ORDER-RESTRICTED COVARIATE EFFECTS AND HAZARD REGRESSION MODELS

Ph.D. Thesis

ARNAB BHATTACHARJEE

INDIAN STATISTICAL INSTITUTE

APPLIED STATISTICS UNIT

KOLKATA, INDIA.

February 2008

ORDER-RESTRICTED COVARIATE EFFECTS AND HAZARD REGRESSION MODELS

Arnab Bhattacharjee

B.Stat. (Hons.), Indian Statistical Institute, 1991
M.Stat., Indian Statistical Institute, 1993
PhD in Economics, University of Mumbai, 2002

Thesis submitted to the Indian Statistical Institute
in partial fulfillment of the requirement
for
the award of
DOCTOR OF PHILOSOPHY

Supervisor: Professor Debasis Sengupta

INDIAN STATISTICAL INSTITUTE

APPLIED STATISTICS UNIT
KOLKATA, INDIA.

February 2008

List of papers by the author included in this thesis

	Title, Co-author(s), Publication	References in the thesis	
		Discussed in	**Referred in**
1.	Testing for the proportionality of hazards in two samples against the increasing cumulative hazard ratio alternative. (Co-authors: D. Sengupta, B. Rajeev). Publication: *Scandinavian Journal of Statistics* **25**(4), 1998, 637–647.	Chapter 2	Chapters 1, 3, 4, 5, 6 and 8.
2.	Estimation in hazard regression models under ordered departures from proportionality. (Co-authors: None). Publication: *Computational Statistics and Data Analysis* **47**(3), 2004, 517–536.	Chapter 4	Chapters 1, 3, 6, 7 and 8.
3.	Macroeconomic conditions and business exit: determinants of failures and acquisitions of UK firms. (Co-authors: C. Higson, S. Holly, P. Kattuman). Publication: *Economica*, forthcoming.	Chapter 7	Chapters 1, 3, 4, 5, 6 and 8.
4.	Macroeconomic instability and corporate failure. (Co-authors: C. Higson, S. Holly, P. Kattuman). Publication: *Review of Law and Economics*, forthcoming.	Chapter 7	Chapters 1, 3, 5, 6 and 8.
5.	Monotone departures from proportional hazards with respect to continuous covariates: inference procedures and applications. (Co-authors: None). Publication: *Mimeo*, 2003. Proceedings of 54th Binennial Session of the International Statistical Institute.	Chapter 4	Chapters 1, 6, 7 and 8.
6.	Testing the Proportional Hazards Model with Continuous Covariates in Duration Models against Monotone Ordering. (Co-authors: None). Publication: *Mimeo*, Current version: 2007. Working Paper, University of Cambridge, UK.	Chapter 3	Chapters 1, 4, 5, 6, 7 and 8.
7.	A simple test for the absence of covariate dependence in hazard regression models. (Co-authors: None). Publication: *Mimeo*, Current version: 2007. Working Paper, University of St Andrews, UK.	Chapter 5	Chapters 1, 4, 6, 7 and 8.
8.	Bayesian analysis of hazard regression models under order restrictions on covariate effects and ageing. (Co-authors: M. Bhattacharjee). Publication: *Mimeo*, Current version: 2007. Working Paper, University of St Andrews, UK.	Chapter 6	Chapters 1, 4, 5, 6 and 8.
9.	Models of firm dynamics and the hazard rate of exits: is unobserved heterogeneity really that important? (Co-authors: None). Publication: *Mimeo*, Current version: 2007. Working Paper, University of St Andrews, UK.	Chapter 7	Chapters 1, 4, 5 and 8.

ACKNOWLEDGEMENT

This thesis represents a rather long and personal journey. It has taken me through widely different environments and experiences, and encompassing considerable expanses of space and time. I started my research work in statistics, as a research student at the Indian Statistical Institute, way back in 1993. In 1994, I moved into the professional field, joining the Reserve Bank of India, and subsequently moved to a different academic discipline – economics. Therefore, the work here represents a line of research I have followed, as a leisure activity, over a long period of time.

Many people – friends, family, colleagues and senior professionals have encouraged me to continue with my research work in statistics. I particularly express my deep sense of appreciation for my research supervisor, Prof. Debasis Sengupta for his guidance and persuation at each stage of this thesis work. I am also deeply indebted to Elja Arjas, R.B. Barman, Sushama Bendre, Arup Bose, Sanjoy Bose, Madhuchhanda Bhattacharjee, Snigdhansu Chatterjee, Prabal Chaudhuri, Samarjit Das, Jayant Deshpande, Anup Dewanji, Chris Higson, Sean Holly, (Late) Chris Jensen-Butler, Peter Jupp, Ajit Karnik, Paul Kattuman, Roderick McCrorie, Mukul Majumdar, Sumit Majumdar, Rahul Mukerji, Charles Nolan, Abhay Pethe, Gavin Reid, P.G. Sankaran and David Ulph for providing many suggestions and help in academic matters. Sincere thanks are also due to my dear friends, particularly Subhomoy Bhattacharyya, Jean Bonnet, Bram Boskamp, Jagjit Chadha, Subhasis Chakraborty, Eduardo Anselmo de Castro, Jim Jin, Caroline Moore, Debdulal Roy, Christoph Thoenissen, Carme Vila and Andrew Vivian, who provided the support that was necessary to sustain my passion for this long period of time, often over tumultuous periods. Without their active help and understanding over all these years, this work would never have materialised.

I am thankful to many people at the Indian Statistical Institute for their help and support over the years. In addition to seniors and colleagues mentioned above, particular thanks are due Prof. Pradipta Banerjee, Prof. Alok Goswami, Prof. Bimal Roy, Dr. Rudrapada Sarkar and Prof. Ashis Sengupta, and Aparesh-da, Ashok-da and Sangal-da.

Over the years, Madhuchhanda, and of late, little Ausija have borne much of the adverse shocks related to my research. Without their sincere support and love, and immense patience, this work would never have been complete. Special thanks are due to my friend Sabarna, but particularly Nandini and Sudur, for being fantastic company and providing the support that was required during the stressful final stages of this journey.

I gratefully acknowledge financial support from the Leverhulme Trust and KPMG UK for funding part of the work included in this thesis (particularly Chapters 4 and 7). Last but not the least, sincere thanks and gratitude are due to Luc Bauwens, John Beath, David Dunson, Peter Hart, Erricos Kontoghiorges, Manfredi La Manna, Geoff Meeks, Oliver Linton, Hashem Pesaran, Elvezio Ronchetti, Ananda Sen, Gerard van den Berg and Melvyn Weeks, in addition to fellow academics mentioned above, for helpful comments and suggestions on various parts of this research.

Arnab Bhattacharjee
Kolkata, February 2008

Contents

Chapter 1

Introduction

The proportional hazards (PH) model, but more specifically its special case the Cox regression model (Cox, 1972), plays an important role in the theory and practice of lifetime and duration data analysis. This is because the PH model (and the Cox regression model) provides a convenient way to evaluate the influence of one or several covariates on the probability of conclusion of lifetime or duration spells. However, the PH specification substantially restricts interdependence between the explanatory variables and the lifetime in determining the hazard. In particular, the Cox regression model model restricts the coefficients of the regressors in the logarithm of the hazard function to be constant over the lifetime. This restriction may not hold in many situations, or may even be unreasonable from the point of view of relevant theory. Further, this and other kinds of misspecification often lead to misleading inferences about the effects of explanatory variables and the shape of the baseline hazard.

Testing the Cox PH model, particularly against the omnibus alternative, has therefore been an area of active research. However, the omnibus tests do not offer much clarity regarding the nature of departure from underlying assumptions. As a result, these tests do not provide useful inference for further modeling covariate effects when the Cox regression model does not hold. For example, it is often of interest to explore whether the hazard rate for one level of the covariate increases in lifetime relative to another level (i.e., the hazard ratio increases/decreases with lifetime). Ordered departures from proportionality of this and related types are useful in the two-sample (or binary covariate) setup for studying commonly observed features like crossing hazards. Similar situations also occur quite frequently in the k-sample setup and when the covariate is continuous. Throughout this thesis, we call such ordered departures generically as "*order restrictions on covariate dependence*", as distinct from "*order restrictions on ageing*" which refers to restrictions on the shape of the baseline hazard function (or, on duration dependence).

The work included in this thesis develops analytical and graphical inference on covariate effects in situations when the Cox regression model, or more generally the PH model, may not hold. In particular, we develop methods to study covariate effects in the presence of potentially order restricted departures from proportionality. The thesis places emphasis on both theory and applications, and extends the literature along both these dimensions in several ways. In this sense, the work is firmly set within the tradition of research in applied statistics and econometrics.

In the following section (Section 1.1), we motivate our research on order restrictions on covariate dependence using a few real life examples, focusing on some useful ways in which order restrictions can be characterised and hazard regression models accommodating order restricted covariate effects. Next, in Section 1.2, we review recent research on hazard regression models, which are useful for modeling and estimation of covariate dependence under order restrictions, particularly when the covariate is continuous. The review is selective, focusing largely on order restrictions in these models and aimed at identifying gaps in the literature. As we proceed, we place the main contributions made in the thesis within the context of the literature. Finally, we outline the new research and describe the chapter scheme for the rest of the thesis (Section 1.3).

1.1 Motivation for the research

The main focus of our research is on the way covariate effects deviate from the proportional hazards assumption. We first discuss the two sample setup, where the covariate under consideration is binary. Following this, we discuss continuous covariates and finally, a regression model for nonproportional hazards. In each of these main themes, we motivate our research using real life applications.

1.1.1 Two sample setup

In the two sample setup, Gill and Schumacher (1987) and Deshpande and Sengupta (1995) consider departures where the ratio of hazard rates in the two samples is monotonically increasing or decreasing with the lifetime[1]. They develop tests of the null hypothesis of proportionality against the increasing (decreasing) hazard ratio alternative. Sengupta and Deshpande (1994) show that this kind of departure is equivalent to convex-ordering of the lifetime distributions in the two samples (Kalashnikov and Rachev, 1986). Denoting by λ_F and λ_G (correspondingly, Λ_F and Λ_G) the hazard functions (cumulative hazard functions) in the two samples,

$$\frac{\lambda_F(t)}{\lambda_G(t)} \uparrow t \text{ on } [0,\infty) \Longleftrightarrow F \underset{c}{\prec} G, \tag{1.1}$$

where convex ordering of the lifetime distributions is defined as the condition that $\Lambda_F \circ \Lambda_G^{-1}$ is a convex function on $[0,\infty)$. Similarly, Deshpande and Sengupta (1994) also show that star-ordering of the lifetime distributions is equivalent to monotone ratio of cumulative hazard functions

$$\frac{\Lambda_F(t)}{\Lambda_G(t)} \uparrow t \text{ on } (0,\infty) \Longleftrightarrow F \underset{*}{\prec} G, \tag{1.2}$$

where star-ordering of the lifetime distributions is defined by $\Lambda_F \circ \Lambda_G^{-1}$ being a star-shaped function from $[0,\infty)$ to $[0,\infty)$[2].

One important starting point for our work is the analysis, in Gill and Schumacher (1987), of the Veterans' Administration data (Detre *et al.*, 1977) on a controlled clinical trial in chronic stable angina. The main purpose of the analysis is to compare survival times of patients receiving coronary artery bypass graft surgery and of patients receiving a conservative medical treatment. The tests proposed by Gill and Schumacher (1987) fail to reject the null hypothesis of proportionality against the alternative of decreasing hazard ratio for surgery versus medical treatment. However, a plot of the empirical trend function (Lee and Pirie, 1981) with log rank weight function (Gill and Schumacher (1987), shown in Figure 1-1 with axes interchanged)

[1] The word "increasing" would mean "non-decreasing" throughout this thesis; similarly "decreasing" will mean "non-increasing".

[2] A non-negative valued function is star-shaped function if any straight line through the origin intersects it at most once and from above; a negative star-shaped function has the opposite property. Note that star-shapedness is a weaker property than convexity; similarly negative star-shapedness is weaker than concavity.

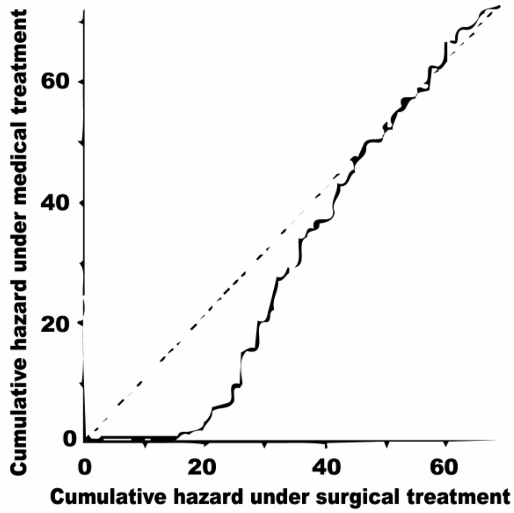


Figure 1-1: Lee-Pirie plot for Veterans' Administration data (Figure 5, Gill and Schumacher (1987), with axes interchanged)

demonstrate clear evidence of ordered departure from proportionality. In fact, the star-shaped pattern (Kalashnikov and Rachev, 1986) suggests that a monotone cumulative hazard ratio alternative, which is weaker than concave ordering, may characterise the nature of depature from proportionality more accurately.

This example motivated us to develop tests for proportionality against the weaker monotone cumulative hazard ratio alternative. Further, the success of alternatives such as convex and star ordering in describing ordered departures from proportional hazards in the two-sample setup also motivate our work on extending the tests to a competing risk framework.

1.1.2 Continuous covariates

While the above characterisation of covariate dependence in nonproportional hazards situation is useful in studying commonly observed phenomena like crossing hazards, the two sample setup itself is rather restrictive in its application. At the same time, similar evidence of ordered departures from proportionality with respect to continuous covariates are quite common in applications.

For survival with malignant melanoma, for example, Andersen *et al.* (1993) observe that, while "hazard seems to increase with tumor thickness" (pp. 389), the plot of estimated cumulative baseline hazards for patients with '2 mm $\leq$ tumor thickness $<$ 5 mm' and 'tumor thickness $\geq$ 5 mm' against that of patients with 'tumor thickness $<$ 2 mm' reveal "concave looking curves indicating that the hazard ratios decrease with time" (pp. 544–545). Similarly, Jayet and Moreau (1991), using data on French employment durations, find that the ratio of hazard function for individuals in the age groups $24-28$ years to that for $37-40$ years is increasing upto a duration of approximately 120 days.

Motivated by applications like the above, we extend the notions of order restricted covariate dependence to the case of continuous covariates. For example, we define the lifetime random variable T to have *increasing hazard ratio for continuous covariate (IHRCC)* with respect to a continuous covariate X if,

$$\text{whenever} \ \ x_1 > x_2, \qquad (T|X=x_1) \underset{c}{\prec} (T|X=x_2). \tag{1.3}$$

Within this very general framework, we develop tests of the proportional hazards assumption against ordered alternatives. These tests are powerful and can detect departures not only in the direction of alternatives like $IHRCC$, but also violations of the proportional hazards hypothesis where the covariate effects change at an unknown changepoint.

1.1.3 Modeling nonproportional hazards

For further inference and modeling of covariate effects in the presence of such non-proportionality of conditional hazard functions, it is useful to consider appropriate alternative hazard regression models. We argue the use of a multiplicative hazard regression model with time-varying coefficients

$$\lambda\left(t|x\right) = \lambda_0(t).\exp\left(\beta(t).x\right)$$

for modeling the nature of nonproportionality. Under this model, monotonicity of the coefficients as a function of lifetime is equivalent to ordered departures of the $IHRCC$ type (1.3):

$$\beta(t) \uparrow t \text{ on } \ [0,\infty) \Longleftrightarrow (T|X=x_1) \underset{c}{\prec} (T|X=x_2) \ \text{ whenever } \ x_1 > x_2,$$

while changepoint departures like in the unemployment duration data (Jayet and Moreau, 1991) correspond to the time-varying coefficients increasing over one range of lifetimes and decreasing over another.

To demonstrate the flexibility and usefulness of this approach, we consider an application to exits due to bankruptcy among listed firms in the UK (Bhattacharjee *et al.*, 2008a, 2008b). The main purpose of our focus is to study the effect of macroeconomic instability on business failure. Based on economic theory and prior evidence, we expect instability to have an adverse effect on firm survival, and therefore a positive covariate effect on bankruptcy hazard. Further, the effect of instability is expected to reduce with the age of the firm, implying time varying coefficients that decrease to zero. Inference on the strength of the effect of instability and its variation with age have important relevance for economic and legislative policy implications. In fact, using exchange rate volatility as a measure of instability, the null hypothesis of proportional hazards is rejected against a *decreasing hazard ratio for continuous covariate (DHRCC)* alternative.

Our research demonstrates that biased bootstrap methods, such as data tilting and particularly local adaptive bandwidths provide useful order restricted estimates of such hazard regression models. The plot of adaptive bandwidth estimates (Figure 1-2a) demonstrate the variation in and strength of covariate effects over the lifetime of the firm, and provide useful and policy relevant inference. Figure 1-2b reports bayesian order restricted inference on the covariate effect of instability on discrete failure time data in the presence of arbitrary frailty. These estimates provide similar inference, though accounting for frailty somewhat reduces the inferred strength of the effect of instability.

The above applications motivate the main research ideas developed in this thesis. We develop analytical and graphical inference tools to examine evidence of nonproportional hazards in two sample and continuous covariate setups, and to infer on the nature of nonproportionality. Further, we develop frquentist and bayesian inference in regression models admitting a variety of nonproportional hazard situations. Finally, we make contributions towards developing economic theory and applications for understanding macroeconomic effects on the survival of firms.

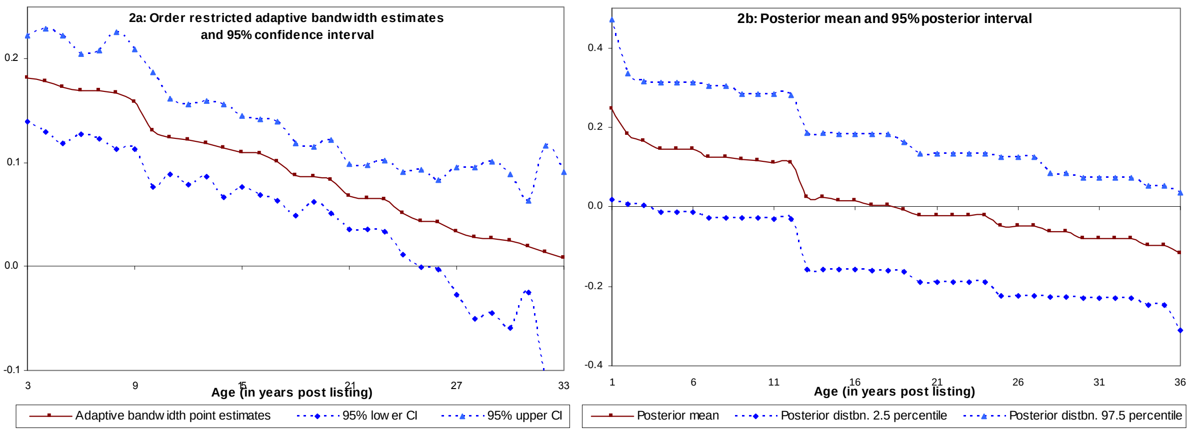


Figure 1-2: Effect of exchange rate volatility on bankruptcy exit hazard

1.2 Order restrictions in hazard regression models

The Cox regression model (Cox, 1972) has been the workhorse of hazard regression models and played an important role in the theory and practice of lifetime and duration data analysis over the past few decades. This is because this model (and more generally the PH model) provides a convenient way to evaluate the influence of one or several covariates on the probability of termination of lifetime or duration spells. Limitations and extensions of this model provide the context of the present review of recent research.

1.2.1 The Cox regression model

The model[3] specifies that the hazard function of the failure time conditional on a set of possibly time varying covariates is the product of an arbitrary baseline hazard function and a regression function of the covariates. For a failure time variable T associated with an experimental unit with vector of possibly time-dependent covariates $\underline{X}(t)$, this model postulates that the conditional hazard rate function of T at time t, given $\underline{X}(t)$, is

$$\lambda\left(t|\underline{X}(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}^T.\underline{X}(t)\right], \tag{1.4}$$

[3] The discussion of the Cox regression model here is largely based on Andersen *et al.* (1993).

where $\lambda_0(.)$ is some baseline hazard function, $\underline{\beta}$ is a vector of regression coefficients, and superscript T (^{T}) denotes vector/matrix transpose[4]. An alternative, and often more convenient, way of representing the Cox regression model is in the form of a linear transformation model, in terms of the baseline cumulative hazard function $\Lambda_0(t) = \int_0^t \lambda_0(s)ds$:

$$\ln \Lambda_0(t) = -\underline{\beta}^T.\underline{X}(t) + \varepsilon,$$

where $\ln \Lambda_0(t)$ is a positive-valued but arbitrary increasing function and ε has the usual extreme value distribution[5].

The Cox regression model allows the baseline hazard function to assume a completely unrestricted shape. The multiplicative separation of the effect of lifetime from that of the other covariates has a very important implication. The hazard functions conditional on two different values of the covariate vector is independent of the lifetime

$$\frac{\lambda\left(t|\underline{X}_i(t)\right)}{\lambda\left(t|\underline{X}_j(t)\right)} = \exp\left[\underline{\beta}^T.\left(\underline{X}_i(t) - \underline{X}_j(t)\right)\right].$$

In other words, the conditional hazard functions are proportional to each other. The proportionality also holds under the more general PH model

$$\lambda\left(t|\underline{X}(t)\right) = \lambda_0(t).\phi\left[\underline{\beta}^T.\underline{X}(t)\right], \tag{1.5}$$

where $\phi(.)$ is an arbitrary (smooth) monotone function. An alternative interpretation of this result is that the impact of the covariate vector on the conditional hazard is the same multiplicative factor $\exp\left[\underline{\beta}^T.\underline{X}(.)\right]$ at any lifetime; this interpretation will be important in subsequent developments in this thesis.

Cox (1972, 1975) introduced the ingenious partial likelihood principle to eliminate the infinite dimensional baseline hazard function from the estimation of regression parameters with censored data and potentially time varying covariates. For untied failure time data with time-

[4]Please note that throughout this thesis, the failure time variable will be denoted by T while superscript T (^{T}) will denote the tanspose of a matrix or vector.

[5]When ε has a standard logistic distribution, we get the proportional odds model.

varying covariates, the Cox partial likelihood has the form

$$PL\left(\underline{\beta}\right)=\prod_{i=1}^{n}\prod_{t\geq 0}\left\{\frac{Y_i(t).\exp\left[\underline{\beta}^T.\underline{X}_i(t)\right]}{\sum_j Y_j(t).\exp\left[\underline{\beta}^T.\underline{X}_i(t)\right]}\right\}^{dN_i(t)}, \tag{1.6}$$

where $Y_i(t)$ is the at-risk indicator taking value 1 if individual i is under observation and at risk at time t (zero otherwise), and $N_i(t)$ denotes the number of observed failures for individual i over the interval $[0,t]$; $dN_i(t)$ takes the value 1 if the individual has failed at time t (zero otherwise). The beauty of the above formulation is in that the infinite dimensional baseline hazard function is not included in the partial likelihood function at all. Having obtained the maximum partial likelihood estimator, $\widehat{\underline{\beta}}$, by maximising the partial likelihood (1.6), the baseline cumulative hazard function is estimated using the Aalen-Breslow estimator (Breslow, 1975; Aalen, 1993)

$$\begin{aligned}\widehat{\Lambda_0}\left(t,\widehat{\underline{\beta}}\right) &= \int_0^t \frac{d\overline{N}(s)}{\sum Y_i(s).\exp\left[\widehat{\underline{\beta}}^T.\underline{X}_i(s)\right]}, \\ d\overline{N}(s) &= \sum_{i=1}^{n} dN_i(s).\end{aligned} \tag{1.7}$$

If there are no covariates, this estimator reduces to the familiar Nelson-Aalen estimator (Nelson, 1969, 1972; Aalen, 1975, 1978)

$$\widehat{\Lambda_0}\left(t\right)=\int_0^t \frac{d\overline{N}(s)}{\overline{Y}(s)}, \qquad d\overline{N}(s)=\sum_{i=1}^{n} dN_i(s), \overline{Y}(s)=\sum_{i=1}^{n} Y_i(s). \tag{1.8}$$

for the cumulative hazard function of a lifetime or duration variable.

In a seminal paper, Andersen and Gill (1982) extended the Cox regression model to general counting processes and established the asymptotic properties of the maximum partial likelihood estimator and the associated Breslow (1974) estimator of the cumulative baseline hazard function via the elegant counting process martingale theory. This follows from the representation of the log partial likelihood as

$$l\left(\underline{\beta}\right)=\sum_{i=1}^{n}\int_0^\infty \left[Y_i(t).\underline{\beta}^T.\underline{X}_i(t)-\ln\left(\sum_j Y_j(t).\exp\left[\underline{\beta}^T.\underline{X}_i(t)\right]\right)\right]dN_i(t), \tag{1.9}$$

and the Doob-Meyer decomposition of the counting process $N_i(t)$

$$dM_i(t) = dN_i(t) + Y_i(t).\lambda_0(t).\exp\left[\underline{\beta}^T.\underline{X}_i(t)\right].dt \tag{1.10}$$

where $M_i(t)$ is a standard counting process martingale (for details, see Andersen *et al.*, 1993).

The partial likelihood argument follows through in the case of staggered entry (sometimes also called delayed entry) where some individuals are not observed from time zero. This kind of situation is present in some of the empirical applications included later in the thesis. Large sample theory for this case has been developed in Tsiatis (1981) and Sellke and Siegmund (1983).

These contributions render the Cox regression model very convenient for empirical analysis while at the same time retaining the flexibility of a fully nonparametric baseline hazard function. This flexibility comes at a cost – the partial likelihood estimates of the covariate effects as well as the shape of the baseline hazard function are known to be highly sensitive to violation of the model's various assumptions. This issue has been discussed in the literature, for example, in the work of Johnson *et al.* (1982), Lagakos and Schoenfeld (1984), Solomon (1984), Struthers and Kalbfieisch (1986) and Lagakos (1988). A large simulation study reported in Li *et al.* (1996) highlight these issues quite strongly.

1.2.2 Effect of misspecification

There are several basic and important features of the Cox regression model,.tests for the underlying assumptions for many of which are critical for the model's use in empirical studies. These various aspects of the model as well as the corresponding assumptions also suggest directions for extending the model (Therneau and Grambsch, 2000)[6]. The first of these, and perhaps most crucial, is the assumption that the hazard functions conditional on different values of the covariate vector are proportional to each other. This PH specification substantially restricts interdependence between the explanatory variables and the lifetime in determining the hazard (Gill and Schumacher, 1987; Kiefer, 1988; Neumann, 1997).

Proportionality of hazards is not consistent with the crossing hazards or converging/ diverg-

[6]This Section borrows heavily from Therneau and Grambsch (2000), and also Andersen *et al.* (1993).

ing hazards phenomena frequently observed in empirical studies; see Stablein *et al.* (1981), Han and Hausman (1990), Manton *et al.* (1991), Caplehorn and Bell (1991) and Liu *et al.* (2007) for some examples from biomedicine, economics and demography. Further, the assumption may even be unreasonable from the point of view of relevant theory. In many applications in the medical field, one expects the prognostic relevance of some covariates to decay, or even disappear, in the long run (Gill and Schumacher, 1987; Therneau and Grambsch, 2000); evidence of such decay can be found, for example, in Pocock *et al.* (1982), Champlin *et al.* (1983) and Begg *et al.* (1984). Predictions of non-proportional hazards can also be found in economic theory. For example, Mortensen (1977) and Burdett (1979) developed theoretical models where unemployment benefits have different effects on the hazard from unemployment as the spell lengthens; using British data, Atkinson *et al.* (1984) and Narendranathan and Stewart (1993) find evidence of such non-proportional hazards in unemployment duration. Similarly, in this thesis, we develop a model of firm exits through competing routes of bankruptcy and acquisition (Bhattacharjee *et al.*, 2008a, 2008b), where adverse macroeconomic effects decay with the age of the firm.

Since violation of the PH assumption leads to inaccurate inference on covariate effects and the baseline hazard (Breslow *et al.*, 1984; Stablein and Koutrouvelis, 1985; Schemper, 1992; Tubert-Bitter *et al.*, 1994; Hsieh, 1996), testing the PH model has been an area of active research. The main focus of this thesis is in developing methods to detect departures from the proportional hazards assumption, as well as modeling and estimation when proportionality does not hold.

The second main assumption, that of no frailty, is violated when there are omitted covariates. With scalar multiplicative frailty[7], u, representing the combined effect of unobserved covariates independent of included regressors, we have the standard frailty model

$$\lambda\left(t|\underline{X}_i(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}^T.\underline{X}_i(t)\right].u_i, \qquad u_i \epsilon \left(0, \infty\right) \overset{iid}{\sim} F_U.$$

[7] Frailty is alternatively called unobserved heterogeneity, particularly in the econometrics literature. We find this definition slightly ambiguous; while unobserved heterogeneity can be both in the nature of random and fixed effects, frailty usually refers to multiplicative random effects unobserved heterogeneity in the Cox regression model. We use the term frailty throughout this thesis.

The model[8] was first used in the econometrics literature by Lancaster (1979) and Nickell (1979), and Vaupel *et al.* (1979) introduced it in demography. With unrestricted individual level frailty, the above model can be expressed as a linear transformation model

$$\ln \Lambda_0(t) = -\underline{\beta}^T.\underline{X}(t) - U + \varepsilon, \qquad U \sim F_U. \tag{1.11}$$

where $\ln \Lambda_0(t)$ is a positive-valued but arbitrary increasing function, ε has the usual extreme value distribution and the log-frailty $U = \ln(u)$ has an arbitrary distribution.

The presence of multiplicative frailty invalidates partial likelihood inference, both on the covariate effects and the baseline hazard function (Lancaster; 1985, 1990; Struthers and Kalbfleisch, 1986; Henderson and Oman, 1999); see Hougaard (2000) and van den Berg (2001) for discussion. Research, based on both simulations (Bretagnolle and Huber-Carol, 1988; Baker and Melino, 2000) and empirical applications (Heckman and Singer, 1984b; Trussell and Richards, 1985; Hougaard *et al.*, 1994; Keiding *et al.*, 1997), also suggests that inference is usually sensitive to the choice of the frailty distribution. Therefore, Kiefer (1988) argues that it may be preferable to model frailty using the nonparametric approach of Heckman and Singer (1984a, 1984b), where a sequence of discrete multinomial distributions with a progressively increasing number of support points is used to approximate the unknown frailty distribution.

Further, like the proportionality assumption, the assumption that frailty is absent is also frequently violated in applications, and is often even unjustifiable from theoretical considerations. The shared frailty model, where individuals are clustered *a priori* based on the value of their shared but unobserved frailty, is commonly used in biomedical applications (Lin, 1994; Andersen *et al.*, 1999; Hougaard, 2000). However, many economic applications have strong reasons, both theoretical and empirical, to anticipate unobserved heterogeneity at the individual level. The work in this thesis incorporates such univariate frailty, either with a known frailty distribution or with a completely nonparametric treatment of unobserved heterogeneity.

The Cox regression model incorporates two further important features: (a) multiplicative separability of the effect of the baseline hazard and of each individual covariate, and (b) the exponential link function. By representing violation of proportionality through interaction

[8] Also known as the Mixed Proportional Hazards (MPH) model.

between the lifetime and the covariate effects, the work in this thesis develops a richer model of covariate dependence as compared with multiplicative separability. This line of inference follows the work of Mau (1986), who demonstrated the use of the additive hazard model (Aalen, 1980) in detecting possible time dependent effect of a covariate. Further, Pettitt and Bin Daud (1990) show that, when covariate effects are not very large, the hazard regression model with time varying coefficients provides a first order Taylor approximation to other popular alternatives – the additive hazard and the accelerated failure time models; see Therneau and Grambsch (2000) for further discussion.

The issue of finding adequate covariates with loglinear effects is highly specific to any application, and is therefore a matter of empirical modeling. Further, we do not directly discuss the problem of inferring on an appropriate functional form or transformation through which a covariate's effect is expressed in the regression model. We, however, take on board several contributions to this line of research, including Lagakos (1988), Lin *et al.* (1993), Grambsch *et al.* (1995) and Holländer and Schumacher (2006).

The above literature highlights the importance of the special features of the Cox regression model, particularly the proportionality and the no frailty assumptions, for obtaining useful inference on covariate effects. The presence of censoring exacerbates the effect of model misspecification, particularly when there are omitted covariates (Andersen *et al.*, 1996). Perhaps most importantly, the literature suggests extensions that would make the Cox regression model more useful for studying the prognostic relevance of various regressors.

As regards estimation of the baseline hazard function, the effects of misspecification are even more severe. First, under the partial likelihood approach, estimation of the baseline hazard function depends on the estimates of the covariate effects (1.7). Hence, any violation of assumptions that are crucial for covariate effects are also important for inference on the baseline hazard function. Second, and perhaps more importantly, several studies of real-life single-spell failure time data find that estimates of both the covariate effects and the shape of the baseline hazard function depends crucially on appropriate modeling of frailty; see, for example, Heckman and Singer (1984a, 1984b), Hougaard *et al.* (1994) and Keiding *et al.* (1997). This is true even when the overall fit of the model does not change with inclusion of frailty in any substantial way. The crucial nature of the no frailty assumption can also be easily seen within a model without covariates, which satisfies the PH assumption by definition, even though the observed hazard rates will not be proportional across the different levels of frailty. In fact, identification of the unknown frailty distribution in the proportional hazards frailty model (1.11) comes from this nonproportionality of the observed conditional hazard functions (see Hougaard, 1991; van den Berg, 1992; and Keiding, 1998).

Unfortunately, the presence of frailty can often be confused with interaction between the failure time and the explanatory variables (Elbers and Ridder, 1982; Aalen, 1994). Andersen et al. (1993, pp. 550–551) present similar evidence, in that a model omitting an important covariate appears to exhibit evidence of non-proportional hazards. In a similar vein, Abbring and van den Berg (2007) show how tests for proportional hazards can be adjusted to test for the no frailty hypothesis when the PH assumption holds. This observation of the close relationship between non-proportionality and unobserved covariates is a major motivation behind our treatment of frailty in this thesis. Specifically, we will consider hazard regression models for single-spell failure time data with potentially non-proportional hazards and frailty having an unknown distribution.

1.2.3 Goodness-of-fit tests of the PH assumption

Given the crucial nature of the proportionality assumption, an important focus of research has been in the development of analytical and graphical tests for the proportional hazards hypothesis. Most of these tests are based on goodness-of-fit, testing for proportional hazards either against an omnibus alternative or an alternative within which the Cox regression model is nested[9].

Many of the available tests, both graphical and analytical, set the Cox regression model as the null hypothesis and use an omnibus alternative. For example, Kay (1977), Crowley and Hu (1977) and Crowley and Storer (1983) used cross-plots of the estimated martingale residuals

$$\widehat{M_i}(t) = N_i(t) - \int_0^t Y_i(s).\exp\left[\underline{\widehat{\beta}}^T.\underline{X}_i(s)\right].d\widehat{\Lambda_0}(s),$$

either against a covariate value or against a set of order statistics from the unit exponential

[9] The review here is largely based on Therneau and Grambsch (2000) and Andersen *et al.* (1993).

distribution. Lagakos (1981) proposed a method based on permuted rank statistics of such residuals. Andersen (1982) developed graphical methods similar to Kay (1977) and a goodness-of-fit test that involves the estimation of a piecewise constant baseline hazard. Cox (1979) suggested a graphical method using total cumulative baseline hazard between k successive order statistics; if the Cox model is correct, this statistic has a Gamma distribution with shape parameter k and unit scale parameter. Schoenfeld (1982) suggested plotting the differences between the actual value of the covariate for the individual who fails at time t_j and the expected value over all individuals at risk at that time. Consider the score process for the i-th individual

$$U_i(\beta, t) = \int_0^t \left[\underline{X}_i(s) - \overline{x}\left(\underline{\beta}, s\right)\right].dM_i(s),$$

where $M_i(t)$ is the counting process martingale

$$M_i(t) = N_i(t) - \int_0^t Y_i(s).\exp\left[\underline{\beta}^T.\underline{X}_i(s)\right].d\Lambda_0(s)$$

and

$$\overline{x}\left(\underline{\beta}, s\right) = \frac{\sum Y_i(s).\exp\left[\underline{\beta}^T.\underline{X}_i(s)\right].\underline{X}_i(s)}{\sum Y_i(s).\exp\left[\underline{\beta}^T.\underline{X}_i(s)\right]}.$$

Then the Schoenfeld residual at the k-th failure time is given by

$$\begin{aligned} s_k &= \int_{t_{k-1}}^{t_k} \sum_i \left[\underline{X}_i(s) - \overline{x}\left(\widehat{\underline{\beta}}, s\right)\right].d\widehat{M_i}(s) \\ &= \int_{t_{k-1}}^{t_k} \sum_i \left[\underline{X}_i(s) - \overline{x}\left(\widehat{\underline{\beta}}, s\right)\right].dN_i(s). \end{aligned}$$

This is a useful way to detect departures from the proportional hazards model. Other useful graphical tools for assesing the PH assumption under the Cox regression model have been developed in Arjas (1988) and O'Quigley (2003).

There are many analytical tests of the Cox regression model against the omnibus alternative, often based on the graphical tools for model validation. Schoenfeld (1980) proposed a goodness-of-fit statistic for the Cox proportional hazards model by partitioning the subjects into mutually exclusive regions based on their covariate values. The goodness-of-fit statistic is then calculated as a sum of squared differences between the observed and predicted number of failures in these regions. Other goodness-of-fit statistics are proposed by Kalbfleisch and Prentice (1980), Andersen (1982), Lancaster (1983), Gray and Pierce (1985), Lancaster and Chesher (1985), Arjas (1988), Barlow and Prentice (1988), Hjort (1990), Lin and Wei (1991), McKeague and Utikal (1991), Chen and Wang (1991), Henderson and Milner (1991), Andersen et al. (1993, pp. 545-550), Li and Doss (1993), and Grambsch and Therneau (1994); the testing procedures differ mainly in the notions of goodness-of-fit (usual χ^2, Kolmogorov-Smirnov or Cramér-von Mises) and the definition of residuals (martingale or generalised residuals, Schoenfeld residuals, Arjas (1988) type residuals, etc.). Nagelkerke et al. (1984) propose a goodness-of-fit test based on the autocovariance of successive contributions to the log likelihood. Similar tests were proposed by Therneau et al. (1990), Horowitz and Neumann (1992) and Lin et al. (1993) based on cumulative sums of martingale-based residuals or on maximum deviation of the score process from the zero line. A different approach to the omnibus alternative is to assume a more general model within which the Cox regression model can be nested. Such nested tests have been considered by Aranda-Ordaz (1983), O'Quigley and Moreau (1984, 1986), Moreau et al. (1985, 1986). Excellent summaries of some of these procedures have been given by Andersen et al. (1993) and Fleming and Harrington (1991).

There are two main aspects where the above tests of the proportionality assumption can be improved. First, the available choices are often too extreme; the omnibus tests have very low power against many alternatives of interest, while the nested tests are very limited in the dimensions along which departures from the PH model is allowed. Therefore, there is need to develop a trade-off between the two approaches. Second, and more importantly, these tests provide little assistance in understanding the nature of the regression relationship when the null hypothesis of proportional hazards is rejected. Graphical tools may be useful in these situatuions to identify suitable alternatives (see, for example, Sengupta 1996). We address these two issues in this thesis. We develop new methods, both analytical and graphical, that allow for modeling the nature of nonproportionality in the two-sample (binary covariate) setup, as well as when the covariates are continuous and time-varying.

1.2.4 Order restrictions on covariate dependence

As opposed to broad alternatives like the omnibus alternative or alternatives considered in the general nested tests discussed above, it is often of interest to understand the nature of departures from the proportionality assumption. As discussed earlier, the PH assumption essentially implies that hazard functions conditional on different covariate values are proportional to each other. It is therefore of interest to identify which of the explanatory factors have nonproportional effects and to examine the nature of the covariate effect.

A useful approach is based on checking proportional hazards within the context of a hazard regression model

$$\lambda\left(t|\underline{X}(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}(t)^T.\underline{X}(t)\right], \tag{1.12}$$

where proportionality corresponds to the condition that the time varying regression coefficient process, $\underline{\beta}(t)$, is constant over time: $\underline{\beta}(t) \equiv \underline{\beta}$. Estimators for $\underline{\beta}(t)$ have been developed using the histogram sieve (Murphy and Sen, 1991; Gore *et al.*, 1984), spline models (Hess, 1994; Abrahamowicz *et al.*, 1996), local partial likelihood (Valsecchi *et al.*, 1996), penalized partial likelihood (Zucker and Karr, 1990; Gray, 1992; Hastie and Tibshirani, 1993; Verweij and van Houwelingen, 1995), kernel-weighted partial likelihood (Tian *et al.*, 2005), local linear estimation (Cai and Sun, 2003) and recursive estimation using Schoenfeld residuals (Winnett and Sasieni, 2003). Starting from any initial consistent estimator for the time-varying coefficients, Martinussen and Scheike (2002) and Martinussen *et al.* (2002) propose a one-step estimation procedure for the cumulative coefficient $B(t) = \int_0^t \beta(s)ds$.

In an important contribution, Grambsch and Therneau (1994) show that a plot of the Schoenfeld residuals for covariate l, properly scaled, gives a first order approximation to $\beta_l(t), l = 1, \ldots, p$. Expressing $\beta_l(t)$ as a regression on some function $g_l(t)$ (either a known fixed function of time or a specified predictable process)

$$\beta_l(t) = \beta_l + \theta_l.g_l(t),$$

they derive a test for proportionality where $\theta_l = 0$ constitutes the null hypothesis of proportional hazards. Andersen *et al.* (1993, pp. 539-545) describe an alternative strategy where stratified Cox regression models are fitted based on a pre-specified partition of the covariate space, and proportionality is checked by a test that the baseline hazard for all the strata are equal. Marzec and Marzec (1997) follow the histogram sieve approach of Murphy and Sen (1991) in defining an increasingly fine partition of the time scale and using the Arjas (1988) methodology to develop a test that the covariate effect is constant across time. Kvaløy and Neef (2004) and Kraus (2007) develop similar tests based on cumulative sums of Schoenfeld residuals that test for the constancy of the covariate effect. Scheike and Martinussen (2004) base their test of the PH assumption on the fact that under the null hypothesis, the time-integrated covariate effect process estimated using the methodology developed in Martinussen *et al.* (2002), should be a straight line through the origin.

The above tests help in understanding of the nature of covariate effects in situations when the PH assumption does not hold. Further, by assuming the time-varying coefficient model, these tests also obtain higher power than the earlier tests for proportionality. However, many empirical applications involve specific order restrictions on the covariate effect which is not explicitly incorporated in these tests. For example, it is often of interest to explore whether the hazard rate for one level of the covariate increases in lifetime, relative to another level (*i.e.*, the hazard ratio increases/ decreases with lifetime), particularly when the covariate is discrete (two-sample or k-sample setup).

As opposed to omnibus alternatives, it is therefore often of interest to consider more specific situations where the covariate effect is order-restricted. In the two-sample setup, Wei (1984), Gill and Schumacher (1987) and Deshpande and Sengupta (1995) have constructed analytical tests of the PH hypothesis against the alternative of 'increasing hazard ratio'; Lin (1993) extends the Gill and Schumacher (1987) test to the Cox regression model The alternative hypothesis accomodates the commonly observed phenomenon of 'crossing hazards', and is a useful ordered alternative to the proportional hazards model in the two-sample setup. Empirical evidence of such ordering is abundant in the literature on empirical survival analysis, demography and economic duration models. Besides, this framework permits more explicit modeling of the covariate effect in the two sample setup, when the covariate effect is time-varying and ordered.

The test developed in Gill and Schumacher (1987) is particularly interesting, and motivates much of the work in this thesis. For censored data in a two-sample setup, they develop a test for the hypotheses

$$\begin{aligned}\mathbb{H}_0 \quad &: \quad \lambda_2(t)/\lambda_1(t) = \theta \text{ for some positive } \theta \\ \text{versus } \mathbb{H}_1 \quad &: \quad \lambda_2(t)/\lambda_1(t) \neq \theta \text{ for any positive } \theta,\end{aligned}$$

where $\lambda_1(.)$ and $\lambda_2(.)$ are the hazard functions in the two samples. Gill and Schumacher (1987) construct their test statistic based on the intuition that under $\mathbb{H}_0$, the contrast between two different estimators of the hazard ratio, θ, should be close to zero. They choose the two estimators as generalised rank estimators (Begun and Reid, 1983; Andersen, 1983) using different predictable weight functions. The test is particularly useful in detecting ordered departures from the PH assumption (i.e., when the hazard ratio is monotone), for which it is unbiased when the ratio of the weight functions is monotone.

The Gill and Schumacher (1987) test is also motivated through a graphical tool developed by Lee and Pirie (1981), the so-called trend function $\gamma(u)$

$$\gamma(u) = \Lambda_2\left(\Lambda_1^{-1}(u)\right), \qquad u \epsilon \left(0, \Lambda_1\left(\tau\right)\right),$$

where $\Lambda_1(.)$ and $\Lambda_2(.)$ are the cumulative hazard functions in the two samples, and $\Lambda_1^{-1}(.)$ is the functional inverse of the cumulative hazard function in sample 1. Gill and Schumacher (1987) show that their test statistic is a weighted measure of the area between the straight line through the origin and the empirical trend function.

The Lee-Pirie plot is a powerful tool to graphically detect proportionality of hazards as well as different kinds of partial orders in failure time distributions[10]. It is a straight line through the origin under proportionality – $\mathbb{H}_0$ in this case. Whenever the hazard ratio is monotonically increasing[11] (in other words, the failure time distribution in sample 2 is convex ordered with respect to sample 1, denoted $T_2 \underset{c}{\succ} T_1$), the Lee-Pirie plot is convex. The converse is also true – the plot is concave whenever $\lambda_2(t)/\lambda_1(t)$ is monotonically decreasing (concave ordering,

[10] Convex ordering and star ordering (and their duals – concave and negative star ordering respectively) are two important partial orders in this context. See Kalashnikov and Rachev (1986) and Sengupta and Deshpande (1994) for definition and further discussion of their properties.

[11] Throughout this thesis, the word ‘increasing’ would mean ‘non-decreasing’, and ‘decreasing’ would mean ‘non-increasing’.

denoted $T_1 \succ_c T_2$). The property of convexity of a function from $[0,\infty)$ to $[0,\infty)$ is a special case of a weaker property called star-shapedness. If the trend function is star-shaped, then the survival distribution of sample 2 is star-ordered with respect to that of sample 1. This happens if and only $\Lambda_2(t)/\Lambda_1(t)$ is increasing (Sengupta and Deshpande, 1994). The above two concepts of partial ordering of failure time distributions are very useful in applications, and represent meaningful ordered alternatives to proportionality.

The Gill and Schumacher (1987) test provides a logical starting point for the work in this thesis for two important reasons. First, unlike the goodness-of-fit tests discussed earlier, this test has demonstrated unbiasedness against ordered alternatives where the hazard ratio is either increasing or decreasing (convex/ concave ordering). Further, examination of the Lee-Pirie plot in combination with rejection of the null hypothesis of PH provides additional information on the nature of covariate dependence. This additional information can be used to model the ordered nature of covariate dependence more precisely. Second, as discussed in Section 1.1.1, Gill and Schumacher (1987) present an application of their methodology to data comparing surgery and medical treatment for patients with chronic stable angina (Detre *et al.*, 1977). While their tests fail to reject the null hypothesis of proportionality, the Lee and Pirie (1981) empirical trend plot (Figure 5 in Gill and Schumacher, 1987) show evidence that the hazard functions are not proportional. The plot is not convex but apears to be star-shaped. This suggests monotone cumulative hazard ratio, which is a weaker order than the monotone hazard ratio.

We extend the Gill and Schumacher (1987) work on testing the PH model with respect to a binary covariate in two ways. First, motivated by the above example, we develop censored data tests of proportional hazards in two samples against the alternative hypothesis of 'increasing ratio of cumulative hazards' (star ordering of the two samples). Second, we extend the Gill and Schumacher (1987) test and the Lee and Pirie (1981) plot to the case of two competing risks; the test against monotone ratio of cumulative hazards can also be extended in a similar way to the competing risks situation. These developments are useful in many applications. The alternative hypothesis of 'increasing ratio of cumulative hazards' provide an explanation for the phenomenon of 'crossing hazards' often observed in applications. In fact, in the empirical literature on survival analysis, convex-ordering/star-ordering of one sample with respect to another in the two-sample setup, or one cause of failure to another in the competing risks setup, as well as their duals (the concave-ordering/ negative-star-ordering hypotheses), can be useful for modeling the ordered nature of covariate effects. Empirical evidence of such ordering are abundant in the literature on survival analysis, demography and economic duration models.

The above tests and graphical tools are potentially useful in analysing lifetime/ duration data because, not only do they detect departures from proportionality, they also provide further clues about the nature of covariate dependence. However, their practical usefulness is limited by the fact that many of the important covariates in biomedical/ economic applications are continuous in nature (Horowitz and Neumann, 1992). Similar ordered departures are also common and potentially meaningful alternatives to the PH model in the case of continuous covariates. If, for example, the time-varying coefficient corresponding to a covariate X is increasing in age, the distribution of the lifetime T conditional on a higher value of the covariate (x_2) would be convex ordered with respect to the lifetime distribution conditional on a smaller covariate value (x_1). Notationally expressed as

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

this provides a useful notion of ageing order with respect to a continuous covariate. The higher the covariate, the faster the ageing of the individual – a situation which is empirically not an uncommon experience.

In biomedical applications, such monotonically time-dependant covariate effects have been noted in the literature, both under additive hazard models (Aalen, 1980; Mau, 1986) and multiplicative hazard models (Anderson and Senthilselvan, 1982; Andersen *et al.*, 1993). In Section 1.1.2, we have discussed evidence of ageing order in data on survival with malignant melanoma (Andersen *et al.*, 1993) and unemployment durations (Jayet and Moreau, 1991). Decay or even disappearence of covariate effects with time (age) has been observed in several other medical applications. For example, Sather *et al.* (1981), in studying survival from childhood acute lymphoblastic leukemia, observe that the strong prognostic effect of lymphocyte count at diagnosis disappears with time. Gore *et al.* (1984), in a study of breast cancer, provide another demonstration of decay in covariate effect. Similarly, O'Quigley and Natarajan (2004) observe that the effect of histology grade, one of the main factors affecting survival and recurrence rates of breast cancer, have its influence significantly dimisnished with time. While the literature addresses time-varying coefficients in various ways, order restrictions on covariate effects for continuous covariates has not been discussed.

This thesiscontributes in this area in several ways. First, we develop a suggestion in Fleming and Harrington (1991) and suggest several notions of ordered departure from proportionality with respect to continuous covariates. We use these notions to propose tests for the PH hypothesis against such ordered departures. As in the two sample case, these tests are based on comparing estimates of the cumulative baseline hazard functions conditional on different covariate values. Second, we propose the time-varying coefficients model as an alternative hazard regression model under which such departures can be studied. The model's usefulness in studying order restricted covariate effects is highlighted, with special focus on continuous covariates. Third, we consider estimation of hazard regression models with continuous covariates under ordered departures from the PH relation. Here, we propose biased bootstrap methods such as data tilting (Hall and Presnell, 1999; Hall and Huang, 2001) and local adaptive bandwidths (Brockmann *et al.*, 1993; Schucany, 1995; Hermann, 1997) for order restricted inference; local adaptive bandwidths are also closely related to SiZer maps (Chaudhuri and Marron, 1999, 2000).

1.2.5 Order restrictions on ageing

In the literature, the Cox regression model has been used mostly to study the prognostic effect of the covariate(s) on the hazard rate, leaving the baseline hazard function $\lambda_0(t)$ (1.4) completely unspecified. In fact, an important feature of the partial likelihood approach (1.6) is that inferences on the covariate effects can be drawn, while the baseline hazard is treated as an infinite-dimensional nuisance parameter. Indeed, the flexibility to leave $\lambda_0(t)$ completely unrestricted is a major advantage of the Cox regression model over parametric hazard regression models, in that it provides robustness against violations of any maintained assumptions on the baseline hazard.

Therefore, the Cox regression model offers the possibility of inference on the shape of the baseline hazard function, though this line of enquiry is largely unexplored in empirical studies. Estimates of $\lambda_0(t)$ are typically used either to test for the proportionality assumption by stratification over the range of covariate values (as in Andersen *et al.*, 1993, pp. 539–545), or to predict survival probabilities. While in many applications, the baseline hazard function is expected to be constant (exponential regression), systematic departures from this pattern is often observed in practice (see, for example, Andersen *et al.*, 1993, pp. 533–535; Baltazar-Aban and Peña, 1995). Such departures may be due to omitted covariates or other kinds of model misspecifications, or may reflect genuine underlying patterns of ageing in the conditional baseline hazard functions. In many applications, it is therefore of interest to understand ageing properties structural or inherent in the baseline hazard, after accommodating covariate effects in an appropriate way. For example, in the popular passive learning economic model of firm dynamics (Jovanovic, 1982; Lippman and Rumelt, 1982), hazard rates of firm exits are often non-increasing with age conditional on the main covariate, size (for further discussion, see Pakes and Ericson, 1998).

Ageing properties like the above may be conveniently studied using the notions of positive and negative ageing in reliability theory. The most commonly used classes describing notions of positive ageing are increasing failure rate (IFR), increasing failure rate in average ($IFRA$), new better than used (NBU); these are defined as follows. Let T be a non-negative (failure time) random variable with survival function $\overline{F}(t) = P\left[T > t\right]$ and hazard rate $\lambda(t)$ and assume, for simplicity, that $\overline{F}(0) = 1$. Then T is IFR if

$$\lambda(t) \uparrow t \iff \frac{\overline{F}(s+t)}{\overline{F}(t)} \downarrow t \text{ ..for all } \; s \geq 0;$$

T is $IFRA$ if

$$\frac{1}{t}\int_0^t \lambda(s)ds \uparrow t \iff \overline{F}(\alpha t) \geq \overline{F}^{\alpha}(t) \text{ ..for all } \; t \geq 0, 0 < \alpha < 1;$$

and T is NBU if

$$\overline{F}(s+t) \leq \overline{F}(s).\overline{F}(t) \text{ ..for all } \; s, t \geq 0.$$

The corresponding negative ageing classes (decreasing failure rate (DFR), decreasing failure rate in average ($DFRA$) and new worse than used (NWU)) are similarly defined. There are also weaker notions of ageing like new better than used in expectation ($NBUE$), harmonic new better than used in expectation ($HNBUE$), $\mathcal{L}$-class and decreasing mean residual life ($DMRL$), as well as their duals.

The above ageing notions have been very useful in reliability theory; see Barlow and Proschan (1975) for detailed discussion of their properties. There are several reasons why these classes of distributions are useful for characterising ageing in the context of hazard regression models. First, these ageing classes all include the exponential distribution on the boundary, and further most of them form a nice sequence of nested classes for positive ageing

$$Exponential \subset IFR \subset IFRA \subset NBU \subset NBUE \subset HNBUE \subset \mathcal{L}\text{-}class$$

as well as for negative ageing

$$Exponential \subset DFR \subset DFRA \subset NWU \subset NWUE \subset HNWUE \subset \overline{\mathcal{L}}\text{-}class.$$

This structure makes tests for the Exponential distribution against these ageing classes useful for inference on the nature of ageing present in the data. A large literature has, therefore, evolved on such tests; see Klefsjö (1983) and Doksum and Yandell (1984) for good reviews of the literature and Ahmad (2001) a recent contribution. These tests can be adapted to the hazard regression context, as in Chang and Chung (1998).

Second, the above notions of ageing suggest partial orders of failure time distributions. For example, the IFR and $IFRA$ classes have important connections with convex and star ordering respectively and their duals (Sengupta and Deshpande, 1994); similar connections exist between DFR and concave ordering and $DFRA$ and negative star ordering. For two failure time distributions F and G,

$$\begin{array}{lll} F \underset{c}{\prec} G \iff F \circ G^{-1} \text{ is } IFR; & \qquad & F \underset{c}{\succ} G \iff F \circ G^{-1} \text{ is } DFR \\ F \underset{*}{\prec} G \iff F \circ G^{-1} \text{ is } IFRA; & \qquad & F \underset{*}{\succ} G \iff F \circ G^{-1} \text{ is } DFRA. \end{array}$$

As we have discussed, it is useful to test the nature of partial order of lifetime distributions using notions of positive and negative ageing. Therefore, the above ageing classes provide a unified framework where order restrictions on both covariate dependence and ageing can be characterised and empirically studied.

Finally, this framework provides a convenient way to integrate study of ageing properties of hazard functions conditional on any covariate values with the shape of the baseline hazard function. This follows from the observation that, if the baseline distribution is $IFR/\ DFR/\ IFRA/\ DFRA/\ NBU/\ NWU$, then the failure time distribution at other values of the covariate also has the same ageing property, provided the covariate effect is not time-varying[12]. This is because these properties correspond to various geometric shapes of the cumulative hazard function $\Lambda(t) = -\ln\left[\overline{F}(t)\right]$, which continue to hold after multiplication of the function by a scalar. Thus, looking into the ageing property of the baseline hazard amounts to looking into that of an entire class of distributions over *all* covariate values. The PH assumption in the above argument is not very restrictive; if nonproportionality is present, the relevant covariates can be interacted with histogram sieves (Murphy and Sen, 1991) appropriately constructed to reflect the time-varying nature of coefficients.

Because of the above reasons, the proposed framework is convenient and useful for inference on order restrictions on ageing, once covariate dependence has been appropriately modeled. Though not in the context of order restrictions on covariate effects, inference on ageing in hazard regression models has been studied in a couple of previous contributions to the literature. Peña (1998) develops inference on the baseline hazard function, by considering the goodness-of-fit problem of testing whether $\lambda_0(.)$ is equal to some specified hazard rate function. He uses this methodology to test for exponentiality in the baseline hazard function against the omnibus alternative. In research more closely related to our work, Chang and Chung (1998) develop an estimator for monotone baseline hazard function under the Cox regression model. However, ageing properties in the baseline hazard function in the presence of potential nonproportionality in the hazards remains largely unexplored in the literature.

In this thesis, we advance research in this area in two ways. First, we use tests for the exponential distribution to identify departures along the dimension of specific ageing classes. These tests are applied to baseline cumulative hazard functions estimated after taking into account

[12] Note that the covariate can be time-varying, but all covariates should have proportional hazard effects.

order restrictions on the nature of covariate dependence. Second, in addition to order restrictions related to nonproportionality, it is of interest to build in order restrictions on ageing in the estimation of hazard regression models. However, inference under multiple order restrictions using biased bootstrap methods turns out to be computationally very challenging. We develop a Bayesian modeling framework to understand these ageing properties, in the presence of order restrictions on covariate dependence and unrestricted multiplicative frailty.

1.2.6 Individual level frailty

In many applications of hazard regression models, there is reason to suspect the influence of unobserved random variables or frailty. Since Lancaster (1979) and Vaupel *et al.* (1979), there has been general recognition of the need to account for frailty in models for lifetime and duration data. Failure to consider unobserved random covariates causes the estimated hazard rate to decrease more with the duration than the hazard rate of a randomly selected member of the population. Moreover, the estimated proportional effect of explanatory variables on the population hazard rate is smaller in absolute value than that on the hazard rate of the average population member and decreases with the duration; see van den Berg (2001) for further discussion.

In the PH model, a scalar frailty variable uncorrelated with the included covariates is usually assumed to have multiplicative effect on the hazard rate. Inference in this mixed proportional hazards (MPH) model,

$$\lambda\left(t|\underline{X}_i(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}^T.\underline{X}_i(t)\right].u_i, \qquad u_i\epsilon\left(0,\infty\right)\overset{iid}{\sim}F_U, \tag{1.13}$$

is complicated by the fact that the popular counting process methodology does not apply here (Petersen *et al.*, 1996). Recent research, for example in Spiekerman and Lin (1998) and Kosorok *et al.* (2004), has developed an approach based on empirical process theory for the asymptotic analysis of frailty models.

In many applications, particularly in biomedicine or when there are repeated failure time data, it is reasonable to assume a classification of the data based on the magnitude of the unobserved frailty variable; these are called shared frailty models. However, in other applications, particularly in economic duration data, the frailty variable is unique to each individual. Given the nature of most of the important applications considered in this thesis, we focus primarily on univariate (or individual level) frailty, and do not discuss shared frailty models in detail. We note, however, that Spiekerman and Lin (1998) have proposed estimation, based on "*quasi-partial likelihood*" estimating equations, of the Cox PH model in a multivariate duration model setting with shared frailty. A special case of this setup is a competing risk model with unrestricted frailty at the individual level, but the frailty random effect is shared between the two competing risks; see also Wei *et al.* (1989). In this thesis, we use this approach for inference on business failure in UK and US firms through competing exit routes of bankruptcy and acquisition (Bhattacharjee *et al.*, 2008a, 2008b).

In this thesis, our primary interest lies in inference on order restrictions on covariate dependence and ageing in the presence of individual level frailty. Inference under both these types of order restrictions rest crucially on good estimates of the baseline cumulative hazard function. Below we review research on identifiability and estimation under univariate frailty, focusing mainly on the cumulative baseline hazard function[13]. Within the class of individual-level frailty models, we distinguish between estimation under a known (parametric) frailty distribution and nonparametric treatment of frailty.

Known distribution of individual frailty

Several parametric continuous distributions for individual-level frailty have been considered in the literature: the gamma frailty (Lancaster, 1979; Vaupel *et al.*, 1979); the inverse Gaussian frailty (Hougaard, 1984), the positive stable frailty (Hougaard, 1986), the log-normal frailty (McGilchrist and Aisbett, 1991), the power variance frailty (Aalen, 1988), the uniform frailty (Lee and Klein, 1988) and the threshold frailty (Lindley and Singpurwalla, 1986). While theory provides little insight about the appropriate frailty distribution, the choice of gamma frailty in most empirical studies is driven by analytical and computational ease[14].

Nielsen *et al.*, (1992) showed that the partial likelihood estimator of Cox (1972) can be generalized to the frailty model with Gamma distributed frailty. Their estimator is semiparametric

[13] The review is partly based on van den Berg (2001).

[14] Abbring and van den Berg (2007) provide partial justification for this choice, showing that the frailty distribution of survivors asymptotically converges to the gamma distribution. This insight is, however, not particularly useful in our applications, where inference on survival at lower durations is important.

in that it uses parametric specifications of the regression function and the frailty distribution, but an unrestricted baseline hazard. Han and Hausman (1990) and Meyer (1990) proposed estimators assuming a piecewise-constant baseline hazard and frailty having a gamma distribution. In an important recent contribution, Kosorok *et al.* (2004) provide a rigorous foundation for inference within a wide class of parametric frailty models, and propose robust estimates of the cumulative baseline hazard function, the regression parameters and the parameter describing the frailty distribution. They extend their results to one-jump frailty intensity models with time dependent covariates, including the gamma, the lognormal and the generalized inverse Gaussian frailty intensity models.

