View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by University of Dundee Online Publications

University of Dundee

Partial Orders with Respect to Continuous Covariates and Tests for the Proportional Hazards Model

Bhattacharjee, Arnab

Publication date:
2008

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):
Bhattacharjee, A. (2008). Partial Orders with Respect to Continuous Covariates and Tests for the Proportional Hazards Model. (Discussion Paper Series; No. 0807). St. Andrews: Department of Economics, University of St. Andrews.

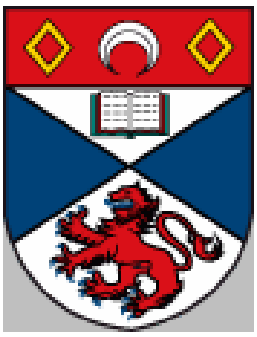
General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

ISSN 0962-4031

UNIVERSITY OF ST. ANDREWS



Partial Orders with Respect to Continuous Covariates and Tests for the Proportional Hazards Model

Arnab Bhattacharjee

No. 0807

DISCUSSION PAPER SERIES

SCHOOL OF ECONOMICS & FINANCE
St. Salvator's College
St. Andrews, Fife KY16 9AL
Scotland

Partial Orders with Respect to Continuous Covariates and Tests for the Proportional Hazards Model

Arnab Bhattacharjee*
University of St. Andrews, UK.

Revised, November 2007

Abstract

Several omnibus tests of the proportional hazards assumption have been proposed in the literature. In the two-sample case, tests have also been developed against ordered alternatives like monotone hazard ratio and monotone ratio of cumulative hazards. Here we propose a natural extension of these partial orders to the case of continuous covariates. The work is motivated by applications in biomedicine and economics where covariate effects often decay over lifetime. We develop tests for the proportional hazards assumption against ordered alternatives and propose a graphical method to identify the nature of departures from proportionality. The proposed tests do not make restrictive assumptions on the underlying regression model, and are applicable in the presence of multiple covariates and frailty. Small sample performance and applications to real data highlight the usefulness of the framework and methodology.

Keywords: Two-sample tests; Increasing hazard ratio; Continuous covariate; Proportional hazards; Frailty; Partial orders; Time varying coefficients.
JEL Classification: C12, C14, C41.

*Address: School of Economics and Finance, University of St. Andrews, Castlecliffe, St. Andrews KY16 9AL, UK. Tel: +44 1334 462423. e-mail: ab102@st-andrews.ac.uk
This work was supported by KPMG. The author thanks Samarjit Das for helpful discussions, and Elja Arjas, Anup Dewanji, Sean Holly, Hashem Pesaran and Debasis Sengupta for useful comments and suggestions. The usual disclaimer applies.
A very early version was circulated as: Bhattacharjee, A. and Das. S. (2002). Testing proportionality in duration models with respect to continuous covariates. DAE Working Paper no. **0220**, Department of Applied Economics, University of Cambridge, UK.

1 Introduction

The Cox regression model (Cox, 1972) plays a very prominent role in the theory and practice of survival analysis. Indeed, the model, and more generally the proportional hazards (PH) model, provides a convenient way to evaluate the influence of one or several covariates on the probability of termination of lifetime or duration spells. However, the PH specification substantially restricts interdependence between the explanatory variables and the lifetime in determining the hazard. In particular, the Cox PH model restricts coefficients of the regressors in the logarithm of the hazard function to be constant over the lifetime. This may not hold in many situations, or may even be unreasonable from the point of view of relevant theory. Since such misspecifications lead to misleading inferences about the effects of explanatory variables and shape of the baseline hazard, testing the PH assumption has been an area of active research.

Most of the analytical tests are either omnibus tests or tests in which the PH model is embedded in a larger class of semiparametric models. However, many of these tests are not satisfactory. While the omnibus tests usually have low power, the semiparametric alternatives typically make unverifiable assumptions about the shape of the regression function. Further, when the PH assumption does not hold, applied researchers require additional information regarding the nature of the covariate effects. In this context, it is often useful to explore whether the hazard rate for one level of the covariate increases in lifetime relative to another level, particularly when the covariate is discrete (two-sample or k-sample setup).[1]

In the two-sample setup, Gill and Schumacher (1987) and Deshpande and Sengupta (1995) developed analytical tests of the PH hypothesis against the alternative of 'increasing hazard ratio', which is equivalent to convex partial order of the lifetime distribution in the two samples.[2] Under the same setup, Sengupta *et al.* (1998) proposed a test of the PH model against the weaker alternative hypothesis of 'increasing ratio of cumulative hazards' (star ordering of the two samples). The above alternative hypotheses ('increasing hazard ratio' and 'increasing ratio of cumulative hazards') provide explanations for the phenomenon of 'crossing hazards' often found in applications.

These two-sample tests are useful for analysing survival data because, not only are they powerful in detecting departures from proportionality, they also provide further clues about the nature of covariate dependence. However,

[1] This kind of situation could arise, for example, if the coefficient of the covariate is not constant over time, or is dependent on some other (possibly unobserved) covariate.

[2] Throughout this paper, the word 'increasing' means 'non-decreasing', and 'decreasing' means 'non-increasing'.

their applicability is limited because many important covariates in biomedical or economic applications are continuous in nature (Horowitz and Neumann, 1992).

In this paper, we extend partial orders in the above two-sample problems to the case of continuous covariates. Based on examples from the applied literature as well as new applications, we argue that similar ordered departures are also common in the case of continuous covariates and provide meaningful alternatives to the PH model. We propose tests of the PH model against such ordered departures and study their asymptotic properties. Our framework does not assume any specific underlying regression model, and the tests are applicable in the presence of additional covariates – observed or unobserved. Monte Carlo studies and applications to real data highlight the advantages of the proposed methods.

In Section 2, we develop notions of ordered alternatives to the PH model in the case of continuous covariates. Tests of the PH assumption against such partial orders are constructed and their asymptotic properties studied in Section 3, and issues regarding implementation and extensions are discussed in Section 4. Small sample properties are studied in Section 5, while two real life applications are presented in Section 6. We also discuss modeling non-proportional covariate effects and develop a related graphical test. Section 7 concludes.

2 Partial orders with respect to a continuous covariate

Partial orders of lifetime distributions are commonly used in theory and applications. The two most popular notions of partial ordering, namely convex ordering and star ordering (Kalashnikov and Rachev, 1986; Sengupta and Deshpande, 1994), offer useful interpretations in terms of monotonicity of ratios of hazard and cumulative hazard functions respectively over time. Therefore, they describe useful and intuitively appealing ways to characterise departures from the PH model in two samples and in the competing risks framework. Gill and Schumacher (1987), Deshpande and Sengupta (1995) and Sengupta *et al.* (1998) consider several empirical applications where the departure from the PH model in two samples is evident from the fact that the ratio of the hazard rates is not constant over the lifetime; see also Andersen (1998).

For the two-sample setup, Gill and Schumacher (1987) and Deshpande and Sengupta (1995) developed tests of the PH model against the "increas-

ing hazard ratio" alternative, which is equivalent to convex ordering of the life-time distribution in one sample with respect to the other. Sengupta *et al.* (1998) constructed a test against the weaker alternative hypothesis of "increasing ratio of cumulative hazards" (star ordering of the two samples).[3]

The following definitions describe natural extensions of the above partial orders to the continuous covariate case. Let T be a lifetime variable, X a continuous covariate and let $\lambda(t|x)$ denote the hazard rate of T, given $X = x$, at $T = t$.[4]

Definition 1. *The lifetime random variable T is defined to be increasing hazard ratio for continuous covariate ($IHRCC$) with respect to the covariate X if, whenever $x_1 > x_2$, $\lambda(t|x_1)/\lambda(t|x_2) \uparrow t$. In other words, the lifetime distribution conditional on the lower covariate value is convex ordered with respect to that conditional on the higher value:*

$$(T|X = x_1) \underset{c}{\prec} (T|X = x_2)..$$

The dual decreasing hazard ratio for continuous covariate ($DHRCC$) is correspondingly defined.

Definition 2. *The lifetime random variable T is defined to be increasing cumulative hazard ratio for continuous covariate ($ICHRCC$) with respect to X if, whenever $x_1 > x_2$,*

$$\Lambda(T|x_1)/\Lambda(t|x_2) \uparrow t \ (\equiv (T|X = x_1) \underset{*}{\prec} (T|X = x_2),$$

where $\underset{*}{\prec}$ denotes star ordering of the conditional lifetime distributions. The dual decreasing cumulative hazard ratio for continuous covariate ($DCHRCC$) is correspondingly defined.

Definition 3. *The lifetime random variable T is defined to be increasing then decreasing hazard ratio for continuous covariate ($IDHRCC$) with respect to the covariate X if, there exists a point x within the range of X such that, T is $IHRCC$ on the interval $(-\infty, x)$ and $DHRCC$ on the interval (x, ∞). Similarly, we can define decreasing then increasing hazard ratio for continuous covariate ($DIHRCC$).*

Definitions 1 and 2 describe notions of positive ageing with respect to a continuous covariate. The higher the covariate, the faster the ageing of the individual – a situation which is common in empirical studies. In biomedical

[3]Sengupta and Bhattacharjee (1994), Deshpande and Sengupta (1995) and Dauxois and Kirmani (2004) extend these tests to the competing risks problem.

[4]See Fleming and Harrington (1991) for related discussion.

applications, such monotonically time-dependant covariate effects have been discussed both under additive hazard models (Aalen, 1980; Mau, 1986) and multiplicative models (Anderson and Senthilselvan, 1982; Andersen *et al.*, 1993).

Examples of such partial orders are common in applications. For example, while analysing of survival with malignant melanoma, Andersen *et al.* (1993) observe that, while "hazard seems to increase with tumor thickness" (pp. 389), the plot of estimated cumulative baseline hazards for patients with '2mm $\leq$ tumor thickness $<$ 5mm' and 'tumor thickness $\geq$ 5mm' against that of patients with 'tumor thickness $<$ 2mm' reveal "concave looking curves indicating that the hazard ratios decrease with time" (pp. 544–545). In fact, it is commonly observed in medical settings that treatment effects of an active drug decays with time (Therneau and Grambsch, 2000; Scheike and Martinussen, 2004). Similar evidence has also been noted in the applied econometrics literature. Using French data on unemplyment durations, Jayet and Moreau (1991) observe that the ratio of hazard function for individuals in the age groups 24–28 years to that for 37–40 years increases with duration of unemployment upto approximately 120 days.

Definition 3 describes a notion of non-monotonic departure from the PH model, with respect to the effect of a continuous covariate. An application considered later in the paper demonstrate evidence of such non-monotonic departures. The following examples illustrate some simple data generation processes (DGPs) that generate monotone and non-monotonic departures from the PH assumption with respect to a continuous covariate.

Example 1. Consider a hazard regression model with time varying coefficients (Murphy and Sen, 1991; Martinussen *et al.*, 2002), with the hazard function $\lambda(t|x) = \lambda_0(t).\exp(\beta(t).x)$, where x is a continuous covariate and $\beta(.)$ is an increasing function of lifetime t. This model is appropriate when the influence (prognostic value) of the covariate is expected to be higher at higher lifetimes. Then, if $x_1 > x_2$, $\lambda(t|x_1)/\lambda(t|x_2) = \exp(\beta(t).(x_1 - x_2))$ is increasing in t. In other words, the lifetime random variable T is $IHRCC$ with respect to the covariate X. Conversely, if $\beta(.)$ is an decreasing function of the lifetime, T would be $DHRCC$ with respect to X, a feature commonly observed in empirical studies.

