

University of Dundee

A Simple Test for the Absence of Covariate Dependence in Duration Models

Bhattacharjee, A.

Publication date:
2004

Document Version
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

Citation for published version (APA):
Bhattacharjee, A. (2004). *A Simple Test for the Absence of Covariate Dependence in Duration Models*. (Cambridge Working Papers in Economics; No. 0401). Faculty of Economics, University of Cambridge. <http://ideas.repec.org/p/cam/camdae/0401.html>

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.

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.

A Simple Test for the Absence of Covariate Dependence in Duration Models

Arnab Bhattacharjee

January 2004

CWPE 0401

Not to be quoted without permission

A SIMPLE TEST FOR THE ABSENCE OF COVARIATE DEPENDENCE IN DURATION MODELS

December 17, 2003

ARNAB BHATTACHARJEE¹
Department of Applied Economics,
University of Cambridge,
Sidgwick Avenue,
Cambridge CB3 9DE, UK.

Abstract

This paper describes a simple extension of popular tests of equality of hazard rates in a two-sample or k -sample setup to a situation where the covariate under study is continuous. In other words, we test the hypothesis $H_0 : \lambda(t|x) = c(t)$ for all x against the omnibus alternative $H_1 : \text{not } H_0$, where the covariate x is continuous. The tests developed are also useful in detecting trend in the underlying hazard rates (*i.e.*, when the alternative hypothesis is $H_1^* : \lambda(t|x_1) \geq \lambda(t|x_2)$ for all t whenever $x_1 > x_2$, the strict inequality holding for atleast one covariate pair $(x_1, x_2), x_1 > x_2$), or changepoint trend alternatives ($H_1^{**} : \text{there exists } x^* \text{ such that } \lambda(t|x) \uparrow x \text{ whenever } x < x^* \text{ and } \lambda(t|x) \downarrow x \text{ whenever } x > x^*$). Asymptotic distribution of the test statistics are established using counting process techniques. Small sample properties of the tests are studied, and the use of the tests in empirical applications is illustrated.

Key words: Covariate dependence; Continuous covariate; Two-sample tests; Trend tests.

1 Introduction

Understanding the nature of covariate dependence is one of the main objectives of regression analysis of duration data. The usual method of assessing the strength of covariate dependence is by conducting tests of the hypothesis $H_0 : \lambda(t|x) = c(t)$ for all x against different kinds of alternatives, the choice of the alternative depending on the nature of covariate dependence hypothesized. When the covariate is dichotomous or categorical, a test for absence of covariate effects against the omnibus alternative $H_1 : \text{not } H_0$ would be equivalent to testing that the hazard rates or survival functions in the two (or k) samples are the same. For this situation, there are several censored-data rank tests for the above hypothesis available in the literature. The Mantel-Haenszel or logrank test (Mantel, 1966; Peto and Peto, 1972) has been one of the most popular in empirical applications. This test has optimal power if the two compared groups have proportional hazard functions (Peto and Peto, 1972). The Gehan or

¹This work was supported in part by the KPMG. An earlier version of the paper was presented at the 58th European Meeting of the Econometric Society, Stockholm, August 2003, and the author thanks the participants in that conference for their valuable comments. The usual disclaimer applies.

Breslow test (Gehan, 1965; Breslow, 1970) and Prentice (1978) test generalised the Wilcoxon and Kruskal-Wallis tests to right censored data. Tarone and Ware (1977) and Harrington and Fleming (1982) have proposed weighted log-rank tests that are also frequently used in applications. The theoretical properties of these tests, as well as their performance in small samples, have been elaborately discussed elsewhere (Fleming and Harrington, 1991; Andersen *et al.*, 1992).