However, both simulations (Bretagnolle and Huber-Carol, 1988; Baker and Melino, 2000) and empirical studies (Heckman and Singer, 1984b; Trussell and Richards, 1985; Hougaard *et al.*, 1994; Keiding *et al.*, 1997; Hausman and Woutersen, 2005) with these models reveal that the estimates are rather sensitive to the assumed frailty distribution. Hence, we take the view that, in the absence of a strong justification for an assumed parametric form, nonparametric specification of the distribution of unobserved heterogeneity is preferable.

Arbitrary distribution of individual-level frailty

Before discussing estimation and inference in models with unrestricted frailty distribution, it is important to consider identifiability of such models. Infact, constructive identification can also point to useful methods of inference in these models. Elbers and Ridder (1982) showed that the standard frailty model (1.13) is semiparametrically identified if there is minimal variation in the regression function. A single indicator variable in the regression function suffices to recover the regression function, the baseline hazard, and the distribution of the frailty, provided that frailty is independent of the included covariates. Semi-parametric identification means that semiparametric estimation is feasible; however, their proof of identifiability is not constructive, and therefore does not suggest an estimation method.

Heckman and Singer (1984a) derived the non-parametric maximum likelihood estimator of the MPH model with a parametric baseline hazard and regression function. Based on prior work by Laird (1978) and Lindsay (1983a, 1983b), they approximate the frailty distribution by a discrete mixture of degenerate distributions. Starting with the no frailty case (single mass point degenerate distribution), the number of support points is sequentially increased until the in-sample fit cannot be improved any further. This method is very useful in that it approximates the nonparametric frailty distribution using an increasing sequence of parametric distributions. However, the rate of convergence and the asymptotic distribution of the Heckman and Singer (1984a) estimator are not known. Another estimator that does not require the specification of the unobserved heterogeneity distribution was suggested by Honoré (1990). This estimator assumes a Weibull baseline hazard and only uses very short durations to estimate the Weibull parameter. The main limitation of both these estimators lies in the strong parametric assumptions imposed on the baseline hazard function (ageing).

Under an arbitrary heterogeneity distribution, Melino and Sueyoshi (1990) provide a constructive proof of identifiability in the MPH model for the continuous regressor case. Their proof relies heavily on the observed duration density at $t = 0$, and therefore cannot be used to devise an attractive estimation strategy. Kortram *et al.* (1995) provided a constructive proof for the two-sample (binary regressor) case (i.e., where βx can take only two distinct values), and Lenstra and Van Rooij (1998) used this to construct a consistent nonparametric model estimator for the two-sample case. This idea is potentially useful; however, the asymptotic distribution of their estimator of the baseline cumulative hazard function is unknown.

Horowitz (1996, 1999) consider the representation of the MPH model in a linear transformation model form (1.11),

$$\ln \Lambda_0(t) = -\underline{\beta}^T.\underline{X}(t) - U + \varepsilon, \qquad\qquad U \sim F_U,$$

where $\ln \Lambda_0(t)$ is an increasing function, log-frailty U has an arbitrary distribution that is independent of the covariates, and ε has the usual extreme value distribution; see Cheng *et al.* (1995) for related work. Since U has an arbitrary distribution, so does $-U + \varepsilon$, and hence this is a standard transformation model. Under some additional smoothness assumptions, Horowitz (1999) proposes a nonparametric kernel-based estimator for the regression coefficients $\underline{\beta}$, the baseline cumulative hazard function $\Lambda_0(t) = \displaystyle\int_0^t \lambda_0(s).ds$ and the distribution function of the scalar frailty F_U, and derives asymptotic distributions. It is important to note that this model is identified only upto a location and scale transformation (see Horowitz, 1999), and therefore has to be normalised before estimation. The location normalisation sets $\ln \Lambda_0(t)$ to zero at a given duration t:

$$\ln \Lambda_0(t) \equiv 0 \Rightarrow \Lambda_0(t) \equiv 1.$$

The model also requires a scale transformation, which requires setting the absolute value of the regression coefficient for a given covariate to a given fixed positive number. Unfortunately, the Horowitz (1999) estimates are sensitive to the choice of bandwidths, rendering this methodology difficult to employ in practice. In an alternative approach, Hausman and Woutersen (2005) consider discrete failure time data and treat the frailty distribution as nuisance parameters. They propose estimators for the other parameters of the MPH model under an unspecified frailty distribution, based on the maximum rank correlation methodology of Han (1987) and Sherman (1993).

McCall (1996) establish identifiability of the standard frailty model with time-varying coefficients

$$\lambda\left(t|\underline{X}_i(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}(t)^T.\underline{X}_i(t)\right].u_i, \qquad u_i \epsilon\,(0,\infty) \overset{iid}{\sim} F_U, \tag{1.14}$$

under the condition that at least one of the included regressors has unbounded support. The result is not constructive, but McCall (1996) suggests using the histogram sieve estimator (Murphy and Sen, 1991) to estimate the time-varying coefficients, in combination with the Heckman and Singer (1984a) methodology for nonparametric estimation of the unknown frailty distribution.

The existing literature on frailty models reviewed above does not consider order restrictions on covariate dependence or ageing. However, there is some empirical work to suggest that the issue of time-varying coefficients may be confounded with omitted random variables. For example, Andersen *et al.* (1993, Examples 7.3.1 and 7.3.4) achieve a good fit to a decaying treatment effect by introducing a frailty parameter. This thesis augments the existing literature in two ways. First, we develop tests for the proportional hazards assumption against order restricted covariate effects in the presence of frailty. We consider individual-level frailty with completely arbitrary distribution, and also the simpler case of.shared frailty. Second, we develop Bayesian inference where there are order restrictions on covariate dependence and ageing, as well as individual level frailty modeled using a degenerate mixture distribution.

1.2.7 Other hazard regression models

As we have discussed, a major issue with the use of the Cox regression model in empirical studies is that inferences are highly sensitive to the model's various assumptions, particularly proportional hazards and no frailty. This has encouraged development of many alternative hazard regression models; the main competitors are briefly discussed below[15]. It has, however, come to be generally acknowledged that an important advantage of the Cox regression model lies in that it offers simple semiparametric analysis of covariate effects. The main aim of our discussion will be to motivate extensions to the Cox regression model that would allow incorporation of different aspects of the association between covariates and failure time.

Additive hazards model

In contrast to the PH model, the additive hazards (AH) model (Aalen, 1978, 1980; Lin and Ying, 1994),

$$\lambda\left(t|\underline{X}_i(t)\right) = \lambda_0(t) + \underline{\beta}^T.\underline{X}_i(t), \tag{1.15}$$

specifies that covariates have additive rather than multiplicative effect on the hazard function. If the additive hazards model holds then the difference of hazards rates under constant covariates does not depend on the failure time.

As compared with the PH model, the main difference in inference arises from the property that the additive hazards model is not rank invariant; therefore, partial likelihood inference is no longer applicable. Using the counting process martingale approach, Lin and Ying (1994) obtain closed form estimators for the covariate effects, $\underline{\beta}$, and the baseline hazard, $\lambda_0(t)$.

Like the PH model, the additive hazards model also has the so-called 'absence of memory property' (Bagdonavičius and Nikulin, 2004), in that the hazard rate at the given time does not depend on the past values of time varying covariates. However, this is not a major issue – if such dependence were postulated, appropriate covariates that incorporate such effects can be constructed.

The choice between the proportional and additive hazards models is often empirical, but may also be guided by relevant theory. Nevertheless, in many applications, the additive versus

[15] Part of our discussion borrows from Fleming and Lin (2000) and Bagdonavičius and Nikulin (2004).

multiplicative model choice is not very clear. Lin and Ying (1996) and Sasieni (1996) have proposed models that provide a synthesis between the two.

Such amalgam models are similar in spirit to hazard rgression models with time varying coefficients. In fact, one of the ways the additive hazards model has been motivated is through its use in detecting nonproportional hazards (Mau, 1986). However, a wide range of nonproportional hazards situations can be accomodated within the multiplicative hazard framework by considering time varying coefficients. Further, the additive hazards model is well-approximated by the time varying coefficients model if covariate effects are not very large (Pettitt and Bin Daud, 1990).

Accelerated failure time (AFT) models

The accelerated failure time (AFT) model was first considered in Bagdonavičius (1978) and was motivated by the idea that conditional hazard rates may incorporate the effect of past covariate values:

$$\overline{F}_{X(.)}(t) = \overline{F}_0\left(\int_0^t r\left[X(u)\right] du\right),$$

where $r(.)$ is a known decreasing function and the baseline survival function $\overline{F}_0(.)$ does not depend on the covariate values, $X(.)$. With constant covariates and the usual choice

$$r(x) = \exp\left[-\underline{\beta}^T.\underline{x}\right],$$

we obtain the standard AFT model

$$\overline{F}_{\underline{X}}(t) = \overline{F}_0\left(\exp\left[-\underline{\beta}^T.\underline{x}\right].t\right).$$

The above representation suggests the interpretation that the effect of the covariates is to transform the time scale: $t \to \exp\left[-\underline{\beta}^T.\underline{x}\right].t$ – hence the name accelerated failure time model (Cox and Oakes, 1984). For any given function $r(.)$, the model can be represented as a linear transformation model

$$h(t) = \underline{\beta}^T.\underline{X}_i + \varepsilon_i, \tag{1.16}$$

where $h(.)$ is a known monotone function and the errors ε_i are IID with known or unknown distribution; for the standard AFT model, $h(t) = \ln(t)$. Note that, if instead $h(.)$ were arbitrary and the errors followed an extreme value distribution, this would give the Cox regression model. The transformation model representation above (1.16) highlights an important advantage of the AFT model, in that frailty due to independent unobserved covariates can be easily accomodated by considering a completely unrestricted error distribution. On the other hand, the model can be rather restrictive because the function $h(.)$ (equivalently, $r(.)$) has to be completely specified *a priori*.

With censored data, several approaches have been proposed for the estimation and inference on the AFT model. Rank-based methods were developed, among others, in Tsiatis (1990), Wei *et al.* (1990), Lai and Ying (1991), Ying (1993) and Fygenson and Ritov (1994); Koul *et al.* (1981), Ritov (1990) and Lai and Ying (1991) have considered least squares based and M-estimation methods. Robins and Tsiatis (1992) make an useful extension to the AFT model, where time-varying covariates are included through the modified mapping $t \to \left(\int \exp\left[-\underline{\beta}^T.\underline{x}\right] du\right).t$. However, since the estimating functions are typically neither differentiable nor monotone, the methods are numerically complicated and difficult to implement, particularly when the number of covariates is large. Further, the covariance matrices of the estimators are rather difficult to obtain because they involve nonparametric estimation of the underlying unknown error distribution. Recently proposed counting process (Lin *et al.*, 1998) and rank-based (Jin *et al.*, 2003) methods address some of these issues.

In biostatistics, lack of robustness of the Cox regression model combined with methodological developments has led to a renewed interest in the AFT models for the analysis of single-spell failure time data. While the AFT model provides a convenient framework for dealing with unobserved heterogeneity, restrictive assumptions on $h(.)$ (closely related to the shape of the baseline hazard) is a serious mitigating factor; Bagdonavičius and Nikulin (2004) argue that the AFT model may be a good choice when the researcher has a good idea about the nature of ageing. Even when unobserved covariates are expected to be important, many practitioners may prefer the frailty model to the AFT model. This is particularly true if they are interested in either disentangling genuine ageing from the effect of frailty, or in quantifying the effect of covariates on the individual hazard as opposed to the observed hazard, with univariate failure time data. Finally, like the additive hazards model, the AFT model is similar to the model with age varying coefficients when covariate effects are not very large (Pettitt and Bin Daud, 1990).

Generalized proportional hazards model

Generalized Proportional Hazards (GPH) models (Bagdonavičius and Nikulin, 1999, 2004) are powerful alternatives to the PH and AFT models, and are also a potentially useful alternative to the time varying coefficients model. With constant covariates, the GPH models allow ratios of the hazard rates to be not only constant but also increasing or decreasing. They include AFT and PH models as particular cases. The models are defined by postulating that the hazard rate at any failure time is proportional not only to a function of the covariate at this moment and to a baseline rate, but also to a function of the probability of survival until that time (or, equivalently, to the cumulative hazard function):

$$\begin{aligned}\overline{F}_{X(.)}(t) &= G\left(\int_0^t r\left[X(s)\right] d\Lambda_0(s)\right),\\ \Lambda_0(t) &= \int_0^t \lambda_0(s)ds, \qquad G = H^{-1}, \qquad H(u) = \int_0^{-\ln(u)} \frac{dv}{q(v)},\end{aligned}$$

where $r(.)$ is a known function and H^{-1} denotes the inverse function of $H(.)$. The models obtained by completely specifying $q(.)$ are rather narrow in their applicability; more useful models are obtained by either by parametrizing $q(.)$ or by leaving it completely unspecified (Bagdonavičius and Nikulin, 2004).

The most useful property of these models is that they can accommodate a wide range of hazard ragression models considered in the literature. Particular cases are the PH model ($q(u) \equiv 1$) and the AFT model ($\lambda_0(u) \equiv \lambda_0$), with $r(x) = \exp\left[\underline{\beta}^T.\underline{x}\right]$ in either case. Further, assuming constant covariates and $r(x) = \exp\left[\underline{\beta}^T.\underline{x}\right]$, various families of models with monotone hazard ratios are obtained by specifying $q(u) = (1+u)^{-\gamma+1}$ (first GPH model), $q(u) = (1+\gamma u)^{-1}$ (second GPH model) or $q(u) = \exp(-\gamma u)$ (third GPH model); see Bagdonavičius and Nikulin (1999, 2004) for further discussion.

Bagdonavičius and Nikulin (2004) discuss the connection of these models with frailty models, models with time varying coefficients and additive multiplicative hazard models, as well as heteroscedastic hazard regression models (Hseih, 2001) and models with crossing survival functions. Bagdonavičius and Nikulin (1999, 2004) also discuss semi-parametric inference under the GPH models. The main limitations are, however, the lack of flexibility in dealing with time-varying covariates and unrestricted frailty distributions.

Time varying coefficients model

In this thesis, we mainly focus on the hazard regression model (1.14)

$$\lambda\left(t|\underline{X_i}(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}(t)^T.\underline{X_i}(t)\right].u_i, \qquad u_i\epsilon\,(0,\infty) \overset{iid}{\sim} F_U,$$

with time varying coefficients and individual level frailty. The covariates, $\underline{X_i}(t)$, if time varying, have a clear interpretation as the values of prognostic factors measured over time, so that $\underline{\beta}(t)$ is precisely identified as the regression effect of $\underline{X_i}(t)$ on the log hazard at failure time t. As discussed earlier, this model provides a simple framework that accommodates order restrictions on both covariate effects and ageing, and permits inference on the nature of these order restrictions. Besides, the model is identified under the assumption that one of the covariates has unbounded support (McCall, 1996).

The time varying coefficients model, without frailty, has been widely used in the biomedical literature for modeling covariate dependence under nonproportional hazards; for a representative selection, see Moreau *et al.* (1985), Zucker and Karr (1990), Liang *et al.* (1990), Murphy and Sen (1991), O'Quigley and Pessione (1991), Gray (1992), Hastie and Tibshirani (1993), Verweij and van Houwelingen (1995), Lausen and Schumacher (1996), Marzec and Marzec (1997), Martinussen *et al.* (2002) and Schieke and Martinussen (2004). Inference under the model is more complicated than the Cox regression model because the additional infinite dimensional parameter $\underline{\beta}(t)$ does not admit to a simple partial likelihood treatment. As discussed earlier, many different estimation methodologies have been proposed, of which the most convenient and attractive one is based on histogram sieves (Murphy and Sen, 1991). Sieve methods are typically used to estimate an infinite dimensional parameter (Grenander, 1981). The essence of the method is that a sequence of increasing subspaces (sieves) is used to estimate a large parameter space such that, asymptotically, the closure of the limiting subspace contains the original parameter space. The histogram sieve implementation (Murphy and Sen, 1991) assumes that $\underline{\beta}(t)$ is a time-varying function that is constant over consecutive intervals $L_1, L_2, \ldots, L_k$ spanning the sample space of the failure time variable T:

$$\underline{\beta}(t) = \underline{\beta}_1.\boldsymbol{I}\left\{t \in L_1\right\} + \underline{\beta}_2.\boldsymbol{I}\left\{t \in L_2\right\} + \ldots + \underline{\beta}_k.\boldsymbol{I}\left\{t \in L_k\right\},$$

where $\boldsymbol{I}\{.\}$ denotes the indicator function. The asymptotic setup is one where the partition $L_{1^{(n)}}, L_{2^{(n)}}, \ldots, L_{k^{(n)}}$ becomes finer and finer as more data become available. Given a partition and in the absence of frailty, the time varying coefficients can be estimated by maximising the Cox partial likelihood (1.6) with modified time varying covariates

$$\underline{X_i}(t).\boldsymbol{I}\left\{t \in L_1\right\}, \underline{X_i}(t).\boldsymbol{I}\left\{t \in L_2\right\}, \ldots, \underline{X_i}(t).\boldsymbol{I}\left\{t \in L_k\right\},$$

and the baseline hazard function can then be estimated in the usual way (1.7). In the presence of frailty with unrestricted distribution, we can combine the histogram sieve approach with either the Horowitz (1999) kernel based estimator, or with the discrete multinomial mixing distribution proposed in Heckman and Singer (1984a). In the second approach, the distinct mass-points $m_1 \equiv 1, m_2, \ldots, m_J$ as well as the corresponding probabilities $\pi_1, \pi_2, \ldots, 1 - \sum_{j=1}^{J-1} \pi_j$ $(0 < \pi_j < 1, j = 1, 2, \ldots, J)$ are estimated from the data.

One of the main challenges in inference on age-varying coefficients is that the estimates can be quite volatile and unsmooth, particularly when there are limited data in some time intervals. This issue prompted Martinussen and Scheike (2002) and Martinussen *et al.* (2002) to propose inference on the cumulative coefficients $B(t) = \int_0^t \beta(s)ds$. An alternative approach, proposed by Gamerman (1991), postulates a dynamic model for the time variation through a Markov structure; see Sargent (1997) for a refinement.

Overall, the time varying coefficients model is an useful framework combining many of the strengths of the Cox regression model with the possibility of nonproportional hazards. the main advantages of the approach lie in the complete flexibility in the patterns of duration dependence and ageing, as well as in allowing the presence of frailty. In this thesis, we will explore the usefulness of the model in studying order restrictions, both on covariate effects and in the shape of the baseline hazard function.

Discrete duration data regression models

Many real life applications, including some of the data analysed in this thesis, have reported failure times grouped into time intervals – days, months, years, etc. This motivates the use of discrete time hazard regression models, since models for continuous failure times are rendered inadequate because of the large number of ties. Discrete failure time models have a long history in biostatistics, and several models have been proposed in the literature. The two most popular models are:

(a) the grouped time version of the Cox PH model, also called the complementary log-log model or discrete PH model (Cox, 1972; Kalbfleisch and Prentice, 1973; Prentice and Gloeckler, 1978; Cox and Oakes, 1984)

$$\ln\left[-\ln\left\{1-h_j\left(\underline{X}\right)\right\}\right]=\underline{\beta}^T.\underline{X}+\gamma_j, \tag{1.17}$$

where the time intervals are indexed by $j = 1, 2, \ldots$and h_j denotes the discrete hazard rate in interval j (assumed constant over the interval); and

(b) the proportional odds, or the logistic hazard, model (Cox, 1972, 1975; Arjas and Haara, 1987; McCullagh and Nelder, 1989)

$$\ln\left[\frac{h_j\left(\underline{X}\right)}{1-h_j\left(\underline{X}\right)}\right]=\operatorname{logit}\left[h_j\left(\underline{X}\right)\right]=\underline{\beta}^T.\underline{X}+\alpha_j. \tag{1.18}$$

While the discrete PH model (1.17) assumes that latent continuous failure times have a proportional hazards specification but are grouped into intervals, the proportional odds model (1.18) offers a specific interpretation with regard to the relative odds of failure in period j conditional on survival upto the previous period. There are some important connections between the two models. Sueyoshi (1995) shows that, like the discrete PH model, the proportional odds model can also be consistent with an underlying continuous time PH model. In practice, the two models share similar ageing properties in the baseline hazard function and yield similar estimates of covariate effects, so long as the hazard rate is relatively small. On the other hand, Chen and Manatunga (2007) point out important differences between the two models and caution against use of the proportional odds model when the PH assumption holds.

Sengupta and Jammalamadaka (1993) develop a counting process formulation of the proportional odds and discrete PH models without any assumptions on the origin of the discrete data. While this approach is potentially useful, asymptotic results can only be derived undera strict *iid* assumption. This is not reasonable in some of our applications. Following Jenkins (1995), the effect of a scalar unobserved covariate can be incorporated by estimating a model where the frailty distribution is chracterised by the nonparametric approach of Heckman and Singer (1984a).

The above literature has considered several alternatives to the Cox PH model; many of these models are aimed at characterising the nature of nonproportionality in the data. However,.the literature is not very informative when there are potential order restrictions or omitted covariates. We advance the literature on regression models for failure time data by proposing the use of the model with time varying coefficients (1.14) in these situations. Specifically, we argue that this model is useful for studying both covariate dependence and ageing under order restrictions and in the presence of unrestricted frailty. We also demonstrate the usefulness of the model by considering several real life applications. In the case of discrete failure time data, we take a similar view, advocating the use of the discrete PH model with a nonparametric multinomial frailty distribution, allowing the number of mass points to increase sequentially. Finally, we demonstrate how this model can be used to address the question as to what explains nonproportionality better – time varying coefficients or frailty?

1.2.8 Bayesian semiparametric inference

Bayesian semiparametric modeling and inference in the context of hazard regression models, with order restrictions on covariate dependence and ageing and in the presence of frailty, offers several important advantages over frequentist inference (Sinha and Dey, 1997; Sinha *et al.*, 1999). First, Bayesian methods enable exact small-sample inference from moderately sized data sets on parameters of interest which are themselves either high-dimensional or in the presence of infinite dimensional nuisance parameters. This is important in this thesis, where parametric assumptions are not imposed on either time variation in the covariate effects or the baseline hazard function. Second, powerful computational tools enable remarkably complex Bayesian models to be fitted with relative ease, and facilitate the choice of suitably parsimonious models. This is particularly true of semiparametric hazard regression models where nonparametric frequentist estimation of the frailty distribution presents severe computational challenges, especially in a model with a flexible baseline hazard function (Campolieti, 2001). The computational issues are compounded even further when we attempt joint inference on order restrictions in covariate effects and ageing in the presence of unobserved covariates. Third, prior beliefs about order restrictions on parameters can often be expressed in a way that places no support on part of the parameter space. In this case, the posterior also exhibits the same property; we exploit this useful property of Bayesian inference for studying order restrictions on covariate dependence and ageing.

While there is little prior work on order restricted Bayesian modeling and inference in hazard regression models, there has been some research on several related areas. We survey related literature briefly with a view towards placing our work within the context of the literature and highlighting the distinctive nature of our approach. The survey is partly based on several review papers: Sinha and Dey (1997), Ibrahim *et al.* (2001) and Damien (2005).

Bayesian inference in the Cox PH model

Semiparametric approaches to Bayesian inference in hazard regression models usually assume the Cox proportional hazards model (1.4)

$$\lambda\left(t|\underline{X}(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}^T.\underline{X}(t)\right],$$

where $\lambda_0(.)$ is some baseline hazard function, $\underline{X}(t)$ is a vector of (possibly time varying) covariates, and $\underline{\beta}$ is a vector of corresponding regression coefficients. Various Bayesian formulations of the model differ mainly in the nonparametric specification of $\lambda_0(t)$.

A model based on an independent increments gamma process was proposed by Kalbfleisch (1978) who studied its properties and estimation. In the context of multiple event time data, Sinha (1993) considered an extension of Kalbfleisch's (1978) model for $\lambda_0(t)$. The proposal assumes the events are generated by a counting process with intensity given by a multiplicative expression similar to (1.4), but including an indicator of the censoring process, and individual frailties to accommodate the multiple events occurring per subject.

Several other modeling approaches based on the Cox PH model have been studied in the literature. Laud *et al.* (1998) consider a Beta process prior for $\Lambda_0(t)$ and propose an MCMC implementation for full posterior inference. Nieto-Barajas and Walker (2002a) propose their flexible Lévy driven Markov process to model $\lambda_0(t)$, and allowing for time dependent covariates. Full posterior inference is achieved via substitution sampling.

Other Bayesian survival data models

While Bayesian formulation of the Cox proportional hazards model has been rather narrow in the specification of the baseline hazard function, several other models have been used more generally in Bayesian survival analysis. These models can be used in the context of hazard regression models to specify the baseline hazard or baseline cumulative hazard functions.

Many stochastic process priors that have been proposed as nonparametric prior distributions for survival data analysis belong to the class of neutral to the right (NTTR) processes. A random probability measure $F(t)$ is an NTTR process on the real line, if it can be expressed as $F(t) = 1 - exp(-Y(t))$, where $Y(t)$ is a stochastic process with independent increments, almost surely right-continuous and non-decreasing with $P\{Y(0) = 0\} = 1$ and $P\{\lim_{t \longrightarrow \infty} Y(t) = 1\} = 1$ (Doksum 1974). The posterior for a NTTR prior and i.i.d. sampling is again a NTTR process. Ferguson and Phadia (1979) showed that for right censored data the class of NTTR process priors remains closed, i.e., the posterior is still a NTTR process.

NTTR processes are used in many approaches that construct probability models for the hazard function $\lambda(t)$ or the cumulative hazard function $\Lambda(t)$. Dykstra and Laud (1981) define the extended gamma process as a model for $\lambda(t)$, generalizing the independent gamma increments process studied in Ferguson (1973). Dykstra and Laud (1981) show that the resulting function $\lambda(t)$ is monotone, making it useful for modeling ageing in the nature of monotone hazard rates.

An alternative Beta process prior on $\Lambda(t)$ was proposed by Hjort (1990), where the baseline hazard comprises piecewise constant independent beta distributed increments. This model is closed under prior to posterior updating as the posterior process is again of the same type. Full Bayesian inference for a model with a Beta process prior for the cumulative hazard function using Gibbs sampling can be found in Damien *et al.* (1996). Walker and Mallick (1997) specify a similar structure for the prior, but use independently distributed gamma hazards.

While the above models for $\Lambda(t)$ are based on independent hazard increments $\{\lambda_j\}$, considering dependence provides a different modeling perspective. A convenient way to introduce dependence is a Markovian process prior on $\{\lambda_j\}$. In a model with time-varying coefficients, Gamerman (1991) proposes the following characterization for the baseline hazard function: $\ln(\lambda_j) = \ln(\lambda_{j-1}) + \varepsilon_j$ for $j \geq 2$, where $\{\varepsilon_j\}$ are independent with $E(\varepsilon_j) = 0$ and $Var(\varepsilon_j) = \sigma^2 < \infty$. Later, Gray (1994) used a similar prior process but directly on the hazards $\{\lambda_j\}$, without the log transformation. A further generalization involving a martingale process was proposed in Arjas and Gasbarra (1994). More recently, Nieto-Barajas and Walker (2002b) proposed a model based on a latent process $\{u_j\}$ such that $\{\lambda_j\}$ is included as

$$\lambda_1 \longrightarrow u_1 \longrightarrow \lambda_2 \longrightarrow u_2 \longrightarrow \ldots$$

and the pairs (u, λ) are generated from conditional densities $f(u|\lambda)$ and $f(\lambda|u)$ implied by a specified joint density $f(u, \lambda)$. The main idea is to ensure linearity in the conditional expectation: $E(\lambda_{j+1}|\lambda_j) = a_j + b_j\lambda_j$. Both the gamma process of Walker and Mallick (1997) and the discrete Beta process of Hjort (1990) are obtained as special cases of the above construction.

Frailty

Accounting for unobserved covariates is important in the analysis of hazard regression models. With single survival data and individual-level frailty, estimation of the frailties is not possible but their distribution can be inferred on. Clayton (1991) and Walker and Mallick (1997) both consider Bayesian inference in the Cox proportional hazards model with gamma frailty distribution, but with different priors on the baseline hazard function. Sinha (1993) also assumes gamma distributed frailties, but in multiple event survival data. Extensions of this model to the case of positive stable frailty distributions and a correlated prior process with piecewise exponential hazards can be found in Qiou et al. (1999).

In its ability to deal with potentially large number of latent variables, the Bayesian framework is convenient for nonparametric modeling of individual level frailty. Based on repeated failures data, Bhattacharjee *et al.* (2003) and Arjas and Bhattacharjee (2003) have proposed a hierarchical Bayesian model based on a latent variable structure for modeling unobserved heterogeneity; the model is very powerful and shown to be useful in applications.

Order restricted inference

The literature on order restricted Bayesian inference, with restrictions either on the shape of the baseline hazard function or on the nature of covariate depence, is indeed very sparse. Notable contributions to the literature in this area are Arjas and Gasbarra (1996), Sinha *et al.* (1999), Gelfand and Kottas (2001) and Dunson and Herring (2003).

Arjas and Gasbarra (1996) develop models of the hazard rate processes in two samples under the restriction of stochastic ordering. They define their prior on the space of pairs of hazard rate functions; the unconstrained prior in this space consists of piecewise constant gamma distributed hazards which incorporate path dependence. The constrained prior is then constructed by restricting to a subspace on which the maintained order restriction holds. In their work, Arjas and Gasbarra (1996) propose a coupled and constrained Metropolis-Hastings algorithm for posterior elicitation based on the order restriction and also for Bayesian evaluation of the stochastic ordering assumed in the analysis. For the same problem, Gelfand and Kottas (2001) developed an alternative prior specification and computational algorithm. The Bayesian model in Arjas and Gasbarra (1996), in combination with the general treatment of Bayesian order restricted inference (for example, in Gelfand et al., 1992), is related to the current chapter.

Sinha *et al.* (1999) develop Bayesian analysis and model selection tools using interval censored data where covariate dependence is possibly nonproportional. They model the baseline hazard function using an independent Gamma prior and time varying coefficients are endowed with a Markov type property $\beta_{k+1}|\beta_1, \ldots, \beta_k \sim N(\beta_k, 1)$ .While Sinha *et al.* (1999) do not explicitly consider order restrictions either on covariate dependence or on ageing, they provide Bayesian inference procedures to infer on the validity of the proportional hazards assumption.

In research closely related to this thesis, Dunson and Herring (2003) consider an order restriction on covariate dependence in hazard regression models. They develop Bayesian methods for inferring on the restriction that the effect of an ordinal covariate is higher for higher levels of the covariate; in other words, they conduct inference on trend in conditional hazard functions. We work with restrictions on covariate dependence which are different in two respects. First, the covariate is continuous in our case and not categorical. Second, our order restriction is related to convex/ concave partial ordering of conditional hazard functions rather than trend. Consequently, we express our constraints in terms of monotonic time varying coefficients, and propose a different methodology for Bayesian inference.

Our work extends the literature on Bayesian modeling and inference in hazard regression models by considering order restrictions in covariate dependence and ageing as well as individual level frailty. We propose Bayesian models in which order restrictions on both the covariate effects and the shape of the baseline hazard can be studied. Since our applications are based on single failure per subject data, we use a latent variable structure for inferring on the frailty distribution rather than the latent variables themselves. We model frailty in two different ways. First, we divide the subjects into groups and incorporate fixed effects unobserved heterogeneity across these different groups. Second, we model individual level frailty in a more nonparametric tradition (Heckman and Singer, 1984a) by introducing a sequence of multinomial frailty distributions with increasing number of support points; for a related Bayesian implementation, see Campolieti (2001).

1.3 Outline of the thesis

As discussed above, this thesis makes several contributions to order restricted inference and modeling in hazard regression models. In addition to order restrictions on covariate dependence, we consider order restrictions on ageing and individual level frailty with unrestricted distributions.

1.3.1 Testing proportionality with respect to a binary covariate

In Chapter 2, we develop new tests for the proportional hazards assumption in the two-sample setup where existing tests are either too general (like the omnibus tests) or the alternative hypothesis is too strong (like monotone hazard ratio). Specifically, our method test for proportionality against the monotone cumulative hazard ratio alternative, which is a weaker partial order (Sengupta *et al.*, 1998). Asymptotic distribution of the test is derived and small sample properties studied. The new tests as well as existing inference procedures for the two-sample case are also extended to the competing risks framework (Sengupta and Bhattacharjee, 1994). Several examples are considered.

1.3.2 Testing proportionality with respect to continuous covariates

In Chapter 3, we extend partial orders in two samples, like convex/ concave ordering or star/ negative star ordering, to the continuous covariate case. Tests for proportionality against such ordered alternatives are developed, combining evidence from two-sample tests based on failure times conditional on different pairs of covariate values (Bhattacharjee, 2007a)[16]. Asymptotic distributions are derived and finite sample performance of the tesst are explored. Usefulness of the tests is demonstrated using real life applications.

1.3.3 Estimation under order restrictions on covariate dependence

The context of Chapter 4 is estimation and modeling of order restricted covariate effects using hazard regression models. Following Bhattacharjee (2003), we argue that the time varying coefficients model is useful for the study of order restricted covariate effects. Under this model, we propose estimation of hazard regression models with continuous covariates under ordered departures using various biased bootstrap techniques (Bhattacharjee, 2004a). We find that kernel estimation with locally adaptive bandwidths is particularly useful for such order restricted inference and modeling. An application to data on firm exits shows decay with age in the adverse effect of macroeconomic instability on firm survival. We also discuss the potential usefulness of the time varying coefficients model for studying order restrictions on ageing as well as incorporating the effect of unobserved covariates.

1.3.4 Testing proportionality with unrestricted frailty

Chapter 5 extends our testing problem to a model with unrestricted frailty. The tests developed earlier can be extended to frailty with known distribution and to shared frailty models. In this chapter, we develop tests for proportional hazards against ordered alternatives, where the distribution of frailty is completely unrestricted (Bhattacharjee, 2007b)[17]. As an extension, we also develop tests for the related hypothesis of no covariate effect with respect to continuous covariates. The asymptotic properties of the tests are studied and their use is demonstrated using real applications.

[16] A previous version was circulated as Bhattacharjee and Das (2001).
[17] A previous version was circulated as Bhattacharjee (2004b).

1.3.5 Order restrictions on both covariate dependence and ageing

In Chapter 6, we develop Bayesian inference in hazard regression models under potential order restrictions on both covariate dependence and ageing, and there is multiplicative frailty with arbitrary distribution (Bhattacharjee and Bhattacharjee, 2007). We find strong evidence of decay in the effect of macroeconomic instability on firm exits with age. However, evidence on any ageing pattern in the baseline hazard is very weak. While the data demonstrates fixed effects heterogeneity at the industry level, evidence of frailty is not found. The inference highlights the strengths of the proposed modeling framework.

1.3.6 Applications to firm dynamics

Three applications of the methods developed in the thesis to study of firm exits is presented in Chapter 7. In this chapter, we use our methodology and framework to make important contributions to the theory and application surrounding an important research problem in applied economics – the study of firm dynamics.

First, following Bhattacharjee *et al.* (2008a), we develop a theoretical framework for studying macroeconomic influences on firm exits through dependent competing routes – bankruptcy and acquisitions. Our empirical work shows the importance of macroeconomic stability, as well as evidence of ordering in both covariate dependence and ageing among listed firms in the UK. In addition to order restrictions on covariate dependence, we find evidence of negative ageing of the new worse than used type in the shape of the baseline hazard, and relative ageing in the nature of convex ordering between two baseline hazards of two competing risks.

Second, in Bhattacharjee *et al.* (2008b), we take a similar approach to data on US firms and find that macroeconomic influences have been less important since the introduction of a new bankruptcy code (called Chapter 11) in 1979. The work has important policy implications with regard to the design and conduct of legislative procedures related to bankruptcy codes. Additional issues addressed in the work relate to unobserved heterogeneity and to robustness from truncation potentially dependent on the exit process.

Finally, in Bhattacharjee (2007c)[18], we consider a model developed in Bhattacharjee *et al.*

[18] A previous version of the paper was circulated as Bhattacharjee (2005).

(2006) where individuals in the labour market make an endogenous decision to become entrepreneurs and the survival of their firms is conditioned on their, potentially partly unobserved, human capital. The model indicates potential unobserved covariates as well as order restricted covariate effects. Applied work incorporating such completely arbitrary frailty is usually very demanding on the data and on computing facilities, and such inference is therefore often not very useful – discrete life history data offers considerable simplification. Data on new firms incorporated by French entrepreneurs demonstrate the importance of segregating these two issues, both theoretically and empirically.

1.3.7 Real data and applications

Throughout the thesis, we retain a strong applied focus and develop various applications, particularly from biomedicine and economic duration data. Further, as discussed above, the thesis makes particularly important contributions to developing one particular area of applications – firm exits. We add to the literature on firm dynamics by developing an economic models for the competing risks of bankruptcy and acquisitions, and emphasize the distinction between unobserved heterogeneity and order restrictions on covariate dependence. Order restrictions on both covariate dependence and ageing are important in our estimated models. The various data sets and applications used in the thesis are summarised below.

Ovarian cancer data (Fleming *et al.*, 1980)

The two-sample censored data, drawn from a study at Mayo Clinic, on patients having limited low-grade (Stage II, 15 patients) or high-grade (Stage IIIA, 20 patients) ovarian carcinoma are reported and analysed in Fleming *et al.* (1980); further analysis are reported in Gill and Schumacher (1987) and Deshpande and Sengupta (1995). The purpose of the analyses are to study the dependence of time to progression of disease on the grade.

Our analysis in Chapter 2, based on analytic and graphical procedures, suggests that the cumulative hazard ratio of high-grade to low-grade tumour has an increasing trend. This supports earlier findings that the hazard ratio is increasing.

Veteran's administration data (Detre *et al.*, 1977)

Gill and Schumacher (1987) analyse these two-sample data, on survival times of patients receiving coronary artery bypass graft surgery and of patients receiving a conservative medical treatment, based on a controlled clinical trial in chronic stable angina. They fail to reject the hypothesis of proportional hazards against a monotone hazard ratio alternative.

We do not include any new analyses of the data. However, based on graphical evidence, we argue (Chapter 2; see also discussion in Chapter 1, Section 1.1.1) that the departure from oproportionality may be weaker. We use this example to motivate our tests of proportionality against the monotone ratio of cumulative hazards alternative.

Unemployment duration data (Han and Hausman, 1990)

These are US data on a sample of 1051 heads of households between the ages of 20 and 65, from waves 14 and 15 (1980 and 1981) of the Panel Study of Income Dynamics (PSID), on duration of unemployment in weeks and whether the reason for the end of the spell is a new job, recall, or censoring; see Han and Hausman (1990) for detailed discussion of the sampling scheme including potential sample selection biases. 58 per cent of the spells end in recall, 23 per cent in a new job, while the remaining 19 per cent are censored. An important feature observed in previous work is significantly high exits from unemployment at 26 and 39 weeks, which correspond to exhaustion points of unemployment insurance benefits.

We analyse competing risks of recall to old job and new job, and find evidence of nonproportional hazards in the nature of concave ordering (Chapter 2).

Mice cancer data (Hoel, 1972)

The competing risks data pertain to 99 male mice examined after exposure to 300 rads of radiation. There are 60 deaths due to cancer and 39 deaths attributed to other causes; there is no censoring in the data. Previous analyses of the data are reported in Hoel (1972) and Bagai *et al.* (1989a).

Our analysis of the data in Chapter 2 uncover evidence that the risk due to cancer increases in the long run relative to the other competing causes. The nature of departure from proportionality is in the nature of monotone hazard ratio. This evidence adds additional dimension to previous findings that the hazard due to cancer is smaller than the other hazards combined.

Survival with malignant melanoma (Drzewiecki and Andersen, 1982)

These data pertain to 205 patients (148 of these are censored) with malignant melanoma (cancer of the skin) on whom a radical operation was performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark in the period 1962-77. Andersen *et al.* (1993) include detailed analyses of the data, and identify tumour thickness as one of the main prognostic factors for survival.

One of the main motivations of our work on continuous covariates (particularly Chapters 3 and 4) is the observation, in Andersen *et al.* (1993), of order restrictions in the covariate effect of tumour thickness; see also Chapter 1, Section 1.1.2. We analyse these data in Chapter 3 and find evidence of order restrictions in that the covariate effect of tumour thickness decreases with time since surgery. There is also limited evidence that the above evidence is strongly supported only for large tumours. In Chapter 4, we also find strong support for the above order restrictions and obtain biased bootstrap estimates of time varying coefficients.

Data on Strike Durations (Kennan, 1985)

The data pertain to durations of 566 contract strikes in the U.S., each involving 1000 workers or more, beginning during the period January 1968 to December 1976. Several authors have analysed these data, including Kennan (1985), Kiefer (1988), Horowitz and Neumann (1992), and Neumann (1997), a major focus of the analysis being on theeffect of the business cycle (measured by production index) on strike duration. Given that, strike durations are also known to exhibit some seasonal effects (Neumann, 1997), we use only the data on 292 strikes beginning in the first half of each year (none of these failure times are censored).

Our analysis of these data in Chapter 3, both graphical and analytical, show evidence of order restrictions on the effect of production index on the hazard rate of strike termination.

Child mortality in rural India (Bhalotra and Bhattacharjee, 2001)

The data are extracted from the National Family Health Survey 1992-93 for the study of mortality outcomes of children in rural India. In particular, we are interested in understanding the relationship between mortality hazards and mother's age at child birth, which is one of the most important (physiological) determinants of child mortality.

Our analysis of the data, reported in Chapter 3, provides evidence of order restrictions in the nature of covaraite dependence for mother's age. Mortality decreases with mother's age upto about 24 years and declines thereafter; the magnitude of the prognostic effect, however, declines with age of the child. The data are reanalysed in Chapter 5, this time with emphasis on changepoint trend in the covariate effect of mother's age.

Listed UK firms

The dataset pertains to firms quoted in the UK, constructed by combining the Cambridge-DTI, DATASTREAM and EXSTAT databases of firm accounts. The relevant failure time variable is age since listing, where listing dates are compiled by merging these data with the London Share Price Database. The main objective of our analyses are to evaluate the impact of macroeconomic fluctuations on business exits due to competing risks of bankruptcy and acquisition, which requires data running over several business cycles. The combined firm level accounting data provides an unbalanced panel of about 4,100 UK listed companies over the period 1965 to 2002. Data on macroeconomic conditions, macroeconomic stability and firm specific accounting information are used in the analyses. There were 206 instances of bankruptcy and 1858 acquisitions among 48,046 firm years over the 38 year period.[19] In terms of life history analysis, the data are right-censored and left-truncated[20].

The data are used at several places in the thesis. In Chapters 4 and 6, order restricted covariate effects of macroeconomic instability on bankruptcy hazard is studied (see also Chapter 1, Section 1.1.2), and Chapter 5 includes studies on whether aggregate Q has any impact on bankruptcies, and on testing for proportional hazards against monotone alternatives. Chapter

[19] A firm that has irretrievably entered the path to bankruptcy may, in a precursor phase of distress, stop publishing accounts one or two years prior to actually being declared bankrupt. From the point of view of econometrically modelling bankruptcy it is sensible to reassign the date of "real" bankruptcy to the year of last published accounts when the firm has been declared legally bankrupt within a 2 year period. Our assignment of a bankruptcy to a particular point in time captures the date of economic bankruptcy rather than declaration of bankruptcy. We assign accounting data for each company fiscal year to the calendar year that covers the majority of the accounting year corresponding to the fiscal year.

[20] The data used pertain to years, since 1965, during which each company is listed in the London Stock Exchange. Hence, for each company, the available data are left-truncated, and do not pertain to the entire period that it is listed.

7 includes integrated study of several issues including comparison with US firms. The above analyses draw on methods developed in Chapters 2, 3 and 4, as well as frequentist inference on estimated baseline hazard functions using notions of ageing and ageing orders.

Listed US firms

These data are constructed by matching the Compustat accounting database with the CRSP database to identify all listed firms[21] and to extract listing dates. This gives an unbalanced panel of about 13,700 US industrial and commercial firms over the period 1969 to 2000. There were 566 exits due to bankruptcy and 2,529 acquisitions in around 133,000 firm years over the 32 year period. Failure time data, measuring the postlisting lifetime of each firm, are augmented by annual indicators of macroeconomic conditions, as well as firm and industry-specific factors. These variables constitute the time-varying covariates used to explain exit-probabilities or hazard rates. The lifetime data are left-truncated, randomly right censored by potentially dependent competing risks, and the covariates explaining the nature of the cause-specific hazards are time-varying.

These data are analysed in Chapter 7 to understand the impact of macroeconomic instability on competing risks of exit due to bankruptcy and acquisitions, using methods developed in Chapters 2, 3 and 4. Further, comparison with exits of UK firms is conducted and the effect of Chapter 11 introduction is studied. The analysis also includes empirical investigation of potentially dependent left truncation and unrestricted frailty shared between competing risks.

French new firms

The data are extracted from the SINE 94[22] survey, which was conducted by the French National Institute of Statistical and Economic Studies [23] in 1994. It provides qualitative data on entrepreneurship and, more specifically, variables pertaining to the entrepreneur and the circumstances in which entrepreneurship occurred. A second survey carried out in 1997 (SINE 97) gives information about the situation of the same firms (closed down or still running; when

[21] Listed on the NYSE/AMEX, NASDAQ, Over-the-Counter or any of the regional exchanges (Boston, Midwest, Montreal, Pacific or Philadelphia).

[22] "Système d'informations sur les nouvelles entreprises" (Information system on new firms)

[23] Insee (Institut National des Statistiques et des Etudes Economiques).

closed down, the date of the discontinuation). The surveyed units belong to the private productive sector in the field of industry, building, commerce and services. These data are merged with an individual-level survey database on French entepreneurs, to extract information on the entrepreneur's education level, previous situation in the labor market, financial endowments etc.

The data are analysed in Chapter 7 to evaluate the relative importance of time varying coefficients and frailty due to unobserved human capital on the survival of firms. We take the framework developed in Chapters 3, 4 and 5, particularly in the context of discrete failure time data.

Chapter 2

Testing for the Proportionality of Hazards in Two Samples Against Ordered Alternatives

2.1 Chapter summary

A number of tests of the proportional hazards hypothesis have been proposed in the past. Previous researchers have proposed tests geared specially for the alternative hypothesis of "increasing hazard ratio", keeping in mind the case of crossing hazards (see Gill and Schumacher, 1987; Deshpande and Sengupta, 1995; and Lin, 1991). This alternative may be too restrictive in many situations. In this chapter, based on Sengupta *et al.* (1998), we develop a test of the proportional hazards model for the weaker "increasing *cumulative* hazard ratio" alternative. The work is motivated by a data analytic example given by Gill and Schumacher (1987) where their test fails to reject the null hypothesis of proportional hazards even though the faster ageing of one group is quite apparent from a plot. The normalised test statistic proposed here has an asymptotically normal distribution under either hypothesis. We also present two graphical methods related to our analytical test. We also adapt these methods, as well as those of Gill and Schumacher (1987), to the competing risks setup, where one cause-specific hazard is sometimes oberved to overtake another.

2.2 Introduction

The proportional hazards (PH) model has played an instrumental part in data analysis in such areas as survival analysis, reliability, economics, demography and environmental studies. The validity of the PH assumption in a two-sample problem may be checked through one of the traditional graphical methods proposed, among others, by Cox (1972), Kay (1977), Andersen (1982) or Arjas (1988) (see Sengupta (1995) for a review). Several analytical tests are also available; see, for example, Schoenfeld (1980), Andersen *et al.* (1982), Wei (1984), Nagelkerke *et al.* (1984), Breslow *et al.* (1984) and Ciampi and Etezadi-Amoli (1985). Gill and Schumacher (1987) and Deshpande and Sengupta (1995) proposed analytical tests of the PH hypothesis against the alternative of "increasing hazard ratio", which may account for the "crossing hazards" phenomenon (Lin, 1991).

If F_1 and F_2 are two life distributions on the positive real line with hazard rates λ_1 and λ_2 and cumulative hazard functions Λ_1 and Λ_2, respectively, then the condition Λ_1/Λ_2 increasing is equivalent to the composition $\Lambda_1 \circ \Lambda_2^{-1}$ being convex on $[0, \infty)$. Using this equivalence, Lee and Pirie (1981) suggested plotting an estimator of Λ_1 (e.g. the Nelson-Aalen estimator) against that of Λ_2. It is expected that the graph would be approximately convex when the hazard ratio is increasing, and a straight line through the origin when the ratio is constant.

The "increasing hazard ratio" alternative may be too strong in some cases. Consider the situation where the hazard rate Λ_2 has jump discontinuities. The ratio Λ_1/Λ_2 cannot be increasing unless Λ_1 also has a jump of adequate size at every point of discontinuity of Λ_2. On the other hand, the consistency of an "omnibus" test is not guaranteed. It would be nice to have a test which is consistent for a weaker alternative hypothesis.

We consider a weaker form of relative ageing represented by the condition "$\Lambda_1 \circ \Lambda_2^{-1}$ is star-shaped", that is, $\Lambda_1 \circ \Lambda_2^{-1}$ intersects any straight line passing through the origin at most once and from below. Convexity is a special case of star-shapedness. Sengupta and Deshpande (1994) showed that the above condition holds if and only if the cumulative hazard ratio (CHR) Λ_1/Λ_2 is an increasing function. Thus, the plot of Λ_1 against Λ_2 is star-shaped if and only if Λ_1/Λ_2 is increasing. The empirical plot of Lee and Pirie (1981) should also be approximately star-shaped when the CHR for the two groups is increasing. Such a phenomenon is indeed observed in the case of the Veterans' Administration data (Detre *et al.*, 1977). The plot given by Gill

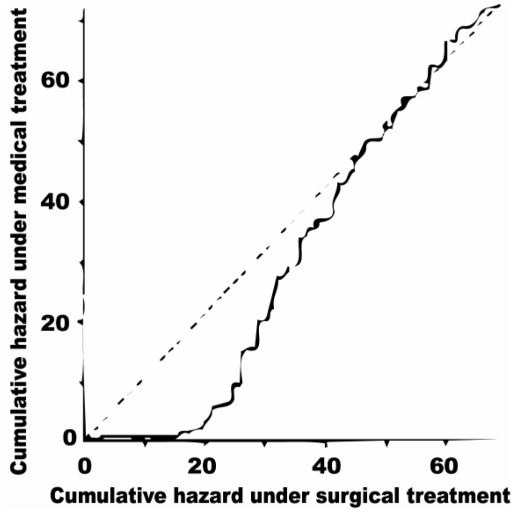


Figure 2-1: Lee-Pirie plot for Veterans' Administration data (Figure 5, Gill and Schumacher (1987), with axes interchanged)

and Schumacher (1987) (with the coordinates interchanged) is star-shaped, but not convex (for discussion on star-shaped and convex function, see Kalashnikov and Rachev, 1986); see Figure 2-1. Hence, it is not surprising that the analytical tests proposed by Gill and Schumacher (1987) failed to reject the PH hypothesis in favour of the increasing hazard ratio alternative. Perhaps a test designed for the increasing CHR alternative would have been able to reject the PH hypothesis.

In this chapter we propose a family of tests for the null and alternative hypotheses

$$\begin{aligned} \mathbb{H}_0 \quad &: \quad \Lambda_1(t)/\Lambda_2(t) = a \qquad\qquad \text{for all } t > 0, \text{ for some } a > 0 \\ \mathbb{H}_1 \quad &: \quad \Lambda_1(t)/\Lambda_2(t) \text{ is a non-constant increasing function of } t \text{ over } [0, \infty). \end{aligned}$$

The family of statistics presented here are consistent for testing $\mathbb{H}_0$ vs $\mathbb{H}_1$. The asymptotic distribution of a suitably normalized form of the test statistic is standard normal both under $\mathbb{H}_0$ and $\mathbb{H}_1$. While the results are obtained in the general context of comparing two counting processes, the case of censored survival data is given special consideration. We also present two graphical methods related to our analytical test. Finally, we adapt these methods, as well as those developed by Gill and Schumacher (1987) to the competing risks setup.

The chapter is organised as follows. In Section 2.2, we develop the test statistics, followed by consideration of consistency and asymptotic distributions in Section 2.3. In Section 2.4, we explore related graphical methods and develop real applications in Section 2.5, while in Section 2.6 we discuss the choice of weight functions in the proposed tests. Sections 2.1 through 2.6 are based on Sengupta *et al.* (1998). In Section 2.7, based on Sengupta and Bhattacharjee (1994), we extend these methods to the competing risks setup. Finally, we provide some concluding remarks in Section 2.8.

2.3 Development of the test statistic

Let $N_j(t)$ for $j = 1, 2$ and $t\epsilon[0, \infty)$ represent two components of a bivariate counting process. Let the Doob-Meier decomposition of the processes be of the form

$$dM_j(t) = dN_j(t) - Y_j(t)d\Lambda_j(t), \qquad j = 1, 2$$

where $\Lambda_j(.), j = 1, 2$ are deterministic functions on $[0, \infty)$ and $Y_j(.), i = 1, 2$ are non-negative processes which are predictable with respect to the filtration on which the martingales on the left hand side are defined. The above coincides with the "multiplicative intensity" model of the compensator process (see Aalen, 1978). When $N_j(t)$ corresponds to the number of failures or deaths up to time t in the j-th group consisting of individuals with i.i.d. life distributions, $\Lambda_j(t)$ is the cumulative hazard rate corresponding to this distribution. In general, $N_j(t)$ may be the number of type j transitions in a Markov chain, $Y_j(t)$ the number at risk for type j transition and $\Lambda_j(t)$ the integrated transition rate.

Under $\mathbb{H}_1$, it is expected that $\Lambda_1(y).\Lambda_2(x) - \Lambda_1(x).\Lambda_2(y)$ would be non-negative for all $x < y$ and positive for some $x < y$. If the ratio Λ_1/Λ_2 is a fast increasing function, the above difference would be generally large. This fact may be used to define a measure of non-proportionality of the cumulative hazard functions,

$$q(w) = \int\int_{0<x<y<\tau} w(x, y). \left[\Lambda_1(y)\Lambda_2(x) - \Lambda_1(x)\Lambda_2(y)\right].dx.dy, \tag{2.1}$$

where $w(x, y)$ is a positive weight function and τ is a large positive number such that $\Lambda_j(\tau) < \infty$

for $j = 1, 2$. The idea is similar to that of Deshpande and Sengupta (1995), who considered a measure of non-proportionality of the hazard rates. The double integral may be reduced to products of single integrals by choosing the weight function $w(x, y) = k_1(y)k_2(x) - k_1(x)k_2(y)$, $k_1(.)$ and $k_2(.)$ being positive weight functions with an increasing ratio. With this choice, the above measure simplifies to

$$q(k_1, k_2) = t_{11}t_{22} - t_{12}t_{21}, \tag{2.2}$$

where

$$t_{ij} = \int_0^\tau k_i(s)\Lambda_j(s)ds, \qquad i = 1, 2, j = 1, 2.$$

Clearly, $q(k_1, k_2)$ is positive under $\mathbb{H}_1$ and zero under $\mathbb{H}_0$. Therefore a consistent estimator of this difference can serve as a test statistic for the problem at hand. Suppose for $j = 1, 2$, $\widehat{\Lambda}_j(t)$ be the Nelson-Aalen estimator of $\Lambda_j(t)$ given by $\int_0^t dN_j(s)/Y_j(s)$ where the reciprocal of $Y_j(s)$ is defined to be 0 whenever $Y_j(s)$ is 0. Let $K_i(.)$, $i = 1, 2$ be right-continuous functions with left limits (rcll) converging in probability to $k_i(.), i = 1, 2$, respectively. We define the test statistic as

$$Q_{K_1K_2} = T_{11}T_{22} - T_{12}T_{21},$$

where $T_{ij} = \int_0^\tau K_i(s)\widehat{\Lambda}_j(s)ds, i = 1, 2, j = 1, 2$. It is shown in the appendix that a consistent estimator of the variance of the test statistic under the null hypothesis is

$$\widehat{\mathrm{Var}}(Q_{K_1K_2}) = T_{21}T_{22}V_{11} - T_{21}T_{12}V_{12} - T_{11}T_{22}V_{12} + T_{11}T_{12}V_{22}, \tag{2.3}$$

where

$$V_{ij} = \int_0^\tau \int_0^\tau K_i(t)K_j(s)V(s \wedge t)dsdt \qquad i = 1, 2, j = 1, 2,$$

and

$$V(t) = \int_0^t \frac{dN_1(s) + dN_2(s)}{Y_1(s)Y_2(s)}.$$

Note that the form of $Q_{K_1K_2}$ is similar to the statistic proposed by Gill and Schumacher (1987). In fact, if the cumulative hazard functions are replaced by the corresponding hazard rates, $q(k_1, k_2)$ becomes a measure of non-proportionality of the hazard rates. The family of statistics given by Gill and Schumacher (1987) may be motivated by this measure, although they did not mention it. An important difference between these two families is that the tests proposed here are not functions of the ranks alone; the actual lengths of time between successive jumps are made use of.

The weight functions $K_1(t)$ and $K_2(t)$ may be chosen so that $K_1(t)/K_2(t)$ is an increasing function, in order to make sure that $k_1(t)/k_2(t)$ is increasing. Gill and Schumacher (1987) have indicated several choices of weight functions for their family of statistics. Some of the choices are suitably normalized versions of

$$\begin{aligned}
K_a(t) &= Y_1(t)Y_2(t) \\
K_b(t) &= Y_1(t)Y_2(t)\,[Y_1(t)+Y_2(t)]^{-1} \\
K_c(t) &= Y_1(t)Y_2(t)\,[Y_1(t)+Y_2(t)]^{-1}\,\widehat{S}(t) \\
K_d(t) &= Y_1(t)Y_2(t)\,[Y_1(t)+Y_2(t)]^{-1}\left[\widehat{S}(t)\right]^{1/2}
\end{aligned}$$

where $\widehat{S}(t)$ is the Kaplan-Meier estimator computed from the combined sample. One may choose any pair of weight functions from the above that have an increasing ratio. All these weight functions are predictable, and hence satisfy the conditions of Gill and Schumacher (1987). Being rcll, these may also be used in the test statistic proposed here. In fact, the usable class of weight functions is larger here, because predictability is not required. For instance, one may replace the Kaplan-Meier estimator in the expression of $K_c(t)$ or $K_d(t)$ by a smoothed estimator.

2.4 Consistency and asymptotic normality

The form of the test statistic $Q_{K_1K_2}$ is similar to that of Gill and Schumacher (1987). However, here T_{ij} is not a stochastic integral but rather an ordinary Stieljes integral of a stochastic process. Therefore we take the following route to obtain the convergence results: (a) we show the convergence of the integral T_{ij} from that of the corresponding integrand (obtained from standard martingale convergence results); (b) subsequently we obtain the convergence of the test statistic by arguing that it is a constant function of the T_{ij}'s.

The first step comes from the following theorem.

Theorem 2.3.1

Let $\boldsymbol{K}_n$ and $\boldsymbol{X}_n$ be vector stochastic processes with sample paths in $D[0,\infty)^p$ and $D[0,\infty)^q$, such that $\boldsymbol{K}_n \xrightarrow{P} \boldsymbol{k}$ and $\boldsymbol{X}_n \xrightarrow{D} \boldsymbol{X}$, where $\boldsymbol{k}$ is a deterministic function in $C[0,\infty)^p$ and $\boldsymbol{X}$ is a stochastic process with sample paths in $D[0,\infty)^q$. Then for every positive constant τ,

$$\int_0^\tau \boldsymbol{K}_n(t) \otimes \boldsymbol{X}_n(t)dt \xrightarrow{D} \int_0^\tau \boldsymbol{k}(t) \otimes \boldsymbol{X}(t)dt. \tag{2.4}$$

(In the above, "$\otimes$" indicates the Kronecker product.)

Proof. See the appendix to this chapter.

In order to study the convergence of $T_{ij}, i = 1,2, j = 1,2$, we replace $\boldsymbol{K}_n(t)$ and $\boldsymbol{X}_n(t)$ in the above theorem by $[K_1(t) : K_2(t)]^T$ and a suitably normalized version of $\left[\widehat{\Lambda}_1(t) - \Lambda_1(t) : \widehat{\Lambda}_2(t) - \Lambda_2(t)\right]^T$, respectively. The latter process can be written as

$$\begin{pmatrix} \widehat{\Lambda}_1(t) - \Lambda_1(t) \\ \widehat{\Lambda}_2(t) - \Lambda_2(t) \end{pmatrix} = \begin{pmatrix} \int_0^t Y_1^{-1}(s)dM_1(s) \\ \int_0^t Y_2^{-1}(s)dM_2(s) \end{pmatrix}.$$

We denote this vector martingale by $\boldsymbol{M}(t)$. Further, let

$$\boldsymbol{K}(.) = \begin{pmatrix} K_1(.) \\ K_2(.) \end{pmatrix}, \boldsymbol{k}(.) = \begin{pmatrix} k_1(.) \\ k_2(.) \end{pmatrix}, \boldsymbol{\Lambda}(.) = \begin{pmatrix} \Lambda_1(.) \\ \Lambda_2(.) \end{pmatrix}, \widehat{\boldsymbol{\Lambda}}(.) = \begin{pmatrix} \widehat{\Lambda}_1(.) \\ \widehat{\Lambda}_2(.) \end{pmatrix}$$

where $K_i(.), k_i(.), \Lambda_i(.)$ and $\widehat{\Lambda}_i(.)$ for $i = 1,2$ are as defined in the Section 2.2. Finally, let $\boldsymbol{T} = (T_{11}T_{12}T_{21}T_{22})^T$ and $\boldsymbol{t} = (t_{11}t_{12}t_{21}t_{22})^T$. Note that the dependence of each of these quantities on n is suppressed here for notational simplicity. The convergence of the integral takes place as indicated below.

Corollary 2.3.2

Suppose there is a positive sequence $\{a_n\}$, approaching infinity as n goes to ∞, such that the following three conditions hold for $j=1,2$:

$$a_n \int_0^s \frac{d\Lambda_j(u)}{Y_j(u)} \xrightarrow{P} \int_0^s \frac{d\Lambda_j(u)}{y_j(u)} \qquad \forall s \in [0,\tau], \tag{2.5}$$

$$a_n \int_0^\tau Y_j^{-1}(u).I\left(\left|\frac{a_n}{Y_j(u)}\right| > \epsilon\right) d\Lambda_j(u) \xrightarrow{P} 0 \qquad \forall \epsilon > 0, \tag{2.6}$$

$$\sqrt{a_n} \int_0^\tau I\left(Y_j(u)=0\right) d\Lambda_j(u) \xrightarrow{P} 0, \tag{2.7}$$

where y_1^{-1} and y_2^{-1} are bounded on $[0,\tau]$. Then

$$\overline{\boldsymbol{T}} = \int_0^\tau \boldsymbol{K}(t) \otimes \boldsymbol{\Lambda}(t)dt \xrightarrow{P} \boldsymbol{t}, \tag{2.8}$$

$$\boldsymbol{T} = \int_0^\tau \boldsymbol{K}(t) \otimes \widehat{\boldsymbol{\Lambda}}(t)dt \xrightarrow{P} \boldsymbol{t}, \tag{2.9}$$

$$\sqrt{a_n}\left(\boldsymbol{T} - \overline{\boldsymbol{T}}\right) = \sqrt{a_n} \int_0^\tau \boldsymbol{K}(t) \otimes \boldsymbol{M}(t)dt \xrightarrow{D} \int_0^\tau \boldsymbol{K}(t) \otimes \boldsymbol{W}(t)dt, \tag{2.10}$$

where $\boldsymbol{W}(.)$ is a vector of two independent Gaussian processes $W_1(.)$ and $W_2(.)$ with zero mean, independent increments and variance function $\int_0^\tau y_j^{-1}(s)d\Lambda_j(s), j=1,2$, respectively.

