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

Testing for Proportional Hazards with Unrestricted Univariate Unobserved Heterogeneity

Bhattacharjee, Arnab

Publication date:
2009

Document Version
Peer reviewed version

Link to publication in Discovery Research Portal

Citation for published version (APA):
Bhattacharjee, A. (2009). Testing for Proportional Hazards with Unrestricted Univariate Unobserved Heterogeneity. (Discussion Paper Series; No. 0904). St. Andrews: Department of Economics, University of St. Andrews.

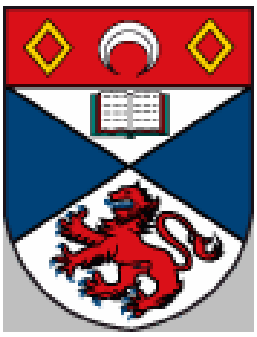
General rights
Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- 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



Testing for Proportional Hazards with Unrestricted Univariate Unobserved Heterogeneity

Arnab Bhattacharjee

No. 0904

DISCUSSION PAPER SERIES

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

Testing for Proportional Hazards with Unrestricted Univariate Unobserved Heterogeneity

Arnab Bhattacharjee*
University of St. Andrews, UK.

Abstract

We develop tests of the proportional hazards assumption, with respect to a continuous covariate, in the presence of unobserved heterogeneity with unknown distribution at the individual observation level. The proposed tests are specially powerful against ordered alternatives useful for modeling non-proportional hazards situations. By contrast to the case when the heterogeneity distribution is known up to finite dimensional parameters, the null hypothesis for the current problem is similar to a test for absence of covariate dependence. However, the two testing problems differ in the nature of relevant alternative hypotheses. We develop tests for both the problems against ordered alternatives. Small sample performance and an application to real data highlight the usefulness of the framework and methodology.

Keywords: Two-sample tests; Increasing hazard ratio; Trend tests; Partial orders; Mixed proportional hazards model; Time varying coefficients.
JEL Classification: C12, C14, C24, C41.

*Address: School of Economics & Finance, University of St. Andrews, Castlecliffe, St. Andrews KY16 9AL, UK. Tel: +44 1334 462423. e-mail: ab102@st-andrews.ac.uk.
Some of the results reported here were previously included in "A Simple Test for the Absence of Covariate Dependence in Hazard Regression Models", Discussion Paper **0708**, School of Economics & Finance, University of St. Andrews, UK. The current paper is a substantially extended and thoroughly revised version of the previous paper. The author thanks Laurie Davies, Anup Dewanji, Sean Holly, Rod McCrorie, Geert Ridder and Debasis Sengupta for useful comments and suggestions. The usual disclaimer applies.

1 Introduction

The main contribution of this paper is to develop tests for proportional hazards in the presence of individual level unobserved heterogeneity with completely unrestricted and unknown distribution. The proportional hazards assumption, routinely made in empirical duration studies, is often violated in applications. Further, allowing for an arbitrary unobserved heterogeneity distribution is useful in the hazard regression analyses. Our tests are developed in the context of a very general hazard regression model with no restrictions on duration dependence, minimal structure on the covariate effects, and with arbitrary unobserved heterogeneity. Importantly, the form of our test statistic is new to the literature, and the applicability of our asymptotic results go beyond the scope of the current problem.

The Cox regression model (Cox, 1972), and more generally the proportional hazards (PH) model, provides a convenient way to evaluate the influence of one or several covariates on the hazard rate of duration spells, and thus plays an important role in the theory and practice of hazard regression models. However, the PH specification substantially restricts interdependence between the explanatory variables and duration 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 duration. 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 seek 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 duration 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

[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.

the alternative of 'increasing hazard ratio', which is equivalent to convex partial order of the duration 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). These two-sample tests are useful for analysing duration data – not only are they powerful in detecting departures from proportionality, they also provide further clues about the nature of covariate dependence. However, their use in applications is limited because important covariates are often continuous (Horowitz and Neumann, 1992).

Bhattacharjee (2008) extended the notion of monotone hazard ratio in two samples to the continuous covariate case, and proposed tests for proportionality against ordered alternatives. These tests are useful when there is random effects heterogeneity in the nature of shared frailties, or when the distribution of individual level unobserved heterogeneity belongs to a known finite dimensional family. However, this approach is not applicable when the heterogeneity distribution is arbitrary and unknown.

In this paper, we develop tests for proportional hazards while allowing for an arbitrary unobserved heterogeneity distribution. This is very important, since both simulations (Bretagnolle and Huber-Carol, 1988; Baker and Melino, 2000) and empirical applications (Heckman and Singer, 1984a,b; Keiding *et al.*, 1997) show that inference is sensitive to the choice of the unobserved heterogeneity distribution. Further, characterising the unobserved heterogeneity distribution is a notoriously difficult problem (Horowitz, 1999). Also, our tests are applicable in very general settings, with no restrictions on duration dependence and only weak separability assumptions on the nature of covariate dependence.

The paper is organised as follows. In Section 2, we formulate the proposed test for proportional hazards under the mixed non-proportional hazards (MNPH) model incorporating unrestricted univariate heterogeneity. Identifying conditions of the MNPH model imply that testing for the PH assumption is the same testing as testing for equality of conditional hazard functions. Therefore, we extend tests for equality of hazard rates in two samples to testing for absence of covariate dependence with respect to continuous covariates, and then adapt these tests to our main testing problem. In Section 3, we develop the tests, outlining the relevant alternative hypotheses, assumptions and asymptotic properties, and discuss choice of weight functions. We present results of a Monte Carlo study in Section 4, followed by a real life application in Section 5. Finally, Section 6 concludes.

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

2 Formulation of the testing problems

2.1 Testing proportional hazards

We first consider the standard mixed proportional hazards (MPH) model:

$$\begin{aligned} \lambda\left(t|X=x,Z=z,U=u\right) &= \lambda_0\left(t\right)\exp\left[\beta_X.x+\beta_Z^T.z+u\right] \\ &\iff \ln\Lambda_0(T) = -\left(\beta_X.x+\beta_Z^T.z+U+\varepsilon\right) \quad (1) \end{aligned}$$

where $\Lambda_0(t)=\int_0^t\lambda_0(s).ds$ is an increasing function of arbitary shape (the cumulative baseline hazard function), X is the covariate under test and Z the vector of other covariates, log-heterogeneity U has an arbitrary distribution that is independent of the covariates X and Z, and ε has an extreme value distribution; see, for example, Horowitz (1999). Since U has an arbitrary distribution, so does $U+\varepsilon$, and hence this is a special case of the monotonic transformation model considered, for example, by Han (1987), Sherman (1993) and Horowitz (1996).

Since our interest here is in testing whether the hazard functions conditional on different values of the covariate X are proportional, we now consider a more general mixed non-proportional hazards (MNPH) model where in addition to unrestricted univariate heterogeneity, the covariates have potentially non-proportional hazard effects modeled through time varying coefficients. The model is described by

$$\lambda\left(t|X=x,Z=z,U=u\right)=\lambda_0\left(t\right)\exp\left[\beta_X\left(t\right).x+\beta_Z(t)^T.z+u\right], \quad (2)$$

with covariates X and Z, which are both allowed to have potentially time varying effects ($\beta_X\left(t\right)$ and $\beta_Z(t)$).[3] Under this model, the null hypothesis of proportional hazards corresponds to covariate effects constant over duration

$$\mathbb{H}_{0,PH}:\beta_X(t)\equiv b, \quad (3)$$

and the ordered alternative of monotone covariate effects

$$\frac{\lambda\left(t|X=x_2,Z=z,U=u\right)}{\lambda\left(t|X=x_1,Z=z,U=u\right)}\uparrow t \quad \text{whenever} \quad x_2>x_1, \quad \text{for all} \quad z,u, \quad (4)$$

corresponds to increasing time varying coefficients

$$\mathbb{H}_{1,PH}:\beta_X(t)\uparrow t. \quad (5)$$

[3]While we assume fixed covariates for simplicity, time varying covariates can be considered by a simple extension. Specifically, we can place a histogram sieve (Grenander, 1981) on the covariate over the duration scale, and restrict the time varying coefficient corresponding to every time interval to be zero except on the specific interval considered.

While, we assume the MNPH model (2) for expositional simplicity, the methods developed here are valid within the context of the model

$$\lambda\left(t|X=x,Z=z,U=u\right)=\lambda_0\left(t\right)\exp\left[\beta_X\left(x,t\right)+\beta_Z\left(z,t\right)+u\right],$$

where the covariate effects are completely unrestricted. This is about the most general hazard regression model that can be considered in this problem.[4]

Sufficient conditions for identifiability of the MNPH model with individual level unobserved heterogeneity and time varying coefficients (2) is the inclusion of a covariate with proportional hazards that has support over the whole real line (McCall, 1996). We feel this condition may be justifiable in empirical applications. McCall (1996) suggests estimation of the model using the histogram sieve estimator (Murphy and Sen, 1991) for time varying coefficients, in combination with unrestricted heterogeneity distribution modeled as a sequence of discrete multinomial mixtures with increasing number of support points (Heckman and Singer, 1984a).

The alternative hypothesis (4, 5) is similar to the *increasing hazard ratio for continuous covariate* ($IHRCC$) condition (Bhattacharjee, 2008), suggesting that tests similar to Bhattacharjee (2008) may be useful here. However, the formulation of our testing problem has to be modified to reflect the identifying restrictions of the transformation model (1, 2).[5] Specifically, since the MNPH model still continues to hold if a constant is added to both sides, a location normalisation is required for identification. This can be achieved by setting

$$\Lambda_0(t_0)\equiv 1$$

for some fixed and finite $t_0>0$.[6] In fact, our tests of the PH assumption will be based on the shape of the estimated baseline hazard function conditional on different values of the covariate X. Accordingly, the above normalisation

[4]The main assumption underlying this model is that of multiplicative separability of the effect of X, Z and U on the conditional hazard rate.

[5]Strictly speaking, the MNPH model (2) is not a linear transformation model. However, it can be cast as a transformation model, if one makes the assumption that the time varying coefficients are piecewise constant, changing values across known intervals. In the histogram sieve implementation, the width of these intervals is allowed to decrease to zero with sample size.

[6]Note that, the MNPH model has an important distinction from the standard transformation model. Here, β_X and β_Z are exactly identified by the fact that ε has the extreme value distribution. Since the scale of ε is fixed, a scale transformation is not required in this case. However, the scale parameter is 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 $n^{-2/5}$ (Ishwaran, 1996), which is slower than the usual convergence rate of $n^{-1/2}$; see Horowitz (1999) for further discussion.

takes the form

$$\Lambda_0(t_0|X=x) = \int_0^t \lambda_0(s).\exp\left[\beta_X(s).x\right].ds \equiv 1, \tag{6}$$

conditional on every covariate value $X = x$

Because of the scale normalisation, the baseline cumulative hazard function in (6) is only identified upto a factor of proportionality, restricting it to take the value unity at a fixed duration t_0. As a result, if the covariate X has proportional hazards effect, the constrained baseline cumulative hazard function conditional on different covariate values will be equal. Correspondingly, nonproportional covariate effects imply that cumulative baseline hazard functions conditional on different covariate values, while constrained to be equal at t_0, will be different at other durations. Therefore, nonproportionality implies violation of equality of the cumulative baseline hazard functions condtional on different covariate values.