However, it is often important to know not only whether the covariate dependence is significant, but the direction of the covariate effect, *i.e.*, whether an increase in covariate value increases or decreases the duration stochastically (or, according to some other notion of ageing). In the k -sample setup, we have the so-called trend tests, *i.e.* tests for equality of hazards against the alternatives $H_1 : \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_k$ or $H_1 : S_1 \leq S_2 \leq \dots \leq S_k$ (with one or more of the inequalities being strict), where λ_j and S_j are the hazard and survival functions respectively in the j -th sample. Modified score tests that detect trend in hazard functions have been proposed by Tarone (1975) and Tarone and Ware (1977), while Liu *et al.* (1993) and Liu and Tsai (1999) have proposed ordered weighted logrank tests to detect similar trend in survival functions. Mau (1988) proposed trend tests for censored duration data by applying isotonic regression to scores from existing k -sample tests. However, these two-sample and k -sample tests are of limited use in econometric and biomedical applications where covariates typically are continuous (Horowitz and Neumann, 1992; Neumann, 1997). The usual way of extending these inference procedures to the case of continuous covariates is by stratifying the covariate, and then apply existing inference procedures for k samples. The outcome of these inference procedures is highly sensitive to the choice of such intervals, and relevant procedures for optimally choosing these intervals are not generally available (Horowitz and Neumann, 1992).

There are some other tests that may be more suitable to applications involving continuous covariates, where the equivalent alternative hypothesis is $H_1^* : \lambda(t|x_1) \geq \lambda(t|x_2)$ for all t whenever $x_1 > x_2$ (or vice versa). Under the assumption of an appropriate hazard regression model (like the Cox proportional hazards (PH) model or the accelerated failure time model), one can use score tests for the significance of the regression coefficient (Cox, 1972; Prentice, 1978). There are several other tests that assume a known covariate label function. Brown *et al.* (1974) derived a permutation test based on ranking of both the covariate values and the observed durations, and O'Brien (1978) has proposed inverse normal and logit rank tests using the respective transformations of the ranked covariates. Jones and Crowley (1989, 1990) considered a general class of test statistics based on a known covariate label function; this test nests most of the other trend tests as well as robust versions of these tests.

However, all of these available test procedures for trend with respect to continuous covariates suffer from the limitation that they assume either validity of a specified regression model, or a known covariate label function. Hence, these tests fail to retain the attractive nonparametric properties of the corresponding two-sample or k -sample tests, and are not useful in many situations. In particular, these tests would not be able to detect presence of covariate dependence in changepoint trend situations (*i.e.*, when the alternative is of the form $H_1^{**} : \text{there exists } x^* \text{ such that } \lambda(t|x) \uparrow x \text{ whenever } x < x^* \text{ and } \lambda(t|x) \downarrow x \text{ whenever } x > x^*$). Jespersen (1986) has proposed inference procedures in the context of a single changepoint regression model; however, the assumptions of a specified regression model and a single changepoint appear to be quite restrictive. Thus, appropriate tests for absence of covariate dependence for continuous covariates are not available in the literature, in applications where either the form of the regression relationship or an appropriate covariate label function cannot be assumed. Insignificance of the estimated parameter in a Cox regression model is of-

ten (inappropriately) interpreted as a test for covariate dependence. This is because, such insignificance can arise from many sources, including lack of proportionality, and model misspecification.²

This article develops tests for the absence of covariate dependence that are useful in detecting trend (and changepoint trend) with respect to a continuous covariates, by a simple extension of the tests available in the two-sample setup. This is achieved by conducting the usual two-sample tests conditional on several pairs of distinct covariate values, and using the maxima/ minima or average of these individual test statistics to combine the results. Similar techniques have been used in Bhattacharjee and Das (2002) to construct tests of proportionality of hazards with respect to continuous covariates. Section 2 elaborates on the construction of the test statistics and their asymptotic distributions. Small sample properties of the tests are discussed in Section 3 through a small simulation study, and two empirical applications are presented in Section 4. Section 5 gives concluding remarks.