Proof. The definition of $\boldsymbol{M}(.)$ implies that its components are orthogonal martingales with variation processes $\int_0^\tau Y_j^{-1}(s)d\Lambda_j(s), j=1,2$. Therefore the conditions (2.5)–(2.7) ensure, by a version of Rebolledo's martingale central limit theorem (see th. IV1.2 of Andersen *et al.*, 1993), that

$$\sqrt{a_n}\boldsymbol{M}(t)dt \xrightarrow{D} \boldsymbol{W}(t).$$

The results (2.8), (2.9) and (2.10) follow from theorem by replacing $\boldsymbol{X}_n(t)$ with $\boldsymbol{\Lambda}(t)$, $\widehat{\boldsymbol{\Lambda}}(t)$ and $\sqrt{a_n}\boldsymbol{M}(t)$, respectively.

Remark. The stronger condition

$$\sup_{0\le t\le\tau}\left|\frac{Y_j(t)}{a_n} - y_j(t)\right| \xrightarrow{P} 0 \text{ as } n \longrightarrow \infty, \qquad j=1,2 \tag{2.11}$$

implies the conditions (2.5)–(2.7).

The second step in the asymptotic argument is similar to that of Gill and Schumacher (1987). The results (2.9) and (2.10), coupled with the version of the delta-method given by Gill and Schumacher (1985) imply that

$$Q_{K_1K_2} \xrightarrow{P} q\left(k_1, k_2\right),$$
$$\sqrt{a_n}\left(Q_{K_1K_2} - q\left(k_1, k_2\right)\right) \xrightarrow{D} \int_0^\tau \left[c(t)W_1(t) - d(t)W_2(t)\right] dt,$$

where

$$\begin{aligned} c(t) &= t_{22}k_1(t) - t_{12}k_2(t), \\ d(t) &= t_{21}k_1(t) - t_{22}k_2(t). \end{aligned}$$

The limiting distribution is therefore Gaussian with zero mean, while the variance is given by

$$\int_0^\tau \int_0^\tau \left[c(t)c(s)V_1\left(s \wedge t\right) + d(t)d(s)V_2\left(s \wedge t\right)\right] dsdt,$$

where

$$V_j(t) = \int_0^t \frac{d\Lambda_j(s)}{y_j(s)}, \qquad j = 1, 2.$$

Under the null hypothesis, the ratio $\Lambda_2(.)/\Lambda_1(.)$ is a constant θ, which can also be called the hazard ratio. Further, $c(.)/d(.)$ is also equal to θ under $\mathbb{H}_0$. Thus an alternative expression for the asymptotic null variance is

$$\begin{aligned} \operatorname{var}\left(\sqrt{a_n}Q_{K_1K_2}\right) &= \int_0^\tau \int_0^\tau c(t)d(s)\left[\theta V_1\left(s \wedge t\right) + \theta^{-1}V_2\left(s \wedge t\right)\right] dsdt \\ &= \int_0^\tau \int_0^\tau c(t)d(s) \int_0^{s\wedge t} \left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)}\right] dsdt \\ &= t_{21}t_{22}v_{11} - t_{21}t_{12}v_{12} - t_{11}t_{22}v_{12} + t_{11}t_{12}v_{22}, \end{aligned}$$

where

$$v_{ij} = \int_0^\tau \int_0^\tau k_i(t)k_j(s) \int_0^{s\wedge t} \left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)}\right] dsdt, \qquad i = 1, 2, j = 1, 2.$$

This variance is estimated consistently by a_n times the expression given in (2.3), as shown in the appendix to this chapter, provided $\frac{Y_j(t)}{a_n} \xrightarrow{P} y_j(t)$ pointwise on $[0, \tau]$. Since $q(K_1, K_2)$ is zero under $\mathbb{H}_0$ and positive under $\mathbb{H}_1$, the normalized statistic can be used for a one-sided test.

2.5 Graphical methods

The following three graphical procedures are of special interest here:

(a) the plot of $\widehat{\Lambda}_1(t)$ vs $\widehat{\Lambda}_2(t)$, proposed by Lee and Pirie (1981),

(b) the plot of $\left(\widehat{\Lambda}_1(t) - \widehat{\Lambda}_2(t)\right)$ vs t, due to Dabrowska *et al.* (1989) and

(c) the plot of the log cumulative hazard difference $\ln\left(\widehat{\Lambda}_1(t) - \widehat{\Lambda}_2(t)\right)$ against t, suggested by Dabrowska *et al.* (1992).

A monotone trend in any of the last two plots suggests a monotone CHR of the two samples, while no trend corresponds to the PH model. Plot (a) is expected to be close to a straight line in the PH case and star-shaped when the CHR is (monotone) increasing. Thus, all the three plots are expected to bring out monotone CHR-type departures from the PH model, although they have so far been used to look for monotone hazard ratio.

The above plots can be quite unstable. Plots (b) and (c) can have wild fluctuations for small values of t (see Dabrowska *et al.*, 1989), while plot (a) may lack precision for large values of t. Gill and Schumacher (1987) suggested a modification of plot (a), replacing $\widehat{\Lambda}_j(t)$ with $\widehat{\Lambda}_j^K(t) = \int_0^t K(s)d\widehat{\Lambda}_j(s), j = 1, 2$, where $K(.)$ is a predictable weight function (see Section 2.2). This modification can also be used in plots (b) and (c). The modified plots have the same characteristic features when the hazard ratio is constant or monotone, but such a feature no longer exists for monotone CHR.

To overcome this problem, we propose two graphical tests based on the estimated functions $T_j^K(t) = \int_0^t K(s)d\widehat{\Lambda}_j(s), j = 1, 2$, where $K(.)$ is now a rcll weight function. The plot of $T_1^K(t)/T_2^K(t)$ against t is expected to be like a horizontal straight line when the PH model holds. On the other hand, a monotone ratio of the cumulative hazards of the two populations is expected to produce a monotone trend in the plot, irrespective of the choice of the weight function. Since $\widehat{\theta}_K = T_1^K(\tau)/T_2^K(\tau)$ is a consistent estimator of the hazard ratio in the PH case, the horizontal straight line passing through the right end-point of the graph serves as a reference corresponding to the PH hypothesis.

The other suggested plot is that of $T_1^K(.)$ against $T_2^K(.)$. This graph is expected to be close to a straight line when the PH model holds and approximately convex or concave when the CHR is monotone. The straight line joining the origin with the end-point of the graph $\big(T_2^K(\tau), T_1^K(\tau)\big)$, may serve as a reference for the PH hypothesis. The two suggested plots are expected to be smoother and more stable than their unweighted counterparts.

2.6 Data Analysis

The analytic and graphical procedures proposed in Sections 2.2 and 2.4 were used to analyse the ovarian cancer data set reported by Fleming *et al.* (1980), which describes the number of days from treatment to progression of disease. Here, groups 1 and 2 consist of 20 patients with high-grade tumor (stage IIA) and 15 patients with low-grade tumor (stage II), respectively. The statistic $Q_{K_b K_a}$ (after normalization) is 2.258. The corresponding two-sided p-value is 0.024, suggesting an increasing trend of the ratio $\Lambda_1(t)/\Lambda_2(t)$. This supports the findings of Gill and Schumacher (1987) and Deshpande and Sengupta (1995) that the hazard ratio is increasing.

The plot of $\widehat{\Lambda}_1(t)/\widehat{\Lambda}_2(t)$ vs. t, shown in Figure 2-2 has by and large an increasing trend, but the fluctuations are substantial. Figure 2-3 shows the plot of $T_1^{K_b}(t)/T_2^{K_b}(t)$ against t which was suggested in Section 2.4. This graph is smoother and more clearly suggestive of an increasing trend of the CHR.

The plot of $\widehat{\Lambda}_1^{K_b}(t)$ vs. $\widehat{\Lambda}_2^{K_b}(t)$ shown in Figure 2-4 is approximately convex, indicating an increasing hazard ratio. However, the plot of $T_1^{K_b}(t)$ against $T_2^{K_b}(t)$ shown in Figure 2-5 is smoother and clearly convex, suggesting an increasing CHR.

2.7 Choice of weight functions

The role of the weight functions in the family of tests proposed here is crucial. An interesting question that can be posed in this connection is: "Can the weight functions be chosen 'optimally' according to some chosen criterion?" We have no clear answer to this question as yet. If a sequence of alternative hypotheses converging to $\mathbb{H}_0$ at a suitable rate is considered, it can be shown that the asymptotic relative efficiency is of the form

$$\frac{\left[\int_0^\tau l(t)g(t)dt\right]^2}{\int_0^\tau \int_0^\tau l(s)W(t,s)dsdt},$$

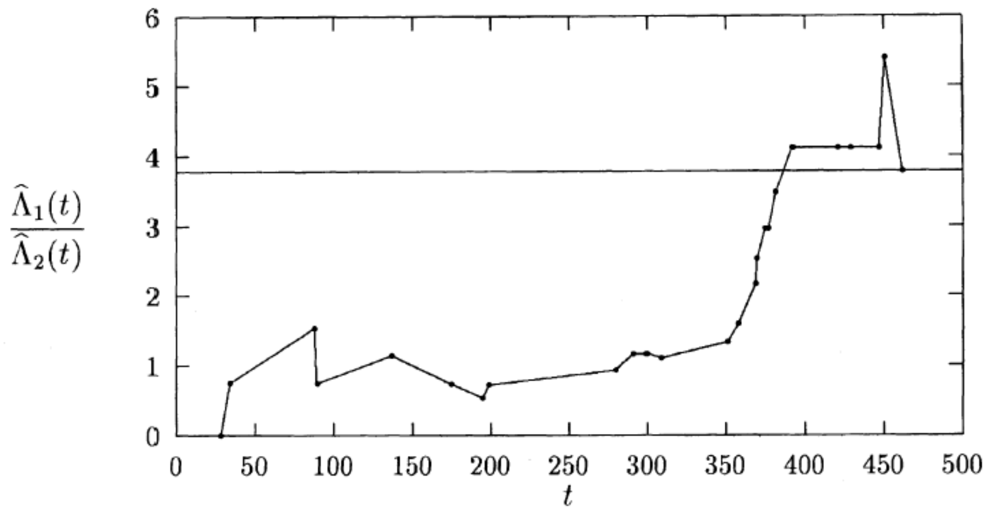


Figure 2-2: Plot of $\widehat{\Lambda}_1(t)/\widehat{\Lambda}_2(t)$ vs t for the ovarian cancer data.

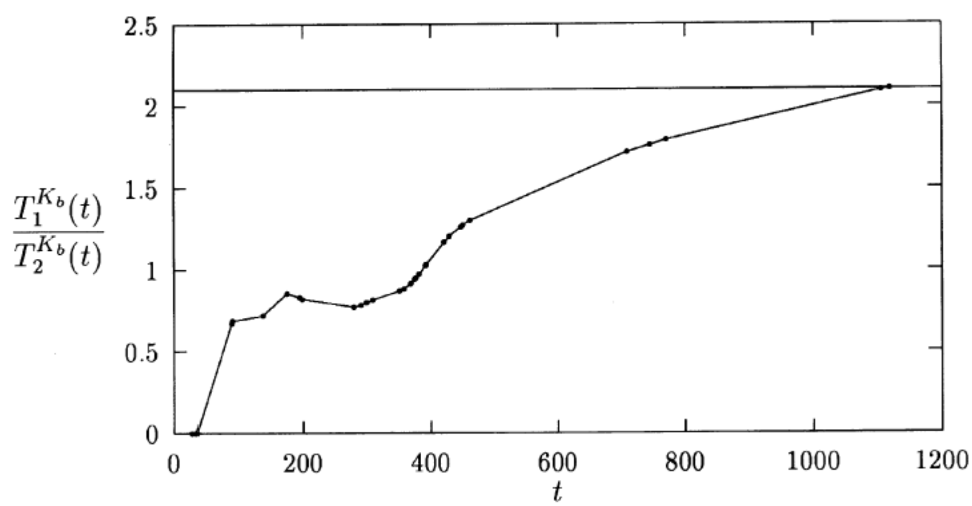


Figure 2-3: Plot of $T_1^{K_b}(t)/T_2^{K_b}(t)$ vs t for the ovarian cancer data.

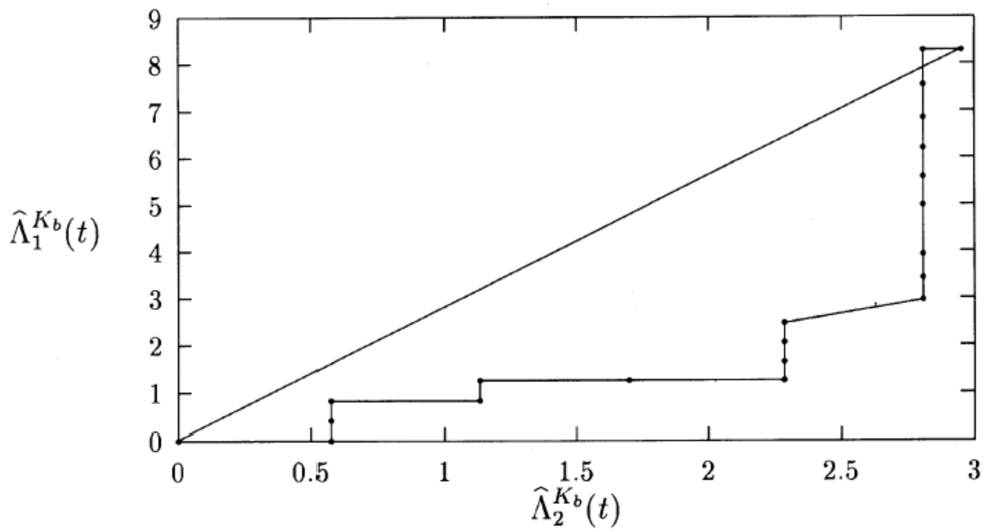


Figure 2-4: Plot of $\widehat{\Lambda}_1^{K_b}(t)$ vs $\widehat{\Lambda}_2^{K_b}(t)$ for the ovarian cancer data.

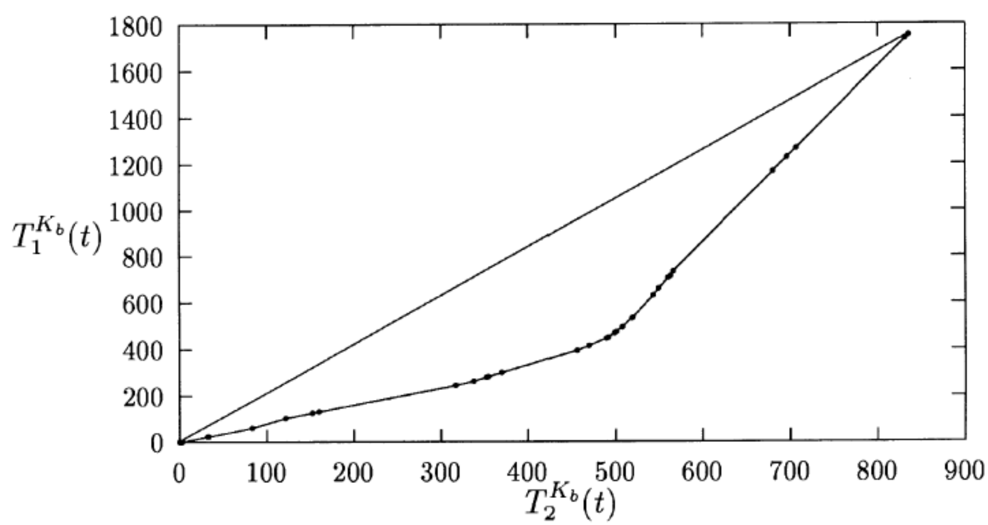


Figure 2-5: Plot of $T_1^{K_b}(t)$ vs $T_2^{K_b}(t)$ for the ovarian cancer data.

where $l(t)$ is the probability limit of the ratio of the weight functions, $g(t)$ is a function determined by $\Lambda_1(t)$ and $\Lambda_2(t)$, and $W(t,s)$ is a positive definite function of two variables, also determined by $\Lambda_1(t)$ and $\Lambda_2(t)$. A function $l(t)$ that maximizes this expression would lead to a suitable weight function. Unfortunately a closed form solution to this problem is not available. This is in contrast to the similar problem addressed by Gill and Schumacher (1987), where the "optimal" solution could be obtained in closed form through the Cauchy-Schwartz inequality. This is not a major issue since the role of optimal weight functions is rather limited in applications where a precise idea of departures from $\mathbb{H}_0$ is rarely available *a priori*; for further discussion, see Section 2.7.3.

Here, we explore the role of the weight functions in the two-sample testing problem by performing a small-scale simulation study. The two samples were generated from an exponential distribution and a piecewise exponential distribution, respectively. Several combinations of weight functions were tried out. Out of these, the combination $Y_1(t)Y_2(t)exp[-t/TTT]$ and $Y_1(t)Y_2(t)$, where TTT is the total time on test statistic for the combined sample, yielded the highest power. The former weight function could not have been used for the family of tests proposed by Gill and Schumacher (1987), since it is not predictable. This underscores the wide scope of the class of rcll weight functions considered here.

2.8 Testing Proportionality of Hazards due to Competing Risks

Consider a system or unit which is exposed to several risks that can induce failure. The system can be an individual suffering from more than one disease, as found commonly in survival data. In Reliability, a series system fits well into the competing risk framework. Different types of employment of an unemployed individual can serve as a third example of this model.

The nonparametric analysis of competing risks data is often made tractable by means of simplifying assumptions such as the independence of risks. Comparing one risk with another is an important problem in this context. Bagai *et al.* (1989a) studied a test for equality of two risks against the alternative of one hazard dominating the other. In another paper (Bagai *et al.*, 1989b) they considered the alternative of stochastic dominance. In these works the independence of risks is a key assumption and the presence of a third risk is ruled out.

Deshpande and Sengupta (1995) proposed a test, based on U-statistics, for the hypothesis that the hazards due to two risks are proportional to each other, against the alternative that the hazard ratio is monotonically increasing. As discussed earlier in the thesis, this problem is important because of two reasons. The proportional hazards assumption is commonly used in applied work in survival analysis and reliability models, as well as duration models in econometrics (see Kalbfleisch and Prentice, 1980 and Kiefer, 1988). Secondly, the alternative provides a reasonable description for the 'crossing hazards' situation often observed in empirical studies. Crossing hazards represent the situation where the importance of one risk as compared to another becomes noticeable only in the long run. An interesting additional aspect of their work is that the presence of a third risk is also taken into account. This is a crucial generalisation because all the risks which are different from the risks being compared can be pooled to form the third group of risks.

In this Section, we propose an alternative methodology, by adapting the family of tests proposed by Gill and Schumacher (1987) and in Section 2.2, originally proposed for two-sample data, to the competing risks situation. The asymptotic properties of the tests follow from the counting process theory, which hold even when the risks are not independent. A related graphical technique is also discussed.

2.8.1 A graphical method

Consider Gill and Schumacher's (1987) modification of the plot of Lee and Pirie (1981), where the Nelson-Aalen estimators of the cumulative hazards of two samples be plotted against one another. A convex or concave trend would indicate a monotone hazard ratio. In a competing risks situation, the Nelson-Aalen estimators of the integral of each cause-specific hazard rate can be easily computed. Let the estimators be $\widehat{\Lambda}_j(.)$, where j is 1 for risk 1 and 2 for risk 2. Therefore a convex or concave trend in the plot of $\widehat{\Lambda}_1$ vs $\widehat{\Lambda}_2$ would indicate that $\lambda_1(t)/\lambda_2(t)$ is a monotonic function of t, where $\lambda_j(.)$ is the j-th cause-specific hazard. When the risks are independent, each cause-specific hazard reduces to a simple hazard rate due to a given risk. On the other hand, the proportional hazards model would correspond to approximately a straight line passing through origin on the above graph. These properties continue to hold if, in order to ensure stability of the plot in the tail region, one uses the generalised plot (see Gill and

Schumacher, 1987) of $\widehat{\Lambda}_2^K(.)$ vs. $\widehat{\Lambda}_1^K(.)$, where

$$\widehat{\Lambda}_j^K(t) = \int_0^t K(s)\frac{dN_j(s)}{Y(s)}, \qquad j = 1, 2,$$

and $K(s)$ is a predictable weight function. As usual, $Y(s)$ is the number at risk at time s and $N_j(s)$ is the number of type j absorptions up to time s. A simple weight function such as $K(s) = Y(s)$ would put greater weight on the more reliable part of the estimate, thus producing a smoother plot.

The graph can be quite revealing when the sample size is large. Take for example the unemployment duration data due to Han and Hausman (1990). The data consists of the duration of unemployment of 1051 individuals, 603 of whom are recalled to the old job, 245 eventually get a new job, while the remaining 203 are censored at various points of the study. If absorption into a new job_ and recall to the old job are taken as risks 1 and 2, respectively, then the plot of $\widehat{\Lambda}_2^K(t)$ vs. $\widehat{\Lambda}_1^K(t)$ (with $K(t) = Y(t)$) is as shown in Figure 2-6. The concave trend of the plot is quite clear. It shows that the rate of recall has a decreasing ratio with the rate of getting a new job. The latter becomes more significant in the long run. It is surprising that Han and Hausman used the proportional hazards model for this data. Although they did this in the presence of a binary covariate (unemployment insurance coverage), their basic assumption seems to be wrong.

A second example considered here is that of the male mice cancer data due to Hoel (1972). The group of 99 mice are examined after exposure to 300 rads of radiation. The 60 deaths due to cancer are attributed to risk 1. The other 39 deaths are gut together as deaths due to risk 2. There is no censoring in the data. The plot $\widehat{\Lambda}_2^K(t)$ vs. $\widehat{\Lambda}_1^K(t)$ shown in Figure 2-7 is somewhat convex. indicating that the risk due to cancer increases in the long run.

While the indication from Figure 2-5 is quite clear, the conclusions to be drawn from Figure 2-7 are not as obvious. The latter is a situation where analytical tests, discussed in the next section (Section 2.7.2), can play a particularly decisive role.

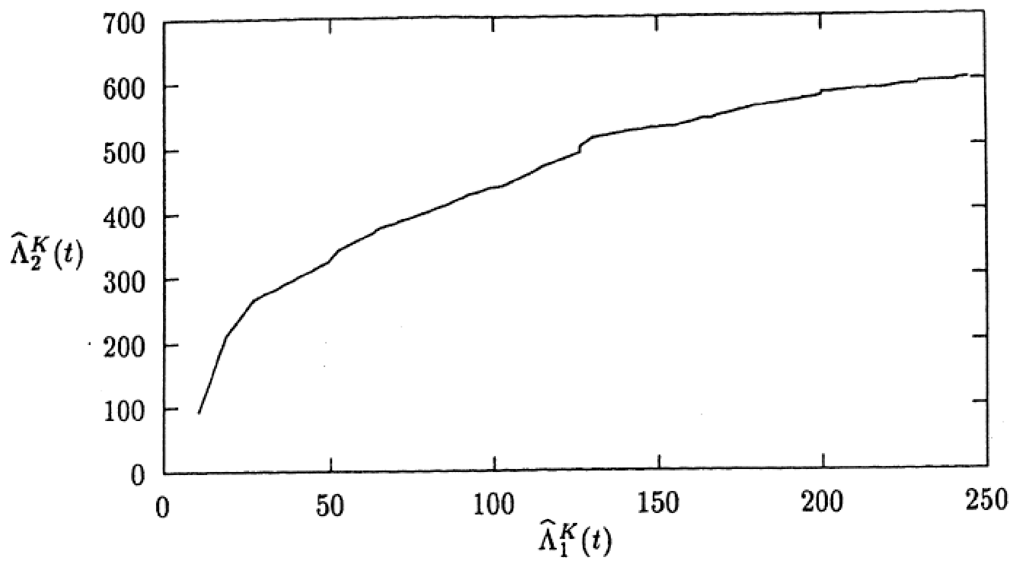


Figure 2-6: Weighted cumulative hazards due to recall and new job for the unemployment data.

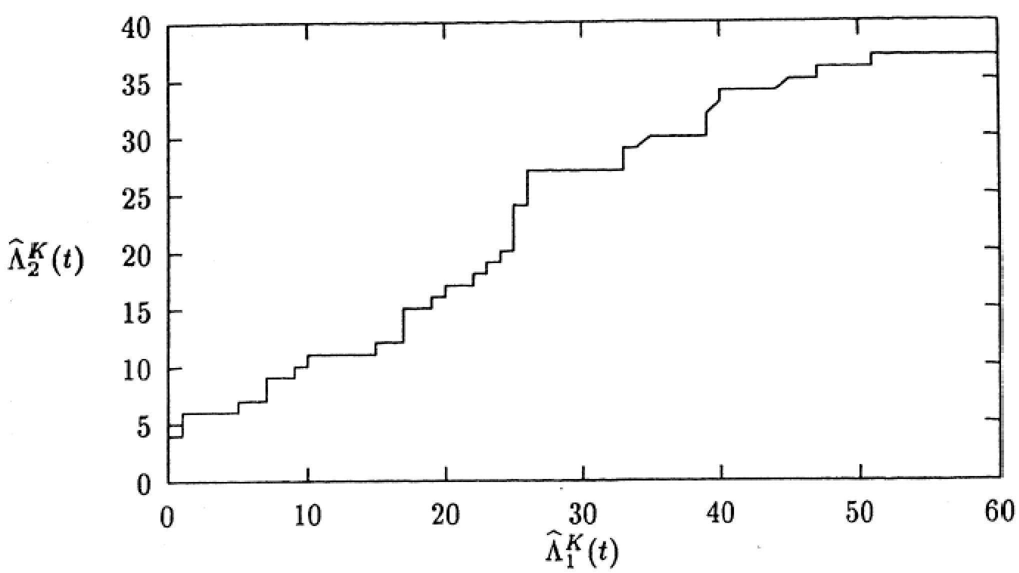


Figure 2-7: Weighted cumulative hazards due to cancer and other deaths for the mice data.

2.8.2 A family of analytical tests

Suppose $\lambda_1(t)$ and $\lambda_2(t)$ are the cause-specific hazard rates due to risk 1 and risk 2, respectively. We focus on the following testing problem:

$$\begin{aligned}
\mathbb{H}_0 &: \lambda_1(t)/\lambda_2(t) = a \qquad\qquad \text{for all } t > 0, \text{ for some } a > 0 \\
\mathbb{H}_1 &: \lambda_1(t)/\lambda_2(t) \text{ is a non-constant increasing function of } t \text{ over } [0, \infty).
\end{aligned}$$

Gill and Schumacher (1987) suggested a test for $\mathbb{H}_0$ vs. $\mathbb{H}_1$ when $\lambda_1(t)$ and $\lambda_2(t)$ are hazard rates of two isolated samples. The idea is that when $\mathbb{H}_0$ is true, different estimators of the hazards ratio should be close to each other. This leads to the statistic

$$R_{K_1K_2} = \widehat{K}_{11}\widehat{K}_{22} - \widehat{K}_{21}\widehat{K}_{12},$$

where $\widehat{K}_{ij} = \int_0^\tau K_i(t) d\widehat{\Lambda}_j(t), i = 1, 2, j = 1, 2,$ $K_1(t)$ and $K_2(t)$ are two different predictable weight functions, and τ is a stopping point. The statistic is directly applicable to the competing risks framework where $\widehat{\Lambda}_j(t)$ for $j = 1, 2$ are the Nelson-Aalen estimator of the cause-specific cumulative hazard functions. Specifically, $\widehat{\Lambda}_j(t) = \int_0^t [Y(s)]^{-1} dN_j(s)$, where $N_j(s)$ and $Y(s)$ have their usual interpretations. In this competing risks setup, the following weight functions may be used:

$$\begin{aligned}
K_e(t) &= Y(t), \\
K_f(t) &= Y(t)S(t), \\
K_g(t) &= Y^2(t), \\
K_h(t) &= Y(t)S^{1/2}(t),
\end{aligned}$$

where $S(t)$ is the Kaplan-Meier estimator for the entire sample, treating the absorptions due to the third risk as censored observations. In the two-sample case, the above four weight functions correspond to the logrank test, the Prentice-Wilcoxon test, the Gehan test and the Harrington-Fleming test (see Gill and Schumacher, 1985). In the competing risks situation these weight functions do not have any such interpretations, but they are easy to use anyway.

The estimated variance of the test statistic is given by

$$\widehat{\mathrm{Var}}\,(R_{K_1K_2}) = \widehat{K}_{21}\widehat{K}_{22}\widehat{V}_{11} - \widehat{K}_{21}\widehat{K}_{12}\widehat{V}_{12} - \widehat{K}_{11}\widehat{K}_{22}\widehat{V}_{21} + \widehat{K}_{11}\widehat{K}_{12}\widehat{V}_{22},$$

where

$$\widehat{V}_{ij} = \int_0^\tau K_i(t)K_j(t)\frac{d\,(N_1(t)+N_2(t))}{Y^2(t)}, \qquad i,j = 1,2.$$

Let us assume that as the sample size goes to infinity, the censoring proportion and the proportion of each type of absorption stabilises. The asymptotic normality of the statistic

$$T_{K_1K_2} = \frac{R_{K_1K_2}}{\sqrt{\mathrm{Var}(R_{K_1K_2})}}$$

under the null hypothesis follow from the counting process theory; see Gill and Schumacher (1987). The arguments given by Gill and Schumacher (1985) for consistency also go through in the competing risks case. A sufficient condition for consistency is that the ratio of the weight functions $K_1(t)/K_2(t)$ should be increasing in the limiting sense, as the number of subjects go to infinity. Several pairs of weight functions from the above list satisfy this criterion.

Note that when the risks are assumed to be independent, the statistic $T_{K_1K_2}$ becomes a competitor of the statistic V_0 of Deshpande and Sengupta (1995). In fact, the class of statistics discussed here have a wider range of application because the assumption of independence is not needed.

2.8.3 Choice of the weight functions

Suppose $K_1(t)$ and $K_2(t)$ converge in probability, as $n \longrightarrow \infty$, to $k_1(t)$ and $k_2(t)$, respectively. Further, let $k(t) = k_2(t)$ and $l(t) = k_1(t)/k_2(t)$. Gill and Schumacher (1987) considered the issue of asymptotic relative efficiency (ARE) by taking a sequence of alternatives approaching $\mathbb{H}_0$ in a specific way. Specifically, they assumed that: as sample size $n \longrightarrow \infty$, the hazard rates indexed on sample size, $\lambda_j^{(n)}(t)$, approaches $\lambda_j(t)$, $j = 1,2$, in such a way that

$$\lim_{n\longrightarrow\infty} \sqrt{n}\left(\frac{\lambda_2^{(n)}(t)}{\lambda_1^{(n)}(t)} - \theta\right) = m(t), \qquad \text{where } \theta = \frac{\lambda_2(t)}{\lambda_1(t)}.$$

Under these assumptions, they showed that the normalized test statistic in the two-sample case converges to a normal distribution with unit variance and mean given by

$$-\theta^{1/2}\frac{\int lk\,(m-\overline{m})\,d\Lambda_1}{\sqrt{\left(l-\bar{l}\right)^2 k^2 d\Lambda_1/y}},$$

where

$$\bar{l}=\frac{\int lkd\Lambda_1}{\int kd\Lambda_1} \text{ and } \overline{m}=\frac{\int mkd\Lambda_1}{\int kd\Lambda_1}.$$

Here

$$y=\frac{y_1y_2}{(y_1+\theta y_2)},$$

where $y_j(t)$ is the probability limit of $Y_j(t)/n, j=1,2$, where $Y_1(t)$ and $Y_2(t)$ correspond to the numbers at risk in the two samples at time t. From this result they concluded, through the use of Cauchy-Schwartz inequality, that the Pitman efficacy is maximised (within the family of tests considered here) by choosing $k(l-\bar{l})\propto y(m-\overline{m})$. However, this argument is somewhat misleading, since the optimization should have been carried out explicitly under the constraint $\int k(l-\bar{l})d\Lambda_1=0$. This constraint introduces a correction term, suggesting that the optimal choice would be

$$k(l-\bar{l})\propto y\left[m-\frac{\int myd\Lambda_1}{\int yd\Lambda_1}\right].$$

If k is chosen as y, the right hand side becomes $y(m-\overline{m})$, leading to the choice $l=m$. Happily this pair of solutions coincides with the suggestion of Gill and Schumacher (1985, 1987).

The same argument also holds in the competing risks situation with $y(t)$ as the probability limit of $Y(t)/n$. The highest ARE is achieved by choosing

$$K_2(t)=\frac{Y(t)}{n}, \qquad K_1(t)=n^{-1}Y(t)m(t).$$

Naturally the latter function can only be computed with a particular alternative in mind. Nevertheless, the 'optimal' pair of weight functions may serve as a benchmark for comparing the performance of other tests within the family. Its role is similar to that of the locally most powerful rank test in examining the performance of rank tests.

We illustrate the above result by computing the 'optimal' weight function in the case of three parametric families, assuming the risks to be independent.

Example 2.7.3.1 (Weibull): Let $\overline{F}_j(t) = \exp\left\{-\alpha_j t^{\theta_j}\right\}$ for $j = 1, 2$. Then $\lambda_2(t)/\lambda_1(t)$ is an increasing function of t if and only if $\theta_1 \leq \theta_2$, irrespective of α_1 and α_2 which are positive. Holding α_1, α_2 and θ_1 fixed, let $\theta_2^{(n)} = \theta_1\left(1 + n^{-1/2}\right)$. Then the above conditions are satisfied, with $m(t) = \frac{\alpha_2}{\alpha_1}(1 + \theta_1 \ln t)$.

Example 2.7.3.2 (Linear Failure Rate or LFR): Let $\overline{F}_j(t) = \exp\left\{-\alpha_j\left(t + \frac{1}{2}\theta_j t^2\right)\right\}$ for $j = 1, 2$. In this case, $\lambda_2(t)/\lambda_1(t)$ is an increasing function of t if and only if $\theta_1 \leq \theta_2$, irrespective of α_1 and α_2 which are positive. In this case the same configuration of the parameters as above produces $m(t) = \frac{\alpha_2}{\alpha_1}\left(1 + \frac{1}{\theta_1 t}\right)^{-1}$.

Example 2.7.3.3 (Pareto): Let $\overline{F}_j(t) = (1 + t/\theta_j)^{-\alpha_j}$ for $j = 1, 2$. In this case, $\lambda_2(t)/\lambda_1(t)$ is an increasing function of t if and only if $\theta_1 \leq \theta_2$, irrespective of α_1 and α_2 which are positive. Once again let α_1, α_2 and θ_1 be fixed, let $\theta_2^{(n)} = \theta_1\left(1 + n^{-1/2}\right)$. The resulting $m(t)$ is $-\frac{\alpha_2}{\alpha_1}\frac{\theta_1}{\theta_1 + t}$.

We emphasize that the above choice is optimal within a given family only in the context of ARE. Choosing the appropriate weight function does not ensure highest power for a fixed sample size and an alternative well separated from $\mathbb{H}_0$.

2.8.4 Monte Carlo study

We examine the following test statistics:

1. T_1: the statistic with weight functions K_e and K_f;

2. T_2: the statistic with weight functions K_e and K_g;

3. T_3: the statistic with weight functions K_e and K_h;

4. T_4: the statistic with weight functions K_h and K_f;

5. T_5: the statistic with weight functions K_e and K_o, where the latter is optimal for a given family of distributions, as described in Section 2.7.3;

6. V_0: the U-statistic proposed by Deshpande and Sengupta (1995) normalized by the asymptotic variance $4\widehat{E}_2/n$, where $\widehat{E}_2$ is the U-statistic estimator of E_2 (Deshpande and Sengupta, 1995, p. 257).

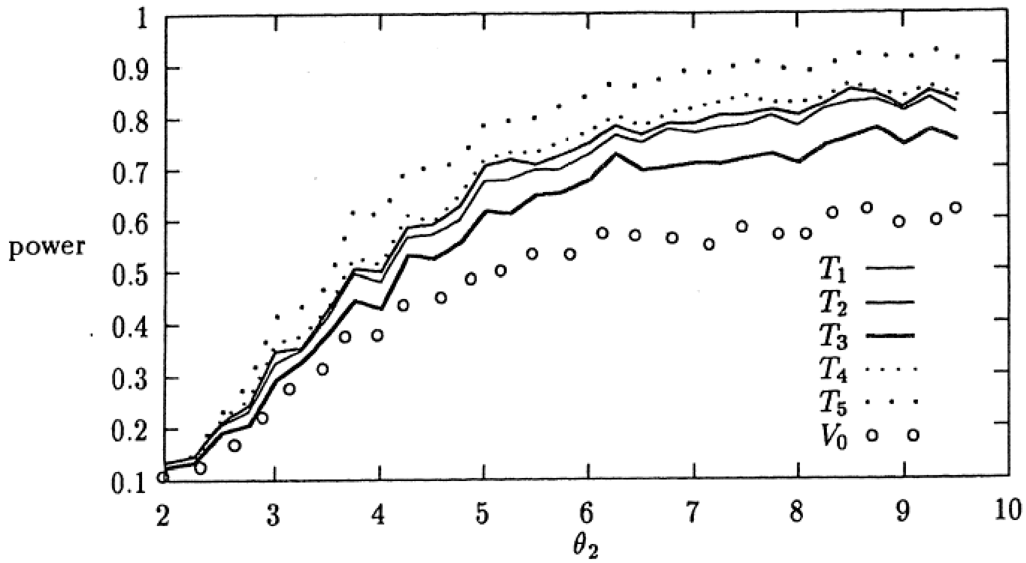


Figure 2-8: Empirical power curves when risks have Weibull distribution.

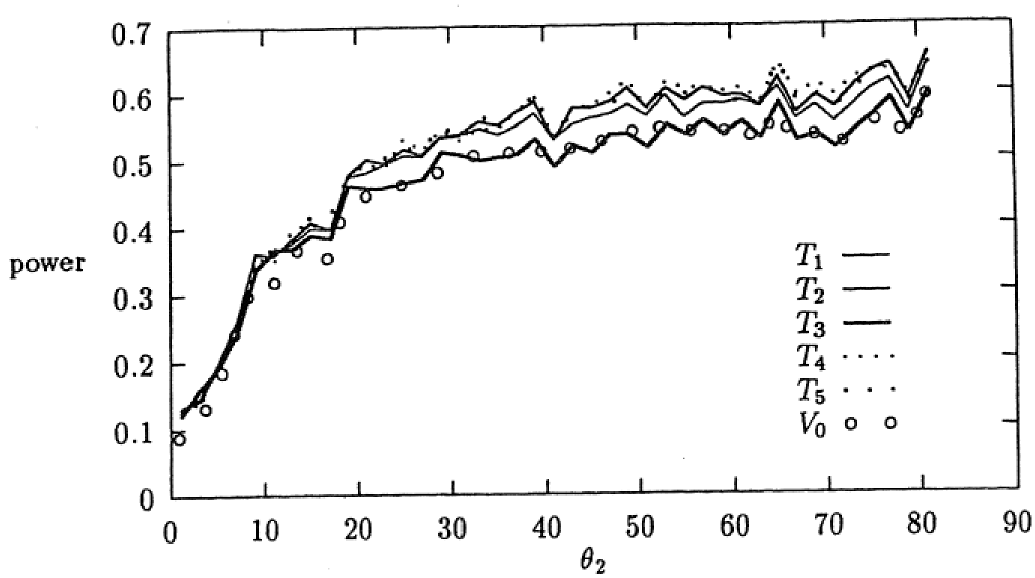


Figure 2-9: Empirical power curves when risks have LFR distribution.

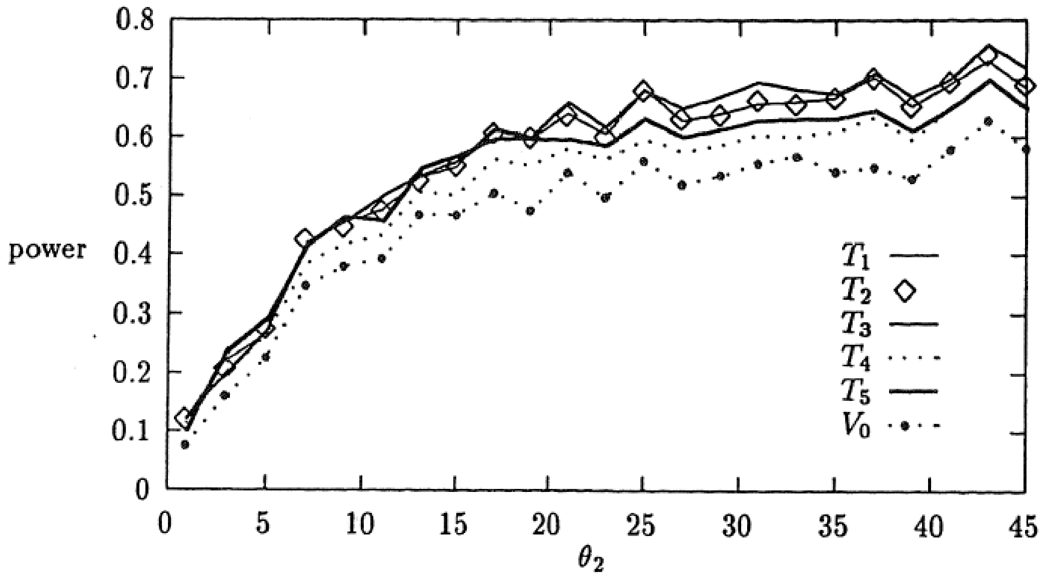


Figure 2-10: Empirical power curves when risks have Pareto distribution.

The null distributions of these statistics are checked first. The empirical distributions of the six statistics from 1000 monte carlo simulations were compared to the standard normal cdf. The sample size for each experiment was 40. Each risk corresponded to a Weibull distribution of the notional lifetime. The parameters of the distributions were: $\alpha_1 = 1$, $\alpha_2 = 0.5$ and $\theta_1 = \theta_2 = 2$. The censoring distribution was chosen to be exponential with mean 10. All the empirical cdf's showed reasonable closeness to the theoretical curve. For the sake of brevity we are not reproducing these plots. Instead, the Shapiro-Francia (1972) test of normality (from 100 monte carlo runs) is carried out. The Shapiro-Francia statistics for the six tests mentioned above turn out to be 0.9893, 0.9875, 0.9923, 0.9851, 0.9857 and 0.9884, respectively. These may be compared to the percentage points 0.980($p = 0.1$), 0.984($p = 0.2$), 0.989($p = 0.5$) and 0.993($p = 0.8$). The results are quite satisfactory. The reason for using only 100 runs for the analytical tests is that the percentage points are not readily available for higher sample sizes.

The same experiments were also carried out for the LFR and Pareto distributions of notional lifetimes. The parameters in the LFR case were: $\alpha_1 = 1$, $\alpha_2 = 0.1$ and $\theta_1 = \theta_2 = 1$, while the sample size was 50. The parameters in the Pareto case were: $\alpha_1 = 0.5$, $\alpha_2 = 1$ and $\theta_1 = \theta_2 = 1$, while the sample size was 40. The same censoring distribution (exponential with mean 10) was used. The empirical distribution of all the six test statistics from 1000 monte carlo runs showed closeness to the standard normal cdf. The Shapiro-Francia statistics from 100 runs in the LFR case were 0.9856, 0.9848, 0.9842, 0.9796, 0.9889 and 0.9834, respectively. In the Pareto case the statistics were 0.9856, 0.9849, 0.9811, 0.9823, 0.9834 and 0.9911, respectively. The results assure us that the cut-off points from the asymptotic null distribution may be used for further study.

Empirical power computations from monte carlo experiments were also made. Figures 2-8 to 2-10 show the empirical power curves from 500 experiments for the three families of distributions. In each case the parameter θ_2 was gradually increased, holding the other parameters fixed at their respective values in the previous experiment. The plots generally reveal the superiority of the family of tests considered here over the U-statistic. The 'optimal' choice of weight functions from ARE considerations does not always lead to the best power, as expected. Compared to this benchmark, the performance of T_2 appears to be good in all the cases. The performances of T_1 and T_4 are also quite good.

2.8.5 Data analysis

The statistics T_1 to T_4, when evaluated for the unemployment duration data of Han and Hausman (1990), turn out to be 11.83, 11.93, 11.96 and 11.48, respectively. The U-statistic V_0 in this case is 10.63. All these strongly indicate that the rates of absorption into new and old jobs are not proportional to each other. This is in accordance with the indications from the plot of Section 2.7.1 (Figure 2-6), and contradicts the basic assumption of Han and Hausman. Consequently a fresh analysis of the data may be in order.

The mice cancer data due to Hoel (1972) has no censored observations. Hence T_1 and T_2 are identical, while the computation of the null variance of the U-statistic is simplified (see Deshpande and Sengupta, 1995). The statistics are $T_1 = T_2 = 1.895$, $T_3 = 1.823$, $T_4 = V_0 = 1.87$. The corresponding one-sided p-values are 0.029, 0.032 and 0.031, respectively. Once again all the test statistics tell the same story – that the hazard due to cancer increases with time when compared to the hazard due to other causes. This adds a significant dimension to the analysis of the same data by Bagai *et al.* (1989a) who found the hazard due to cancer to be smaller than the other hazards combined.

2.8.6 Testing against the monotone cumulative hazard ratio alternative

As in the two sample case, the monotone CHR alternative may be more appealing and appropriate than increasing/ decreasing hazard ratio in many competing risks applications. Let $\lambda_1(t)$ and $\lambda_2(t)$ denote the cause-specific hazard rates due to risks 1 and 2, with the corresponding cumulative cause-specific hazard rates denoted by $\Lambda_j(t) = \int_0^t \lambda_j(s)ds, j = 1, 2$. Therefore, our testing problem is:

$$\begin{aligned} \mathbb{H}_0 &: \Lambda_1(t)/\Lambda_2(t) = a \qquad \text{for all } t > 0, \text{ for some } a > 0 \\ \mathbb{H}_1 &: \Lambda_1(t)/\Lambda_2(t) \text{ is a non-constant increasing function of } t \text{ over } [0, \infty). \end{aligned}$$

The analytical test proposed in Section 2.2 can easily be adapted to this competing risks situation. The test statistic $Q_{K_1K_2}$ will now be based on two distinct rcll weight functions K_1 and K_2 appropriate for the competing risks setup. An important advantage of this test is that an independence assumption on the risks is not necessary. Further, the presence of other risks can be accommodated in a way similar to that discussed in Section 2.7.2. Consistency and asymptotic distributions, as well as variance estimation, follows exactly in the same way as in the two sample case (see Section 2.3), and the graphical tests proposed in Section 2.4 also work perfectly well. Dauxois and Kirmani (2004) have developed a closely related test for proportionality of cumulative incidence functions.

2.9 Concluding remarks

In this chapter, we proposed tests of the proportional hazards assumption in two samples against the monotone cumulative hazard ratio alternative. This partial order is weaker than the monotone hazard ratio hypothesis considered in the literature. The use of the proposed graphical and analytical tests are illustrated with several applications. Further, we extend tests for both the above kinds of partial orders to a competing risks setup. This extension has the important advantages of relaxing the common assumption of independence of competing risks and in allowing the presence of other risks. In the competing risks context, we also clarify the nature of optimal tests in the sense of efficacy.

The tests proposed here can be generalised in three ways. First, an useful and interesting research problem is to extend the partial orders describing nonproportional hazards situations and the corresponding tests to the case of continuous covariates. This line of research will be developed in the following chapters of the thesis. Second, the effect of covariates can be taken into consideration in a manner similar to Breslow (1974) and Dabrowska *et al.* (1992). The null hypothesis would then be equivalent to checking the proportionality of the effect of a binary covariate (such as a group indicator or a discretised covariate), assuming the other covariate effects to be proportional. However, an extension to the Cox regression model with continuous covariates along the lines of Lin (1991) may not be possible. Instead, as indicated above, we extend the framework to continuous covariates by defining new notions of partial order in this case. The third generalisation may involve the cumulative γ-rate functions considered by Dabrowska *et al.* (1989), which includes as a special case the cumulative hazard function and the odds ratio function.

A nice feature of the graphical methods suggested here is that they produce smooth plots, even for small sample sizes. Thus the user need not be wary of reading too much from the shape of the plot. Further, as demonstrated by the examples, these graphical tools are quite powerful in detecting departures from the PH assumption in the direction of ordered alternatives.

Other researchers have used the work in this chapter to advance the literature in different ways. The measure of non-proportionality of hazards developed here (2.1) and the main result on weak convergence and asymptotic distribution of ordinary Stieljes integral of a stochastic process (Theorem 2.3.1; Theorem 3.1 of Sengupta *et al.*, 1998) have been particularly useful in this context. Specifically, tests for the proportional odds model (Dauxois and Kirmani, 2003), relative risk (Kirmani and Dauxois, 2003), of the Koziol-Green model (Koziol and Green, 1976) against monotone conditional odds for censoring (Kirmani and Dauxois, 2004), proportionality of cumulative incidence functions of competing risks (Dauxois and Kirmani, 2004), of equality of survival functions against monotone ratio (Dauxois and Kirmani, 2005), and of proportional odds with interval-censored data (Sun *et al.*, 2007) have made good use of the above ideas. The work has also been discussed in Alvarez-Andrade *et al.* (2007a) and in a review article on hazard ratios (Andersen, 1998).

Appendix to Chapter 2

Proof of Theorem 2.3.1

Consider the function $h : D[0,\infty)^p \times D[0,\infty)^q \longrightarrow D[0,\infty)^{pq}$ defined by $h(\boldsymbol{k},\boldsymbol{x})(t) = \boldsymbol{k}(t) \otimes \boldsymbol{x}(t)$. It is easy to show that h is continuous at all points $(\boldsymbol{k},\boldsymbol{x})$ such that $\boldsymbol{k}$ is rcll and $\boldsymbol{x}$ is continuous. The probability that $(\boldsymbol{k},\boldsymbol{X})$ does not belong to the continuity set of h is the same as the probability that $\boldsymbol{X}$ does not belong to $D[0,\infty)^q - C[0,\infty)^p$. The assumptions of the theorem ensures that this probability is zero. Therefore $\boldsymbol{K}_n(.) \otimes \boldsymbol{X}_n(.) \xrightarrow{D} \boldsymbol{k}(.) \otimes \boldsymbol{X}(.)$ by virtue of the continuous mapping theorem.

Now consider the function $f : D[0,\infty)^{pq} \longrightarrow \mathbb{R}^{pq}$ defined by $f(\boldsymbol{x}) = \int_0^\tau \boldsymbol{x}(t)dt$. To show that f is continuous, let $\boldsymbol{x}_n \longrightarrow \boldsymbol{x}$ in $D[0,\infty)^{pq}$ and notice that every component of $f(\boldsymbol{x}_n)$ converges to the corresponding component of $f(\boldsymbol{x})$ by the dominated convergence theorem. Since the domain and range of f are spaces equipped with product topologies, this implies that $f(\boldsymbol{x}_n)$ converges to $f(\boldsymbol{x})$. Therefore f is continuous and the result of the theorem follows from the continuous mapping theorem.

Consistency of the variance estimator (2.3)

Assuming that $Y_j(t)/a_n \xrightarrow{P} y_j(t)$ for $j = 1,2$ pointwise on $[0,\tau]$, we have $a_n V(t) \xrightarrow{P} v(t)$ in $D[0,\infty)$ under the usual Rebolledo conditions, where

$$v(t) = \int_0^t \left[\frac{d\Lambda_2(u)}{y_1(u)} + \frac{d\Lambda_1(u)}{y_2(u)} \right].$$

Let us also assume that $K_i \xrightarrow{P} k_i$ for $i = 1,2$, and that each of the functions v, k_1 and k_2 is continuous. In view of (2.9), we only have to show that $a_n V_{ij} \xrightarrow{P} v_{ij}$, $i = 1,2, j = 1,2$. We write v_{ij} as $\psi(k_{ij}, \phi(k_j, \nu))$, where ψ and ϕ are functions from $D[0,\infty) \times D[0,\infty)$ to $\mathbb{R}$ and $D[0,\infty)$, respectively, defined as

$$\begin{aligned} \psi(k,l) &= \int_0^\tau k(s)l(s)ds, \\ \phi(k,l)(t) &= \int_0^\tau k(s)l(s \wedge t)ds. \end{aligned}$$

In such a case $a_n V_{ij} = \psi(K_i, \phi(K_j, a_n V))$. The convergence of $a_n V_{ij}$ to v_{ij} in distribution is proved by showing that $\phi(K_j, a_n V) \xrightarrow{D} \phi(k_j, \nu)$. Since the limit of convergence in either step is deterministic, we can invoke the continuous mapping theorem and show that the functions ϕ and ψ are continuous at the limit points. To show the continuity of ϕ, let (k_{jn}, ν_n) be a sequence in $D[0,\infty) \times D[0,\infty)$ converging to (k_j, ν). Thus $k_{jn} \longrightarrow k_j$ and $v_n \longrightarrow v$ in $D[0,\infty)$. Since k_j and v are assumed to be continuous, prop. 1.17(b) of Jacod and Shiryayev (1980, p. 292) ensures that for each t, $\sup\limits_{s\le t}|k_{jn}(s) - k_j(s)| \longrightarrow 0$ and $\sup\limits_{s\le t}|v_n(s) - v(s)| \longrightarrow 0$. Note that $\phi(k,v) \in C[0,\infty)$. It follows that for $s \in [0,\tau]$,

$$\begin{aligned}
|\phi(k_{jn}, \nu_n)(s) - \phi(k_j, v)| &= \left|\int_0^\tau [k_{jn}(t)(v_n(s\wedge t) - v(s\wedge t)) + v(s\wedge t)(k_{jn}(t) - k_j(t))]\, dt\right| \\
&\le \sup_{s\in[0,\tau]} |v_n(s) - v(s)| \,. \int_0^\tau |k_{jn}(t)|\, dt + \tau. \sup_{s\in[0,\tau]} |v(s)| \,. \sup_{s\in[0,\tau]} |k_{jn}(s) - k_j(s)| \,.
\end{aligned}$$

Thus $\phi(k_{jn}, \nu_n)$ converges to $\phi(k_j, \nu)$ locally uniformly. Therefore $\phi(k_{jn}, \nu_n)$ converges to $\phi(k_j, \nu)$ in $D[0,\infty)$, and ϕ is continuous at (k_j, ν). The continuity of ψ at $(k_i, \phi(k_j, \nu))$ is proved in a similar manner.

Chapter 3

Testing for Proportional Hazards against Ordered Alternatives with respect to Continuous Covariates

3.1 Chapter summary

Several two-sample tests of the proportional hazards assumption against ordered alternatives have been proposed; see Chapter 1 for discussion. Gill and Schumacher (1987) and Deshpande and Sengupta (1995) considered the monotone hazard ratio alternative, while we (Sengupta *et al.* (1998), our Chapter 2) developed a test against the weaker alternative of monotone ratio of cumulative hazards. In this chapter, based on Bhattacharjee (2007a), we propose a natural extension of these partial orders to the case of continuous covariates. We develop tests for the proportional hazards assumption against ordered alternatives and a graphical method to identify the nature of departures from proportionality. The proposed tests do not make restrictive assumptions on the underlying regression model, and are applicable in the presence of multiple covariates and frailty. Small sample performance and applications to real data highlight the usefulness of the framework and methodology.

3.2 Introduction

As discussed in Chapter 1, testing the proportional hazards assumption is important for empirical studies and has been an active area of 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. The omnibus tests usually have low power, while the semiparametric alternatives typically make unverifiable assumptions about the shape of the regression function. Further, when the PH assumption does not hold, applied researchers require additional information regarding the nature of the covariate effects. In this context, it is often useful to explore whether the hazard rate for one level of the covariate increases in lifetime relative to another level, particularly when the covariate is discrete (two-sample or k-sample setup); for further discussion, see Section 1.2.4.

In the two-sample setup, Gill and Schumacher (1987) and Deshpande and Sengupta (1995) developed analytical tests of the PH hypothesis against the alternative of 'increasing hazard ratio', which is equivalent to convex partial order of the lifetime distribution in the two samples. In Chapter 2 (Sengupta *et al.*, 1998), we proposed a two-sample test of the PH model against the weaker alternative hypothesis of 'increasing ratio of cumulative hazards' (star ordering of the two samples). As discussed earlier (Section 1.2.4 and Chapter 2), the above alternative hypotheses ('increasing hazard ratio' and 'increasing ratio of cumulative hazards') provide explanations for the phenomenon of 'crossing hazards' often found in applications. These two-sample tests are useful for analysing survival data because, not only are they powerful in detecting departures from proportionality, they also provide further clues about the nature of covariate dependence. However, their applicability is limited because many important covariates in biomedical or economic applications are continuous in nature (Horowitz and Neumann, 1992).

In this chapter, we extend partial orders in the above two-sample problems to the case of continuous covariates. This extension is particularly motivated by applications in biomedicine and economics where covariate effects typically change monotonically over lifetime.[1] Based on examples from the applied literature as well as new applications, we argue that the proposed partial orders provide meaningful alternatives to the PH model in the continuous covariate case.

[1] Often the prognostic effects decay over time, but sometimes they also increase for certain range of covariate values.

We propose tests of the PH model against such ordered departures and study their asymptotic properties. Our framework does not assume any specific underlying regression model, and the tests are applicable in the presence of additional covariates – observed or unobserved. Monte Carlo studies and applications to real data highlight the advantages of the proposed methods.

The current chapter, based on Bhattacharjee (2007a), is organised as follows. In Section 3.2, we develop notions of ordered alternatives to the PH model in the case of continuous covariates. Tests of the PH assumption against such partial orders are constructed and their asymptotic properties studied in Section 3.3, and issues regarding implementation and extensions are discussed in Section 3.4. Small sample properties are studied in Section 3.5, while two real life applications are presented in Section 3.6. We also discuss modeling non-proportional covariate effects and develop a related graphical test. Section 3.7 concludes.

3.3 Partial orders with respect to a continuous covariate

Partial orders of lifetime distributions are commonly used in theory and applications. The two most popular notions of partial ordering, namely convex ordering and star ordering (Kalashnikov and Rachev, 1986; Sengupta and Deshpande, 1994), offer useful interpretations in terms of monotonicity of ratios of hazard and cumulative hazard functions respectively over time; for further discussion, see Section 1.1.1. Therefore, they describe useful and intuitively appealing ways to characterise departures from the PH model in two samples and in the competing risks framework. Gill and Schumacher (1987), Deshpande and Sengupta (1995) and Sengupta *et al.* (1998) (our Chapter 2) consider several empirical applications where the departure from the PH model in two samples is evident from the fact that the ratio of the hazard rates is not constant over the lifetime; see also Andersen (1998).

For the two-sample setup, Gill and Schumacher (1987) and Deshpande and Sengupta (1995) developed tests of the PH model against the "increasing hazard ratio" alternative, which is equivalent to convex ordering of the life-time distribution in one sample with respect to the other. In Chapter 2 (based on Sengupta *et al.*, 1998), we constructed a test against the weaker alternative hypothesis of "increasing ratio of cumulative hazards" (star ordering of the two samples). Sengupta and Bhattacharjee (1994) (Section 2.7), Deshpande and Sengupta (1995) and Dauxois and Kirmani (2004) extend these tests to the competing risks problem.

The following definitions describe natural extensions of the above partial orders to the continuous covariate case. Let T be a lifetime variable, X a continuous covariate and let $\lambda(t|x)$ denote the hazard rate of T, given $X = x$, at $T = t$.[2]

Definition 3.2.1. *The lifetime random variable T is defined to be increasing hazard ratio for continuous covariate ($IHRCC$) with respect to the covariate X if, whenever $x_1 > x_2$, $\lambda(t|x_1)/\lambda(t|x_2) \uparrow t$. In other words, the lifetime distribution conditional on the lower covariate value is convex ordered with respect to that conditional on the higher value:*

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

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

Definition 3.2.2. *The lifetime random variable T is defined to be increasing cumulative hazard ratio for continuous covariate ($ICHRCC$) with respect to X if, whenever $x_1 > x_2$,*

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

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

Definition 3.2.3. *The lifetime random variable T is defined to be increasing then decreasing hazard ratio for continuous covariate ($IDHRCC$) with respect to the covariate X if, there exists a point x within the range of X such that, T is $IHRCC$ on the interval $(-\infty, x)$ and $DHRCC$ on the interval (x, ∞). Similarly, we can define decreasing then increasing hazard ratio for continuous covariate ($DIHRCC$).*

Definitions 3.2.1 and 3.2.2 describe notions of positive ageing with respect to a continuous covariate. The higher the covariate, the faster the ageing of the individual – a situation which is

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

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

Examples of such partial orders are common in applications. In Section 1.1.2, we have discussed an application to survival with malignant melanoma. Analysing these data, Andersen *et al.* (1993) observe that, while "hazard seems to increase with tumor thickness" (pp. 389), the plot of estimated cumulative baseline hazards for patients with '2 mm $\leq$ tumor thickness $<$ 5 mm' and 'tumor thickness $\geq$ 5 mm' against that of patients with 'tumor thickness $<$ 2 mm' reveal "concave looking curves indicating that the hazard ratios decrease with time" (pp. 544–545). In fact, it is commonly observed in medical settings that treatment effects of an active drug decays with time (Therneau and Grambsch, 2000; Scheike and Martinussen, 2004). Similar evidence has also been noted in the applied econometrics literature. Using French data on unemplyment durations, Jayet and Moreau (1991) observe that the ratio of hazard function for individuals in the age groups 24–28 years to that for 37–40 years increases with duration of unemployment upto approximately 120 days.

Definition 3.2.3 describes a notion of non-monotonic departure from the PH model, with respect to the effect of a continuous covariate. An application considered later in the chapter demonstrate evidence of such non-monotonic departures. The following examples illustrate some simple data generation processes (DGPs) that generate monotone and non-monotonic departures from the PH assumption with respect to a continuous covariate.

Example 3.2.1. Consider the hazard regression model with time varying coefficients (Murphy and Sen, 1991; Martinussen *et al.*, 2002) discussed in Sections 1.2.4 and 1.2.7.4. Assume the hazard function $\lambda(t|x) = \lambda_0(t).\exp(\beta(t).x)$, where x is a continuous covariate and $\beta(.)$ is an increasing function of lifetime t (1.12). This model is appropriate when the prognostic value of the covariate is expected to be higher at higher lifetimes. Then, if $x_1 > x_2$, $\lambda(t|x_1)/\lambda(t|x_2) = \exp(\beta(t).(x_1 - x_2))$ is increasing in t. In other words, the lifetime random variable T is $IHRCC$ with respect to the covariate X. Conversely, if $\beta(.)$ is a decreasing function of the lifetime, T would be $DHRCC$ with respect to X, a feature commonly observed in empirical studies. Put differently, the hazard regression model with time varying coefficients exhibits $IHRCC$ ($DHRCC$) partial order if and only if the integrated (or cumulative) regression effect $B(t) =$

$\int_0^t \beta(s)ds$ is a convex (concave) function over the lifetime.