In other words, the above normalisation renders testing for proportionality equivalent to testing the equality of hazard functions conditional on different values of the chosen covariate, X. Based on the above argument, our modified null hypothesis is

$$\begin{aligned} \mathbb{H}_{0,PH} \quad &: \quad \Lambda_0(t|X=x) = \Lambda_0(t) \quad \text{for all} \;\; x \\ &\iff \quad \lambda_0(t|X=x_1) = \lambda_0(t|X=x_2) \quad \text{for all} \;\; x_1 \neq x_2, \end{aligned} \tag{7}$$

where $\lambda_0(t|X=x) = \lambda_0(t).\exp\left[-\left(\beta_X(t).x\right)\right]$. The proposed test will extend two sample tests for equality of hazard functions to the continuous covariate setup. The relevant alternative hypotheses, discussed in Section 3, will determine the appropriate choice of the underlying two-sample test statistics from various options available from the literature.

2.2 Testing absence of covariate dependence

We now turn to a related testing problem suggested by the modified null hypothesis (7). Since the null hypothesis of proportional hazards is equivalent to equality of baseline hazard rates conditional on different values of the index covariate X, the above testing problem is closely related to testing for the absence of covariate dependence. This itself is an important inference problem, particularly since understanding the nature of covariate dependence is one of the main objectives of regression analysis of duration data.

We consider the general hazard regression model (without unobserved heterogeneity)

$$\lambda\left(t|X=x, Z=z\right) = \lambda_0\left(t\right)\exp\left[\beta_X\left(x,t\right) + \beta_Z\left(z,t\right)\right],$$

where, as before, X and Z are covariates with completely unrestricted covariate effects. Our interest is to test whether the covariate X has any effect on the hazard rate. As discussed in Bhattacharjee (2004), by suitable transformations and use of the histogram sieve, the effect of the other covariates Z can be approximated by time varying effects:

$$\beta_Z(z,t) = \beta_Z(t)^T.z,$$

which is a convenient form for regression modeling.

Within the context of the above model, absence of covariate dependence can be assessed by conducting a test of the hypothesis

$$\begin{aligned} \mathbb{H}_{0,Eq} \quad &: \quad \lambda_0(t|X=x) = c(t) \quad \text{for all } x \\ &\iff \quad \lambda_0(t|X=x_1) = \lambda_0(t|X=x_2) \quad \text{for all } \; x_1 \neq x_2 \end{aligned} \tag{8}$$

against relevant alternatives. The similarity between the above null hypothesis (8) and that for testing proportional hazards (7) suggests that similar tests can be developed for either case.

The choice of the alternative hypothesis usually depends on the expected nature of covariate dependence. We propose tests for the null hypothesis of absence of covariate dependence where the covariate is continuous and the alternative hypothesis is either omnibus

$$\mathbb{H}_{1,Eq}: \text{ not } \mathbb{H}_{0,Eq}, \tag{9}$$

or trended (when the covariate has positive or negative effect), or changepoint trended (when the sign of the covariate effect, positive or negative, varies over different regions of the sample space). We will focus mainly on trended and changepoint trended alternatives since these are more useful in regression modeling; we discuss relevant alternative hypotheses in Section 3.

Finally, note that we have not considered unrestricted unobserved heterogeneity in our regression model specification for the test for absence of covariate dependence. In fact, an important implication of the location normalisation (6) inherent in the corresponding MNPH model is that, absense of covariate dependence cannot be tested in this model. This is because equality of the conditional hazard rates also results from proportional (but unequal) conditional hazard functions. However, models with either shared frailty or with finite dimensional heterogeneity distributions are accommodated easily within our framework. Further, the case of unrestricted unobserved heterogeneity distribution can be addressed under the time varying coefficients model, by developing tests for the condition $\beta_X(t) = 0$ for all t; we do not discuss this case here.

2.3 Estimation of baseline hazard functions

Our proposed inference procedures for the above two testing problems will be based on estimates of the conditional baseline cumulative hazard functions. Various candidate estimators for the cumulative baseline hazard $\widehat{\Lambda}_0(t|x_1)$, $\widehat{\Lambda}_0(t|x_2), \ldots$, conditional on different covariate values $X = x_1, x_2, \ldots$, in models including additional covariates, Z, and possibly unrestricted univariate unobserved heterogeneity, are available in the literature.

For the hazard regression model with time varying coefficients but without unobserved heterogeneity, the histogram sieve estimator (Murphy and Sen, 1991) can be used. While several alternative estimators have been proposed in the literature, including the ones proposed by Zucker and Karr (1990) and Martinussen *et al.* (2002), we use the histogram sieve estimator in our tests for absence of covariate dependence. The choice is based on simplicity for use and interpretation.

For duration data with shared frailties, one can either use the marginal modeling approach with unrestricted heterogeneity distribution (Spiekerman and Lin, 1998), or assume gamma frailties and use the efficient estimator proposed by Parner (1998). Kosorok *et al.* (2004) have proposed another estimator, which can be used when the distribution of individual level unobserved heterogeneity can be assumed to belong to a given one-parameter family of continuous distributions.

For the tests of proportional hazards, we focus on the unrestricted univariate unobserved heterogeneity case. Contributions in this area are rather limited. Of particular interest are the kernel-based estimators of the baseline cumulative hazard function proposed in Horowitz (1999) and Gørgens and Horowitz (1999), in the presence of scalar unobserved heterogeneity with completely unrestricted distribution. The proposed estimators for the baseline hazard function and baseline cumulative hazard function can be made to converge at a rate arbitrarily close to the optimal $n^{-2/5}$ by suitable choice of bandwidths (Horowitz, 1999). However, the choice of bandwidths and other tuning parameters is itself a difficult problem in implementation. Further, the methods do not allow for time varying covariates. While an extension to this case is certainly possible, the properties of such estimators are yet to be studied.

With discrete duration data, estimation of the baseline hazard function reduces to a simpler problem. Further, if one approximates the unknown heterogeneity distribution by a sequence of discrete mixtures of degenerate distributions (Heckman and Singer, 1984a), estimation of the heterogeneity distribution also becomes a parametric problem. The approach, proposed by Jenkins (1995), of considering the grouped time proportional hazards

model (Prentice and Gloeckler, 1978) in combination with discrete mixture unobserved heterogeneity is therefore an attractive strategy.[7]

In summary, a variety of estimators of the conditional baseline hazard function are available. Most of these estimators, suitably normalised, converge weakly to a Gaussian processes under appropriate assumptions. For the construction of our proposed tests, we assume that such an estimator has been chosen. In practise, an appropriate choice will depend both on the assumed underlying model and properties of the estimator.

3 Proposed tests

In this Section, we discuss test procedures for the two testing problems, (8) and (7). We first describe the alternative hypotheses, followed by the test for absence of covariate effect, and then the test for proportionality. The proposed class of tests allow the user to choose a relevant weight function, the choice depending on the relevant null and alternative hypotheses. We establish the statistical properties of the proposed tests, and discuss the choice of weight functions.

3.1 Alternative hypotheses

As discussed in the previous Section, the null hypothesis for both the tests posit that the hazard functions conditional on different covariate values are the same. However, our alternative hypotheses in these two cases are different, and reflect the expected nature of departures from the null.

Consider first the problem of testing whether the covariate X has proportional hazard effects against ordered alternatives. Specifically, we consider alternatives defined by nonproportional partial orders, specifically $IHRCC$ or *decreasing hazard ratio with continuous covariate* ($DHRCC$):

$$\begin{aligned} IHRCC \quad &: \quad \text{whenever } x_1 > x_2, \lambda(t|x_1)/\lambda(t|x_2) \uparrow t \\ &\qquad \left[\equiv (T|X=x_1) \underset{c}{\prec} (T|X=x_2)\right], \end{aligned} \tag{10}$$

$$\begin{aligned} DHRCC \quad &: \quad \text{whenever } x_1 > x_2, \lambda(t|x_2)/\lambda(t|x_1) \uparrow t \\ &\qquad \left[\equiv (T|X=x_2) \underset{c}{\prec} (T|X=x_1)\right], \end{aligned} \tag{11}$$

where we supress dependence of other covariates Z and heterogeneity U for notational convenience.

[7] An alternative approach based on maximum rank correlations (Han, 1987), proposed by Hausman and Woutersen (2005), may also be useful. This method treats the unknown heterogeneity distribution as nuisance parameters.

Let us initially consider two distinct covariate values, x_1 and x_2. Our strategy is to first test for proportional hazards against the partial order conditional on these two values, and then extend the test by considering multiple covariate pairs. Without loss of generality, the two distinct values of the covariate X, $x_1 > x_2$, can be set to $x_1 = 1$ and $x_2 = 0$. In this binary covariate case, the most general model is the time varying covariate effects model

$$\lambda\left(t|x,z,u\right) \equiv \lambda_{0,x_2}(t).\exp\left[\beta_{(x_1>x_2)}(t).x\right].\exp\left[\beta_Z\left(z(t),t\right)+u\right],$$

under the assumption of multiplicative separability in the effects of X, Z and U. This implies that the PH null hypothesis (7) can be restated as

$$\mathbb{H}_{0,PH,(x_1>x_2)} : \beta_{(x_1>x_2)}(t) = 0, \quad \text{for all } \ t,$$

where we add $(x_1 > x_2)$ to the index set to emphasize that the statement of the null is specific to this covariate pair.

Now, under the MNPH model, the ordered alternative $IHRCC$

$$\mathbb{H}_{1,PH,(x_1>x_2)} : \lambda(t|x_1)/\lambda(t|x_2) \uparrow t$$

holds if and only if $\beta_{(x_1>x_2)}(t) \uparrow t$. Since identifiability restrictions under the model require that $\Lambda\left(t_0|x\right) \equiv 1$, the following conditions must therefore hold

$$\int_0^{t_0} \lambda_{0,x_2}(s)ds = 1 \ \text{ and } \ \int_0^{t_0} \lambda_{0,x_2}(s)\exp\left[\beta_{(x_1>x_2)}(s)\right]ds = 1.$$

In other words, under the alternative hypothesis $\beta_{(x_1>x_2)}(t)$ starts from a negative value at $t = 0$, rises to a positive value at $t = t_0$ such that the above relationship holds, and continues to rise thereafter.

Thus, the PH assumption is represented by a null hypothesis of equal conditional hazards, and the alternative posits monotone covariate effect with crossing hazards character. We consider tests of the above hypotheses by extending two sample tests for equality of hazard functions. Several tests of this hypothesis will be conducted, corresponding to different pairs of covariate values. Our tests for proportionality of hazards will be based on a combination of several two sample tests.

The underlying two sample tests are rank tests of the form

$$T_{2s} = \int_0^{\tau} L(t)d\widehat{\Lambda}_1(t) - \int_0^{\tau} L(t)d\widehat{\Lambda}_2(t), \tag{12}$$

where $L(.)$ is some appropriate weight function and τ is a large duration, either fixed or random. Most of the standard censored data two sample tests

for equality of hazard functions belong to this general class with different choices of the weight function.[8] Later, we discuss how these tests can be adapted to our specific testing problem, and the related issue of choice of weight functions.

When the covariate is binary or categorical, the above tests are often used to test the null hypothesis of absence of covariate dependence (8) against the omnibus alternative (9). However, the omnibus alternative is often too broad and does not convey sufficient information about the nature of covariate dependence. In many empirical applications, it is important to infer not only whether there is 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 duration spell.