Example 2. Consider a changepoint survival model given by the cumulative hazard function $\Lambda(t|x) = \Lambda_0(t).\exp(I\,(t > t^*)\,.\beta x)$, where x is the covariate, $I(.)$ the indicator function, and t^* is a lifetime in the interior of the sample space. This is a model where initially the covariate has no effect on the lifetime. The effect of the covariate begins as soon as the lifetime crosses a certain threshold t^*, and it lifts the distribution function upto a level where

it would have been, if the effect of the covariate would have persisted over the entire past life of the lifetime variable. If $\beta > 0$, this model is $ICHRCC$, but not $IHRCC$.[5] This kind of model may be useful in analysing the effect of active labour market programmes on unemployment duration, where the effect may become significant only around the time when unemployment benefits are terminated; see, for example, Narendranathan and Stewart (1993).

Example 3. Consider the hazard function $\lambda(t|x) = \lambda_0(t).\exp(\beta(t).|x-a|)$, where x is the covariate, a is a point on the covariate space, and $\beta(.)$ is an increasing function of lifetime t. This model is neither $IHRCC$ nor $DHRCC$, but it is $DIHRCC$; it is $IHRCC$ on one region of the covariate space ($x > a$), and $DHRCC$ on another region ($x < a$). An application where such a feature is observed is the effect on mother's age on infant mortality. Because of physiological reasons, mortality is lowest around an optimal childbearing age; however, keeping mother's age fixed, the effect itself declines with age of the child (Bhalotra and Bhattacharjee, 2001). Another application is considered later in the paper (Section 6).

As the above examples illustrate, the notions of ordering introduced in Definitions 1, 2 and 3 encompass a wide range of non-PH situations, and are potentially useful in many empirical applications. There may be a number of different explanations for changes in the covariate effects over lifetime. In fact, in many applications, monotone departures from the PH model may be more reasonable even from a theoretical point of view. Examples include medical applications where one expects the prognostic relevance of some covariates to decay, or even disappear, in the long run (Pocock *et al.*, 1982; Therneau and Grambsch, 2000). Similar decline in covariate effects are observed in economic studies on the effect of benefits on unemployment duration (Narendranathan and Stewart, 1993) and on the effect of macroeconomic conditions on firm exits (Bhattacharjee *et al.*, 2007). Construction of tests of the PH model against monotone alternatives with respect to continuous covariates is therefore important.

The above examples also demonstrate typical patterns of time varying coefficients when proportionality does not hold. These are useful for modeling ordered departures ($IHRCC$ or $DHRCC$) as well as non-monotonic violations ($IDHRCC$ or $DIHRCC$) of the PH assumption. Using the empirical applications (Section 6), we will demonstrate how such time varying covariate effects can be used, in combination with the proposed tests, to draw useful inferences in non-PH situations.

[5] The distribution function here has a jump discontinuity, but one can construct examples where $ICHRCC$ holds, and the distribution function is absolutely continuous.

3 Test statistics

Several two-sample tests of the PH model against monotone alternatives exist in the literature. For a continuous covariate, a natural approach for testing the PH assumption against ordered alternatives $IHRCC$ and $ICHRCC$ (and their duals) would be repeated applications of the corresponding tests in the two-sample setup. In this paper, we consider the two-sample test statistics proposed in Gill and Schumacher (1987) (T_{GS}) and Sengupta *et al.* (1998) (T_{SBR}).

Taking this approach, we propose a simple construction of our tests as follows. First, we fix a positive integer $r > 1$, and randomly select r pairs of distinct points on the covariate space. Next, for each pair, we construct the two-sample standardised test statistics (T_{GS} and T_{SBR}) based on counting processes conditional on the two distinct covariate values. Finally, our test statistics are constructed by taking maxima, minima or average of these basic test statistics over the r pairs.

3.1 Monotone hazard ratio

For the alternative of 'increasing hazard ratio' (convex partial order) in two samples, Gill and Schumacher (1987) propose the test statistic

$$T_{GS,std} = \frac{T_{GS}}{\sqrt{\widehat{\mathrm{Var}}\left[T_{GS}\right]}}, \tag{1}$$

where

$$\begin{aligned}
T_{GS} &= T_{11}T_{22} - T_{12}T_{21}, && (2)\\
\widehat{\mathrm{Var}}\left[T_{GS}\right] &= T_{21}T_{22}V_{11} - T_{21}T_{12}V_{12} - T_{11}T_{22}V_{21} + T_{11}T_{12}V_{22}, && (3)\\
T_{ij} &= \int_0^\tau L_i(t)d\widehat{\Lambda}_j(t), (i,j=1,2),\\
V_{ij} &= \int_0^\tau L_i(t)L_j(t)\{Y_1(t)Y_2(t)\}^{-1}d\left(N_1+N_2\right)(t), (i,j=1,2),
\end{aligned}$$

τ is a random stopping time,[6] $L_1(t)$ and $L_2(t)$ are two predictable processes, and for the j-th sample ($j = 1, 2$), $\Lambda_j(t)$ is the cumulative hazard function and $\widehat{\Lambda}_j(t)$ its Nelson-Aalen estimator, $Y_j(t)$ denotes the number of individuals on test at time t, and $N_j(t)$ the counting process for the number of failures in the sample at time t.

[6] For example, τ may be taken as the time at the final observation in the combined sample.

Gill and Schumacher (1987) show that the unstandardised test statistic (T_{GS}) has mean zero under the null hypothesis (PH) and positive (negative) mean if the hazard ratio $\lambda_1(t)/\lambda_2(t)$ is monotonically increasing (decreasing) in t on $[0, \infty)$ and $L_1(.)$ and $L_2(.)$ are so chosen that $L_1(t)/L_2(t)$ is monotonically decreasing, and that its standard error falls to zero as sample size increases to ∞ under both the null and alternative hypotheses. Hence, while the standardized test statistic $T_{GS,std}$ is asymptotically standard normal under the null hypothesis, the mean increases (decreases) to ∞ ($-\infty$) under the alternative hypotheses of monotonically increasing (decreasing) hazard ratio. In many applications, L_1 and L_2 are chosen corresponding to the Gehan-Wilkoxon and log rank tests, where $L_1 = Y_1 Y_2$ and $L_2 = Y_1 Y_2 (Y_1 + Y_2)^{-1}$, so that $L_1(t)/L_2(t)$ is monotonically decreasing in t.

For testing $H_0 : PH$ vs. $H_1 : IHRCC$, we propose the following procedure. We fix $r > 1$, and select $2r$ distinct points $\{x_{11}, x_{21}, \ldots, x_{r1}, x_{12}, x_{22}, \ldots, x_{r2}\}$ on the covariate space $\mathcal{X}$, such that $x_{l2} > x_{l1}, l = 1, \ldots, r$. We then construct our test statistics $T_{GS}^{(\max)}, T_{GS}^{(\min)}$ and $\overline{T}_{GS}$ based on the r statistics $T_{GS,std}(x_{l1}, x_{l2}), l = 1, \ldots, r$ (each testing convexity with respect to the pair of counting processes $N(t, x_{l1})$ and $N(t, x_{l2})$), where

$$\begin{aligned} T_{GS,std}(x_{l1}, x_{l2}) &= \frac{T_{GS}(x_{l1}, x_{l2})}{\sqrt{\widehat{\mathrm{Var}}\left[T_{GS}(x_{l1}, x_{l2})\right]}}, \\ T_{GS}(x_{l1}, x_{l2}) &= T_{l11} T_{l22} - T_{l12} T_{l21}, \\ \widehat{\mathrm{Var}}\left[T_{GS}(x_{l1}, x_{l2})\right] &= T_{l21} T_{l22} V_{l11} - T_{l21} T_{l12} V_{l12} - T_{l11} T_{l22} V_{l21} + T_{l11} T_{l12} V_{l22}, \end{aligned}$$

$$T_{lij} = \int_0^\tau L_i(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{lj}),$$

and

$$V_{lij} = \int_0^\tau L_i(x_{l1}, x_{l2})(t) L_j(x_{l1}, x_{l2})(t) \frac{d\left[N(t, x_{l1}) + N(t, x_{l2})\right]}{Y(t, x_{l1}) Y(t, x_{l2})}$$

for $i, j = 1, 2$.

Therefore, our test statistics are:

$$T_{GS}^{(\max)} = \max\left\{T_{GS,std}(x_{11}, x_{12}), T_{GS,std}(x_{21}, x_{22}), \ldots, T_{GS,std}(x_{r1}, x_{r2})\right\}, \quad (4)$$

$$T_{GS}^{(\min)} = \min\left\{T_{GS,std}(x_{11}, x_{12}), T_{GS,std}(x_{21}, x_{22}), \ldots, T_{GS,std}(x_{r1}, x_{r2})\right\}, \quad (5)$$

and

$$\overline{T}_{GS} = \frac{1}{r} \sum_{l=1}^{r} T_{GS,std}(x_{l1}, x_{l2}). \quad (6)$$

For the choice of L_1 and L_2 mentioned above, these statistics are close to zero under the null hypothesis. Under the alternative hypothesis $IHRCC$, $\overline{T}_{GS}$ and $T_{GS}^{(\max)}$ increases to ∞ as sample size increases, while under $DHRCC$, $\overline{T}_{GS}$ and $T_{GS}^{(\min)}$ decreases to $-\infty$. Under $IDHRCC$ or $DIHRCC$, $T_{GS}^{(\max)}$ and $T_{GS}^{(\min)}$ will both diverge, to ∞ and $-\infty$ respectively, as sample size increases to ∞.

3.2 Monotone cumulative hazard ratio

The form of the test statistic proposed by Sengupta *et al.* (1998), for testing the proportional hazards model against the 'increasing cumulative hazard ratio' (star partial order) alternative, is similar to $T_{GS,std}$, and given by

$$T_{SBR,std} = \frac{T_{SBR}}{\sqrt{\widehat{\mathrm{Var}}\left[T_{SBR}\right]}}, \tag{7}$$

where

$$\begin{aligned}
T_{SBR} &= S_{11}S_{22} - S_{12}S_{21}, && (8)\\
\widehat{\mathrm{Var}}\left[T_{SBR}\right] &= S_{21}S_{22}W_{11} - S_{21}S_{12}W_{12} - S_{11}S_{22}W_{21} + S_{11}S_{12}W_{22}, && (9)\\
S_{ij} &= \int_0^{\tau^*} K_i(t).\widehat{\Lambda}_j(t).dt, (i,j=1,2),\\
W_{ij} &= \int_0^{\tau^*}\int_0^{\tau^*} K_i(t).K_j(s).K_j(s).W\left(\min(s,t)\right)dsdt, (i,j=1,2),\\
W(t) &= \int_0^{t} \left(Y_1(s)Y_2(s)\right)^{-1}.d\left(N_1+N_2\right)(s),
\end{aligned}$$

τ^* is a large lifetime with $\Lambda_j(\tau^*) < \infty, j = 1, 2$,[7] and $K_j(t)(j = 1, 2)$ are right continuous functions with left limits (rcll functions) that need not be predictable processes.

This standardised test statistic is also asymptotically standard normal under the null hypothesis of proportional hazards, and asymptotically normal with mean diversing to ∞ ($-\infty$) accordingly as the cumulative hazard ratio $\Lambda_1(t)/\Lambda_2(t)$ is monotonically increasing (decreasing) in t on $[0, \infty)$ and K_1 and K_2 are so chosen that $K_1(t)/K_2(t)$ is a decreasing process.

