A Simple Test for the Absence of Covariate Dependence in Hazard Regression Models

Arnab Bhattacharjee

University of St Andrews

2004

A Simple Test for the Absence of Covariate Dependence in Hazard Regression Models

Arnab Bhattacharjee*
School of Economics and Finance
University of St. Andrews, UK.

Abstract

This paper extends commonly used tests for equality of hazard rates in a two-sample or k-sample setup to a situation where the covariate under study is continuous. In other words, we test the hypothesis $H_0 : \lambda(t|x) = c(t)$ for all x against the omnibus alternative (H_1 : not H_0) as well as more specific alternatives, when the covariate X is continuous. The tests developed are particularly useful for detecting trend in the underlying conditional hazard rates (i.e., when the alternative hypothesis is $H_1^* : \lambda(t|x_1) \geq \lambda(t|x_2)$ for all t whenever $x_1 > x_2$), or changepoint trend alternatives (such as H_1^{**} : there exists x^* such that $\lambda(t|x) \uparrow x$ whenever $x < x^*$ and $\lambda(t|x) \downarrow x$ whenever $x > x^*$). Asymptotic distribution of the test statistics are established and small sample properties of the tests are studied. An application to the effect of aggregate Q on corporate failure in the UK shows evidence of trend in the covariate effect, whereas a Cox regression model failed to detect evidence of any covariate effect. Finally, we discuss an important extension to testing for proportionality of hazards in the presence of individual level frailty with arbitrary distribution.

Keywords: Covariate dependence; Continuous covariate; Two-sample tests; Trend tests, Proportional hazards, Frailty, Linear transformation model..

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 the KPMG. The author thanks Elja Arjas, Sean Holly, Hashem Pesaran, Debasis Sengupta, and participants at the 58th European Meeting of the Econometric Society (Stockholm, 2003) for their valuable comments and suggestions.

1 Introduction

Understanding the nature of covariate dependence is one of the main objectives of regression analysis of lifetime/ duration data. The strength of covariate dependence is usually assessed by conducting tests of the hypothesis

$$H_0 : \lambda(t|x) = c(t) \text{ for all } x \tag{1}$$

against different kinds of alternatives, the choice of the alternative hypothesis depending on the expected nature of covariate dependence. Here we propose tests for the null hypothesis of absence of covariate dependence where the covariate is continuous and the alternative hypothesis is either omnibus i.e.

$$H_1 : \text{ not } H_0, \tag{2}$$

or trended, or changepoint trended.

The trended alternative means that the covariate has a positive or negative effect on the hazard function. In other words, the alternative hypothesis is

$$H_1^{(t)} : \lambda(t|x_1) \geq \lambda(t|x_2) \text{ for all } t \text{ whenever } x_1 > x_2 \text{ (or its dual)}, \tag{3}$$

the strict inequality holding for at least one covariate pair (x_1, x_2). The changepoint trended alternative implies that the covariate has a positive effect on the hazard rate over one region of the sample space and negative effect over another. A typical example is:

$$\begin{aligned} H_1^{(c)} \quad &: \quad \text{there exists } x^* \text{ such that } \lambda(t|x) \uparrow x \text{ whenever } x < x^* \quad (4) \\ \text{and } \lambda(t|x) \quad &\downarrow \quad x \text{ whenever } x > x^* \text{ (or its dual).} \end{aligned}$$

When the covariate is dichotomous or categorical, a test for absence of covariate effects against the omnibus alternative (2) is equivalent to testing that the hazard rates or survival functions in the two (or k) samples are the same. There are several censored-data rank tests appropriate for this situation; the Mantel-Haenszel or logrank test (Mantel, 1966; Peto and Peto, 1972) is one of the most popular in empirical applications. This test has optimal power if the two compared groups have proportional hazard functions (Peto and Peto, 1972). The Gehan or Breslow (Gehan, 1965; Breslow, 1970) and Prentice (1978) tests generalise the Wilcoxon and Kruskal-Wallis tests to right censored data. Tarone and Ware (1977) and Harrington and Fleming (1982) have proposed weighted log-rank tests. The theoretical properties of these tests and their performance in small samples has been discussed elsewhere (Fleming and Harrington, 1991; Andersen et al., 1992).

The omnibus alternative in the above tests is often too broad and does not convey sufficient information about the nature of covariate dependence. In empirical applications, it is often important to infer not only whether there is significant covariate dependence, but also about the direction of the covariate effect, *i.e.*, whether an increase in covariate value is expected to increase or decrease the lifetime/ duration, according to some notion of relative ageing. In the k-sample setup, several trend tests have been proposed; these procedures test for equality of hazards against the alternatives $H_1 : \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$ or $H_1 : S_1 \leq S_2 \leq \ldots \leq S_k$ (one or more of the inequalities being strict), where λ_j and S_j are the hazard and survival functions respectively in the j-th sample. Modified score tests that detect trend in hazard functions have been proposed by Tarone (1975) and Tarone and Ware (1977), while Liu et al. (1993) and Liu and Tsai (1999) have proposed ordered weighted logrank tests to detect similar trend in survival functions. Mau (1988) proposed trend tests for censored lifetime/ duration data by applying isotonic regression to scores from existing k-sample tests. These two-sample and k-sample tests are, however, of limited use in econometric and biomedical applications where the covariates are typically continuous (Horowitz and Neumann, 1992; Neumann, 1997). The usual method of extending these inference procedures to the case of continuous covariates involves stratification with respect to the covariate, followed by application of existing inference procedures for k samples. The outcomes of these inference procedures are highly sensitive to the choice of such intervals, and relevant procedures for optimally choosing these intervals are not available in general (Horowitz and Neumann, 1992).

There are some trend tests in the literature that are more appropriate for applications involving continuous covariates. Here the alternative hypothesis is (3). If one assumes an appropriate hazard regression model (like the Cox proportional hazards (PH) model or the accelerated failure time model), then one can use score tests for the significance of the regression coefficient (Cox, 1972; Prentice, 1978). Several tests assume a known covariate label function. Brown et al. (1974) derive a permutation test based on ranking of both the covariate values and the observed lifetimes, and O'Brien (1978) propose inverse normal and logit rank tests using the respective transformations of the ranked covariates. Jones and Crowley (1989, 1990) consider a general class of test statistics based on a known covariate label function; this test nests most of the other trend tests as well as robust versions of these tests.