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 against trend in hazard functions have been proposed by Tarone (1975), while Liu and Tsai (1999) have proposed ordered weighted logrank tests to detect similar trend in survival functions. As discussed earlier, such two-sample and k-sample tests are of limited use in applications.

In our setting, the trended alternative that the covariate has a positive or negative effect on the hazard function can be represented by the hypothesis

$$\begin{aligned} \mathbb{H}_{1,Eq}^{(t)} \quad &: \quad \lambda(t|x_1, z) \geq \lambda(t|x_2, z) \qquad \text{for all } z \text{ and } t \\ \text{whenever } x_1 \quad &> \quad x_2 \text{ (or its dual)}, \end{aligned} \tag{13}$$

the strict inequality holding for at least one covariate pair (x_1, x_2). The changepoint trended alternative posits 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:

$$\begin{aligned} \mathbb{H}_{1,Eq}^{(c)} \quad &: \quad \text{there exists } x^* \text{ such that } \lambda(t|x) \uparrow x \qquad \text{for all } z \text{ and } t \\ \text{whenever } x \quad &< \quad x^*, \text{ and } \lambda(t|x) \downarrow x \text{ whenever } x > x^* \text{ (or its dual).} \end{aligned} \tag{14}$$

[8] The Mantel-Haenszel or logrank test (Peto and Peto, 1972), one of the most popular tests in this class, has optimal power if the two compared groups have proportional hazard functions under the alternative (Peto and Peto, 1972). The Gehan-Breslow (Breslow, 1970) and Prentice (1978) tests generalise the Wilcoxon and Kruskal-Wallis tests to right censored data. The theoretical properties of these and other related rank tests and their use in applications have been discussed elsewhere (Fleming and Harrington, 1991; Andersen *et al.*, 1993).

Some trend tests in the literature are specific to continuous covariates and consider (13) as the alternative hypothesis. If an underlying hazard hazard regression model is assumed (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). Jones and Crowley (1990) consider a general class of test statistics which nests most of the other trend tests as well as their robust versions.

All the above tests are rather restrictive since, they assume either validity of a specified regression model, or a known covariate label function. Therefore, they fail to retain the attractive nonparametric flavour of the corresponding two-sample or k-sample tests. Further, these tests are not useful when covariate dependence is in the nature of a changepoint trend (14).

3.2 Testing absence of covariate dependence

First, we consider the single covariate case. Let T be a duration variable, X a continuous covariate and let $\lambda(t|x)$ denote the hazard rate of T at $T = t$, given $X = x$. We intend to test the hypothesis (8) against the alternative $\mathbb{H}_{1,Eq} : \lambda(t|x_1) \neq \lambda(t|x_2)$ for some $x_1 \neq x_2$. In particular, we are interested in test statistics that would be useful in detecting trend departures from $\mathbb{H}_{0,Eq}$ of the form $\mathbb{H}^{(t)}_{1,Eq}$ (13), and changepoint trend departures like $\mathbb{H}^{(c)}_{1,Eq}$ (14).

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

$$T_{2s,std} = \frac{T_{2s}}{\sqrt{\widehat{\mathrm{Var}}\left[T_{2s}\right]}}, \tag{15}$$

where

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

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,

$$K(t) = I\left[Y_1(t) + Y_2(t) > 0\right], \tag{16}$$

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

$$K(t) = I\left[Y_1(t) + Y_2(t) > 0\right].\{Y_1(t) + Y_2(t)\}. \tag{17}$$

These standardised two sample test statistics have zero mean under the null hypothesis of equal hazards and positive (negative) mean accordingly as the hazard functions are trended upwards (downwards). Further, they are asymptotically normally distributed under the null hypothesis.

Based on the above test statistics, we propose a simple construction of our tests as follows. We first select a fixed number, r, of pairs of distinct points on the covariate space, and construct the standard two-sample test statistics ($T_{2s,std}$) for each pair, based on counting processes conditional on two distinct covariate values. We then construct our test statistics, by taking maximum, minimum 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 $T_{2s}^{(\max)}$, $T_{2s}^{(\min)}$ 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

$$\begin{aligned}
T_{2s,std}(x_{l1}, x_{l2}) &= \frac{T_{2s}(x_{l1}, x_{l2})}{\sqrt{\widehat{\mathrm{Var}}\left[T_{2s}(x_{l1}, x_{l2})\right]}}, \\
T_{2s}(x_{l1}, x_{l2}) &= \int_0^\tau L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l1}) - \int_0^\tau L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l2}), \qquad (18) \\
\widehat{\mathrm{Var}}\left[T_{2s}(x_{l1}, x_{l2})\right] &= \int_0^\tau L^2(x_{l1}, x_{l2})(t)\{Y(t, x_{l1})Y(t, x_{l2})\}^{-1}.d\left[N(t, x_{l1}) + N(t, x_{l2})\right],
\end{aligned}$$

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:

$$T_{2s}^{(\max)} = \max_{l=1,\ldots,r} T_{2s,std}(x_{l1}, x_{l2}), \tag{19}$$

$$T_{2s}^{(\min)} = \min_{l=1,\ldots,r} T_{2s,std}(x_{l1}, x_{l2}), \tag{20}$$

$$\text{and } \overline{T}_{2s} = \frac{1}{r}\sum_{l=1}^{r} T_{2s,std}(x_{l1}, x_{l2}). \tag{21}$$

We now derive the asymptotic distributions of these test statistics.

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 $\lambda(t|x) = \lambda(t)$ for all t and x (under the null hypothesis of equal hazards). Let, as before, $L(x_1, x_2)(.)$ be a process indexed on a pair of distinct values of the continuous covariate x (*i.e.,* indexed on $(x_1, x_2), x_1 \neq x_2, x_1, x_2\epsilon\mathcal{X}$). 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, $l = 1, 2, \ldots, r$, let $L(x_{l1}, x_{l2})(t)$ be a predictable process indexed on the pair of fixed covariate values (x_{l1}, x_{l2}).*
Assumption 2 *Let τ be a random stopping time.*[9]
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 $l = 1, \ldots, r$, $L(x_{l1}, x_{l2})(t)$ is 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*

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

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

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

Remark 1. Assumptions 1 through 4 constitute an extension, to the continuous covariate framework, of the standard set of assumptions for the counting process formulation of duration data (see, for example, Andersen *et al.*, 1993). As discussed in Sengupta *et al.* (1998), the condition on probability limit of $Y(t,x)$ in Assumption 4 can be replaced by a set of weaker conditions. All the assumptions are satisfied by the standard random censorship model.

Let the test statistics $T_{2s}^{(\max)}, T_{2s}^{(\min)}$ and $\overline{T}_{2s}$ be as defined earlier (19 – 21).

[9] 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 $T_{2s,std}(x_{l1}, x_{l2}), l = 1, \ldots, r$.

Theorem 1. *Let Assumptions 1 through 4 hold. Then, under* $\mathbb{H}_{0,Eq}$: $\lambda_0(t|X=x)=c(t)$ for all x, *as* $n\to\infty$,
(a) $P\left[T_{2s}^{(\max)}\leq z^*\right]\to[\Phi(z^*)]^r$,

(b) $P\left[T_{2s}^{(\min)}\geq -z^*\right]\to[\Phi(z^*)]^r$, *and*

(c) $\sqrt{r}.\overline{T}_{2s}\xrightarrow{D}N(0,1)$,
where $\Phi(z^*)$ *is the distribution function of a standard normal variate.*
(Proof in Appendix 1).

Corollary 1.

$$\begin{aligned} P\left[a_r\left\{T_{2s}^{(\max)}-b_r\right\}\leq z^*\right] &\to \exp\left[-\exp(-z^*)\right] \text{ as } r\to\infty \\ \text{and } P\left[a_r\left\{T_{2s}^{(\min)}+b_r\right\}\geq z^*\right] &\to \exp\left[-\exp(z^*)\right] \text{ as } r\to\infty, \end{aligned}$$

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 1).

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*

$$\begin{aligned} T_{2s,\underline{w}}^{(\max)} &= \max_{l=1,\ldots,r}\left\{w_l.T_{2s,std}(x_{l1},x_{l2})\right\}, \\ T_{2s,\underline{w}}^{(\min)} &= \min_{l=1,\ldots,r}\left\{w_l.T_{2s,std}(x_{l1},x_{l2})\right\}, \\ \text{and } \overline{T}_{2s,\underline{w}} &= \frac{\sum_{l=1}^r w_l.T_{2s,std}(x_{l1},x_{l2})}{\sum_{l=1}^r w_l}. \end{aligned}$$

Let Assumptions 1 through 4 hold. Then, under $\mathbb{H}_{0,Eq}$, *as* $n\to\infty$,
(a) $P\left[T_{2s,\underline{w}}^{(\max)}\leq z^*\right]\to\prod_{l=1}^r[\Phi(z^*/w_l)]$,

(b) $P\left[T_{2s,\underline{w}}^{(\min)}\geq -z^*\right]\to\prod_{l=1}^r[\Phi(z^*/w_l)]$, *and*

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

Remark 2. 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 $\{(x_1,x_2):\ x_2>x_1,x_1,x_2\epsilon\mathcal{X}\}$ is pointwise standard normal and independent, but do not have a well-defined limiting process.

Therefore, if r is allowed to grow, the maximum (minimum) diverges to $+\infty$ $(-\infty)$.

Remark 3. Corollary 1 provides a simple way to compute p-values for the test statistics when r is reasonably large.[10] 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.

Remark 4. 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. In our empirical implementation, we consider "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; see also Bhattacharjee (2008). This issue can be addressed by using a jacknife estimator for the variance of the average estimator.

Remark 5. By construction, the tests are consistent against the trended alternative (13). The average test statistic $\overline{T}_{2s}$ has asymptotically Gaussian distributions under both the null and alternative hypothesis, with mean zero under the null and positive mean under the alternative. Under the null hypothesis of absence of covariate effect, the maxima test statistic $T_{2s}^{(\max)}$ has the extreme value distribution given in Theorem 1, whereas under the trended alternative (13), it diverges to $+\infty$; therefore, the test is consistent. Similarly, the average and minima test statistics are consistent when departures are trended in the opposite direction: $\lambda(t|x_1) \leq \lambda(t|x_2)$ whenever $x_1 > x_2$. Further, both the maxima and the minima test statistics are consistent when there is a changepoint trend in the covariate effect (14). The ability to detect both trended and changepoint trended covariate effects highlights an important advantage of the proposed tests. The power of the tests depend on the choice of weight functions, which we discuss in Section 3.4.

Remark 6. The choice of the r pairs of covariate values may be important in applications. The issues regarding this choice are similar to stratification in goodness-of-fit tests. Quantiles of the cross-sectional distribution of the covariate can be used to select non-overlapping intervals and corresponding covariate pairs.[11]

[10] Note that r is fixed and finite; however, if it assumes a large enough value (say, 20 or higher), the approximation can be useful.