As before, we construct our test statistics $T_{SBR}^{(\max)}, T_{SBR}^{(\min)}$ and $\overline{T}_{SBR}$ based on the r statistics $T_{SBR,std}(x_{l1}, x_{l2}), l = 1, \ldots, r$ (each testing star-ordering

[7] Note that, unlike τ in the Gill-Schumacher statistic T_{GS}, τ^* need not be a stopping time.

with respect to the pair of counting processes $N(t, x_{l1})$ and $N(t, x_{l2})$). Thus, we have:

$$T_{SBR}^{(\max)} = \max\left\{T_{SBR,std}(x_{11}, x_{12}), T_{SBR,std}(x_{21}, x_{22}), \ldots, T_{SBR,std}(x_{r1}, x_{r2})\right\}, \tag{10}$$

$$T_{SBR}^{(\min)} = \min\left\{T_{SBR,std}(x_{11}, x_{12}), T_{SBR,std}(x_{21}, x_{22}), \ldots, T_{SBR,std}(x_{r1}, x_{r2})\right\}, \tag{11}$$

and

$$\overline{T}_{SBR} = \frac{1}{r}\sum_{l=1}^{r} T_{SBR,std}(x_{l1}, x_{l2}). \tag{12}$$

3.3 Large sample results

We now derive the large sample results for the proposed test statistics, using the counting process methods (Gill and Schumacher, 1987; Andersen *et al.*, 1993) and an useful result on convergence of ordinary Stieljes integral of a stochastic process (Theorem 3.1 in Sengupta *et al.*, 1998). It is also indicated how these results can be used, in combination with extreme value theory, to obtain p-values of $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$.

Consider a counting processes $\{N(t,x) : t\epsilon[0,\tau], x\epsilon\mathcal{X}\}$, indexed on a continuous covariate x, with intensity processes $\{Y(t,x).\lambda(t,x)\}$ such that $\lambda(t,x) = \theta_x\lambda(t)$ for all t (under the null hypothesis of proportional hazards). As before, L_1 and L_2 denote two predictable processes, each indexed on a pair of distinct values of the continuous covariate x (*i.e.,* indexed on (x_1, x_2), $x_1 \neq x_2,\ x_1, x_2\epsilon\mathcal{X}$), and let τ be a stopping time. Similarly, let K_1 and K_2 be right continuous functions with left limits, which are each indexed on $\{(x_1, x_2),\ x_1 \neq x_2,\ x_1, x_2\epsilon\mathcal{X}\}$, and τ^* is a large positive time such that $\Lambda(\tau^*, x_i) < \infty$, $i = 1, 2$. Now, let r be a fixed positive integer ($r > 1$) and $\{x_{11}, x_{21}, \ldots, x_{r1}, x_{12}, x_{22}, \ldots, x_{r2}\}$ are $2r$ points on the covariate space $\mathcal{X}$, such that $x_{l2} > x_{l1}, l = 1, \ldots, r$.

Assumption 1 *For each $l, l = 1, 2, \ldots, r$, let $L_1(x_{l1}, x_{l2})(t)$ and $L_2(x_{l1}, x_{l2})(t)$ be predictable processes indexed on the pair of fixed covariate values (x_{l1}, x_{l2}).*
Assumption 2 *Let τ be a random stopping time. In particular, τ may be taken as the time at the final observation of the counting process $\Sigma_{l=1}^{r}\Sigma_{j=1}^{2}N(t, x_{lj})$. In principle, one could also have different stopping times $\tau(x_{l1}, x_{l2}), l = 1, \ldots, r$ for each of the r basic test statistics $T_{GS,std}(x_{l1}, x_{l2}), l = 1, \ldots, r$.*
Assumption 3 *The sample paths of $L_i(x_{l1}, x_{l2})$ and $Y(t, x_{li})^{-1}$ are almost surely bounded with respect to t, for $i = 1, 2$ and $l = 1, \ldots, r$. Further, for each $l = 1, \ldots, r$, $L_1(x_{l1}, x_{l2})$ and $L_2(x_{l1}, x_{l2})$ are both zero whenever $Y(t, x_{l1})$ or $Y(t, x_{l2})$ are.*

Assumption 4 *There exists a sequence* $a^{(n)}$, $a^{(n)} \longrightarrow \infty$ *as* $n \longrightarrow \infty$, *and fixed functions* $y(t,x)$, $l_1(x_{l1}, x_{l2})(t)$ *and* $l_2(x_{l1}, x_{l2})(t)$, $l = 1, \ldots, r$ *such that*

$$\sup_{t\epsilon[0,\tau]} \left| Y(t,x)/a^{(n)} - y(t,x) \right| \xrightarrow{P} 0 \qquad \text{as } n \to \infty, \quad \forall x \epsilon \mathcal{X}$$

$$\sup_{t\epsilon[0,\tau]} \left| L_i(x_{l1}, x_{l2})(t) - l_i(x_{l1}, x_{l2})(t) \right| \xrightarrow{P} 0 \quad \text{as } n \to \infty, \quad i = 1,2, l = 1, \ldots, r$$

where $|l_i(x_{l1}, x_{l2})(.)|$ *are bounded on* $[0,\tau]$ for each $i = 1,2$ and $l = 1, \ldots, r$, *and* $y^{-1}(., x)$ *is bounded on* $[0,\tau]$, *for each* $x \epsilon X$.[8]

Let the test statistics $T_{GS}^{(\max)}, T_{GS}^{(\min)}$ and $\overline{T}_{GS}$ be as defined earlier (4 – 6).

Theorem 1. *Let Assumptions 1 through 4 hold. Then, under* $H_0 : PH$, *as* $n \to \infty$,

(a) $P\left[T_{GS}^{(\max)} \leq z\right] \to [\Phi(z)]^r$,

(b) $P\left[T_{GS}^{(\min)} \geq -z\right] \to [\Phi(z)]^r$,

and

(c) $\sqrt{r}.\overline{T}_{GS} \xrightarrow{D} N(0,1)$,

where $\Phi(z)$ *is the distribution function of a standard normal variate.*

(Proof in Appendix.)

Corollary 1.

$$P\left[a_r \left\{T_{GS}^{(\max)} - b_r\right\} \leq z\right] \to \exp\left[-\exp(-z)\right] \text{ as } r \to \infty$$

and

$$P\left[a_r \left\{T_{GS}^{(\min)} + b_r\right\} \geq z\right] \to \exp\left[-\exp(z)\right] \text{ as } r \to \infty,$$

where $a_r = (2\ln r)^{1/2}$ *and* $b_r = (2\ln r)^{1/2} - \frac{1}{2}(2\ln r)^{-1/2}(\ln\ln r + \ln 4\pi)$.

(Proof in Appendix).

Corollary 2. *Given a vector* $\underline{w} = (w_1, w_2, \ldots, w_r)$ *of* r *weights, each possibly dependent on* x_{lj} $(l = 1,2,\ldots,r; j = 1,2)$ *but not on the counting processes* $N(t, x_{lj})$, *let us define the test statistics*

$$T_{GS,\underline{w}}^{(\max)} = \max_{l=1,\ldots,r} \{w_l.T_{GS,std}(x_{l1}, x_{l2})\},$$

$$T_{GS,\underline{w}}^{(\min)} = \min_{l=1,\ldots,r} \{w_l.T_{GS,std}(x_{l1}, x_{l2})\},$$

[8] The condition on probability limit of $Y(t,x)$ can be replaced by a set of weaker conditions. See, for example, Sengupta *et al.* (1998).

and

$$\overline{T}_{GS,\underline{w}} = \frac{\sum_{l=1}^{r} w_l . T_{GS,std}(x_{l1}, x_{l2})}{\sum_{l=1}^{r} w_l}.$$

Let Assumptions 1 through 4 hold. Then, under $H_0 : PH$*, as* $n \to \infty$*,*
(a) $P\left[T_{GS,\underline{w}}^{(\max)} \leq z\right] \to \prod_{l=1}^{r} [\Phi(z/w_l)]$*,*

(b) $P\left[T_{GS,\underline{w}}^{(\min)} \geq -z\right] \to \prod_{l=1}^{r} [\Phi(z/w_l)]$*,*

and

(c) $\frac{\sum_{l=1}^{r} w_l}{\sqrt{\sum_{l=1}^{r} w_l^2}} . \overline{T}_{GS,\underline{w}} \xrightarrow{D} N(0,1)$*,*

where $\Phi(z)$ *is the distribution function of a standard normal variate.*
(Proof in Appendix).

Theorem 1, along with Corollaries 1 and 2, establish the asymptotic results for testing proportionality against monotone hazard ratio alternatives ($IHRCC$ and $DHRCC$) as well as non-monotonic violations ($IDHRCC$ or $DIHRCC$) of the PH assumption.

Next, we derive similar results for partial orders based on cumulative hazard ratios.

Assumption 5 For each $l, l = 1, 2, \ldots, r$, let $K_1(x_{l1}, x_{l2})(t)$ and $K_2(x_{l1}, x_{l2})(t)$ be stochastic processes with sample paths in $D[0, \infty)$ (*i.e.*, are right continuous and have left limits).
Assumption 6 Let τ^* be a positive lifetime such that $\Lambda(\tau^*, x_{lj}) < \infty, l = 1, 2, \ldots, r, j = 1, 2$.
Assumption 7 There exists a sequence $a^{(n)}$, $a^{(n)} \to \infty$ as $n \to \infty$, and deterministic functions $y(t, x)$, $k_1(x_{l1}, x_{l2})(t)$ and $k_2(x_{l1}, x_{l2})(t)$, $l = 1, \ldots, r$ such that

$$\sup_{t \epsilon [0, \tau^*]} \left| Y(t,x)/a^{(n)} - y(t,x) \right| \xrightarrow{P} 0 \qquad \text{as } n \to \infty, \quad \forall x \epsilon \mathcal{X}$$

$$\sup_{t \epsilon [0, \tau^*]} \left| K_i(x_{l1}, x_{l2})(t) - k_i(x_{l1}, x_{l2})(t) \right| \xrightarrow{P} 0 \quad \text{as } n \to \infty, \quad i = 1, 2, l = 1, \ldots, r$$

where $k_1(x_{l1}, x_{l2})(t)$ and $k_2(x_{l1}, x_{l2})(t)$, $l = 1, \ldots, r$ are continuous functions with respect to t, and $y^{-1}(., x)$ is bounded on $[0, \tau]$, for each $x \epsilon \mathcal{X}$.

Let the test statistics $T_{SBR}^{(\max)}, T_{SBR}^{(\min)}$ and $\overline{T}_{SBR}$ be as defined earlier (10 – 12).
Theorem 2. *Let Assumptions 5 through 7 hold. Then, under* $H_0 : PH$*, as* $n \to \infty$*,*

(a) $P\left[T_{SBR}^{(\max)} \leq z\right] \rightarrow [\Phi(z)]^r$,

(b) $P\left[T_{SBR}^{(\min)} \geq -z\right] \rightarrow [\Phi(z)]^r$,

and
(c) $\sqrt{r}\overline{T}_{SBR} \xrightarrow{D} N(0,1)$,
where $\Phi(z)$ *is the distribution function of a standard normal variate.*
(Proof in Appendix.)

Corollary 3.

$$\begin{aligned} P\left[a_r\left\{T_{SBR}^{(\max)} - b_r\right\} \leq z\right] &\rightarrow \exp\left[-\exp(-z)\right] \text{ as } r \rightarrow \infty \quad \text{and} \\ P\left[a_r\left\{T_{SBR}^{(\min)} + b_r\right\} \geq z\right] &\rightarrow \exp\left[-\exp(z)\right] \text{ as } r \rightarrow \infty, \\ \text{where } a_r &= (2\ln r)^{1/2}, \\ \text{and } b_r &= (2\ln r)^{1/2} - \frac{1}{2}(2\ln r)^{-1/2}(\ln\ln r + \ln 4\pi). \end{aligned}$$

(Proof in Appendix.)