All these test procedures for trend with respect to continuous covariates suffer from the limitation that they assume either validity of a specified regression model, or a known covariate label function. Hence, these tests fail to retain the attractive nonparametric flavour of the corresponding two-sample or k-sample tests, and are not useful in many situations. For example, these tests would not be able to detect presence of covariate dependence in changepoint trend situations (4). Jespersen (1986) has proposed inference procedures in the context of a single changepoint regression model; however, the assumptions of a specified regression model and a single changepoint are quite restrictive. Thus, appropriate tests for absence of covariate dependence for continuous covariates are not available in the literature, in applications where neither the form of the regression relationship nor an appropriate covariate label function are known, *a priori*. In many applications, insignificance of the estimated parameter in a Cox regression model is interpreted as a test for covariate dependence. Such an implication is inappropriate, since lack of significance can be due to other reasons, like violation of proportionality or model misspecification[1].

This article develops tests for the absence of covariate dependence that are useful in detecting trend (and changepoint trend) with respect to a continuous covariate, by a simple extension of the tests available in the two-sample setup. The usual two-sample tests are first conducted conditional on several pairs of distinct covariate values, and then the results of these tests are combined using the maxima/ minima or average of these individual test statistics to combine the results[2]. Section 2 describes construction of the test statistics and derives their asymptotic properties. Small sample properties of the tests are discussed in Section 3 through a simulation study, and an empirical application is presented in Section 4. In Section 5, an important extension to testing for proportionality in the presence of individual level frailty with a completely unrestricted distribution is developed. Finally, Section 6 collects concluding remarks.

2 Proposed tests for covariate dependence

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 $T = t$. We intend to test the hypothesis $H_0 : \lambda(t|x_1) = \lambda(t|x_2)$ for all x_1, x_2 against the alternative $H_1 : \lambda(t|x_1) \neq \lambda(t|x_2)$ for some $x_1 \neq x_2$. In particular, we are interested in test statistics that would be useful in detecting trend departures from H_0 of the form $H_1^{(t)}$ (3), and changepoint trend departures like $H_1^{(c)}$ (4).

[1] A large simulation study reported in Li et al. (1996) highlight the severe consequences of these issues in the context of the Cox PH model.

[2] A similar approach was adopted in Bhattacharjee (2006) to construct tests of proportionality of hazards with respect to continuous covariates.

As mentioned earlier, several two-sample tests of the equality of hazards hypothesis exist in the literature. Most of these tests are of the form:

$$T_{2s,std} = \frac{T_{2s}}{\sqrt{\widehat{Var}\left[T_{2s}\right]}}, \tag{5}$$

where

$$\begin{aligned}
T_{2s} &= \int_0^\tau L(t)d\widehat{\Lambda}_1(t) - \int_0^\tau L(t)d\widehat{\Lambda}_2(t), \\
\widehat{Var}\left[T_{2s}\right] &= \int_0^\tau L^2(t)\{Y_1(t)Y_2(t)\}^{-1}d\left(N_1 + N_2\right)(t), \\
L(t) &= K(t)Y_1(t)Y_2(t)\{Y_1(t) + Y_2(t)\}^{-1},
\end{aligned}$$

τ is a random stopping time (in particular, τ may be taken as the time at the final observation in the combined sample), $K(t)$ is a predictable process depending on $Y_1 + Y_2$, but not individually on Y_1 or Y_2, $\widehat{\Lambda}_j(t)$ is the Nelson-Aalen estimator of the cumulative hazard function in the j-th sample ($j = 1, 2$), $Y_j(t)$ (for $j = 1, 2$) denote the number of individuals on test in sample j at time t, and N_1, N_2 are counting processes counting the number of failures in either sample.

In particular, for the logrank test,

$$K(t) = I\left[Y_1(t) + Y_2(t) > 0\right], \tag{6}$$

and for the Gehan-Breslow modification of the Wilcoxon test,

$$K(t) = I\left[Y_1(t) + Y_2(t) \ > 0\right].\{Y_1(t) + Y_2(t)\}. \tag{7}$$

In the two sample setup, these standardised test statistics have zero mean under the null hypothesis of equal hazards and positive/ (negative) mean if the hazards are trended. Further, they are asymptotically normally distributed under the null hypothesis.

Based on these test statistics, we propose a simple construction of our tests as follows. We first select a number of pairs of distinct points on the covariate space, and construct the usual two-sample test statistics ($T_{2s,std}$) for each pair, based on counting processes conditional on these two distinct covariate values. We then construct our test statistics, by taking supremum, infimum or average of these basic test statistics over the fixed number of pairs.

Thus, 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_{2s}^{*}, T_{2s}^{**} and $\overline{T}_{2s}$ based on the r statistics $T_{2s,std}(x_{l1}, x_{l2})$, $l = 1, \ldots, r$ (each testing equality of hazard rates for the pair of counting processes $N(t, x_{l1})$ and $N(t, x_{l2})$), where

$$\begin{aligned}
T_{2s,std}(x_{l1}, x_{l2}) &= \frac{T_{2s}(x_{l1}, x_{l2})}{\sqrt{\widehat{Var}\left[T_{2s}(x_{l1}, x_{l2})\right]}}, \\
T_{2s}(x_{l1}, x_{l2}) &= \int_0^{\tau} L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l1}) - \int_0^{\tau} L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l2}), \\
\widehat{Var}\left[T_{2s}(x_{l1}, x_{l2})\right] &= \int_0^{\tau} L^2(x_{l1}, x_{l2})(t) \{Y(t, x_{l1}) Y(t, x_{l2})\}^{-1}. \\
&\qquad\qquad d\left(N(t, x_{l1}) + N(t, x_{l2}\right),
\end{aligned}$$

where $L(x_{l1}, x_{l2})(t)$ is a random (predictable) process indexed on the pair of covariate values x_{l1} and x_{l2}, and $\widehat{\Lambda}(t, x_{l1})$ and $\widehat{\Lambda}(t, x_{l2})$ are the Nelson-Aalen estimators of the cumulative hazard functions for the respective counting processes.

Then, our test statistics are:

$$T_{2s}^{(\max)} = \max\left\{T_{2s,std}(x_{11}, x_{12}), T_{2s,std}(x_{21}, x_{22}), \ldots, T_{2s,std}(x_{r1}, x_{r2})\right\}, \quad (8)$$

$$T_{2s}^{(\min)} = \min\left\{T_{2s,std}(x_{11}, x_{12}), T_{2s,std}(x_{21}, x_{22}), \ldots, T_{2s,std}(x_{r1}, x_{r2})\right\}. \quad (9)$$

$$\text{and } \overline{T}_{2s} = \frac{1}{r} \sum_{l=1}^{r} T_{2s,std}(x_{l1}, x_{l2}). \quad (10)$$

We now derive the asymptotic distributions of these test statistics.