[11] This procedure adjusts for variations in the design density (none of the intervals are too sparse) and ensures that intervals corresponding to each pair of covariate values do

When other covariates, Z, are also present, we start by estimating a modified Cox model, with possibly time varying coefficients on Z, by the histogram sieve method (Murphy and Sen, 1991). The regression coefficients are estimated by partial likelihood estimators, $\widehat{\beta}_Z$, and the baseline cumulative hazard function by the standard Breslow estimator (Breslow, 1974), $\widehat{\Lambda}(t, x_{lj}, \widehat{\beta}_Z)$. This estimator of the baseline cumulative hazard function is plugged into the two sample test statistic (12) in place of the Nelson-Aalen estimator of the cumulative baseline hazard function, giving

$$T_{2s}^{(Z)}(x_{l1}, x_{l2}) = \int_0^\tau L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l1}, \widehat{\beta}_Z) - \int_0^\tau L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l2}, \widehat{\beta}_Z).$$

The asymptotic properties follow in a similar way as above, by noting that, in place of the usual counting process martingale, we now have

$$\widehat{M}(t, x_{lj}) = N(t, x_{lj}) - \int_0^t Y(s, x_{lj}) \,.\, \exp\left[\widehat{\beta}_Z(s)^T .z(s)\right] .d\widehat{\Lambda}(t, x_{lj}, \widehat{\beta}_Z)$$

which is a local martingale (Andersen *et al.*, 1993).

When there is shared heterogeneity or parametric frailty, the tests are constructed as above, using an appropriate estimator for the baseline cumulative hazard function. Though martingale based arguments are no longer valid, the asymptotic arguments continue to hold, with some minor modifications; see Spiekerman and Lin (1998). For the MPH model with univariate parametric heterogeneity, a procedure similar to continuously distributed unrestricted heterogeneity distribution (Section 3.3) can be used.

Finally, in applications with multiple covariates, the tests developed here can be used to sequentially evaluate the absence of covariate dependence for the covariates. This provides an intuitive and convenient way to build an appropriate hazard regression model in such cases (Scheike and Martinussen, 2004).

3.3 Testing the proportional hazards assumption

Our proposed tests for proportional hazards are similar to the above problem. Here, too, we estimate the baseline cumulative hazard function under maintained assumptions on the model and nature of unobserved heterogeneity, and plug these estimators into the two sample test statistic (12) in place of

not overlap. In our simulation study, we divided the sample into deciles by the magnitude of the covariate, and based our tests on the corresponding $\binom{10}{2} = 45$ covariate pairs, while for the empirical application we used 20 covariate pairs obtained by random sampling.

the Nelson-Aalen estimator. The asymptotic properties are similar to Theorem 1 and Corollaries 1 and 2. However, the assumptions underlying the tests reflect the differences in the models and methods, and similarly there are important differences in the asymptotic arguments. We first discuss continuous duration data with arbitrary continuous unobserved heterogeneity, followed by discrete duration data with discrete mixture unobserved heterogeneity distribution.

3.3.1 Arbitrary continuous heterogeneity distribution

First consider the estimation procedure proposed by Horowitz (1999) for the continuous duration MPH model with unrestricted continuously distributed unobserved heterogeneity. The estimator for the baseline hazard function extends an estimator for the transformation model (Horowitz, 1996), accounting for censoring and the fact that the scale of the MPH model with time varying coefficients (2) is fixed by the extreme value distribution for ε. Horowitz (1999) proposed estimating the scale separately and plugging this into the transformation model estimator for the baseline cumulative hazard function.

We assume that the effect of the other covariates Z has been modeled *a priori* and a well-specified MNPH model with time varying coefficients (2),

$$\lambda\left(t|X=x,z,u\right)=\lambda_0\left(t,x\right)\exp\left[\beta_Z(t)^T.z(t)+u\right],$$

has been found. This model is then estimated, conditional on various covariate values. We denote by $\widehat{\lambda}_{0,H}\left(t,x\right)$ the corresponding estimator of the baseline hazard function, incorporating unrestricted unobserved heterogeneity and conditional on $X=x$.

Here, our test is similar to Section 3.2, starting with the choice of $r>1$ and selection of $2r$ (r is a fixed positive integer, $r>1$) distinct points, $\left\{x_{11},x_{21},\ldots,x_{r1},x_{12},x_{22},\ldots,x_{r2}\right\}$, $x_{l2}>x_{l1}, l=1,\ldots,r$ on the covariate space $\mathcal{X}$. Next, we construct the basic statistics as

$$\begin{aligned}
T_{H,std}(x_{l1},x_{l2}) &= \frac{T_H(x_{l1},x_{l2})}{\sqrt{\widehat{\mathrm{Var}}\left[T_H(x_{l1},x_{l2})\right]}}, \qquad (22)\\
T_H(x_{l1},x_{l2}) &= \int_0^{\tau^*} L(x_{l1},x_{l2})(t).\widehat{\lambda}_{0,H}\left(t,x_{l1}\right).dt\\
&\quad -\int_0^{\tau^*} L(x_{l1},x_{l2})(t).\widehat{\lambda}_{0,H}\left(t,x_{l2}\right).dt,\\
\widehat{\mathrm{Var}}\left[T_H(x_{l1},x_{l2})\right] &= \int_0^{\tau^*}\int_0^{\tau^*} \widehat{c}(t).\widehat{c}(s).\widehat{\sigma}^2_L(x_{l1},x_{l2})(s\wedge t).ds.dt,
\end{aligned}$$

where $L(x_{l1}, x_{l2})(t)$ is a random process indexed on the pair of covariate values x_{l1} and x_{l2}, $\widehat{\sigma}^2_L(x_{l1}, x_{l2})(t)$ is the sample variance (pointwise) of $L(x_{l1}, x_{l2})(t)$, and

$$\widehat{c}(t) = \left[\widehat{\lambda}_{0,H}(t, x_{l1}) - \widehat{\lambda}_{0,H}(t, x_{l2})\right].$$

As in (19 – 21), these basic statistics are combined to construct our maxima, minima and average test statistics ($T_H^{(\max)}, T_H^{(\min)}$ and $\overline{T}_H$, respectively).

We now state the assumptions required for our asymptotic results. The first two assumptions pertain to our testing procedure, while the following three relate to the estimator for baseline hazard function under unrestricted unobserved heterogeneity. For the sake of brevity, we give only a brief flavour of the kind of assumptions required for estimation, and refer to Horowitz (1999) for technical details.

The duration data $(T_i, \delta_i, X_i, Z_i(t), U_i)$ are independently and identically sampled from the MNPH model (2), for $i = 1, \ldots, n$. Here, T_i denotes the observed duration, δ_i is the censoring indicator, X_i and $Z_i(t)$ are covariates, and U_i is the unobserved heterogeneity. The following additional assumptions apply.

Assumption 5 *The cut-off duration, $\tau^* > t_0 > 0$, is a (large) positive duration such that $\Lambda_0(\tau^*, x_{lj}) < \infty, l = 1, 2, \ldots, r, j = 1, 2$. The intermediate duration t_0 is specified in Assumption 7 (b) below.*

Assumption 6 *For each $l, l = 1, 2, \ldots, r$, let $L(x_{l1}, x_{l2})(t)$ be a monotonic stochastic process with sample paths in $D[0, \infty)$ (i.e., right continuous with left limits), and with pointwise finite first and second moments over the interval $[0, \tau^*]$.*

Assumption 7 *(Identifiability conditions)*

(a) *Unobserved heterogeneity U is independent of covariates Z and censoring, and there is a tail restriction on the heterogeneity distribution.*[12]

(b) *For every covariate value $X = x$, $\Lambda_0(t, x)$ is strictly increasing on $[0, \infty)$ and takes value unity at a fixed t_0 (location normalisation).*

(c) *The covariate effect of at least one of the covariates, say Z_1, is non-zero and spans the whole of the real line. The distribution of Z_1 is absolutely continuous with respect to all the others. There is no perfect multicollinearity amongst the covariates Z.*

[12] The tail condition is stronger than Heckman and Singer (1984a), but facilitates achieving a faster convergence rate (Horowitz, 1999).

(d) *Censoring is random, and possibly dependent on Z, but only through the single index $\beta_Z(t)^T.z(t)$. In particular, censoring can be dependent on X, the covariate under test.*

Assumption 8 *(Smoothness conditions and kernel properties)*

(a) *Smoothness conditions involving several bounded derivatives for the unknown heterogeneity distribution, the baseline cumulative hazard function, the regression single index, $\beta_Z(t)^T.z(t)$, and the distribution of the leading covariate Z_1.*

(b) *Several technical restrictions on admissible kernel functions and bandwidths.*

Assumption 9 *(Conditions on regression estimator) The underlying regression estimator for the transformation model converges at $n^{-1/2}$ rate and has bounded second moments.*

Remark 7. Dependence between unobserved heterogeneity and X is not ruled out. In fact, the setup allows censoring to depend on the covariates through the single index. This, in our view represents a strength of the methodology, particularly in allowing censoring to depend on the covariate under test.

Remark 8. The methodology does not directly allow for time varying covariates. However, if the regression coefficient is fixed, a time varying covariate can be naturally accomodated by replacing the time varying covariate by its average value over the observed duration. A similar approach can also be easily applied if the covariate has time varying coefficients modeled using a histogram sieve (*i.e.*, the coefficient is constant over time intervals).[13]

Remark 9. Standard regression estimators for the transformation model satisfy the covergence rate and finite second moments conditions. Further, the smoothness and kernel conditions are satisfied by the Horowitz (1999) estimator. However, appropriate choice of bandwidths and other tuning parameters is very important for good performance of the estimator. Finally, Assumptions 7 through 9 ensure pointwise consistency of the baseline hazard estimator, which is required for our tests.[14]

Additional conditions required for the test are given in Assumptions 5 and 6. These comprise a deterministic cut-off at a duration where the cumulative

[13] A standard assumption in the literature, that of bounded total variation in the time varying coefficients, is not required in the current setup.

[14] Horowitz (1999) shows that the estimator is uniformly consistent and pointwise asymptotically Gaussian.

hazard function is finite, and existence of second moments and monotonicity of the stochastic weight function. Another required assumption, that of continuity of the baseline hazard rate, is already assumed in the estimation procedure. Under the above assumptions, we have the following asymptotic results.

Theorem 2. *Let Assumptions 5 through 9 hold. Then, under* $\mathbb{H}_{0,PH}$: $\lambda_0(t|X=x) = c(t)$ for all x, *as* $n \to \infty$,

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

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

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

(Proof in Appendix 1).

Corollary 3.

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

$$\textit{and } P\left[a_r\left\{T_H^{(\min)} + b_r\right\} \geq z^*\right] \;\to\; \exp\left[-\exp(z^*)\right] \text{ as } r \to \infty,$$

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 1).

Remark 10. A result similar to Corollary 2 on covariate dependent weighted tests is also available here, details of which are omitted.

Remark 11. In Appendix 2, we consider an alternative estimator (Gørgens and Horowitz, 1999), where the above approach is not directly applicable, since an estimator is available only for the baseline cumulative hazard function. Hence, we propose an alternative strategy, which is intuitive and potentially promising, but nevertheless poses substantial challenges in variance estimation. We speculate that the bootstrap developed in Kosorok *et al.* (2004) should work in this case, but this requires further investigation.

Importantly, while the form of the above test is similar to the test for absence of covariate effects (Section 3.2), as well as the test in Bhattacharjee (2008), there is a major point of difference. The asymptotics here is derived by interpreting the test statistic as an integral of the baseline hazard function with respect to the weight function, which is exactly the opposite from our previous approach. This is because, in this case, the weight functions are independent while the baseline hazard estimates are dependent across the sample points. Different asymptotic arguments, based on empirical processes

rather than counting process martingales, are therefore required. Specifically, our proof of Theorem 2 establishes weak convergence results for statistics of the form

$$\int K^{(n)}(t).H^{(n)}(t).dt,$$

where $K^{(n)}(.)$ and $H^{(n)}(.)$ are stochastic processes involving data from all the n observations, $K^{(n)}(.)$ is manageable, and $H^{(n)}(.)$ is a plug-in estimator with a continuous probability limit. Asymptotic theory for these new kinds of statistics are derived by combining modern empirical process theory (Pollard, 1990) with Sengupta *et al.* (1998). While the derivations in our case are made simpler by the fact that the $K^{(n)}(.)$ process is monotonic (and therefore has pseudodimension 1 and is manageable), the applications of this result goes well beyond the current problem.[15]

In summary, we propose tests of the PH assumption based on the Horowitz (1999) estimator of the baseline hazard function under arbitrary continuous unobserved heterogeneity. An alternative approach based on the Gørgens and Horowitz (1999) estimator is also potentially attractive (Appendix 2).