Example 3.2.2. Consider a changepoint survival model given by the cumulative hazard function $\Lambda(t|x) = \Lambda_0(t).\exp(I\,(t > t^*)\,.\beta x)$, where x is the covariate, $I(.)$ the indicator function, and t^* is a lifetime in the interior of the sample space. This is a model where initially the covariate has no effect on the lifetime. The effect of the covariate begins as soon as the lifetime crosses a certain threshold t^*, and it lifts the distribution function upto a level where it would have been, if the effect of the covariate would have persisted over the entire past life of the lifetime variable. If $\beta > 0$, this model is $ICHRCC$, but not $IHRCC$.[3] This kind of model may be useful in analysing the effect of active labour market programmes on unemployment duration, where the effect may become significant only around the time when unemployment benefits are terminated; see, for example, Narendranathan and Stewart (1993). More generally, the hazard regression model with time varying coefficients has $ICHRCC$ partial order if and only if the integrated (or cumulative) regression effect is star-shaped; the converse holds for $DCHRCC$ partial order.

Example 3.2.3. Consider the hazard function $\lambda(t|x) = \lambda_0(t).\exp(\beta(t).\,|x - a|)$, where x is the covariate, a is a point on the covariate space, and $\beta(.)$ is an increasing function of lifetime t. This model is neither $IHRCC$ nor $DHRCC$, but it is $DIHRCC$; it is $IHRCC$ on one region of the covariate space $(x > a)$, and $DHRCC$ on another region $(x < a)$. An application where such a feature is observed is the effect on mother's age on infant mortality. Because of physiological reasons, mortality is lowest around an optimal childbearing age; however, keeping mother's age fixed, the effect itself declines with age of the child (Bhalotra and Bhattacharjee, 2001). Another application is considered later in the chapter (Section 3.6).

As the above examples illustrate, the notions of ordering introduced in Definitions 3.2.1, 3.2.2 and 3.2.3 encompass a wide range of non-PH situations, and are potentially useful in many empirical applications. There may be a number of different explanations for changes in the covariate effects over lifetime. In fact, in many applications, monotone departures from the PH model may be more reasonable even from a theoretical point of view. Examples include medical

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

applications where one expects the prognostic relevance of some covariates to decay, or even disappear, in the long run (Pocock *et al.*, 1982; Therneau and Grambsch, 2000). Similar decline in covariate effects are observed in economic studies on the effect of benefits on unemployment duration (Narendranathan and Stewart, 1993) and on the effect of macroeconomic conditions on firm exits (Bhattacharjee *et al.*, 2008a). Construction of tests of the PH model against monotone alternatives with respect to continuous covariates is therefore important.

The above examples also demonstrate typical patterns of time varying coefficients when proportionality does not hold. These are useful for modeling ordered departures ($IHRCC$ or $DHRCC$) as well as non-monotonic violations ($IDHRCC$ or $DIHRCC$) of the PH assumption. Using the empirical applications (Section 3.6), we will demonstrate how such time varying covariate effects can be used, in combination with the proposed tests, to draw useful inferences in non-PH situations.

3.4 Test statistics

Several two-sample tests of the PH model against monotone alternatives exist in the literature. For a continuous covariate, a natural approach for testing the PH assumption against ordered alternatives $IHRCC$ and $ICHRCC$ (and their duals) would be repeated applications of the corresponding tests in the two-sample setup. In this chapter, we consider the two-sample test statistics proposed in Gill and Schumacher (1987) (T_{GS}) and Section 2.2 (also Sengupta *et al.*, 1998) (T_{SBR}).

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

3.4.1 Monotone hazard ratio

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

$$T_{GS,std} = \frac{T_{GS}}{\sqrt{\widehat{\text{Var}}\left[T_{GS}\right]}}, \tag{3.1}$$

where

$$\begin{aligned}
T_{GS} &= T_{11}T_{22} - T_{12}T_{21}, && (3.2)\\
\widehat{\text{Var}}\left[T_{GS}\right] &= T_{21}T_{22}V_{11} - T_{21}T_{12}V_{12} - T_{11}T_{22}V_{21} + T_{11}T_{12}V_{22}, && (3.3)\\
T_{ij} &= \int_0^{\tau} L_i(t)d\widehat{\Lambda}_j(t), (i,j=1,2), &&\\
V_{ij} &= \int_0^{\tau} L_i(t)L_j(t)\{Y_1(t)Y_2(t)\}^{-1}d\left(N_1+N_2\right)(t), (i,j=1,2), &&
\end{aligned}$$

τ is a random stopping time,[4] $L_1(t)$ and $L_2(t)$ are two predictable processes, and for the j-th sample ($j = 1, 2$), $\Lambda_j(t)$ is the cumulative hazard function and $\widehat{\Lambda}_j(t)$ its Nelson-Aalen estimator, $Y_j(t)$ denotes the number of individuals on test at time t, and $N_j(t)$ the counting process for the number of failures in the sample at time t.

Gill and Schumacher (1987) show that the unstandardised test statistic (T_{GS}) has mean zero under the null hypothesis (PH) and positive (negative) mean if the hazard ratio $\lambda_1(t)/\lambda_2(t)$ is monotonically increasing (decreasing) in t on $[0, \infty)$ and $L_1(.)$ and $L_2(.)$ are so chosen that $L_1(t)/L_2(t)$ is monotonically decreasing, and that its standard error falls to zero as sample size increases to ∞ under both the null and alternative hypotheses. Hence, while the standardized test statistic $T_{GS,std}$ is asymptotically standard normal under the null hypothesis, the mean increases (decreases) to ∞ ($-\infty$) under the alternative hypotheses of monotonically increasing (decreasing) hazard ratio. In many applications, L_1 and L_2 are chosen corresponding to the Gehan-Wilkoxon and log rank tests, where $L_1 = Y_1Y_2$ and $L_2 = Y_1Y_2(Y_1 + Y_2)^{-1}$, so that $L_1(t)/L_2(t)$ is monotonically decreasing in t.

For testing $H_0 : PH$ vs. $H_1 : IHRCC$, we propose the following procedure. We fix $r > 1$,

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

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

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

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

and

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

for $i, j = 1, 2$.

Therefore, our test statistics are:

$$T_{GS}^{(\max)} = \max\left\{T_{GS,std}(x_{11}, x_{12}), T_{GS,std}(x_{21}, x_{22}), \ldots, T_{GS,std}(x_{r1}, x_{r2})\right\}, \tag{3.4}$$

$$T_{GS}^{(\min)} = \min\left\{T_{GS,std}(x_{11}, x_{12}), T_{GS,std}(x_{21}, x_{22}), \ldots, T_{GS,std}(x_{r1}, x_{r2})\right\}, \tag{3.5}$$

and

$$\overline{T}_{GS} = \frac{1}{r}\sum_{l=1}^{r} T_{GS,std}(x_{l1}, x_{l2}). \tag{3.6}$$

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

3.4.2 Monotone cumulative hazard ratio

The form of the test statistic proposed in Section 2.2 (also Sengupta *et al.*, 1998), for testing the proportional hazards model against the 'increasing cumulative hazard ratio' (star partial order) alternative, is similar to $T_{GS,std}$. The standardised statistic[5] is given by

$$T_{SBR,std} = \frac{T_{SBR}}{\sqrt{\widehat{\text{Var}}\left[T_{SBR}\right]}}, \tag{3.7}$$

where

$$\begin{aligned}
T_{SBR} &= S_{11}S_{22} - S_{12}S_{21}, &(3.8)\\
\widehat{\text{Var}}\left[T_{SBR}\right] &= S_{21}S_{22}W_{11} - S_{21}S_{12}W_{12} - S_{11}S_{22}W_{21} + S_{11}S_{12}W_{22}, &(3.9)\\
S_{ij} &= \int_0^{\tau^*} K_i(t).\widehat{\Lambda}_j(t).dt, (i,j=1,2),\\
W_{ij} &= \int_0^{\tau^*}\int_0^{\tau^*} K_i(t).K_j(s).K_j(s).W\left(\min(s,t)\right)dsdt, (i,j=1,2),\\
W(t) &= \int_0^t \left(Y_1(s)Y_2(s)\right)^{-1}.d\left(N_1+N_2\right)(s),
\end{aligned}$$

τ^* is a large lifetime with $\Lambda_j(\tau^*) < \infty, j = 1,2,$[6] and $K_j(t)(j = 1,2)$ are right continuous functions with left limits (rcll functions) that need not be predictable processes.

As shown in Section 2.3, this standardised test statistic is also asymptotically standard normal under the null hypothesis of proportional hazards. Under the monotone cumulative hazard ratio alternative, it is asymptotically normal with mean diversing to ∞ ($-\infty$) accordingly as the cumulative hazard ratio $\Lambda_1(t)/\Lambda_2(t)$ is monotonically increasing (decreasing) in t on $[0,\infty)$ and K_1 and K_2 are so chosen that $K_1(t)/K_2(t)$ is a decreasing process.

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

[5] The notation for the test statistic in Chapter 2 is $Q_{K_1K_2}$, which emphasizes the important role for the weight functions. Since our focus here is on values of the test statistic conditional on different covariate pairs, we choose the simpler notation T_{SBR} and suppress the dependence on weight functions.

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

processes $N(t, x_{l1})$ and $N(t, x_{l2})$). Thus, we have:

$$T_{SBR}^{(\max)} = \max\left\{T_{SBR,std}(x_{11}, x_{12}), T_{SBR,std}(x_{21}, x_{22}), \ldots, T_{SBR,std}(x_{r1}, x_{r2})\right\}, \tag{3.10}$$

$$T_{SBR}^{(\min)} = \min\left\{T_{SBR,std}(x_{11}, x_{12}), T_{SBR,std}(x_{21}, x_{22}), \ldots, T_{SBR,std}(x_{r1}, x_{r2})\right\}, \tag{3.11}$$

and

$$\overline{T}_{SBR} = \frac{1}{r}\sum_{l=1}^{r} T_{SBR,std}(x_{l1}, x_{l2}). \tag{3.12}$$

3.4.3 Large sample results

We now derive the large sample results for the proposed test statistics, using the counting process methods (Gill and Schumacher, 1987; Andersen *et al.*, 1993) and a result on convergence of ordinary Stieljes integral of a stochastic process proved earlier (Theorem 2.3.1 in the thesis and Theorem 3.1 in Sengupta *et al.*, 1998). It is also indicated how these results can be used, in combination with extreme value theory, to obtain p-values of $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$.

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

Assumption 3.3.1 *For each l, $l = 1, 2, \ldots, r$, let $L_1(x_{l1}, x_{l2})(t)$ and $L_2(x_{l1}, x_{l2})(t)$ be predictable processes indexed on the pair of fixed covariate values (x_{l1}, x_{l2}).*

Assumption 3.3.2 *Let τ be a random stopping time. In particular, τ may be taken as the time at the final observation of the counting process $\Sigma_{l=1}^{r}\Sigma_{j=1}^{2}N(t, x_{lj})$. In principle, one could also have different stopping times $\tau(x_{l1}, x_{l2}), l = 1, \ldots, r$ for each of the r basic test statistics $T_{GS,std}(x_{l1}, x_{l2}), l = 1, \ldots, r$.*

Assumption 3.3.3 *The sample paths of* $L_i(x_{l1}, x_{l2})$ *and* $Y(t, x_{li})^{-1}$ *are almost surely bounded with respect to* t*, for* $i = 1, 2$ *and* $l = 1, \ldots, r$*. Further, for each* $l = 1, \ldots, r$*,* $L_1(x_{l1}, x_{l2})$ and $L_2(x_{l1}, x_{l2})$ *are both zero whenever* $Y(t, x_{l1})$ *or* $Y(t, x_{l2})$ *are.*

Assumption 3.3.4 *There exists a sequence* $a^{(n)}$*,* $a^{(n)} \longrightarrow \infty$ *as* $n \longrightarrow \infty$*, and fixed functions* $y(t, x)$*,* $l_1(x_{l1}, x_{l2})(t)$ *and* $l_2(x_{l1}, x_{l2})(t)$*,* $l = 1, \ldots, r$ *such that*

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

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

where $|l_i(x_{l1}, x_{l2})(.)|$ *are bounded on* $[0, \tau]$ for each $i = 1, 2$ and $l = 1, \ldots, r$*, and* $y^{-1}(., x)$ *is bounded on* $[0, \tau]$*, for each* $x \epsilon X$*.*[7]

Let the test statistics $T_{GS}^{(\max)}, T_{GS}^{(\min)}$ and $\overline{T}_{GS}$ be as defined earlier (3.4 – 3.6).

Theorem 3.3.1. *Let Assumptions 3.3.1 through 3.3.4 hold. Then, under* $H_0 : PH$*, as* $n \to \infty$*,*

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

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

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

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

(Proof in Appendix.)

Corollary 3.3.1.

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

and

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

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

(Proof in Appendix).

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

Corollary 3.3.2. *Given a vector* $\underline{w} = (w_1, w_2, \ldots, w_r)$ *of* r *weights, each possibly dependent on* x_{lj} $(l = 1, 2, \ldots, r; j = 1, 2)$ *but not on the counting processes* $N(t, x_{lj})$*, let us define the test statistics*

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

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

and

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

Let Assumptions 3.3.1 through 3.3.4 hold. Then, under $H_0 : PH$*, as* $n \to \infty$*,*

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

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

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

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

(Proof in Appendix).

Theorem 3.3.1, along with Corollaries 3.3.1 and 3.3.2, establish the asymptotic results for testing proportionality against monotone hazard ratio alternatives ($IHRCC$ and $DHRCC$) as well as non-monotonic violations ($IDHRCC$ or $DIHRCC$) of the PH assumption.

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

Assumption 3.3.5 For each $l, l = 1, 2, \ldots, r$, let $K_1(x_{l1}, x_{l2})(t)$ and $K_2(x_{l1}, x_{l2})(t)$ be stochastic processes with sample paths in $D[0, \infty)$ (*i.e.*, are right continuous and have left limits).

Assumption 3.3.6 Let τ^* be a positive lifetime such that $\Lambda(\tau^*, x_{lj}) < \infty, l = 1, 2, \ldots, r, j = 1, 2$.

Assumption 3.3.7 There exists a sequence $a^{(n)}$, $a^{(n)} \to \infty$ as $n \to \infty$, and deterministic functions $y(t, x)$, $k_1(x_{l1}, x_{l2})(t)$ and $k_2(x_{l1}, x_{l2})(t)$, $l = 1, \ldots, r$ such that

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

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

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

Let the test statistics $T_{SBR}^{(\max)}, T_{SBR}^{(\min)}$ and $\overline{T}_{SBR}$ be as defined earlier (3.10 – 3.12).

Theorem 3.3.2. *Let Assumptions 3.3.5 through 3.3.7 hold. Then, under* $H_0 : PH$*, as* $n \to \infty$*,*

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

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

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

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

(Proof in Appendix.)

Corollary 3.3.3.

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

(Proof in Appendix.)

Corollary 3.3.4. *Given a vector* $\underline{w} = (w_1, w_2, \ldots, w_r)$ *of* r *weights, each possibly dependent on* x_{lj} $(l = 1, 2, \ldots, r; j = 1, 2)$ *but not on the counting processes* $N(t, x_{lj})$*, let us define the test statistics*

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

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

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

Let Assumptions 3.3.5 through 3.3.7 hold. Then, under $H_0 : PH$*, as* $n \to \infty$*,*

(a) $P\left[T_{SBR,\underline{w}}^{(\max)} \leq z\right] \to \prod_{l=1}^{r}[\Phi(z/w_l)]$,

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

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

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

(Proof in Appendix).

Remark 3.3.1. Restricting the statistics $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$ to depend on a fixed number (r) of distinct pairs of points is crucial for the asymptotic results. This is because, the processes $T_{GS,std}(x_1, x_2)$ and $T_{SBR,std}(x_1, x_2)$ on the space $\{(x_1, x_2) : x_2 > x_1, x_1, x_2 \epsilon \mathcal{X}\}$ are pointwise standard normal and independent, and therefore the maxima (minima) diverges to $+\infty(-\infty)$ without having well-defined asymptotic distributions.

Remark 3.3.2. Corollaries 3.3.1 and 3.3.3 provide simple ways to calculate the p-values for the extremal test statistics $T_{GS}^{(\max)}$ and $T_{GS}^{\min}$ (and similarly, $T_{SBR}^{(\max)}$ and $T_{SBR}^{(\min)}$) provided r is reasonably large. Note that since r is held fixed it cannot increase to ∞, but with a value large enough (say, 20 or higher) the approximation is quite accurate.

Remark 3.3.3. Corollaries 3.3.2 and 3.3.4 can be used to weight the underlying test statistics by some measure of the distance between x_{l1} and x_{l2}. In other words, one can give higher weights to a covariate pair where the covariates are further apart. In practice, this is expected improve the empirical performance of the tests. We have, however, not used these weights in the empirical work in Sections 3.5 and 3.6.

3.5 Implementation and extensions

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

3.5.1 Small sample correction

Since the covariate under consideration is continuous, it is not feasible to construct the basic tests (T_{GS} and T_{SBR}) based solely on two distinct fixed points on the covariate space. In our implementation, we consider "small" intervals around the (randomly) chosen points, assuming the hazard function within these intervals to be approximately constant over covariate values. While the asymptotic distributions in Section 3.3 are based on specified points in the covariate space, the tests will be valid for small intervals around these points, provided the hazard function (for $T_{GS}^{(\max)}$, $T_{GS}^{(\min)}$ and $\overline{T}_{GS}$) or the cumulative hazard function (for $T_{SBR}^{(\max)}$, $T_{SBR}^{(\min)}$ and $\overline{T}_{SBR}$) is continuous at these points.

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

3.5.2 Choice of r and covariate pairs

The proposed tests take r, the number of covariate pairs, as fixed *a priori*. If the chosen value is sufficiently high (say, 20 or more), Corollaries 3.3.1 and 3.3.3 can be used to compute p-values very easily; the choice of r is not very critical otherwise. For the Monte Carlo study reported in Section 3.5, we choose $r = 45$.

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

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

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

3.5.3 Comparison with other tests

As discussed earlier, a convenient way to interpret the ordered alternatives considered here is through time varying coefficients in a multiplicative hazard regression model. In this sense, our tests are somewhat related to other analytical tests of time-dependant covariate effects proposed in the literature.

However, our approach embodies several important points of departure. First, our tests are based on the partial orders defined in Section 3.2 and not on any restrictive regression model. Second, some of the available analytical tests are based on partitioning the sample space of the lifetime variable into intervals (Anderson and Senthilselvan, 1982; Murphy, 1993) and consequently do not make use of the full information that the data offers. Our tests do not have this shortcoming. Third, unlike some other tests (Grambsch and Therneau, 1994; Scheike and Martinussen, 2004), our methods enable us to identify useful non-monotonic departures from the PH model, like $IDHRCC$ and $DIHRCC$. Fourth, while the previous tests merely identify violation of the constancy of covariate effects over the lifetime, our tests are based on explicit partial orders and provide additional insight into the nature of the regression relationship. This is useful for further inference and modeling. Finally, along with the test proposed by Scheike and Martunussen (2004), our tests have the advantage that tests of proportionality can be conducted sequentially for different covariates. This is often very useful in applications.

Notwithstanding these iportant differences, we compare the performance of the proposed tests against the popular test for time constant effects (PH model) due to Grambsch and Therneau (1994), using a simulation study (Section 3.5).

3.5.4 Choice between the proposed tests

The choice between the maxima, minima and average test statistics can be important in practice. The maxima and minima tests detect more complicated departures from the PH model ($IDHRCC$, $DIHRCC$, and their counterparts based on the cumulative hazard functions), and thereby facilitate detailed investigation of ordered covariate effects. On the other hand, as we shall see in the Monte Carlo simulations (Section 3.5), the adjusted average statistics outperform the maxima and minima tests in terms of power.

3.5.5 Extensions

The proposed ethodology offers several straightforward extensions.

k-sample problem

The proposed tests can be used to study monotone departures in k-sample (discrete covariate) problems. In this case, an *a priori* ordering of the k samples can be obtained using estimators of hazard ratio proposed in Gill and Schumacher (1987) or Chapter 2 (Sengupta *et al.*, 1998), or using the tree-structured modeling approach (Ahn and Loh, 1994). One can then easily apply the test for the PH model proposed here. The tests can also be similarly extended to the competing risks problem with more than 2 competing risks.

Different censoring and sampling plans

While our proposed methods are developed under the standard random censorship model (Fleming and Harrington, 1991; Andersen *et al.*, 1993), these can be easily extended to other censoring and sampling plans. For example, Bordes (2004) and Alvarez-Andrade *et al.* (2007) extend the counting process approach to estimation of the cumulative hazard function and proportional hazards regression based on progressive type-II censoring. Their results can be easily used to extend our results to this setup. Similarly, Sellke and Siegmund (1983) extend partial likelihood inference under the Cox regression model to the case of staggered (delayed) entry. Here, the counting process approach does not work. However, large sample results for our tests can still be derived using Theorem 2.3.1 in combination with Theorem 3.3.2.

Frailty

Like in the case of staggered entry, the counting process approach is not applicable in the presence of frailty. Under the shared frailty model, where individuals are clustered *a priori* based on the value of their shared but unobserved frailty, "quasi partial likelihood" inference was developed in Spiekerman and Lin (1998) based on empirical process theory. Similar theory for the univariate frailty model with a known one-parameter frailty distribution is developed in Kosorok *et al.* (2004). In either case, combining Theorem 2.3.1 with Theorem 3.3.2 gives us asymptotic results for the test statistics.

Presence of other covariates

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

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

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

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

3.6 Monte Carlo study

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

$$\lambda(t, x) = \lambda_0(t). \exp\left[\beta(t, x)\right], \tag{3.13}$$

where $\lambda_0(t)$ and $\beta(t, x)$ are chosen to assume a variety of functional forms. Note that, under model (3.13), the PH assumption holds if and only if $\beta(t, x)$ depends only on x. If, for fixed x, $\beta(t, x)$ increases (decreases) in t, we have the $IHRCC$ and $ICHRCC$ ($DHRCC$ and $DCHRCC$) alternatives. If, on the other hand, $\beta(t, x)$ increases in t over some range of the covariate space, and decreases over another (as in Example 3), the alternatives $IDHRCC$ or $DIHRCC$ may hold. While the proposed average tests are consistent for ordered alternatives to the null hypothesis of proportional hazards, our maxima and minima tests are consistent in both monotonic and non-monotonic cases.

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

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

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

and two specifications of the baseline hazard function $\lambda_0(t)$ $(= 2, 12t)$; see Table 3.5.1 for definitions and notations for the DGPs. Randomly right-censored data are generated using the Gauss 386 random number generator, where the covariate X is i.i.d. $U(-1,1)$, and the censoring time C is i.i.d. $U(0.2, 2.2)$. Of the eight, four DGPs belong to the null hypothesis of PH, and two have $IHRCC$ (also $ICHRCC$ specifications). The two remaining models, with $\beta(t,x) = \ln(t).|x|$, have $DIHRCC$ specifications, being $IHRCC$ and $ICHRCC$ over the range $x\epsilon[0,1]$ and $DHRCC$ and $DCHRCC$ over the range $x\epsilon[-1,0]$.

TABLE 3.5.1: DATA GENERATING PROCESSES

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

Table 3.5.2 reports, for each of the above eight data generation processes, the observed rejection rates (in percentage) of each of the test statistics, at 5 per cent confidence level, for different sample sizes. The reported percentages of rejection are based on 1000 Monte Carlo simulations in each case, and asymptotic distributions are used to compute the cut-offs. The covariate values considered are midpoints of each decile of the empirical distribution of realised covariate samples. Our test statistics are computed based on 45 random pairs of points on the covariate space ($r = 45$) in each case, given by each distinct combination of the above covariate values. Conditional on each covariate value, a sample of 20 nearest neighbour data points are used to construct the underlying two-sample test statistics T_{GS} and T_{SBR}.

TABLE 3.5.2: Rejection Rates (%) at the 5% Asymptotic Confidence Level

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

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

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

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

3.7 Empirical applications

Now, we illustrate the use of the tests with three applications: to (a) durations of contract strikes in the US (Kennan, 1985), (b) survival with malignant melanoma (Drzewiecki and Andersen, 1982; Andersen *et al.*, 1993), and (c) infant mortality in India (Bhalotra and Bhattacharjee, 2001).

3.7.1 Data on Strike Durations

The data, reported in Kennan (1985), pertain to durations of 566 contract strikes in the U.S., each involving 1000 workers or more, beginning during the period January 1968 to December 1976. Several authors have analysed these data, including Kennan (1985), Kiefer (1988),

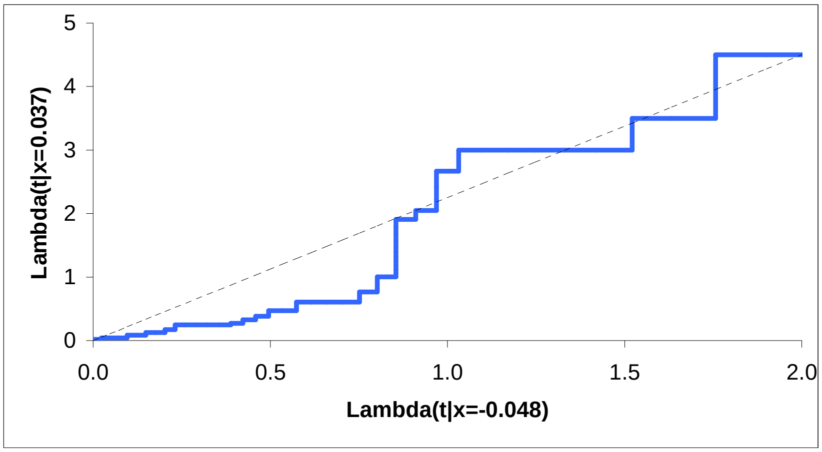


Figure 3-1: Lee-Pirie Plot of $\widehat{\Lambda}(t|x = 0.037)$ versus $\widehat{\Lambda}(t|x = -0.048)$.

Horowitz and Neumann (1992), and Neumann (1997). A important question of research interest, and of previous analyses, is the effect of business cycles (measured by production index) on strike duration. This production index represents the continuous covariate in our application. Since strike durations are also known to exhibit seasonal effects (Neumann, 1997), we use only the data on 292 strikes beginning in the first half of each year.

Empirical investigations of Kennan's strike data by previous authors suggest that the level of production index significantly affects strike duration (Kennan, 1985; Neumann, 1997). Higher values of the production index were associated with higher conditional probability of ending the strike, implying significant counter cyclical pattern of strike duration. However, the PH model specifies much more than merely the sign of the covariate effect. In order to graphically explore whether the data exhibit monotone departures from the PH model, we use Lee-Pirie plots (Lee and Pirie, 1981) of cumulative hazard functions conditional on various randomly chosen pairs of covariate values. Many of these plots indicate an increasing ratio of the hazards, as evident from the convexity (in some cases, star-shapedness) of the plot lending credence to *a priori* suspicion of monotone ordering of the $IHRCC$ type; as an illustration, see Figure 3-1, the Lee-Pirie plot conditional on covariate values -0.048 and 0.037).

TABLE 3.6.1: Tests of the PH model: Strike Duration data

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

Next, we apply our tests to these data (Table 3.6.1). Each of the tests were based on 150 pairs of distinct covariate values. The results of the tests confirm our *a priori* notion based on the above plots. The null hypothesis of PH model is rejected in favour of the alternative $IHRCC$ (and $ICHRCC$), at 5% level, with production index as the continuous covariate. This implies that the covariate effect of production index is such that, the duration distribution conditional on a higher value of the covariate is convex-ordered with respect to that conditional on a lower production index. In other words, the impact of production index on the hazard rate of strike duration increases in the duration of the strike.

Further, the maxima and minima tests provide additional information on the covariate pairs for which the basic test statistics attain their extreme values, which may be useful for modeling the nature of departures from proportionality. The maxima test-statistic $T_{GS}^{(\max)}$ is attained for the covariate pair $\{-0.0478, 0.0371\}$. The test statistic $T_{GS}^{(\min)}$ (covariate pair 0.0371 and 0.0675) has a p-value of 0.054, which provides some evidence of concave-ordering towards the upper end of the covariate space ($IDHRCC$).

To illustrate how this $IDHRCC$ nature can be incorporated into a regression model of strike durations, we present parameter estimates for three different models in Table 3.6.2. Model 1 is a simple Cox PH model, with production index as the continuous covariate. In Model 2, we allow for time-varying coefficients using the histogram sieve estimator proposed in Murphy and Sen (1991).[8] This model accomodates monotone departures from proportionality, in the nature of $IHRCC$ or $DHRCC$. In Model 3, we allow the coefficient of the covariate to vary not only over failure time, but also for covariate values. More specifically, we allow the coefficients to be

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

different for covariate values below and above 0.0371, enabling us to model departures of the $IDHRCC$ or $DIHRCC$ type. Here again, we use the estimators given by Murphy and Sen (1991) for inference.

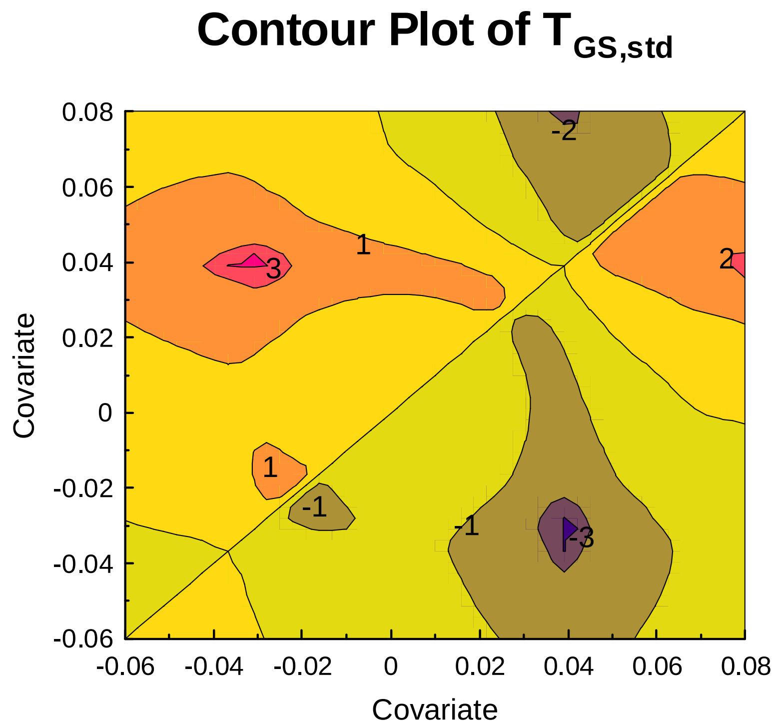
Model 1 indicates a significant impact of production index on the hazard rate of strike durations. However, this evidence is misleading. Model 3 estimates show that the true nature of covariate dependence is strikingly different. These time- and covariate-varying nature of the parameter estimates closely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

TABLE 3.6.2: Model Estimates: Strike Duration data

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

3.7.2 A related graphical test

Plotting the contours of the underlying standardised test statistics on a covariate $\times$ covariate two-dimensional plane provides an useful graphical tool for inference on monotonic and non-monotonic departures considered in this chapter. Figure 3-2 shows a contour diagram of the standardized test statistic $T_{GS,std}$ (smoothed using the Epanechnikov kernel) for the strike duration data. The significant height of the peaks and troughs indicate nonproportionality, and the shift in the slopes about the covariate value of approximately 0.04 indicate non-monotonic departures from proportionality about this point. The use of the plot here confirms the inference drawn from our analytical tests, and in particular helps in choosing the changepoint for the $IDHRCC$ pattern.



In applications with multiple covariates, similar graphical analysis can also provide valuable insights into the interaction between different covariates. With two continuous covariates x and z, one can obtain similar plots for different candidate functions $h(x, z)$ (see Section 3.4.5) to examine which of these provides the sharpest slopes in the contour plot. The candidate functions can sometimes be implied by the relevant application. For example, in survival of a series system with covariates measuring proneness to failure of the two components, the relevant function may be $\max(x, z)$. In other situations where there is no a priori knowledge about $h(.,.)$, one can either hypothesize linear functions of the form $x + \gamma z$, or find the function using regression methods. The identity of the covariate pairs with high (low) values for the maxima (minima) test statistics can be very helpful in such analyses.

3.7.3 Survival with Malignant Melanoma

The data pertain to 205 patients (148 of these are censored) with malignant melanoma (cancer of the skin) on whom a radical operation was performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark. The analysis of these data in Andersen *et al.* (1993) identifies tumor thickness as one of the most important prognostic factors. Further, Andersen *et al.* (1993) show that the Lee-Pirie plots of Nelson-Aalen estimates of the cumulative hazard functions for patients with '2 mm $\leq$ tumor thickness $<$ 5 mm' and 'tumor thickness $\geq$ 5 mm' against that of patients with 'tumor thickness $<$ 2 mm' are "concave looking curves", indicating possible violation of the PH model in favour of $DHRCC$. Similarly, the plot of the cumulative regression functions for log-thickness (Martinussen *et al.*, 2002) also indicate a distinct concave shape. However, surprisingly, the constant coefficient estimate lies almost entirely within the 95 percent confidence band of their estimates of the cumulative regression function $\int_0^t \beta(s)ds$.

Our analytical tests (Table 3.6.3) based on 100 pairs of distinct covariate values show that $T_{GS}^{(\min)}$ and $T_{SBR}^{(\min)}$ are significant at 1 percent level and $T_{GS}^{(\max)}$ is significant at 5 percent level, but $\overline{T}_{GS,Adj}$ and $\overline{T}_{SBR,Adj}$ are not significant. Further, $T_{GS}^{(\min)}$ and $T_{GS}^{(\max)}$ are attained for covariate pairs $\{1.9, 7.7\}$ and $\{1.0, 1.8\}$ respectively. This provides partial support for the observation in Andersen *et al.* (1993), in that the null of PH is rejected in favour of the alternatives $DHRCC$ and $DCHRCC$ over the upper range of the covariate space. However, in patients with small tumors, there is some evidence of an $IHRCC$ pattern (probably the reason why $\overline{T}_{GS,Adj}$ and $\overline{T}_{SBR,Adj}$ are not significant). The inference from the Murphy-Sen histogram sieve estimators (Table 3.6.4) is similar.

TABLE 3.6.3: Tests of the PH model: Malignant Melanoma Data

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

TABLE 3.6.4: MODEL ESTIMATES: MALIGNANT MELANOMA DATA

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

This demonstrates the usefulness of the proposed methods for detecting non-proportional covariate effects which previous tests fail to identify.

3.7.4 Child mortality in India

The third application is adapted from a study (Bhalotra and Bhattacharjee, 2001) of child mortality across the three Indian states of Kerala, West Bengal and Uttar Pradesh. As argued by Sen (1998), infant and child mortality are important indicators of quality of life, in that they vary widely across space and time, they contain substantial information about and social inequalities (including gender bias), and are quite strongly influenceable by economic policy. The literature highlights a host of determinants that affect child mortality – economic, socio-cultural and physiological, and identifies the importance of provision and access to welfare measures and community infrastructure.

There is substantial spatial variation in infant and child mortality within India. Here we consider data for the state of Kerala which has demographic features more typical of a middle-income country than of a poor developing country. A large number of covariates are included, covering economic, socio-cultural and physiological determinants of child mortality. The data are from the National Family Health Survey of 1992-93 and we use retrospective data for ten years for each ever-married women who had at least one live-birth during the ten years preceding the date of survey.

Our main focus is on the mother's age at childbirth, which has been identified in previous research as an important covariate (Martin *et. al.*, 1983; Trussell and Hammerslough, 1983; Pebley and Stupp, 1987; Guo and Rodríguez, 1992). However, the effect of maternal age on mortality outcomes depends critically on the age of the child. Children born to very young (teenage) mothers are expected to be disadvantaged, both at birth (because of physiological reasons) and during early childhood because very young mothers may not be able to provide adequate childcare. Similarly, the effect of a higher maternal age on child survival may be mixed; while children born to older mothers may be physiologically disadvantaged at birth, such mothers may be more experienced and better able to provide adequate childcare. Therefore, a priori, we expect the covariate effect to be negative and falling to zero at lower ranges of maternal age ($IHRCC$). The effect may be non-monotonic ($IDHRCC$) if the age benefits of better childcare provision at higher maternal ages are not strong.

As expected for a state with good socio-economic conditions (including maternal education and post-natal childcare provision), $T_{GS}^{(\max)}$ and $\overline{T}_{GS,Adj}$ are significant at 1% level, while $T_{GS}^{(\min)}$ is not ($r = 150, T_{GS}^{(\max)} = 6.72$ – covariate pair 23, 29 years, $T_{GS}^{(\min)} = -2.52, \overline{T}_{GS,Adj} = 0.30$). The Grambsch and Therneau (1994) test fails to reject the PH assumption, though the p-value is fairly small at 0.079.

We use a sequential testing procedure to identify all covariates with non-PH effects. Two other covariates, preceding and succeeding birth intervals, also demonstrated monotone covariate effects. Interestingly, our tests fail to reject the null hypothesis for one covariate, distance to nearest town, which was identified by the Grambsch and Therneau (1994) test to have time varying coefficients. However, the histogram sieve estimates of age-varying coefficients strongly support the inferences drawn using our tests for all the covariates.

The three applications considered here demonstrate the value of studying departures from the PH model with respect to continuous covariates in terms of monotonicity of the covariate effects. These examples also illustrate the use of our test statistics in identifying monotonic and non-monotonic structures in the data. Similar inference has been used in Bhattacharjee *et al.* (2008a, 2008b) in applications to business failures in the UK and the US, which we discuss in Chapter 7 (Sections 7.2 – 7.4).

3.8 Conclusion

In this chapter, we develop notions of partial ordering of lifetime distributions with respect to continuous covariates and propose tests of the PH model against such monotone or ordered departures. Departures of these kinds are common in applications. Therefore, both empirical and theoretical work in lifetime models need to have a framework flexible enough to accomodate these kinds of covariate dependence. Unlike other tests available in the literature, the proposed methodology works in very general situations and does not require any assumptions on the underlying regression models. Further, the methods offer a great deal of flexibility in terms of accommodating the effects of other covariates, both observed and unobserved.

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

Several promising areas of future research emerge from the research in this chapter. First, in the derivation of asymptotic results, we show that the basic underlying two-sample test statistics for distinct covariate pairs are independent of each other. This fact can be exploited to extend many familiar two-sample inference techniques to the case of continuous covariates. In Chapter 5, we will take this approach in developing tests for the absence of covariate dependence. Second, research can be directed towards extension of the proposed tests to models with unrestricted univariate frailty. The notions of partial ordering introduced in this chapter will be valid in this case, and we can in principle construct similar tests using estimators of the cumulative hazard function under such models. However, this inference problem is quite distinct from the one addressed here, because of identifiability restrictions and the different nature of estimators proposed in the literature (see, for example, Horowitz, 1999). In Chapter 5, we show that this problem is related to testing for the absence of covariate dependence, and develop tests for the PH model in the univariate frailty case. Third, estimation of semiparametric regression models under order restrictions motivated by the current work is an area of considerable research potential. In Chapter 4 (also Bhattacharjee, 2004a), we develop biased bootstrap methods for such order restricted inference on covariate effects, while Bayesian inference under oder restrictions on both covariate effects and ageing, and in the presence of frailty, is developed in Chapter 6.

Fourth, it will be useful to develop further inference on the changepoint in non-monotonic models using covariate pairs corresponding to the maxima and minima tests. Fifth, a somewhat related problem is inference on the unknown $h(.,.)$ function in the multiple covariate case. These problems will be retained for future work.

Appendix to Chapter 3

Proof of Theorem 3.3.1: It follows from Gill and Schumacher (1987) that, under PH, as $n \longrightarrow \infty$,

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

where

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

so that,

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

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

are asymptotically independent. In other words,

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

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

Following Gill and Schumacher (1987), let

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

Then

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

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

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

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

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

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

where

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

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

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

It follows that

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

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

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

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

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

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

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

$\square$

Proof of Corollary 3.3.1: 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.

$\square$

Proof of Corollary 3.3.2: From Theorem 3.3.1, we have:

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

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

The proof follows immediately.

$\square$

Proof of Theorem 3.3.2: It follows from Section 2.3 that, under H_0, as $n \longrightarrow \infty$,

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

where

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

so that,

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

Like Theorem 3.3.1, the proof will follow, if it further holds that

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

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

The essential difference in the arguments required to establish asymptotic distributions here, from those in Theorem 3.3.1, lie in the fact that the integrals considered in Theorem 3.3.1 are transformations of stochastic integrals, while here, they are functions of ordinary Steiljes integrals of stochastic processes.

Let us define

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

Then, by Rebolledo's central limit theorem and Theorem 2.3.1 (Theorem 3.1 in Sengupta *et al.*, 1998), we have, as $n\to\infty$,

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

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

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

Now, as in Theorem 3.3.1, invoking the δ-method of Gill and Schumacher (1987), it follows that

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

where

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

and under H_0,

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

It follows that

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

where $\sum = diag\left((\sigma^2_{SBR,l})\right), l=1,\dots,r$, and following similar arguments as Appendix to Chapter 2, it can be shown that $\sigma^2_{SBR,l}$ can be consistently estimated by $\widehat{\text{Var}}\left[T_{SBR}\left(x_{l1},x_{l2}\right)\right]$. Hence, it follows that

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

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

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

□

Proof of Corollary 3.3.3: Proof follows from extreme value theory and the δ-method, as in Corollary 3.3.1.

$\square$

Proof of Corollary 3.3.4: From Theorem 3.3.2, we have:

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

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

The proof follows immediately.

$\square$

Chapter 4

Estimation in nonproportional hazard regression models with monotone covariate effect

4.1 Chapter summary

In this chapter, based on Bhattacharjee (2003, 2004a), we build on the notion of ordered departures from proportionality introduced in Chapter 3, and propose estimation methods for hazard regression models under such order restrictions. In our proposed test of the proportional hazards assumption (Bhattacharjee, 2007a – our Chapter 3), the ordered alternative of special interest was that the lifetime distribution conditional on a higher covariate value was convex (concave) ordered with respect to that conditional on a lower value. Here we argue that hazard regression models with time varying coefficients provide an appropriate framework for studying such order restrictions. Building on a natural interpretation of these alternatives in terms of monotonicity of time varying coefficients, we use biased bootstrap methods to estimate the covariate effects when such monotone departures are known to hold. In particular, it is shown how order restricted estimation can be performed using biased bootstrap techniques such as adaptive bandwidth kernel estimators (Brockmann *et al.*, 1993) or data tilting (Hall and Presnell, 1999). The performance of the estimators is compared using simulated data, and their use is illustrated with applications from biomedicine and economic duration data. The methods are relatively simple to implement, and provide useful inference in nonproportional hazard situations.

4.2 Introduction

As discussed earlier in the thesis (Chapters 1 through 3), the proportional hazards assumption is often violated in applications, or may even be unreasonable from the point of view of relevant theory. Further, such violation leads to misleading inferences based on the Cox PH model. Johnson *et al.* (1982), Lagakos and Schoenfeld (1984), Solomon (1984), Struthers and Kalbfieisch (1986) and Lagakos (1988) discuss the effect of misspecification on inferences about the effects of explanatory variables, and Li *et al.* (1996) reports results of a large simulation study. Similarly, inaccurate inferences on the shape of the baseline hazard function has been discussed by Breslow *et al.* (1984), Stablein and Koutrouvelis (1985), Schemper (1992), Tubert-Bitter *et al.* (1994) and Hsieh (1996). Testing the PH model, particularly against the omnibus alternative, has therefore been an area of active research.

Recently, specific attention has focussed on testing the PH assumption against ordered alternatives. As discussed earlier, it is often of interest to explore whether the hazard rate for one level of the covariate increases in lifetime, relative to another level (*i.e.*, the hazard ratio increases/ decreases with lifetime). Such tests have the important advantage of providing inferences useful for regression modeling when the proportionality assumption does not hold. In the two-sample setup, Gill and Schumacher (1987) and Deshpande and Sengupta (1995) have constructed analytical tests of the PH hypothesis against the alternative of 'increasing hazard ratio', while Sengupta *et al.* (1998) (our Chapter 2) have proposed a test of the PH model against the weaker alternative hypothesis of 'increasing ratio of cumulative hazards'. In Chapter 3 (Bhattacharjee, 2007a), we developed a natural extension of such monotone ordering to the case of continuous covariates, and constructed tests for the proportional hazards model against these alternatives.

It is observed that monotone departures are common in economic and biomedical applications (Bhattacharjee, 2007a; Scheike, 2004), and provide useful information about the nature of covariate dependence. Besides, they may also be suggested by theory. Estimation of hazard regression models under such order restrictions is therefore important. These methods are appropriate when either analytical tests or relevant theory suggests order restrictions rather than proportionality.

A popular approach in the literature is to interpret violations of the PH model in terms of time varying coefficients (for a review, see Scheike, 2004). Several authors have suggested validation of the PH assumption by testing for time varying coefficients (see, for example, Grambsch and Therneau, 1994; Scheike and Martinussen, 2004), and several methods for estimation of these time varying coefficients have been proposed (see, for example, Zucker and Karr, 1990; Murphy and Sen, 1991; Martinussen *et al.*, 2002). However, none of these estimators consider the case when the covariate effects are order restricted. As discussed in Chapter 3, Bhattacharjee (2007a) considers testing the proportionality assumption against the alternative of order restricted covariate effects in a more general framework. This framework includes the hazard regression model with time varying coefficients as a special case.

In this chapter, we build on the notion of ordered departures from proportionality introduced in Bhattacharjee (2007a) and propose estimation methods for hazard regression models under such order restrictions. We argue that a natural framework for hazard regression models in such situations is the one that allows time varying coefficients. Further, building on a natural interpretation of these alternatives in terms of monotonicity of time varying coefficients (see Example 3.2.1), we use biased bootstrap methods to estimate the covariate effects when such monotone departures from proportional hazards hold. Small sample properties of the estimators are explored using simulated data. Algorithms are developed for using these methods, and the nature of inference derived is demonstrated using a couple of applications.

The chapter is organised as follows. In Section 4.2, we motivate modeling ordered depertures from PH by a hazard regression model with monotonically time-varying coefficients, while Section 4.3 briefly reviews some alternative methods for order restricted inference that can be potentially used in the current context. Sections 4.2 and 4.3 are based on Bhattacharjee (2003). Next, following Bhattacharjee (2004a), we discuss estimation of hazard regression models with order restricted covariate effects. The proposed estimation methodologies based on biased bootstrap methods are developed in Section 4.4. In Section 4.5, we illustrate the use of the estimators using simulations and two real life applications. Section 4.6 collects the concluding remarks.

4.3 A hazard regression model admitting order restrictions in covariate effects

As discussed in Chapter 3, Bhattacharjee (2007a) extended to the continuous covariate setup the notion of monotone hazard ratio in two samples developed in Gill and Schumacher (1987), Sengupta and Deshpande (1994) and Deshpande and Sengupta (1995). Let T be a lifetime variable, X a continuous covariate and let $\lambda\left(t|x\right)$ denote the hazard rate of T, given $X = x$, at $T = t$. Then, T is defined to be *increasing (decreasing) hazard ratio for continuous covariate ($IHRCC$ ($DHRCC$))* with respect to X if, whenever $x_1 > x_2$, the ratio $\lambda\left(t|x_1\right)/\lambda\left(t|x_2\right)$ is increasing (decreasing) in t (Definition 3.2.1).

Further, Bhattacharjee (2007a) showed that, within the context of the hazard regression model with time varying coefficients (Murphy and Sen, 1991; Martinussen *et al.*, 2002)

$$\lambda(t|x) = \lambda_0(t).\exp(\beta(t).x),$$

the lifetime random variable T is $IHRCC$ with respect to the covariate X if and only if the time varying covariate effect $\beta(.)$ is an increasing function of lifetime t (Example 3.2.1); this is also true in the presence of additional covariates or frailty. The converse holds for the partial order $DHRCC$.

The above result suggests that the time varying coefficients model (1.12) may be useful for regression modeling in situations where covariate effects are non-proportional. However, it is also clear that the partial orders $IHRCC$ and $DHRCC$ are defined in more general settings than the time varying coefficients model. Therefore, before considering estimation under order restrictions, we characterise the additional assumptions embodied in the time varying coefficients hazard regression model.

First, we consider the single covariate case. For simplicity, we assume that the the lifetime variable T is discrete and takes values $0, 1, \ldots$, and that the covariate X takes three possible values: $l = 0$ (low), $m = 1$ (medium) and $h > m$ (high). Here, the most general hazard regression model is given by

$$\lambda(t|x) = \lambda_0(t). \exp\left[\gamma\left(t, x\right)\right], \qquad \gamma\left(t, 0\right) \equiv 0. \tag{4.1}$$

Without loss of generality, the exponential function can be substituted by any other monotonic positive valued function. We treat the hazard rate corresponding to the lowest covariate value, $\lambda(t|X = 0)$, as the baseline hazard rate $\lambda_0(t)$. Then, the time varying covariate effects, $\beta(t)$, implied by the general model (4.1) for various combinations of T and X are as follows:

Covariate	**Lifetime,** $T = 0$	**Lifetime,** $T = 1$	...	**Lifetime,** $T = k$	...
$X = l(= 0)$	$\beta(0)$ unrestricted	$\beta(1)$ unrestricted		$\beta(k)$ unrestricted	
$X = m(= 1)$	$\beta(0) = \gamma\left(0, 1\right)$	$\beta(1) = \gamma\left(1, 1\right)$		$\beta(k) = \gamma\left(k, 1\right)$	
$X = h(> 1)$	$\beta(0) = \gamma\left(0, h\right)/h$	$\beta(1) = \gamma\left(1, h\right)/h$		$\beta(k) = \gamma\left(k, h\right)/h$	

When the covariate is zero, the conditional hazard rate is the same as the baseline hazard, and $\beta\left(.\right)$ is completely unrestricted. Clearly, when the covariate is binary,[1] the time varying coefficients model coincides with the most general model, and an exact correspondence $\beta(t) = \gamma\left(t, 1\right)$ holds. However, when the covariate is not binary (takes more than two possible values), the time varying effect model holds only when two conditions are satisfied. First, the following scaling condition holds:

$$\frac{\gamma\left(t, x_1\right)}{x_1} = \frac{\gamma\left(t, x_2\right)}{x_2},$$

where x_1 and x_2 are any two non-zero values of the covariate X. Second, the logarithm of conditional hazard rates for different covariate values have to be proportional to each other.[2]

The first condition can be addressed by suitable transformations of the covariate X. The second assumption is more critical. However, if it fails to hold, we can use a histogram sieve (Grenander, 1981) to divide the covariate space into disjoint intervals within which the shape of $\beta\left(.\right)$ is approximately similar.[3] Thus, we can still construct an appropriate time varying

[1] In this case, without loss of generality, X can be assumed to take values 0 and 1.

[2] Note that proportionality of hazards implies the much stronger condition that logarithm of conditional hazard rates are constant over lifetime.

[3] See, for example, the applications reported in Sections 3.6.1 and 3.6.3, particularly Model 3 estimates in Tables 3.6.2 and 3.6.4

coefficients model and estimate the model using methods similar to those proposed in Murphy and Sen (1991); see discussion in Section 1.2.7. In fact, the above assumption can also be tested using methods similar to those developed by Murphy (1993) in the context of testing the PH model.

The above arguments can be simply extended to the continuous failure time case, as well as the continuous covariate case.

Next, let us consider the case when there are multiple covariates. From the above argument, it is clear that, even in the simplest case with 2 binary covariates the time varying coefficients model (1.12) may fail to hold. This is because of potential interaction between the covariates. However, in this case, the time varying coefficients model

$$\lambda(t|x,z) = \lambda_0(t).\exp\left[\beta_X(t).x + \beta_Z(t).z + \beta_{XZ}(t).x.z\right]$$

is exactly equivalent to the most general case.

Carrying this intuition to the case of several continuous covariates, it can be seen that the time varying coefficients model is valid under the additional assumption of additive covariate effects on the logarithm of conditional hazard rates. The assumption too can be tested. For example, one can use a Hausman-type test (Hausman, 1978), based on the difference between a consistent unrestricted nonparametric estimate and an efficient estimate under the additivity assumption. Under additivity, both the Murphy and Sen (1991) histogram sieve estimator and the Martinussen *et al.* (2002) estimator of the cumulative coefficients are efficient in that they attain the bounds based on efficient influence functions given by Sasieni (1992).[4] Further, as before, even when additivity is rejected, a valid time varying coefficients model can be built by placing histogram sieves on the product space of the covariates.

Thus, the time varying coefficients model incorporates two important assumptions relating to variation in the shape of the time varying coefficients over covariate values and to additivity of covariate effects. At the same time, even when these conditions are invalid, one can work with a modified model where the time varying coefficients are allowed to also vary with covariate

[4]Gozalo and Linton (2001) developed a similar test for additivity in nonparametric regression which can be modified to the current framework.

values and may include interaction between covariates. Further, it is easy to estimate these models, either by using histogram sieves along the lines of Murphy and Sen (1991) or using kernel based methods. All the above arguments also hold in the presence of frailty and time varying covariates.

In summary, the time varying coefficients model is potentially a very useful hazard regression model for order restricted inferences of the kind developed in this chapter and thesis.

4.4 Estimation under order restrictions

Curve estimation under shape constraints is of considerable interest in many applications. In the context of nonparametric regression, typical examples include the study of dose response experiments in medicine and the study of indirect utility, cost and production functions in economics, and pricing of options in finance, among others. In the context of hazard regression models, monotonicity of time varying coefficients provides a useful way to express order restrictions on covariate effects; see also discussion in Chapter 3 and Section 4.2.

Starting from the classic works of Hildreth (1954) and Brunk (1955), there exists a large literature on the problem of estimating monotone, concave or convex regression functions; for further discussion, see Barlow *et al.* (1972) and Robertson *et al.* (1988). More recently, attention has focussed on simple, smooth and efficient estimation of shape restricted regression functions. In the following subsections, we provide a very brief overview of some of these methods, focussing mainly on monotonicity (or isotonic regressions) and methods that are particularly attractive for hazard regression modeling. We also discuss our choice of estimation methodologies.

4.4.1 Isotonic regression approach

The isotonic regression approach (Barlow *et al.*, 1972; Hanson *et al.*, 1973) represents the most traditional method for estimating a nonparametric regression function under monotonicity constraints. The method obtains a least squares solution under the monotonicity restriction by the pool adjacent violators (PAV) algorithm. Whenever monotonicity is violated at a particular data point, the algorithm averages over neighbouring data, expanding the neighbourhood until monotonicity is restored. Newer generations of isotonic regression methods and their properties have been studied by Mukerjee (1988), Mammen (1991a, 1991b), Qian (1994), and others. In the sense that pooling is similar to expanding the bandwidth in the kernel regression framework, this method is similar to local adaptive bandwidths. The idea is also related to taut strings.

4.4.2 Estimation based on projections

Mammen *et al.* (2001) developed a general framework where the constrained smoothing problem can be interpreted as a projection of the unconstrained estimator in an appropriate Hilbert space. Special cases include smoothing spline and local polynomial methods; see also Ramsay (1988), Tantiyaswasdikul and Woodroofe (1994), Mammen and Thomas-Agnan (1999) and Mammen *et al.* (1999). In fact, Mammen *et al.* (2001) also show how the usual Nadaraya Watson nonparametric kernel regression estimator can also be interpreted as a projection.

4.4.3 Taut string method

The taut string method has its origins in isotonic regression, and specifically the familiar result that the greatest convex minorant of the data is a taut string and its derivative is the isotonic estimator (Barlow *et al.*, 1972; Leurgans, 1982). The taut string method is also related to the notion of excess mass which motivated the development of scale space view of kernel smoothing (SiZer maps) by Chaudhuri and Marron (1999, 2000).

Mammen and van de Geer (1997) developed the method further and extended its use to locally adaptive nonparametric regression. In recent times, the method has attracted substantial attention because of two main reasons. First, the work of Dümbgen (1998) and Davies and Kovac (2001) established a connection between taut string and the number and location of local extremes. This has important implications for testing qualitative order restrictions as well as nonparametric smoothing under order restrictions. Second, the method has been found to offer good control over the number of local extreme values. In other words, it does very well in detecting even low peaks without picking up artificial peaks.

4.4.4 Density regression approach

The density regression method proposed by Dette *et al.* (2006) is implemented in two steps. In the first step, the unconstrained Nadaraya Watson nonparametric regression estimator is obtained. In the second step, an isotonic estimator of the inverse regression function is obtained, using a different kernel from the first step. The resulting estimator has the desirable property that it is of the same order of smoothness as the unconstrained estimator.

The above four approaches are all based on intuitive and attractive ideas. However, we find the biased bootstrap methods discussed below more intuitively appealing, particularly because they are easy to implement and provide smooth estimates that agree with the corresponding unconstrained (kernel or sieve) estimators most of the time. Besides, there are issues relating to computational intensity and the lack of a simple way to visualise departures from hypothesized order restrictions.

4.4.5 Biased bootstrap methods

"Biased bootstrap" is usually taken to mean a weighted bootstrap procedure where the weights are chosen to satisfy the constraints imposed by the statistical model. Following Hall and Turlach (1999), we adopt a slightly different interpretation, where the notion is enlarged to also include reweighting data in a neighbourhood of the covariate space.

Data tilting

The data tilting method (Hall and Presnell, 1999; Hall and Huang, 2001) starts with an unconstrained estimator, and then reduces the relative weights on observations influential for violation of the maintained order restrictions. In this way, the method preserves the smoothness of the unconstrained estimator in large samples.

This idea is very attractive in the current context. It is closely related to influence functions and identification of influential observations, and can provide valuable information about the strength of the maintained order restriction. We also find it quite convenient to use in combination with the histogram sieve estimator (Murphy and Sen, 1991).

Local adaptive bandwidths

Local adaptive bandwidths (Brockmann *et al.*, 1993; Schucany, 1995) are based on a similar idea of leaving the unrestricted nonparametric kernel estimator unchanged at most places, and only reweighting in regions where the monotonicity property is violated. The reweighting is implemented by adjusting the bandwidth locally.

Like data tilting, the degree to which local adjustments are required can provide insights into the validity of the hypothesized order restriction. In this sense, the idea of local adaptive bandwidths is similar to SiZer maps (Chaudhuri and Marron, 1999, 2000. A closely related idea is adaptive weights smoothing, introduced by Polzehl and Spokoiny (2003) in the context of image denoising.

Data sharpening

In data sharpening (Choi and Hall, 1999), the idea is to modify the data just that little bit so that the estimates maintain the hypothesized order restrictions. Since such data modification works a bit like changing the bandwidth, the method is related to local adaptive bandwidths. Potentially, the approach is useful in our context. Extensions to hazard rate estimation with censored data were developed by Claeskens and Hall (2002). In a nonparametric regression context, data sharpening has been often used to adapt to sparse design density in certain regions (Choi *et al.*, 2000), and can be similarly used for estimating monotone regression curves. This is usually implemented by adjusting both the explanatory and the response variables prior to substitution into a local linear estimator. However, since most of our applications include time varying covariates, it is difficult to implement data sharpening methods directly. Specifically, if the response variable (lifetime) is raised, it will generate missing values in the time varying covariates.

4.4.6 Choice of estimation methods

There are several reasons guiding our choice of biased bootstrap methods, particularly data tilting and local adaptive bandwidths, for estimating hazard regression models with ordered covariate effects. First, as we will demonstrate in the later sections, they are simple to implement with the kernel and histogram sieve estimators. They also offer intuitively appealing interpretation in terms of local bandwidths and sampling weights.

Second, since they modify the underlying unrestricted estimator only in regions where they are non-monotone (Hall and Turlach, 1999), the biased bootstrap methods preserve the degree of smoothness in the original estimator.[5] Smoothness is an attractive property which is not shared by some other estimation methods, such as those based on projections (Hall and Huang, 2001). Effectively, the adaptive bandwidth estimator smoothes away "spurious wiggles" by increasing the local bandwidth at the middle of the wiggles, and reducing the bandwidth towards the boundaries. Data tilting estimators achieve a similar objective by reducing the sampling weight on observations that are atypical, and help create an illusion of non-monotonic covariate effects. While in large samples, the monotone nature of the data would dominate, and biased bootstrap estimators may not be necessary, these methods usually produce more visually appealing curve estimates in small samples (Farmen and Marron, 1999).

Thirdly, the biased bootstrap methods facilitate inferences relating to influence functions (influential observations) or local violations of maintained order restrictions. In case the maintained order restrictions do not hold, this helps in understanding why this might be so. Further, since both these methods are based on modifying unconstrained estimators only in "small" regions where they are non-monotone, they also provide means for testing the strength of the maintained monotone relationship. With respect to adaptive bandwidth estimation, this testing philosophy is very similar to SiZer maps (Chaudhuri and Marron, 1999, 2000; see also Bowman *et al.*, 1998 and Fisher *et al.*, 1994). Similarly, in the case of data tilting, the power measure of divergence (Cressie and Reid, 1984) can be used to construct a measure of the strength of the monotonic relationship.

Last, but not the least, the two biased bootstrap methods are quite popular in the literature on nonparametric and semiparametric curve estimation under order restrictions. This is not only because of their convenient application and easier interpretation, but also because recent research is facilitating better appreciation of their attractive theoretical properties; see, for example, Hall and Huang (2001) and Prewitt (2003). For these compelling reasons we choose to focus on these two biased bootstrap techniques.

[5] The density-regression method is based on a different approach, but it also achieves good smoothness properties.

In the following section (Section 4.4), we describe how biased bootstrap methods can be applied to unconstrained kernel and histogram sieve estimators to restore monotonicity in the estimated time varying coefficients.

4.5 Estimation procedures based on biased bootstrap techniques

We consider a age-varying covariate effect regression model $\lambda\,(t|x) = \lambda_0(t).\,\exp\,(\beta(t).x)$, where $\beta(t)$ is known to increase or decrease in t. The basis for this monotonicity assumption can either be tests of proportionality against monotone alternatives (Bhattacharjee, 2007a, our Chapter 3), or theoretical considerations, or prior knowledge. As discussed in Section 4.2, with adequately defined covariates, this model can be very general. Nonmonotonic covariate effects, discussed in Chapter 3, can also often be expressed in this form, in terms of auxilliary covariates. For example, if $\beta(t)$ increases in t over one range of the covariate space, say $x \leq x_0$, and decreases in t otherwise, we can write the regression model as

$$\begin{aligned}\lambda\,(t|x) &= \lambda_0(t).\,\exp\,(\beta_1(t).x_1 + \beta_2(t).x_2)\,, \\ x_1 &= x.I\,(x \leq x_0)\,, x_2 = -x.I\,(x > x_0)\,,\end{aligned}$$

where $\beta_1(t)$ and $\beta_2(t)$ both increase in t.

In this section, we discuss estimation of $\beta(t)$ under such models. We consider two biased bootstrap methods by which usual kernel regression or sieve estimators can be monotonised to obtain the required order-restricted estimators. In Section 4.3, we have discussed several other ways by which order restricted estimates of hazard regression models can be obtained – namely isotonic regression, projection on to constrained subspaces, taut strings and density regression approach. However, we choose biased bootstrap methods because of their ease of interpretation and implementation, as well as their attractive smoothness properties.