2 Construction and asymptotic distributions of the test statistics

Let T be a duration variable, X a continuous covariate and let $\lambda(t|x)$ denote the hazard rate of T , given $X = x$, at $T = t$. We intend to test the hypothesis $H_0 : \lambda(t|x_1) = \lambda(t|x_2) \forall x_1, x_2$ against the alternative $H_1 : \lambda(t|x_1) \neq \lambda(t|x_2)$ for some $x_1 \neq x_2$. In particular, we are interested in test statistics that would be useful in detecting trend departures from H_0 of the form $H_1^* : \lambda(t|x_1) \geq \lambda(t|x_2) \forall x_1 > x_2$ (with the strict inequality holding for some $x_1 > x_2$) or its dual, and changepoint trend departures like $H_1^{**} : \text{there exists } x^* \text{ such that } \lambda(t|x) \uparrow x \text{ whenever } x < x^* \text{ and } \lambda(t|x) \downarrow x \text{ whenever } x > x^* \text{ (or its dual)}$.

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

$$\begin{aligned}
 T_{2s, std} &= \frac{T_{2s}}{\sqrt{\widehat{Var}[T_{2s}]}} \text{, where} \\
 T_{2s} &= \int_0^\tau L(t) d\widehat{\Lambda}_1(t) - \int_0^\tau L(t) d\widehat{\Lambda}_2(t), \\
 \widehat{Var}[T_{2s}] &= \int_0^\tau L^2(t) \{Y_1(t)Y_2(t)\}^{-1} d(N_1 + N_2)(t), \\
 L(t) &= K(t)Y_1(t)Y_2(t)\{Y_1(t) + Y_2(t)\}^{-1}, \\
 \tau &\text{ is a random stopping time (in particular, } \tau \text{ may be taken as the} \\
 &\text{time at the final observation in the combined sample),} \\
 K(t) &\text{ is a predictable process depending only on } Y_1 + Y_2, \\
 &\text{but not individually on } Y_1 \text{ or } Y_2, \\
 \widehat{\Lambda}_j(t) &\text{ (} j = 1, 2\text{), is the Nelson-Aalen estimator of the cumulative hazard} \\
 &\text{function in the } j^{\text{th}} \text{ sample,} \\
 Y_j(t) &\text{ (} j = 1, 2\text{), is the number of individuals on test in sample } j \text{ at time } t, \\
 \text{and } N_1, N_2 &\text{ are the counting processes counting the number of failures} \\
 &\text{in each sample.}
 \end{aligned}$$

²Refer to Li *et al.* (1996) for results of a large simulation study.

In particular, for the logrank test, $K(t) = I[Y_1(t) + Y_2(t) > 0]$, and for the Gehan-Breslow modification of the Wilcoxon test, $K(t) = I[Y_1(t) + Y_2(t) > 0] \cdot \{Y_1(t) + Y_2(t)\}$. In the two sample setup, these standardised test statistics have mean zero under the null hypothesis of equal hazards and positive/ (negative) mean if the hazards are trended. Further, they are asymptotically normally distributed under the null hypothesis.

Based on these test statistics, we propose a simple construction of our tests as follows. As in Bhattacharjee and Das (2002), we first select a fixed number of pairs of distinct points on the covariate space, and construct the usual two-sample test statistics ($T_{2s,std}$) for each pair, based on counting processes conditional on these two distinct covariate points. We then construct our test statistics, by taking supremum, infimum or average of these basic test statistics over these fixed number of pairs.

Thus, we fix $r > 1$, and select $2r$ distinct points $\{x_{11}, x_{21}, \dots, x_{r1}, x_{12}, x_{22}, \dots, x_{r2}\}$ on the covariate space \mathcal{X} , such that $x_{l2} > x_{l1}, l = 1, \dots, r$. We then construct our test statistics T_{2s}^* , T_{2s}^{**} and \bar{T}_{2s} based on the r statistics $T_{2s,std}(x_{l1}, x_{l2}), l = 1, \dots, r$ (each testing equality of hazard rates for the pair of counting processes $N(t, x_{l1})$ and $N(t, x_{l2})$), where