3.3.2 Arbitrary discrete mixture heterogeneity

We now turn to an alternative nonparametric procedure to accomodate unrestricted unobserved heterogeneity. This is based on the Heckman and Singer (1984a,b) idea of chracterising the unknown heterogeneity distribution by discrete mixtures of degenerate distributions in a sequence with increasing number ($s = 2, 3, \ldots$) of components:

$$u_i \in \{m_1 = 0, m_2, \ldots, m_s\} = \begin{cases} m_1 & \text{with prob. } \pi_1 \\ m_2 & \text{with prob. } \pi_2 \\ \vdots & \\ m_s & \text{with prob. } \pi_s \end{cases}, \quad s = 2, 3, \ldots$$

The sequential procedure is terminated when subsequent steps lead to degeneracy or no improvement in the maximised likelihood. The Heckman-Singer methodology is very attractive in that it approximates the nonparametric frailty distribution by an increasing sequence of parametric distributions and

[15] For example, piecewise monotonic processes with finitely many turning points, or drawn from a finite mixture distribution of such processes, also have finite pseudodimension and are therefore manageable (Pollard, 1990). Further, linear combinations of such manageable processes are also manageable (Bilias *et al.*, 1997).

produces robust estimates of regression parameters and the baseline hazard function.[16]

In our implementation, we follow Jenkins (1995) in combining discrete mixture heterogeneity with the grouped duration proportional hazards (complementary log-log) model (Cox, 1972; Prentice and Gloeckler, 1978):

$$\ln\left[-\ln\left\{1-h_t\left(X=x_{lj},Z=z,U=u\right)\right\}\right]=\gamma_{t,x_{lj}}+\beta_{Z,t}^{T}.z_t+u,$$

where the time intervals are indexed by $t\ (=1,2,\ldots)$, h_t denotes the discrete hazard rate in interval t conditional on $X=x_{lj}$, $Z=z$ and $U=u$, and $\gamma_{t,x_{lj}}$ denotes the baseline hazard rate conditional on $X=x_{lj}$. The model can be estimated using parametric maximum likelihood, for each chosen covariate value $X=x_{lj}$, to obtain the estimates $\widehat{\gamma}_{t,x_{lj}}$, $\widehat{\beta}_{Z,t}$, $\widehat{s}$, $\{m_1=0,\widehat{m}_2,\ldots,\widehat{m}_s\}$ and $\{\widehat{\pi}_1,\widehat{\pi}_2,\ldots,\pi_s=1-\widehat{\pi}_1-\widehat{\pi}_2-\ldots-\widehat{\pi}_{s-1}\}$.

For our tests, the above estimation methodology is combined with a choice of covariate pairs as before. The data generating process and assumptions are as follows.

The discrete duration data $(T_i,\delta_i,X_i,Z_i\left(t\right),U_i)$ are independently and identically sampled from the above complementary log-log model with discrete mixture frailties, for $i=1,\ldots,n$. The following assumptions hold.

Assumption 10 *The cut-off duration $0<T<\infty$ is large but finite, and subject to the condition that, for each $x=x_{lj},j=1,\ldots,r,j=1,2$, and for each t, $t=1,2,\ldots,T$, the baseline hazard rate is positive: $\gamma_{t,x}>0$.*

Assumption 11 *For each $l\ (=1,2,\ldots,r)$, let $L_t(x_{l1},x_{l2})$ be a monotonic discrete time stochastic process with finite first and second moments for each $t=1,\ldots,T$.*

Assumption 12 *(Identifiability conditions)*

(a) *Unobserved heterogeneity U is independent of covariates Z and censoring. A tail restriction holds for the heterogeneity distribution. Further, for the test, we also assume independence between unobserved heterogeneity and the index covariate X.*

(b) *There is minimal variation in covariate effect for each covariate in Z. There is at least one covariate effect that spans the whole of the real line. There is no perfect multicollinearity amongst the covariates.*

(c) *Censoring is random, and independent of Z and X.*

[16]However, the method often suggests frailty distributions with only 2 or 3 support points even when the original is known to be a well dispersed continuous distribution. This could be because estimation of the frailty distribution is a very difficult problem, with well discussed convergence problems (Horowitz, 1999) .

Assumption 13 *(Identification of finite mixture heterogeneity distribution) The conditions, originally given by Lindsay (1983a,b), state that the density of the data at each mass point of the heterogeneity distribution is a bounded function of the regression parameters.*

Assumption 14 *Boundedness and right continuity of the baseline hazard function and the regression parameters.*

The above assumptions are less restrictive than the previous case, since estimation here is a finite dimensional parametric problem, for each candidate value of $s \geq 1$. However, like most other problems with mixture distributions, convergence is slow, whether one uses gradient based methods or the EM (Expectations-Maximisation) algorithm. Having obtained estimates under an appropriate model with time varying coefficients, the test statistics are constructed as before. The basic statistics are

$$\begin{aligned} T_{HS,std}(x_{l1}, x_{l2}) &= \frac{T_{HS}(x_{l1}, x_{l2})}{\sqrt{\widehat{\text{Var}}\left[T_{HS}(x_{l1}, x_{l2})\right]}}, \\ T_{HS}(x_{l1}, x_{l2}) &= \sum_{t=1}^{T} L_t(x_{l1}, x_{l2}).\left[\widehat{\gamma}_{t,x_{l1}} - \widehat{\gamma}_{t,x_{l2}}\right], \qquad (23) \\ \widehat{\text{Var}}\left[T_{HS}(x_{l1}, x_{l2})\right] &= \sum_{t=1}^{T}\sum_{s=1}^{T}\left[\widehat{\gamma}_{t,x_{l1}} - \widehat{\gamma}_{t,x_{l2}}\right].\left[\widehat{\gamma}_{s,x_{l1}} - \widehat{\gamma}_{s,x_{l2}}\right].\widehat{\sigma}^2_{s\wedge t}(x_{l1}, x_{l2}), \end{aligned}$$

where $\widehat{\sigma}^2_{s\wedge t}(x_{l1}, x_{l2})$ is the (pointwise) sample variance of the weight process $L_t(x_{l1}, x_{l2})$. As in (19 – 21), these basic statistics are combined to construct our test statistics $T_{HS}^{(\max)}, T_{HS}^{(\min)}$ and $\overline{T}_{HS}$.

Then, we have the following asymptotic results.

Theorem 3. *Let Assumptions 10 through 14 hold. Then, under* $\mathbb{H}_{0,PH}$: $\gamma_{t,x} = c_t$ for all x, *as* $n \to \infty$,

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

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

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

(Proof in Appendix 1).

Corollary 4.

$$\begin{aligned} P\left[a_r\left\{T_{HS}^{(\max)} - b_r\right\} \leq z^*\right] &\to \exp\left[-\exp(-z^*)\right] \text{ as } r \to \infty \\ \textit{and } P\left[a_r\left\{T_{HS}^{(\min)} + b_r\right\} \geq z^*\right] &\to \exp\left[-\exp(z^*)\right] \text{ as } r \to \infty, \end{aligned}$$

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 1).

Remark 12. As in the continuous unobserved heterogeneity case, covariate dependent weighted tests can also be employed. Details are omitted here.

This completes our description of the proposed tests. A final point to note is that, the discrete mixture unobserved heterogeneity can also be used to model the heterogeneity distribution in the continuous time MPH model. This approach may have some advantages both in ease of implementation and computational effort. Similarly, the method based on maximum rank correlations, recently proposed by Hausman and Woutersen (2005), may be useful in the discrete duration setting, particularly if we are not as such interested in estimating the unobserved heterogeneity distribution. We have not pursued either of these approaches here.

3.4 Choice of weight functions

As emphasized earlier, the form of the null hypothesis in the two testing problems considered here are remarkably similar, and so are the test statistics proposed in Sections 3.2 and 3.3. However, the nature of departures from the null hypothesis that we are interested in is different for the two problems. Further, the choice of weight functions for the tests is left unspecified, and will depend on the type of violations expected in either case.

In our tests for absence of covariate dependence, the relevant null hypothesis is (8) and the alternatives of special interest are either trended (13) or changepoint trended (14). For the corresponding two sample tests, the logrank weight function, given by $L(t) = Y_1(t).Y_2(t)$, is optimal for proportional hazards alternatives; see, for example, Gill and Schumacher (1987) and Andersen *et al.* (1993). The proportional hazards model describes in a natural and intuitive way the notion of trend, as represented in the alternative hypothesis (13). However, a one-sided score test for $\beta_X = 0$ under the null hypothesis may be too restrictive, as demonstrated in an application considered later. Further, the log rank weight function is also useful for a changepoint trend alternative of the kind (14), because both positive and negative trends are evident on different regions of the sample space. In other words, the log rank weight function is appropriate for the proposed test for absence of covariate dependence if the suspected alternative is of a PH nature. The Gehan-Breslow weight function (Breslow, 1970), given by $L(t) = Y_1(t).Y_2(t).[Y_1(t) + Y_2(t)]$, may also be useful if censoring is high. Compared to the logrank test, this weight function places higher weight on differences in the hazard function at shorter durations (Andersen *et al.*, 1993).

By contrast, the two sample Peto-Prentice generalisation of the Wilcoxon test (Peto and Peto, 1972; Prentice, 1978) is optimal for a time-transformed logistic location family (Andersen *et al.*, 1993), and has higher power against alternatives with hazard ratio ordering (convex or concave ordering); see Prentice (1978) and Gill and Schumacher (1987) for further discussion. The above weight function is given by $L(t) = Y_1(t).Y_2(t)\left[Y_1(t) + Y_2(t)\right]^{-1}\widehat{S}(t)$, where $\widehat{S}(t)$ is a predictable analogue for the Kaplan Meier estimator. Our interest here is in tests for proportional hazards against order restricted covariate dependence, where the two sample representation of order restrictions $IHRCC$ and $DHRCC$ is described by convex or concave ordering of the two duration distributions. Hence, the Prentice weight function is appropriate for testing proportionality against these ordered alternatives.[17]

4 Simulation study

The asymptotic distributions of the proposed test statistics were derived in Section 3. Here, we report results of a two simulation studies exploring the performance of the proposed tests for absence of covariate effect and proportional hazards respectively, with respect to a continuous covariate.

For absence of covariate dependence, 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 represent different shapes of the baseline hazard function and patterns of covariate dependence. In all cases, the null hypothesis of absence of covariate dependence, $\mathbb{H}_{0,Eq}$ (8), holds if and only if $\beta(t,x) = 0$. If, for fixed x, $\beta(t,x)$ increases (or decreases) in x, we have trended alternatives of the type $\mathbb{H}^{(t)}_{1,Eq}$ (13). 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 $\mathbb{H}^{(c)}_{1,Eq}$ (14).