4.5.1 Data tilting

We begin with a suitable estimator of time varying coefficients at r distinct ordered lifetimes $t^{(1)} < t^{(2)} < \ldots < t^{(r)}$. Denote this estimator $\hat{\beta}$,

$$\hat{\beta}_{r\times 1} = \begin{bmatrix} \hat{\beta}\left(t^{(1)}; t_{n\times 1}, x_{n\times 1}, Y_{n\times k}, \delta_{n\times 1}, p_{n\times 1}\right) \\ \hat{\beta}\left(t^{(2)}; t_{n\times 1}, x_{n\times 1}, Y_{n\times k}, \delta_{n\times 1}, p_{n\times 1}\right) \\ \vdots \\ \hat{\beta}\left(t^{(r)}; t_{n\times 1}, x_{n\times 1}, Y_{n\times k}, \delta_{n\times 1}, p_{n\times 1}\right) \end{bmatrix},$$

where the observed (possibly censored) data are of the form $\left(t_i, x_i, y_{i(1\times k)}, \delta_i\right)$, $i = 1, 2, \ldots, n$, and $p_{n\times 1}$ ($p_i \geq 0, \sum p_i = 1$) represents the weights assigned to the n data points. Here, $x_{n\times 1}$ represents the covariate for which the age-varying effects are under study, $Y_{n\times k}$ denotes other covariates (whose effects are assumed to be age-constant, for simplicity), and $\hat{\beta}$ may be taken as one of the usual estimators of time varying coefficients, like the ones proposed by Zucker and Karr (1990), Murphy and Sen (1991) or Martinussen *et al.* (2002).

Following Hall and Huang (2001), and taking $p = p_{unif} = (1/n, 1/n, \ldots, 1/n)'$ as the base case, the objective of the data tilting methodology is to find $p = p^*$ that minimises a power measure of divergence (Cressie and Read, 1984) from p_{unif} among all p's for which the constraint is satisfied, *i.e.*, for which

$$\hat{\beta}\left(t^{(1)}; t, x, Y, \delta, p\right) \leq \hat{\beta}\left(t^{(2)}; t, x, Y, \delta, p\right) \leq \ldots \leq \hat{\beta}\left(t^{(r)}; t, x, Y, \delta, p\right).$$

The usual measure of divergence used is

$$\begin{aligned} D_\rho(p) &= \left\{n - \sum_{i=1}^{n}(np_i)^\rho\right\} / \left\{\rho(1-\rho)\right\}, \rho \neq 0, 1, \\ D_0(p) &= -\sum_{i=1}^{n}\log(np_i) \\ \text{and} \quad D_1(p) &= -\sum_{i=1}^{n} p_i \log(np_i). \end{aligned}$$

Then, the estimator is given by

$$\hat{\beta}_{r\times 1}^{DT} = \begin{bmatrix} \hat{\beta}\left(t^{(1)}; t_{n\times 1}, x_{n\times 1}, Y_{n\times k}, \delta_{n\times 1}, p^{*}_{n\times 1}\right) \\ \hat{\beta}\left(t^{(2)}; t_{n\times 1}, x_{n\times 1}, Y_{n\times k}, \delta_{n\times 1}, p^{*}_{n\times 1}\right) \\ \vdots \\ \hat{\beta}\left(t^{(r)}; t_{n\times 1}, x_{n\times 1}, Y_{n\times k}, \delta_{n\times 1}, p^{*}_{n\times 1}\right) \end{bmatrix}.$$

It is reasonably straightforward to abstract to an estimator over a continuous range on the lifetime axis, instead of the discrete set of points $t^{(1)}, t^{(2)}, \ldots, t^{(r)}$. In this case, one can have the constraint as

$$L\left(p; t, x, Y, \delta\right) = \int_0^T \hat{\beta}^T\left(s; t, x, Y, \delta, p\right).I\left(\hat{\beta}^T\left(s; t, x, Y, \delta, p\right) < 0\right) ds = 0,$$

where $I(.)$ is the indicator function.

Hall and Huang (2001) have discussed estimation of order restricted regression functions using data tilting when the regression function is monotonically increasing or decreasing. The extension of the procedure to the case of hazard regression models is conceptually similar. However, while this is theoretically an appealing estimation procedure, there are some issues regarding its implementation in the general form.

First, the likelihood function is complicated, and the influence function (measuring the influence of each observation on $L(p)$) is not available in closed form.[6] However, estimates of the influence of each observation can be estimated, either by row-deletion (jacknife) of each observation by turn, or by computing partial likelihood estimators for different weighting of the observations using the method proposed recently in Cai and Sun (2003).

Second, following estimation of the influence of each observation, a typical application of the data tilting procedure would involve convex optimisation (with linear constraints) in very high dimensions. This dimension increases with sample size, making the procedure computationally very demanding.

In order to proceed, we restrict attention to the class of estimators for which $p_j = i_j/n$ where $i_j \geq 0$ are integers. This reduces the problem to a discrete optimization problem, though

[6] For notational convenience, here and in the following, we denote L (a function of five parameters) as a function only of p. The other parameters of L are held constant throughout the estimation process.

the number of candidate p's, $2^n - 1$, is still very large. We now propose an algorithm, in two simpler steps, to obtain data tilting estimates within this class of weighting vectors.

In the first step, we fix $n_0 \geq 1$, initialize the weighting vector $\tilde{p}^{(0)} = p_{unif}$ and adopt the following iterative procedure. At iteration r, the procedure modifies the weighting vector from $\tilde{p}^{(r-1)}$ to $\tilde{p}^{(r)}$ by increasing the weight of the n_0 observations with the highest influence on $L(\tilde{p}^{(r-1)})$ by $1/n$ each, and correspondingly reduce the weight of the n_0 lowest influence observations. This iterative procedure is continued till we achieve $L(p) = 0$. Let the weighting vector at this stage be denoted p^*; the corresponding divergence is $D_\rho(p^*)$. This gives us one potential estimate.

In the second step, we enumerate $L(p)$ for all p's (within the class $p_j = i_j/n, i_j \geq 0$) for which $D_\rho(p) \leq D_\rho(p^*)$. We then estimate the final solution p^{**} as the one in this class for which $L(p) = 0$ and $D_\rho(p)$ is the minimum:

$$p^{**} = \arg\min_p \{D_\rho(p) : D_\rho(p) \leq D_\rho(p^*), L(p) = 0\}.$$

The search involved in this step is considerably less computation-intensive than what would be necessary if we were to optimise over all p's (instead of only over p's for which $D_\rho(p) \leq D_\rho(p^*)$, even after taking into account the computations in the first step of the algorithm. Thus, the division of the algorithm into these two steps reduces the computational complexity of the estimation procedure substantially. This procedure does not necessarily produce an unique solution; however, in our simulations, the effect of this on the final estimates was negligible. The steps of this procedure are summarized in Algorithm 4.4.1.

It must be mentioned here that this algorithm does not strictly give data tilting estimators, since we restrict to the set of p's for which $p_j = i_j/n, i_j \geq 0$. For a reasonable sample size, however, this is not likely to be an issue. An attractive feature of this algorithm is that the most computation-intensive sub-steps of the procedure (Step 2a and Step 3a) are amenable to parallel computation. The computation-intensity of the whole algorithm depends to a large extent on how large the search procedure in Step 3a is, which in turn depends critically on the choice of n_0. If n_0 is too large, the algorithm can rapidly reduce effective sample size, by reducing a sizeable number of the p_j's to nil, in which case the set P will also be very large. On the other hand, if n_0 is too small, a large number of iterations will be required in Step2 to get a feasible solution. In fact, the algorithm is particularly useful if sample size reduction is matched with fast convergence towards monotonicity. The choice of n_0 and the effectiveness of the algorithm in applications are important empirical issues which will be addressed in the next Section.

Algorithm 4.4.1

Computation of data tilting estimates

Step 1.	**Initialize**: Fix $\tilde{p}^{(0)} = p_{unif}$ and n_0 (the number of p_j's reduced at each iteration). Compute age-varying coefficients and $L(\tilde{p}^{(0)})$.
Step 2.	**Loop**: Do while $L(\tilde{p}^{(r)}) < 0$
a)	**Computation of influence functions** (for each observation for which $\tilde{p}_j^{(r-1)} > 0$): This can be done by actual row-deletion (jacknife) followed by estimation of age-varying regression coefficients.
b)	**Compute** $\tilde{p}^{(r)}$: Increase p_j by $1/n$ for the n_0 data points with highest influence on $L(\tilde{p}^{(r-1)})$, and correspondingly reduce p_j for the n_0 data points with lowest influence.
c)	**Compute** $L(\tilde{p}^{(r)})$ **and age-varying coefficients**.
	Endo: End of loop. Return p^* and $D^* = D_\rho(p^*)$.
Step 3.	**Find** p^{**}.
a)	Construct the set $P = \{p : D_\rho(p) \leq D^*, L(p) = 0\}$.
b)	Find $p \epsilon P$ for which $D_\rho(p)$ is minimum. Set $p^{**} = p$.
c)	Return final p^{**} and age-varyng regression coefficients.

4.5.2 Local adaptive bandwidth

Adaptive bandwidth selection has a long and established tradition in nonparametric regression; some recent contributions to this literature are Brockmann *et al.* (1993), Schucany (1995), Hermann *et al.* (1995) and Hermann (1997). In addition to the ability to adapt to the density of design points, and to the presence of heteroscedasticity, adaptive bandwidth regression estimators also have the advantage that they can adapt readily to the structure of the regression function, smoothing more in flat parts of the curve and less in peaky parts (Brockmann *et al.*, 1993). This feature suggests the use of adaptive bandwidth estimators for order restricted inference in regression models, including hazard regressions. If one were to smooth more in peaky parts rather than the flat ones, adaptive bandwidth would be useful in estimating regression functions under order restrictions in the nature of monotonicity of shape or slope parameters.

This method involves reweighting of the original data in a particular way, and in this sense, it falls within the general class of biased bootstrap methods (Hall and Turlach, 1999). The estimation procedure may be considered richer than the data tilting method in the sense that it offers the possibility of choosing different bandwidths at different age levels, instead of choosing a general overall reweighting of the whole data.

Adaptive bandwidth estimation is also similar in spirit to the way in which the location and scale view (SiZer maps) has been proposed as an attractive way for exploring structures in curves (scale is interpreted here as the "level of resolution" or "bandwidth") (Chaudhuri and Marron, 1999, 2000). However, while Chaudhuri and Marron (1999, 2000) focus on identifying features of a nonparametric curve that are relatively more robust to changes in bandwidth (in a sense, their focus is on testing), we propose to use adaptive bandwidths to perform kernel regression estimation subject to some maintained monotone structure.

In order to implement an adaptive bandwidth estimation algorithm, we require, for each lifetime t and local bandwidth $h(t)$, an estimator for the local kernel regression age-varying covariate effect $\hat{\beta}\left(t, h(t)\right)$ in the neighbourhood of t. This can be appropriately estimated by putting the weights from the kernel function on the corresponding term in the partial likelihood function, and then obtain partial likelihood estimates of the time varying coefficients (Cai and Sun, 2003).

The estimation procedure begins with choosing a global bandwidth $h_{(0)}$ (which provides an initial kernel estimator that is reasonably smooth), and several candidate bandwidths $h_1 < \ldots < h_{(0)} < \ldots < h_r$, both above and below $h_{(0)}$. The choice and range of these bandwidths depend on the particular context and application, and the degree of smoothness desired. For each of these candidate bandwidths, we estimate the kernel regression time varying coefficients $\hat{\beta}\left(t, h_i\right), i = 1, \ldots, r$ and $\hat{\beta}\left(t, h_{(0)}\right)$ using the methodology proposed in Cai and Sun (2003).

The objective of estimation is to achieve monotonicity with minimum deviation from the baseline bandwidth $h_{(0)}$. Hence, our adaptive bandwidth kernel estimator will be given by

$\hat{\beta}(t, h^*(t))$, where $h^*(.)$ minimises $\int_0^T ||h(t) - h_{(0)}||.dt$ within the class of $h(t)$'s satisfying

$$M(h(.); t, x, Y, \delta) = \int_0^T \hat{\beta}^T(s, h(.)) .I\left(\hat{\beta}^T(s, h(.)) < 0\right) ds = 0.$$

Note that, the adaptive bandwidth $h^*(t)$ varies with age t, and, at each t, is equal to one of the $r+1$ candidate bandwidths $h_{(0)}, h_1, \ldots, h_r$. If none of the candidate bandwidths gives $M(h(.)) = 0$, the choice of bandwidths has to be extended. This extension would usually be more towards the higher side.[7] A lower bandwidth will typically compromise the desirable smoothness properties of the estimator and, besides, monotonicity will always be achieved if the bandwidth is increased sufficiently.[8]

On the other hand, even when a feasible $h(.)$ has been identified, one may decide to extend candidate bandwidths over finer grids for one of the two following reasons. Either, if one finds multiple $h(.)$'s having $M(h(.)) = 0$ and the same divergence measure $\int_0^T ||h(t) - h_{(0)}||.dt$, so that finetuning the bandwidths around the potential candidate adaptive bandwidths is necessitated. Or, if one wishes to finetune the grids further to ensure that the estimated adaptive bandwidth does indeed minimise the divergence from $h_{(0)}$ within the class of adaptive bandwidths for which $M(h(.)) = 0$. The process of selecting candidate bandwidths and estimating the kernel regression time varying coefficients for these bandwidths is continued until a suitable adaptive bandwidth h^{**} is found. The steps of this proposed estimation procedure are outlined in Algorithm 4.4.2.

The algorithm is quite easy to implement and involves only moderate computational intensity. The choice of candidate bandwidths is a critical issue. Choice of $h_{(0)}$ will depend on the smoothness desired in any particular application, and the choice may be made using available methods for choosing an optimal bandwidth.[9] The other bandwidths may initially be chosen on the basis that they do not compromise the smoothness of the estimates too much (on the lower side), and include some bandwidths that give reasonably flat estimates (on the higher side), so that the initial admissible h^* can be identified quickly. In the refinement stage of the

[7] For notational convenience, we denote M as a function only of h, and supress the other four parameters. All parameters in M, other than h, are held constant throughout the estimation procedure.

[8] This is because a very large bandwidth will effectively reduce the age-varying kernel estimates to the usual (age-constant) partial likelihood estimate of the covariate effect, which is monotone by default.

[9] See Delaigle and Gijbels (2003) for an excellent practitioner-oriented review of the techniques in the context of density estimation with contaminated data; similar methods apply to kernel regression applications.

procedure, the aim of the choice of further bandwidths is to minimise the divergence from $h_{(0)}$, so the additional bandwidths chosen would typically be closer to $h_{(0)}$ than those in the current h^* at that stage.

Algorithm 4.4.2

Computation of local adaptive bandwidth estimates

Step 1.		**Fix** $h_{(0)}, h_1, \ldots, h_r$: Choose a bandwidth $h_{(0)}$ that gives age-varying coefficient estimates that are as smooth as desired, and several other candidate bandwidths $h_1, \ldots, h_r$ that also give reasonably smooth estimates. Set $h^*(.) = h_{(0)}$.
Step 2.		**Loop**: Do while $M\,(h^*(.)) < 0$
	a)	**Search for admissible adaptive bandwidth estimators**: Enumerate the set H of $h(.)$'s as $H = \left\{h(.) : h(t) \epsilon \left\{h_{(0)}, h_1, \ldots, h_r\right\}, M\,(h(.)) = 0\right\}$.
	b)	**Select** h^*: If H is not empty, set $h^*(.) = \arg\min_{h \epsilon H} \int_0^T \lvert\lvert h(t) - h_{(0)} \rvert\rvert .dt$.
	c)	**Expand choice of candidate bandwidths**: If H is empty, select a larger set of candidate bandwidths, particularly including higher bandwidths that would flatten out the kernel estimates of age-varying coefficients.
		Endo: End of loop. Return h^* and $\hat{\beta}\,(t, h^*(t))$.
Step 3.		**Refine estimates**: If h^* given in Step 2 is unique, consider other candidate bandwidths close to this and refine estimates. If there are multiple candidate h^*'s, choose other candidate bandwidths close to these, and resolve the tie. Return final h^{**} and age-varying regression coefficients.

The search for admissible adaptive bandwidths (Step 2a) is the most computation intensive step in the algorithm, and it is useful to keep the number of candidate bandwidths within bounds. The ways by which this can be achieved depend largely on the particular context of the application at hand. In our experience, it is often apparent that some of the bandwidths are not useful in achieving monotonicity, and these may then be omitted in favour of more useful bandwidths at that stage.

Since this method allows the choice of widely different bandwidths at different points on the duration scale, it can be less parsimonious than data tilting. Consequently, the method offers more choice and makes it easier to attain the desired monotonicity. Further, the adaptive bandwidth method is easier to implement, being less computation intensive than data tilting.

Further, we find adaptive bandwidth estimators easier to interpret than data tilting. Since a higher bandwidth flattens out the kernel estimates, the adaptive bandwidth method is expected to give higher bandwidths to points on the lifetime scale that are either peaky in terms of the time varying coefficients, or where the data are sparse. As mentioned earlier, this feature of the estimation procedure has the potential of being interpreted as a strength of the maintained monotone relationship, much in the same way as SiZer maps (Chaudhuri and Marron, 1999, 2000).

Also, standard confidence intervals are easier to construct, and provide useful inference about the strength of the maintained order restriction at different ages. These confidence intervals are, however, not proper confidence intervals of the adaptive bandwidth estimator since they are not adjusted for pretesting. Pretesting-adjusted confidence intervals can be constructed by resampling (bootstrap or jacknife) from the original sample – such computations are, however, quite intensive.

4.6 Applications and simulations

In this section, we explore empirical performance of the two proposed biased bootstrap estimation methods, based on a small simulation study and two real applications – one each from biomedicine and economic duration data. We also provide some practical guidance as to the choice of parameters during implementation of the algorithms.

4.6.1 Simulation study

In the simulation study, we use the histogram sieve estimator (Murphy and Sen, 1991) to benchmark the performance of the proposed estimators. It is worth noting that, while the application of biased bootstrap methods to order-restricted inference in the hazard regression context is new, several other papers have reported simulation studies in the linear regression context. Hall and Huang (2001) have examined the performance of data tilting estimators under monotonicity constraints, the performance of spline-based nonparametric regression estimators has been empirically evaluated in Lee (2003), and Lee and Solo (1999) have compared the empirical performance of bandwidth selection methods for local linear regression.

Randomly right censored data are generated from the following age-varying coefficient hazard regression model:

$$\lambda\left(t|X=x\right)=2.\exp\left(tx\right),$$

where X are generated from $Uniform[1,2]$, and the censoring random variable C has distribution function $F(c)=(c-0.005)^3, c\epsilon[0.005,1.005]$. 100 random samples of 500 observations each were generated from this data generating process and Murphy-Sen histogram sieve estimation and the two biased bootstrap techniques described in Sections 4.3 and 4.4 were applied to each. The estimates of age-varying coefficients based on the three estimators were evaluated at 10 equidistant lifetimes 0.06 through 0.60 with increments of 0.06.

For the implementation of Algorithm 4.4.1 (data tilting), we set n_0 to 10 and ρ to unity. The choice of ρ has been discussed by previous authors (Hall and Presnell, 1999; Hall and Huang, 2001), and we do not have anything new to add to the discussion in the present context. As discussed in Section 4.4, an effective choice of n_0 is necessary to strike a good balance between sample size reduction (and therefore enlargement of the set P) and fast convergence towards monotonicity. We experimented with several values of n_0 and decided on 10 based on our experience (for a sample size of 100, the choice of $n_0=5$ appeared to work well). With $n_0=10$, it took an average of 21 iterations to achieve monotonicity. The effective sample size reduced to 432 on average; on average, 378 observations had frequency 1, 41 had frequency 2 and 10 had frequency 3. Hence, construction of the set P was quite computation intensive; however, optimising for the lowest divergence $D_\rho(p)$ (Step 3b) did not change the age-varying coefficients substantially. For many empirical applications, therefore, we feel that one may terminate the algorithm at Step 2.

For our implementation of Algorithm 4.4.2, we used an Epanechnikov kernel, and the Cai and Sun (2003) method was used to estimate the time varying coefficients. Our choice of initial bandwidths were $h_{(0)} = 0.11$, and 0.07, 0.09, 0.13, 0.15, 0.19 and 0.25. This, in our opinion, constituted a good mix of candidate bandwidths that gave smooth coefficient estimates over lifetime, and made the time-variation quite flat at the upper end.

The performance of the data tilting method was the worst of the three methods under study. As mentioned earlier, the effective sample size was reduced to about 430 on average, over the 100 samples. The final estimates were poor, particularly towards the boundaries of the sample space. For $t = 0.06$, $t = 0.30$ and $t = 0.60$, for example, the average estimates were -0.675 (quartiles -1.08, -0.65 and -0.21), 0.253 (quartiles 0.11, 0.22 and 0.32) and 4.283 (quartiles 2.89, 4.22 and 5.55) respectively, as compared to parameter values of $\beta(0.06) = 0.06$, $\beta(0.30) = 0.30$ and $\beta(0.60) = 0.60$ respectively under the model. The problem with this implementation of the data tilting method appeared to be that the algorithm systematically reduced the weights on observations having high influence towards the boundaries, with the result that estimates in these neighbourhoods were pushed too far out of sync.

The adaptive bandwidth estimator performed the best of the three, the average estimates for $t = 0.06$, $t = 0.30$ and $t = 0.60$ being -0.021 (quartiles -0.11, 0.04 and 0.16), 0.292 (quartiles 0.18, 0.30 and 0.39) and 0.631 (quartiles 0.41, 0.62 and 0.83) respectively. The Murphy-Sen estimator by contrast had average estimates for $t = 0.06$, $t = 0.30$ and $t = 0.60$ of 0.057 (quartiles -0.16, 0.05 and 0.32), 0.195 (quartiles -0.17, 0.20 and 0.55) and 0.675 (quartiles 0.22, 0.65 and 1.31) respectively.

On the basis of the simulations, the adaptive bandwidth estimator was the most efficient of the three estimators considered, with the estimates for the different samples tightly clustered together, as seen from the box plots in Figure 4-1 (adaptive bandwidth) and Figure 4-2 (Murphy-Sen histogram sieve estimator). The average absolute deviation of the estimates from actual values for these 10 points was, on average over the 100 samples, 0.150 for adaptive bandwidth and 0.468 for Murphy-Sen, while this measure was as high as 0.805 for our implementation of the the data tilting estimator.

Therefore, on the basis of our simulation study, the adaptive bandwidth estimator appears to work better in terms of empirical performance, and we concentrate on this estimator in the following two applications.

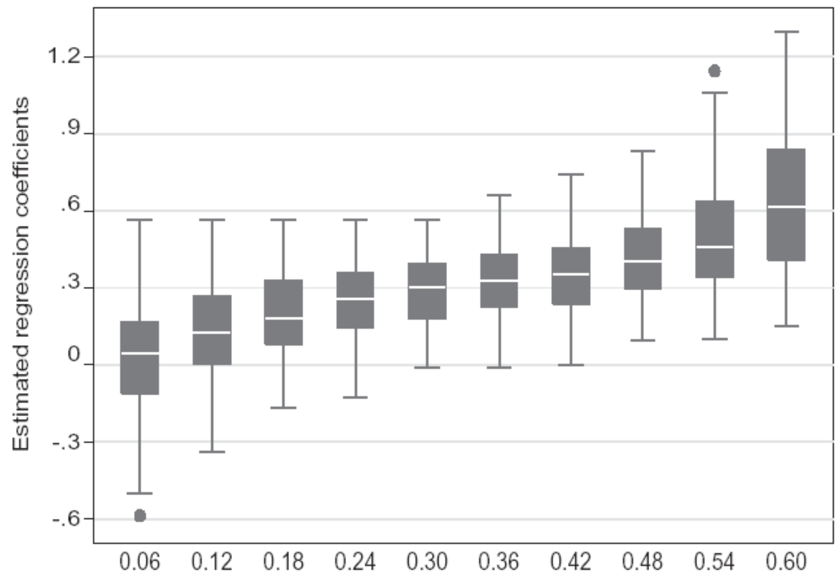


Figure 4-1: Box plot of adaptive bandwidth estimates (duration $0.06, 0.12, \ldots, 0.60$; 100 samples; sample size 500).

4.6.2 Example: Malignant melanoma data

These data pertain to 205 patients (148 of these are censored) with malignant melanoma (cancer of the skin) on whom a radical operation was performed at the Department of Plastic Surgery, University Hospital of Odense, Denmark. Andersen *et al.* (1992) have reproduced the data and elaborately analysed it, and have discussed the findings of several other researchers who have worked on these data. One of the strongest prognostic factors in malignant melanoma identified in the literature is tumor thickness. As discussed in Chapters 1 and 3 (Sections 1.1.2 and 3.6.3 respectively), Andersen *et al.* (1992) find possible violation of the PH model in these data, particularly in favour of alternatives like $DHRCC$. Further, the plot of the cumulative regression functions for log-thickness (Martinussen *et al.*, 2002) also indicate a distinct concave shape, though the constant coefficient estimate lies almost entirely within the 95 percent confidence band of the cumulative regression function.

Bhattacharjee (2007a, our Chapter 3) showed that the null hypothesis of proportional hazard was rejected in these data, in favour of the alternative $DHRCC$ over the upper range of the covariate space, while for patients with small tumors, there was some evidence of an $IHRCC$

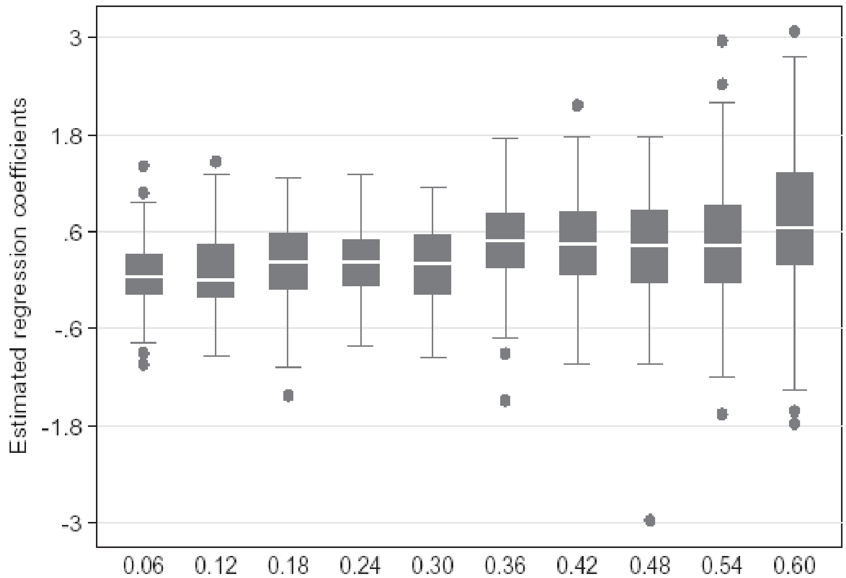


Figure 4-2: Box plot of Murphy-Sen estimates (duration $0.06, 0.12, \ldots, 0.60$; 100 samples; sample size 500).

pattern (this was also confirmed by the Murphy-Sen histogram sieve estimators). Figures 4-3 and 4-4 show kernel estimators of the time varying coefficients for various bandwidths, for patients with tumor thickness less than, and greater than 1.8 mm respectively. One can see that the monotonicity evident from the tests emerge prominently in these plots, and that constrained estimation using adaptive bandwidth selection can be used to obtain estimates of order-restricted covariate effects for tumor thickness.

4.6.3 Example: Macroeconomic instability and business failure

We analyse data from Bhattacharjee *et al.* (2008a) on firm exits in the UK over the period 1965 to 1998; see Sections 1.1.3 and 1.3.7.1 for previous discussions of the data and the application. A major focus of the analysis is on the effect of macroeconomic instability on business failure. Two measures of macoeconomic instability are considered: turnaround in business cycle (a measure of the curvature of the Hodrick-Prescott filter of output per capita) and volatility in exchange rates (maximum monthly change in exchange rates over a year). Theory suggests that the effect of the first measure on bankruptcy may be negative, and the second one positive.

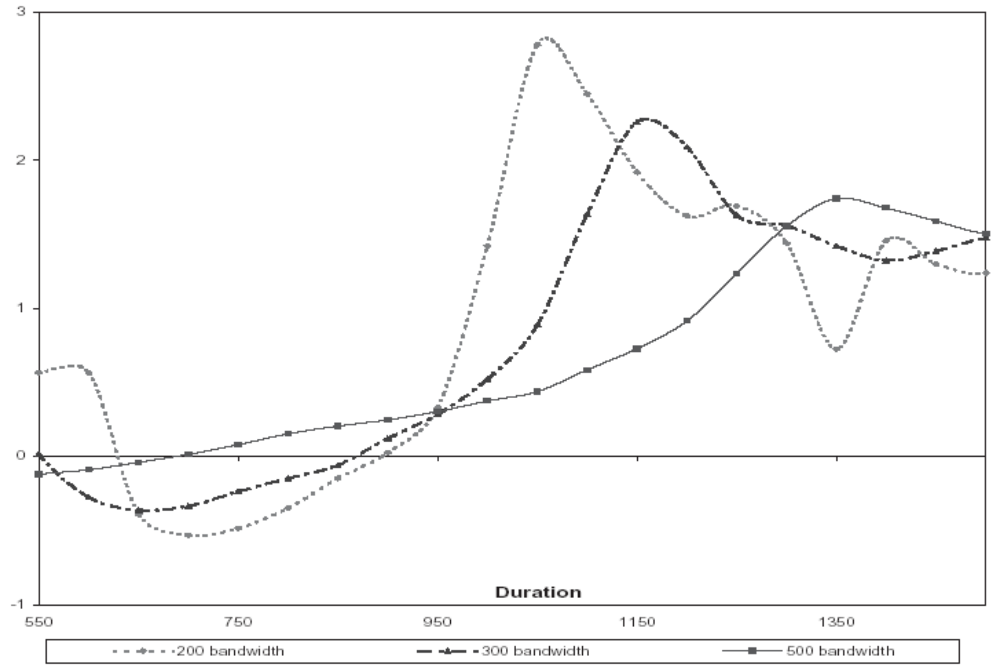


Figure 4-3: Age varying covariate effects: $\ln(\text{Thickness}) \times 1\,(\text{Thickness} \leq 1.8)$.

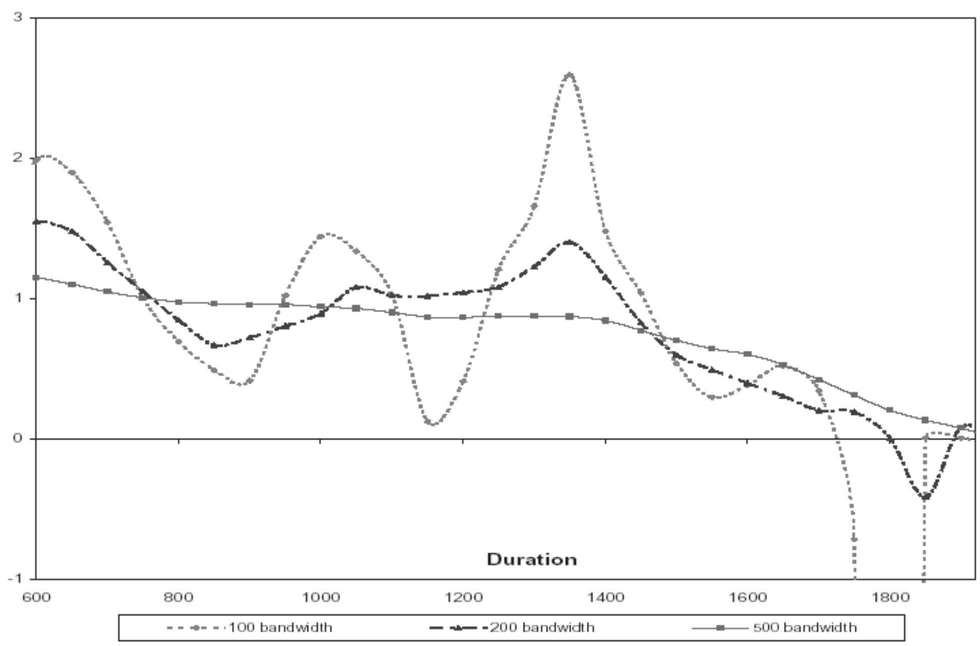


Figure 4-4: Age varying covariate effects: $\ln(\text{Thickness}) \times 1\,(\text{Thickness} > 1.8)$.

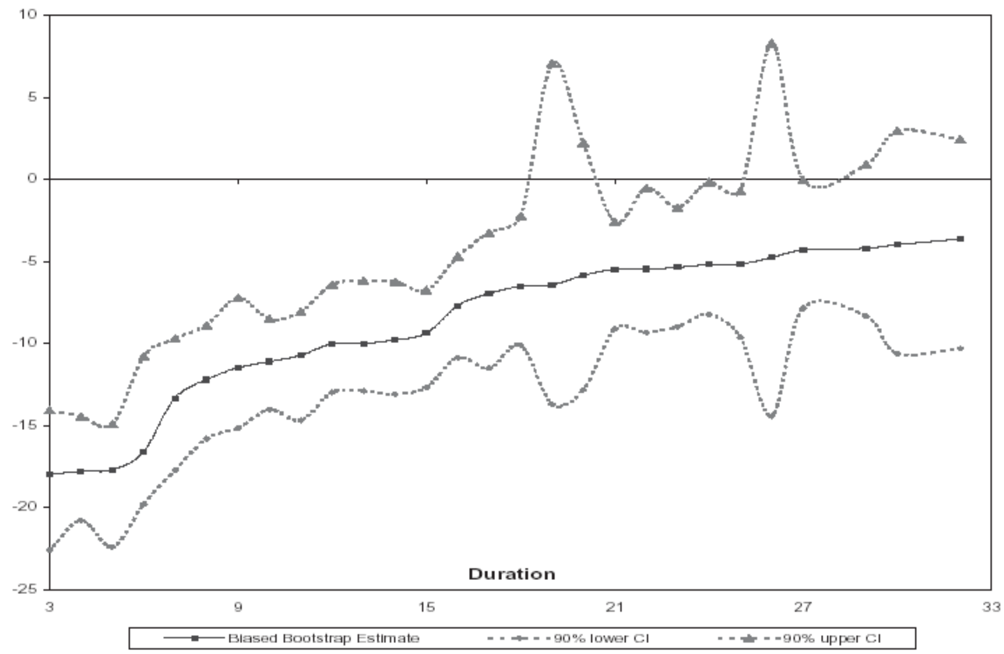


Figure 4-5: Age varying covariate effects: turnaround in business cycle.

Because of learning effects, the adverse impact of instability is expected to decline in the age of the firm, post-listing.

The tests of proportional hazards against monotone departures proposed in Bhattacharjee (2007a) indicate monotone departures in both cases, and this is also confirmed by the Murphy-Sen estimates (see Chapter 7), after conditioning on industry dummies and firm level factors like size, profitability and cash flow.

The kernel estimates of time varying coefficients for several candidate bandwidths confirm that the detrimental effect of uncertainty diminishes with the age of the firm, post-listing. The adaptive bandwidth estimators along with 90 per cent confidence bands (not adjusted for pre-testing) confirm these findings (Figures 4-5 and 4-6), and provide useable and meaningful estimates of the prognostic impact of instability on corporate failure. The confidence bands also provide useful inference about the strength of the monotonicity relationship, in that they depend closely on the magnitude of the bandwidth given by the estimator, which in turn depends on the peakedness feature of the kernel estimates at different durations and on the density of data

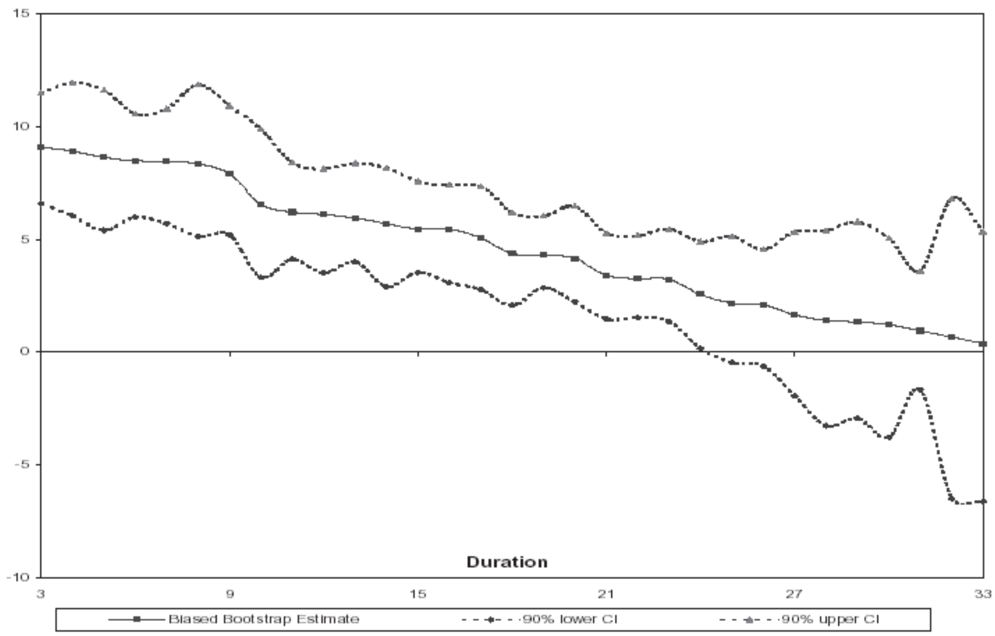


Figure 4-6: Age varying covariate effects: volatility in exchange rate.

around these durations.

In summary, the adaptive bandwidth estimators appear to be a convenient way to estimate hazard regression models under monotone departures from proportionality. Their empirical performance is good, and they provide useful inference in applications. By contrast, data tilting methods are comparatively more difficult to implement, and their performance in the simulation study was poor.

4.7 Concluding remarks

In this chapter, we discussed estimation in hazard regression models under order restrictions, where the time varying coefficients are known to be monotonically increasing or decreasing. Such situations occur frequently in applications, and encompass a wide range of data generating processes. We consider estimation using two biased bootstrap methods, one based on data tilting and the other on local adaptive bandwidths.

The adaptive bandwidth estimator performed much better than the histogram sieve estimator and the data tilting estimator in simulations, and was useful in applications. In combination with research reported earlier in Chapter 3 (Bhattacharjee, 2007a), on testing proportionality against monotone alternatives in hazard regression models, these inference techniques provide a new and useful way to analyse covariate dependence in hazard regression models when the PH assumption does not hold.

Several lines of potential further research emerge from our work. First, while our focus here was on biased bootstrap methods, estimation under constraints using other methods (particularly taut string and density regression approach) may be useful. Second, while we point out the usefulness of the proposed methods in detecting departures from monotonicity, more work needs to be done on formal testing for order restrictions using these approaches. Third, many applications imply order restrictions on ageing in addition to those on covariate effects. In Chapter 6 (Bhattacharjee and Bhattacharjee, 2007), we develop Bayesian methods for analysis in these situations; frequentist inference may be developed in future work. Fourth, our discussion of the time varying coefficients model highlighted additional assumptions relating to additivity and proportional variation in time varying coefficients. Development of formal tests for the time varying coefficients model in these respects will be an useful research direction. Finally, it is well acknowledged that monotone covariate effects may often be confounded with frailty. Inference on frailty models to address this issue will be reported in Chapter 5 (Bhattacharjee, 2007b), Chapter 6 (Bhattacharjee and Bhattacharjee, 2007) and Section 7.4 (Bhattacharjee, 2007c).

Chapter 5

Testing for Proportional Hazards with Unrestricted Univariate Frailty

5.1 Chapter summary

Based on Bhattacharjee (2007b), here we develop tests of the proportional hazards assumption, with respect to a continuous covariate, in the presence of individual level frailty with unknown distribution. Unlike the case where the frailty distribution is known upto finite dimensional parameters (Chapter 3, Bhattacharjee, 2007a), 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 first develop tests for absence of covariate dependence, particularly against trending alternatives, by extending two-sample tests for equality of hazard rates. Next, we adapt the above methods to testing for proportional hazards by making suitable choice of weight functions. The proposed tests are particularly useful for detecting trend in the underlying conditional hazard rates, and for testing proportionality against ordered alternatives, respectively. Asymptotic distribution of the test statistics are established, followed by a Monte Carlo study. An application to the effect of aggregate Q on corporate failure in the UK shows evidence of trend in the covariate effect, and violation of proportional hazards assumption, whereas a traditional score test under the Cox regression model failed to detect evidence of any covariate effect.

5.2 Introduction

In Chapter 3 (Bhattacharjee, 2007a), we 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 frailties belongs to a known finite dimensional family. However, the above inferential approach is not applicable when there is individual level frailty with arbitrary distribution. Our contribution here is to develop tests for proportional hazards in the presence of individual level unobserved heterogeneity with completely unrestricted and unknown frailty distribution. Allowing for an arbitrary frailty distribution is particularly important in the hazard regression context, since the frailty distribution assumptions are very important. Both simulations (Bretagnolle and Huber-Carol, 1988; Baker and Melino, 2000) and empirical applications (Heckman and Singer, 1984b; Trussell and Richards, 1985; Hougaard *et al.*, 1994; Keiding *et al.*, 1997) show that inference is sensitive to the choice of the frailty distribution.

The chapter is organised as follows. In Section 5.2, we formulate the proposed test for proportional hazards under the mixed proportional hazards (MPH) model incorporating unrestricted univariate frailty. Identifying conditions of the MPH 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 5.3, 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 5.4, followed by a real life application in Section 5.5. Finally, Section 5.6 concludes.

5.3 Formulation of the testing problems

5.3.1 Testing proportional hazards

We first consider the standard mixed proportional hazards (MPH) model (introduced in Section 1.2.2 and discussed elaborately in Section 1.2.6)

$$\begin{aligned} \lambda\left(t|X=x,Z=z,U=u\right) &= \lambda_0\left(t\right)\exp\left[\beta_X.x+\beta_Z^T.z+u\right] \\ &\iff \ln\Lambda_0(T) = -\left(\beta_X.x+\beta_Z^T.z+U+\varepsilon\right), \end{aligned} \tag{5.1}$$

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-frailty 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), Härdle and Stoker (1989), Sherman (1993), Cheng *et al.* (1995) 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 MPH model with time varying coefficients

$$\lambda\left(t|X=x,Z=z,U=u\right)=\lambda_0\left(t\right)\exp\left[\beta_X\left(t\right).x+\beta_Z(t)^T.z+u\right], \tag{5.2}$$

with covariates X and Z, which are both allowed to have potentially time varying effects ($\beta_X\left(t\right)$ and $\beta_Z(t)$).[1] Under this model, the null hypothesis of proportional hazards corresponds to covariate effects constant over lifetime

$$\mathbb{H}_{0,PH}:\beta_X(t)\equiv b, \tag{5.3}$$

[1] While we assume fixed covariates for the sake of 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 lifetime scale, and restrict the time varying coefficient corresponding to every time interval to be zero except on the specific interval considered.

and the ordered alternative of monotone covariate effects

$$\frac{\lambda\left(t|X=x_2,Z=z,U=u\right)}{\lambda\left(t|X=x_1,Z=z,U=u\right)}\uparrow t \quad \text{whenever} \quad x_2>x_1, \quad \text{for all} \quad z,u, \tag{5.4}$$

corresponds to increasing time varying coefficients

$$\mathbb{H}_{1,PH}:\beta_X(t)\uparrow t. \tag{5.5}$$

While, we assume the above MPH model with time varying coefficients (5.2) 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 frailty model that can be considered in this problem.[2]

As shown by McCall (1996), sufficient conditions for identifiability of the MPH model with individual level frailty and time varying coefficients (5.2) is the inclusion of a covariate with proportional hazards that has support over the whole real line. 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 frailty distribution modeled as a sequence of discrete multinomial mixtures with increasing number of support points (Heckman and Singer, 1984a).

The alternative hypothesis (5.4, 5.5) is the $IHRCC$ condition introduced in Definition 3.2.1 and developed in Chapters 3 and 4. This suggests that tests similar to those developed in Chapter 3 may be useful here. However, the formulation of our testing problem has to be modified to reflect the identifying restrictions of the transformation model (5.1, 5.2).[3] Specifically, since the MPH model still continues to hold if a constant is added to both sides, a location

[2] The main assumption underlying this model is that of multiplicative separability of the effect of X, Z and U on the conditional hazard rate; see also discussion in Section 4.2.

[3] Strictly speaking, the MPH model with time varying coefficients (5.2) is not a linear transformation model. However, it can be cast as a transformation model, if one makes the (histogram sieve) assumption that the time varying coefficients are piecewise constant, changing values across known or hypothesized intervals. The width of these intervals will be allowed to decrease to zero with sample size.

normalisation is required for identification. This can be achieved by setting

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

for some fixed and finite $t_0 > 0$.[4] 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 here takes the form

$$\Lambda_0(t_0|X = x) = \int_0^t \lambda_0(s). \exp\left[\beta_X(s).x\right].ds \equiv 1, \tag{5.6}$$

conditional on every covariate value $X = x$

Because of the above scale normalisation, the baseline cumulative hazard function in (5.6) is only identified upto a factor of proportionality, restricting it to take the value unity at a fixed failure time 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 the same. 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 failure times. 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

$$\begin{aligned} \mathbb{H}_{0,PH} \quad &: \quad \Lambda_0(t|X = x) = \Lambda_0(t) \quad \text{for all } \; x \\ &\Longleftrightarrow \quad \lambda_0(t|X = x_1) = \lambda_0(t|X = x_2) \quad \text{for all } \; x_1 \neq x_2, \end{aligned} \tag{5.7}$$

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

[4] Note that, the MPH model has an important distinction from the standard transformation model, in that the usual scale normalisation is not necessary here. In other words, 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.

for equality of hazard functions to the continuous covariate setup. The relevant alternative hypothesis, discussed in Section 5.3, will determine the appropriate choice of the underlying two-sample test statistics.

5.3.2 Testing absence of covariate dependence

We now turn to a related testing problem suggested by the modified null hypothesis (5.7). Since, in the formulation above, the null hypothesis of proportional hazards is that of 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 itelf is an important inference problem, particularly since understanding the nature of covariate dependence is one of the main objectives of regression analysis of failure time data.

We consider the general hazard regression model

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

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 Section 4.2, by suitable transformations and use of the histogram sieve, the effect of the other covariates Z can be approximated by time varying effects:

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

which is a convenient form for regression modeling.

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

$$\begin{aligned} \mathbb{H}_{0,Eq} \quad &: \quad \lambda_0(t|X=x)=c(t) \qquad \text{for all } x \\ &\iff \quad \lambda_0(t|X=x_1)=\lambda_0(t|X=x_2) \quad \text{for all} \;\; x_1\neq x_2 \end{aligned} \tag{5.8}$$

against relevant alternatives. The similarity between the above null hypothesis (5.8) and that for testing proportional hazards (5.7) 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

$$\mathbb{H}_{1,Eq}: \text{ not } \mathbb{H}_{0,Eq}, \tag{5.9}$$

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 5.3.

Finally, note that we have not considered unrestricted frailty in our regression model specification for the test for absence of covariate dependence. In fact, an important implication of the location normalisation (5.6) inherent in the corresponding MPH model with unrestricted frailty distribution is that, absense of covariate dependence cannot be tested in this model. This is because equality of the conditional hazard rates is also outcome of proportional hazards. However, models with either shared frailty or with finite dimensional frailty distributions are accommodated easily within our framework. Further, the case of unrestricted frailty 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.

5.3.3 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 and hazard functions. For this purpose, we consider estimators for the cumulative baseline hazard $\widehat{\Lambda}_0\left(t|x_1\right), \widehat{\Lambda}_0\left(t|x_2\right), \ldots$, conditional on different covariate values $X = x_1, x_2, \ldots$, in models including additional covariates, Z, and possibly unrestricted univariate frailty. Various candidate estimators are avaiable in the literature.

For the hazard regression model with time varying coefficients but without frailty, 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 lifetime data with shared frailties, one can either use the marginal modeling approach with unrestricted frailty 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 frailty can be assumed to belong to a given one-parameter family of continuous distributions.

For the tests of proportional hazards, we focus lies in the unrestricted univariate frailty case. Contributions in this area, reviewed earlier in Section 1.2.6, 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, based on previous work on the transformation model (Breiman and Friedman, 1985; Horowitz, 1996), can be made to converge at a rate arbitrarily close to the optimal $n^{-2/5}$ by suitable choice of bandwidths. 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 is yet to be studied.

With discrete lifetime data (discussed earlier in Section 1.2.7) over a finite dimensional sample space, estimation of the baseline hazard function reduces to a simpler problem. Further, if one approximates the unknown frailty distribution by a sequence of discrete mixtures of degenerate distributions (Heckman and Singer, 1984a), estimation of the frailty 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 frailty is therefore an attractive strategy.

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 frailty distribution as nuisance parameters.

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 an appropriate estimator has been chosen. In practise, an appropriate choice will have to be made based on both the assumed underlying model and properties of the estimator itself.

5.4 Proposed tests

In this Section, we discuss test procedures for the two testing problems. We first describe the alternative hypotheses, followed by the test for absence of covariate effect, and then the test for proportionality. Like the two sample tests on which they are based, a class of tests are proposed in either case, where the user can specify a relevant weight function; see also Chapters 2 and 3. We establish the statistical properties of the tests, and discuss the choice of weight functions.

5.4.1 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 of the kind considered in Chapters 3 and 4. Specifically, we consider alternatives defined by nonproportional partial orders, specifically $IHRCC$ or $DHRCC$ (Definition 3.2.1):

$$IHRCC \quad : \quad \text{whenever } x_1 > x_2, \lambda(t|x_1)/\lambda(t|x_2) \uparrow t \left[\equiv (T|X = x_1) \underset{c}{\prec} (T|X = x_2)\right], \tag{5.10}$$

$$DHRCC \quad : \quad \text{whenever } x_1 > x_2, \lambda(t|x_2)/\lambda(t|x_1) \uparrow t \left[\equiv (T|X = x_2) \underset{c}{\prec} (T|X = x_1)\right], \tag{5.11}$$

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

Let us initially consider two distinct covariate values, x_1 and x_2. As in Chapter 3, our strategy will be 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 (see discussion in Section 4.2) 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 above null hypothesis 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.

As discussed in Chapter 4 (Section 4.2) and Section 4 above, under the time varying coefficients 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, under this model with individual level frailty, the PH assumption is represented by a null hypothesis of equal conditional hazards, and the alternative posits monotone covariate effect with crossing hazards character. Following the approach in Chapter 3 (Bhattacharjee, 2007a), we will 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

$$T_{2s} = \int_0^{\tau} L(t) d\widehat{\Lambda}_1(t) - \int_0^{\tau} L(t) d\widehat{\Lambda}_2(t), \tag{5.12}$$

where $L(.)$ is some appropriate weight function and τ is a large failure time, 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. The Mantel-Haenszel or logrank test (Mantel, 1966; 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 (Gehan, 1965; Breslow, 1970) and Prentice (1978) tests generalise the Wilcoxon and Kruskal-Wallis tests to right censored data. Tarone and Ware (1977) and Harrington and Fleming (1982) have proposed weighted log-rank tests. The theoretical properties of these tests and their use in applications has been discussed elsewhere (Fleming and Harrington, 1991; Andersen *et al.*, 1993). Later in the chapter, we will 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 (5.8) against the omnibus alternative (5.9). However, the omnibus alternative in the above tests 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 significant covariate dependence, but also about the direction of the covariate effect, *i.e.*, whether an increase in covariate value is expected to increase or decrease the lifetime, according to some notion of relative ageing. In the k-sample setup, several trend tests have been proposed; these procedures test for equality of hazards against the alternatives $H_1 : \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_k$ or $H_1 : S_1 \leq S_2 \leq \ldots \leq S_k$ (one or more of the inequalities being strict), where λ_j and S_j are the hazard and survival functions respectively in the j-th sample.

Modified score tests against 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 failure time data by applying isotonic regression to scores from existing k-sample tests. These two-sample and k-sample tests are, however, of limited use in applications. The usual method of extending these inference procedures to the case of continuous covariates involves stratification with respect to the covariate, followed by application of existing inference procedures for k samples. The outcomes of these inference procedures are highly sensitive to the choice of such intervals, and relevant procedures for optimally choosing these intervals are not available in general (Horowitz and Neumann, 1992; Neumann, 1997).

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

$$\mathbb{H}_{1,Eq}^{(t)} : \lambda(t|x_1, z) \geq \lambda(t|x_2, z) \qquad \text{for all } z \text{ and } t \qquad \text{whenever } x_1 > x_2 \text{ (or its dual)}, \tag{5.13}$$

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:

$$\begin{aligned} \mathbb{H}_{1,Eq}^{(c)} \quad &: \quad \text{there exists } x^* \text{ such that } \lambda(t|x) \uparrow x \qquad \text{for all } z \text{ and } t \\ \text{whenever } x \quad &< \quad x^*, \text{ and } \lambda(t|x) \downarrow x \text{ whenever } x > x^* \text{ (or its dual)}. \end{aligned} \tag{5.14}$$

Some trend tests in the literature are specific to continuous covariates and consider (5.13) 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). Other tests assume a known covariate label function. Brown *et al.* (1974) developed a permutation test based on ranking of both the covariate values and the observed lifetimes, and O'Brien (1978) proposed inverse normal and logit rank tests using the respective transformations of the ranked covariates. Jones and Crowley (1989, 1990) consider a more 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 (5.14). Jespersen (1986) has proposed inference procedures in the context of a single changepoint regression model; however, the assumptions of a specified regression model and a single changepoint are quite restrictive. Thus, appropriate tests for absence of covariate dependence for continuous covariates are not available in the literature, in applications where neither the form of the regression relationship nor an appropriate covariate label function are known *a priori*. In many applications, insignificance of the estimated parameter in a Cox regression model is interpreted as a test for covariate dependence. Such an implication is inappropriate, since lack of significance can be due to other reasons, like violation of proportionality or model misspecification[5].

5.4.2 Testing absence of covariate dependence

First, we consider the single covariate case. Let T be a lifetime 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 (5.8) 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 $\mathbb{H}^{(t)}_{1,Eq}$ (5.13), and changepoint trend departures like $\mathbb{H}^{(c)}_{1,Eq}$ (5.14).

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:

$$T_{2s,std} = \frac{T_{2s}}{\sqrt{\widehat{\mathrm{Var}}\left[T_{2s}\right]}}, \tag{5.15}$$

where

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

[5] A large simulation study by Li *et al.* (1996) highlights the serious consequences of these issues in the context of the Cox PH model.

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,

$$K(t) = I\left[Y_1(t) + Y_2(t) > 0\right], \tag{5.16}$$

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

$$K(t) = I\left[Y_1(t) + Y_2(t) > 0\right].\{Y_1(t) + Y_2(t)\}. \tag{5.17}$$

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

$$\begin{aligned}
T_{2s,std}(x_{l1}, x_{l2}) &= \frac{T_{2s}(x_{l1}, x_{l2})}{\sqrt{\widehat{\text{Var}}\left[T_{2s}(x_{l1}, x_{l2})\right]}}, \\
T_{2s}(x_{l1}, x_{l2}) &= \int_0^\tau L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l1}) - \int_0^\tau L(x_{l1}, x_{l2})(t) d\widehat{\Lambda}(t, x_{l2}), \qquad (5.18) \\
\widehat{\text{Var}}\left[T_{2s}(x_{l1}, x_{l2})\right] &= \int_0^\tau L^2(x_{l1}, x_{l2})(t)\{Y(t, x_{l1})Y(t, x_{l2})\}^{-1}.d\left(N(t, x_{l1}) + N(t, x_{l2})\right),
\end{aligned}$$

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:

$$T_{2s}^{(\max)} = \max\left\{T_{2s,std}(x_{11}, x_{12}), T_{2s,std}(x_{21}, x_{22}), \ldots, T_{2s,std}(x_{r1}, x_{r2})\right\}, \qquad (5.19)$$

$$T_{2s}^{(\min)} = \min\left\{T_{2s,std}(x_{11}, x_{12}), T_{2s,std}(x_{21}, x_{22}), \ldots, T_{2s,std}(x_{r1}, x_{r2})\right\}, \qquad (5.20)$$

$$\text{and } \overline{T}_{2s} = \frac{1}{r}\sum_{l=1}^{r} T_{2s,std}(x_{l1}, x_{l2}). \qquad (5.21)$$

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 5.3.1 *For each l, $l = 1, 2, \ldots, r$, let $L(x_{l1}, x_{l2})(t)$ be a predictable processes indexed on the pair of fixed covariate values (x_{l1}, x_{l2}).*

Assumption 5.3.2 *Let τ be a random stopping time. In particular, τ may be taken as the time at the final observation of the counting process $\Sigma_{l=1}^r\Sigma_{j=1}^2 N(t, x_{lj})$. In principle, one could also have different stopping times $\tau(x_{l1}, x_{l2}), l = 1, \ldots, r$ for each of the r basic test statistics $T_{2s,std}(x_{l1}, x_{l2}), l = 1, \ldots, r$.*

Assumption 5.3.3 *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 5.3.4 *There exists a sequence $a^{(n)}$, $a^{(n)} \longrightarrow \infty$ as $n \longrightarrow \infty$, and fixed functions $y(t, x)$, $l_1(x_{l1}, x_{l2})(t)$ and $l_2(x_{l1}, x_{l2})(t)$, $l = 1, \ldots, r$ such that*

$$\sup_{t \epsilon [0,\tau]} \left| Y(t,x)/a^{(n)} - y(t,x) \right| \xrightarrow{P} 0 \qquad \text{as } n \to \infty, \quad \forall x \epsilon \mathcal{X}$$
$$\sup_{t \epsilon [0,\tau]} \left| L(x_{l1}, x_{l2})(t) - l(x_{l1}, x_{l2})(t) \right| \xrightarrow{P} 0 \quad \text{as } n \to \infty, \quad l = 1, \ldots, r$$

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 $x \epsilon X$.*

Assumptions 5.3.1 through 5.3.4 constitute a simple extension, to the continuous covariate framework, of the standard set of assumptions for the counting process formulation of lifetime data (see, for example, Andersen *et al.*, 1993). As discussed in Chapter 2 (Sengupta *et al.*, 1998), the condition on probability limit of $Y(t, x)$ in Assumption 5.3.4 can be replaced by a set of weaker conditions. All the assumptions are satisfied in the random censorship model with continuous failure times, for any choice from the predictable weight functions discussed earlier.

Let the test statistics $T_{2s}^{(\max)}, T_{2s}^{(\min)}$ and $\overline{T}_{2s}$ be as defined earlier (5.19 – 5.21).

Theorem 5.3.1. *Let Assumptions 5.3.1 through 5.3.4 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.)

Corollary 5.3.1.

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

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

(Proof in Appendix).

Corollary 5.3.2. *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*

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

Let Assumptions 5.3.1 through 5.3.4 hold. Then, under $\mathbb{H}_{0,Eq}$*, as* $n \to \infty$*,*
(a) $P\left[T^{(\max)}_{2s,\underline{w}} \leq z^*\right] \to \prod_{l=1}^{r} [\Phi(z^*/w_l)]$*,*
(b) $P\left[T^{(\min)}_{2s,\underline{w}} \geq -z^*\right] \to \prod_{l=1}^{r} [\Phi(z^*/w_l)]$*, 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).

The above results establish the asymptotic properties of the proposed tests. Some other features of the testing procedure (similar to Chapter 3) merit further discussion. First, the number of covariate pairs, r, on which the statistics ($T^{(\max)}_{2s}$, $T^{(\min)}_{2s}$ 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 $+\infty$ $(-\infty)$. Second, Corollary 5.3.1 provides a simple way to compute p-values for the test statistics when r is reasonably large.[6] Third, Corollary 5.3.2 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

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

to a covariate pair where the covariates are further apart. In practice, this is expected to improve the empirical performance of the tests.

Fourth, since the covariate under consideration is continuous, it may not be feasible to construct the basic tests $T_{2s,std}$ based exactly on two distinct fixed points on the covariate space. 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 Chapter 3, Bhattacharjee, 2007a). This issue can be resolved by using a jacknife estimator for the variance of the average estimator.

Fifth, following arguments in Gill and Schumacher (1987) as well as Chapters 2 and 3 (Sengupta *et al.*, 1998; Bhattacharjee, 2007a), the tests are consistent against the trended alternative (5.13). 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 5.3.1, whereas under the trended alternative (5.13), 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 (5.14). 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 5.3.4.

Finally, the choice of the r pairs of covariate values may be important in applications. The issues regarding this choice are similar to those relating to stratification in goodness-of-fit tests. Quantiles of the cross-sectional distribution of the covariate can be used to select these covariate pairs and to construct the "small" intervals around the covariate values – this, in a simple way, ensures that variations in the density of design points are adjusted for (none of the intervals are too sparse) and that the intervals corresponding to each pair of covariate values do not overlap. In our simulation studies (Section 5.4), we divided the sample into deciles by the magnitude of the covariate, and based our tests on the $\binom{10}{2} = 45$ covariate pairs generated by this construction, while for the empirical application (Section 5.5), we used 20 covariate pairs obtained by random sampling.

Extension of the tests to the case when other covariates, Z, are also present is straightforward. Here, we build an appropriate Cox model, possibly with time varying coefficients on Z, and estimate the model 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 (5.12) 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 frailty or parametric frailty, the tests are constructed as above, using an appropriate estimator for the baseline cumulative hazard function. Though martingale based arguments are not valid any more, the asymptotic arguments still hold, with some minor modifications. For the shared frailty model, results from Spiekerman and Lin (1998) demonstrate this; see Theorems 1 and 2 in Spiekerman and Lin (1998). For the parametric individual level frailty model, a procedure similar to continuously distributed unrestricted frailty can be used. This is discussed below (Section 5.3.3) in the context of the Horowitz (1999) estimator.

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; see also Scheike and Martinussen (2004).

5.4.3 Testing the proportional hazards assumption

Our proposed tests for proportional hazards are similar to those for the previous testing problem. Here, too, we estimate the baseline cumulative hazard function under maintained assumptions on the model and nature of frailty, and plug these estimators into the two sample test statistic (5.12) in place of the Nelson-Aalen estimator. The asymptotic properties are similar to those given by Theorem 5.3.1 and Corollaries 5.3.1 and 5.3.2. However, the assumptions underlying the tests reflect the differences in the models and methods, and similarly there are important differences in the asymptotic arguments. Below, we discuss continuous failure time data with arbitrary continuous frailty, followed by discrete failure time data combined with a discrete mixture frailty distribution.

We first consider the kernel-based estimation procedure proposed by Horowitz (1999) under the continuous failure time MPH model with unrestricted continuously distributed frailty. 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 (5.2) 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 MPH model with time varying coefficients (5.2),

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

has been found. The 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 frailty and conditional on $X=x$.