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

Then, our test statistics are:

$$\begin{aligned} T_{2s}^* &= \max \{T_{2s,std}(x_{11}, x_{12}), T_{2s,std}(x_{21}, x_{22}), \dots, T_{2s,std}(x_{r1}, x_{r2})\}, \\ T_{2s}^{**} &= \min \{T_{2s,std}(x_{11}, x_{12}), T_{2s,std}(x_{21}, x_{22}), \dots, T_{2s,std}(x_{r1}, x_{r2})\}. \\ \text{and } \bar{T}_{2s} &= \frac{1}{r} \sum_{l=1}^r T_{2s,std}(x_{l1}, x_{l2}). \end{aligned}$$

We now derive the asymptotic distributions of these test statistics.

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

Assumption 1 For each l , $l = 1, 2, \dots, r$, let $L(x_{l1}, x_{l2})(t)$ be a predictable processes (predictable with respect to t) of the form

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

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

Assumption 2 Let τ be a random stopping time. In particular, τ may be taken as the time at the final observation of the counting process $\sum_{l=1}^r \sum_{j=1}^2 N(t, x_{lj})$. In

principle, one could also have different stopping times $\tau(x_{l1}, x_{l2}), l = 1, \dots, r$ for each of the r basic test statistics.

Assumption 3 The sample paths of $L(x_{l1}, x_{l2})$ and $Y(t, x_{li})^{-1}$ are almost surely bounded with respect to t , for $i = 1, 2$ and $l = 1, \dots, r$. Further, for each $l = 1, \dots, r$, $L(x_{l1}, x_{l2})$ is zero whenever $Y(t, x_{l1})$ or $Y(t, x_{l2})$ are.

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

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

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

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

Let the test statistics T_{2s}^*, T_{2s}^{**} and \bar{T}_{2s} be as defined earlier.

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

- (a) $P [T_{2s}^* \leq z] \rightarrow [\Phi(z)]^r$,
- (b) $P [T_{2s}^{**} \geq -z] \rightarrow [\Phi(z)]^r$,

and

$$(c) \sqrt{r} \bar{T}_{2s} \xrightarrow{D} N(0, 1),$$

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

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

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

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

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

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

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

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

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

This follows from a version of Rebolledo's central limit theorem (see Andersen *et. al.*, 1992), noting that the innovation martingales corresponding to components of a vector counting process are orthogonal, and the vector of these martingales asymptotically converge to a Gaussian martingale. A similar argument in a different context can be found in Bhattacharjee and Das (2002).

It follows that

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

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

Proofs of (a), (b) and (c) follow. ◇

Corollary 1:

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

Proof: 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. ◇

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

$$\begin{aligned} T_{2s,\underline{w}}^* &= \max \{w_1 \cdot T_{2s,std}(x_{11}, x_{12}), w_2 \cdot T_{2s,std}(x_{21}, x_{22}), \dots, w_r \cdot T_{2s,std}(x_{r1}, x_{r2})\}, \\ T_{2s,\underline{w}}^{**} &= \min \{w_1 \cdot T_{2s,std}(x_{11}, x_{12}), w_2 \cdot T_{2s,std}(x_{21}, x_{22}), \dots, w_r \cdot T_{2s,std}(x_{r1}, x_{r2})\}. \\ \text{and } \bar{T}_{2s,\underline{w}} &= \left[\sum_{l=1}^r w_l \cdot T_{2s,std}(x_{l1}, x_{l2}) \right] / \left[\sum_{l=1}^r w_l \right]. \end{aligned}$$

Let Assumptions 1 to 4 hold. Then, under H_0 , as $n \rightarrow \infty$,

$$(a) P \left[T_{2s,\underline{w}}^* \leq z \right] \longrightarrow \prod_{l=1}^r [\Phi(z/w_l)],$$

and

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

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

Proof: Proof follows from Theorem 1. ◇

Remark 1: Restricting the statistics T_{2s}^* , T_{2s}^{**} and \bar{T}_{2s} to depend on a fixed number (r) of distinct pairs of points is a crucial step in the construction of the test statistics. This is because, the process $T_{2s, std}(x_1, x_2)$ on the space $\{(x_1, x_2) : x_2 > x_1, x_1, x_2 \in \mathcal{X}\}$, being pointwise standard normal and independent, do not have well-defined limiting processes, and the supremum (infimum) diverges to $+\infty$ ($-\infty$).