The tests discussed in Section 3.2 are consistent against the global alternative $\mathbb{H}_{1,Eq}$ (9), but are also expected to be powerful against the above kinds of specific alternatives to the null hypothesis. Specifically, we consider 2 different specifications of the baseline hazard function in combination with 3 patterns of covariate dependence. The Monte Carlo simulations are based on independent right-censored data from the following 6 data generating processes described in Table 1.

[17]Note that, because of unobserved heterogeneity, a martingale based framework is not applicable here and predictability is therefore not relevant. The cadlag nature of $\widehat{S}(t)$ makes the weight function itself cadlag, which is required for our tests.

TABLE 1: Data Generating Processes
(Test for absence of covariate dependence)

Model	$\boldsymbol{\lambda}_0(\mathbf{t})$	$\boldsymbol{\beta}(\mathbf{t},\mathbf{x})$	**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 distributed as $Uniform(-1,1)$. The independent censoring variable C is 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 (8), 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 ($T_{2s}^{(\max)}, T_{2s}^{(\min)}$ and $\overline{T}_{2s}$, respectively). Table 2 presents simulation results for $1,000$ simulations from the above data generating processes with sample sizes of 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 changepoint trended, so that when a pair of points are drawn at random from the covariate space, only a quarter of these pairs reflect the increasing nature of covariate dependence, and another quarter reflects the decreasing trend. However, the results also demonstrate the strength of the maxima and minima test statistics ($T_{2s}^{(\max)}$ and $T_{2s}^{(\min)}$ respectively) 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 (1999) tests. For the models DGP_{22} and DGP_{23}, and sample size 200, the Tarone (1975) test had rejection rates at the 5% confidence level, of 73 and 7 per cent respectively. The corresponding figures for the test proposed by Liu and Tsai (1999) were 81 and 9 per cent respectively.

TABLE 2: Test for absence of covariate dependence
(Rejection Rates (%) at 5 % and 1 % Asymptotic Confidence 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

Next, we examine the performance of the tests for the proportional hazards assumption in the presence of unobserved heterogeneity (3) against ordered alternatives of the IHRCC type (4, 5). The design of the data generating process is a combination of Horowitz (1999) and Bhattacharjee (2008). Samples are generated from the model

$$\lambda\left(t|x,z,u\right)=\lambda_0(t).\exp\left[-\left(\beta_X(t).x+\beta_Z.z+u\right)\right],$$

with two scalar covariates X and Z, and independent unobserved heterogeneity U. The covariate Z has proportional hazards effect, $\beta_Z=1$, while X has potentially time varying coefficients. In the experiments, $Z\sim N(0,1)$ while X has a right censored normal distribution with mean zero, variance 0.25 and censoring point 1.9.[18] We consider a single specification of the baseline hazard function as

$$\lambda_0(t)=0.087t,$$

[18]The censoring addressed a discontinuity in the inverse of the distribution function at $x=2$, and makes simulations easier; this adjustment should not affect our results.

and 2 different patterns of covariate dependence

$$\beta_X(t) = \left\{ \begin{array}{c} 1 \\ \ln(t) \end{array} \right. ,$$

in combination with 2 heterogeneity distributions. One unobserved heterogeneity distribution is continuous and defined by the distribution function

$$F(u) = \exp\left[-\exp\left(-u\right)\right],$$

so that $\exp\left(-U\right)$ has the unit exponential distribution, while the other is a discrete mixture with masspoints at 0.48 and 0.64, and corresponding probabilities 0.6 and 0.4. The simulated duration data are right censored by independent censoring times distributed as $Uniform\left(0.5, 25.5\right)$.

Therefore, these Monte Carlo simulations are based on independent right-censored data from 4 data generating processes (DGPs), defined by combinations of 2 specifications of the regression function and 2 specifications of the heterogeneity distribution. The description of the DGPs and expected results are summarised in Table 3. The two DGPs with $\beta_X(t) = 1$ belong to the null hypothesis of proportional hazards, while the other two, with $\beta_X(t) = \ln(t)$, are of the $IHRCC$ type. There is substantial censoring, around 25 per cent, in each of the four models.

TABLE 3: Data Generating Processes
(Test for proportional hazards, with unobserved heterogeneity)

Model	$\boldsymbol{\beta_X(t)}$	**Heterogeneity**	**Median cens.dur.**	% **cens.**	**Expd. significance**
DGP_{31}	1	Continuous	5.23	23.4	None
DGP_{32}	$\ln(t)$	Continuous	5.37	25.8	$T_{HS}^{(\max)}, \overline{T}_{HS}$
DGP_{41}	1	Mixture	5.16	23.6	None
DGP_{42}	$\ln(t)$	Mixture	5.37	25.4	$T_{HS}^{(\max)}, \overline{T}_{HS}$

For constructing the test statistics, we divide the sample into deciles by the value of the covariate X. The 45 pairwise combinations of these 10 deciles are used to construct the maxima, minima and average tests.

However, implementing the test procedures for continuous unrestricted unobserved heterogeneity using the Horowitz (1999) estimator turned out to be very challenging. The main problem was finding appropriate bandwidths and tuning parameters in a consistent manner to make the Monte Carlo useful.[19] Horowitz (1999) suggests the use of cross-validation or bootstrap

[19]The critical issue is that estimation of the scale parameter is a difficult problem. Further, attempts to estimate this parameter well compromises the baseline hazard estimate, which is the main input for our tests.

for this purpose. Using cross-validation, we could implement the method fairly well for individual samples, but not consistently over repeated runs of the Monte Carlo experiment. How far the bootstrap procedures suggested in Kosorok *et al.* (2004) are useful remains a research question. On the positive side, our study shows that, using cross-validation, the method can be easily implemented in individual applications.

TABLE 4: Test for Proportional Hazards, with Unobs. Het. (Rejection Rates (%) at 5 % and 1 % Asymptotic Confidence Levels)

Model	**Test statistic**	**Sample size, Confidence level**	
		10000, 5%	**10000, 1%**
DGP_{31}	$T_{HS}^{(\max)}$	8.5	0.5
	$T_{HS}^{(\min)}$	3.0	1.0
	$\overline{T}_{HS}$	3.5	2.0
DGP_{32}	$T_{HS}^{(\max)}$	91.0	61.5
	$T_{HS}^{(\min)}$	1.5	0.0
	$\overline{T}_{HS}$	100.0	100.0
DGP_{41}	$T_{HS}^{(\max)}$	7.5	0.5
	$T_{HS}^{(\min)}$	3.5	1.5
	$\overline{T}_{HS}$	5.5	1.5
DGP_{42}	$T_{HS}^{(\max)}$	96.5	69.0
	$T_{HS}^{(\min)}$	2.0	0.0
	$\overline{T}_{HS}$	100.0	99.5

Implementing the Heckman and Singer (1984a) method was relatively straightforward. For this purpose, we transformed our data into grouped data form by censoring over unit intervals. As noted in the literature (see, for example, Jenkins, 1995), the maximum likelihood procedure had convergence problems. Making use of multiple starting values, different candidate maximisation algorithms, and by adjusting tolerance levels on the Hessian, we were able to implement the procedure with sample sizes upwards of 1000.[20] The results presented in Table 4 are based on a larger sample size of $10,000$, which was convenient for working with repeated Monte Carlo samples. Our exercise also suggests that it may be useful to use the entire data to estimate the unobserved heterogeneity distribution, while using data for each decile to estimate the baseline hazard function; we have not investigated this approach further.

Considering the challenges noted above, and slow convergence of the maximum likelihood procedure, we report results based on a modest 200 Monte

[20] With a sample size of 1000, each decile has only 100 data points, which makes estimation of the heterogeneity distribution quite challenging.

Carlo replications for each of the four DGPs. The performance of the tests is encouraging, in that nominal sizes are approximately maintained, and power is very good.

Overall, our Monte Carlo study confirms the usefulness of the proposed tests for both the testing problems considered. In the next Section, we apply our methods to real data.

5 An application

Now, we illustrate the use of the proposed tests through an application based on real data. The objective is to study the effect of aggregate Q on the hazard rate of corporate failure in the UK. The data on firm exits through bankruptcy, over the period 1980 through 1998, pertain to 2789 listed manufacturing companies, covering 24,034 company years and include 95 bankruptcies. The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1980, and contain staggered entries. The focus of our analysis here is on the impact of aggregate Q on corporate failure. Following usual practice, we consider the reciprocal of Q as the continuous covariate in our regression model.[21]

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 (exponential of the regression coefficient) of 0.92, with p-value 0.156 per cent. Taking this evidence on face value, one might therefore be inclined to believe that covariate dependence is absent. However, such lack of evidence for the covariate effect could also result from model misspecification. This possibility suggests that we could take a more nonparametric approach that does not assume *a priori* the structure of the regression model.

Descriptive graphical tests based on counting processes conditional on several pairs of covariate values indicate significant trend in the hazard functions. Since our tests of absence of covariate dependence are powerful against trended alternatives, we apply the tests to these data (Table 5). Each of the tests were based on 20 pairs of distinct covariate values, drawn at random from the marginal distribution of the covariate. 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, $\mathbb{H}_1^* : \lambda(t|x_1) \leq \lambda(t|x_2)$ for all $x_1 > x_2$ (with strict inequality holding for some

[21] Fore more discussion and analyses based on these data, see Bhattacharjee *et al.* (2008).

$x_1 > x_2$). This implies that, contrary to estimates of a standard Cox regression model, higher aggregate Q significantly depresses the hazard of business exit due to bankruptcy.

TABLE 5: 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

Further, the maxima and minima test statistics provide additional information on the covariate pairs for which the basic test statistics assume their extreme values, which may be useful for investigating the nature of departures from proportionality.[22] For example, the significant test-statistics $T_{2s}^{(\min)}$ are attained for the covariate pairs $\{-0.058, 0.116\}$ (7th and 63rd percentile) for the logrank weight function (and $\{-0.017, 0.098\}$ (10th and 50th percentile) for the Gehan-Breslow weight function). This provides further evidence of trend.

TABLE 6: Time Varying Coefficients Model
(Estimates based on UK Corporate Bankruptcy Data)

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

To explore whether this apparent trend in conditional hazard functions was masked in the Cox regression model (and the score test) by lack of proportionality, we present in Table 6 a time varying coefficient model for the same data estimated using the histogram sieve estimators (Murphy and Sen, 1991). The results confirm the presence of trend, particularly at higher durations.

However, the above inference can be misleading because of model misspecification, particularly in the form of omitted covariates. In fact, the estimated empirical model, with a single covariate, is rather simplistic and it is quite likely that unobserved heterogeneity is present in these data. Therefore, we

[22] This is in line with the way in which Bhattacharjee (2008) approximately locates changepoints.

include firm size (measured by logarithm, of fixed assets divided by 10 and incremented by one), as an additional covariate and apply the proposed tests for proportional hazards allowing for unrestricted unobserved heterogeneity. The measure of size considered assumes both positive and negative values, and is expected to be an important firm level covariate. We allow size to have age varying coefficients, model unobserved heterogeneity using the Heckman and Singer (1984a) procedure, and estimate grouped duration proportional hazards models conditional on various values of the covariate under test, Q. As expected, there is significant unobserved heterogeneity in the data. Further, the tests for proportional hazards, based on 20 randomly chosen covariate pairs, reject the null hypothesis in favour of a $DHRCC$ alternative (11). Both $T_{HS}^{(\min)}$ and $\overline{T}_{HS}$ are significant, at the 5 per cent and 1 per cent levels of significance respectively.