Consider a counting processes $\{N(t, x) : t \in [0, \tau], x \in \mathcal{X}\}$, indexed on a continuous covariate x, with intensity processes $Y(t, x).\lambda(t|x)$ such that $\lambda(t|x) = \lambda(t)$ for all t and x (under the null hypothesis of equal hazards). Let, as before, L be a process indexed on a pair of distinct values of the continuous covariate x (*i.e.*, indexed on $(x_1, x_2), x_1 \neq x_2, x_1, x_2 \in \mathcal{X}$). Now, let $\{x_{11}, x_{21}, \ldots, x_{r1}, x_{12}, x_{22}, \ldots, x_{r2}\}$ be $2r$ (r is a fixed positive integer, $r > 1$) distinct points on the covariate space $\mathcal{X}$, such that $x_{l2} > x_{l1}, l = 1, \ldots, r$.

Assumption 1 For each l $(= 1, 2, \ldots, r)$, let $L(x_{l1}, x_{l2})(t)$ be a predictable process of the form

$$L(x_{l1}, x_{l2})(t) = K(x_{l1}, x_{l2})(t).Y(t, x_{l1}).Y(t, x_{l2}).\left[Y(t, x_{l1}) + Y(t, x_{l2})\right]^{-1},$$

where $K(x_{l1}, x_{l2})(t)$ depends on $[Y(t, x_{l1}) + Y(t, x_{l2})]$ but not individually on $Y(t, x_{l1})$ or $Y(t, x_{l2})$.

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 $\tau\left(x_{l1}, x_{l2}\right), l = 1, \ldots, r$ for each of the r basic test statistics.

Assumption 3 The sample paths of $L(x_{l1}, x_{l2})$ 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 $l = 1, \ldots, r$, $L(x_{l1}, x_{l2})$ is zero whenever $Y\left(t, x_{l1}\right)$ or $Y\left(t, x_{l2}\right)$ are.

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

$$\begin{array}{lll} sup_{t\in[0,\tau]} \left|Y(t,x)/a^{(n)} - y(t,x)\right| \xrightarrow{P} 0 & \text{as } n \longrightarrow \infty, & \forall x \in \mathcal{X} \\ sup_{t\in[0,\tau]} \left|L(x_{l1}, x_{l2})(t) - l_i(x_{l1}, x_{l2})(t)\right| \xrightarrow{P} 0 & \text{as } n \longrightarrow \infty, & l = 1, \ldots, r \end{array}$$

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

Assumptions 1-4 constitute a simple extension, to the continuous covariate framework, of the standard set of assumptions for the counting process formulation of lifetime data (see, for example, Andersen et al., 1992). The condition on probability limit of $Y(t, x)$ in Assumption 4 can be replaced by a set of weaker conditions (Sengupta et al., 1998).

Let the test statistics $T_{2s}^{(\max)}$, $T_{2s}^{(\min)}$ and $\overline{T}_{2s}$ be as defined earlier.

Theorem 1 *Let Assumptions 1 to 4 hold. Then, under H_0, as $n \to \infty$,*
(a) $P\left[T_{2s}^{(\max)} \leq z\right] \longrightarrow [\Phi(z)]^r$,
(b) $P\left[T_{2s}^{(\min)} \geq -z\right] \longrightarrow [\Phi(z)]^r$,
and
(c) $r^{1/2}\overline{T}_{2s} \xrightarrow{D} N(0, 1)$,
where $\Phi(z)$ is the distribution function of a standard normal variate.
(Proof in the Appendix).

***Corollary* 1**

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

(Proof in the 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 statistic*

$$\overline{T}_{2s,\underline{w}}=\frac{\sum_{l=1}^{r}w_l.T_{2s,std}(x_{l1},x_{l2})}{\sum_{l=1}^{r}w_l}$$

as a weighted average of the r *individual standardised test statistics. Let Assumptions 1 to 4 hold. Then, under* H_0, *as* $n\to\infty$,

$$\frac{\sum_{l=1}^{r}w_l}{\left[\sum_{l=1}^{r}w_l^2\right]^{1/2}}.\overline{T}_{2s,\underline{w}}\xrightarrow{D}N(0,1),$$

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

Proof follows from Theorem 1.

The above results establish the asymptotic properties of our proposed tests. Some other features of the testing procedure merit further discussion. First, the number of covariate pairs, r, on which the statistics ($T_{2s}^{(\max)}$, $T_{2s}^{(\min)}$ and $\overline{T}_{2s}$) are based is fixed *a priori*. This is crucial, since the process $T_{2s,std}(x_1,x_2)$ on the space

$$\{(x_1,x_2):\ x_2>x_1,x_1,x_2\in\mathcal{X}\},$$

is pointwise standard normal and independent, but do not have a well-defined limiting process. Therefore, if r is allowed to grow, the supremum (infimum) will diverge to $+\infty$ $(-\infty)$.

Second, Corollary 1 provides a simple way to compute p-values for the test statistics when r is reasonably large. Note that r is fixed and finite; however, if it assumes a large enough value (say, 20 or higher), the approximation can be used.

Third, Corollary 2 shows that one can weight the underlying test statistics by some measure of the distance between x_{l1} and x_{l2}. For example, one can give higher weight to a covariate pair where the covariates are further apart. In practice, this is expected to improve the empirical performance of the tests. We have not used such weighting in the empirical exercise in Sections 3 and 4.

Fourth, since the covariate under consideration is continuous, it may not be feasible to construct the basic tests $T_{2s,std}$ based exactly on two distinct fixed points on the covariate space. We have considered "small" intervals around these chosen points, such that the hazard function within these intervals is approximately constant (across covariate values). The average test statistics constructed in this way, however, sometimes fail to maintain their nominal sizes under the null hypothesis because of correlation between statistics based on overlapping intervals (Bhattacharjee, 2006). This issue can be resolved by using a jacknife estimator of the variance of this average estimator.

Fifth, the choice of the r pairs of covariate values may be important in applications. The issues regarding this choice are similar to those relating to stratification in goodness-of-fit tests. Quantiles of the cross-sectional distribution of the covariate can be used to select these covariate pairs and to construct the "small" intervals around the covariate values – this, in a simple way, ensures that variations in the density of design points are adjusted for (none of the intervals are too sparse) and that the intervals corresponding to each pair of covariate values do not overlap.

