. Kirmani: Testing the proportional odds model under random censoring. Biometrika *90* (2003), 913–922.

[10] T. R. Fleming and D. P. Harrington: Counting Processes and Survival Analysis. Wiley, New York 1991.

[11] R. D. Gill and M. Schumacher: A simple test of the proportional hazards assumption. Biometrika *74* (1987), 289–300.

[12] D. Kraus: Data-driven smooth tests of the proportional hazards assumption. Lifetime Data Anal. *13* (2007), 1–16.

[13] T. Ledwina: Data-driven version of Neyman's smooth test of fit. J. Amer. Statist. Assoc. *89* (1994), 1000–1005.

[14] D. Y. Lin, L. J. Wei, and Z. Ying: Checking the Cox model with cumulative sums of martingale-based residuals. Biometrika *80* (1993), 557–572.

[15] T. Martinussen and T. H. Scheike: Dynamic Regression Models for Survival Data. Springer, New York 2006.

[16] S. A. Murphy, A. J. Rossini, and A. W. van der Vaart: Maximum likelihood estimation in the proportional odds model. J. Amer. Statist. Assoc. *92* (1997), 968–976.

[17] J. C. W. Rayner and D. J. Best: Smooth Tests of Goodness of Fit. Oxford University Press, New York 1989.

[18] D. Sengupta, A. Bhattacharjee, and B. Rajeev: Testing for the proportionality of hazards in two samples against the increasing cumulative hazard ratio alternative. Scand. J. Statist. *25* (1998), 637–647.

[19] C. A. Struthers and J. D. Kalbfleisch: Misspecified proportional hazard models. Biometrika *73* (1986), 363–369.

[20] A. W. van der Vaart and J. A. Wellner: Weak Convergence and Empirical Processes. With Applications to Statistics. Springer, New York 1996.

[21] L. J. Wei: Testing goodness of fit for proportional hazards model with censored observations. J. Amer. Statist. Assoc. *79* (1984), 649–652.

*David Kraus, Institute of Information Theory and Automation – Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 4, 182 08 Praha 8. Czech Republic.*
*e-mail: david.kraus@matfyz.cz*