The above application demonstrates the use of the proposed test statistics. The first set of 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. Similarly, the proposed tests for proportional hazards are powerful against ordered covariate effects, in the presence of arbitrary unobserved heterogeneity. These tests are useful not only for detecting violation of the proportional hazards assumption, but also for understanding the nature of departures from proportionality and for subsequent modeling.

6 Conclusion

In summary, the tests described in this paper add important tools to the armoury of a duration data analyst. Our work extends an important class of two sample tests for equality of hazards to a continuous covariate framework, both for discrete and continuous duration data, with and without the presence of unobserved heterogeneity.

The proposed tests for absence of covariate effect are powerful against trended and changepoint trended alternatives. Hence, they allow more precise inferences on the direction of covariate effects. Perhaps most importantly, the methods do not make any strong assumptions regarding the underlying regression model, and thereby provide robust inference. Using simulated data and a real life application, the strength of the tests is demonstrated and more specific inferences are derived regarding the nature of covariate dependence.

Further, our main contribution here is in extending tests for proportionality with respect to a continuous covariate against ordered alternatives in

the presence of individual level unobserved heterogeneity with unrestricted distribution. Here, counting process arguments do not hold, but we use empirical process theory to extend standard two sample tests to this setup. In conjunction with Bhattacharjee (2008), this work therefore extends many of the two sample tests to the continuous covariate setup, and thereby makes these tests more readily usable in real life applications.

Our test statistics for proportional hazards have the form

$$\int K^{(n)}(t).H^{(n)}(t).dt, \tag{24}$$

where $K^{(n)}(.)$ and $H^{(n)}(.)$ are stochastic processes involving data from all the n observations, $K^{(n)}(.)$ is manageable and $H^{(n)}(.)$ has a continuous probability limit. We show how modern empirical process theory (Pollard, 1990) in combination with Sengupta *et al.* (1998) can be used to derive asymptotic theory for the statistics like (24). By contrast, for testing absence of covariate dependence, we used statistics like

$$\sum_{i=1}^{n} \int K^{(n)}(t).dM_i(t), \tag{25}$$

which are standard in the analysis of failure time data based on counting processes. Asymptotic results typically follow from counting process theory (Andersen *et al.*, 1993), under the conditions that $M_i(.)$ are martingales (or local martingales) and $K^{(n)}(.)$ is a predictable process.

Several areas of further research emerge from our work. First, the development of asymptotic arguments for statistics like (24) is useful in contexts well beyond the current application.[23] For example, in the context of unobserved heterogeneity models, one can think of alternate statistics constructed by plugging-in the estimated heterogeneity distribution in the counting process martingale. Exploration of these and other applications is beyond our current scope. Second, the proposed methods raise important new research questions relating to inference on the changepoint in hazard regression models, and on effective and efficient ways to conduct joint inference on several continuous covariates. These problems will be retained for future work. Third, we consider the possibility of new inference tools for duration data, where inferences on duration dependence over subsamples can be potentially combined

[23] In fact, the tests proposed here do not fully use the strengths of this methodology. While monotonicity of $K_i(.)$ simplifies our arguments, the condition required is that the process has a finite pseudodimension (Pollard, 1990). Similarly, the main condition required of $H(t)$ is that it has a continuous probability limit.

with pooled inference for unoberved heterogeneity across the whole sample. Development of such methods will be useful extensions of the current work. Fourth, the proposed tests for proportional hazards in the presence of unobserved heterogeneity, together with the application considered here, further emphasize the importance of considering such heterogeneity together with monotonic covariate effects in empirical studies. Joint modeling of non-proportional covariate effects and unrestricted unobserved heterogeneity is therefore important. Finally, our work highlights the limitations of currently available approaches for inference on unrestricted univariate unobserved heterogeneity in duration models. This is currently an active research area, and further developments will emerge in coming years.

Appendix 1: Proofs of the Results

Proof of Theorem 1

It follows from standard counting process arguments (see, for example, Andersen *et al.*, 1993) that, under $H_{0,Eq}$ (8), for $l = 1, \ldots, r$,

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

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

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

Since $\widehat{\text{Var}}\left[T_{2s}\left(x_{l1}, x_{l2}\right)\right]$ is a consistent estimator for the variance of $T_{2s}\left(x_{l1}, x_{l2}\right)$ (Gill and Schumacher, 1987; Andersen *et al.*, 1993), we have as $n \longrightarrow \infty$,

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

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.*, 1993), 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.

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 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:

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

where I_r is the identity matrix of order r. The proof follows straightaway.

□

Proof of Theorem 2

Recall that our basic statistics, conditional on the covariate pair (x_{l1}, x_{l2}), are

$$T_H(x_{l1},x_{l2}) = \int_0^{\tau^*} L(x_{l1},x_{l2})(t).\widehat{\lambda}_{0,H}\left(t,x_{l1}\right).dt - \int_0^{\tau^*} L(x_{l1},x_{l2})(t).\widehat{\lambda}_{0,H}\left(t,x_{l2}\right).dt.$$

We first show that the above statistic converges weakly to a mean zero normal distribution under the null hypothesis, then show that the variance estimator is consistent, so that the standardised statistic is asymptotically standard normal, and finally that the statistics are aysmptotically independent for different pairs of covariate values. The proof then follows from Theorem 1 above.

For proving weak convergence of the basic statistic, we make use of Theorem 3.1 (Sengupta *et al.*, 1998). In order to study the convergence of $T_i, i = 1, 2$, we

replace $K_n(t)$ and $X_n(t)$ in the above theorem by $\left[\widehat{\lambda}_{0,H}(t, x_{l1}) : \widehat{\lambda}_{0,H}(t, x_{l2})\right]^T$ and $[L(x_{l1}, x_{l2})(t)]$, respectively.

It follows from Horowitz (1999) (Corollary 1.1) that

$$\begin{pmatrix} \widehat{\lambda}_{0,H}(t, x_{l1}) \\ \widehat{\lambda}_{0,H}(t, x_{l2}) \end{pmatrix} \xrightarrow{P} \begin{pmatrix} \lambda_{0,H}(t, x_{l1}) \\ \lambda_{0,H}(t, x_{l2}) \end{pmatrix},$$

for $t \in [0, \tau^*]$, and by our assumptions, $\lambda_{0,H}(t, x_{lj})$ are continuous functions on $[0, \tau^*]$.

Now, by our assumption, the weight function $L(x_{l1}, x_{l2})(t)$ is monotone. Since monotone functions have pseudodimension 1, the process $L(x_{l1}, x_{l2})(t)$ is manageable (Pollard, 1990; Bilias *et al.*, 1997). It then follows from the functional central limit theorem (Pollard, 1990) that $L(x_{l1}, x_{l2})(t)$ converges weakly to a Gaussian process.

Now, by applying Theorem 3.1 of Sengupta *et al.* (1998), we have

$$\begin{aligned} &\begin{pmatrix} \int_0^{\tau^*} n^{-1/2} \left[L(x_{l1}, x_{l2})(t) - l(x_{l1}, x_{l2})(t)\right] . \widehat{\lambda}_{0,H}(t, x_{l1}) . dt \\ \int_0^{\tau^*} n^{-1/2} \left[L(x_{l1}, x_{l2})(t) - l(x_{l1}, x_{l2})(t)\right] . \widehat{\lambda}_{0,H}(t, x_{l2}) . dt \end{pmatrix} \\ &\xrightarrow{D} \begin{pmatrix} \int_0^{\tau^*} \lambda_{0,H}(t, x_{l1}) . W(x_{l1}, x_{l2})(t).dt \\ \int_0^{\tau^*} \lambda_{0,H}(t, x_{l2}) . W(x_{l1}, x_{l2})(t).dt \end{pmatrix}, \end{aligned}$$

where $l(x_{l1}, x_{l2})(t)$ is the asymptotic mean process corresponding to $L(x_{l1}, x_{l2})(t)$, and $W(x_{l1}, x_{l2})(t)$ is a Gaussian process. It follows that

$$\begin{aligned} &\int_0^{\tau^*} n^{-1/2} \left[L(x_{l1}, x_{l2})(t) - l(x_{l1}, x_{l2})(t)\right] . \left[\widehat{\lambda}_{0,H}(t, x_{l1}) - \widehat{\lambda}_{0,H}(t, x_{l2})\right] . dt \\ &\xrightarrow{D} \int_0^{\tau^*} \left[\lambda_{0,H}(t, x_{l1}) - \lambda_{0,H}(t, x_{l2})\right] . W(x_{l1}, x_{l2})(t).dt. \end{aligned}$$

This completes the first part of the proof.

The above limiting distribution is Gaussian with mean zero, and variance

$$\int_0^{\tau^*} \int_0^{\tau^*} c(t).c(s).V(s \wedge t) . ds dt,$$

where

$$c(t) = \left[\lambda_{0,H}(t, x_{l1}) - \lambda_{0,H}(t, x_{l2})\right]$$

and $V(.)$ is the variance process of the limiting distribution of

$$n^{-1/2} \left[L(x_{l1}, x_{l2})(t) - l(x_{l1}, x_{l2})(t)\right].$$

Since, conditional on the covariate pair (x_{l1}, x_{l2}), $c(t)$ is consistently estimated by

$$\left[\widehat{\lambda}_{0,H}\left(t, x_{l1}\right) - \widehat{\lambda}_{0,H}\left(t, x_{l2}\right)\right],$$

and $V(t)$ is estimated consistently (pointwise) by the sample variance of $L(x_{l1}, x_{l2})(t)$, $\widehat{\text{Var}}\left[T_H(x_{l1}, x_{l2})\right]$ is a consistent estimator of the variance of $T_H(x_{l1}, x_{l2})$.

Since $\widehat{\text{Var}}\left[T_H\left(x_{l1}, x_{l2}\right)\right]$ is a consistent estimator for the variance of $T_H\left(x_{l1}, x_{l2}\right)$, we have as $n \longrightarrow \infty$,

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

The proof of the Theorem will now follow, if it further holds that $T_{H,std}\left(x_{l1}, x_{l2}\right)$, $l = 1, \ldots, r$ are asymptotically independent. This follows because sampling is independent for the counting processes $N\left(t, x_{lj}\right)$ conditional on different covariate values x_{lj} $(l = 1, \ldots, r; j = 1, 2)$.

It follows that

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

where I_r is the identity matrix of order r. Proofs of (a), (b) and (c) follow.

□

Proof of Corollary 3

Proof follows exactly in the same way as Corollary 1.

□

Proof of Theorem 3

With discrete data, the problem is finite dimensional, and therefore the proofs are simpler. Our basic statistics, conditional on the covariate pair (x_{l1}, x_{l2}), are

$$T_{HS}(x_{l1}, x_{l2}) = \sum_{t=1}^{T} L_t(x_{l1}, x_{l2}).\left[\widehat{\gamma}_{t,x_{l1}} - \widehat{\gamma}_{t,x_{l2}}\right]$$

We follow a similar approach to the proof of Theorem 1, first showing that the above statistic converges weakly to a mean zero normal distribution under the null hypothesis, then showing that the variance estimator is consistent, so that the standardised statistic is asymptotically standard normal, and finally that the statistics are aysmptotically independent for different pairs of covariate values. The proof then follows from Theorem 1.