Corollary 4. *Given a vector* $\underline{w} = (w_1, w_2, \ldots, w_r)$ *of* r *weights, each possibly dependent on* x_{lj} *(*$l = 1, 2, \ldots, r; j = 1, 2$*) but not on the counting processes* $N(t, x_{lj})$*, let us define the test statistics*

$$T_{SBR,\underline{w}}^{(\max)} = \max_{l=1,\ldots,r}\left\{w_l.T_{SBR,std}(x_{l1}, x_{l2})\right\},$$

$$T_{SBR,\underline{w}}^{(\min)} = \min_{l=1,\ldots,r}\left\{w_l.T_{SBR,std}(x_{l1}, x_{l2})\right\},$$

$$\text{and } \overline{T}_{SBR,\underline{w}} = \frac{\sum_{l=1}^r w_l.T_{SBR,std}(x_{l1}, x_{l2})}{\sum_{l=1}^r w_l}.$$

Let Assumptions 5 through 7 hold. Then, under $H_0 : PH$*, as* $n \rightarrow \infty$*,*
(a) $P\left[T_{SBR,\underline{w}}^{(\max)} \leq z\right] \rightarrow \prod_{l=1}^r [\Phi(z/w_l)]$,

(b) $P\left[T_{SBR,\underline{w}}^{(\min)} \geq -z\right] \rightarrow \prod_{l=1}^r [\Phi(z/w_l)]$,

and
(c) $\frac{\sum_{l=1}^r w_l}{\sqrt{\sum_{l=1}^r w_l^2}}\overline{T}_{SBR,\underline{w}} \xrightarrow{D} N(0,1)$,

where $\Phi(z)$ *is the distribution function of a standard normal variate.*
(Proof in Appendix).

Remark 1. Restricting the statistics $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$ to depend on a fixed number (r) of distinct pairs of points is crucial for the asymptotic results. This is because, the processes $T_{GS,std}(x_1, x_2)$ and $T_{SBR,std}(x_1, x_2)$ on the space $\{(x_1, x_2) : x_2 > x_1, x_1, x_2 \epsilon \mathcal{X}\}$ are pointwise standard normal and independent, and therefore the maxima (minima) diverges to $+\infty(-\infty)$ without having well-defined asymptotic distributions.

Remark 2. Corollaries 1 and 3 provide simple ways to calculate the p-values for the extremal test statistics $T_{GS}^{(\max)}$ and $T_{GS}^{\min}$ (and similarly, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$) provided r is reasonably large. Note that since r is held fixed it cannot increase to ∞, but with a value large enough (say, 20 or higher) the approximation is quite accurate.

Remark 3. Corollaries 2 and 4 can be used to weight the underlying test statistics by some measure of the distance between x_{l1} and x_{l2}. In other words, one can give higher weights to a covariate pair where the covariates are further apart. In practice, this is expected improve the empirical performance of the tests[9]. We have, however, not used these weights in the empirical work in Sections 5 and 6.

4 Implementation and extensions

In this Section, we discuss some issues regarding implementation of the proposed tests, particularly in small samples, and extensions to other cases.

4.1 Small sample correction

Since the covariate under consideration is continuous, it is not feasible to construct the basic tests (T_{GS} and T_{SBR}) based solely on two distinct fixed points on the covariate space. In our implementation, we consider "small" intervals around the (randomly) chosen points, assuming the hazard function within these intervals to be approximately constant over covariate values. While the asymptotic distributions in Section 3 are based on specified points in the covariate space, the tests will be valid for small intervals around these points, provided the hazard function (for $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$ and $\overline{T}_{GS}$) or the cumulative hazard function (for $T_{SBR}^{(\max)}$, $T_{SBR}^{(\min)}$ and $\overline{T}_{SBR}$) is continuous at these points.

However, in small samples, these intervals often overlap, causing independence of the basic test statistics to be violated. Our Monte Carlo studies

[9]The author is grateful to Elja Arjas for pointing out the usefulness of the weighted average in this context.

suggest that the average test statistics are susceptible to this problem, resulting in a sample variance larger than $1/r$. We suggest making a small sample correction in such cases, by normalizing the average statistic using a jacknife or bootstrap (subsample) estimate of the standard error. In this paper, we have used the Quenouille-Tukey jacknife variance estimator for this purpose. This adjustment improves the performance of the tests in small samples, and does not affect our asymptotic results. We denote these adjusted test statistics as $\overline{T}_{GS,Adj}$ and $\overline{T}_{SBR,Adj}$ respectively.

4.2 Choice of r and covariate pairs

The proposed tests take r, the number of covariate pairs, as fixed *a priori*. If the chosen value is sufficiently high (say, 20 or more), Corollaries 1 and 3 can be used to compute p-values very easily; the choice of r is not very critical otherwise. For the Monte Carlo study reported in Section 5, we choose $r = 45$.

However, the choice of covariate pairs can be quite critical for the performance of the tests. Typically, the choice will have to take account of the design density in an appropriate way. This is to ensure that the underlying two sample tests (T_{GS} and T_{SBR}) are based on reasonable sample sizes and on representative samples of the covariate values.

We considered three methods to choose covariate pairs. In the first aproach, we resample from the realised covariate distribution using a simple bootstrap. Once covariate values are selected, we computed T_{GS} and T_{SBR} based on small samples of 20 nearest neighbour observations corresponding to each chosen value. Our second approach was the nonparametric bootstrap using a kernel estimate of the design density. This should work better particularly in regions where covariate values are sparse. The samples were constructed as in the previous approach. Third, we divided the sample observations into deciles based on the covariate values, and then chose the $\binom{10}{2} = 45$ combinations given by the partition.

All the three approaches gave comparable results in our Monte Carlo experiments. We, however, prefer the third approach because of its simplicity and its advantages of generating non-overlapping intervals and adequately covering the covariate space.

4.3 Comparison with other tests

As discussed earlier, a convenient way to interpret the ordered alternatives considered here is through time varying coefficients in a multiplicative hazard

regression model. In this sense, our tests are somewhat related to other analytical tests of time-dependant covariate effects proposed in the literature.

However, our approach embodies several important points of departure. First, our tests are based on the partial orders defined in Section 2 and not on any restrictive regression model. Second, some of the available analytical tests are based on partitioning the sample space of the lifetime variable into intervals (Anderson and Senthilselvan, 1982; Murphy, 1993) and consequently do not make use of the full information that the data offers. Our tests do not have this shortcoming. Third, unlike some other tests (Grambsch and Therneau, 1994; Scheike and Martinussen, 2004), our methods enable us to identify useful non-monotonic departures from the PH model, like $IDHRCC$ and $DIHRCC$. Fourth, while the previous tests merely identify violation of the constancy of covariate effects over the lifetime, our tests are based on explicit partial orders and provide additional insight into the nature of the regression relationship. This is useful for further inference and modeling. Finally, along with the test proposed by Scheike and Martunussen (2004), our tests have the advantage that tests of proportionality can be conducted sequentially for different covariates. This is often very useful in applications.

Notwithstanding these iportant differences, we compare the performance of the proposed tests against the popular test for time constant effects (PH model) due to Grambsch and Therneau (1994), using a simulation study (Section 5).

4.4 Choice between the proposed tests

The choice between the maxima, minima and average test statistics can be important in practice. The maxima and minima tests detect more complicated departures from the PH model ($IDHRCC$, $DIHRCC$, and their counterparts based on the cumulative hazard functions), and thereby facilitate detailed investigation of ordered covariate effects. On the other hand, as we shall see in the Monte Carlo simulations (Section 5), the adjusted average statistics outperform the maxima and minima tests in terms of power.

4.5 Extensions

The proposed ethodology offers several straightforward extensions.

4.5.1 k-sample problem

The proposed tests can be used to study monotone departures in k-sample (discrete covariate) problems. In this case, an *a priori* ordering of the k

samples can be obtained using estimators of hazard ratio proposed in Gill and Schumacher (1987) or Sengupta *et al.* (1998), or using the tree-structured modeling approach (Ahn and Loh, 1994). One can then easily apply the test for the PH model proposed here. The tests can also be similarly extended to the competing risks problem with more than 2 competing risks.

4.5.2 Different censoring and sampling plans

While our proposed methods are developed under the standard random censorship model (Fleming and Harrington, 1991; Andersen *et al.*, 1993), these can be easily extended to other censoring and sampling plans. For example, Bordes (2004) and Alvarez-Andrade *et al.* (2007) extend the counting process approach to estimation of the cumulative hazard function and proportional hazards regression based on progressive type-II censoring. Their results can be easily used to extend our results to this setup. Similarly, Sellke and Siegmund (1983) extend partial likelihood inference under the Cox regression model to the case of staggered (delayed) entry. Here, the counting process approach does not work. However, large sample results for our tests can still be derived using Theorem 3.1 of Sengupta *et al.* (1998) in combination with our Theorem 2.

4.5.3 Frailty

Like in the case of staggered entry, the counting process approach is not applicable in the presence of frailty. Under the shared frailty model, where individuals are clustered *a priori* based on the value of their shared but unobserved frailty, "quasi partial likelihood" inference was developed in Spiekerman and Lin (1998) based on empirical process theory. Similar theory for the univariate frailty model with a known one-parameter frailty distribution is developed in Kosorok *et al.* (2004). In either case, combining Theorem 3.1 in Sengupta *et al.* (1998) with our Theorem 2 gives us asymptotic results for the test statistics.

4.5.4 Presence of other covariates

While the proposed method is presented in the context of a single covariate, it can be extended to a multiple covariate setup in several ways. First, we may assume that the other covariates have proportional effects on the hazard function, as in the Cox regression model. In this case, the usual Aalen-Breslow estimator of the cumulative baseline hazard function, conditional on different values of the index covariate, can be used to construct the tests. Large sample results follow in the same way as before.

Second, if it is suspected that some of the other covariates may have nonproportional effects, these can be accommodated by incorporating time varying coefficients for these covariates. In this case, the tests can be constructed using estimates of the cumulative baseline hazard function based on estimated cumulative baseline hazard function using the histogram sieve estimator proposed by Murphy and Sen (1991). The asymptotic arguments described above still follow. In fact, in general, we recommend starting with a model where all the covariates are allowed to have time varying effects, and then reduce the model by sequentially testing for proportionality of each covariate. This is similar to the approach in Scheike and Martinussen (2004).

Third, the proposed method can be used to nonparametrically study covariate effects in the context of more general regression models, without the assumption of time varying coefficients. For example, one could define the lifetime T to be $IHRCC$ with respect to continuous covariates X and Z if, whenever $x_1 > x_2$ and $z_1 > z_2$, $\lambda(t|x_1, z_1)/\lambda(t|x_2, z_2) \uparrow t$. More generally, one may define T to be $IHRCC$ with respect to X and Z if, given some function $h(.,.)$, $\lambda(t|x_1, z_1)/\lambda(t|x_2, z_2) \uparrow t$ whenever $h(x_1, z_1) > h(x_2, z_2)$. Further, the appropriate specification of the function $h(.,.)$, which will be typically application-specific, can be made from the values of the underlying two sample test statistics. A proposed graphical method, discussed later, may be particularly useful in this situation. This demonstrates the versatility of the proposed framework and methodology for studying covariate effects.

It is clear from the above discussion that, though the testing procedure is applied sequentially to individual covariates or a small number of covariates, its applicability is almost universal. This outlines the usefulness of the proposed methods.

5 Monte Carlo study

In this Section, we explore the finite sample performance of the tests for different specifications of the baseline hazard function and covariate dependence. The selected data generation processes are similar to those used in Horowitz (1999) and Martinussen *et al.* (2002). In particular, we consider models of the form

$$\lambda(t, x) = \lambda_0(t).\exp\left[\beta(t, x)\right], \tag{13}$$

where $\lambda_0(t)$ and $\beta(t, x)$ are chosen to assume a variety of functional forms. Note that, under model (13), the PH assumption holds if and only if $\beta(t, x)$ depends only on x. If, for fixed x, $\beta(t, x)$ increases (decreases) in t, we have the $IHRCC$ and $ICHRCC$ ($DHRCC$ and $DCHRCC$) alternatives. If, on the other hand, $\beta(t, x)$ increases in t over some range of the covariate space,

and decreases over another (as in Example 3), the alternatives $IDHRCC$ or $DIHRCC$ may hold. While the proposed average tests are consistent for ordered alternatives to the null hypothesis of proportional hazards, our maxima and minima tests are consistent in both monotonic and non-monotonic cases.