Remark 2: The significance of the Corollary 1 is that it gives a simple way of calculating the p-values for the test statistics, if r is reasonably large. Note that r is held fixed, and hence cannot increase to ∞ , but then it can be fixed at a large enough value (say, 20 or higher), so that the approximation can be fruitfully used.

Remark 3: Corollary 2 shows that one can weight the underlying test statistics by some measure of the distance between x_{l1} and x_{l2} . In other words, one can give higher weights to a covariate pair where the covariates are further apart. In practice, this would improve the empirical performance of the tests. We have, however, not used such weighting in the empirical work in Sections 3 and 4.

Since the covariate under consideration is continuous, it is not feasible to construct the basic tests $T_{2s, std}$ based on two distinct fixed points on the covariate space. We have considered “small” intervals around these (randomly) chosen points, such that the hazard function within these intervals can be approximately regarded as constant (across covariate values). While the derived distributions are for counting processes pertaining to specified pairs of points in the covariate space, the tests would go through for small intervals around these points, provided the covariate values are so chosen that they are continuity points of the hazard function. The average test statistics constructed in this way, however, often fail to maintain their nominal sizes under the null hypothesis, because of correlation between statistics based on overlapping intervals. As in Bhattacharjee and Das (2002), this issue can be resolved by using a jackknife estimator of the variance of this average estimator.

3 Simulation Results

The asymptotic distributions of the proposed test statistics have been derived in the previous Section. In this Section, we shall explore the finite sample performance of the tests for different specifications of the baseline hazard function and covariate dependence, by means of a small simulation study. In particular, we consider models of the form

$$\lambda(t, x) = \lambda_0(t) \cdot \exp[\beta(t, x)],$$

where $\lambda_0(t)$ and $\beta(t, x)$ are chosen to assume a variety of functional forms. The null hypothesis of absence of covariate dependence holds if and only if $\beta(t, x) = 0$. If, for fixed x , $\beta(t, x)$ increases/ decreases in x , we have trended alternatives of the type H_1^* . If, on the other hand, $\beta(t, x)$ increases in x over some range of the covariate space, and decreases over another, we have changepoint trend departures of the type H_1^{**} . In addition to the global alternative, our tests would also be consistent against both these kinds of alternatives to the null hypothesis.

The Monte Carlo simulations are based on independent right-censored data from the following 6 data generation processes, generated using the Gauss 386 random number generator, where the covariate X are i.i.d. $U(-1, 1)$.

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

The censoring duration C are i.i.d. *Exponential*(6) for DGP_{11} , DGP_{12} and DGP_{13} and *Exponential*(2) for DGP_{21} , DGP_{22} and DGP_{23} . The data generating processes DGP_{11} and DGP_{21} belong to the null hypothesis of absence of covariate dependence, DGP_{12} and DGP_{22} belongs to the category of monotone trended alternatives, and DGP_{13} and DGP_{23} are changepoint trended. We have used the logrank test to construct the basic test statistics, and used 100 distinct pairs of covariate values. The following Table presents simulation results for 10,000 simulations from the above data generating processes with samples of size 100 and 200.