The testing procedure will be similar to Section 5.3.2, starting with the choice of $r>1$ and selection of $2r$ (r is a fixed positive integer, $r>1$) distinct points, $\{x_{11},x_{21},\ldots,x_{r1},x_{12},x_{22},\ldots,x_{r2}\}$,

$x_{l2} > x_{l1}, l = 1, \ldots, r$ on the covariate space $\mathcal{X}$. Next, we construct the basic statistics as

$$
\begin{aligned}
T_{H,std}(x_{l1}, x_{l2}) &= \frac{T_H(x_{l1}, x_{l2})}{\sqrt{\widehat{\text{Var}}\left[T_H(x_{l1}, x_{l2})\right]}}, \qquad (5.22)\\
T_H(x_{l1}, x_{l2}) &= \int_0^{\tau^*} L(x_{l1}, x_{l2})(t).\widehat{\lambda}_{0,H}(t, x_{l1}).dt - \int_0^{\tau^*} L(x_{l1}, x_{l2})(t).\widehat{\lambda}_{0,H}(t, x_{l2}).dt,\\
\widehat{\text{Var}}\left[T_H(x_{l1}, x_{l2})\right] &= \int_0^{\tau^*}\int_0^{\tau^*} \widehat{c}(t).\widehat{c}(s).\widehat{\sigma}_L^2(x_{l1}, x_{l2})(s \wedge t).ds.dt,
\end{aligned}
$$

where $L(x_{l1}, x_{l2})(t)$ is a random process indexed on the pair of covariate values x_{l1} and x_{l2}, $\widehat{\sigma}_L^2(x_{l1}, x_{l2})(t)$ 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 (5.19 – 5.21), these basic statistics are combined to construct our maxima, minima and average test statistics (denoted $T_H^{(\max)}, T_H^{(\min)}$ 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 frailty. 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 failure time data $(T_i, \delta_i, X_i, Z_i(t), U_i)$ are independently and identically sampled from the MPH model with time varying coefficients (5.2), for $i = 1, \ldots, n$. Here, T_i denotes the observed lifetime, δ_i is the censoring indicator, X_i and $Z_i(t)$ are covariates, and U_i is the unobserved frailty. The following additional assumptions apply.

Assumption 5.3.5 *The cut-off failure time, $\tau^* > t_0 > 0$, is a (large) positive lifetime such that $\Lambda_0(\tau^*, x_{lj}) < \infty, l = 1, 2, \ldots, r, j = 1, 2$. The intermediate lifetime t_0 is specified in Assumption 5.3.7 (b) below.*

Assumption 5.3.6 *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 5.3.7 *(Identifiability conditions)*

(a) *Frailty U is independent of covariates Z and censoring, and there is a tail restriction on the frailty distribution.*[7]

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

(c) *The covariate effect of at least one of the covariates, say Z_1, is significant 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.*

(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 5.3.8 *(Smoothness conditions and kernel properties)*

(a) *Smoothness conditions involving several bounded derivatives for the unknown frailty 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 5.3.9 *(Conditions on regression estimator) The underlying regression estimator for the transformation model converges at $n^{-1/2}$ rate and has bounded second moments.*

Some qualifying comments are necessary. First, dependence between frailty and X is not ruled out. However, we view testing for ptoportional hazards as a step towards appropriate specification of a regression model. The additional assumption of independence may be required for further modeling. Second, unlike the standard literature (see, for example, Andersen *et al.*, 1993), the setup in Horowitz (1999) allows censoring to depend on the covariates through the single index. This, in our view represents a strength of the methodology, particularly allowing censoring to depend on the covariate under test. Third, the methodology does not directly allow for time varying covariates. However, if the regression coefficient is fixed, a time varying

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

covariate can be naturally accomodated by replacing the time varying covariate by its average value over the observed lifetime. 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).[8] Fourth, standard regression estimators for the transformation model satisfy the covergence rate and finite second moments conditions. Fifth, the smoothness and kernel conditions are satisfied by the Horowitz (1999) estimator. Further, it turns out that appropriate choice of bandwidths and other tuning parameters is very important for good performance of the estimator. Finally, the Assumptions 5.3.7 through 5.3.9 ensure pointwise consistency of the baseline hazard estimator, which is required for our tests.[9]

Additional conditions required for the test are given in Assumptions 5.3.5 and 5.3.6. These comprise a deterministic cut-off at a failure time where the cumulative 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.

We are now ready to state the asymptotic results.

Theorem 5.3.2. *Let Assumptions 5.3.5 through 5.3.9 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.)

Corollary 5.3.3.

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

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

[9] Horowitz (1999) actually shows that the estimator is uniformly consistent and pointwise asymptotically gaussian.

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).

A result similar to Corollary 5.3.2 on covariate dependent weighted tests is also available. Details are very similar to Section 5.2, and are omitted. A final point worth noting is that, while the form of the above test is similar to the test for absence of covariate effects (Section 5.2), as well as the test in Chapter 3 (Bhattacharjee, 2007a), 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 earlier 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 are therefore required.

For the alternative estimator, proposed by Gørgens and Horowitz (1999), which we consider next, the above approach is not directly applicable, since an estimator is available only for the baseline cumulative hazard function.[10] On the other hand, this estimator has the advantage of convergence to a Gaussian process with continuous sample paths. Despite this, it seems inevitable that either a mixing or a m-dependence kind of assumption would be necessary for the asymptotics in this case. This appears to be too strong a condition. Hence, we attempted an alternative strategy, which is intuitive and potentially promising. Though this approach is not entirely satisfactory, we report this below for the sake of completeness.

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

[10] However, as in Chapter 3 (Bhattacharjee, 2007a), a related test based on the cumulative hazard function can be developed. The natural alternative hypothesis here will be based on star (or negative) star ordering.

appropriate MPH model with time varying coefficients (5.2) has been found. Next, we adjust the model in a way suitable for our test. Specifically, what we require are estimators of the processes

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

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 model

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

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 the estimation method does not directly allow for time varying covariates. This is because the MPH model with time varying covariates is not a transformation model. But, in the case that the corresponding coefficient is fixed, this can be addressed by substituting the covariate value by an average over the lifetime of the time varying covariate. This procedure can be followed for the additional covariate above, by substituting for it the average value $\int_0^\tau \tau^{-1} \ln L(x_{l1}, x_{l2})(t).dt$. Since time varying coefficients are incorporated in the model using histogram sieves, a similar procedure can also be followed for all other time varying covariates.

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 5.3.7a *(Identifiability conditions)*

(a) *In addition to covariates and censoring, frailty 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 frailty and the weight function is a strong assumption in our case. We take the view that the relevant component of frailty here is its projection onto the orthogonal space of the covariates and the weight function. This is in line with interpretation of frailty 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 a bit more critical. For this purpose, we suggest the weighted and nonparametric bootstrap procedures developed in Kosorok *et al.* (2004). These methods are valid under a wide class of continuous frailty distributions, but under some additional assumptions; see Kosorok *et al.* (2004) for details.

In summary, for arbitrary continuous frailties, the tests based on the estimator proposed by Horowitz (1999) is implementable. We have proposed an alternative procedure based on the Gørgens and Horowitz (1999) which, though potentially attractive, requires some further development before it can be implemented in real data situations.

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

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

The sequential procedure is terminated when subsequent steps lead to degeneracy or no improvement in the maximised likelihood. This methodology for approximating any arbitrary frailty distribution is very useful in that it approximates the nonparametric frailty distribution by an increasing sequence of parametric distributions, and it produces robust estimates of regression parameters and the baseline hazard function.[11]

In our implementation, we follow Jenkins (1995) in combining the above frailty distribution with a discrete grouped failure time version of the proportional hazards model (1.17), or the complementary log-log model (Cox, 1972; Kalbfleisch and Prentice, 1973; Prentice and Gloeckler, 1978; Cox and Oakes, 1984), discussed previously in Section 1.2.7.5

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

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 $\{\widehat{\pi}_1, \widehat{\pi}_2, \ldots, \pi_s = 1 - \widehat{\pi}_1 - \widehat{\pi}_2 - \ldots - \widehat{\pi}_{s-1}\}$. The covariate pairs are chosen as before.

The assumptions underlying the testing procedure, and a brief description of assumptions for estimation are as follows.

The discrete failure time data $(T_i, \delta_i, X_i, Z_i(t), U_i)$ are independently and identically sampled from the above complementary log-log model with discrete mixture frailties, for $i =$

[11] 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 documented convergence problems (Horowitz, 1999) .

$1, \ldots, n$. The following additional assumptions hold.

Assumption 5.3.10 *The cut-off failure time $0 < T < \infty$ 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, $t = 1, 2, \ldots, T$, we have a positive baseline hazard rate: $\gamma_{t,x} > 0$.*

Assumption 5.3.11 *For each $l, l = 1, 2, \ldots, r$, 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 5.3.12 *(Identifiability conditions)*

(a) *Frailty U is independent of covariates Z and censoring. A tail restriction is required on the frailty distribution, for both discrete and continuous failure times. For the test, we also assume independence between frailty 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.*

Assumption 5.3.13 *(Identification of finite mixture frailty distribution) The conditions, originally given by Lindsay (1983a, 1983b), state that the density of the data at each mass point of the frailty distribution is a bounded function of the regression parameters.*

Assumption 5.3.14 *Boundedness and right continuity of the baseline hazard function and the regression parameters.*

The above assumptions are fairly standard. They are also 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 as

$$\begin{aligned}
T_{HS,std}(x_{l1},x_{l2}) &= \frac{T_{HS}(x_{l1},x_{l2})}{\sqrt{\widehat{\text{Var}}\left[T_{HS}(x_{l1},x_{l2})\right]}}, \\
T_{HS}(x_{l1},x_{l2}) &= \sum_{t=1}^{T} L_t(x_{l1},x_{l2}).\left[\widehat{\gamma}_{t,x_{l1}} - \widehat{\gamma}_{t,x_{l2}}\right], \qquad (5.23) \\
\widehat{\text{Var}}\left[T_{HS}(x_{l1},x_{l2})\right] &= \sum_{t=1}^{T}\sum_{s=1}^{T}\left[\widehat{\gamma}_{t,x_{l1}} - \widehat{\gamma}_{t,x_{l2}}\right].\left[\widehat{\gamma}_{s,x_{l1}} - \widehat{\gamma}_{s,x_{l2}}\right].\widehat{\sigma}^2_{s\wedge t}(x_{l1},x_{l2}),
\end{aligned}$$

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 (5.19 – 5.21), these basic statistics are combined to construct our maxima, minima and average test statistics (denoted $T^{(\max)}_{HS}, T^{(\min)}_{HS}$ and $\overline{T}_{HS}$, respectively).

Then, we have the following asymptotic results.

Theorem 5.3.3. *Let Assumptions 5.3.10 through 5.3.14 hold. Then, under* $\mathbb{H}_{0,PH}: \gamma_{t,x} = c_t$ for all x, *as* $n \to \infty$,

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

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

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

(Proof in Appendix.)

Corollary 5.3.4.

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

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).

As in the continuous frailty 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 frailty can also be used to model the frailty 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 failure time setting, particularly if we are not as such interested in estimating the frailty distribution. We have not pursued either of these approaches here.

5.4.4 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 5.3.2 and 5.3.3. 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 (5.8) and the alternatives of special interest are either trended (5.13) or changepoint trended (5.14). For the corresponding two sample tests, the logrank weight function (Mantel, 1966; Cox, 1972; Peto and Peto, 1972), 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 (5.13). 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 (5.14), because both positive and negative trends are evident on different regions of the sample space. In other words, the log rank weight function may be quite appropriate for the proposed test for absence of covariate dependence, particularly if the suspected alternative is of a PH nature. The Gehan-Breslow weight function (Gehan, 1965; Breslow, 1970), given by $L(t) = Y_1(t).Y_2(t).\,[Y_1(t) + Y_2(t)]$, may also be useful, particularly if censoring is high. Compared to the logrank test, this weight function places higher weight on differences in the hazard function at shorter failure times (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). This property of the Prentice weight function is discussed in Prentice (1978) and Gill and Schumacher (1987), and demonstrated in simulation studies (Krogen and Magel, 2000; Jung and Jeong, 2003). 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 failure time distributions. Hence, the Prentice weight function will be appropriate for testing proportionality against these ordered alternatives.[12]

5.5 Simulation study

The asymptotic distributions of the proposed test statistics were derived in Section 5.3. 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, $\mathbb{H}_{0,Eq}$ (5.8), 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 $\mathbb{H}^{(t)}_{1,Eq}$ (5.13). 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 $\mathbb{H}^{(c)}_{1,Eq}$ (5.14). The tests discussed in Section 5.3.2 are consistent against the global alternative $\mathbb{H}_{1,Eq}$ (5.9), but are also expected to be powerful

[12] Note that, because of frailty, a martingale based framework is not available here and the predictability property is therefore not useful. However, the cadlag nature of $\widehat{S}(t)$ makes the weight function itself cadlag, which is required for our tests.

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 5.4.1.

TABLE 5.4.1: Data Generating Processes

(Test for absence of covariate dependence)

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

The covariate X is 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 (5.8), 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 5.4.1 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 them will reflect the increasing nature of covariate dependence, and another quarter reflect the decreasing trend. The results also reflect 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.

TABLE 5.4.2: 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

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.

Next, we examine the performance of the tests for the proportional hazards assumption in the presence of frailty (5.3) against ordered alternatives of the IHRCC type (5.4, 5.5). The design of the data generating process is a combination of Horowitz (1999) and our Chapter 3 (Bhattacharjee, 2007a), and 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 frailty 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.[13] We consider a single specification of the baseline hazard function as

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

and 2 different patterns of covariate dependence

$$\beta_X(t) = \begin{cases} 1 \\ \ln(t) \end{cases},$$

in combination with 2 frailty distributions. One frailty 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 lifetime data are right censored by independent censoring times distributed as $Uniform\,(0.5, 25.5)$.

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 frailty distribution. The description of the DGPs and expected results are summarised in Table 5.4.3. 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.

[13] 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.

TABLE 5.4.3: Data Generating Processes

(Test for proportional hazards, with frailty)

Model	$\lambda_0(t)$	$\beta_X(t)$	**Frailty**	**Median cens.dur.**	**% cens.**	**Expected significance**
DGP_{31}	$0.087t$	1	Continuous	5.23	23.4	None
DGP_{32}	$0.087t$	$\ln(t)$	Continuous	5.37	25.8	$T_{HS}^{(\max)}, \overline{T}_{HS}$
DGP_{41}	$0.087t$	1	Mixture	5.16	23.6	None
DGP_{42}	$0.087t$	$\ln(t)$	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 frailty 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.[14] Horowitz (1999) suggests the use of cross-validation or bootstrap 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 implemented in individual applications fairly well.

Implementing the Heckman and Singer (1984a) method was relatively more 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.[15] The results presented in Table 5.4.4 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 frailty distribution, while using data for each decile to estimate the baseline

[14] 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.

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

hazard function; we have not investigated this approach further.

TABLE 5.4.4: Test for Proportional Hazards, with Frailty

(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

Considering the challenges noted above, and slow convergence of the maximum likelihood procedure, we report results based on a modest 200 Monte 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 put our methods to test on real life data.

5.6 An application

Here, we illustrate the use of the tests proposed in this chapter using an application based on real life data. The objective is to study the effect of aggregate Q on the hazard rate of corporate failure in the UK. The data are on firm exits through bankruptcy over the period 1980 to 1998 and pertain to 2789 listed manufacturing companies, covering 24,034 company years and includes 95 bankruptcies. The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1980, and contain staggered entries. Here the focus of our analysis 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.[16]

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.

TABLE 5.5.1: 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

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 5.5.1). 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 $x_1 > x_2$). 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.

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

[16] The dataset constitutes the empirical context behind much of the work in this thesis. It has been discussed earlier in Chapters 3 and 4, and will be used in Chapter 6 too. More detailed analysis of these data, based on Bhattacharjee *et al.* (2008a, 2008b) will be discussed in Chapter 7.

useful for investigating the nature of departures from proportionality.[17] 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 5.5.2: 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 5.5.2 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 ages. Similarly, tests for proportional hazards against order restricted covariate effects in the absence of frailty, discussed in Chapter 3 (Bhattacharjee, 2007a), reject the null hypothesis of proportionality against a $DHRCC$ (5.11) alternative.[18]

However, the above inference could also 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 frailty is present in these data. Therefore, we 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 frailty. 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 frailty using the Heckman and Singer (1984a) procedure, and estimate grouped failure time proportional hazards models conditional on various values

[17] This is in line with the way we approximately located changepoints in Chapter 3 (Bhattacharjee, 2007a).

[18] Specifically, the test statistics $T_{GS}^{(\min)}$ and $\overline{T}_{GS}$ are both significant at the 1 per cent level of significance.

of the covariate under test, Q. As expected, there is significant frailty in the data. However, the tests for proportional hazards, based on 20 randomly chosen covariate pairs, produce the same inference as before. The null hypothesis of proportionbal hazards is rejected in favour of a $DHRCC$ alternative. Both $T_{GS}^{(\min)}$ and $\overline{T}_{GS}$ 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. Further, these tests can provide clues about the approximate location of such changepoints, when present. Similarly, the proposed tests for proportional hazards are powerful against ordered covariate effects, in the presence of arbitrary frailty. 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.

5.7 Conclusion

In summary, the tests described in this chapter add important tools to the armoury of a lifetime or 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 failure time data, and with and without the presence of frailty. The work extends the horizon of inference procedures beyond martingale based continuous failure time methods described in Fleming and Harrington (1991) and Andersen *et al.* (1993), extensions to discrete life history data (Hjort, 1985; Sengupta and Jammalamadaka, 1993), shared frailty models (Spiekerman and Lin, 1998; Andersen *et al.*, 1999) and recurrent failure time data (Lin *et al.*, 2000; Lin and Ying, 2001).

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 frailty 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 Chapter 3 (Bhattacharjee, 2007a), 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.

The basic statistics encountered in our tests for proportional hazards are of the form

$$\sum_{i=1}^{n} \int K_i(t).H(t).dt, \tag{5.24}$$

where $K_i(.)$ $(i = 1, \ldots, n)$ are iid copies of stochastic processes, and $H(.)$ involves data from all the n observations. By contrast, for testing absence of covariate dependence, we used statistics like

$$\sum_{i=1}^{n} \int K_i(t).dM_i(t), \tag{5.25}$$

which are standard in the analysis of failure time data based on counting processes. For (5.25), asymptotic results typically follow from martingale theory, under the conditions that $M_i(.)$ are martingales and $K_i(.)$ are predictable processes. Using empirical process arguments, Lin *et al.* (2000) and Lin and Ying (2001) have extended inference methods for (5.25) to statistics where the $K_i(.)$ are replaced by a process $H(.)$ involving data from all the observations. We show how modern empirical process theory in combination with Theorem 2.3.1 (Sengupta *et al.*, 1998) can be used to derive asymptotic theory for the statistics like (5.24).

Several areas of further research emerge from our work. First, the development of asymptotic arguments for statistics like (5.24) is useful in contexts well beyond the current application. In fact, the tests proposed here do not fully use the strengths of this methodology. While the fact that $K_i(.)$ are monotonic simplifies arguments in our case, the condition required is that the process has a finite pseudodimension, as defined by Pollard (1990). Similarly, the main condition required of $H(t)$ is that it has a continuous probability limit. For example, in the context of frailty models, one can think of alternate statistics constructed by plugging-in the estimated frailty distribution in the counting process martingale. Exploration of these and other applications is beyond our current scope.

Second, the tests for absence of covariate dependence extend a well-known family of two-sample tests to the continuous covariate setup. Together with related tests for proportional hazards developed in Chapter 2, these methods raise important new research questions, particularly 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, development of new tests for the proportional hazards assumption using either the Gørgens and Horowitz (1999) estimator, or by pooled estimation of the frailty distribution using the Heckman and Singer (1984a) approach, will be useful extensions of the current work.

Fourth, the proposed tests for proportional hazards in the presence of frailty, together with the application considered here, further emphasize the importance of considering frailty together with monotonic covariate effects in empirical studies. In Section 7.4 (Bhattacharjee, 2007c), we return to this issue and consider joint modeling of nonproportional covariate effects and unrestricted frailty.

Fifth, our work here demonstrates that appropriate specification of the frailty distribution is important not only for inference on the nature of and order restrictions on the covariate effects, but also on the shape of the baseline hazard function. In Chapter 6 (Bhattacharjee and Bhattacharjee, 2007), we develop Bayesian methods to address the issue of joint inference on potentially nonproportional covariate effects and order restrictions on ageing, in the presence of unrestricted frailty.

Finally, and perhaps most importantly, our simulation study as well as the application considered point to the need to make important progress in the estimation of hazard regression models under unrestricted frailty. This is statistically a difficult problem. As we have discussed, currently available methods are not satisfactory, either due to convergence issues or the degree of specific tuning required for their implementation. We have discussed some alternative approaches earlier. In particular, developing an appropriate bootstrap procedure for the Horowitz (1999) estimator and the maximum rank correlation approach of Hausman and

Woutersen (2005) may both be useful.[19] This is currently an active research area, and further developments will emerge in coming years.

Appendix to Chapter 5

Proof of Theorem 5.3.1

It follows from standard counting process arguments (see, for example, Andersen *et. al*, 1993) that, under $\mathbb{H}_{0,Eq}$ (5.8), for $l = 1, \ldots, r$,

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

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 $\widehat{\mathrm{Var}}\left[T_{2s}\left(x_{l1}, x_{l2}\right)\right]$ is a consistent estimator for the variance of $T_{2s}\left(x_{l1}, x_{l2}\right)$, we have as $n \longrightarrow \infty$,

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

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 count-

[19] Another recent approach developed in Zeng and Lin (2007) makes it possible to computationally address frailty issues in much more challenging models; the statistical content of their work uses empirical process methods. However, their development appears to be heavily specific to the assumption of the lognormal frailty distribution, which appears to be quite restrictive (Bickel, 2007; Horowitz, 2007).

ing process are orthogonal, and the vector of these martingales asymptotically converge to a Gaussian martingale. A similar argument in the context of testing for proporional hazards is given in Chapter 3 (Bhattacharjee, 2007a).

It follows that

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

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

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

□

Proof of Corollary 5.3.1

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 5.3.2

From Theorem 5.3.1, 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 $\mathbf{I_r}$ is the identity matrix of order r.

The proof follows straightaway.

□

Proof of Theorem 5.3.2

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 5.3.1 above.

For proving weak convergence of the basic statistic, we make use of Theorem 2.3.1 (Sengupta *et al.*, 1998). In order to study the convergence of $T_i, i = 1, 2$, we replace $\boldsymbol{K}_n(t)$ and $\boldsymbol{X}_n(t)$ in the above theorem by $\left[\widehat{\lambda}_{0,H}\left(t, x_{l1}\right) : \widehat{\lambda}_{0,H}\left(t, x_{l2}\right)\right]^T$ and $[L(x_{l1}, x_{l2})(t)]$, respectively.

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

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

for $t \in [0, \tau^*]$, and by our assumptions, $\lambda_{0,H}\left(t, x_{lj}\right)$ 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. Example 2.11.16 in van der Vaart and Wellner (1996) can also be slightly modified to show that a monotonic process with finite first and second moments on an interval converges weakly to a Gaussian process. However, we prefer the first approach because it can be used in other applications where the process is not necessarily monotonic.

Now, by applying Theoren 2.3.1, we have

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

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

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

This completes the first part of the proof.

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

$$\int_0^{\tau^*}\int_0^{\tau^*} c(t).c(s).V\left(s \wedge t\right).dsdt,$$

where

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

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

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

□

Proof of Corollary 5.3.3

Proof follows exactly in the same way as Corollary 5.3.1.

□

Proof of Theorem 5.3.3

With discrete data, the problem here 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 5.3.1, 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 5.3.1.

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 5.3.2, 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 routine application of Slutsky's theorem, continuous mapping theorem and the multivariate central limit theorem.

As in proof of Theorem 5.3.2, the variance of the limiting distribution is given by

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

where $\sigma_t^2(x_{l1}, x_{l2})$ 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 $\sigma_t^2(x_{l1}, x_{l2})$ 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 $T_{HS}(x_{l1}, x_{l2})$.

Further, since $\widehat{\mathrm{Var}}\left[T_{HS}\left(x_{l1}, x_{l2}\right)\right]$ is a consistent estimator for the variance of $T_{HS}\left(x_{l1}, x_{l2}\right)$, we have 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} $(l = 1, \ldots, r; j = 1, 2)$.

It follows that

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

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

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

□

Proof of Corollary 5.3.4

Proof follows exactly in the same way as the proof of Corollary 5.3.1.

□

Chapter 6

Bayesian Analysis of Hazard Regression Models under Order Restrictions on Covariate Effects and Ageing

6.1 Chapter summary

In this chapter, based on Bhattacharjee and Bhattacharjee (2007), we propose Bayesian inference in hazard regression models where the baseline hazard is unknown, covariate effects are possibly non-proportional, and there is multiplicative frailty with unknown distribution. The covariate effects, which are potentially ordered rather than proportional, are estimated and evaluated using time varying coefficients. In addition, we consider restrictions on ageing, specifically in the nature of a decreasing baseline hazard function. Thus, the proposed framework enables evaluation of order restrictions in the nature of both covariate and duration dependence (ageing), and in the presence of unrestricted frailty. The usefulness of the proposed Bayesian model and inference methods are illustrated with an application to corporate bankruptcies in the UK.

6.2 Introduction

Understanding the nature of covariate dependence and ageing are the main objectives of regression analysis of lifetime data. In many applications, relevant underlying theory or preliminary analysis may suggest that there are important order restrictions on either covariate dependence, or the shape of the baseline hazard, or both. Parametric inference in such situations can be conducted by making functional form or distributional assumptions that impose the above order restrictions. However, such assumptions can be very restrictive and may lead to weak inference. Instead, one may aim to conduct order restricted nonparametric analysis under the constraints implied by theory or past experience. In fact, such inference can also be used to judge the validity of the order restrictions themselves.

In this chapter, we propose Bayesian models to conduct order restricted nonparametric inference in applications with single spell lifetime data. Specifically, our framework for inference in hazard regression models incorporates three important features. First, we do not assume proportional hazards with respect to all covariates included in the analysis. As discussed earlier in Chapters 2, 3, 4 and 5 (Sengupta *et al.*, 1998; Bhattacharjee, 2004a, 2007a, 2007b), the proportionality assumption underlying the Cox regression hazards model does not hold in many applications. At the same time, credible inference under the model depends crucially on the validity of the proportionality assumption. Further, the effect of a covariate is often monotone, in the sense that the lifetime (or duration) conditional on a higher value of the covariate ages faster or slower than that conditional on a lower value (Chapter 3, Bhattacharjee, 2007a). In particular, we consider relative ageing in the nature of convex or concave ordering (Kalashnikov and Rachev, 1986) of lifetime distributions conditional on different values of the covariate in question. Ordered departures of this kind are common in applications, and the models provide useful and intuitively appealing descriptions of covariate dependence in non-proportional situations. Further, as discussed in Chapter 4, ordered departures of the above kind can be convenienty studied in a Cox type regression model with time varying coefficients (Bhattacharjee, 2003, 2004a), where positive ageing for higher covariate values implies that the time varying effect of the covariate is a nondecreasing function of lifetime. In other words, our hypothesized covariate efects are of the $IHRCC$ or $DHRCC$ type (Definition 3.2.1), with monotone time varying coefficients for some selected covariates, in cases when the proportional hazards assumption fails to hold.

Second, in addition to order restricted covariate dependence, we allow for constraints on the shape of the baseline hazard function. These order restrictions will typically be in the nature of monotone increasing or decreasing baseline hazard rates. They could also be characterised by weaker notions of ageing, such as "new better than used". As discussed above, these kinds of ordering are important in many applications, and reflect the inherent structural nature of the ageing process, irrespective of differences in observed or unobserved covariates.

The third feature of our work is in the treatment of frailty. In our approach, unobserved covariates induce hazard rates to vary across individuals in two different ways. Unobserved covariates that act at the group level (and are therefore identified by group membership) are incorporated in our model as fixed effects heterogeneity. In addition, as in Chapter 5 (Bhattacharjee, 2007b), we allow a scalar unobserved covariate independent of the included regressors which has a completely unspecified distribution. Our approach is in contrast of much of the literature that specifies a parametric frailty distribution. Our chosen nonparametric approach to modeling frailty (Heckman and Singer, 1984a) operates through a sequence of discrete multinomial distributions. Each of these distributions comprises a set of mass points along with the probabilities of a subject being located at each mass point. By progressively increasing the number of mass points, we are able to approximate any arbitrary frailty distribution; see also Section 1.2.6 and Chapter 5 (Bhattacharjee, 2007b) for previous discussion of the Heckman and Singer (1984a) approach.

The Bayesian approach adopted in the current work offers three main advantages. First, as discussed above, we develop Bayesian inference incorporating order restrictions jointly on covariate and duration dependence, in the presence of unrestricted univariate frailty. As discussed in Chapter 5 (Bhattacharjee, 2007b), frequentist inference on nonproportional covariate effects with an unrestricted frailty distribution is itself a challenging problem. As one may imagine, the computational challenges would be further enhanced in the presence of additional order restrictions on ageing. In this regard, the Bayesian framework, with its associated efficient MCMC implementations, offers an attractive and implementable approach. Second, the framework enables prior beliefs to be explicitly incorporated in the model, particularly beliefs regarding the assumed order restrictions. Therefore, this constitutes a natural and particularly attractive approach for inference under order restrictions. Further, if these prior beliefs can be represented by models that place zero mass on specific regions of the parameter space, the posterior distributions too would have the same property. This feature of the Bayesian approach makes it an useful framework for studying order restrictions. Third, the Bayesian framework offer the important advantage of accommodating, in a natural way, parameter uncertainty involved in the inference process. As discussed in Chapter 4 (Bhattacharjee, 2004a), it is difficult in the frequentist approach to adjust estimation procedures for such uncertainty, and obtain standard errors accounting for pretesting. This issue is addressed in a very natural way in the Bayesian approach adopted in this chapter.

The chapter is organised as follows. Section 6.2 presents a selective review of the literature. We describe our model in Section 6.3 and our application is presented and discussed in Section 6.4. Finally, Section 6.5 concludes.

6.3 Background

Here, our context is order restricted Bayesian semiparametric inference for hazard regression models. Specifically, we consider the MPH model with time varying coefficients (1.14, 5.2)

$$\lambda\left(t|\underline{X}_i(t)\right) = \lambda_0(t).\exp\left[\underline{\beta}(t)^T.\underline{X}_i(t)\right].u_i, \qquad u_i \epsilon\,(0,\infty) \overset{iid}{\sim} F_U,$$

where order restrictions on covariate dependence posit monotone time varying coefficients for some of the covariates, and order restricitions on ageing imply shape constraints on $\lambda_0(t)$. Further, the distribution of individual level frailty is completely unrestricted.

The work here is quite unique in that there is very little prior literature in this specific area. However, there is literature in several related areas, both in a Bayesian paradigm as well as frequentist inference. Earlier, in Section 1.2.8, we have surveyed the related literature on Bayesian semiparametric inference in the hazard regression models. Similarly, the relevant literature in the frequentist framework has been discussed in Chapters 1, 2 and 3.

Here, we briefly survey the literature on order restricted Bayesian and frequentist inference with a view to place our current work within the context of the literature and to highlight the distinctive nature of our approach.

6.3.1 Bayesian semiparametric inference

Semiparametric approaches to Bayesian inference in hazard regression models usually assume the standard Cox proportional hazards model with (time varying) covariates. The covariate effects are held constant over the lifetime, but the baseline hazard function is unrestricted. Various Bayesian formulations of the model differ mainly in the nonparametric specification of $\lambda_0(t)$; see Section 1.2.8 for further discussion. A notable departure is the work of Gamerman (1991), where time varying coefficients are modeled using a Markov process, and a subsequent refinement proposed by Sargent (1997).

As discussed in Section 1.2.8.3, Bayesian approaches have genrally addressed the presence of frailty using various parametric distributions. At the same time, in its ability to deal with potentially large number of latent variables, the Bayesian framework offers the possibility of a more nonparametric approach to modeling individual level frailty. Based on repeated failures data, Bhattacharjee *et al.* (2003) and Arjas and Bhattacharjee (2003) have proposed a hierarchical Bayesian model based on a latent variable structure for modeling unobserved heterogeneity; the model is very powerful and shown to be useful in applications. Since our application here is based on single failure per subject data, we use a latent variable structure but with the objective of inferring on the frailty distribution rather than the latent variables themselves. We model frailty in two different ways. First, we divide the subjects into groups and incorporate fixed effects unobserved heterogeneity across these different groups. Second, as in Chapter 5 (Bhattacherjee, 2007b), we model individual level frailty in a more nonparametric tradition (Heckman and Singer, 1984a) by introducing a sequence of multinomial frailty distributions with increasing number of support points; for a related Bayesian implementation, see Campolieti (2001).

However, our main focus in this chapter is on order restricted inference in hazard regression models. The literature on order restricted Bayesian inference, with restrictions either on the shape of the baseline hazard function or on the nature of covariate depence, is rather limited. Contributions in this area relevant to our current work include Arjas and Gasbarra (1996), Sinha *et al.* (1999), Gelfand and Kottas (2001) and Dunson and Herring (2003).

Arjas and Gasbarra (1996) developed models of the hazard rate processes in two samples under the restriction of stochastic ordering. They defined their prior on the space of pairs of hazard rate functions; the unconstrained prior in this space consists of piecewise constant gamma distributed hazards which incorporate path dependence. The constrained prior is then constructed by restricting to a subspace on which the maintained order restriction holds. In their work, Arjas and Gasbarra (1996) propose a coupled and constrained Metropolis-Hastings algorithm for posterior elicitation based on the order restriction and also for Bayesian evaluation of the stochastic ordering assumed in the analysis. For the same problem, Gelfand and Kottas (2001) developed an alternative prior specification and computational algorithm. The Bayesian model in Arjas and Gasbarra (1996), in combination with the general treatment of Bayesian order restricted inference (for example, in Gelfand *et al.*, 1992), is related to the current work.

Sinha *et al.* (1999) proposed Bayesian analysis with interval censored data where covariate dependence is possibly nonproportional. They modeled the baseline hazard function using an independent Gamma prior and the time varying coefficients were endowed with a Markov type property $\beta_{k+1}|\beta_1,\ldots,\beta_k \sim N(\beta_k, 1)$.While Sinha *et al.* (1999) did not explicitly consider order restrictions either on covariate dependence or on ageing, they provide Bayesian inference procedures to infer on the validity of the proportional hazards assumption.

In another important contribution related to our work, Dunson and Herring (2003) considered order restriction on covariate dependence, and developed Bayesian methods for inferring on the restriction that the effect of an ordinal covariate is higher for higher levels of the covariate. In other words, similar to our tests on absence of covariate effects (Chapter 5, Bhattacharjee, 2007b), they conducted inference on trend in conditional hazard functions. By contrast, our current work with restrictions on covariate dependence is different in two respects. First, in our case the covariate is continuous and not categorical. Second, our order restriction is related to convex or concave partial ordering of conditional hazard functions rather than trend. Consequently, we express our constraints in terms of monotonic time varying coefficients, and propose a different methodology for Bayesian inference.

6.3.2 Order restricted frequentist inference

Order restrictions relating both to the shape of the baseline hazard function (ageing) as well as the effect of covariates (covariate dependence) are important in the study of hazard regression models. However, as discussed in Sections 1.2.4 and 1.2.5, the literature on frequentist order restricted inference in hazard regression models mainly addresses covariate dependence. In the two sample (binary covariate) setup, testing for proportionality of hazards against different notions of relative ageing has been an active area of research. The monotone hazard ratio alternative was considered by Gill and Schumacher (1987) and Deshpande and Sengupta (1995), while in Chapter 2 (Sengupta *et al.*, 1998), we develop tests for proportionality against the weaker alternative hypothesis positing a monotone ratio of cumulative hazards. Order restricted estimation in two samples under the corresponding partial orderings (convex ordering and star ordering) has not been explicitly considered in the literature, though methods developed in Chapter 3 (Bhattacharjee, 2004a) can be adapted to this problem. Further, estimation in two samples with right-censored survival data under the stronger constraint of stochastic ordering has been considered in Dykstra (1982), and extended to uniform conditional stochastic ordering in the k-sample setup by Dykstra *et al.* (1991). These inference procedures are, however, not very useful in the hazard regression context, where covariates are typically continuous in nature.

In Chapter 3 (Bhattacharjee, 2007a), we extended the notion of monotone hazard ratio in two samples to the situation when the covariate is continuous, and proposed tests for proportional hazards against ordered alternatives. Specifically, the alternative hypothesis here states that, lifetime conditional on a higher value of the covariate is convex (or concave) ordered with respect to that conditional on a lower covariate value:

$$\begin{aligned} IHRCC &: \text{ whenever } x_1 > x_2, \lambda(t|x_1)/\lambda(t|x_2) \uparrow t (\equiv (T|X = x_1) \underset{c}{\prec} (T|X = x_2), \\ DHRCC &: \text{ whenever } x_1 > x_2, \lambda(t|x_2)/\lambda(t|x_1) \uparrow t (\equiv (T|X = x_2) \underset{c}{\prec} (T|X = x_1), \end{aligned} \tag{6.1}$$

where x_1 and x_2 are two distinct values of the covariate under study, $\underset{c}{\prec}$ denotes convex ordering, and $IHRCC$ ($DHRCC$) are acronyms for "Increasing (Decreasing) Hazard Ratio for Continuous Covariates" (Definition 2.3.1). In Section 4.2 (Bhattacharjee, 2003), we showed that monotone covariate dependence of this type can be naturally represented by monotonic time varying coefficients, so that

$$\begin{aligned} IHRCC \quad &: \quad \lambda(t|x_i) = \lambda_0(t).\exp\left[\beta(t).x_i\right], \beta(t) \uparrow t, \\ DHRCC \quad &: \quad \lambda(t|x_i) = \lambda_0(t).\exp\left[\beta(t).x_i\right], \beta(t) \downarrow t. \end{aligned} \tag{6.2}$$

Thus, the above partial orders (6.1) can be studied using time varying coefficients. In Chapter 4 (Bhattacharjee, 2004a), we used this representation to propose biased bootstrap methods (like data tilting and local adaptive bandwidths) to estimate hazard regression models under these order restrictions. Finally, in Chapter 5 (Bhattacharjee, 2007b), we extended the test for proportionality to a regression model with individual level unobserved heterogeneity with unrestricted frailty distribution.

In this chapter, we consider order restrictions on the shape of the baseline hazard function in addition to constraints on covariate dependence. This kind of ordering is relevant in many applications. For example, relevant theory may suggest that the the effect of a covariate is positive but decreases to zero with age. In addition, the baseline hazard function may be expected to decrease with age.

6.4 Our Bayesian model

As discussed above (Section 6.1), the Bayesian framework offers several advantages, including computational convenience, opportunity to incorporate beliefs into prior distributions, accounting for parameter uncertainty. The major challanges, on the other hand, are (a) appropriate representation of prior beliefs in the model, and (b) ensuring numerical tractability of posterior simulations. Here, we describe how our model specification takes account of the specific empirical features of our application, and addresses the challenges mentioned above.

As discussed earlier, our proposed inference procedures are illustrated by an application to firm exits due to bankruptcy in the UK. In this context, the major objective of our empirical analysis is to understand the effect of macroeconomic conditions on business failure. Age of the firms is measured in years post-listing. The lifetime data are right censored, left truncated and contain staggered entries. Most of the covariates included in the regression model (firm-specific and macroeconomic) are time-varying. In addition, our data includes industry dummies which are fixed over age.

Initially, we consider the Cox proportional hazards model with time varying covariates, fixed regression coefficients and completely unrestricted baseline hazard function (1.4)

$$\lambda \left(t|\underline{X}_i(t)\right) = \lambda_0(t). \exp \left[\underline{\beta}^T . \underline{X}_i(t)\right], \tag{6.3}$$

where $\underline{X}_i(t)$ are a set of (potentially time varying) covariates with proportional covariate effects. We will incorporate into the model additional features of our analysis: (a) *order restricted covariate dependence* – time varying (and possibly monotonic) covariate effects, (b) *unobserved heterogeneity* – fixed effects heterogeneity and frailty, and (c) *order restrictions on ageing*.

To facilitate analysis and presentation, we partition the time axis $[0, \infty)$ into a finite number of disjoint intervals (in our case, in years), say $I_1, I_2, \ldots, I_{g+1}$, where $I_j = [a_{j-1}, a_j)$ for $j = 1, 2, \ldots, g+1$ with $a_0 = 0$ and $a_{g+1} = \infty$. We assume the baseline hazard function to be constant within each of these intervals (taking values $\lambda_1, \lambda_2, \ldots, \lambda_{g+1}$), and the time varying coefficients are also similarly piecewise constant.

6.4.1 Order restricted covariate dependence

Like other applied disciplines, economic theory does not usually imply functional forms or exact distributions, but rather order restrictions such as monotonicity, convexity, homotheticity *etc.* In the context of failure time hazard regression models, there are many applications where there is evidence of order restrictions of the kind described by (6.1) or (6.2) on the nature of covariate dependence.

For example, Metcalf *et al.* (1992) and Card and Olson (1992) observed that the impact of real wage changes varied with duration of strikes, and the variation was in the nature of ordered departures. In particular, Card and Olson (1992) found that, while longer duration strikes (lasting more than 4 weeks) were most common for strikes with wage changes of less than 15 per cent, shorter duration strikes (1 to 3 days) were most frequent for wage changes above 15 per cent. Similarly, Narendranathan and Stewart (1993) found that the effect of unemployment benefits on unemployment durations decreases the closer one gets to the termination of benefits.

Using the current data on firm exits, Bhattacharjee *et al.* (2008a, 2008b) found that the impact of macroeconomic instability on business exit decreases with age of the firm post-listing; these results are presented in further detail in Chapter 7. Further, as discussed in Chapters 1, 3 and 4, such evidence of monotonic covariate effects are not confined to economic applications. For survival with malignant melanoma, for example, Andersen *et al.* (1993) observed that, while conditional hazard rates increase with tumor thickness, the hazard ratios decrease decrease with lifetime.

Based on the above discussion, we allow some covariates in our analysis to have fixed coefficients and some others with time varying coefficients. For some covariates with nonproportional hazards, the time varying coefficients could monotonically increase or decrease with time, accordingly as the covariate effects are $IHRCC$ or $DHRCC$.

6.4.2 Frailty

We account for unobserved covariate effects in two distinct ways. First, there are unobserved covariates at the industry level which create variation in exit rates across industries (other factors remaining constant). Since industry membership is observed for all firms, these factors can be incorporated by including fixed effects heterogeneity. In essence, we include a dummy variable for each industry in our regression model. The estimates for these fixed effects will then be interpreted as the effect of all unobserved regressors at the industry level.

Second, we include scalar multiplicative frailty that is independent of all other covariates. Unlike previous Bayesian studies, the frailty distribution is fully nonparametric in our case. We implement this feature using a method suggested by Heckman and Singer (1984a), where the unknown distribution is approximated by a sequence of multinomial distributions based on progressively increasing number of mass points; see also Chapter 5 (Bhattacharjee, 2007b). For example, with two mass points, log-frailty is assumed to have a two point distribution (say, with mass at $m_1 = 0$ and m_2, and corresponding probabilities π_1 and $\pi_2 = 1 - \pi_1$); one of the mass points is set to zero because of scaling. The number of mass points is increased sequentially until no substantial improvement in the model is observed. At that point, the multinomial distribution approximates the unknown frailty distribution reasonably well.[1]

[1]Modeling frailty distribution in this way offers good opportunities for inference and interpretation. For

6.4.3 Order restrictions on ageing

In addition to covariate dependence, it is often reasonable to expect order restrictions on the shape of the baseline hazard function. For example, in an application based on the current data discussed in Chapter 7 (Bhattacharjee *et al.*, 2008a), we find that the baseline hazard function exhibits some negative ageing. However, this evidence is not in the nature of a decreasing hazard rate, but perhaps a weaker form of partial order, indicating thereby a weak form of learning not related to other observed covariates. This suggests an additional order restriction, perhaps in the nature of a "new worse than used" lifetime distribution. We incorporate such order restrictions in our application to evaluate any evidence on ageing.

Incorporating the above three features in the Cox PH model (6.3), we have the following hazard regression model:

$$\lambda\left(t|\underline{J^{(d)}}_i, \underline{z^{(f)}}_i(t), \underline{z^{(v)}}_i(t), \nu_i\right) = \lambda_0(t).\exp\left[\underline{\beta^{(d)}}^T.\underline{J^{(d)}}_i + \underline{\beta^{(f)}}^T.\underline{z^{(f)}}_i(t) + \underline{\beta^{(v)}}(t)^T.\underline{z^{(v)}}_i(t)\right].\nu_i, \tag{6.4}$$

where $\lambda_0(t)$ is the unknown baseline hazard function which could potentially incorporate order restrictions on ageing, $\underline{J^{(d)}}_i$ is a vector of dummy variables indicating membership in the various industry groups, $\underline{z^{(f)}}_i(t)$ are covariates with proportional effects on the hazard function, $\underline{z^{(v)}}_i(t)$ are covariates with nonproportional effects potentially represented by order restrictions on covariate dependence, and ν_i is a multiplicative frailty variable with arbitrary distribution.

6.4.4 Prior specification

We explore several models with different specifications for the prior distributions. These prior distributions are related to models considered in the literature, for example in Sinha *et al.* (1999). However, our models are unique in that they explicitly consider order restrictions in covariate dependence and ageing, in the presence of individual level multiplicative frailty. Below we describe specification of priors for the three main categories of parameters for our model: covariate effects, baseline hazard and frailty.

example, a two support point distribution with $\pi_1 = 0.25$ would indicate that, with respect to the unobserved covariate, there are two types of subjects. 25% of these subjects draw a lower value from the population and consequently have a lower hazard rate.

Covariate effects

We use three alternative prior distributions for modeling the covariate effects:

1. Truncated normal, with truncation reflecting whether the covariate effect is expected to be positive or negative. For the industry fixed effects, there is no truncation, and the distribution is centered at zero.

2. Truncated normal, with variance proportional to the number at risk (for time varying coefficients).

3. Exponential prior. Like the truncated normal prior above, the shape parameter is proportional to the number at risk (for time varying coefficients).

For the covariates with potentially time varying coefficients, we model order restrictions in three different ways:

1. Initially, no order restriction is imposed, leaving the effects free to assume any value (positive or negative). However, a first order smoothing condition is assumed: $E\left[\beta\left(t_k\right)|\beta\left(t_{k-1}\right)\right] = \beta\left(t_{k-1}\right)$. Further, variance is set at 10 for $\beta\left(t_k\right)$'s up to age 35, and at 1 thereafter. This adjustment is a measure to control for the cumulative uncertainty effect due to the Markov smoothing assumption.

2. Order restrictions in the posterior mean.

3. Stochastic ordering: For example, for decreasing covariate effects, mean is set at a reasonable level initially, decreasing by a step each year. Steps have exponential distributions, with parameter proportional to number at risk.

We make use of the well known consistency property of Bayesian updating procedures that if the prior is supported completely by a subset of the parameter space, then so is the posterior.

Baseline hazard

Four different specifications for the baseline hazard prior are explored.

1. Gamma independent increments.

2. Truncated normal independent increments.

3. Neutral to the right gamma process.

4. Gamma independent increments till age 10, stochastically decreasing thereafter (this reflects a weak form of negative ageing).

Frailty

Our empirical work is based on a two-point support frailty distribution. Since we do not find substantial evidence of individual level frailty, we did not extend the analysis to frailty distributions with higher number of support points.

6.4.5 Model Implementation

We formulate the model in the Bugs language and performed parameter estimation using WinBUGS 1.4 (Spiegelhalter *et al.*, 1999).

6.5 Results and discussion

The data on firm exits due to bankruptcy in the UK, used for our analysis here, pertain to around 4300 listed manufacturing companies over the period 1965 to 2000.[2] The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1965, and contain staggered entries. Age is measured in years post-listig, and all time varying covariatesare measured at an annual frequency. An important focus of the current analysis is the effect of macroeconomic conditions and instability on business failure. Industry dummies are included in the analysis – these are fixed covariates.

To address the issue of staggered entries, we take two distinct approaches, leading to separate analyses. In the first approach, we include covariate information retrospectively for firms that were not covered in the sample for the first few years of their lifetime. Such retrospective data

[2] See Chapters 3, 4 and 5 for previous analyses of these data in the thesis. Further analysis will be reported in Chapter 7, Sections 7.2, 7.3 and 7.4.

on the macroeconomic environment can be collected in this way, but not for the firm level covariate – size, or for industry dummies.[3]

In the second approach, our inference is based solely on the partial likelihood based on an appropriate definition of risk sets, ignoring the past history for the staggered entry firms. This limited information strategy is valid in a wide range of situations with staggered entries (Andersen *et al.*, 1993; Sellke and Siegmund, 1983), even though some standard properties of counting processes do not hold here.

Four measures of macroeconomic conditions and instability are considered: (a) US business cycle (Hodrick-Prescott filter of US output per capita), (b) instability in foreign currency markets (maximum monthly change, year on year for each month, in exchange rates over a year), (c) instability in prices (similar to exchange rates, but measured in terms of RPI inflation), and (d) a measure of business cycle turnaround (measured by the curvature, or second order difference, of the annual Hodrick-Prescott filtered series of UK output per capita). Theory suggests that the effect of the first and the fourth measure on bankruptcy may be negative, and the second and third ones positive. Because of learning effects, the adverse impact of instability is expected to decline in the age of the firm, post-listing. Similarly, the effect of the US business cycle, negative initially, may also rise with age.

A firm level variable – size, measured as logarithm of gross fixed assets in real terms – is also included as a covariate. Industry dummies are used as fixed effects to control for unobserved factors at the industry level.

Next, we report results of the two models under different treatments for staggered entry, as well as different specifications of the prior distribution and order restrictions.

6.5.1 Model using retrospective data

First, we describe our model using retrospective data on a limited set of covariates. For the i-th subject (in this case company), let the corresponding counting process be denoted by $N_i(t)$. We model the process as having increments $dN_i(t)$ in the time interval $[t, t+dt)$ distributed as independent Poisson random variables with means $\Lambda_i(t)dt$.

[3] An alternative approach might be treating the unobserved firm level information as missing at random (MAR) (Little and Rubin, 1987). Adjusting for such missing data is quite convenient in WinBUGS. However, the MAR assumption itself may be rather strong in teh current context.

For computational simplicity we use the conjugacy property of Poisson-Gamma distributions in this context and model the baseline hazard function as a Gamma distributed random variable for each distinct age (measured in years). In our implementation, we model the baseline hazard $\lambda_0(t)$ using a Gamma process prior with unit mean.

Two time varying macroeconomic indicators are included as covariates, namely instability in exchange rates and business cycle turnaround. Note that these indicators are calender time specific, while their effect on a company could potentially depend on the age of the company. Therefore, these two covariates are assumed to have time varying coefficients; we denote the covariates by $Z_e^v(t)$ and $Z_t^v(t)$ respectively.

Because we use retrospective data to account for staggered entry, information on company size and industry dummies cannot be used in this preliminary model. Also, no order restriction on ageing is included in the model.

Annual unbalanced panel data on 4320 listed companies over the period 1965 to 2000 are used for the analysis, accumulating to a total of 45546 company years. The maximum age observed in this data was 50 years. As mentioned above, calender year specific data on exchange rates and US business cycle were included in the analysis.

A total of 166 exits due to bankruptcy (involuntary liquidation) were observed for these 4320 companies. Age at exit ranges form 1 year to 48 years. However, very few exits were observed after the age of 35 years. The lack of failure data on the age range between 35-48 years requires a slightly stronger modeling assumption in order to obtain usable inference.

The distributional assumptions for the likelihood and priors for this model are as follows:

$$\begin{aligned} dN_i(t) &\sim Poisson\left[\Lambda_i(t)dt\right], \\ \Lambda_i(t)dt &= d\Lambda_0(t) \times \exp\left[\beta_e^v(t) \times Z_e^v(t) + \beta_t^v(t) \times Z_t^v(t)\right], \\ d\Lambda_0(t) &\sim Gamma(1,1), \quad \text{for } t = 1, \ldots, 50, \end{aligned} \tag{6.5}$$

where $d\Lambda_0(t) = \Lambda_0(t)dt$ is the increment in the integrated baseline hazard function during the time interval $[t, t+dt)$, with Z's and β's being the corresponding (time varying) covariates and (possibly time varying) regression coefficients.

Economic intuition, and prior empirical evidence, indicates that the effect of the business cycle on bankruptcy hazard is negative while te covariate effect of exchange rate instability is positive. Further, these effects are strong for a newly listed firm but gradually wane off with age; this issue will be discussed in further detail in Chapter 7 (Bhattacharjee *et al.*, 2008a). As mentioned above we will not assume any order restrictions on the covariate effects explicitly, however we would like to infer on the direction of effect and variation of covariate effects with age. This structure is incorporated in the prior distributions as follows:

a) $\beta_e^v(1) \sim Normal(25, 0.1)$ and $\beta_t^v(1) \sim Normal(-25, 0.1)$. Note that the second parameter of normal indicates precision (i.e. inverse variance) and not variance.

b) $\beta_k^v(t) \sim Normal(\beta_k^v(t-1), 0.1)$ where $k = e, t$ and $t = 1, \ldots, 35$.

c) $\beta_k^v(t) \sim Normal(\beta_k^v(t-1), 1)$ where $k = e, t$ and $t = 36, \ldots, 50$. Note that, data for later ages do not contain as much information as earlier ones. The precision is accordingly set at a higher value to adjust for the lack of data and to control the compounding propagation of uncertainty through the first order model.

The posterior distributions for the time varying coefficients and the baseline hazard function offer useful and intuitively appealing interpretation. The baseline hazard estimates do not show any apparent trend. In other words, no substantial ageing is evident in the data, after accounting for covariate effects of exchange rate instability and business cycle turnaround.

However noticeable trend over time is evidenced in the regression coefficients. The posterior estimates strongly reflect time variation on the effect of exchange rate instability (Figure 6-1). There is a strong positive effect on exits when the firm is newly listed, but the effect decreases with age and dies out at about the age of 13 years post-listing.

Similarly, the time varying coefficient of business cycle turnaround is negative initially and rises to zero with age (Figure 6-2).

It is worth noting that these observed trends in the posterior is actually a contribution from the data and not from the prior. In fact, other than setting positive or negative direction for only the initial starting values for regression coefficients of the two covariates no further structural assumptions were made.

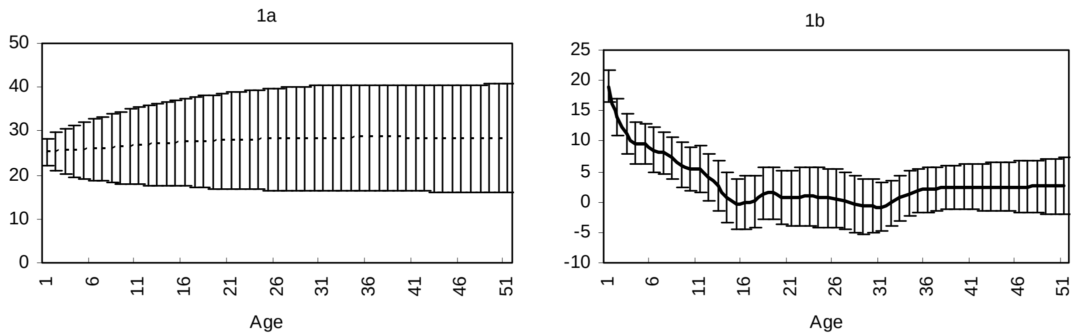


Figure 6-1: Time varying coefficients for exchange rate volatility:(a) Prior (b) Posterior

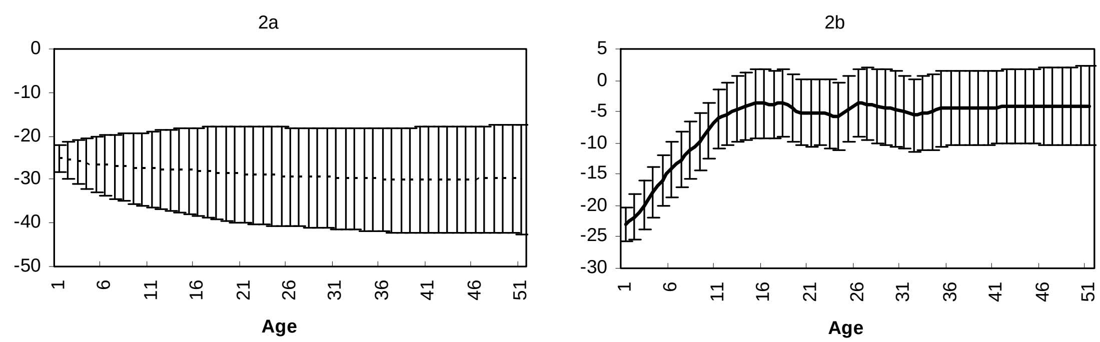


Figure 6-2: Time varying coefficients for business cycle turnaround:(a) Prior (b) Posterior

Therefore the results confirm the economic intuition and prior evidence on order restrictions in covariate dependence. In summary, the model which is rather simplistic nevertheless seems to yield meaningful and useful results.

6.5.2 Model using data with staggered entries

Having experimented with a rather simplistic hazard regression model in the preceding subsection, we now enhance the model in several important ways. First, in addition to macroeconomic factors, we include covariate effect in an important firm level covariate – size (measured by the log of gross fixed assets). Second, we drop business cycle turnaround and include instability in price and the US business cycle as covariates. Third, we include several industry dummies to account for unobserved fixed effects heterogeneity at the industry level. Fourth, and in addition to the above, we include a multiplicative frailty term representing unobserved heterogeneity orthogonal to observed covariates. The frailty distribution is modeled as a two support point multinomial distribution. Fifth, we now measure age in years since inception, rather than years post-listing. This change is motivated partly by the lack of evidence on negative ageing in the baseline hazard function, with age measured in years post listing. The current definition of age is more in line with prior research in empirical industrial organisation, where negative ageing is interpreted as evidence of learning.

Because our model now includes individual level frailty, our dataset needs to be modified to ensure that all included firms contain data for at least two years. We also include two additional years of data on UK listed firms; our data now covers the period 1965 to 2002. Further, as discussed above, we now measure age in years since inception. The data includes 4117 companies with 48176 company years. The maximum age of any company covered in these data is 186 years and maximum exit age is 113 years. The data includes 208 exits due to bankruptcy, of which 203 exits occur by the age of 50 years post listing.

As before we continue to exploit the conjugacy property of Poisson-Gamma distributions and the baseline hazard function is modeled as a Gamma distributed random variable in each year. However the prior distribution for the baseline hazard is adjusted to reflect the availability of information at different ages. This is achieved by allowing the variance to depend on the number at risk at the specified age.

We model the base line hazard $\lambda_0(t)$ using a Gamma process prior, with the parameter depending on the number at risk at each age. The prior distribution is defined as follows:

a) $d\Lambda_0(1) \sim Gamma(1,1)$,

b) $d\Lambda_0(t) \sim Gamma\left[\alpha_1(t), \alpha_2(t)\right]$, for $t = 2, \ldots, 50$ where $\alpha_1(t)$ and $\alpha_2(t)$ such that the mean is $d\Lambda_0(t-1)$ and variance $Y(t)/100$ ($Y(t)$ being the number at risk at age t), and

c) $d\Lambda_0(t) = d\Lambda_0(t-1)$ for $t > 50$.

We implement the hazard regression model with fixed and time varying coefficients, with fixed effects heterogeneity, and with individual level frailty (6.4) as follows:

$$\Lambda_i(t)dt = d\Lambda_0(t) \times exp\left[\begin{array}{c} \sum_{j=1}^{J} \beta_j^{(d)}.J_{ji}^{(d)} + \beta_s^{(f)}.z_{si}^{(f)}(t) + \beta_y^{(f)}.z_{yi}^{(f)}(t) \\ +\beta_e^{(v)}(t).z_{ei}^{(v)}(t) + \beta_\pi^{(v)}(t).z_{\pi i}^{(v)}(t) + \theta_i \end{array}\right] \tag{6.6}$$

The following covariates are included in the model:

1. *Industry dummies*, $J_{ji}^{(d)}$ (J distinct industries, $j = 1, \ldots, J$), are included in the analysis as fixed covariates.with corresponding age constant fixed effects coefficients $\beta_j^{(d)}$,

2. *Covariates with proportional hazards (with coefficients constant over time)*: $z_{si}^{(f)}(t)$ is size of the firm and $z_{yi}^{(f)}(t)$ is a measure of the US business cycle (Hodrick-Prescott filter of output per capita), with corresponding coefficients $\beta_s^{(f)}$ and $\beta_y^{(f)}$,

3. *Covariates with time varying coefficients*: $z_{ei}^{(v)}(t)$ and $z_{\pi i}^{(v)}(t)$ denote exchange rate and price instability, with corresponding nonproportional covariate effects $\beta_e^{(v)}(t)$ and $\beta_\pi^{(v)}(t)$ respectively (the covariate effects are expected to be positive initially and decreasing with age), and

4. $\nu_i = exp(\theta_i)$ is an individual level multiplicative frailty term with a two point support distribution.

The prior distribution for log-frailty (θ_i) is modeled as having two support points $m_1 = 0$ and m_2, with corresponding probabilities p_1 and $p_2 = 1 - p_1$; m_1 is fixed at zero because of scaling. We assume a standard normal distribution for the prior of m_2. The population assignment of a company is then given by a latent variable, here assumed to have a multinomial distribution with a Dirichlet prior for the probability p_1. Our implementation, which is similar to Campolieti (2001), has two major advantages. First, it exploits the Multinomial-Dirichlet conjugacy property which helps in computations. Second, the model is easily extendible to a larger number of support points for the frailty distribution.

Standard normal priors were considered for the industry fixed effects. For the time constant coefficients, nearly half normal distributions were considered as priors, with a slight shift from zero:

$$\beta_s^{(f)}, \beta_y^{(f)} \sim Normal(-0.01, 10) \text{ truncated on } (-\infty, 0).$$

For the time varying coefficients decreasing with age, Gamma distributed increments were taken away from the coefficient at the previous age to maintain monotonicity in the prior distributions:

a) $\beta_k^{(v)}(1) \sim Normal(0.25, 1),\ k = e, \pi;$

b) For $t \epsilon (2, 50)$, $\beta_k^{(v)}(t) = \beta_k^{(v)}(t-1) - \left[b_k^0(t-1) \times \frac{Y(t)}{c}\right]$, where $b_k^0(t-1) \sim Gamma(0.01, 1)$, $Y(t)$ is the number at risk at age t, and c is the maximum number at risk at any age in the data.

c) For $t > 50$ $\beta_k^{(v)}(t) = \beta_k^{(v)}(t-1)$

The posterior estimates for the baseline hazard function (Figure 6-3a) do not show any obvious evidence of ageing. This is a bit surprising since earlier work has found evidence of negative ageing. This observation, however, does not seem to be feature of the current data. In fact, estimates of the baseline hazard function based on the partial likelihood estimates also show a very similar time varying pattern to the posterior mean (Figure 6-3a).

The time varying coefficients for exchange rate and price instability (Figures 6-4 and 6-5 respectively) indicate strong evidence of non-proportionality. The coefficients are positive when the firm is newly listed, but decline to zero as the firm gets older.