Since $T_{HS}(x_{l1}, x_{l2})$ is a finite linear combination of statistics like $L_t(x_{l1}, x_{l2}).\widehat{\gamma}_{t,x_{lj}}$ $(t = 1, \ldots, T;\ l = 1, \ldots, r;\ j = 1, 2)$, weak convergence of the basic statistic follows from weak convergence of a vector comprising all the above statistics to the multivariate normal distribution.

Arguing as in Theorem 2, monotonicity of the weight function $L_t(x_{l1}, x_{l2})$ implies it has pseudodimension 1, and therefore the process $L_t(x_{l1}, x_{l2})$ is manageable (Pollard, 1990; Bilias *et al.*, 1997). It then follows from the functional central limit theorem (Pollard, 1990) that $L_t(x_{l1}, x_{l2})$ converges weakly to a Gaussian process.

Further, $\widehat{\gamma}_{t,x_{lj}}$ are consistent estimators of the corresponding parameters $\gamma_{t,x_{lj}}$, implying that $\widehat{\gamma}_{t,x_{lj}} \xrightarrow{P} \gamma_{t,x_{lj}}$. Weak convergence of $T_{HS}(x_{l1}, x_{l2})$ to a mean zero Gaussian distribution now follows by application of Slutsky's theorem, continuous mapping theorem and the multivariate central limit theorem.

Next, the variance of the limiting distribution is given by

$$\sum_{t=1}^{T}\sum_{s=1}^{T}\left[\gamma_{t,x_{l1}} - \gamma_{t,x_{l2}}\right] \cdot \left[\gamma_{s,x_{l1}} - \gamma_{s,x_{l2}}\right] .\sigma^2_{s\wedge t}(x_{l1}, x_{l2}),$$

where $\sigma^2_t(x_{l1}, x_{l2})$ is the variance process of the limiting distribution of $L_t(x_{l1}, x_{l2})$. Since, conditional on the covariate pair (x_{l1}, x_{l2}), $\left[\gamma_{t,x_{l1}} - \gamma_{t,x_{l2}}\right]$ is consistently estimated by $\left[\widehat{\gamma}_{t,x_{l1}} - \widehat{\gamma}_{t,x_{l2}}\right]$, and $\sigma^2_t(x_{l1}, x_{l2})$ is estimated consistently by the sample variance of $L_t(x_{l1}, x_{l2})$, $\widehat{\mathrm{Var}}\left[T_{HS}(x_{l1}, x_{l2})\right]$ is a consistent estimator of the variance of $T_{HS}(x_{l1}, x_{l2})$. Therefore, as $n \longrightarrow \infty$,

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

The proof of the Theorem would follow, if it further holds that $T_{HS,std}\left(x_{l1}, x_{l2}\right)$, $l = 1, \ldots, r$ are asymptotically independent. This follows because sampling is independent for the counting processes $N\left(t, x_{lj}\right)$ conditional on different covariate values x_{lj} $(l = 1, \ldots, r; j = 1, 2)$. Therefore, we have

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

where I_r is the identity matrix of order r. Proofs of (a), (b) and (c) follow.

□

Proof of Corollary 4

Proof follows exactly in the same way as the proof of Corollary 1.

□

Appendix 2: Test based on Gørgens and Horowitz (1999)

The estimator proposed by Gørgens and Horowitz (1999) is an extension of Horowitz (1996) to include censoring. It is valid for the more general transformation model and imposes the scale normalisation restricting one of the regression coefficients to be unity (positive or negative). Like the estimator in Horowitz (1999), this estimator too cannot directly accomodate time varying covariates. However, an attractive feature of this approach is that the estimator for the baseline cumulative hazard function converges to a Gaussian process with a consistent estimator for the covariance function.

For our purpose, we adjust the Gørgens and Horowitz (1999) estimator in the following way. First, we assume that the effect of the other covariates Z has been modeled *a priori* and an appropriate MNPH model (2) has been found. Next, we adjust the model in a way suitable for our test. Specifically, what we require are estimators of the processes

$$\int_0^\tau L(x_{l1}, x_{l2})(t).\lambda_0(t, x_{lj}).dt, \quad j = 1, 2,$$

where $L(x_{l1}, x_{l2})(t)$ is the random weight function corresponding to the covariate pair (x_{l1}, x_{l2}). Now, $\int_0^\tau L(x_{l1}, x_{l2})(t).\lambda_0(t, x_{lj})$ is the cumulative baseline hazard function in the modified regression model

$$\begin{aligned} \lambda^*(t|X = x_{lj}, z, u) &= [L(x_{l1}, x_{l2})(t).\lambda_0(t, x_{lj})] \\ &\quad .\exp\left[-\ln L(x_{l1}, x_{l2})(t) + \beta_Z(t)^T.z(t) + u\right], \end{aligned}$$

where $\ln L(x_{l1}, x_{l2})(t)$ is an additional time varying covariate. This model can now be estimated using the Gørgens and Horowitz (1999) estimator. An attractive feature of this procedure is that the scale normalisation is automatically satisfied, since the new covariate $\ln L(x_{l1}, x_{l2})(t)$ has a regression coefficient -1.

Note that since the MNPH model with time varying covariates is not a transformation model, the estimation method does not directly allow for time varying covariates. But, in the case that the corresponding coefficient is fixed, this can be addressed by substituting the covariate value by an average over the duration of the time varying covariate. This procedure can be followed for the additional covariate above, by substituting for it the average value $\int_0^\tau \tau^{-1} \ln L(x_{l1}, x_{l2})(t).dt$.

Likewise, time varying coefficients for other covariates are included in the model using histogram sieves, by interacting each such covariate with various duration intervals, and treating the resulting covariates as constant.

We denote the resulting estimator for the baseline cumulative hazard function, conditional on a given value for the index covariate, $X = x$, by $\widehat{\Lambda}_{GH,L(x_{l1},x_{l2})}\left(t, x_{lj}, \widehat{\beta}_Z\right)$.

Similar assumptions are required here as the above method using the Horowitz (1999) estimator, with the following modifications:

Assumption 7a *(Identifiability conditions)*

(a) In addition to covariates and censoring, unobserved heterogeneity U is independent of the weight function $L(x_{l1}, x_{l2})(t)$.

(b) The effect of one of the covariates, in our case $L(x_{l1}, x_{l2})(t)$, is scaled to ± 1 (scale normalisation).

(d) Censoring is independent of Z, and possibly depends on X, but only through the weight function $L(x_{l1}, x_{l2})(t)$.

Some qualifying comments are required for our implementation of the Gørgens and Horowitz (1999) estimator. First, dependence between unobserved heterogeneity and the weight function is a strong assumption in our case. We take the view that the relevant component of unobserved heterogeneity here is its projection onto the orthogonal space of the covariates and the weight function. This is in line with interpretation of unobserved heterogeneity as the effect of omitted covariates. Second, Gørgens and Horowitz (1999) allow censoring to depend on the covariates through the single index, which in our case is $-\ln L(x_{l1}, x_{l2})(t) + \beta_Z(t)^T.z(t)$. We assume independent censoring. However, since the weight function itself may depend on the censoring pattern, we allow censoring to depend on X, but only through the weight function. Third, as discussed above, the scale normalisation has a natural interpretation in our case, since the weight function has a regression coefficient of -1. Fourth, like the Horowitz (1999) procedure, appropriate choice of bandwidths and tuning parameters is difficult, and a potential limitation of this approach. Finally, while the test statistic is obtained quite easily using the above procedure, variance estimation is more difficult. For this purpose, we suggest the weighted and nonparametric bootstrap developed in Kosorok *et al.* (2004), which are valid under a wide class of continuous unobserved heterogeneity distributions.

References

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

[2] Baker, M. and Melino, A. (2000). Duration dependence and nonparametric heterogeneity: a Monte Carlo study. *Journal of Econometrics* **96**, 357–393.

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

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

[5] Bhattacharjee, A. (2008). Partial orders with respect to continuous covariates and tests for the proportional hazards model. Discussion Paper **0807**, Department of Economics, University of St. Andrews, UK (http://ideas.repec.org/p/san/wpecon/0807.html). Revise and resubmit: *Journal of Statistical Planning and Inference*.

[6] Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2008). Macroeconomic conditions and business exit: determinants of failures and acquisitions of UK firms. CDMA Working Paper **0713**, University of St Andrews, UK (http://ideas.repec.org/p/san/cdmawp/0713.html). Forthcoming: *Economica*.

[7] Bilias, Y., Gu, M. and Ying, Z. (1997). Towards a general asymptotic theory for the Cox model with staggered entry. *Annals of Statistics* **25**, 662–682.

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

[9] Breslow, N.E. (1974). Covariance analysis of censored survival data. *Biometrics* **30**, 89–99.

[10] Bretagnolle, J. and Huber-Carol, C. (1988). Effect of omitting covariates in Cox's model for survival data. *Scandinavian Journal of Statistics* **15**, 125–138.

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

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

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

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

[15] Gørgens, T. and Horowitz, J.L. (1999). Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable. *Journal of Econometrics* **90**, 155–191.

[16] Grenander, U. (1981). *Abstract Inference*. Wiley: New York.

[17] Han, A.K. (1987). Non-parametric analysis of a generalized regression model. *Journal of Econometrics* **35**, 303–316.

[18] Hausman, J.A. and Woutersen, T.M. (2005). Estimating a semiparametric duration model without specifying heterogeneity. CeMMAP Working Paper **CWP11/05**, Institute of Fiscal Studies, University College London, London.

[19] Heckman, J.J. and Singer, B. (1984a). A method for minimising the impact of distributional assumptions in econometric models for duration data. *Econometrica* **52**, 271–320.

[20] Heckman, J.J. and Singer, B. (1984b). Econometric duration analysis. *Journal of Econometrics* **24**, 63–132.

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

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

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

[24] Ishwaran, H. (1996). Identifiability and rates of estimation for scale parameters in location mixture models. *Annals of Statistics* **24**, 1560–1571.

[25] Jenkins, S.P. (1995). Easy estimation methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics* **57**(1), 129–138.

[26] Jones, M.P. and Crowley, J.J. (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Annals of Statistics* **18**, 1203–1220.

[27] Keiding, N., Andersen, P.K. and Klein, J.P. (1997). The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* **16**, 215–224.

[28] 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.

[29] Lindsay, B.G. (1983a). The geometry of mixture likelihoods: a general theory. *Annals of Statistics* **11**(1), 86–94.

[30] Lindsay, B.G. (1983b). The geometry of mixture likelihoods, part II: the exponential family. *Annals of Statistics* **11**(3), 783–792.

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

[32] 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.

[33] McCall, B.P. (1996). The identifiability of the mixed proportional hazards model with time-varying coefficients. *Econometric Theory* **12**(4), 733–738.

[34] 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.

[35] Parner, E. (1998). Asymptotic theory for the correlated Gamma-frailty model. *Annals of Statistics* **26**, 183–214.

[36] 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.

[37] Pollard, D. (1990). *Empirical Processes: Theory and Applications*. Institute of Mathematical Statistics: Hayward.

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

[39] Prentice, R.L. and Gloeckler, L. (1978). Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* **34**, 57–67.

[40] 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.

[41] 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.

[42] Sherman, R.P., 1993. The limiting distribution of the maximum rank correlation estimator. *Econometrica* **61**, 123–137.

[43] 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.

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

[45] Zucker, D.M. and Karr, A.F. (1990). Nonparametric survival analysis with time-dependent covariate effects: A penalized partial likelihood approach. *Annals of Statistics* **18**, 329–353.