Finally, the tests can be applied in situations where we have more than one covariate. If the interest is in testing for covariate dependence with respect to a single covariate, and covariate dependence with respect to the other covariates are known to follow some regression model (such as the Cox regression model, or a nonproportional hazards model with age-varying covariate effects), the usual partial likelihood estimates of baseline cumulative hazard from a regression model including the other covariates can be used to construct the test statistics. On the other hand, if one is interested is in testing for covariate dependence with respect to a collection of different covariates, pairs of covariate vectors can be used to construct a test statistic for the joint test of absence of covariate dependence over all these covariates.

3 Simulation study

The asymptotic distributions of the proposed test statistics were derived in Section 2. Here, we report results of a simulation study exploring the finite sample performance of the proposed tests, for different specifications of the baseline hazard function and covariate dependence. In particular, we consider models of the form

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

where $\lambda_0(t)$ and $\beta(t,x)$ are chosen to reflect a variety of baseline hazards and patterns of covariate dependence. In all cases, the null hypothesis of absence of covariate dependence holds if and only if $\beta(t,x)=0$. If, for fixed x, $\beta(t,x)$ increases/ decreases in x, we have trended alternatives of the type $H_1^{(t)}$ (3). If, on the other hand, $\beta(t,x)$ increases in x over some range of the covariate space, and decreases over another, we have changepoint trend departures of the type $H_1^{(c)}$ (4). In addition to the global alternative H_1 (2), our tests are consistent against both these kinds of alternatives to the null hypothesis.

The Monte Carlo simulations are based on independent right-censored data from the following 6 data generating processes, simulated using the Gauss 386 random number generator.

Model	$\lambda_0(t)$	$\beta(t,x)$	**Median cens.dur.**	**% cens.**	**Expected significance**
DGP_{11}	2	0	0.32	7.7	None
DGP_{12}	2	x	0.30	9.2	$T_{2s}^{(\max)}, \overline{T}_{2s}$
DGP_{13}	2	$\lvert x\rvert$	0.20	6.6	$T_{2s}^{(\max)}, T_{2s}^{(\min)}$
DGP_{21}	$20t$	0	0.17	9.4	None
DGP_{22}	$20t$	x	0.16	10.4	$T_{2s}^{(\max)}, \overline{T}_{2s}$
DGP_{23}	$20t$	$\lvert x\rvert$	0.14	7.4	$T_{2s}^{(\max)}, T_{2s}^{(\min)}$

The covariate X is independently distributed as $Uniform(-1,1)$. The censoring variable C is independent of the lifetime and distributed as $Exp(6)$ for DGP_{11}, DGP_{12} and DGP_{13} and $Exp(2)$ for DGP_{21}, DGP_{22} and DGP_{23}. The data generating processes DGP_{11} and DGP_{21} belong to the null hypothesis of absence of covariate dependence, DGP_{12} and DGP_{22} are trended, and DGP_{13} and DGP_{23} are changepoint trended alternatives. We use the logrank test to construct the basic test statistics, and 100 distinct pairs of covariate values are used to construct the maxima, minima and average test statistics. Table 1 presents simulation results for 10,000 simulations from the above data generating processes with samples of size 100 and 200.

The nominal sizes are approximately maintained in the random samples, and the tests have good power, with the exception of DGP_{13} and DGP_{23}.

This is not surprising, since these two data generation processes are change-point trended, so that when a pair of points are drawn at random from the covariate space, only a quarter of them reflect the increasing nature of covariate dependence, and another quarter reflect the decreasing trend. The results also reflect the strength of the supremum/ infimum test statistics in their ability to detect non-monotonic departures (DGP_{13} and DGP_{23}) from the null hypothesis of absence of covariate dependence.

Though the tests proposed here are not directly comparable with other trend tests, we have examined how these two categories of tests compare in terms of power. For the purpose of applying the trend tests in the current context, we had to stratify the samples with respect to the value of the covariate. This comparison shows our tests to perform favourably in comparison with the Tarone (1975) and Liu and Tsai tests (the results are not reported here, but are available from the author).

TABLE 1:
REJECTION RATES (%) AT 5% AND 1% ASYMPTOTIC CONF. LEVELS

Model	**Test statistic**	**Sample size, Confidence level**			
		100, 5%	**200, 5%**	**100, 1%**	**200, 1%**
DGP_{11}	$T_{2s}^{(\max)}$	3.76	5.59	0.67	1.08
	$T_{2s}^{(\min)}$	7.23	5.66	1.18	0.88
	$\overline{T}_{2s}$	5.46	5.35	1.19	0.99
DGP_{12}	$T_{2s}^{(\max)}$	95.46	100.00	82.98	100.00
	$T_{2s}^{(\min)}$	2.43	1.91	0.41	0.80
	$\overline{T}_{2s}$	96.82	100.00	87.95	100.00
DGP_{13}	$T_{2s}^{(\max)}$	26.06	63.30	5.67	29.41
	$T_{2s}^{(\min)}$	38.19	70.62	12.29	40.40
	$\overline{T}_{2s}$	5.67	4.83	1.23	0.94
DGP_{21}	$T_{2s}^{(\max)}$	3.90	5.51	0.53	1.61
	$T_{2s}^{(\min)}$	7.24	6.12	1.45	0.79
	$\overline{T}_{2s}$	5.62	5.68	0.92	1.35
DGP_{22}	$T_{2s}^{(\max)}$	97.18	100.00	86.03	99.87
	$T_{2s}^{(\min)}$	2.69	1.85	0.41	0.82
	$\overline{T}_{2s}$	97.71	100.00	92.02	100.00
DGP_{23}	$T_{2s}^{(\max)}$	21.26	54.50	4.39	23.04
	$T_{2s}^{(\min)}$	36.44	69.35	11.64	37.73
	$\overline{T}_{2s}$	7.18	6.96	1.56	2.06

4 An application

In this section, we illustrate the use of the tests proposed in this paper by way of an application, in which we study the effect of aggregate Q on the hazard rate of corporate failure in the UK. The data are on firm exits through bankruptcy over the period 1980 to 1998 and pertain to 2789 listed manufacturing companies, covering 24,034 company years and includes 95 bankruptcies. The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1980, and contain delayed entries. Here the focus of our analysis is on the impact of aggregate Q on corporate failure (more detailed analysis of these data are reported elsewhere (Bhattacharjee et al., 2002)). Following usual practice, we consider the reciprocal of Q as the continuous covariate in our regression model.

A priori, we expect periods with higher values of the covariate to correspond to lower incidence of bankruptcy. However, estimates of the Cox proportional hazards model on these data reports a hazard ratio of 0.92, with p-value 0.156 per cent. One would then be tempted to believe that covariate dependence is absent. However, such lack of evidence of covariate effect could also arise from model misspecification. This possibility suggests that we could take a completely nonparametric approach that does not assume any *a priori* knowledge of the nature of covariate dependence.