The usefulness of our model of unobserved heterogeneity, in terms of fixed effects heterogeneity at the industry level combined with individual level frailty with distribution on a finite

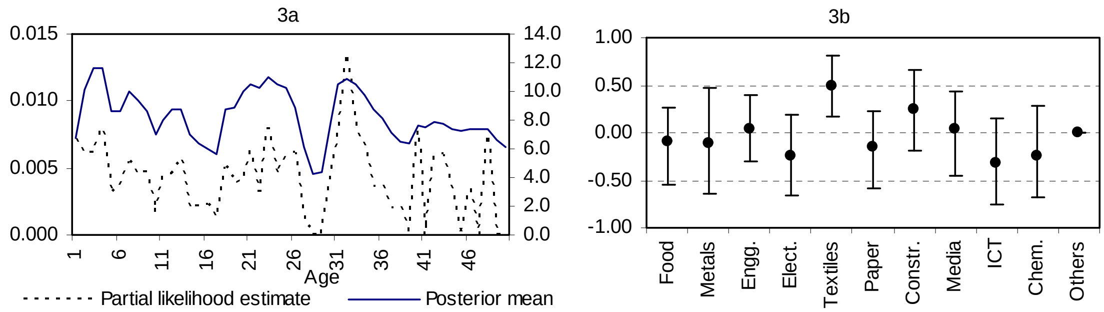


Figure 6-3: Posterior Estimates: (3a) Baseline hazard, (3b) Industry fixed effects (with 95% posterior intervals)

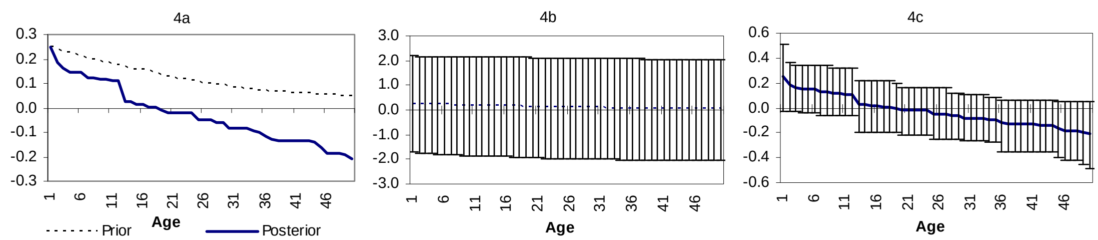


Figure 6-4: Time varying coefficients for exchange rate volatility:
(a) Prior and posterior mean (b) prior mean and 95% interval (c) posterior mean and 95% interval

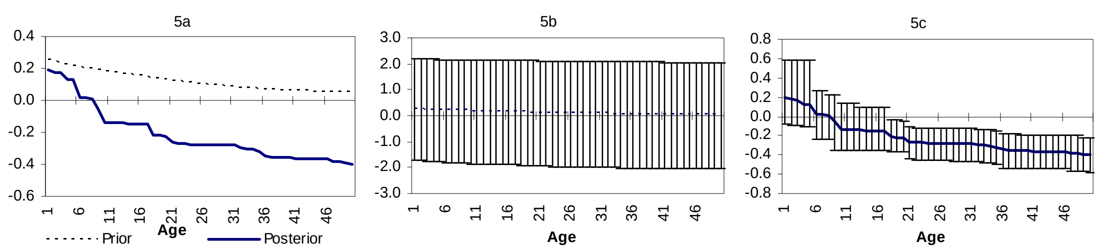


Figure 6-5: Time varying coefficients for price instability:
(a) Prior and posterior mean (b) prior mean and 95% interval (c) posterior mean and 95% interval

number of support points, is emphasized by the empirical results. The posterior distributions of the industry level fixed effects demonstrate evidence of substantial unobserved heterogeneity (Figure 6-3b). Other factors being equal, high technology industries such as "ICT" and "Electronics and Electricals" have a lower hazard rate of exit due to bankruptcy, while the "Textiles" industry attracts a substantially higher hazard. This is in reasonable agreement with economic intuition and prior empirical evidence.

At the same time, we do not find evidence of multiplicative frailty at the level of the individual firm. In fact, the posterior distribution of frailty converges to a single mass point. From an economic point of view, this evidence is not surprising, because unobserved human capital may be rather homogeneous in a sample of successful listed firms.

In summary, we find strong support for the order restrictions on covariate dependence, but not much evidence of expected shape in the baseline hazard function. We also find that the models and priors developed here are useful for inference on order restricted covariate dependence and ageing, as well as on the effect of unobserved heterogeneity.

6.6 Conclusion

Research on order restricted Bayesian inference in the context of hazard regression models has been rather limited. In this chapter, we make contributions to this literature by proposing a Bayesian framework for order restricted inference in hazard regression models in the presence of unrestricted univariate frailty. We consider constraints on covariate dependence; these constraints are in the nature of convex (concave) ordering of lifetime distributions conditional on distinct covariate values. Our proposed methods are very useful in understanding covariate dependence in situations where the proportional hazards assumption does not hold.

In addition to covariate dependence, we also discuss order restrictions on the shape of the baseline hazard function. These order restrictions inform about ageing properties of the lifetime distributions, holding observed covariates and frailty constant.

Our methodology pays special attention to the modeling of frailty. In addition to fixed effects unobserved heterogeneity, we model individual level frailty nonparametrically using a sequence of discrete mixture of multinomial distributions with increasing number of mass points

(Heckman and Singer, 1984a). This is in sharp contrast to much of the existing literature where frailties are assumed to have parametric distributions that do not offer additional insights.

The analysis of corporate failure data using our methodology offers interesting new evidence on the nature of covariate dependence. In particular, we find that the macroeconomic environment has a strong effect on the hazard rate of firm exits due to bankruptcy. Further, the effect of adverse economic conditions which is quite drastic on young firms decreases to zero as the firm gains in experience. However, in our application, we do not find much evidence on ageing characteristics in the baseline hazard function. Further, while we observe substantial fixed effects unobserved heterogeneity at the industry level, evidence points to absence of significant multiplicative frailty at the level of the individual firm.

In the context of the thesis, the current work extends research reported in Chapters 4 and 5 in several ways. In Chapter 4 (Bhattacharjee, 2004a), we proposed estimation of hazard regression models under monotone time varying covariate effects, while tests proposed in Chapter 5 (Bhattacharjee, 2007b) extended tests for proportional hazards against order restrictions on covariate effects (Chapter 3, Bhattacharjee, 2007a) to the case when there is arbitrary univariate frailty. Here, we extend the above work by developing Bayesian inference on order restricted covariate effects in the presence of unrestricted frailty, but also potentially order restricted ageing patterns. The methods developed here are simple to use, and offer inferences on both the strength of order restrictions, and estimates under such assumed shape constraints. Importantly, we consider order restrictions on both covariate dependence and duration dependence (ageing), in the presence of unrestricted frailty.

The work reported in this chapter raise several important research questions. It has been noted in the literature that the presence of frailty can often be confused with nonproportional covariate effects (Elbers and Ridder, 1982; Andersen *et al.*, 1993; Aalen, 1994). This issue is further emphasized by the application considered here, in that, when the macroeconomic variables are allowed to have order restricted covariate effects, no evidence for frailty is found. This observation has two important implications for further research. First, estimation of the unknown frailty distribution has to be carefully conditioned on the possibility that nonproportional covariate effects may be present. This is an approach that we explore further in Section 7.4 (Bhattacharjee, 2007c). Second, it implies that the approach proposed in Abbring and van den Berg (2007), of using hazard ratios to test for the significance of frailty, is invalidated in the presence of nonproportional covariate effects. Developing appropriate inference procedures in this situation will be an useful direction of research. Third, there is substantial current research on order restricted Bayesian inference that aims to place the approaches under an appropriate theoretical setting. In particular, recent research in the area, both in the context of stochastic ordering (Dunson and Peddada, 2007; Wang and Dunson, 2007) and trend ordering in hazard rates (Gunn and Dunson, 2007), are developing methods of Bayesian inference for testing of partial ordering across groups, and estimation under partial orders. Extending the methods developed here, in terms of convex ordering and order restrictions on ageing, to the above framework will be a challenging and useful direction of future research.

Chapter 7

Applications to Firm Dynamics

In this Chapter, we use the concepts and methods developed in the thesis to understand the dynamic nature of a firm's survival rate. The empirical framework draws on the theoretical economic literature on firm dynamics, and places it within the context of hazard regression models under order restrictions.

While the theory of survival analysis provides a multitude of analytical tools, their application to the analysis of economic duration data is somewhat limited. The main reason is that most economic theoretical models do not make verifiable predictions about the distributions of duration variables (van den Berg, 2001). Notable exceptions are economic models of unemployment and strike durations. Several theoretical models of search for jobs by individual agents in the labour market have been developed in the labour economics literature; see van den Berg (2001) for further discussion. These job search models have been very popular as explanatory theoretical frameworks for reduced-form econometric duration analyses of unemployment spells (Devine and Kiefer, 1991; van den Berg, 2001). Similarly, Kennan and Wilson (1989) describe bargaining models for strike durations. While the hazard regression models implied by the above economic theories do not postulate nonproportional effects of the covariates, there are shape restrictions on the baseline hazard function[1] and omitted covariate represented by univariate frailties.

Another promising area of application is the theory of firm dynamics, which gives rise

[1] The above models for unemployment durations imply a constant baseline hazard rate, while the baseline hazard for strike durations can be U-shaped.

to several testable predictions about the age at which different firms exit, through failure or censoring, from a cohort of new firms that entered the population at the same time. The primary object of our empirical analysis in this chapter is to understand how the hazard rate of firm exits depends on the level of efficiency of the firm and on exogenous macroeconomic shocks, as well as appropriate modeling of frailty in such applications. Theoretical research facilitating the empirical study of such hazard regression models has been advanced in work on learning models; see, for example, Jovanovic (1982), Lambson (1991), Hopenhayn (1992), Ericson and Pakes (1995), Pakes and Ericson (1998), and Bergin and Bernhardt (2004). However, a major challenge in this line of research is that, unlike job search mdels, theoretical models of firm dynamics do not predict proportional hazard functions for different values of the covariates. This chapter presents new empirical evidence on exits of firms, through competing exit routes of bankruptcies (liquidations) and acquisitions. The theoretical framework is based on models of firm dynamics that typically imply nonproportional hazards of firm exits and an important role for frailty, and empirical inference is obtained using statistical methods developed in this thesis.

The chapter is organised as follows. We briefly review existing and new theoretical economic models in Section 7.1, focusing mainly on the effect of macroeconomic conditions on business exit (Bhattacharjee *et al.*, 2008a) and on the role of unobserved human capital of entrepreneurs (Bhattacharjee *et al.*, 2006). This is followed by analysis of firm exits in the UK through competing routes of liquidation (bankruptcies) and acquisitions (Section 7.2, Bhattacharjee *et al.*, 2008a) and on comparative analysis for US firms (Section 7.3, Bhattacharjee *et al.*, 2008b). The work highlights the important impact of macroeconomic instability on business failures in the UK, while differences in covariate effects in the US can be partly attributed to legislative reforms in the late 1970s. The empirical work in Sections 7.2 and 7.3 incorporate shared (proportional) frailties between the competing risks. In Section 7.4, we consider the UK data on firm exits in combination with French data on new firms to ask the question as to how far the issues of frailty and nonproportional covariate effects are related. While the analyses point to important frailty effects on the UK data in addition to order restrictions in covariate effects, the French data are better explained by a model incorporating nonproportional hazards only. Thus, the three applications points to several important aspects of modeling order restricted covariate effects using nonproportional hazard regression models proposed in the thesis, as well as the importance of modeling frailty appropriately. Notably, economic inferences drawn form our analyses have important policy implications.

7.1 Firm dynamics and the hazard rate of firm exits

7.1.1 Active and Passive Learning

The literature on industrial organisation proposes several theoretical models of the dynamics of firm behaviour that incorporate heterogeneity among firms, different sources of uncertainty (either firm-specific or idiosyncratic) and exit/ entry outcomes[2]. Two of these models are popular: the "passive learning" model (Jovanovic, 1982; Lippman and Rumelt, 1982) and the "active learning" (also known as active exploration) model (Ericson and Pakes, 1995; Pakes and Ericson, 1998). Consistent with evidence that firms make their entry investments unsure of their success, both these models assume that new firms that make their entries are uncertain of their quality and use "noisy" cost and profit signals to learn about their true efficiency or productivity levels. However, while the passive learning model assumes that the state variable (efficiency) remains constant over the lifetime of the firm, the active learning model allows firms to change the level of their stochastic state variable through potentially quality-enhancing investments. There are many other theoretical contributions to this literature, including models proposed by Lambson (1991), Hopenhayn (1992), Cooley and Quadrini (2001) and Asplund and Nocke (2003). The empirical implications of these models are similar to each other in some respects, and different in others. Studying these empirical implications is important for understanding the nature of firm survival in different industries, as well as their market structure, attrition, and response to possible changes in policy or other environmental conditions.

The starting point of the current literature on stochastic dynamic industry equilibria with heterogenous firms is the seminal paper by Jovanovic (1982). In this model of "passive learning" (see also Lippman and Rumelt, 1982; Hopenhayn, 1992; and Cabral, 1993), the potential entrant into a perfectly competitive industry with heterogeneous but time-invariant efficiency

[2] Caves (1998) provides an extensive survey of the theoretical and empirical literature on turnover and mobility of firms.

levels is assumed to know the distribution of the state variable across all firms, but not its own realisation. Upon paying a (nonrecoverable) entry fee, it starts to receive noisy information on its true efficiency. Firms which learn that they are efficient grow and survive, while firms that obtain consistently negative signals decline and eventually leave the market. The model produces a rich array of empirical predictions on the relationship between firm growth and survival on the one hand and firm age and size on the other. However, under passive learning all firms eventually learn their efficiency level, and so there is no firm turnover in the long run.

By contrast, in "active learning" models such as Ericson and Pakes (1995) (see also Pakes and Ericson, 1998), entrants invest in uncertain but expectedly profitable innovations or cost reductions. Here, firms entering an industry have efficiency varying over time due to stochastic market changes, their own investment decisions and those of other market participants. The firm grows if successful, shrinks or exits if unsuccessful. Thus, the passive learning model by Jovanovic (1982) differs from the active learning model in that the stochastic process generating the size of a firm is non-ergodic. Pakes and Ericson (1998) use this difference to develop empirical tests to distinguish between the two models.

Hopenhayn (1992) considers a perfectly competitive industry. The main prediction of his model is that firm turnover is negatively related to entry costs. Due to the absence of the price competition effect, however, market size has no effect on entry and exit rates. An extension of the model to an imperfectly competitive market with monopolistic competition is considered in Asplund and Nocke (2003); the model generates implications of sunk costs and market size on firm exits and the size distribution of surviving firms. Bergin and Bernhardt (2006) consider business cycle effects in a similar model of perfect competition.

Lambson (1991) considers a model with atomistic price takers, where there are no idiosyncratic shocks but instead common shocks to input price (and demand). In equilibrium, firms may choose different technologies and hence be affected differently by the common shocks. The model predicts that the variability of firm values is negatively related to the level of sunk costs.

The large literature on empirical industrial organisation has collected several regularities regarding firm efficiency, size, growth and exit rates. Following Pakes and Ericson (1998), we outline some of the most important stylised relationships (R1, R2, R3a, R3b, R4a and R4b) and relate these to theoretical models of firm dynamics.

$R1$ Conditional on age, the hazard rate is decreasing in current size.

$R2$ The size distribution of the firms that survive from a cohort of firms increases, in a stochastic dominance sense, with age.

$R3$ Hazard rate (unconditional, and conditional on size):

(a) The hazard rate is decreasing in age conditional on size (current size and/or initial size). Sometimes, the hazard rate increases with age initially, but decreases at older ages.

(b) The unconditional hazard rate may also decrease with age, at least at older ages.

$R4$ Effect of initial size:

(a) The initial value of the state variable may also matter; hazard rate may decrease in initial size (proxy for efficiency).

(b) The effect of initial size may persist even at an older age.

Pakes and Ericson (1998) show that the first two relationships ($R1$ and $R2$) hold both for the passive and the active learning models.

The third relationship ($R3a$ and $R3b$) implies that younger firms experience higher hazard rates, and that the hazard rate declines with age. Empirical studies have shown consistent evidence of declining hazard rates at higher ages, though the hazard rate for entrants sometimes increases with age. This is true for both unconditional hazard rates and hazard rates conditioned on initial size. Dunne *et al.* (1989) and related studies have advanced the view that a monotonically decreasing hazard function provides evidence in favour of the passive learning model. However, Pakes and Ericson (1998) show that the passive learning model does not necessarily predict hazard rates falling from the outset. They could rise at first, if ill-fated firms need some experience to be sure of their low efficiency. Thus, a monotonically decreasing hazard rate, and hazard rate that increases upto a certain age and then decreases, may be consistent with both active and passive learning models.

The fourth relationship ($R4a$ and $R4b$) is the most useful for distinguishing between the active and passive learning models. Both $R4a$ and $R4b$ hold for the passive learning model, since this model does not allow the firm an opportunity to change its efficiency through investment. However, $R4b$ does not hold for the active learning model; the efficiency of a firm depends on investment and evolves over time, with the result that the relationship between the hazard rate of exit and initial size diminishes with age, and finally dies out. Pakes and Ericson (1998) use this difference to construct empirical tests to distinguish between the active and passive learning models.

Other models of firm dynamics are also consistent with some of these stylised relationships (Pakes and Ericson, 1998). The model of Hopenhayn (1992) satisfies $R1$ but not $R4$, while Lambson's (1991) model satisfies the $R4$, and not $R1$. This, in a sense, reinforces the view of Caves (1998) that tests of persistence of the impact of initial size do not necessarily validate specific theoretical models, such as the passive or the active learning model.

In the context of the hazard regression model with time varying coefficients, the above observation implies that the covariate effect of initial size is negative in both the passive and active learning models. While in the active learning model, this effect drops to zero with age of the firm, the effect remains persistent in the passive learning model. Further, the above economic models imply negative ageing in the shape of the baseline hazard function. As discussed above, this negative ageing may not always be in the nature of a decreasing hazard rate (Pakes and Ericson, 1998), but can perhaps be described by a weaker ageing class (like DFRA, NWU, etc.).

7.1.2 Impact of macroeconomic shocks

This section presents an economic framework, in the spirit of Jovanovic and Rousseau (2001, 2002), for analysing the manner in which the competing risks of bankruptcy and acquisition are influenced by macroeconomic conditions, specifically, macroeconomic instability. We use the terms macroeconomic instability and macroeconomic uncertainty interchangibly.

An important innovation in the theoretical model developed here lies in explicitly incorporating macroeconomic effects within a model of firm exits through bankruptcy and acquisitions. The existing theoretical and empirical literature has pointed out several ways in which firm performance, including exits, are related to changes in the macroeconomic environment (see, for example, Wadhwani, 1986; Cuthbertson and Hudson, 1996; Higson *et al.*, 2002; and Delli Gatti *et al.*, 2001). These macroeconomic conditions can be characterised both by the aggregate level of economic activity (broadly speaking, economic expansions and downturn) as well as the degree of instability in the macroeconomic environment.

While firm exits through compulsory liquidation increase during periods of severe downturn in the aggregate economy (see, for example, Caballero and Hammour, 1994), merger waves are broadly procyclical (Shleifer and Vishny, 2003). On the other hand, instability potentially affects firm exits in two ways. First, lenders are less willing to lend when there is higher instability (Greenwald and Stiglitz, 1990); this channel increases credit constraints on firms and leads some firms to bankruptcy. Second, uncertainty can induce growing firms to delay their decisions to invest (Dixit, 1989).

Further, the effect of uncertainty on business performance may be asymmetric, particularly in the presence of credit constraints (Bernanke and Gertler, 1989; Kiyotaki and Moore, 1997). This asymmetry has important implications for the way instability is measured in our empirical work discussed later. Our economic model, which takes explicit account of the above macroeconomic effects on firm exits, is described below.

At any time, t, each firm, i, is at some risk of exit through bankruptcy or by being acquired. Figure 7-1 gives a schematic representation of the way macroeconomic conditions affect exit risks. On one side are firms that exit due to financial distress (through bankruptcy or by being acquired), or choose to exit even though they are not distressed. Adverse macroeconomic conditions increase the pool of firms in financial distress. On the other side are investor firms who are in the market for acquired capital. Any firm that is not distressed chooses an optimal level of investment I_{it}, conditional on the level and stability of the macroeconomy. This optimal investment, which maximises the expected present value of the firms' future cash flows, typically comprises both investment in new capital, X, and acquired capital, Y. The balance between X and Y depends on the relative prices of acquired and new capital, as well as on the fixed and adjustment cost of acquisitions.

Let the i-th firm's state of technology (or efficiency) at time t be denoted by z_{it} and its capital by K_{it}. Firms operate under an AK type production function which takes the form $f(z)K$. Here $f(z)$ is akin to the output-capital ratio and depends on firm efficiency: $\partial f(z)/\partial z > 0$. We assume that the dynamics in z and the economy wide macro-environment variable of interest, u (denoting instability) are each governed by Markov transition processes. z and u are each assumed to be positively autocorrelated, and independent of each other. Hence,

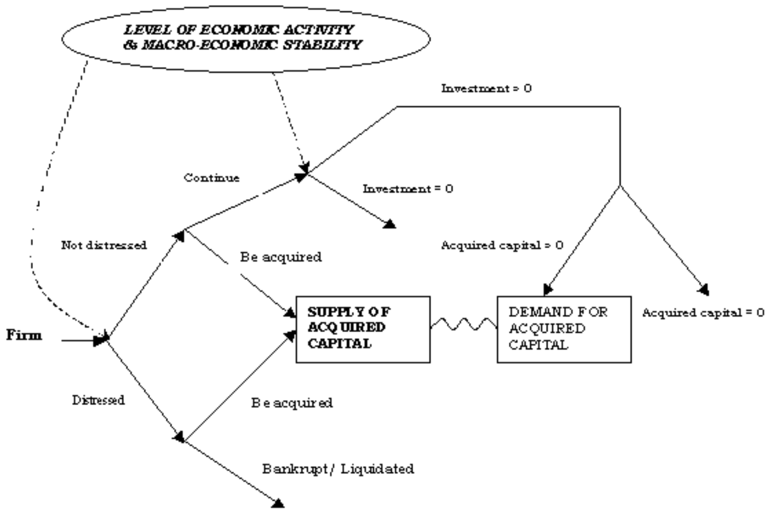


Figure 7-1: Macroeconomic conditions and firm exits

z and u are jointly Markov. We denote

$$Pr[z_{i,t+1} \leq z', u_{t+1} \leq u' | z_{it} = z, u_t = u] = F(z', u' | z, u).$$

Profits can then be written as $[f(z) - C(x, y) - x - qy - g(u)]K$, where x and y are the (per unit capital) investments in new and acquired capital respectively ($i = x + y$), $C(x, y)$ is the (per unit capital) adjustment cost of investment, and $g(u)$ is the firm specific impact of macro-environment on profits. $g(u)$ is increasing and convex in u, and $g(0) = 0$. The price of new capital is normalised to unity, and q denotes the price of acquired capital ($q < 1$). Then, the market value of the firm per unit of capital under the optimal investment plan is:

$$Q(z, u) = max_{x \geq 0, y \geq 0}\{f(z) - C(x, y) - x - qy - g(u) + (1 - \delta + x + y)Q'(z, u)\},$$

where

$$Q'(z, u) = \frac{1}{1 + r} \int \int max\{q, Q(z', u')dF(z', u' | z, u)\}$$

is the expected present value of capital in the next period given the firm's z and the economy's u today, and δ is the rate of depreciation. Since z and u are independent and positively autocorrelated, $Q(z, u)$ is increasing in z and decreasing in u. Denote by $z_e(u)$ the level of z

at which the firm is indifferent between exiting and staying in business, given macroeconomic conditions, and by $z^*(u)$, the level of technology at which the firm is indifferent between staying out of the acquisitions market or entering it.[3]

In a period of economic stability, when demand is more predictable, the incidence of financial distress will be lower. The smaller pool of distressed firms may also face a larger number of potential acquirers whose investment policies are encouraged by macroeconomic stability. Thus firms that are on the verge of bankruptcy will have a higher probability of being rescued, and the observed bankruptcy rates can be expected to be lower. Further, in such periods the hazard of acquisitions will be higher, even though there are fewer distressed firms that are candidates for acquisition. With the boost to investment in more stable periods, the market for acquisitions may tighten, driving up the market price of acquired assets. This can be expected to induce a larger number of non-distressed firms to enter the pool of potential acquirees. These would be firms who find the offers from potential acquirers to be higher than their own continuation values (Jovanovic and Rosseau, 2001).

The implications of changes in u for firm exits and acquisitions can be understood with reference to a plot of the four regions of z, (Jovanovic and Rousseau, 2002). Let $\overline{u} > u$, then $z_e(\overline{u}) > z_e(u)$ and $z^*(\overline{u}) > z^*(u)$ (Figure 2). In a period with higher u, a larger number of firms decide to exit, and fewer firms decide to acquire. Hence, a larger number of firms are likely to go bankrupt during such periods.

Overall, in a period of economic stability, the propensity for bankruptcy will be lower, and the propensity for acquisitions will be higher. A testable implication of the model is that the impacts of macroeconomic instability on the likelihood of bankruptcy and acquisition are of opposite signs.

Some other economic models have studied the effect of macroeconomic environment on exit decisions of firms. Campbell (1998) studies a general equilibrium model of industry dynamics, where aggregate uncertainty through innovations drives the correlation between current exit and future output growth. In his model, future anticipated technical innovations lead to more firm exit in earlier periods because consumers respond by increasing savings and reducing

[3] Assuming a fixed deadweight cost of investing in acquired capital ensures existence of a threshold level z^*, above which a firm invests in acquired capital, and below which it does not (Jovanovic and Rousseau, 2002).

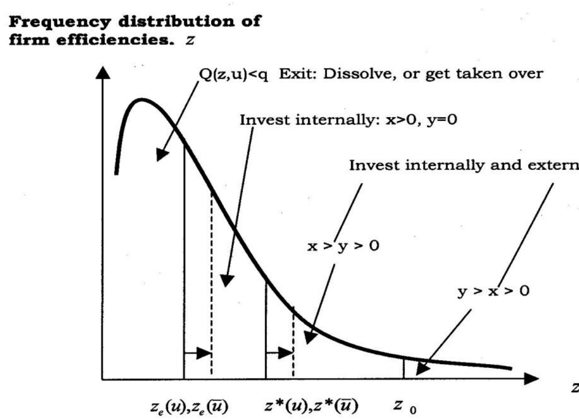


Figure 7-2: The four regions of z (modified from Jovanovic and Rousseau, 2002)

current consumption. Delli Gatti *et al.* (2001) develop a theoretical model linking the macroeconomic environment, financial fragility and the entry and exit of firms. In a setting where all prices are constant, Cooley and Quadrini (2001) explores the impact of interest rate shocks on the entry and exit of firms and aggregate output. Bergin and Bernhardt (2006) study the dynamics of an industry subject to aggregate demand shocks where the productivity of a firm's technology evolves stochastically over time. Their model offers an alternative explanation to the relationship between exit rates and expected economic growth.

Compared to the above, our model draws into sharper focus the effect of economic instability and explicitly considers exits through alternative competing routes. In the following Sections 7.2 and 7.3, we use this model as the theoretical framework for empirically studying the impact of macroeconomic conditions and instability on business failure of quoted firms in the UK and the US, through competing routes of bankruptcy and acquisitions.

7.1.3 Unobserved heterogeneity

The empirical literature on firm dynamics has generally acknowledged the importance of frailty in understanding firm exits. For example, in the US shipbuilding industry, Thompson (2005) finds an important role for unobserved heterogeneity related to variation in initial experience. However, the precise role and effect of frailty has not been adequately addressed in most economic models of firm dynamics.

An exception is Bhattacharjee *et al.* (2006), who study the role of unobserved human capital in entrepreneurial choice and its impact on the survival of newly created firms. Here, we outline the main features of their model, omitting details of the theoretical framework. Bhattacharjee *et al.* (2006) consider a setting where, when starting a new business, an entrepreneur's labour market situation (e.g. employed or not) reflects how her human capital may be valued through salaried labour. To the extent that this valuation affects the entrepreneurial decision, an entrepreneur's human capital is correlated with the state at which she decided to start a new firm. Their model of entrepreneurial choice provides predictions about an entrepreneur's actual human capital as a function of human capital observed by the econometrician as well as the individual's state in the labour market when the firm was created. The model allows for information asymmetry on the labour market as well as other sources of inefficiencies such as incentive problems due to moral hazard. It also allows for dynamic considerations on the part of the entrepreneur regarding potential depreciation of her human capital. In a situation where employer's information on employee's human capital is sufficiently poor and where there is a strong concern about human capital depreciation for those with a high level of observed human capital, the model predicts an important role of unobserved human capital on the survival of firms.

In Section 7.4, we use the above model as a framework for empirical study of the role of frailty in survival of new firms created by French entrepreneurs. The evidence is contrasted with firm exits among quoted firms in the UK to discuss the relative importance of frailty and nonproportional covariate effects in empirical studies on firm survival.

7.2 Macroeconomic conditions and business exit: determinants of failures and acquisitions of UK firms

Based on Bhattacharjee *et al.* (2008a), here we empirically study the impact of macroeconomic conditions on business exit in a setup where acquisition and bankruptcy are co-determined. We use data on UK quoted firms over a thirty-seven year period, spanning several business cycles, and estimate a competing-risks hazard regression model to determine how processes that determine bankruptcies and acquisitions depend on the macroeconomic environment, conditional on the age of the firm since listing and other firm- and industry-specific factors. We show that adverse macroeconomic conditions both increase bankruptcy hazard while at the same time decreasing acquisition hazard. Moreover, the US business cycle is a better predictor of UK acquisitions and bankruptcies than the UK cycle itself; bankruptcies are counter-cyclical and acquisitions are pro-cyclical. Importantly, macroeconomic instability has time varying coefficients, with the advrse covariate effects decreasing to zero as the age of the firm increases. The baseline hazard function shows evidence of weak negative ageing of the NWU type. The empirical results are robust to model specification in several ways, including the effect of potentially dependent left truncation.

7.2.1 Data

The evaluation of the impact of macroeconomic fluctuations on business exits requires data running over several business cycles. We use a database of firm quoted in the UK, constructed by combining the Cambridge-DTI, DATASTREAM and EXSTAT databases of firm accounts. The combined firm level accounting data provides an unbalanced panel of about 4,100 UK listed companies over the period 1965 to 2002. There were 206 instances of bankruptcy and 1858 acquisitions among 48,046 firm years over the 38 year period.[4] In terms of hazard model analysis, the data are right-censored and left-truncated.[5]

We use the term 'bankruptcy' to denote the event of compulsory liquidation. We use the term 'acquisition' to denote the event of business combination, which may take the form of a merger, an acquisition or a takeover. Interchangeable use of these words is standard in this literature.[6]

[4] A firm that has irretrievably entered the path to bankruptcy may, in a precursor phase of distress, stop publishing accounts one or two years prior to actually being declared bankrupt. From the point of view of econometrically modelling bankruptcy it is sensible to reassign the date of "real" bankruptcy to the year of last published accounts when the firm has been declared legally bankrupt within a 2 year period. Our assignment of a bankruptcy to a particular point in time captures the date of economic bankruptcy rather than declaration of bankruptcy. We assign accounting data for each company fiscal year to the calendar year that covers the majority of the accounting year corresponding to the fiscal year.

[5] The data used pertain to years, since 1965, during which each company is listed in the London Stock Exchange. Hence, for each company, the available data are left-truncated, and do not pertain to the entire period that it is listed.

[6] It is somewhat rare for a business combination to be a 'merger of equals'. These are, in practice, effectively unobservable to the extent that even case-based contextual research struggles to identify them. 'Merger of equals' is not proxied by other apparently related constructs sometimes used in the literature, such as 'friendly/hostile' or 'equity/cash consideration' – nor is it proxied by the use of pooling (merger) rather than purchase accounting

Measures of macroeconomic conditions

We use the following empirical proxies for macroeconomic conditions:

- As a measure of the business cycle (BC_t), we use a quarterly Hodrick-Prescott-filtered series of UK output per capita ($\lambda = 100$).

  Given the strong trading linkages of the UK industrial sector with the global economy, and particularly with the US economy, it is likely that the global economic environment will affect the exit decisions and outcomes for UK firms.

- We allow for the possible impact of the global economy by including a similar measure of the US business cycle.

- Real interest rates are measured as the yield on 20-year sovereign bonds, less the annual rate of inflation.

- The average annual real effective exchange rate is used to measure the exchange rate environment. Goudie and Meeks (1991) have found that a stronger pound sterling raises the propensity of firms to go bankrupt.

Figures 7-3 and 7-4 plot the annual incidence of bankruptcies and of acquisitions, as well as the business cycle indicator for the year. Incidence is measured as the ratio of the number of companies that went bankrupt (or the number that were acquired) during the year, to the total number of listed companies. Quoted firm bankruptcies were particularly high during years when the economy turned down after a peak, and were lower during upturns in the business cycle (Figure 7-3). The growth rates in firm registration hint at a plausible mechanism for this; entries are pro-cyclical and it is possible that the larger number of entries during the upturn of the business cycle force some firms out of business when the economy turns down.

Figure 7-4 indicates that acquisitions were procyclical. Research on aggregate mergers and acquisitions activity has found aggregate market capitalisation to be a determinant of

for the transaction.

In our data, firm B was considered to have exited the industry if it was acquired by firm A. If, at the same time, firm A changed its name to C, we treated A as remaining in in the industry.

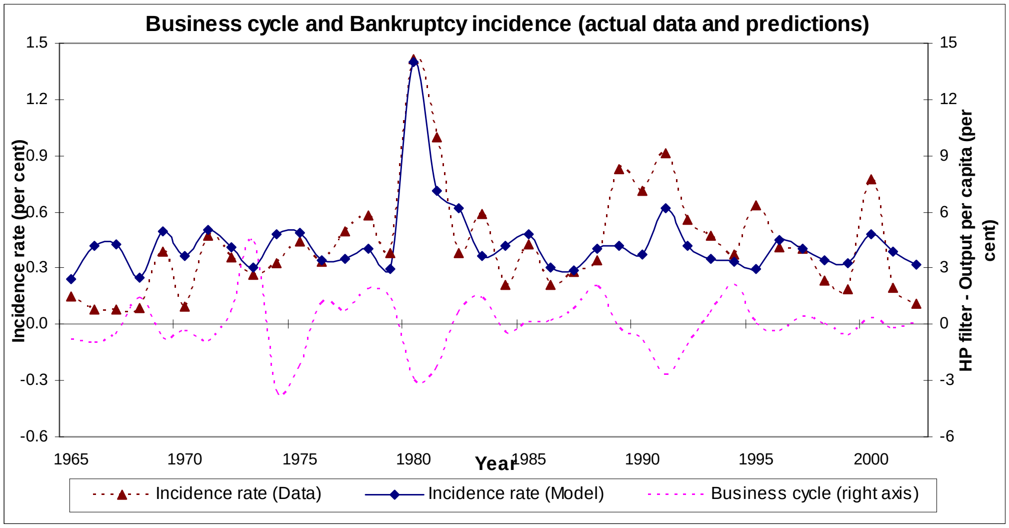


Figure 7-3:

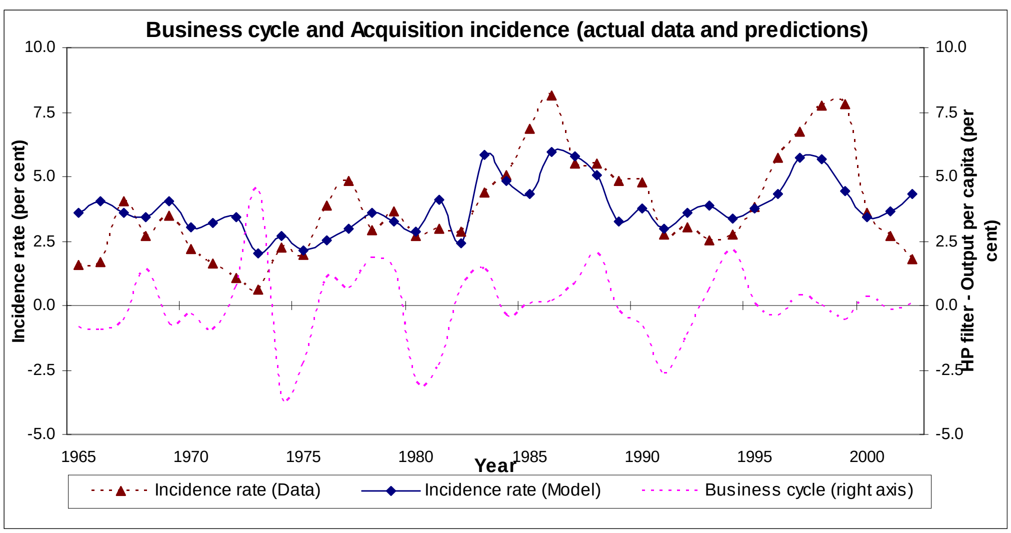


Figure 7-4:

acquisition demand. Similarly, earlier research on firm exits have found explanatory power in other measures of aggregate economic activity. We experimented with several other measures, such as Tobin's q (see also Chapter 5), industrial production, stance of monetary policy and capacity utilisation, and found the substantive conclusions of our estimated models to be robust to variable selection. Our final choice of macroeconomic variables was guided by availability of consistent data over the 38-year period, as well as by statistical variable selection methods.

Measures of macroeconomic stability

Figures 7-3 and 7-4 also suggest that the incidence of bankruptcy and acquisition vary substantially over time. While a part of the aggregate movement may be explained by the business cycle, macroeconomic stability may also have a role to play. It has been argued that the impact of uncertainty on business performance is essentially asymmetric. For example, in economies with credit constraints, credit imperfections generate a transmission mechanism through which a small, temporary shock can generate large, persistent and asymmetric domestic balance sheet effects.[7]

Traditional measures of instability, for example those based on standard deviations, are not able to capture these asymmetric effects. We use signed gradients in monthly measures of macroeconomic indicators to identify sharp variations. We use the following measures of macroeconomic instability:

- To measure exchange rate instability we use year-on-year variations in the exchange rate.
- Price instability is measured by the largest month-to-month rate of variation of the retail price index within the calendar year.
- Instability in long term interest rate is measured by the largest month-to-month rate of variation within the calendar year, of yield rates on 20-year sovereign bonds.

[7] This feature has motivated financial accelerator-type models (Bernanke *et al.*, 1996), including the borrowing constraint in Kiyotaki and Moore (1997), costly state verification in Bernanke and Gertler (1989) and sudden stops in Calvo (2000). The amplification effect can explain why a small fundamental problem can evolve into a large-scale deterioration of economic performance. The credit constraint, interacting with aggregate economic activity over the business cycle can generates asymmetric effects in response to unexpected productivity shocks. While a positive shock has only a small effect, a negative shock (even if temporary) can reduce the value of collateral to a discounted liquidation value. Since the liquidated assets cannot be restored when the shock is over, the amplification effect becomes persistent.

Firm-level and industry-level characteristics

We include a number of variables characterising the firm and its financial performance, and controls for unobserved heterogeneity at the industry level.

- Firm size is measured as the logarithm of fixed capital in real terms, incremented by unity.
- Profitability is measured by the ratio of cash flow to one-year-lagged total assets.
- We use current ratio, the ratio of current assets to current liabilities, as a measure of liquidity.
- Debt sustainability is measured using interest cover, the ratio of interest expenses to profits before interest and tax.
- We measure the firm's financial structure in terms of its gearing ratio, which is the ratio of debt to the sum of debt and equity.

We experimented extensively with alternative firm-level measures, but the substantive conclusions from our models were robust to choice of variables. In addition to the usual ratios, we estimated our model using lagged average sales growth over the past 3 and 5 years, as a proxy for demand conditions. Again, conclusions were robust, though the sample sizes were substantially reduced.

7.2.2 Econometric Methodology

There are a few empirical studies on firm exits based on discrete outcome or scoring models such as probit or logit, but the larger part of the literature have relied on hazard regression models for inference. In our context, there are two advantages to the use of hazard models.

First, these models explicitly incorporate the timing of alternative outcomes, and therefore adequately account for sample selection due to censoring. For example, the likelihood contribution for a firm that went bankrupt in 1980 would incorporate not only the information that the firm went bankrupt, and was not acquired in 1980; but also the fact that it neither went bankrupt nor was acquired in any of the previous years of its existence.

Second, hazard regression models can be used to explicitly segregate the age aspect of the propensity to survive (or exit) from the effect of other covariates. At the same time, this framework allows the effect of age on the hazard to be completely flexible, and the effect of other covariates to possibly vary with age of the firm. This is important in disentangling the influence of macroeconomic conditions on business exit from the influence of firm-specific and industry factors, as well as for understanding the role of learning in mature firms.

The model places the risks of bankruptcy and acquisitions in a unified framework. Each firm is conceived as being concurrently under risk of bankruptcy and acquisition during each year of its life. In other words, bankruptcy and acquisitions are mutually exclusive outcomes, influenced by their own determinants, competing to restrict the survival of an operating firm.

In a hazard model framework, this data generating process can be parametrised using a competing risk model where inference is based on the cause-specific intensity (hazard) rates

$$\lambda_h(t;\theta) = \lim_{\epsilon \to 0} \frac{1}{\epsilon} P[T < t + \epsilon; H = h | t \geq t; \theta] \tag{7.1}$$

where $h = 1, ...k$ are the k competing causes of failure, and $\lambda_h(0;\theta) = 0; h = 1, ...k$. The Cox Proportional Hazards (PH) model provides a convenient description of the regression relationship between the cause-specific hazard rates (Equation 7.1) corresponding to the competing causes of failure, and various explanatory variables (covariates) describing the firm's endowments ($\underline{x}_{i,t}$), the macroeconomic environment ($\underline{m}_t$) and macroeconomic instability ($\underline{u}_t$), given the age (lifetime) of the firm ($a_{i,t}$). The model postulates that the logarithm of the cause-specific hazard function is a linear function of the covariates:

$$\lambda_h(a_{i,t}, \underline{z}_{i,t}; \underline{\theta}_h) = \lambda_{0h}\left(a_{i,t}\right) . \exp\left[\underline{\theta_h}' . \underline{z}_{i,t}\right] \tag{7.2}$$

where $\lambda_{0h}(.)$ is the baseline hazard function corresponding to the h-th cause of failure (in the present case, h takes two values – bankruptcy or acquisition) at age $a_{i,t}$, $\underline{z}$ is the vector of covariates (comprising $\underline{x}_{i,t}$, $\underline{m}_t$ and $\underline{u}_t$), and $\underline{\theta_h}$ are the vectors of coefficients corresponding to the h-th cause of failure.

The parameters of the model are (a) the two baseline hazard functions, $\lambda_{0h}\left(.\right)$, corresponding to the two competing causes of failure, and (b) the distinct vectors of covariate effects ($\underline{\theta_h}$) for the two causes. In the following subsections, we consider estimation in the simple case when proportionality holds, and explain some additional features of our estimation procedure. These include discussion of:

(a) the assumption of conditional independence of the two competing exit routes required for estimation,

(b) violation of the PH assumption and modeling nonproportionality through age-varying covariate effects, and

(c) the effect of left truncation on the estimates.

Further checks on the robustness of our model estimates are discussed later.

Estimation under PH assumption

As emphasized throughout the thesis (see, for example, Chapters 1, 3 and 4), the assumption of proportional hazards is often violated in application and sometimes contested by relevant theory. Therefore, we allow the covariates to have potentially time varying coefficients.

Estimation of a Cox PH model in a multivariate duration model setting is discussed in Wei *et al.* (1989) and Spiekerman and Lin (1998); their model is similar to our regression model for cause-specific hazard rates (Equation 7.2). Inference is based on "*quasi-partial likelihood*" estimating equations with a working assumption of independence (Spiekerman and Lin, 1998). In our competing risks setting, this assumption stipulates that censoring by the competing risks must be independent of the age of the firm at exit, conditional on the observed covariates $\underline{z}$ [8]. In essence, this requires the selection of covariates such that, after conditioning on them, the competing exit processes are independent of each other. We discuss this conditional independence assumption in more detail in the following subsection.

Following Spiekerman and Lin (1998), we express the log- "quasi-partial likelihood" of $\underline{\theta_h}$,

[8] Note that the competing risks model is actually identified under a weaker condition that the two competing exit processes are "*non-informative*" about each other (Arjas and Haara, 1987; Andersen *et al.*, 1993). However, asymptotic results are easier to derive under independence, which we assume.

under the independence assumption, as:

$$l\left(\underline{\theta_h}\right) = \sum_{i=1}^{n}\sum_{h=1}^{k}\int_{0}^{\tau}\left[\underline{\theta_h}'.\underline{z}_{iu} - \ln\left\{\sum_{j=1}^{n} Y_{jh}(u).\exp\left(\underline{\theta_h}'.\underline{z}_{ju}\right)\right\}\right].dN_{ih}(u), \tag{7.3}$$

where $N_{ih}(u)$ denotes the counting process for exits corresponding to the h-th competing risk, and $Y_{jh}(u)$ denotes the corresponding at-risk indicator function (see Andersen *et al.*, 1993). The above expression is the same as the partial likelihood for a stratified Cox model with two independent strata and independent observations in each strata.

Our estimates of covariate effects, $\widehat{\underline{\theta_h}}$, are the ones that maximise the above log-"quasi-partial likelihood" (Equation 7.3)

$$\widehat{\underline{\theta_h}} = \arg\max_{\underline{\theta_h}} l\left(\underline{\theta_h}\right), \tag{7.4}$$

and the estimates of the baseline cumulative hazard functions

$$\Lambda_{0h}(t) = \int_{0}^{t}\lambda_{0h}(u).du$$

are the corresponding Aalen-Breslow type estimators:

$$\widehat{\Lambda}_{0h}(t;\widehat{\underline{\theta_h}}) = \int_{0}^{t}\frac{\sum_{i=1}^{n} dN_{ih}(u)}{\sum_{i=1}^{n} Y_{ih}(u).\exp\left(\widehat{\underline{\theta_h}}'.\underline{z}_{iu}\right)}. \tag{7.5}$$

There are several notable features of this estimation methodology. First, the quasi-partial likelihood (Equation 7.3) is valid under certain forms of unobserved heterogeneity. Specifically, estimation based on this quasi-partial likelihood accounts for frailty arising from a common scalar index of unobserved regressors for the two competing risks (Spiekerman and Lin, 1998)[9].

Second, estimation of the model is straightforward. It can be seen from the form of the quasi-partial likelihood that estimating this model is equivalent to estimating two separate

[9] Because of possible correlation between exit events in the presence of unobserved heterogeneity, asymptotic results cannot be established using standard counting process martingale theory approach (Andersen *et al.*, 1993). One of the main contributions of Spiekerman and Lin (1998) is to provide rigorous statistical results for this case.

univariate Cox regression models corresponding to the two causes of failure – acquisitions and bankruptcies. Therefore, the model can be estimated by maximising the usual stratified partial likelihood function (Cox, 1972). This implies estimation of two separate Cox PH models, one for exits due to bankruptcy and the other one for acquisitions. In each case we treat exits due to the other cause as censored cases. However, unlike the univariate hazard regression model, the interpretation of our parameter estimates will relate to the cause-specific hazard functions rather than the hazard functions themselves.

Third, the data allow us to observe the year a firm is listed and its year of exit. Since failure time is recorded only in years, the latent data are continuous, but observed lifetimes are interval censored. However, since there is considerable variation in the ages of the firms included in the sample,[10] we estimate the model in a continuous time framework using Cox partial likelihood estimates of the regression models (Cox, 1972), thereby ignoring the interval censored nature of observed data. The Peto-Breslow approximation (Breslow, 1974) is used to adjust for ties in computing the log quasi-partial likelihood and the martingale residuals.

Independence of exits due to competing causes

As in the case of univariate Cox regression models, the inference procedure presented above is valid only under the assumption that censoring is independent of failure conditional on covariates included in the model. In the competing risks model, exits are censored by competing causes of failure and hence we have to explicitly make this assumption. Such independence can be achieved by including all regressors in both the models[11]. This procedure is valid under the assumption that, conditional on the covariates, bankruptcy exits are independent of exits due to acquisitions, and vice versa[12].

In other words, when we consider the hazard regression model for bankruptcy, we include all the factors affecting acquisition hazard, and assume that other forms of censoring are either independent or at least depend on the same covariates. We deal in a similar way with the

[10] For example, the oldest exit due to bankruptcy is observed at an age of 113 years post-listing, while for acquisitions the oldest observed case is 186 years.

[11] See also Andersen *et al.* (1993).

[12] In some cases, there may be frailty, where the dependence between the two exit types is not completely described by observed covariates. As discussed above, our inference procedures are also valid under certain types of frailty.

regression model for exit due to acquisitions.

Time varying coefficients

The implications of the PH assumption and interpretation of the hazard regression model with time varying coefficients in nonproportional hazard situations has been discussed earlier (see Chapters 3 and 4). The usefulness of the histogram-sieve estimators of Murphy and Sen (1991) in this context has also been discussed. This method involves dividing the duration scale into several intervals, and including the continuous covariate interacted with indicator functions corresponding to each of the intervals as covariates in a modified Cox PH model. Since we expect a non-constant covariate effect, we would ideally like to have a large number of intervals to capture this feature. An alternative would be to use kernel based methods to estimate the covariate effect continuously over duration (see Bhattacharjee, 2004a). We divide the range in which the ages of firms fall into four intervals – the choice of the number of intervals and the cut-off ages was determined by considerations of parsimony and the requirement that each interval should include sufficient number of exits (of each competing type) and a balanced number of firm-years (observations)[13].

Our chosen 4 intervals are age $0-4$ years, age $5-15$ years, age $16-25$ years and age > 25 years, post-listing. Each of these four intervals have reasonable incidence from the total sample, covering 7569 (16 per cent), 13474 (28 per cent), 11817 (25 per cent) and 15234 (32 per cent) company years respectively[14].

Finding covariates that have non-proportional effects is an important step in the implementation of the above methodology. We use two statistical tests to identify covariates with time varying effects on the cause-specific hazard of either exit. One is the test of the PH assumption against ordered alternatives proposed in Bhattacharjee (2007a; our Chapter 3), and the other is a test for proportionality based on martingale residuals (Grambsch and Therneau, 1994). In this particular application, both tests lead to a very similar selection of covariates. Our empirical results demonstrate that several covariates have age-varying covariate effects, and segmentation

[13] We also experimented with 3 and 5 intervals. With 3 intervals, we sacrifice some flexibility in variation of covariate effects over duration, while for 5 intervals some of our estimates are less significant because of lower sample size (number of company-years, but more importantly number of bankruptcies) in each interval.

[14] The incidence in terms of number of bankruptcies is 49, 56, 51 and 50 respectively, and in terms of acquisitions - 379, 555, 455 and 469, respectively.

of the duration scale in characterising effectively the way the impact of a covariate varies over the life of the firm, post-listing.

Left truncation and robustness of estimates

In addition to right-censoring (by dependant competing risks), our duration data are truncated to the left, in that they pertain only to the period after 1965. For a given firm, the age at left-truncation is given by $L = 1965 - B$, where B is the listing-year of the firm. The Cox partial likelihood estimates based on a modified definition of risk sets (delayed entry) are valid if truncation and exits are independent either unconditionally, or at least after conditioning on the included covariates. While there is no simple way to test such conditional independence, the impact of dependence on estimates can be examined by stratifying the sample with respect to truncation time. We evaluate the robustness of our results to dependent truncation by estimating the age at exit models conditioned on different ranges of the age at left-truncation, and examining the sensitivity of model estimates. We also estimate the models for sub-samples of the data based on different starting years. We truncate the sample at 1970 (instead of 1965), and estimate the models for bankruptcy and acquisitions for this sub-sample.

The Insolvency Act of 1986 is likely to have had a mitigating effect on corporate failures (Cuthbertson and Hudson, 1996; Liu, 2004). In order to examine whether this has a significant effect on our results, we also estimate the model for bankruptcy for the period from 1986 onwards.

In addition to evaluating left truncation, we check the robustness of our estimates in other ways. First, we estimate comparable logit models for exits due to bankruptcy and acquisition and compare the results with our hazard model estimates. Second, we compute jackknife estimates of the model to evaluate the robustness of our parameter estimates and their standard errors.

Results of these robustness tests and the check for impact of the Insolvency Act 1986, not reported here, indicate that our estimated hazard regression models for bankruptcies and acquisitions are robust. We do find evidence of effect of the Insolvency Act 1986, but the conclusions from our estimates for the period since 1965 are preserved. In Section 7.3 (Bhattacharjee *et al.*, 2008b), we report similar evidence of the impact of Chapter 11 legislation on bankruptcies and acquisitions in the US.

7.2.3 Results

The maximum partial likelihood model estimates of the two models, for bankruptcies and for acquisitions, are reported in Table 7.2.1. The reported estimates are hazard ratios, which are exponentials of the estimates of the Cox PH model regression coefficients. These estimates are interpreted as the factor by which the hazard would be increased if there were a one unit increase in the covariate under consideration, other things equal.

The reported z-scores are based on robust standard error estimates proposed by Lin and Wei (1989). These are obtained using a sandwich estimator, where clustering by year is adjusted fro by summing the score residuals within each year before applying the sandwich estimator. The fit of models is judged using a Wald chi-square test, and the validity of the proportionality assumption by the tests proposed in Bhattacharjee (2007a) and Grambsch and Therneau (1994). These tests help us identify two regressors with age-varying covariate effects. The effect of these covariates (our measures of instability in exchange rates and inflation) are allowed to vary over the age of the firm using the histogram sieve estimator (Murphy and Sen, 1991). As discussed earlier, our checks for sensitivity of the estimates indicates that the estimated models are quite robust.

Firm and industry specific factors

Industry matters significantly for either form of exit. Textiles and construction companies are more likely to go bankrupt but less likely to be acquired. While firms in the paper/ packaging business are more likely to be acquired, firms in the engineering and ICT industries have a lower acquisition propensity. The broad division appears to fall along the traditional/modern divide.

TABLE 7.2.1: Model Estimates

Variables	Bankruptcy	Acquisitions
Industry Dummies		
(Base = all others)	1.00	1.00
– Food/Breweries	$0.8349(-0.4)$	$1.1755(1.7)^{+}$
– Chem./Pharma.	$0.5888(-1.3)$	$1.1079(1.1)$
– Metals	$0.4341(-0.8)$	$1.0671(0.4)$
– Engineering	$1.2342(0.9)$	$0.7521(-3.4)^{**}$
– Electricals/Electronics	$0.9073(-0.3)$	$1.1333(1.4)$
– Textiles	$2.0297(3.3)^{**}$	$0.8283(-2.1)^{*}$
– Paper/Packaging	$0.9958(-0.0)$	$1.2053(2.2)^{*}$
– Construction	$1.4754(1.7)^{+}$	$0.7650(-3.1)^{**}$
– ICT	$0.4191(-1.7)^{+}$	$0.4400(-5.2)^{**}$
– Trdg./Superstores	$0.9224(-0.3)$	$0.8940(-1.5)$
Firm $\times$ Year Level		
Current size:		
ln(real fixed capital +1)	$1.1935(1.0)$	$1.2390(3.8)^{**}$
Size-squared	$0.9614(-1.9)^{*}$	$0.9757(-4.5)^{**}$
Cash flow to Capital	$1.0086(0.1)$	$1.3683(8.0)^{**}$
Current ratio	$1.0062(1.3)$	$1.0105(8.7)^{**}$
Interest cover	$0.9619(-4.8)^{**}$	$0.9840(-2.2)^{*}$
Gearing ratio	$1.0258(3.3)^{**}$	$0.9978(-0.1)$
Macro-Economic Conditions		
UK Business cycle	$0.9831(-0.2)$	$0.9371(-1.6)$
Long-term real interest rate	$0.9855(-0.5)$	$1.0225(2.1)^{*}$
£ $-$ \$ exchange rate	$1.0383(0.4)$	$1.0216(0.8)$
US business cycle	$0.8515(-2.2)^{*}$	$1.2298(6.2)^{**}$
Macro-Economic Instability		
y-o-y increase in £ $-$ \$ exchange rate $= v$		
– $v\times$ I(age 0-4 yrs.)	$1.2722(1.9)^{*}$	$0.8691(-2.7)^{**}$
– $v\times$ I(age 5-15 yrs.)	$1.2407(1.3)$	$0.8891(-2.6)^{*}$
– $v\times$ I(age 16-25 yrs.)	$1.0437(0.2)$	$1.0051(0.1)$
– $v\times$ I(age $>$ 25 yrs.)	$1.0424(0.3)$	$0.9359(-1.5)$

TABLE 7.2.1: Model Estimates (Contd.)

Variables	**Bankruptcy**	**Acquisitions**
Vol. - RPI inflation = x		
– $x\times$ I(age 0-4 yrs.)	1.3044(1.2)	$0.8644(-1.8)^{+}$
– $x\times$ I(age 5-15 yrs.)	1.0832(0.4)	$0.8326(-2.9)^{**}$
– $x\times$ I(age 16-25 yrs.)	$0.6906(-1.3)$	$0.8055(-4.5)^{**}$
– $x\times$ I(age $>$ 25 yrs.)	$0.6933(-1.9)^{+}$	$0.8254(-3.0)^{**}$
Volatility - Long term int. rate	1.1886(0.9)	$0.7297(-5.8)^{**}$
No. of firms	4,117	4,117
No. of exits	206	1,858
Total time at risk (in firm-yrs.)	48,094	48,094
Log-likelihood	−1357.808	−12661.188
Wald χ^2 goodness-of-fit test	135.11	383.08
d.f. / p-value	29/0.000	29/0.00
χ^2 test (PH assumption)	14.92	34.77
d.f. / p-value	29/0.990	29/0.251
Only macro-variables (log-lik.)	−1399.280	−12780.16
LRT – Joint significance of		
firm/ind. var. (d.f. / p-value)	16/0.000	16/0.000
Only firm/ind.-variables (log-lik.)	−1375.086	−12714.53
LRT – Joint significance of		
macro. var. (d.f. / p-value)	13/0.002	13/0.000

z-scores in parentheses.

Parameters reported are hazard ratios (exponential of the regression coefficient estimates).

Volatility is measured as maximum monthly difference during the year, divided by the no. of intervening mths.

**, *and $^{+}$– Significant at 1%, 5% and 10%level respectively.

Firm specific characteristics have impacts suggested in the literature. The rates of bankruptcy and acquisition decline sharply with size in the higher size-ranges. Figure 7-5 shows the estimated hazard ratios against size-percentiles after conditioning on other covariates. There is a sharp decline of bankruptcy hazard with size. The figure supports the stylised fact from the acquisition literature that quoted firms in the middle range of the size-distribution are considerably more likely to be acquired.

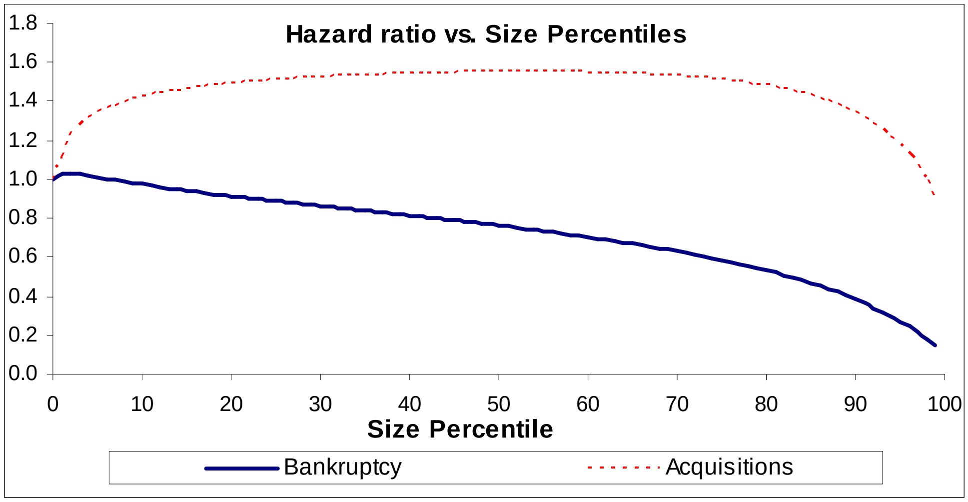


Figure 7-5:

Firms with higher interest cover have a low exit hazard from both bankruptcy and acquisitions. While a higher gearing enhances the risk of bankruptcy, cash rich firms and firms with higher liquidity (with higher cash flow to capital ratio and higher current ratio respectively) are preferred as acquisition targets.

Macroeconomic factors

We conditioned on the long term real interest rate and the sterling-dollar exchange rate. The long term rate has a significant impact only on acquisitions while the exchange rate has no significant impact on either bankruptcy or acquisition. We also conditioned on measures of both the UK and the US business cycle. Only the US business cycle measure has a significant effect on bankruptcies and acquisitions; apparently the US economy is a better predictor of UK bankruptcies and acquisitions than the business cycle in the UK itself. The effect of the US business cycle on acquisitions is particularly strong, indicating the importance of demand for acquired capital from the international capital market in driving merger waves. In the case of bankruptcy, the effect is likely to have been driven by demand for exports.

In comparison to general macroeconomic conditions, the impact of macroeconomic instability on business exits is more pronounced, and depends substantially on the age of the firm since listing, particularly for acquisitions[15]. Newly listed firms are more likely to go bankrupt during the years when exchange rate changes are very sharp. On the other hand, acquisition hazard for younger firms is reduced during these years. Price instability[16] and volatility in long term interest rates subdued acquisition activity significantly.[17] Overall, our findings point to the detrimental impact of macroeconomic instability on survival.

Figure 7-6 plots the baseline cumulative hazard functions of bankruptcy and acquisition against the age of the firm reckoned from listing date. Note that the baseline cause-specific hazard rate of mergers is about four times that of bankruptcy, controlling for covariates. While the baseline hazard due to mergers appears to be constant over the lifetime of a firm, post-listing, the baseline hazard due to bankruptcy decreases with age upto about 20 years post-listing, arguably reflecting a learning effect. In the literature, evidence in favour of learning models has been advanced from cohort studies of new young firms, and it is interesting to note evidence for mature firms. A statistical test, combining the generalised residuals proposed by Peňa (1998) with the test for exponentiality against NWU alternatives (Ahmad, 2001), also confirms the evidence on weak negative ageing.

Figures 7-3 and 7-4 also present the year-wise predicted incidence rates of bankruptcies and acquisitions against the observed incidence rates. The close conformity between the two is noteworthy.

7.2.4 Conclusions

In this section, we used methods developed in this thesis to examine the relationship between business exits and instability associated with the macroeconomic cycle, focussing on large and mature (listed) UK companies, over a long (thirty-four year) period. We disentangled the joint determination of probabilities of two mutually exclusive processes - firms being acquired and firms going bankrupt - by estimating a competing risks model for the probabilities of exit in

[15] The evidence of non-proportionality of hazards underscores the usefulness of the Murphy-Sen histogram sieve estimators for inference in such non-proportional situations.