TABLE
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}^* | 3.76 | 5.59 | 0.67 | 1.08 |
| | T_{2s}^{**} | 7.23 | 5.66 | 1.18 | 0.88 |
| | \bar{T}_{2s} | 5.46 | 5.35 | 1.19 | 0.99 |
| DGP_{12} | T_{2s}^* | 95.46 | 100.00 | 82.98 | 100.00 |
| | T_{2s}^{**} | 2.43 | 1.91 | 0.41 | 0.80 |
| | \bar{T}_{2s} | 96.82 | 100.00 | 87.95 | 100.00 |
| DGP_{13} | T_{2s}^* | 26.06 | 63.30 | 5.67 | 29.41 |
| | T_{2s}^{**} | 38.19 | 70.62 | 12.29 | 40.40 |
| | \bar{T}_{2s} | 5.67 | 4.83 | 1.23 | 0.94 |
| DGP_{21} | T_{2s}^* | 3.90 | 5.51 | 0.53 | 1.61 |
| | T_{2s}^{**} | 7.24 | 6.12 | 1.45 | 0.79 |
| | \bar{T}_{2s} | 5.62 | 5.68 | 0.92 | 1.35 |
| DGP_{22} | T_{2s}^* | 97.18 | 100.00 | 86.03 | 99.87 |
| | T_{2s}^{**} | 2.69 | 1.85 | 0.41 | 0.82 |
| | \bar{T}_{2s} | 97.71 | 100.00 | 92.02 | 100.00 |
| DGP_{23} | T_{2s}^* | 21.26 | 54.50 | 4.39 | 23.04 |
| | T_{2s}^{**} | 36.44 | 69.35 | 11.64 | 37.73 |
| | \bar{T}_{2s} | 7.18 | 6.96 | 1.56 | 2.06 |

The results show that the nominal sizes are maintained, and the tests have reasonable power in small samples, except for DGP_{13} and DGP_{23} . This is not surprising, since these data generation processes are changepoint trended. Hence, when a pair of points are drawn at random from the covariate space, only a quarter of them reflect the increasing nature of covariate dependence, and another quarter reflect the decreasing nature. Overall, the tests are fairly powerful and robust in finite samples. The results also reflect the strength of the supremum/ infimum test statistics in their ability to detect non-monotonic departures from the null hypothesis of absence of covariate dependence (DGP_{13} and DGP_{23}).

The tests are not directly comparable with other trend tests available in the literature. However, we have made some investigations into how these tests compare in terms of power, by stratification with respect to the value of the covariate; our tests perform favourably in this exercise (the results of the exercise are available are not presented in the paper, but are available with the author).

4 Empirical Applications

In this section, we illustrate the tests proposed in this paper by way of two applications, one to corporate failure in the UK and the other to child mortality in rural India.

4.1 Firm Exits in the UK

The data are on firm exits through bankruptcy over the period 1980 to 1998 and pertain to 2789 listed manufacturing companies, covering 24,034 company years and includes 95 bankruptcies. The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1980, and contain delayed entries. Here the focus of our analysis is on the impact of aggregate Q on corporate failure, more detailed analysis of these data are reported elsewhere (Bhattacharjee *et. al.*, 2002). Following usual practice, we consider the reciprocal of Q as the continuous covariate.

A priori, one would expect periods with higher values of this covariate to depress the incidence of bankruptcies. However, an estimate of the Cox proportional hazards model on these data gives a hazard ratio of 0.92, with p-value 15.6 per cent. One would then be tempted to believe that covariate dependence is absent.

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

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

| Test | Test Statistic | P-Value (%) |
|-------------------------------|----------------|-------------|
| T_{2s}^* - Logrank | 0.592 | 100.00 |
| T_{2s}^{**} - Logrank | -3.732 | 1.88 |
| T_{2s}^* - Gehan-Breslow | 0.500 | 100.00 |
| T_{2s}^{**} - Gehan-Breslow | -3.046 | 3.70 |

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

TABLE 3:

MODEL ESTIMATES: CORPORATE BANKRUPTCY DATA

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

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

4.2 Child Mortality in rural India

Here, mortality outcomes of children in rural India are examined, using data from the National Family Health Survey 1992-93 (for further details of the analysis, see Bhalotra and Bhattacharjee (2001)). In particular, we are interested in understanding the relationship between mortality hazards and mother’s age at child birth (one of the most important (physiological) determinants of child mortality).