In addition to the proposed tests, we included in our study the popular test for proportionality proposed by Grambsch and Therneau (1994) (GT). While the GT test is designed for testing specific parametric departures in the single covariate case, it is known to be very powerful in detecting departures from the PH model. A simulation study in Scheike and Martinussen (2004) suggests that a particular implementation of the GT test has higher power than the test proposed in their paper. Hence, the GT test is a good benchmark for comparison.

Our Monte Carlo simulations are based on independent right-censored data from 8 data generating processes (DGPs), defined by combinations of 4 specifications of the regression function

$$\beta(t,x) = \begin{cases} 0 \\ x \\ \ln(t).x \\ \ln(t).|x| \end{cases}$$

and 2 specifications of the baseline hazard function $\lambda_0(t)$ $(= 2, 12t)$; see Table 1 for definitions and notations for the DGPs. Randomly right-censored data are generated using the Gauss 386 random number generator, where the covariate X is i.i.d. $U(-1,1)$, and the censoring time C is i.i.d. $U(0.2, 2.2)$. Of the 8, four DGPs belong to the null hypothesis of PH, and two have $IHRCC$ (also $ICHRCC$ specifications). The two remaining models, with $\beta(t,x) = \ln(t).|x|$, have $DIHRCC$ specifications, being $IHRCC$ and $ICHRCC$ over the range $x\epsilon[0,1]$ and $DHRCC$ and $DCHRCC$ over the range $x\epsilon[-1,0]$.

Table 2 reports, for each of the above 8 data generation processes, the observed rejection rates (in percentage) of each of the test statistics, at 5 per cent confidence level, for different sample sizes. The reported percentages of rejection are based on 1000 Monte Carlo simulations in each case, and asymptotic distributions are used to compute the cut-offs. The covariate values considered are midpoints of each decile of the empirical distribution of realised covariate samples. Our test statistics are computed based on 45 random pairs of points on the covariate space ($r = 45$) in each case, given by each distinct combination of the above covariate values. Conditional on each covariate value, a sample of 20 nearest neighbour data points are used to construct the underlying two-sample test statistics T_{GS} and T_{SBR}. For the maxima and minima tests, the one-sided cut-off for the relevant extreme

value approximation is used, while the average test statistics have the two sided normal cut-offs. As discussed earlier, the average test statistics are standardized using the Quenouille-Tukey jacknife estimator of variance, to account for small sample distortions.

The results show that the proposed tests have good power in small samples, except for DGP_{24}. This is not surprising since DGP_{24} is $DIHRCC$, possessing $IHRCC$ features over one-half of the covariate space, and $DHRCC$ over the other. Hence, when a pair of points are drawn at random from the covariate space, only a quarter of them may be expected to reflect the $IHRCC$ nature of the underlying data generating process, and another quarter would reflect the $DHRCC$ nature. When we increased the sample size to 1500, the rejection rates for $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$ rose to 77, 68, 61 and 83 per cent respectively. The GT test (Grambsch and Therneau, 1994) performed very poorly for both the non-monotonic DGPs (DGP_{14} and DGP_{24}).

Overall, our tests are powerful and maintain their nominal sizes in finite samples. By comparison, the GT test has serious deficiencies in not being able to maintain its nominal size under PH DGPs. However, its power is higher for the monotone alternatives. The results also reflect the strength of the maxima and minima test statistics in their ability to detect non-monotonic departures from the PH model (DGP_{14} and DGP_{24}).

6 Empirical applications

Now, we illustrate the use of the tests with two applications – to durations of contract strikes in the US (Kennan, 1985), and to survival with malignant melanoma (Drzewiecki and Andersen, 1982; Andersen *et al.*, 1993).

6.1 Data on Strike Durations

The data, reported in Kennan (1985), pertain to durations of 566 contract strikes in the U.S., each involving 1000 workers or more, beginning during the period January 1968 to December 1976. Since strike durations are also known to exhibit seasonal effects (Neumann, 1994), we use data on the 292 strikes beginning in the first half of each year (none of these failure times are censored).

Previous research also suggests that the level of production index significantly affects strike duration (Kennan, 1985; Neumann, 1994). Higher values of the production index are associated with higher conditional probability of ending the strike, implying significant counter cyclical pattern of strike du-

ration. However, the PH model specifies much more than merely the sign of the covariate effect. In order to graphically explore whether the data exhibit monotone departures from the PH model, we use Lee-Pirie plots (Lee and Pirie, 1981) of cumulative hazard functions conditional on various randomly chosen pairs of covariate values. Many of these plots indicate an increasing ratio of the hazards, as evident from the convexity (in some cases, star-shapedness) of the plot lending credence to *a priori* suspicion of monotone ordering of the $IHRCC$ type; as an illustration, see Figure 1, the Lee-Pirie plot conditional on covariate values -0.048 and 0.037).

Next, we apply our tests to these data (Table 3). Each of the tests were based on 150 pairs of distinct covariate values. The results of the tests confirm our *a priori* notion based on the above plots. The null hypothesis of PH model is rejected in favour of the alternative $IHRCC$ (and $ICHRCC$), with production index as the continuous covariate.

This implies that the covariate effect of production index is such that, the duration distribution conditional on a higher value of the covariate is convex-ordered with respect to that conditional on a lower production index. In other words, the impact of production index on the hazard rate of strike duration increases in the duration of the strike. Further, the maxima and minima tests provide additional information on the covariate pairs for which the basic test statistics attain their extreme values, which may be useful for modeling the nature of departures from proportionality. The maxima test-statistic $T_{GS}^{(\max)}$ is attained for the covariate pair $\{-0.0478, 0.0371\}$. The test statistic $T_{GS}^{(\min)}$ (covariate pair 0.0371 and 0.0675) has a p-value of 0.054, which provides some evidence of concave-ordering towards the upper end of the covariate space ($IDHRCC$).

To illustrate how this $IDHRCC$ nature can be incorporated into a regression model of strike durations, we present parameter estimates for three different models in Table 4. Model 1 is a simple Cox PH model, with production index as the continuous covariate. In Model 2, we allow for time-varying coefficients using the histogram sieve estimator proposed in Murphy and Sen (1991).[10] This model accomodates monotone departures from proportionality, in the nature of $IHRCC$ or $DHRCC$. In Model 3, we allow the coefficient of the covariate to vary not only over failure time, but also for covariate values. More specifically, we allow the coefficients to be different for covariate values below and above 0.0371, enabling us to model departures of

[10]There are several other estimators for time varying coefficients; see Martinussen *et al.* (2002) for a review. We choose the histogram sieve estimator (Murphy and Sen, 1991) because of its simplicity, intuitive appeal and efficiency in the sense of attaining the variance bound given in Sasieni (1992).

the $IDHRCC$ or $DIHRCC$ type. Here again, we use the estimators given by Murphy and Sen (1991) for inference.

Model 1 indicates a significant impact of production index on the hazard rate of strike durations. However, this evidence is misleading. Model 3 estimates show that the true nature of covariate dependence is strikingly different. These time- and covariate-varying nature of the parameter estimates closely relate to the results of our analytical tests on the nature of covariate dependence. For lower values of the covariate, the coefficient increases with duration, while the opposite holds for higher covariate values.

6.2 A related graphical test

Plotting the contours of the underlying standardised test statistics on a covariate $\times$ covariate two-dimensional plane provides an useful graphical tool for inference on monotonic and non-monotonic departures considered in this paper. Figure 2 shows a contour diagram of the standardized test statistic $T_{GS,std}$ (smoothed using the Epanechnikov kernel) for the strike duration data. The significant height of the peaks and troughs indicate nonproportionality, and the shift in the slopes about the covariate value of approximately 0.04 indicate non-monotonic departures from proportionality about this point. The use of the plot here confirms the inference drawn from our analytical tests, and in particular helps in choosing the changepoint for the $IDHRCC$ pattern.

In applications with multiple covariates, similar graphical analysis can also provide valuable insights into the interaction between different covariates. With two continuous covariates x and z, one can obtain similar plots for different candidate functions $h(x, z)$ (see Section 4.5.4) to examine which of these provides the sharpest slopes in the contour plot. The candidate functions can sometimes be implied by the relevant application. For example, in survival of a series system with covariates measuring proneness to failure of the two components, the relevant function may be $\max(x, z)$. In other situations where there is no a priori knowledge about $h(.,.)$, one can either hypothesize linear functions of the form $x + \gamma z$, or find the function using regression methods. The identity of the covariate pairs with high (low) values for the maxima (minima) test statistics can be very helpful in such analyses.

6.3 Malignant Melanoma Data

These data pertain to 205 patients (148 of these are censored) with malignant melanoma (cancer of the skin) on whom a radical operation was performed

at the Department of Plastic Surgery, University Hospital of Odense, Denmark. The analysis of these data in Andersen *et al.* (1993) identifies tumor thickness as one of the most important prognostic factors. Further, Andersen *et al.* (1993) show that the Lee-Pirie plots of Nelson-Aalen estimates of the cumulative hazard functions for patients with '2mm $\leq$ tumor thickness $<$ 5mm' and 'tumor thickness $\geq$ 5mm' against that of patients with 'tumor thickness $<$ 2mm' are "concave looking curves", indicating possible violation of the PH model in favour of $DHRCC$. Similarly, the plot of the cumulative regression functions for log-thickness (Martinussen *et al.*, 2002) also indicate a distinct concave shape, though the constant coefficient estimate lies almost entirely within the 95 percent confidence band of their cumulative regression function estimates.

Our analytical tests (Table 5) based on 100 pairs of distinct covariate values show that $T_{GS}^{(\min)}$ and $T_{SBR}^{(\min)}$ are significant at 1 percent level and $T_{GS}^{(\max)}$ is significant at 5 percent level, but $\overline{T}_{GS,Adj}$ and $\overline{T}_{SBR,Adj}$ are not significant. Further, $T_{GS}^{(\min)}$ and $T_{GS}^{(\max)}$ are attained for covariate pairs $\{1.9, 7.7\}$ and $\{1.0, 1.8\}$ respectively. This provides partial support for the observation in Andersen *et al.* (1993), in that the null of PH is rejected in favour of the alternatives $DHRCC$ and $DCHRCC$ over the upper range of the covariate space. However, in patients with small tumors, there is some evidence of an $IHRCC$ pattern (probably the reason why $\overline{T}_{GS,Adj}$ and $\overline{T}_{SBR,Adj}$ are not significant). The inference from the Murphy-Sen histogram sieve estimators (Table 6) is similar.

This provides some evidence of the strength of the proposed methods in detecting non-proportional covariate effects which previous tests fail to identify.

The two applications considered here demonstrate the value of studying departures from the PH model with respect to continuous covariates in terms of monotonicity of the covariate effects. These examples also illustrate the use of our test statistics in identifying monotonic and non-monotonic structures in the data. Similar inference has been used in Bhattacharjee *et al.* (2007) and Bhalotra and Bhattacharjee (2001). The former is an application to business failures in the UK, and the latter to child mortality in India.

7 Conclusion

In this paper, we develop notions of partial ordering of lifetime distributions with respect to continuous covariates and propose tests of the PH model against such monotone or ordered departures. Departures of these kinds are

common in applications. Therefore, both empirical and theoretical work in lifetime models need to have a framework flexible enough to accomodate these kinds of covariate dependence. Unlike other tests available in the literature, the proposed methodology works in very general situations and does not require any assumptions on the underlying regression models. Further, the methods offer a great deal of flexibility in terms of accommodating the effects of other covariates, both observed and unobserved.