Descriptive graphical tests based on counting processes conditional on several pairs of covariate values indicate significant trend in the hazards. Hence, we applied our tests of absence of covariate dependence to these data (Table 2). Each of the tests were based on 20 pairs of distinct covariate values. The results of the tests support our *a priori* belief; the null hypothesis is rejected at 5 per cent level of significance in favour of the alternative of negative trend, $H_1^* : \lambda(t|x_1) \leq \lambda(t|x_2)$ for all $x_1 > x_2$ (with strict inequality holding for some $x_1 > x_2$). This implies that, contrary to what the estimates of the Cox regression model indicates, higher aggregate Q significantly depresses the hazard of business exit due to bankruptcy.

TABLE 2:
TESTS FOR ABSENCE OF COVARIATE DEPENDENCE:
UK CORPORATE BANKRUPTCY DATA

Test	**Test Statistic**	**P-Value** (%)
$T_{2s}^{(\max)}$ - Logrank	0.592	1.0000
$T_{2s}^{(\min)}$ - Logrank	-3.732	0.0188
$T_{2s}^{(\max)}$ - Gehan-Breslow	0.500	1.0000
$T_{2s}^{(\min)}$ - Gehan-Breslow	-3.046	0.0370

TABLE 3:
MODEL ESTIMATES: CORPORATE BANKRUPTCY DATA

Model/ Parameter	**Hazard Ratio**	**z-stat.**
$Q.I\,[t \in [0,9)]$	0.947	-0.54
$Q.I\,[t \in [9,17)]$	0.773	-1.30
$Q.I\,[t \in [17,26)]$	0.147	-2.06
$Q.I\,[t \in [26,\infty)]$	0.193	-2.96

Further, these supremum/ infimum test statistics provide additional information on the covariate pairs for which the basic test statistics assume extreme values, which may be useful in further investigating the nature of departures from proportionality. For the bankruptcy data, for example, the significant test-statistics T_{2s}^{**} are attained for the covariate pairs $\{-0.058, 0.116\}$ (7th and 63rd percentile) for the logrank test statistic and $\{-0.017, 0.098\}$ (10th and 50th percentile) for the Gehan-Breslow test. This further indicates a strong evidence of trend.

To explore whether this apparent trend in hazards was masked in the original Cox regression by lack of proportionality, we present in Table 3 a time varying coefficient model for the same data estimated using the histogram sieve estimators proposed by Murphy and Sen (1991). Here, we allow the regression coefficient for the covariate Q to vary over the life of the firm, having different effects over the time ranges '0-8 years', '9-16 years', '17-25 years' and 'above 25 years' of post-listing age. The results confirm the presence of trend, particularly at higher ages.

The above application demonstrates the use of the proposed test statistics. These tests are useful not only for detecting presence of covariate dependence for continuous covariates, but also for detecting trend and changepoint trend in the effect of a covariate. Further, the tests can provide clues about the approximate location of such changepoints, when present.

5 Testing for proportionality with individual level frailty

It is well-known that the proportionality assumption underlying the Cox proportional hazards model does not hold in many applications. At the same time, credible inference under the model depends crucially on the validity of the proportionality assumption. Further, the effect of a covariate is often monotone, in the sense that the lifetime (or duration) conditional on a higher value of the covariate ages faster or slower than that conditional on a lower value (Bhattacharjee, 2004). Ordered departures of this kind are common in applications, and the models provide useful and intuitively appealing descriptions of covariate dependence in non-proportional situations.

Testing for proportionality against such ordered departures is, therefore, an important area of research. For a binary covariate (2 sample setup), tests for proportionality against a monotone hazard ratio alternative have been proposed by Gill and Schumacher (1987) and Deshpande and Sengupta (1995), while Sengupta et al. (1998) proposed a test against the weaker alternative of monotone ratio of cumulative hazards.

In a recent contribution, Bhattacharjee (2006) has extended the notion of monotone hazard ratio in two samples to the situation when the covariate is continuous, and proposed tests for proportionality against ordered alternatives. Specifically, the alternative hypothesis here states that, lifetime conditional on a higher value of the covariate is convex (or concave) ordered with respect to that conditional on a lower covariate value. Bhattacharjee (2004) show that, in the absence of unobserved heterogeneity, monotone covariate dependence of this type can be conveniently studied using age-varying covariate effects and propose biased bootstrap methods to estimate these effects. The above tests are valid when there is no unobserved heterogeneity, or when random effects heterogeneity is in the nature of shared frailties. However, being based on counting process martingales, they are not useful when there is individual level frailty. Our contribution here is to develop tests for proportional hazards in the presence of individual level inobserved heterogeneity with completely unrestricted and unknown frailty distribution.

5.1 The mixed proportional hazards (MPH) model

Monotone covariate effects in the presence of individual level frailty with arbitrary distribution have not been discussed in the literature. In fact, apart from a few important contributions which we discuss below, most of the research has assumed a finite dimensional distribution either for the lifetime or for the frailty. To start with, we consider the following MPH model

$$\ln \Lambda_0(T) = \beta^T.Z + U + \varepsilon, \tag{11}$$

where $\ln \Lambda_0(t)$ is an increasing function of arbitary shape (the log cumulative hazard function), log-frailty U has an arbitrary distribution that is independent of the covariates Z, and ε has an extreme value distribution. Since U has an arbitrary distribution, so does $U + \varepsilon$, and hence this is a special case of the monotonic transformation model considered by Horowitz (1996).

Because the MPH model (Equation 11) still continues to hold if a constant is added to both sides, a location normalisation is required for identification. This is achieved by setting

$$\Lambda_0(t_0) \equiv 1 \tag{12}$$

for some fixed and finite $t_0 > 0$[3]. Our interest here is in making inferences about proportionality using estimates of the baseline cumulative hazard function, $\Lambda_0(t) = \int_0^t \lambda_0(s).ds$. Because of the above normalisation, testing for proportionality becomes equivalent to testing the equality of hazard functions conditional on different values of a chosen covariate (say, X).

The MPH model has an important distinction from the standard transformation model, in that a scale normalisation is not necessary here. In other words, β is exactly identified by the fact that the scale of ε is fixed. However, the scale parameter is very difficult to estimate, which has implications for the rate of convergence of model estimates. The fastest achievable rate of convergence for the cumulative baseline hazard function estimates is only $N^{-2/5}$, which is smaller than the usual convergence rate of $N^{-1/2}$.