Several researchers have included mother’s age as a continuous covariate in Cox regression models. Such an analysis not only relies heavily on model assumptions, but is also seriously flawed from a theoretical point of view. This is because the impact of mother’s age on child mortality is expected to be in the nature of a changepoint trend; as mother’s age rises (from 16 years to say about 25 years), mortality is expected to decline, and when it rises further (say, upto 35 years), mortality is expected to rise again. Therefore, the Cox proportional hazards model would not be expected to give correct indication of the nature of covariate dependence; the hazard ratio corresponding to mother’s age (in months) is estimated at 0.9998, which is not significant even at 10 per cent level.

TABLE 4:

TESTS FOR ABSENCE OF COVARIATE DEPENDENCE:

CHILD MORTALITY IN RURAL INDIA

| Test | Test Statistic | P-Value (%) |
|--------------------------------|----------------|-------------|
| T_{2s}^* - Logrank | 4.103 | 0.90 |
| T_{2s}^{**} - Logrank | -5.763 | 0.00 |
| \bar{T}_{2s} - Logrank | -0.317 | 98.32 |
| T_{2s}^* - Gehan-Breslow | 4.259 | 0.54 |
| T_{2s}^{**} - Gehan-Breslow | -5.946 | 0.00 |
| \bar{T}_{2s} - Gehan-Breslow | -0.412 | 99.77 |

To explore trending in mortality hazard rates with respect to mother’s age at child birth, we apply our test statistics to the data. The results (Table 4) confirm our *a priori* suspicion of changepoint trend. The null hypothesis of absence of covariate dependence is rejected, in favour of both positive and negative trend.

As in the previous example, the covariate pairs corresponding to the supremum and infimum test statistics provide further information about the nature of covariate dependence in different regions on the covariate space, and in particular enable

approximate estimation of the changepoint in changepoint trend models (for similar applications relating to non-monotone departures from proportionality, see Bhattacharjee and Das (2002)). For example, using the logrank test, the hazard increases in mother's age upto the age of about 24 years (the maxima is attained for the covariate pair 184 months and 293 months), and decreases after 27 years (covariate pair 330 months and 479 months correspond to the minima test statistic). The nature of covariate dependence revealed in these data correspond closely with what is observed the biomedical literature.

The data also reveal different patterns in covariate dependence in different regions of India. For example, there is insufficient evidence of negative trending upto the age of 24 years in the Indian state of Kerala. This fact may be explained in terms of better health-care provisions in this region (see Bhalotra and Bhattacharjee (2001) for further details).

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

5 Conclusion

In summary, the tests described in this paper add an important tool to the armoury of a duration data analyst. It extends an important class of two sample tests for equality of hazards to a continuous covariate framework. This also shows that usual statistical treatment of duration data using counting processes are useful in analysing such continuous covariate situations. In conjunction with Bhattacharjee and Das (2002), this paper extends many of these two sample testing procedures to the continuous covariate setup, and thereby makes these tests more readily usable in real life econometric applications.

Though most of the discussion in this paper has been confined to a single continuous covariate setup, the tests can be readily used in applications with multiple continuous covariates. Here, one has two options. The first one is to test the absence of covariate dependence for one covariate, while assuming the other covariates to conform to a Cox regression model framework. Then, one can use the estimates of baseline cumulative hazard functions derived from a Cox regression model (including the other covariates, but not the one under study) to construct the appropriate test statistics. Alternatively, one can proceed to jointly test for covariate dependence for two or more covariates together.

The tests can be extended to duration data with unobserved heterogeneity. Even when this heterogeneity distribution is left unspecified (nonparametric), the cumulative hazard estimates proposed in Horowitz (1999) can be used to construct the tests. Some further work is however required to establish the distribution theory in this situation.

REFERENCES

- Andersen, P.K., Borgan, O., Gill, R.D. and Keiding, N. (1992). *Statistical Models based on Counting Processes*. Springer-Verlag, New York.
- Berman, S.M. (1992). *Sojourns and Extremes of Stochastic Processes*. Wadsworth and Brooks/ Cole, Pacific Grove, CA.