An important advantage of the tests is that they provide valuable insights into the pattern of covariate dependence where the PH assumption does not hold. Unlike other competing tests, this is true for both monotonic and non-monotonic covariate effects. The methods are therefore useful for regression modelling in non-PH cases. Further, since the proposed partial orders can be interpreted in terms of time varying coefficients, existing inference methods can be easily used. Monte Carlo evidence and real life examples demonstrate the strength and usefulness of the proposed framework based on partial orders as well as the tests developed here.

Several promising areas of future research emerge from the work in this paper. First, in the derivation of asymptotic results, we show that the basic underlying two-sample test statistics for distinct covariate pairs are independent of each other. This fact can be exploited to extend many familiar two-sample inference techniques to the case of continuous covariates. Second, research can be directed towards extension of the proposed tests to models with unrestricted univariate frailty. The notions of partial ordering introduced in this paper will be valid in this case, and one can in principle construct similar tests using estimators of the cumulative hazard function under such models. However, this inference problem is quite distinct from the one addressed here, because of identifiability restrictions and the different nature of estimators proposed in the literature (see, for example, Horowitz, 1999). Third, estimation of semiparametric regression models under order restrictions motivated by the current work is an area of considerable research potential. Some research has been reported in this area (Bhattacharjee, 2004), but further useful research can be conducted on classical and Bayesian order restricted inference on covariate effects. Fourth, it will be useful to develop further inference on the changepoint in non-monotonic models using covariate pairs corresponding to the maxima and minima tests. A somewhat related problem is inference on the unknown $h(.,.)$ function in the multiple covariate case. These problems will be retained for future work.

Appendix: Proofs of the Results

Proof of Theorem 1: It follows from Gill and Schumacher (1987) that, under PH, as $n \longrightarrow \infty$,

$$\begin{aligned} \left(a^{(n)}\right)^{1/2} T_{GS}\left(x_{l1}, x_{l2}\right) &\xrightarrow{D} N\left(0, \sigma^2_{GS,l}\right), \text{ and} \\ a^{(n)} \widehat{\text{Var}}\left[T_{GS}\left(x_{l1}, x_{l2}\right)\right] &\xrightarrow{\mathcal{P}} \sigma^2_{GS,l}, \end{aligned}$$

where

$$\begin{aligned} \sigma^2_{GS,l} &= \int_0^\tau \left[\bar{l}_2\left(x_{l1}, x_{l2}\right) l_1\left(x_{l1}, x_{l2}\right)(t) - \bar{l}_1\left(x_{l1}, x_{l2}\right) l_2\left(x_{l1}, x_{l2}\right)(t)\right]^2 \\ &\qquad \theta_{x_{l1}} \theta_{x_{l2}} \left(\frac{d\Lambda(t, x_{l1})}{y(t, x_{l2})} + \frac{d\Lambda(t, x_{l2})}{y(t, x_{l1})}\right), \\ \text{and } \bar{l}_i\left(x_{l1}, x_{l2}\right) &= \int_0^\tau l_i\left(x_{l1}, x_{l2}\right)(t) d\Lambda(t, x_{li}), \qquad i = 1, 2. \end{aligned}$$

so that,

$$T_{GS,std}\left(x_{l1}, x_{l2}\right) = \frac{T_{GS}\left(x_{l1}, x_{l2}\right)}{\sqrt{\widehat{\text{Var}}\left[T_{GS}\left(x_{l1}, x_{l2}\right)\right]}} \xrightarrow{D} N(0,1), \qquad l = 1, \ldots, r.$$

The proof of the Theorem would follow, if it further holds that $T_{GS,std}\left(x_{l1}, x_{l2}\right)$, $l = 1, \ldots, r$ are asymptotically independent. In other words,

$$\begin{bmatrix} T_{GS,std}\left(x_{11}, x_{12}\right) \\ T_{GS,std}\left(x_{21}, x_{22}\right) \\ \vdots \\ T_{GS,std}\left(x_{r1}, x_{r2}\right) \end{bmatrix} \xrightarrow{D} N\left(\underline{\mathbf{0}}, \mathbf{I}_r\right),$$

where $\mathbf{I}_r$ is the identity matrix of order r.

Following Gill and Schumacher (1987), let

$$Z^{(n)}_{lij} = \int_0^\tau L_i(x_{l1}, x_{l2})(t) d\left\{\widehat{\Lambda}(t, x_{lj}) - \Lambda(t, x_{lj})\right\}, \qquad (i, j = 1, 2; l = 1, \ldots, r).$$

Then

$$\begin{aligned} \left(a^{(n)}\right)^{1/2} Z^{(n)}_{lij} &= \left(a^{(n)}\right)^{1/2} \int_0^\tau L_i(x_{l1}, x_{l2})(t) \frac{dN(t, x_{lj}) - Y(t, x_{lj}) d\Lambda(t, x_{lj})}{Y(t, x_{lj})} \\ &\xrightarrow{D} \int_0^\tau l_i(x_{l1}, x_{l2})(t) dM(t, x_{lj}), \end{aligned}$$

where $M(t, x_{lj}), l = 1, \ldots, r, j = 1, 2$ are independent Gaussian processes with zero means, independent increments and variance functions

$$Var\left[M(t, x_{lj})\right] = \int_0^\tau \frac{d\Lambda\left(s, x_{lj}\right)}{y(s, x_{lj})}.$$

This follows from a version of Rebolledo's central limit theorem (see Andersen *et al.*, 1993), which states that the innovation martingales corresponding to components of a vector counting process are orthogonal, and the vector of these martingales asymptotically converge to a Gaussian martingale.

It follows, by a version of the δ-method proved in Gill and Schumacher (1987), that

$$\left(a^{(n)}\right)^{1/2} \begin{bmatrix} T_{GS,std}\left(x_{11}, x_{12}\right) \\ T_{GS,std}\left(x_{21}, x_{22}\right) \\ \vdots \\ T_{GS,std}\left(x_{r1}, x_{r2}\right) \end{bmatrix} \xrightarrow{D} \begin{bmatrix} \sum_{i,j=1}^2 \bar{l}^{1ij} \int_0^\tau l_i(x_{11}, x_{12})(t) dM(t, x_{1j}) \\ \sum_{i,j=1}^2 \bar{l}^{2ij} \int_0^\tau l_i(x_{21}, x_{22})(t) dM(t, x_{2j}) \\ \vdots \\ \sum_{i,j=1}^2 \bar{l}^{rij} \int_0^\tau l_i(x_{r1}, x_{r2})(t) dM(t, x_{rj}) \end{bmatrix}$$

where

$$\begin{aligned} \bar{l}^{lij} &= (-1)^{i+j} \bar{l}_{l,3-i,3-j} \\ \text{and } \bar{l}_{lij} &= \int_0^\tau l_i(x_{l1}, x_{l2})(t) d\Lambda(t, x_{lj}); \qquad l = 1, \ldots, r; i, j = 1, 2. \end{aligned}$$

Now, under $H_0 : PH, \bar{l}_{lij} = \theta_{x_{lj}} \bar{l}_i\left(x_{l1}, x_{l2}\right)$, so that

$$\begin{aligned} & \sum_{i,j=1}^2 \bar{l}^{lij} \int_0^\tau l_i(x_{l1}, x_{l2})(t) dM(t, x_{lj}) \\ = & \int_0^\tau \left[\bar{l}_{l22} l_1(x_{l1}, x_{l2})(t) - \bar{l}_{l12} l_2(x_{l1}, x_{l2})(t)\right] dM(t, x_{l1}) \\ & + \int_0^\tau \left[-\bar{l}_{l21} l_1(x_{l1}, x_{l2})(t) + \bar{l}_{l11} l_2(x_{l1}, x_{l2})(t)\right] dM(t, x_{l2}). \end{aligned}$$

It follows that

$$\begin{bmatrix} T_{GS}\left(x_{11}, x_{12}\right) \\ T_{GS}\left(x_{21}, x_{22}\right) \\ \vdots \\ T_{GS}\left(x_{r1}, x_{r2}\right) \end{bmatrix} \xrightarrow{D} N\left(\underline{\mathbf{0}}, \sum\right),$$

where $\sum = diag\left(\left(\sigma^2_{GS,l}\right)\right), l = 1, \ldots, r$, with

$$\begin{aligned} \sigma^2_{GS,l} &= \int_0^\tau \left[\bar{l}_{l22} l_1(x_{l1}, x_{l2})(t) - \bar{l}_{l12} l_2(x_{l1}, x_{l2})(t)\right]^2 \frac{d\Lambda(t, x_{l1})}{y(t, x_{l1})} \\ & + \int_0^\tau \left[-\bar{l}_{l21} l_1(x_{l1}, x_{l2})(t) + \bar{l}_{l11} l_2(x_{l1}, x_{l2})(t)\right]^2 \frac{d\Lambda(t, x_{l2})}{y(t, x_{l2})}. \end{aligned}$$

Further, following Gill and Schumacher (1987), it can be shown that $\sigma^2_{GS,l}$ can be consistently estimated by $\widehat{Var}\left[T_{GS}\left(x_{l1}, x_{l2}\right)\right]$. Hence, it follows that

$$\left[\begin{array}{c} T_{GS,std}\left(x_{11}, x_{12}\right) \\ T_{GS,std}\left(x_{21}, x_{22}\right) \\ \vdots \\ T_{GS,std}\left(x_{r1}, x_{r2}\right) \end{array}\right] \xrightarrow{D} N\left(\underline{\mathbf{0}}, \mathbf{I}_r\right),$$

where $\mathbf{I}_r$ is the identity matrix of order r.

Proofs of (a), (b) and (c) follow.

□

Proof of Corollary 1: Proof follows from the well known result in extreme value theory regarding the asymptotic distribution of the maximum of a sample of iid $N(0,1)$ variates (see, for example, Berman, 1992), and invoking the δ-method by noting that maxima and minima are continuous functions.

□

Proof of Corollary 2: From Theorem 1, we have:

$$\left[\begin{array}{c} T_{GS,std}\left(x_{11}, x_{12}\right) \\ T_{GS,std}\left(x_{21}, x_{22}\right) \\ \vdots \\ T_{GS,std}\left(x_{r1}, x_{r2}\right) \end{array}\right] \xrightarrow{D} N\left(\underline{\mathbf{0}}, \mathbf{I}_r\right),$$

where $\mathbf{I}_r$ is the identity matrix of order r.

The proof follows immediately.

□

Proof of Theorem 2: It follows from Sengupta *et al.* (1998) that, under H_0, as $n \longrightarrow \infty$,

$$\begin{aligned} \left(a^{(n)}\right)^{1/2} T_{SBR}\left(x_{l1}, x_{l2}\right) &\xrightarrow{D} N\left(0, \sigma^2_{SBR,l}\right), \text{ and} \\ a^{(n)} \widehat{\mathrm{Var}}\left[T_{SBR}\left(x_{l1}, x_{l2}\right)\right] &\xrightarrow{\mathcal{P}} \sigma^2_{SBR,l}, \end{aligned}$$

where

$$\begin{aligned} \sigma^2_{SBR,l} &= \int_0^\tau \int_0^\tau \left[c(t)c(s)V\left(\min(s,t), x_{l1}\right) + d(t)d(s)V\left(\min(s,t), x_{l2}\right)\right] dsdt, \\ V\left(t, x_{lj}\right) &= \int_0^\tau \frac{d\Lambda\left(s, x_{lj}\right)}{y(s, x_{lj})}, \qquad j = 1, 2, \\ c(t) &= s_2\left(x_{l2}\right) k_1\left(x_{l1}, x_{l2}\right)(t) - s_1\left(x_{l2}\right) k_2\left(x_{l1}, x_{l2}\right)(t), \\ d(t) &= s_2\left(x_{l1}\right) k_1\left(x_{l1}, x_{l2}\right)(t) - s_1\left(x_{l1}\right) k_2\left(x_{l1}, x_{l2}\right)(t), \\ \text{and } s_i\left(x_{lj}\right) &= \int_0^\tau k_i\left(x_{l1}, x_{l2}\right)(s).\Lambda(s, x_{lj}) ds, \qquad i = 1, 2, j = 1, 2. \end{aligned}$$

so that,

$$T_{SBR,std}\left(x_{l1},x_{l2}\right)=\frac{T_{SBR}\left(x_{l1},x_{l2}\right)}{\sqrt{\widehat{Var}\left[T_{SBR}\left(x_{l1},x_{l2}\right)\right]}}\xrightarrow{D}N(0,1),\qquad l=1,\ldots,r.$$

Like Theorem 1, the proof will follow, if it further holds that

$$\left[\begin{array}{c} T_{SBR,std}\left(x_{11},x_{12}\right) \\ T_{SBR,std}\left(x_{21},x_{22}\right) \\ \vdots \\ T_{SBR,std}\left(x_{r1},x_{r2}\right) \end{array}\right]\xrightarrow{D}N\left(\underline{\mathbf{0}},\mathbf{I}_r\right),$$

where $\mathbf{I}_r$ is the identity matrix of order r.