Under an arbitrary heterogeneity distribution, Melino and Sueyoshi (1990) provide a constructive proof of identifiability in the MPH model for the continuous regressor case. The identification, however, relies heavily on the observed duration density at $t = 0$, which is in practice very difficult to assess using real-life data. Kortram et al. (1995) provide a constructive proof for the two-sample (binary regressor) case (i.e., where $\beta^T.z$ can take only two distinct values), and Lenstra and Van Rooij (1998) exploit this to construct a consistent nonparametric estimator. This idea is potentially useful; however, the asymptotic distribution of their estimator of the baseline cumulative hazard function is unknown.

We focus instead on a kernel-based estimator of the baseline cumulative hazard function, proposed by Horowitz (1999), in the presence of scalar unobserved heterogeneity with completely unrestricted distribution. The proposed estimators are based on estimates for the scale parameter combined with those for the linear transformation model proposed in Horowitz (1996). Our interest lies in estimates of the cumulative hazard function and the hazard function, the rate of convergence for which can be made arbitrarily close to $N^{-2/5}$ by suitable choice of bandwidths.

Two features of the estimation methodology are relevant to our work. First, while we are interested in estimates of the cumulative baseline hazard $(\widehat{\Lambda}_0(t|x_1), \widehat{\Lambda}_0(t|x_2), \ldots)$ for different values of the chosen covariate, these can

[3]Note that, an important implication of this location normalisation is that absense of covariate dependence cannot be tested in the presence of individual-level frailty.

only be estimated in the presence of at least one other covariate, Z. In practice, this is not a severe restriction because in most applications there would be at least two different covariates. Second, and more crucially, the proportional hazards assumption must hold for the other covariates, Z, included in the model. This restriction is particularly important here because, in any case, some covariates have potentially nonproportional effects on the hazard function. However, nonproportionality can be accommodated in a simple way by allowing the covariate effects vary with age (Bhattacharjee, 2004). In other words, a simple histogram sieve (Murphy and Sen, 1991) can mitigate the nonproportional effects of these covariates. In fact, to counter the potential adverse effect of nonproportionality, it may be more useful to interact most (if not all) covariates with a histogram sieve until proportionality has been tested for.

5.2 Tests for proportionality

Based on the estimates of cumulative baseline hazard and baseline hazard function conditional on different values of the selected covariate, we now discuss our tests for proportionality. If there were no frailty, standard tests from the survival analysis literature could be used for this purpose. The statistical properties for most of these tests are based on the asymptotic joint distribution of processes $\int_0^t K_n(s;x).d\widehat{\Lambda}_{0,n}(s;x)$, where t may either be a fixed time point or a stopping point with respect to an appropriate filtration, K_n is a random process and $\widehat{\Lambda}_{0,n}$ is an estimator of the baseline cumulative hazard function; both K_n and $\widehat{\Lambda}_{0,n}$ are measured at a given fixed value of the covariate x.

For example, like in Section 2, proportionality can be tested against the omnibus alternative (violation of PH assumption) by conducting several log rank tests for different pairs of covariate values, and then combining the standardised tests by taking their average (Equation 10). Note that, under the current setup, testing for trend is no longer relevant. Due of the location normalisation (Equation 12), testing for proportionality is equivalent to testing for absence of covariate dependence (equality of hazards). At the same time, if equality of hazards does not hold, the location mormalisation will force hazard rates to intersect at least once within the duration $(0, t_0)$.

Testing for proportionality against ordered alternatives of monotone covariate dependence, such as

$$\begin{aligned}
IHRCC &: \text{ whenever } x_1 > x_2, \lambda(t|x_1)/\lambda(t|x_2) \uparrow t (\equiv (T|X = x_1) \prec_c (T|X = x_2) \\
DHRCC &: \text{ whenever } x_1 > x_2, \lambda(t|x_2)/\lambda(t|x_1) \uparrow t (\equiv (T|X = x_2) \prec_c (T|X = x_1)
\end{aligned}$$

considered in Bhattacharjee (2006), can be conducted by extending tests such as the one proposed in Gill and Schumacher (1987) to the continuous covariate setup[4]. Following Bhattacharjee (2006), we describe three tests based on maxima, minima or average of the usual two sample tests. For the alternative of 'increasing hazard ratio' (convexity) in two samples (having cumulative hazard functions $\Lambda_1(t)$ and $\Lambda_2(t)$), the test statistic proposed by Gill and Schumacher (1987) is

$$T_{GS,std} = \frac{T_{GS}}{\sqrt{\widehat{Var}\left[T_{GS}\right]}}, \tag{13}$$

where

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

t is a random stopping time (for example, t may be taken as the time at the final observation in the combined sample), $L_1(t)$ and $L_2(t)$ are two predictable processes, and for the j-th sample ($j = 1,2$), $\widehat{\Lambda}_j(t)$ denotes the Nelson-Aalen estimator of the cumulative hazard function, $Y_j(t)$ is the number of individuals on test at time t, and $N_j(t)$ are counting processes counting the number of failures in sample j at time t.

Gill and Schumacher (1987) have shown 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 in t on $[0,\infty)$ and L_1 and L_2 are so chosen that $L_1(t)/L_2(t)$ is monotonically decreasing (increasing), and that its standard error would decrease to zero as sample size increases to ∞ under both the null and alternative hypotheses. Hence, while the standardized test statistic $T_{GS,std}$ would be asymptotically standard normal under the null hypothesis, it's mean would increase (decrease) to ∞ ($-\infty$) under the alternative hypothesis. In many applications, L_1 and L_2 are chosen corresponding to the Gehan-Wilkoxon and log rank tests, where $L_1 = Y_1Y_2$ and $L_2 = Y_1Y_2(Y_1+Y_2)^{-1}$, so that $L_1(t)/L_2(t)$ is monotonically decreasing in t.

[4]IHRCC (DHRCC) are acronyms for "Increasing (Decreasing) Hazard Ratio for Continuous Covariates".