- Bhalotra, S.R. and Bhattacharjee, A. (2001). Understanding regional variations in child mortality in India. Presented at the Workshop on 'Welfare, Demography and Development', University of Cambridge, Cambridge, UK, September 2001. Mimeo.
- Bhattacharjee, A. and Das, S. (2002). Testing Proportionality in Duration Models with Respect to Continuous Covariates. DAE Working Paper No. **0220**, Department of Applied Economics, University of Cambridge.
- , Higson, C., Holly, S. and Kattuman, P. (2002). Macro economic instability and business exit: Determinants of failures and acquisitions of large UK firms. DAE Working Paper No. **0206**, Department of Applied Economics, University of Cambridge.
- Breslow, N.E. (1970). A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika*, **57**, 579–594.
- Brown, B.W., Jr., Hollander, M. and Korwar, R.M. (1974). Nonparametric tests of independence for censored data, with applications to heart transplant studies. In: *Reliability and Biometry, Statistical Analysis of Lifelength* (Eds.) Proschan, F. and Serfling, R.J., Society for Industrial and Applied Mathematics: Philadelphia, 327–354.
- Cox, D.R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society, Series B* **34**, 187–220.
- Fleming, T.R. and Harrington, D.P. (1991). *Counting processes and survival Analysis*. John Wiley and Sons, New York.
- Gehan, E.A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly censored samples. *Biometrika*, **52**, 203–223.
- Harrington, D.P. and Fleming, T.R. (1982). A class of rank test procedures for censored survival data. *Biometrika*, **69**, 133–143.
- Horowitz, J.L. (1999). Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica*, **67**, 5, 1001–1028.
- and Neumann, G.R. (1992). A generalised moments specification test of the proportional hazards model. *Journal of the American Statistical Association*, **87**, 234–240.
- Jespersen, N.C.B. (1986). Dichotomising a continuous covariate in the Cox regression model. Research Report **86/2**, Statistical Research Unit, University of Copenhagen.
- Jones, M.P. and Crowley, J.J. (1989). A general class of nonparametric tests for survival analysis. *Biometrics*, **45**, 157–170.
- and — (1990). Asymptotic properties of a general class of nonparametric tests for survival analysis. *Annals of Statistics*, **18**, 1203–1220.
- Li, Y.-H., Klein, J.P. and Moeschberger, M.L. (1996). Effects of model misspecification in estimating covariate effects in survival analysis for small sample sizes. *Computational Statistics and Data Analysis*, **22**, 177–192.
- Liu, P.Y., Green, S., Wolf, M. and Crowley, J. (1993). Testing against ordered alternatives for censored survival data. *Journal of the American Statistical Association*, **88**, 421, 153–160.
- and Tsai, W.Y. (1999). A modified logrank test for censored survival data under order restrictions. *Statistics and Probability Letters*, **41**, 57–63.
- Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.*, **50**, 163–170.
- Mau, J. (1988). A generalization of a nonparametric test for stochastically ordered distributions to censored survival data. *Journal of the Royal Statistical Society Series B*, **50**, 403–412.
- 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.

- Neumann, G.R. (1997). Search models and duration data. In: *Handbook of Applied Econometrics Volume II: Microeconometrics* (Eds.) Pesaran, M.H., Basil Blackwell: Oxford, Chapter 7, 300–351.
- O'Brien, P.C. (1978). A nonparametric test for association with censored data. *Biometrics* **34**, 243–250.
- 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.
- Prentice, R.L. (1978). Linear rank tests with right censored data. *Biometrika*, **65**, 167–179. Correction: **70**, 304 (1983).
- 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.
- Tarone, R.E. (1975). Tests for trend in life table analysis. *Biometrika*, **62**, 679–682.
- and Ware, J.H. (1977). On distribution-free tests for equality of survival distributions. *Biometrika*, **64**, 156–160.