The essential difference in the arguments required to establish asymptotic distributions here, from those in Theorem 1, lie in the fact that the integrals considered in Theorem 1 are transformations of stochastic integrals, while here they are functions of ordinary Steiljes integrals of stochastic processes.

Let us define

$$Z_{lij}^{*(n)}=\int_0^\tau K_i(x_{l1},x_{l2})(t)\left\{\widehat{\Lambda}(t,x_{lj})-\Lambda(t,x_{lj})\right\}dt,\qquad (i,j=1,2;l=1,\ldots,r).$$

Then, by Rebolledo's central limit theorem and Theorem 3.1 in Sengupta *et al.* (1998), we have, as $n\to\infty$,

$$\left(a^{(n)}\right)^{1/2}Z_{lij}^{*(n)}\xrightarrow{D}\int_0^\tau k_i(x_{l1},x_{l2})(t)M(t,x_{lj})dt,$$

where $M(t,x_{lj}),l=1,\ldots,r,j=1,2$ are independent Gaussian processes with zero means, independent increments and variance functions

$$Var\left[M(t,x_{lj})\right]=\int_0^\tau\frac{d\Lambda\left(s,x_{lj}\right)}{y(s,x_{lj})}.$$

Now, as in Theorem 1, invoking the δ-method of Gill and Schumacher (1987), it follows that

$$\left(a^{(n)}\right)^{1/2}\left[\begin{array}{c} T_{SBR,std}\left(x_{11},x_{12}\right) \\ T_{SBR,std}\left(x_{21},x_{22}\right) \\ \vdots \\ T_{SBR,std}\left(x_{r1},x_{r2}\right) \end{array}\right]\xrightarrow{D}\left[\begin{array}{c} \sum_{i,j=1}^2\overline{k}^{1ij}\int_0^\tau k_i(x_{11},x_{12})(t)M(t,x_{1j})dt \\ \sum_{i,j=1}^2\overline{k}^{2ij}\int_0^\tau k_i(x_{21},x_{22})(t)M(t,x_{2j})dt \\ \vdots \\ \sum_{i,j=1}^2\overline{k}^{rij}\int_0^\tau k_i(x_{r1},x_{r2})(t)M(t,x_{rj})dt \end{array}\right]$$

where

$$\begin{aligned}\overline{k}^{lij} &= (-1)^{i+j}\overline{k}_{l,3-i,3-j} \\ \text{and } \overline{k}_{lij} &= \int_0^\tau k_i(x_{l1},x_{l2})(t)\Lambda(t,x_{lj})dt;\qquad l=1,\ldots,r;i,j=1,2,\end{aligned}$$

and under H_0,

$$\begin{aligned}
&\sum_{i,j=1}^{2} \overline{k}^{lij} \int_0^\tau k_i(x_{l1}, x_{l2})(t) M(t, x_{lj}) dt \\
= &\int_0^\tau \left[\overline{k}_{l22} k_1(x_{l1}, x_{l2})(t) - \overline{k}_{l12} k_2(x_{l1}, x_{l2})(t)\right] M(t, x_{l1}) dt \\
&+ \int_0^\tau \left[-\overline{k}_{l21} k_1(x_{l1}, x_{l2})(t) + \overline{k}_{l11} k_2(x_{l1}, x_{l2})(t)\right] M(t, x_{l2}) dt.
\end{aligned}$$

It follows that

$$\begin{bmatrix} T_{SBR}(x_{11}, x_{12}) \\ T_{SBR}(x_{21}, x_{22}) \\ \vdots \\ T_{SBR}(x_{r1}, x_{r2}) \end{bmatrix} \xrightarrow{D} N\left(\underline{\mathbf{0}}, \sum\right),$$

where $\sum = diag\left(\left(\sigma^2_{SBR,l}\right)\right), l = 1, \ldots, r$, and following Sengupta *et al.* (1998), it can be shown that $\sigma^2_{SBR,l}$ can be consistently estimated by $\widehat{\text{Var}}\left[T_{SBR}(x_{l1}, x_{l2})\right]$. Hence, it follows that

$$\begin{bmatrix} T_{SBR,std}(x_{11}, x_{12}) \\ T_{SBR,std}(x_{21}, x_{22}) \\ \vdots \\ T_{SBR,std}(x_{r1}, x_{r2}) \end{bmatrix} \xrightarrow{D} N\left(\underline{\mathbf{0}}, \mathbf{I}_r\right),$$

where $\mathbf{I}_r$ is the identity matrix of order r.

Proofs of (a), (b) and (c) follow.

□

Proof of Corollary 3: Proof follows from extreme value theory and the δ-method, as in Corollary 1.

□

Proof of Corollary 4: From Theorem 2, we have:

$$\begin{bmatrix} T_{SBR,std}(x_{11}, x_{12}) \\ T_{SBR,std}(x_{21}, x_{22}) \\ \vdots \\ T_{SBR,std}(x_{r1}, x_{r2}) \end{bmatrix} \xrightarrow{D} N\left(\underline{\mathbf{0}}, \mathbf{I}_r\right),$$

where $\mathbf{I}_r$ is the identity matrix of order r.

The proof follows immediately.

□

References

[1] Aalen, O.O. (1980). A model for nonparametric regression analysis of counting processes. *Springer Lecture Notes in Statistics* **2**, 1–25.

[2] Ahn, H. and Loh, W.-Y. (1994). Tree-structured proportional hazards regression modeling. *Biometrics* **50**, 471–485.

[3] Alvarez-Andrade, S., Balakrishnan, N. and Bordes, L. (2007). Proportional hazards regression under progressive type-II censoring. *Annals of the Institute of Statistical Mathematics*, Forthcoming.

[4] Andersen, P.K. (1998). Hazard ratio estimator. *Encyclopedia of Biostatistics* **3**, 1829–1831. Wiley: New York.

[5] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1993). *Statistical Models based on Counting Processes*. Springer-Verlag: New York.

[6] Anderson, J.A. and Senthilselvan, A. (1982). A two-step regression model for hazard functions. *Applied Statistics* **31**, 44–51.

[7] Berman, S.M. (1992). *Sojourns and Extremes of Stochastic Processes*. Wadsworth and Brooks/ Cole, Pacific Grove, CA.

[8] Bhalotra, S.R. and Bhattacharjee, A. (2001). Understanding regional variations in child mortality in India. *Mimeo*. Paper presented at the *Workshop on 'Welfare, Demography and Development'*, University of Cambridge, Cambridge, UK.

[9] Bhattacharjee, A. (2004). Estimation in hazard regression models under ordered departures from proportionality. *Computational Statistics and Data Analysis* **47**, 517–536.

[10] Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2007). Macroeconomic conditions and business exit: determinants of failures and acquisitions of UK firms. *Economica* (forthcoming). Previous version circulated as: DAE Working Paper No. **0206**, Department of Applied Economics, University of Cambridge, UK, 2002.

[11] Bordes, L. (2004). Non-parametric estimation under progressive censoring. *Journal of Statistical Planning and Inference* **119**, 171–189.

[12] Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Scociety, Series B* **34**, 187–220.

[13] Dauxois, J.-Y. and Kirmani, S.N.U.A. (2004). On testing the proportionality of two cumulative incidence functions in a competing risks setup. *Journal of Nonparametric Statistics* **16**(3-4), 479–491.

[14] Deshpande, J.V. and Sengupta, D. (1995). Testing for the hypothesis of proportional hazards in two populations. *Biometrika* **82**, 251–261.

[15] Drzewiecki, K.T. and Andersen, P.K. (1982). Survival with malignant melanoma: a regression analysis of prognostic factors. *Cancer* **49**, 2414–2419.

[16] Fleming, T.R. and Harrington, D.P. (1991). *Counting Processes and Survival Analysis*. John Wiley and Sons: New York.

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

[18] Grambsch, P.M. and Therneau, T.M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, 515–526. Correction: (1995) **82**, pp. 668.

[19] Horowitz, J.L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica* **67**, 1001–1028.

[20] —— and Neumann, G.R. (1992). A generalised moments specification test of the proportional hazards model. *Journal of the American Statistical Association* **87**, 234–240.

[21] Jayet, H. and Moreau, A. (1991). Analysis of survival data: Estimation and specification tests using asymptotic least squares. *Journal of Econometrics* **48**, 263–285.

[22] Kalashnikov, V.V. and Rachev, S.T. (1986). Characterisation of queuing models and their stability. In Prohorov, Yu.K. *et al.* (Eds.), *Probability Theory and Mathematical Statistics*, VNU Science Press **2**, 37–53.

[23] Kennan, J.F. (1985). The duration of contract strikes in U.S. manufacturing. *Journal of Econometrics* **28**, 5–28.

[24] Kosorok, M.R., Lee, B.L. and Fine, J.P. (2004). Robust inference for univariate proportional hazards frailty regression models. *Annals of Statistics* **32**(4), 1448–1491.

[25] Lee, L. and Pirie, W.E. (1981). A graphical method for comparing trends in series of events. *Communications in Statistics - Theory and Methods* **10**, 827–848.

[26] Martinussen, T., Scheike, T.H. and Skovgaard, Ib M. (2002). Efficient estimation of fixed and time-varying covariate effects in multiplicative intensity models. *Scandinavian Journal of Statistics* **29**, 57–74.

[27] Mau, J. (1986). On a graphical method for the detection of time-dependent effects of covariates in survival data. *Applied Statistics* **35**, 245–255.

[28] Murphy, S.A. (1993). Testing for a time dependent coefficient in Cox's regression model. *Scandinavian Journal of Statistics* **20**, 35–50.

[29] Murphy, S.A. and Sen, P.K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochastic Processes and their Applications* **39**,153–180.

[30] Narendranathan, W. and Stewart, M.W. (1993). How does the benefit effect vary as unemployment spells lengthen? *Journal of Applied Econometrics* **8**, 361–381.

[31] Neumann, G.R. (1994). Search models and duration data. In Pesaran, M.H. (Ed.) *Handbook of Applied Econometrics Volume II: Microeconometrics*, Basil Blackwell: Oxford, Chapter 7, 300–351.

[32] Pocock, S.J., Gore, S.M. and Kerr, G.R. (1982). Long-term survival analysis: the curability of breast cancer. *Statistics in Medicine* **1**, 93–104.

[33] Sasieni, P.D. (1992). Information bounds for the additive and multiplicative intensity models (with discussion). In Klein, J.P. and Goel, P.K. (Eds.) *Survival Analysis: State of the Art*, Kluver: Boston, 249–265.

[34] Scheike, T.H. and Martinussen, T. (2004). On estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics* **31**, 51–62.

[35] Sellke, T. and Siegmund, D. (1983). Sequential analysis of the proportional hazards model. *Biometrika* **70**, 315–326.