For testing $H_0 : PH$ vs. $H_1 : IHRCC$, Bhattacharjee (2006) proposed 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

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

$$T_{lij} = \int_0^t L_i(x_{l1}, x_{l2})(t)\, d\widehat{\Lambda}_0(t, x_{lj}),$$

and

$$V_{lij} = \int_0^t L_i(x_{l1}, x_{l2})(t) L_j(x_{l1}, x_{l2})(t) \frac{d\left[N(t, x_{l1}) + N(t, x_{l2})\right]}{Y(t, x_{l1})Y(t, x_{l2})},\ i, j = 1, 2$$

Then, the test statistics are:

$$\begin{aligned}
T_{GS}^{(\max)} &= \max\left\{T_{GS,std}(x_{11}, x_{12}), T_{GS,std}(x_{21}, x_{22}), \ldots, T_{GS,std}(x_{r1}, x_{r2})\right\}, \quad (14) \\
T_{GS}^{(\min)} &= \min\left\{T_{GS,std}(x_{11}, x_{12}), T_{GS,std}(x_{21}, x_{22}), \ldots, T_{GS,std}(x_{r1}, x_{r2})\right\}, \quad (15)
\end{aligned}$$

and

$$\overline{T}_{GS} = \frac{1}{r}\sum\nolimits_{l=1}^{r} T_{GS,std}(x_{l1}, x_{l2}). \quad (16)$$

For the choice of L_1 and L_2 mentioned above, these statistics would be close to zero under the null hypothesis. Under the alternative hypothesis $IHRCC$, $\overline{T}_{GS}$ and $T_{GS}^{(\max)}$ will increase to ∞ as sample size increases, while under $DHRCC$, $\overline{T}_{GS}$ and $T_{GS}^{(\min)}$ will decrease to $-\infty$.

Under the counting process formulation of survival analysis (Andersen et al., 1992), if t is a stopping time and K_n is a locally bounded predictable process, then $\int_0^t K_n(s;x).d\left[\Lambda_0(s;x) - \widehat{\Lambda}_{0,n}(s;x)\right]$ is a local square integrable martingale. Therefore, asymptotic distributions of these kinds of statistics can be obtained by a simple application of martingale central limit theoirem. In the presence of frailty, however, this line of reasoning does not hold, since

$$M_i(t) = N_i(t) - \int_0^t Y_i(s). \exp\left[\left(\beta^T.x\right)\right].d\Lambda_0(s)$$

(where $N_i(t)$ is the counting process for exit and $Y_i(s)$ is the at-risk indicator) is no longer a martingale with respect to the joint filtration generated by all the exits, censoring and covariate histories up to duration s. There are two approaches we can take.

First, using the theory of empirical processes, several researchers have recently shown that the asymptotic properties of statistics like the above can be obtained under much weaker assumptions (see, for example, Spiekerman and Lin, 1998; Lin et al., 2000; and Lin and Ying, 2001). In particular, weak convergence to a Gaussian law can be established if t is a fixed duration, the process K_n is of bounded variation ($\int_0^t |dK_n(s;x)| = O_P(1)$) and the process $\left[\Lambda_0(s;x) - \widehat{\Lambda}_{0,n}(s;x)\right]$ converges weakly to a zero mean process with continuous paths. See, for example, Lemma 1 in Spiekerman and Lin (1998) in combination with Appendix A.2 in Lin et al. (2000). The weak convergence results can be extended to cases where t is the last observed duration, and where durations are discrete (Lin et al., 2000).

We can therefore use the estimator of the cumulative baseline hazard function under frailty proposed by Horowitz (1999) instead of the Nelson-Aalen estimator of the baseline cumulative hazard. Further, we do not require the usual assumption of predictability and local boundedness of K_n used in the counting process approach. Instead, we require that this stochastic process is of bounded variation and has a probability limit which also has bounded variation. Most of the predictable weight functions considered, like the Gehan, log-rank and the Pepe-Fleming weight functions, are of bounded variation. Further, other weight functions like the one considered by Sengupta et al. (1998) ($Y_1(t).Y_2(t).\exp(-t/TTT)$, where TTT denotes the total time on test), which are not predictable either because they are based on all the data or because they are not right-continuous at all points, also have bounded variation and can therefore be used; see Gu et al. (1999) for further discussion. We do not go into further technical details here because the arguments follow in a fairly straightforward way from Appendix A.2 of Lin et al. (2000).

The second approach is to assume that the density function is continuous. Under this assumption, the kernel-based estimator of the baseline hazard function (Horowitz, 1999) also converges weakly to a Gaussian process. If the weight function K_n is cadlag, we can obtain weak convergence results for the above integral by a straighforward application of Theorem 3.1 in Sengupta et al. (1998). Note that, all the weight functions considered above are cadlag.

5.3 Choice of weight functions

We close the section with a brief discussion of the choice of weight functions for the omnibus and ordered tests for proportionality with individual-level frailty. It is well known that, if hazards are proportional in the two samples, the logrank test for equality of hazards is the most optimal (see, for example, Andersen et al., 1992). However, this proportionality assumption does not hold under our alternative hypothesis, so there is no clear winner in terms of asymptotic relative efficiency (ARE).

Gill and Schumacher (1987) discuss the optimal choice of weight functions for their two sample tests of proportionality against convexity. In particular, they show that the logrank weight function $L_1 = Y_1 Y_2 (Y_1 + Y_2)^{-1}$ in combination with Prentice's Wilcoxon generalisation (Prentice, 1978) $L_2 = Y_1 Y_2 (Y_1 + Y_2)^{-1} . \widehat{S}$, where $\widehat{S}$ is the Kaplan-Meier estimate of the survival function in the combined sample, is an optimal choice in terms of ARE under a couple of conditions. First, the hazard ratio under proportionality is unity, which holds in our case (though not necessarily in theirs). Second, the proportion of at risk individuals in each sample is proportional. This is true if censoring is random (which we assume) and if frailty is independent of all other regressors, which is also true in our case. Therefore, the above combination of weight functions is optimal in our case.

It is perhaps possible that the lifetime t_0 can be chosen to achieve further optimality of the testing procedures suggested here. This is, however, a different problem and outside the scope of the current paper.

6 Discussion

In summary, the tests described in this paper add important tools to the armoury of a lifetime/ duration data analyst. Therefore, our work extends an important class of two sample tests for equality of hazards to a continuous covariate framework. This also shows that usual statistical treatment of lifetime/ duration data using counting processes are useful in analysing such continuous covariate situations.

The second important contribution of the paper is in extending tests for proportionality with respect to a continuous covariate against ordered alternatives to the case when there is individual level frailty with completely unrestricted distribution. Here, counting process arguments do not hold, but we can use empirical process theory to extend standard two sample tests to this setup. In conjunction with Bhattacharjee (2006), this paper therefore extends many of these two sample testing procedures to the continuous covariate setup, and thereby makes these tests more readily usable in real life econometric applications.

Though the discussion in this paper has largely focussed on a single continuous covariate, the tests can be readily used in applications with multiple continuous covariates. Here, one can take either of two approaches. The first one is to test the absence of covariate dependence for one covariate, while modeling covariate dependence for other covariates more explicitly using either the Cox regression model or a model with age-varying covariate effects. Then, one can use the estimates of baseline cumulative hazard functions derived from the regression model (including the other covariates, but not the one under study) to construct the appropriate test statistics. Alternatively, one can jointly test for covariate dependence for two or more covariates.