[36] Sengupta, D. and Bhattacharjee, A. (1994). Testing the proportionality of hazards due to competing risks. *Mimeo.* Presented at the Workshop on *Topics in Reliability and Survival Analysis*, Indian Statistical Institute, Kolkata (India).

[37] Sengupta, D., Bhattacharjee, A., and Rajeev, B. (1998). Testing for the proportionality of hazards in two samples against the increasing cumulative hazard ratio alternative. *Scandinavian Journal of Statistics* **25**, 637–647.

[38] Sengupta, D. and Deshpande, J.V. (1994). Some results on the relative ageing of two life distributions. *Journal of Applied Probability* **31**, 991–1003.

[39] Spiekerman, C.F. and Lin, D.Y. (1998). Marginal regression models for multivariate failure time data. *Journal of the American Statistical Association* **93**, 1164–1175.

[40] Therneau, T.M. and Grambsch, P.M. (2000). *Modeling Survival Data: Extending the Cox Model.* Springer-Verlag: New York.

TABLE 1: Data Generating Processes

Model	$\lambda_0(t)$	$\beta(t,x)$	Median cens.	% cens.	Expected significance
DGP_{11}	2	0	0.36	16.4	None
DGP_{12}	2	x	0.30	19.2	None
DGP_{13}	2	$\ln(t).x$	0.25	15.8	$T_{GS}^{(\max)}, \overline{T}_{GS,Adj}, T_{SBR}^{(\max)}, \overline{T}_{SBR,Adj}, GT$
DGP_{14}	2	$\ln(t).\lvert x\rvert$	0.52	26.9	$T_{GS}^{(\max)}, T_{GS}^{(\min)}, T_{SBR}^{(\max)}, T_{SBR}^{(\min)}, GT?$
DGP_{21}	$12t$	0	0.32	8.9	None
DGP_{22}	$12t$	x	0.32	9.6	None
DGP_{23}	$12t$	$\ln(t).x$	0.30	8.9	$T_{GS}^{(\max)}, \overline{T}_{GS,Adj}, T_{SBR}^{(\max)}, \overline{T}_{SBR,Adj}, GT$
DGP_{24}	$12t$	$\ln(t).\lvert x\rvert$	0.42	13.8	$T_{GS}^{(\max)}, T_{GS}^{(\min)}, T_{SBR}^{(\max)}, T_{SBR}^{(\min)}, GT?$

TABLE 2: Rejection Rates (%) at the 5% Asymptotic Confidence Level

Model	Test	Sample size				Model	Test	Sample size			
		100	**200**	**500**	**1000**			**100**	**200**	**500**	**1000**
DGP_{11}	$T_{GS}^{(\max)}$	18.8	7.7	5.5	4.9	DGP_{21}	$T_{GS}^{(\max)}$	13.1	7.3	5.7	5.2
	$T_{GS}^{(\min)}$	23.0	7.5	5.4	5.0		$T_{GS}^{(\min)}$	21.4	8.0	4.5	5.1
	$\overline{T}_{GS,Adj}$	4.1	4.4	4.7	5.2		$\overline{T}_{GS,Adj}$	5.5	5.5	5.4	4.8
	$T_{SBR}^{(\max)}$	13.2	7.8	6.0	4.7		$T_{SBR}^{(\max)}$	11.8	7.0	5.6	4.8
	$T_{SBR}^{(\min)}$	12.9	7.1	5.6	4.9		$T_{SBR}^{(\min)}$	12.9	7.3	5.7	5.2
	$\overline{T}_{SBR,Adj}$	5.5	5.1	5.0	5.1		$\overline{T}_{SBR,Adj}$	15.2	6.0	4.9	5.0
	GT	4.5	4.1	4.7	5.8		GT	3.7	3.7	5.3	4.1
DGP_{12}	$T_{GS}^{(\max)}$	19.6	9.4	6.3	5.4	DGP_{22}	$T_{GS}^{(\max)}$	28.8	8.9	5.6	5.1
	$T_{GS}^{(\min)}$	18.2	7.9	5.7	4.8		$T_{GS}^{(\min)}$	16.4	8.8	6.4	4.6
	$\overline{T}_{GS,Adj}$	12.3	6.3	5.2	5.3		$\overline{T}_{GS,Adj}$	5.7	5.2	5.0	4.8
	$T_{SBR}^{(\max)}$	13.2	6.9	5.4	4.9		$T_{SBR}^{(\max)}$	12.5	7.7	5.5	5.1
	$T_{SBR}^{(\min)}$	16.9	8.1	5.8	5.2		$T_{SBR}^{(\min)}$	12.1	7.0	5.7	4.7
	$\overline{T}_{SBR,Adj}$	5.6	5.5	5.6	4.6		$\overline{T}_{SBR,Adj}$	3.1	3.9	4.4	5.3
	GT	1.6	1.5	2.6	2.3		GT	0.8	1.9	1.7	1.9
DGP_{13}	$T_{GS}^{(\max)}$	52.3	83.8	100.0	100.0	DGP_{23}	$T_{GS}^{(\max)}$	33.1	49.6	100.0	100.0
	$T_{GS}^{(\min)}$	11.9	6.1	0.5	0.0		$T_{GS}^{(\min)}$	13.1	5.4	1.9	2.0
	$\overline{T}_{GS,Adj}$	37.8	100.0	100.0	100.0		$\overline{T}_{GS,Adj}$	75.8	92.3	100.0	100.0
	$T_{SBR}^{(\max)}$	85.2	100.0	100.0	100.0		$T_{SBR}^{(\max)}$	14.8	26.6	98.3	100.0
	$T_{SBR}^{(\min)}$	4.4	0.1	0.0	0.4		$T_{SBR}^{(\min)}$	3.3	1.9	0.0	0.2
	$\overline{T}_{SBR,Adj}$	42.2	100.0	100.0	100.0		$\overline{T}_{SBR,Adj}$	86.1	98.2	100.0	100.0
	GT	99.1	100.0	100.0	100.0		GT	69.0	95.4	100.0	100.0
DGP_{14}	$T_{GS}^{(\max)}$	31.7	33.2	57.9	91.2	DGP_{24}	$T_{GS}^{(\max)}$	24.6	32.1	40.8	46.3
	$T_{GS}^{(\min)}$	29.4	42.1	70.6	94.8		$T_{GS}^{(\min)}$	22.0	29.1	49.5	53.2
	$\overline{T}_{GS,Adj}$	15.4	12.1	7.7	10.1		$\overline{T}_{GS,Adj}$	11.0	10.3	5.5	2.8
	$T_{SBR}^{(\max)}$	10.2	22.4	39.5	87.3		$T_{SBR}^{(\max)}$	11.2	19.8	35.9	45.4
	$T_{SBR}^{(\min)}$	21.1	33.9	75.2	97.8		$T_{SBR}^{(\min)}$	14.4	18.1	27.9	56.3
	$\overline{T}_{SBR,Adj}$	9.2	13.5	9.1	8.3		$\overline{T}_{SBR,Adj}$	13.9	10.2	4.1	4.6
	GT	2.7	2.4	2.4	2.7		GT	1.8	2.1	3.7	3.1

TABLE 3: Tests of the PH model: Strike Duration data

Test	**Test Statistic**	p**-Value** (%)
$T_{GS}^{(\max)}$	3.619	0.030
$T_{GS}^{(\min)}$	−3.426	0.054
$\overline{T}_{GS,Adj}$	4.093	0.000
$T_{SBR}^{(\max)}$	3.415	0.056
$T_{SBR}^{(\min)}$	−2.703	0.420
$\overline{T}_{SBR,Adj}$	3.808	0.000

TABLE 4: Model Estimates: Strike Duration data

Model/ Parameter	**Coefficient**	**z-stat.**
Model 1		
Production Index, x	3.529	3.17
Model 2		
$x.I\left[t\epsilon[0,75)\right]$	5.179	3.90
$x.I\left[t\epsilon[75,150)\right]$	0.360	0.27
$x.I\left[t\epsilon[150,\infty)\right]$	9.416	1.19
Model 3		
$x.I\left[x\epsilon(-\infty,0.037)\right].I\left[t\epsilon[0,75)\right]$	−1.178	−0.75
$x.I\left[x\epsilon(-\infty,0.037)\right].I\left[t\epsilon[75,150)\right]$	9.362	4.32
$x.I\left[x\epsilon(-\infty,0.037)\right].I\left[t\epsilon[150,\infty)\right]$	45.266	3.43
$x.I\left[x\epsilon[0.037,\infty)\right].I\left[t\epsilon[0,75)\right]$	10.173	4.96
$x.I\left[x\epsilon[0.037,\infty)\right].I\left[t\epsilon[75,150)\right]$	−14.910	−5.96
$x.I\left[x\epsilon[0.037,\infty)\right].I\left[t\epsilon[150,\infty)\right]$	−27.619	−5.90

TABLE 5: Tests of the PH model: Malignant Melanoma Data

Test	**Test Statistic**	p**-Value** (%)
$T_{GS}^{(\max)}$	3.462	0.035
$T_{GS}^{(\min)}$	−4.985	0.000
$\overline{T}_{GS,Adj}$	−1.080	0.188
$T_{SBR}^{(\max)}$	2.559	0.420
$T_{SBR}^{(\min)}$	−8.255	0.000
$\overline{T}_{SBR,Adj}$	−1.235	0.249

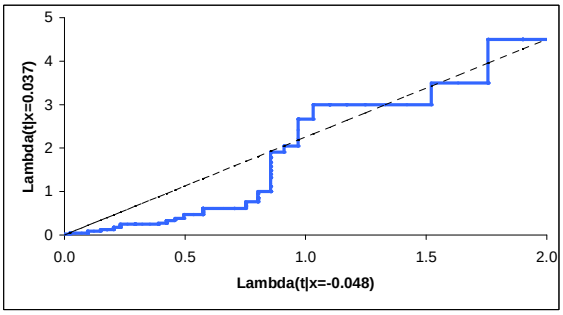


Figure 1: Lee-Pirie Plot of $\widehat{\Lambda}(t|x = 0.037)$ versus $\widehat{\Lambda}(t|x = -0.048)$.

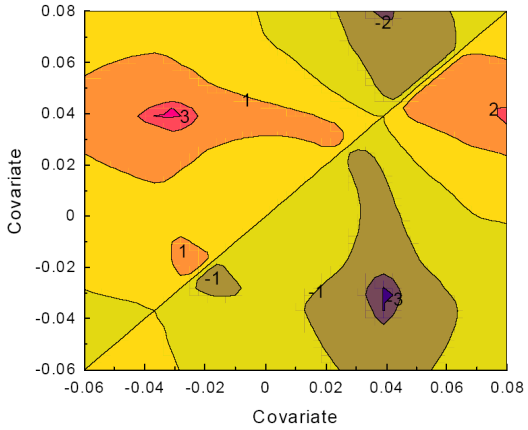


Figure 2: Contour plot of $T_{GS,Std}$

TABLE 6: Model Estimates: Malignant Melanoma Data

Model/ Parameter	**Coefficient**	**z-stat.**
Model 1		
Log Tumor Thickness, $\ln(x)$	0.823	5.49
Model 2		
$\ln(x).I\,[t\epsilon[0, 1062)]$	1.123	5.09
$\ln(x).I\,[t\epsilon[1062, \infty)]$	0.518	2.89
Model 3		
$\ln(x).I\,[x\epsilon(0, 1.9)]\,.I\,[t\epsilon[0, 1062)]$	0.097	0.15
$\ln(x).I\,[x\epsilon(0, 1.9)]\,.I\,[t\epsilon[1062, \infty)]$	1.177	2.39
$\ln(x).I\,[x\epsilon[1.9, \infty)]\,.I\,[t\epsilon[0, 1062)]$	1.184	5.90
$\ln(x).I\,[x\epsilon[1.9, \infty)]\,.I\,[t\epsilon[1062, \infty)]$	0.444	1.99

x