APPENDIX

Proof of Theorem 1: It follows from standard counting process arguments (see, for example, Andersen *et. al*, 1992) that, under H_0, for $l = 1, \ldots, r$,

$$T_{2s}\left(x_{l1}, x_{l2}\right) = \sum_{j=1}^{2} \int_0^{\tau} K(x_{l1}, x_{l2})(t).\left[\delta_{1j} - Y(t, x_{l1})\left\{Y(t, x_{l1}) + Y(t, x_{l2})\right\}^{-1}\right] .dM(t, x_{lj}),$$

where δ is the Kronecker delta function, and $M(t, x_{lj}), l = 1, \ldots, r, j = 1, 2$ are the innovation martingales corresponding to the counting processes $N(t, x_{lj}), l = 1, \ldots, r, j = 1, 2$.

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

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

and we have as $n \longrightarrow \infty$,

$$T_{2s,std}\left(x_{l1}, x_{l2}\right) = \frac{T_{2s}\left(x_{l1}, x_{l2}\right)}{\sqrt{\widehat{Var}\left[T_{2s}\left(x_{l1}, x_{l2}\right)\right]}} \xrightarrow{D} N(0, 1), \qquad l = 1, \ldots, r.$$

The proof of the Theorem would follow, if it further holds that $T_{2s,std}\left(x_{l1}, x_{l2}\right)$, $l = 1, \ldots, r$ are asymptotically independent.

This follows from a version of Rebolledo's central limit theorem (see Andersen *et. al.*, 1992), noting 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. A similar argument in a different context can be found in Bhattacharjee (2006).

It follows that

$$\begin{bmatrix} T_{2s,std}\left(x_{11}, x_{12}\right) \\ T_{2s,std}\left(x_{21}, x_{22}\right) \\ \vdots \\ T_{2s,std}\left(x_{r1}, x_{r2}\right) \end{bmatrix} \xrightarrow{D} N\left(\underline{\mathbf{0}}, \mathbf{I_r}\right),$$

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

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.

□

References

[1] Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1992). *Statistical Models based on Counting Processes.* Springer-Verlag, New York.

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

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

[4] Bhattacharjee, A. (2006). Testing Proportionality in Duration Models with Respect to Continuous Covariates. *Mimeo.*

[5] ——, Higson, C., Holly, S. and Kattuman, P. (2002). Macro economic instability and business exit: Determinants of failures and acquisitions of large UK firms. DAE Working Paper No. **0206**, Department of Applied Economics, University of Cambridge.

[6] Breslow, N.E. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* **57**, 579–594.

[7] Brown, B.W., Jr., Hollander, M. and Korwar, R.M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In: *Reliability and Biometry, Statistical Analysis of Lifelength* (Eds.) Proschan, F. and Serfling, R.J., Society for Industrial and Applied Mathematics: Philadelphia, 327–354.

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

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

[10] Fleming, T.R. and Harrington, D.P. (1991). *Counting processes and survival Analysis.* John Wiley and Sons, New York.

[11] Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika* **52**, 203–223.

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

[13] Gu, M., Follmann, D. and Geller, N.L. (1999). Monitoring a general class of two-sample survival statistics with applications. *Biometrika* **86**, 45–57.

[14] Harrington, D.P. and Fleming, T.R. (1982). A class of rank test procedures for censored survival data. *Biometrika* **69**, 133–143.

[15] Horowitz, J. L. (1996). Semiparametric estimation of a regression model with an unknown transformation of the dependent variable. *Econometrica* **64**, 103–107.

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

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

[18] Jespersen, N.C.B. (1986). Dichotomising a continuous covariate in the Cox regression model. Research Report **86/2**, Statistical Research Unit, University of Copenhagen.

[19] Jones, M.P. and Crowley, J.J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics* **45**, 157–170.

[20] —— and —— (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Annals of Statistics* **18**, 1203–1220.

[21] Kortram, R.A., A.J. Lenstra, G. Ridder, and A.C.M. van Rooij (1995). Constructive identification of the mixed proportional hazards model. *Statistica Neerlandica* **49**, 269–281.

[22] Lenstra, A.J. and van Rooij, A.C.M. (1998). Nonparametric estimation of the mixed proportional hazards model. Working paper, Free University, Amsterdam.

[23] Li, Y.-H., Klein, J.P. and Moeschberger, M.L. (1996). Effects of model misspecification in estimating covariate effects in survival analysis for small sample sizes. *Computational Statistics and Data Analysis* **22**, 177–192.

[24] Lin, D.Y., Wei, L.J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society* Series B **62**, 711–730.

[25] Lin, D.Y. and Ying, Z. (2001). Semiparametric and nonparametric regression analysis of longitudinal data (with discussion and a rejoinder). *Journal of the American Statistical Association* **96**, 103–126.

[26] Liu, P.Y., Green, S., Wolf, M. and Crowley, J. (1993). Testing against ordered alternatives for censored survival data. *Journal of the American Statistical Association* **88**, 421, 153–160.

[27] —— and Tsai, W.Y. (1999). A modified logrank test for censored survival data under order restrictions. *Statistics and Probability Letters* **41**, 57–63.

[28] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemotherapy. Report* **50**, 163–170.

[29] Mau, J. (1988). A generalization of a nonparametric test for stochastically ordered distributions to censored survival data. *Journal of the Royal Statistical Society* Series B **50**, 403–412.

[30] Melino, A. and G.T. Sueyoshi (1990). A simple approach to the identifiability of the proportional hazards model. *Economics Letters* **33**, 63–68.

[31] 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.

[32] Neumann, G.R. (1997). Search models and duration data. In: *Handbook of Applied Econometrics Volume II: Microeconometrics* (Eds.) Pesaran, M.H., Basil Blackwell: Oxford, Chapter 7, 300–351.

[33] O'Brien, P.C. (1978). A nonparametric test for association with censored data. *Biometrics* **34**, 243–250.

[34] Peto, R. and Peto, J. (1972). Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society* Series A **135**, 185–206.

[35] Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika* **65**, 167–179. Correction: **70**, 304 (1983).

[36] 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.

[37] 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.

[38] Tarone, R.E. (1975). Tests for trend in life table analysis. *Biometrika* **62**, 679–682.

[39] —— and Ware, J.H. (1977). On distribution-free tests for equality of survival distributions. *Biometrika* **64**, 156–160.

x