[16] Wadhwani (1986) provides an explanation for how inflation volatility can contribute to bankrupcy. Firms already in a state of financial distress can be tipped over into bankruptcy as higher inflation and higher nominal interest rates increases the service element of debt.

[17] While the effect of instability on bankruptcy hazard is not significant for the entire period under analysis, the effect is more pronounced for the recent period after the introduction of the Insolvency Act of 1986.

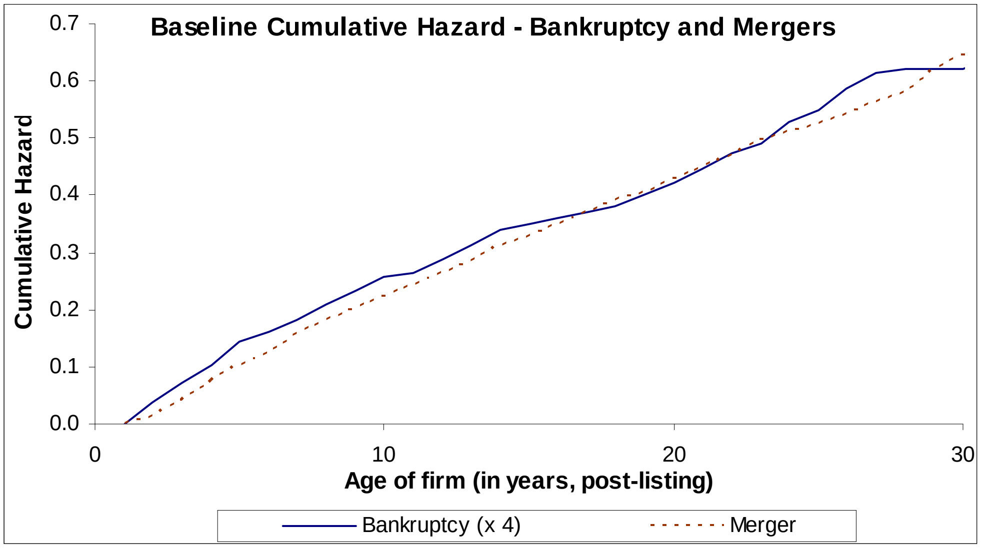


Figure 7-6:

either form, in terms of firm characteristics, industry and features of the business cycle. Our model explains the observed time variation in the incidence of bankruptcy and acquisitions quite well. The two types of exits are marked by differences in the effects of firm-level drivers, industry, macroeconomic conditions as well as macroeconomic instability.

At the firm level our findings corroborate earlier results; the baseline hazard due to bankruptcy and mergers decreases with age after listing. Other factors remaining constant, larger firms and firms with higher interest cover are less likely to go bankrupt or be acquired. Firms with higher liquidity and cash rish firms are more attrative acquisition targets, and firms with higher gearing are more likely to go bankrupt.

Our empirical results on the impact of macroeconomic instability on exits are new. There are notable differences in the way in which recently listed firms, and those listed some years previously respond to changes in the macroeconomic environment. This evidence highlights the usefulness of the framework and methods developed in this thesis for empirical analyses of covariate effects in failure time data.

Uncertainty in the form of sharp increases in inflation and sharp depreciation of the pound sterling affect freshly listed firms adversely - they are more likely to go bankrupt during unstable years. Acquisition activity is also subdued in these years. Further, there are less bankruptcies and more acquisitions during an economic upturn, particularly when measured by the US business cycle. The finding of contemporaneous increase in bankruptcies and decline in acquisitions, in a period of instability or low economic growth, suggests the need for further work on assessing causal relationships between the two processes.

The results reported here underscore the importance of smooth macroeconomic management for the corporate sector. In an era of globalisation they also point to the role that may potentially be played by business cycles in other economic regions in the determination of both forms of business exit. International comparisons, estimating similar models for other economies would aid understanding and policy. Estimates of a similar model for the US (Bhattacharjee *et al.*, 2008b; Section 7.3) also points to an important role for bankruptcy legislation.

7.3 Business failure in UK and US quoted firms: impact of macroeconomic instability and the role of legal institutions

Following our investigation of business exit among quoted firms in the UK (Section 7.2), we examine how macroeconomic instability affects risk of bankruptcy and liquidation in listed US firms. The study is based on Bhattacharjee *et al.* (2008b). In periods of macroeconomic instability more firms become financially distressed while the number of potential acquirers falls. Reorganisation systems such as Chapter 11 can decouple liquidation from macroeconomic conditions. The economic framework behind our analysis is provided by the model presented in Section 7.1.2 (Bhattacharjee *et al.*, 2008a), in which a firm's bankruptcy and acquisition hazards are codetermined by firm-level and sector-level factors, but also by macroeconomic conditions. As a control we use estimates of a similar model for the UK (reported in Section 7.2), which is an economy without an equivalent system to Chapter 11. Differences in responsiveness of bankruptcy to instability are largely attributable to reorganisation under Chapter 11.

7.3.1 Econometric methodology

We employ hazard regression models within a competing risks framework to study the impact of various explanatory factors (covariates) on firm exit. Since the methodology is largely similar to Section 7.2, we emphasize only the main aspects here.

The failure time data considered here are right-censored (by dependant competing risks) and left truncated in 1969 (1965 for the UK firms, and 1980 for competing risks model for Chapters 7 and 11 in Table 7.3.2). Like Section 7.2, we evaluate the robustness of results to dependent truncation by estimating the exit duration models conditioned on different ranges of the truncation duration and comparing estimates for similarity.

Estimates of the regression coefficients are obtained, by maximising the partial log-likelihood of the regression coefficients. We report the maximum partial likelihood model estimates of hazard ratios, which are the exponential of the estimates of the Cox PH model regression coefficients. These estimates are interpreted as the multiplicative factors by which the hazard would be increased if there were a one unit increase in the covariate under consideration, other things equal. The reported z-scores are based on robust standard error estimates proposed by Lin and Wei (1989). These estimates are obtained using the Huber-White sandwich estimator, after adjusting for clustering by year by summing the score residuals within each year.

The above framework allows dependence between the competing exit events. The interaction between the two hazard rates is characterised by variation in covariates included in the analysis. Like Section 7.2, non-identifiability of the hazard rates for the competing causes of exit necessitate partial likelihood inference on cause specific hazard rates. In other words, our approach provides valid inference on the hazard rates for bankruptcy and acquisition only under an important assumption – that exits due to the competing risks are independent of each other after conditioning on all the included covariates. After conditioning on these covariates, the hazard rates for the two competing causes are independent of each other. Further, as shown in Spiekerman and Lin (1998), the above argument also holds in the presence of some forms of unobserved heterogeneity – specifically, when there is a single scalar unobserved heterogeneity term for the two competing risks which acts multiplicatively on the two hazard rates.[18]

Most of the regressors considered, whether firm-level or macroeconomic factors, are time varying covariates. In addition, we explicitly allow for the possibility that the effect of some covariates may change over the lifetime of the firm; in other words, there may be time varying

[18] Similarly, partial likelihood inference for Chapter 7 and Chapter 11 bankruptcies is valid conditional on a suitable selection of covariates.

coefficients (see Chapter 4; Bhattacharjee, 2004a). This constitutes a violation of the proportionality assumption underlying the Cox PH model. For each covariate included in our models, we verify the validity of the proportionality assumption using the tests proposed in Chapter 3 (Bhattacharjee, 2007a) and Grambsch and Therneau (1994), and identify variables with time varying coefficients. Several covariates are identified as having age-varying effects. With the results, we also report tests of the overall validity of the PH assumption (Grambsch and Therneau, 1994).

For estimating the time varying coefficients, we use the intuitive and appealing histogram-sieve estimators of Murphy and Sen (1991). This method entails dividing the lifetime into several intervals and including the covariate interacting with indicator functions for each of the intervals as covariates in a modified Cox PH model. In the analysis that follows, the lives of US firms, post-listing, was divided into four intervals (age 0-8 years, age 9-16 years, age 17-25 years and age $>$ 25 years). This partitioning was chosen in order to allocate similar number of bankruptcies to each age-interval. As our results will demonstrate, several of the covariates have time varying coefficients, and this segregation of the failure time scale helps us to effectively characterise the way the impact of a covariate varies over the life of the firm.

7.3.2 The effect of bankruptcy code

Chapter 11 was instituted in the US on October 1, 1979 as a consequence of the Bankruptcy Reform Act of 1978. Previously the US bankruptcy code remained functionally quite similar to the insolvency system in the UK from which, pre-independence, it derived. A primary aim of the 1978 Act was to make it easier for businesses and individuals to file for bankruptcy in order to reorganise. To facilitate this, the existing management ('the debtor') continues to manage the firm and retain significant rights as debtor-in-possession, and the court mandates the management to propose a reorganisation plan. The initial 120 days to do this can be extended repeatedly by the court, and for larger firms the Chapter 11 process has frequently taken several years (LoPucki and Whitford, 1993). Large listed US firms in distress almost invariably go initially through Chapter 11. The court can then decide that the continuation value of the firm is low and convert a Chapter 11 filing to Chapter 7, which constitutes automatic liquidation. Hence reorganisation systems like Chapter 11 have the potential to offer a safe haven for distressed firms in periods of high macroeconomic instability, enabling some of these firms to recover and perhaps be acquired. In the UK, receivership has offered the principle alternative to immediate liquidation for a distressed firm, but Chapter 11 and receivership are substantially different in their effects. The receiver represents secured (senior) creditors and replaces management. The UK 1986 Insolvency Act introduced the 'administration' process to offer some of the characteristics of Chapter 11 but secured creditors can block the appointment of an administrator by appointing a receiver and, in practice, administration has been rarely used and has not materially changed the creditor orientation of the UK system.

The efficiency implications of deviations from absolute priority in 'debtor-friendly' bankruptcy systems have been the subject of intense debate (Mooradian, 1994; Bebchuk, 2002). However, the stark debtor-friendly/creditor-friendly dichotomy can be exaggerated, and when the legal system imposes costs, actors are likely to mitigate these costs by informal action. In both the US and the UK informal workouts can avoid bankruptcy proceedings altogether and there is evidence that, in the 1990s, large banks became more effective in softening the impact of the UK bankruptcy code through coordination on workouts (Armour *et al.*, 2002). In the US, Baird and Rasmussen (2003) argue that, as investors have become increasingly sophisticated in writing complex contingent contracts, few large Chapter 11 bankruptcies now fit the classic reorganization paradigm in which the court reorganises messy and conflicting claims that threaten the survival of a firm with continuing value. By 2002, in almost all large chapter 11 bankruptcies, effective control was in the hands of senior creditors and that the role of Chapter 11 was to arrange an orderly sale of the firm, in whole or in part. This is consistent with the role we place on Chapter 11 in our analysis here.

7.3.3 Data and construction of variables

We construct the US sample by matching the Compustat accounting database with the CRSP database to identify all listed firms[19] and to extract listing data. This gives an unbalanced panel of about 13,700 US industrial and commercial firms over the period 1969 to 2000. There were 561 exits due to bankruptcy and 2,516 acquisitions in 132,410 firm years over the 32 year

[19] Listed on the NYSE/AMEX, NASDAQ, Over-the-Counter or any of the regional exchanges (Boston, Midwest, Montreal, Pacific or Philadelphia).

period. Figures 7-7 and 7-8 plot the incidence of bankruptcy and acquisition for each year, where incidence is defined as the number of companies that went bankrupt (or were acquired) during the year to the total number of listed companies.

Failure time data, measuring the postlisting lifetime of each firm, are augmented by annual indicators of macroeconomic conditions, as well as firm and industry-specific factors. These variables constitute the time-varying covariates used to explain exit-probabilities or hazard rates. The competing risks framework involves estimation of two separate Cox PH models, one for exits due to bankruptcy and one for acquisitions. In each case we treat exits due to the other cause as censored cases, in addition to observations originally censored due to delisting and other reasons. As in Section 7.2, the data are left-truncated, randomly right censored by potentially dependent competing risks, and the covariates explaining the nature of the cause-specific hazards are time-varying. We obtain estimates of the model parameters making the censoring duration non-informative about the exit duration, after conditioning on an adequate selection of covariates. Further, we take into account possible violation of the proportionality assumption inherent in the Cox regression model, by allowing the covariate effects to vary over the lifetime of the firm (Chapter 4; Bhattacharjee, 2003, 2004a).

The data for the UK firms is discussed in Section 7.2. Below, we describe construction of the macroeconomic covariates, firm-level variables and industry-dummies for the US data; the constructs for the UK data are similar.

Measures of macroeconomic activity.

We use the following empirical proxies for the level of macroeconomic activity:

- The business cycle or output gap (o_t), is measured by the difference between trend output and actual output, using a quarterly Hodrick-Prescott filtered series of output per capita.
- The indicator of business entries is the log-difference of the number of new listed firms in the accounting database for each year.
- Real interest rates are measured as the annual average of monthly 10-year treasury bill rates, minus the annual rate of inflation. The yield on 20-year UK sovereign bonds are used to construct the corresponding measure for the UK.

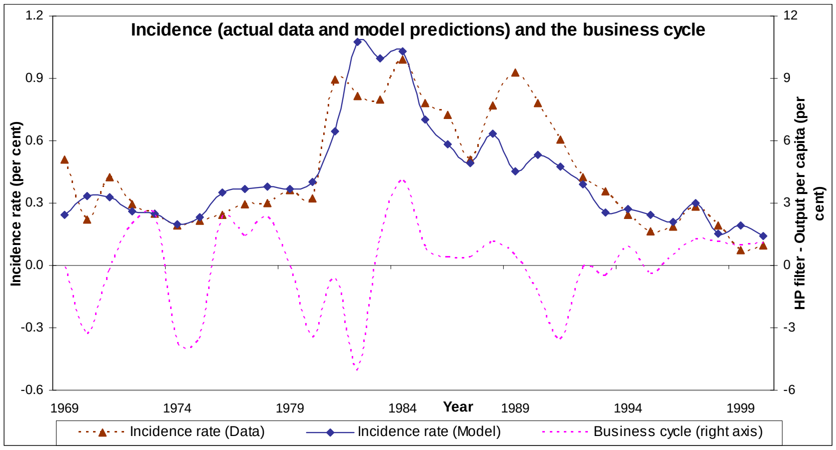


Figure 7-7: US business cycle and corporate bankruptcies.

- The exchange rate is measured by the annual average of monthly nominal broad dollar index (based on trade composition with G-10 economies). For the UK, we use the average annual real effective exchange rate[20].

Figures 7-7 and 7-8 plot the annual incidence of bankruptcies and acquisitions, respectively, for US firms against the business cycle indicator for the year. The corresponding plots for quoted UK firms are given in Figures 7-3 and 7-4, respectively.

As discussed in Section 7.2, the incidence of bankruptcy is high during years when the economy turned down after a peak, and lower during upturns in the business cycle, while acquisitions are procyclical. However, the responsiveness of bankruptcies to turnaround in the business cycle is lower for the US than for the UK. These plots suggest that macroeconomic conditions are important in explaining the survival of listed firms, before conditioning on firm and industry-specific characteristics.

Measures of macroeconomic stability

Figures 7-7 and 7-8 (and corresponding Figures 7-3 and 7-4 for UK firms) also suggest, *a priori*, that even for mature (quoted) firms, the aggregate incidence of bankruptcies and acquisitions shows substantial variation over time. While a part of this aggregate movement

[20] We regard fluctuations in the level of the exchange rate as part of the macroeconomic environment firms face, even though there is a difficulty linking many exchange rate movements to other fundamental macrofactors.

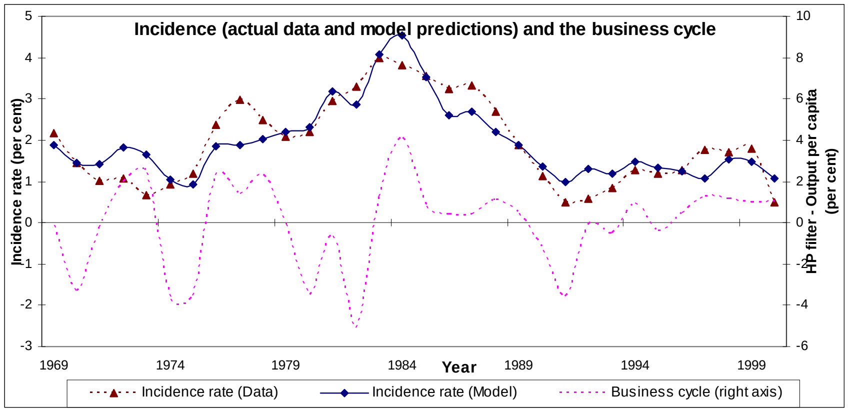


Figure 7-8: US business cycle and corporate acquisitions.

can be explained by the business cycle, macroeconomic stability can also have a role to play. Following Section 7.2, we use signed gradients in monthly measures of macroeconomic indicators to identify sharp changes. We use the following empirical proxies for macroeconomic instability:

- The sharpness of the economic turnaround is measured by $[o_t - o_{t-1}] - [o_{t-1} - o_{t-2}]$, which is the increment of the change in output gap in the current year $(o_t - o_{t-1})$ from that in the previous year. This is a measure of the curvature or second order derivative of the Hodrick-Prescott filter of output per capita. Over a business cycle, this measure would be lowest right after the peak, when the economy turns around downwards, and continue to increase gradually upto its maximum right after the trough, when the economy picks up. Over different business cycles, this measure would be lower (or higher) for a cycle in which the economy turns down sharply after a sharp upturn (or turns up sharply after a sharp downturn.

- Instability in the foreign exchange market is measured by the year-on-year change in the real exchange rate.

- Price instability is measured by the largest month-to-month rate of variation of the retail price index within the calendar year.

- Instability in long term interest rates is measured by the largest month-to-month rate of variation within the calendar year, of yield rates on 20-year sovereign bonds.

This current work focuses on the relationship between the macroeconomic environment (including macroeconomic activity and macroeconomic stability) and exits at the firm level. While Figures 7-7 and 7-8 provide preliminary descriptive idea about the nature of the above relationships, estimates of the partial effects on the hazard rate of exits requires estimation of the econometric models presented in the previous section. In addition to the special role attributed to age of the firm, these models adequately account for simultaneous changes in all the macroeconomic factors as well as firm-level and industry-level characteristics.

Firm-level and industry-level characteristics

The existing theoretical and empirical literature has identified a number of firm and industry-specific features as important determinants of firm exits (Siegfried and Evans, 1994; Caves, 1998). The literature suggests that the age of a firm is an important determinant of survival probabilities of new entrants, though it is not clearly indicated whether the same can also hold for mature (listed) firms. In the hazard model specification, age-since-listing (in years) is used as the measure of firm age to explore this issue. We include dummies to capture industry effects, and a number of variables characterising the firm and its financial performance:

- Firm size is measured as the logarithm of fixed capital in real terms, incremented by unity.
- Profitability is measured by the ratio of cash flow to one-year-lagged total assets.
- Current ratio, which is the ratio of current assets to current liabilities, is used as a measure of liquidity.
- Debt sustainability is measured using interest cover (ratio of interest expenses to profits before interest and tax.
- The firm's financial structure is measured by its gearing ratio, which is the ratio of debt to the sum of debt and equity.

The sample characteristics display significant variability both across firms, and over the period of analysis: the 32-year period 1969 to 2000 for the US, and the 34-year period 1965 to 1998 for the UK. Current ratio, interest cover and gearing ratio are strongly collinear with the macroeconomic variables included in the analysis, and are therefore not included in the estimated hazard regression models. These variable are, however used in the models for bankruptcy exit in the US to correct for potential endogenous selection of the exit route – Chapter 7 or Chapter 11.

7.3.4 Results

Table 7.3.1 presents parameter estimates and goodness-of-fit measures for the estimated models.

TABLE 7.3.1: Model Estimates, UK and US

Variables	**US, Bank.**	**US, Acq.**	**UK, Bank.**	**UK, Acq.**
Industry Dummies				
(not reported)				
Firm × Year Level				
Size $= s$			1.487 (1.6)	1.412 (5.3)**
$- s \times I_1$	0.571 (-1.4)	2.539 (5.1)**		
$- s \times I_2$	0.078 (-6.6)**	1.120 (0.8)		
$- s \times I_3$	0.067 (-5.4)**	1.149 (0.8)		
$- s \times I_4$	0.132 (-6.0)**	0.931 (-0.7)		
Size-squared $= s^2$			0.942 (-2.2)*	0.961 (-6.2)**
$- s^2 \times I_1$	1.058 (0.3)	0.725 (-3.5)**		
$- s^2 \times I_2$	1.549 (5.8)**	0.917 (-1.6)		
$- s^2 \times I_3$	1.463 (3.1)**	0.889 (-1.9)+		
$- s^2 \times I_4$	1.297 (3.6)**	0.950 (-2.0)*		
Cash flow to Capital $= c$	1.000 (1.6)	1.001 (2.4)*		
$- c \times I_1$			0.908 (-3.3)**	4.179 (4.4)**
$- c \times I_2$			0.682 (-4.0)**	1.302 (2.5)*
$- c \times I_3$			0.113 (-1.6)	0.351 (-2.8)**
$- c \times I_4$			0.381 (-3.3)**	0.661 (-2.0)*
Retn. on capital employed	1.000 (-4.2)**	0.999 (-2.3)*	0.997 (-2.2)*	1.001 (0.9)
Macro- Conditions				
Output gap $= o$				
$- o \times I_1$	0.409 (-0.2)	29.60 (1.4)	34901 (1.4)	75131 (4.2)**
$- o \times I_2$	0.001 (-2.0)*	2.231 (0.4)	0.002 (-0.7)	1734 (3.2)**
$- o \times I_3$	0.000 (-1.5)	1920 (2.5)*	2.9e-5 (-1.1)	0.043 (-1.0)
$- o \times I_4$	0.002 (-0.8)	336.9 (1.9)+	4716 (1.2)	28.53 (1.2)

TABLE 7.3.1: Model Estimates, UK and US (Contd.)

Variables	US, Bank.	US, Acq.	UK, Bank.	UK, Acq.
Macro- Conditions				
Entries (y-o-y growth rates)	0.979 (-0.3)	1.043 (1.7)$^{+}$	1.014 (1.7)$^{+}$	0.997 (-1.7)$^{+}$
Long-term real int. rate $= r$				
$- r \times I_1$	1.469 (9.0)**	1.345 (14.1)**	1.163 (1.4)	1.121 (3.6)**
$- r \times I_2$	1.142 (4.2)**	1.246 (12.4)**	1.018 (0.4)	0.945 (-4.0)**
$- r \times I_3$	1.130 (2.4)*	1.133 (4.5)**	0.962 (-0.9)	0.994 (-0.3)
$- r \times I_4$	1.189 (2.4)*	1.167 (5.7)**	1.072 (0.9)	0.973 (-1.4)
Exchange rate $= e$			0.080 (-1.9)$^{+}$	7.048 (5.2)**
$- e \times I_1$	0.954 (-7.1)**	0.963 (-12.6)**		
$- e \times I_2$	0.973 (-7.6)**	0.977 (-10.6)**		
$- e \times I_3$	0.985 (-3.1)**	1.001 (0.7)		
$- e \times I_4$	1.007 (1.2)	1.011 (5.7)**		
Macro- Instability				
Turnaround $= trn$	1.718 (0.4)	0.923 (-0.2)		
$- trn \times I_1$			9.3e-11 (-3.0)**	0.017 (-1.7)$^{+}$
$- trn \times I_2$			0.152 (-0.3)	300.3 (2.6)**
$- trn \times I_3$			0.001 (-1.0)	19.79 (1.4)
$- trn \times I_4$			0.000 (-1.0)	1.834 (0.2)
y-o-y increase in				
exchange rate $= v$	1.002 (0.1)	0.963 (-3.9)**		
$- v \times I_1$			9.6e+5 (3.5)**	0.424 (-0.7)
$- v \times I_2$			289.456 (1.4)	0.322 (-1.1)
$- v \times I_3$			17.577 (0.5)	0.072 (-1.8)$^{+}$
$- v \times I_4$			1305 (1.7)$^{+}$	1.037 (0.0)
Volatility - prices	0.686 (-1.3)	1.355 (2.4)*	1.276 (5.8)**	0.904 (-5.9)**
Volatility - Long term				
int. rate $= l$			0.987 (-0.2)	1.033 (1.7)$^{+}$
$- l \times I_1$	1.065 (1.7)$^{+}$	1.018 (1.0)		
$- l \times I_2$	0.968 (-0.9)	1.007 (0.4)		
$- l \times I_3$	0.947 (-1.4)	1.051 (2.6)**		
$- l \times I_4$	0.901 (-1.8)$^{+}$	1.037 (1.8)$^{+}$		

TABLE 7.3.1: Model Estimates, UK and US (Contd.)

Variables	US, Bank.	US, Acq.	UK, Bank.	UK, Acq.
Macro- Instability				
Volatility - Short term				
int. rate	1.005 (0.9)	1.009 (3.7)**	0.949 (-1.4)	0.991 (-0.8)
No. of firms	13,655	13,655	4,320	4,320
No. of exits	561	2,516	166	1,859
Total time at risk				
(in years)	132,410	132,410	45,527	45,527
Log-likelihood	-4210.73	-19075.2	-1090.59	-12947.1
Chi-square test				
stat.(PH assmp.)	16.52	28.19	29.99	14.36
degrees of freedom	39	39	38	38
p-value (%)	99.9	90.0	82.0	100.0

z-scores in parentheses.

Parameters reported are hazard ratios (exponential of the regression coefficient estimates).

For the UK, I_1, I_2, I_3 and I_4 represent the indicator functions I(age 0-5 yrs.), I(age 6-15 yrs.), I(age 16-25 yrs) and I(age > 25 yrs) respectively. For the US, the same represent I(age 0-8 yrs.), I(age 9-16 yrs.), I(age 17-25 yrs) and I(age > 25 yrs) respectively;

Volatility is measured as maximum monthly difference during the year, divided by the no. of intervening mths.

** , *and $^{+}$– Significant at 1%, 5% and 10%level respectively.

Impact of macroeconomic conditions on firm exit

Controlling for industry and firm-level characteristics, macroeconomic conditions have a significant impact on hazard rates of exit by bankruptcy or acquisition in both economies. But there are considerable differences in the impact of the macroeconomy on business failure in the UK and the US.

Figures 7-9 and 7-10 show hazard ratios against the quantiles of volatility in the different macroeconomic factors, for the US and the UK. The severely traumatic experience that periods of adverse macroeconomic conditions generate for firms is robust, and is one of the main findings of the current work. The dramatic increase in hazard rates during periods of extreme instability is visually demonstrated in Figures 7-9 and 7-10 by the slope of the hazard ratios at the highest

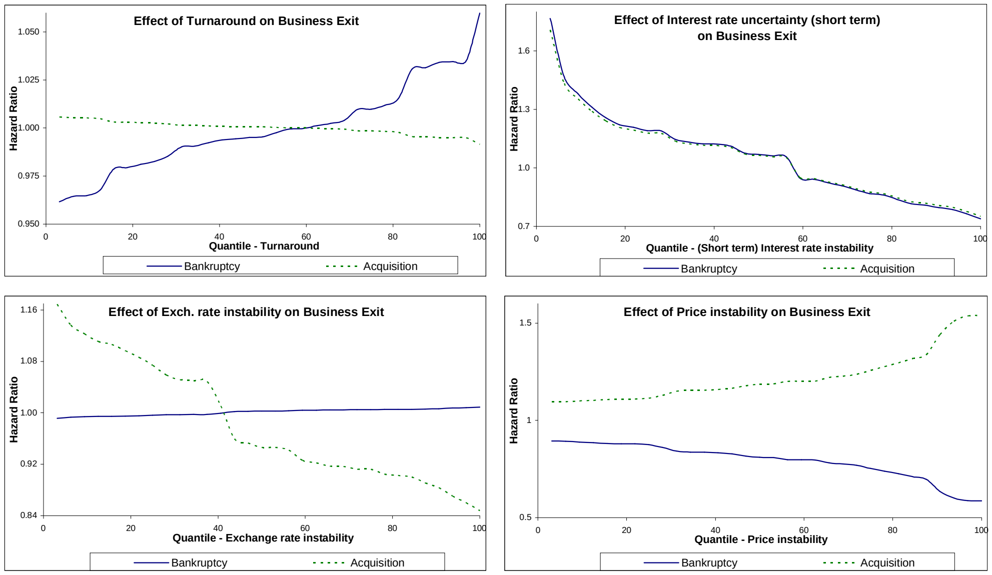


Figure 7-9: US: Effect of macroeconomic instability (hazard ratios).

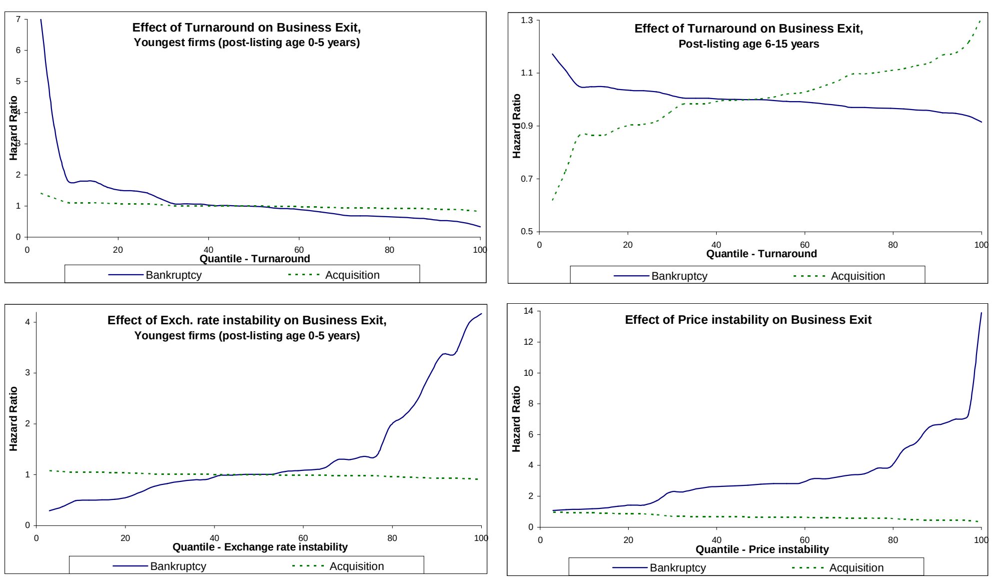


Figure 7-10: UK: Effect of macroeconomic instability (hazard ratios).

and lowest ends.

The effect of macroeconomic instability is non-linear and strong for UK quoted firms, but is smaller for US firms. The youngest UK firms are more likely to go bankrupt immediately after the economy passes its peak; whereas there is no significant effect of economic turnaround on US bankruptcies, after controlling for firm and industry-specific characteristics and other macroeconomic factors. The hazard ratio at the 3rd percentile of stability according to this measure is about 20 times higher than that at the 97th percentile in the UK, while for the US firms the hazards are about the same at both these percentiles.

Similarly, more UK firms go bankrupt in periods when the exchange rate is stronger while no such effect is observed for US firms. Young UK firms are likely to go bankrupt during years when the domestic currency depreciates sharply (the hazard for the 97th percentile is 14 times that for the 3rd percentile) while no such effect can be detected in the US (the hazards are about the same). Price instability increases bankruptcy in the UK, but not in the US. While interest rate instability does not have much effect on bankruptcy; bankruptcy in the US, if anything, is lower in periods of high interest rate instability.

For acquisitions, empirical observations are broadly in line with our prior expectations. Both UK and US firms are more likely to be acquired during growth phases in the economy than during downturns. Price instability has only a marginal effect on acquisitions in either economy. In both economies, firms are more likely to be acquired during periods of higher long-term real rates of interest. However, unlike the UK, acquisitions in the US are depressed in years when real rates of interest are volatile; similarly, in years when the exchange rate increases sharply.

Firm and industry-level factors

Firm-level covariates and industry dummies are also significant in determining exit rates. While exit rate declines in size at higher ranges in the UK, very large US firms, other than the very young are more likely to exit. As expected, in both economies, bankruptcy is declining in profitability and cash flow. Among US firms and younger firms in the UK, those with higher cash flow are more likely to be acquired.

The age of firms, post-listing, significantly affects exit rates due to both bankruptcy and acquisitions. Plots of the baseline cumulative hazard functions of bankruptcy, for the UK and the US, against the post-listing age of the firm show a convex pattern. This indicates that exit rates due to bankruptcy decline with age (learning effect), after controlling for covariates. In the case of quoted US firms, the baseline hazard for bankruptcy seems to be lower in the first 8 years of post-listing life as compared to ages 8-25 years, before declining again after 25 years. This is consistent with what Evans (1987) and Dunne *et al.* (1989) report for new firms in the US. While the baseline hazard due to acquisitions in the UK appears to be constant over the post-listing lifetime of a firm, this shows a declining trend in the US.

Figures 7-7 and 7-8 also show the year-wise predicted incidence rates of bankruptcies and acquisitions in both economies. By incidence rate, we mean the number of firms that fail as a proportion of total firms in business during that year. The close proximity of the predicted and observed incidence rates indicates the ability of the estimated models to reflect aggregate trends in the number of corporate bankruptcies and acquisitions in the US and the UK.

Further, the Chow χ^2 goodness-of-fit tests strongly reject the null hypothesis of no covariate effect (Table 7.3.1). The tests for validity of the proportional hazard assumption (Bhattacharjee, 2007a; Grambsch and Therneau, 1994) indicated non-proportional effects for several covariates. However, after allowing the effects of these covariates to vary with age of the firm, the test does not reject the null of proportionality. We also test for robustness of the results in several ways; see also Section 7.2. First, we estimate logit models for bankruptcy and acquisition exit with a flexible specification for the age effect. Second, we estimate models with different explanatory variables representing firm and industry specific factors, macroeconomic activity and macroeconomic stability. This includes a very parsimonious model with only one variable for each of the three categories. Third, we check robustness of the age varying covariate effects by changing the intervals over which the effects are assumed to be constant. Finally, we also check for the effect of dependent truncation, by restricting the data to shorter sample periods. The estimates are robust to these various specifications and the estimated models offer very similar inferences.

Jovanovic and Rousseau (2002) explain US merger waves in terms of the availability of profitable capital reallocation opportunities, although their model does not explain the 1960s merger wave well. Shleifer and Vishny (2003) stress the role of stock market misvaluations. Our model predicts all the major merger waves in the US - end of the 1960s, the 1980s and

1990s - fairly well, and provides a macroeconomic explanation.

The impact of Chapter 11

Differences in bankruptcy code in the US and the UK can be one reason for the differential impact of instability on bankruptcies. We argue in Section 7.3.2 that US Chapter 11, which has no correlate in the UK, reduces the impact of instability on bankruptcies in the US. Chapter 11 has a second order effect on acquisitions, by providing a ready supply of acquisition candidates during periods of low instability and high demand for acquired capital. To the extent that Chapter 11 shields businesses from bankruptcy during periods of high macroeconomic instability, the detrimental effect of instability on bankruptcies is lower on firms that follow the Chapter 11 route as compared with those that pass through Chapter 7. If this were true, Chapter 7 bankruptcies, like bankruptcies in the UK, can respond more to the macroeconomic instability than Chapter 11 bankruptcies.

There is a self-selection issue here, in that only unviable firms can be sent on the Chapter 7 route. However, if after conditioning on adequate firm and industry-level covariates the exits through Chapter 7 and Chapter 11 are rendered independent of each other, then partial likelihood inference would be valid. This is very similar to the non-informativeness argument made in hazard regression models with censoring due to competing risks. Thus, so far as the impact of macroeconomic conditions on exits through Chapters 7 and 11 go, we can make adequate inference, conditional on firm and industry-level covariates, if the decision process allocating firms to these two routes depend only on these covariates.

In order to explore this issue further, we incorporate a correction for potential endogenous selection into exits through Chapter 7 or Chapter 11. We first estimate a probit model for the Chapter 7 versus Chapter 11 choice, and then include the estimated inverse Mill's ratios in the hazard regression model as an approximate correction for sample selection. Exclusion restrictions are maintained by including firm level regressors like age of the firm, cash flow, gearing etc. in the probit model. The estimates show that sample selection is important, in that the inverse Mill's ratio is highly significant in the hazard regression models for Chapter 7 and Chapter 11. However, the sample selection corrected estimates are almost identical to uncorrected estimates, both in terms of magnitude and direction of effects.

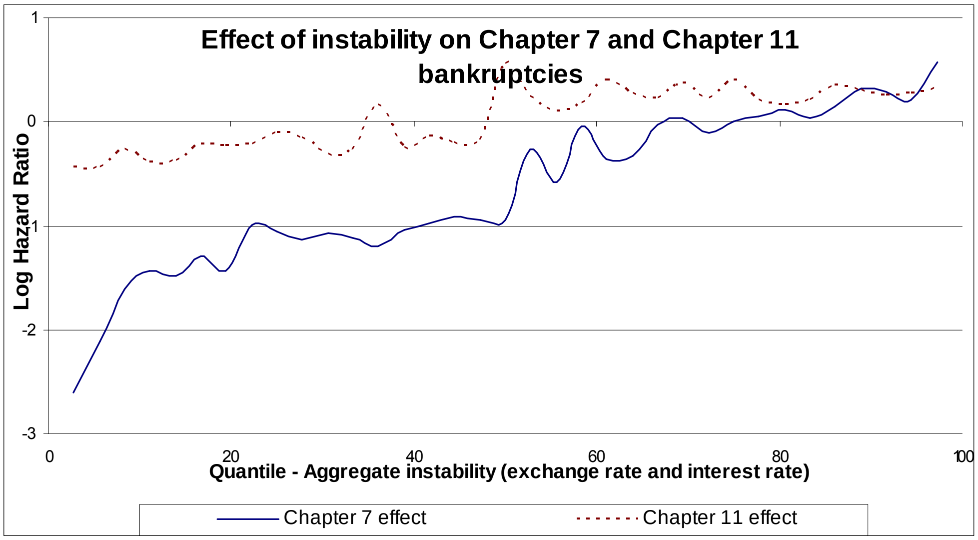


Figure 7-11: Effect of instability on bankruptcies – Chapter 7 and Chapter 11.

We estimate models separately for Chapter 11 and Chapter 7 bankruptcies (Table 7.3.2). The Chapter 11 reorganisation process was instituted in 1979, so all observations on Chapter 11 bankruptcy exits are post 1979. As compared with Chapter 11, Chapter 7 bankruptcies display a higher sensitivity to instability, especially to interest rate and exchange rate volatility. The plot of log-hazard ratios against quantiles of aggregate uncertainty, measured as the linear combination of interest rate and exchange rate volatility that is implied by the estimates for the US bankruptcy model (Figure 7-11) provides further support for this observation. While the hazard of bankruptcy through the Chapter 11 route at the 97th percentile of aggregate uncertainty is only about twice as high as that at the 3rd percentile, the hazard for Chapter 7 bankruptcies at the 97th percentile is 24 times as high as that at the 3rd percentile. For each year, we use these estimated models to predict the proportion of firms that would have failed through the Chapter 11 route as against those failing through Chapter 7. Figure 7-12 shows that the expected number of bankruptcies from Chapter 11 is rather less than those from Chapter 7. A similar test for the effect of Chapter 11 on the number of acquisitions was carried out by estimating models for acquisitions separately for the periods 1969 to 1979, and 1980 to 2000. The estimates for the 1980-2000 period show higher responsiveness to macroeconomic instability, but this difference is not as striking as the difference between Chapter 7 and Chapter

11 bankruptcies.

TABLE 7.3.2: Model Estimates for US Bankruptcy (post-1979)

Variables	**All bankruptcies**	**Chapter 7**	**Chapter 11**
Industry Dummies			
(not reported)			
Firm × Year Level			
Size = s			
$- s \times I_1$	0.767 (-0.6)	0.610 (-1.1)	2.112 (0.8)
$- s \times I_2$	0.084 (-5.6)**	0.146 (-3.7)**	0.036 (-3.6)**
$- s \times I_3$	0.064 (-5.6)**	0.037 (-4.6)**	0.129 (-2.9)**
$- s \times I_4$	0.132 (-5.8)**	0.156 (-4.6)**	0.073 (-3.4)**
Size-squared = s^2			
$- s^2 \times I_1$	1.005 (0.0)	1.092 (0.6)	0.594 (-0.7)
$- s^2 \times I_2$	1.547 (5.3)**	1.338 (2.1)*	1.829 (4.0)**
$- s^2 \times I_3$	1.532 (4.0)**	1.642 (2.8)**	1.338 (1.7)$^+$
$- s^2 \times I_4$	1.300 (3.5)**	1.262 (2.4)*	1.447 (2.7)**
Retn. on cap. empl.	1.000 (-4.0)**	1.000 (-3.7)**	1.000 (-3.5)**
Macro- Conditions			
Output gap	5.271 (0.5)	0.672 (-0.1)	504.1 (1.0)
Long-term real			
int. rate = r			
$- r \times I_1$	1.309 (3.8)**	1.307 (3.3)**	1.360 (2.2)*
$- r \times I_2$	0.916 (-1.3)	0.886 (-1.4)	0.960 (-0.4)
$- r \times I_3$	0.905 (-1.1)	0.820 (-1.6)	0.941 (-0.5)
$- r \times I_4$	0.865 (-1.1)	0.732 (-1.7)$^+$	0.919 (-0.4)
Exchange rate = e			
$- e \times I_1$	0.942 (-5.5)**	0.939 (-4.9)**	0.952 (-2.6)**
$- e \times I_2$	0.956 (-8.7)**	0.962 (-6.6)**	0.943 (-5.6)**
$- e \times I_3$	0.976 (-3.8)**	0.980 (-2.5)*	0.965 (-3.0)**
$- e \times I_4$	0.997 (-0.6)	0.999 (-0.2)	0.976 (-1.7)$^+$

TABLE 7.3.2: Model Estimates for US Bankruptcy (post-1979) (Contd.)

Variables	All bankruptcies	Chapter 7	Chapter 11
Macro- Instability			
Turnaround	0.042 (-2.2)*	0.052 (-1.7)$^{+}$	0.043 (-1.3)
y-o-y incr. in exch. rate	0.998 (-0.1)	1.017 (0.6)	0.963 (-0.9)
Volatility - Long term			
int. rate $= l$			
$- l \times I_1$	1.015 (0.3)	1.037 (0.6)	1.008 (0.1)
$- l \times I_2$	0.883 (-3.0)**	0.900 (-2.0)*	0.832 (-2.7)**
$- l \times I_3$	0.907 (-2.3)*	0.908 (-1.9)$^{+}$	0.885 (-1.6)
$- l \times I_4$	0.866 (-2.4)*	0.930 (-1.0)	0.697 (-2.9)**
Volatility - Short term			
int. rate	1.014 (1.8)$^{+}$	1.027 (2.8)**	0.991 (-0.7)
No. of firms	12,596	12,596	12,596
No. of exits	490	321	169
Total time at risk (in yrs.)	100,487	100,487	100,487
Log-likelihood	-3498.28	-2280.62	-1192.96
χ^2 test – PH assumption	15.09	16.77	14.09
degrees of freedom	33	33	33
p-value	0.9968	0.9914	0.9984
Chow test –			
parameter stability			49.40
degrees of freedom			33
p-value			0.0332

z-scores in parentheses.

Parameters reported are hazard ratios (exponential of the regression coefficient estimates).

I_1, I_2, I_3 and I_4 represent the indicator functions I(age 0-8 yrs.), I(age 9-16 yrs.), I(age 17-25 yrs) and I(age $>$ 25 yrs) respectively;

Volatility is measured as maximum monthly difference during the year, divided by the no. of intervening mths.

** , *and $^{+}$– Significant at 1%, 5% and 10%level respectively.

Table 7.3.2 also includes a Chow-like test of parameter stability, for the post-1980 period, that the covariate effects are the same for Chapter 7 and Chapter 11 bankruptcies. This test is based on the maximised partial log-likelihoods for the three models reported in the Table.

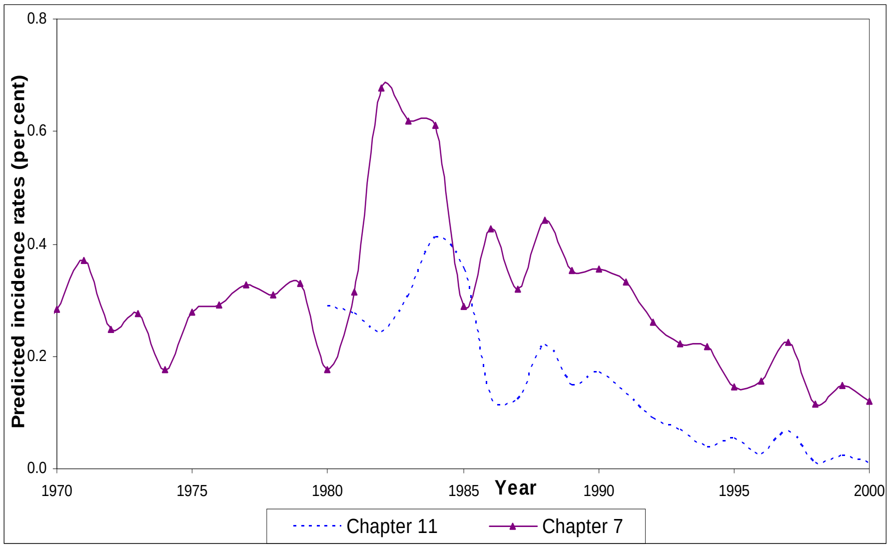


Figure 7-12: Predicted Incidence rates – Chapter 7 and Chapter 11 bankruptcies.

The results show that the null hypothesis of parameter stability is rejected at the 5 percent level of significance, indicating that the various explanatory factors (other than age post-listing) have different effects on the hazard of exit through the Chapter 7 and Chapter 11 routes. Of particular relevance are the effects of the macroeconomic factors, encompassing measures of both the level and instability in the macroeconomic environment. The estimates reinforce the hypothesis that Chapter 11 exits have a lower response to the macroeconomic environment than Chapter 7 bankruptcies. This is particularly evident from the effect of interest rate as well as its volatility.

In summary, we find evidence that the differences in the impact of macroeconomic instability on bankruptcy hazard can be attributed, in significant measure, to the difference between the Chapter 7 and Chapter 11 route. In other words, the legal protection afforded under Chapter 11 in the US appears to reduce the adverse effect of macroeconomic instability on bankruptcies.

7.3.5 Conclusions

In this Section, we examined how macroeconomic instability affects a firm's risk of bankruptcy and liquidation. We developed and tested a model in which a firm's bankruptcy and acquisition hazards are codetermined by firm-level and sector-level factors, and by macroeconomic conditions. We estimate the model on a panel containing well over thirty years of data for US listed firms and covering several business cycles, using a competing risks hazard regression framework. To examine the effect of the legal system on liquidation we also estimate the model for UK listed firms over the same period. The UK is an economy that is institutionally similar to the US, but without any equivalent system to Chapter 11. Effectively, bankruptcy leads directly to liquidation in the UK, whereas bankruptcy has a binary outcome in the US - a failing firm can be liquidated under Chapter 7 or reorganised under Chapter 11.

Like Section 7.2, the framework developed here for order restricted inference on covariate effects is found to be useful. We find that macroeconomic conditions have a significant impact on bankruptcy and acquisition hazard. However, while the impact of instability on bankruptcy is strong in the UK it is much weaker in the US. When we partition US bankruptcies into Chapter 7 and Chapter 11 we find that the difference in responsiveness to macroeconomic instability is largely attributable to the use of reorganisation under Chapter 11.

7.4 Empirics of firm dynamics: modeling the role of frailty

As discussed in Section 7.1, theoretical models of firm dynamics point to a potentially important role for frailty in empirical studies of firm exits. This is primarily related to unobserved heterogeneity in the measurement of initial efficiency, or founding conditions of the firm, in terms of physical capital and intangibles, but particularly unmeasured human capital. The applications in Sections 7.2 and 7.3 address this issue by assuming common shared frailty between the competing risks of bankruptcy and acquisitions. Under this setup, the applications highlight an important role of order restrictions in the nature of covariate effects of macroeconomic instability, as well as negative ageing in the shape of the baseline hazard function.

However, the literature has highlighted the fact that inference on covariate effects are not robust when frailty is not appropriately modeled, and vice versa; see, for example, Andersen *et al.* (1993) and Aalen (1994). In fact, the connection between frailty and nonproportional covariate effects is brought into sharper focus in Abbring and van den Berg (2007), where the test proposed by Gill and Schumacher (1987) for testing the proportional hazards assumption against convex ordering in two samples is modified to test for frailty. It is, therefore, important to ensure that frailty is appropriately modeled before credible inferences on nonproportional covariate effects can be made.

In this Section, we empirically investigate the role of unrestricted univariate frailty in hazard regression models of firm exit when the covariates are allowed to have time varying coefficients. Based on Bhattacharjee (2007c), we model frailty explicitly in two ways. First, we consider a grouped time proportional hazards model, where frailty is modeled by a sequence of discrete mixture distributions with increasing number of mass points (Heckman and Singer, 1984a). Second, we use a continuous failure time Cox regression model with time varying coefficients, with Gamma distributed frailty shared by firms with similar founding conditions.

For our empirical work, we consider two applications. The first is on new French firms, where the objective is to study the effect of entrepreneural human capital and physical endowments on the survival of these firms. Our second application is on firm exits due to bankruptcy for quoted firms in the UK, where (like Sections 7.1 and 7.2) the main objective of analysis is to understand the effect of macroeconomic instability on business failure.

7.4.1 New French firms

The application is based on the model of entrepreneural choice with labour market imperfections briefly discussed in Section 7.1.2. Human capital of entrepreneurs is imperfectly observed, mainly through education and experience. The expression of human capital is also conditioned on the prior state of employment – employed in other firm with same or different activity or unemployed.

The SINE 94/97 database of new French firms established in 1994 is used for our analysis; see Section 1.3.7 for further discussion on these data. Firm characteristics in the database are matched up with individual characteristics of the entrepreneurs; see Bhattacharjee *et al.* (2006) for details. The control variables characterising the entrepreneur are: previous occupation, previous employment status, age, whether belonging to an entrepreneurial "milieu", experience in managerial activities, previous experience in setting up a firm and main motive for the creation. The variables representing the firm are amount of money invested, initial size, receipt of public financial aid, requesting and obtaining bank loans, number of customers, legal status, whether based in highly productive French regions and branch of activity.

Survival of the firm depends on unobserved human capital, in addition to characteristics of the firm and financing constraints. Frailty is likely to be particularly important at the individual level. modeling frailty using a discrete mixture has a natural interpretation in this context. Entrepreneurs with high unobserved human capital are those whose observed human capital falls well below the actual level. By contrast, there are other entrepreneurs whose human capital is valued reasonably appropriately by the labour market, and therefore have low frailty, but who perceive their actual human capital to be higher.

In addition to frailty, covariate effects of firm and entrepreneur variables can have nonproportional covariate effects. These order restrictions are identified using a sequential application of the tests proposed in Chapter 5 (Bhattacharjee, 2007b) with grouped failure time data and discrete mixture frailty. The estimates of three competing models are presented in Table 7.4.1.

The results show evidence of frailty with two support points when covariates are assumed to have proportional hazard effects. An LR test rejects the null hypothesis of "no frailty" at the 1% level of significance. As discussed above, this model has some interesting interpretations which support important economic inferences.

However, testing for proportionality in the presence of arbitrary univariate frailty (Chapter 5; Bhattacharjee, 2007b) points to order restricted covariate effects for several explanatory variables. Once these covariates are allowed to have nonproportional effects (time varying coefficients), there is no longer any evidence of frailty.[21] This observation has the important implication that, in some applications, nonproportional or order restricted covariate effects can be misinterpreted as frailty.

[21] Similar inferences are obtained in a model with frailty shared by entrepreneurs having different combinations of experience and education.

TABLE 7.4.1: ESTIMATES FOR GROUPED TIME PH MODEL

Variables	**No frailty**	**Discrete mix. frailty**	**Non-proportional, no frailty**
LOG BASELINE HAZARD			
– Year 1	−4.993** (−52.2)	−17.227 (−0.6)	−4.667** (−43.2)
– [Year 2 – Year 1]	0.283** (7.0)	0.353** (8.2)	−0.084 (−1.0)
– [Year 3 – Year 1]	0.506** (12.6)	0.646** (12.9)	0.084 (0.9)
– [Year 4 – Year 1]	0.351** (7.7)	0.595** (8.5)	−0.441** (−3.6)
EMPLOYMENT X EDUCATION			
(Base) Same branch, High education	0.00	0.00	0.00
Intermediate education	0.315** (4.9)	0.315** (4.5)	0.330** (5.1)
Low education	0.616** (5.8)	0.602** (5.2)	0.622** (5.9)
Different branch, High education	0.622** (6.4)	0.664** (5.9)	
– $\times I$ [Year 1]			1.007** (7.2)
– $\times I$ [Year 2]			0.797** (5.2)
– $\times I$ [Year 3]			0.250 (1.3)
– $\times I$ [Year 4]			−0.143 (−0.5)
Intermediate education	0.523** (6.4)	0.539** (5.8)	0.539** (6.6)
Low education	0.626** (4.5)	0.821** (4.2)	0.680** (4.9)
Short term unemployed, High education	0.625** (8.1)	0.684** (8.0)	0.637** (8.3)
Intermediate education	0.661** (10.0)	0.710** (9.6)	0.667** (10.0)
Low education	0.692** (7.1)	0.767** (6.8)	0.712** (7.3)
Long term unemployed, High education	0.538** (5.6)	0.614** (5.5)	0.538** (5.6)
Intermediate education	0.749** (10.4)	0.839** (10.3)	0.764** (10.6)
Low education	0.892** (7.9)	0.939** (7.2)	0.897** (7.9)

TABLE 7.4.1: Estimates for Grouped Time PH Model (Contd.)

Variables	No frailty	Discrete mix. frailty	Non-proportional, no frailty
Human Capital			
Experience, same branch			
(Base = 3 – 10 years)	0.00	0.00	0.00
– Less than 3 years	0.180** (3.8)	0.167** (3.1)	0.184** (3.9)
– More than 10 years	−0.290** (−7.4)	−0.331** (−7.5)	−0.290** (−7.4)
Size, prev. same branch firm			
(Base = 10 – 100 employees)	0.00	0.00	0.00
– Less than 10 employees	−0.361** (−9.7)	−0.414** (−9.9)	−0.359** (−9.6)
– More than 100 employees	0.077 (1.6)	0.085 (1.5)	0.082+ (1.7)
Previous professional status			
(Base = Worker)	0.00	0.00	0.00
– Manager/ Executive	−0.086* (−2.0)	−0.103* (−2.1)	−0.082* (−2.0)
– Craftsman/ Middle mgmt.	−0.014 (−0.3)	−0.018 (−0.4)	−0.011 (−0.2)
– Student	0.156* (2.1)	0.271** (3.0)	0.168* (2.3)
Previous setting up of new firms			
(Base = Once or more)	0.00	0.00	0.00
– Never	0.234** (5.7)	0.256** (5.6)	0.226** (5.6)
Entrepreneur Attributes			
Age of entrepreneur (Base = 35 – 40 years)	0.00	0.00	0.00
– 25 – 35 years	0.021 (0.5)	−0.006 (−0.1)	
– $\times I$ [Year 1]			−0.241** (−3.2)
– $\times I$ [Year 2 or Year 3]			0.001 (0.0)
– $\times I$ [Year 4]			0.558** (5.1)
– 40 – 50 years	−0.072+ (−1.7)	−0.100* (−2.0)	
– $\times I$ [Year 1]			−0.442** (−5.4)
– $\times I$ [Year 2 or Year 3]			−0.022 (−0.4)
– $\times I$ [Year 4]			0.386** (3.4)

TABLE 7.4.1: ESTIMATES FOR GROUPED TIME PH MODEL (CONTD.)

Variables	**No frailty**	**Discrete mix. frailty**	**Non-proportional, no frailty**
ENTREPRENEUR ATTRIBUTES			
Entrepreneurship "milieu"			
(Base = Yes)	0.00	0.00	0.00
– No (relatives/close reltns.)	0.100^{**} (3.1)	0.130^{**} (3.4)	
– $\times I$ [Year 1 or Year 2]			0.219^{**} (5.0)
– $\times I$ [Year 3 or Year 4]			−0.053 (−1.1)
Main motivation			
(Base = Unemployed)	0.00	0.00	0.00
– New idea	−0.084 (−1.3)	-0.147^{*} (−2.0)	−0.074 (−1.2)
– Opportunity/ Taste for entrepreneurship	-0.168^{**} (−3.8)	-0.215^{**} (−4.2)	-0.166^{**} (−3.7)
– Entourage example	0.125 (1.5)	0.066 (0.7)	
– $\times I$ [Year 1]			-0.692^{**} (−3.3)
– $\times I$ [Year 2]			0.210 (1.6)
– $\times I$ [Year 3]			0.286^{*} (2.2)
– $\times I$ [Year 4]			0.624^{**} (4.0)
FIRM ATTRIBUTES			
Initial size of enterprise			
(Base = Max. 1 employee)	0.00	0.00	0.00
– More than one employee	0.157^{**} (4.5)	0.114^{**} (2.8)	
– $\times I$ [Year 1]			-0.623^{**} (−8.2)
– $\times I$ [Year 2]			0.170^{**} (2.8)
– $\times I$ [Year 3 or Year 4]			0.495^{**} (10.7)
Initial demand			
(Base = > 10 customers)	0.00	0.00	0.00
– Between 1 – 10 customers	0.133^{**} (4.0)	0.137^{**} (3.6)	0.131^{**} (3.9)

TABLE 7.4.1: ESTIMATES FOR GROUPED TIME PH MODEL (CONTD.)

Variables	**No frailty**	**Discrete mix. frailty**	**Non-proportional, no frailty**
FIRM ATTRIBUTES			
Legal status (Base = Unlimited liability)	0.00	0.00	0.00
– Limited liability	-0.392^{**} (-10.0)	-0.361^{**} (-8.1)	-0.384^{**} (-9.8)
Region of incorporation			
(Base = High entrepreneurship)	0.00	0.00	0.00
– Low entrepreneurship	-0.068^{*} (-2.2)	-0.049 (-1.4)	-0.070^{*} (-2.3)
Industry (Base = Services)	0.00	0.00	0.00
– Catering/ Trade	0.322^{**} (8.0)	0.392^{**} (8.2)	
– $\times I$ [Year 1]			0.524^{**} (8.1)
– $\times I$ [Year 2 or Year 3]			0.269^{**} (5.5)
– $\times I$ [Year 4]			0.115 (1.4)
– Manufacturing	-0.075 (-1.4)	-0.042 (-0.7)	-0.085 (-1.6)
– Construction/ Transport	-0.272^{**} (-5.9)	-0.283^{**} (-5.6)	-0.278^{**} (-6.0)
FINANCING CONSTRAINTS			
Initial capital invested			
(Base = 15245 – 76224 Euros)	0.00	0.00	0.00
– less than 15245 Euros	0.343^{**} (8.3)	0.384^{**} (8.3)	0.331^{**} (8.0)
– more than 76224 Euros	-0.502^{**} (-5.2)	-0.522^{**} (-5.2)	-0.506^{**} (-5.3)
Public financial aid, 1994			
(Base = None)	0.00	0.00	0.00
– Obtained aid	-0.346^{**} (-9.1)	-0.411^{**} (-9.2)	-0.338^{**} (-8.9)
Bank loans, 1994			
(Base = Not applied)	0.00	0.00	0.00
– Applied and refused	0.098 (1.6)	0.171^{*} (2.3)	
– $\times I$ [Year 1 or Year 2]			-0.031 (-0.3)
– $\times I$ [Year 3 or Year 4]			0.259^{**} (3.0)
– Applied and received	-0.299^{**} (-7.6)	-0.354^{**} (-8.0)	-0.302^{**} (-7.7)

TABLE 7.4.1: ESTIMATES FOR GROUPED TIME PH MODEL (CONTD.)

Variables	**No frailty**	**Discrete mix. frailty**	**Non-proportional, no frailty**
MIXTURE FRAILTY DISTBN.			
$m_1 \equiv 0$	–	0.00	–
m_2	–	12.44 (0.5)	–
π_1	–	0.186^{**} (6.3)	–
$\pi_2 = 1 - \pi_1$	–	0.814^{**} (27.5)	–
No. of firms	19,213	19,213	19,213
No. of exits	7,882	7,882	7,882
Log-likelihood	−24593.0	−24583.7	−24422.8

z-scores in parentheses.

**, *and $^{+}$– Significant at 1%, 5% and 10% level respectively.

The Heckman and Singer (1984a) sequential procedure supports a two-point discrete mixture frailty distribution. However, with time varying coefficients, this procedure supports a model with no frailty.

7.4.2 Quoted UK firms

Our second application is based on data on quoted firms in the UK, discussed in Sections 7.2 and 7.3. Since the main implications with regard to nonproportional effects of macroeconomic instability and other covariates are similar to those discussed earlier, we focus our current discussion only on the issue of frailty.

As emphasized in Section 7.1, and discussed in the context of the French data, frailty resulting from incompletely observed founding conditions can be potentially important in these applications. Unfortunately, very little data were available on entrepreneural human and physical capital for these firms at the time of entry. However, listing dates were known, and it is likely that founding environmental conditions may be important for survival of these firms. Therefore, we investigated the role of railty by estimating a model incorporating Gamma distributed frailty shared by firms listed on the same year. The firm, industry and macroeconomic covariates were allowed to have nonproportional effects.

The null hypothesis of "no frailty" was rejected at the 5% level of significance. The model estimates were somewhat different from the comparable estimated model without frailty, but substantive inferences were similar. Thus, this application highlights the importance of modeling frailty apropriately, even when the main object of analysis are the covariate effects.

7.4.3 Conclusions

In summary, the two applications above point to the importance of modeling frailty jointly with potentially nonproportional covariate effects in hazard regression models. While apparent nonproportionality in covariate effects can provide evidence for frailty (Abbring and van den Berg, 2007), assumptions of "no frailty" or proportional hazards can both be violated in applications. This may lead to incorrect and misleading inferences about the nature of frailty and regression coefficients, but also the nature of ageing. The time varying coefficients model, with either unrestricted or order rstricted baseline hazard function, and with appropriately modeled frailty, is an useful framework for analyses of such applications. modeling frailty appropriately within the context of the application considered, and with reference to relevant theory is very important.

Chapter 8

Conclusion

In this thesis, we developed methods for inference on nonproportional covariate effects in hazard regression models, where the effects are potentially order restricted. The use of the proposed methods are illustrated with several applications, mainly from biomedicine and economics. In the following sections, we highlight the main contributions of the thesis, followed by a discussion of potential scope for improvement and directions of future research.

8.1 Contributions of the thesis

The starting point of our work was the monotone hazard ratio alternative to the proportional hazards specification discussed in Gill and Schumacher (1987), which is equivalent to convex ordering of two lifetime distributions (Sengupta and Deshpande, 1994). Tests for the proportional hazards assumption against such order restricted alternatives in two samples were developed in Gill and Schumacher (1987) and Deshpande and Sengupta (1995).

In Chapter 2, we develop alternative tests for proportional hazards in two samples against the weaker alternative where the ratio of cumulative hazard functions is monotone (Sengupta *et al.*, 1998), and extend two sample tests against alternatives positing monotone ratio of hazard and cumulative hazard functions to the competing risks problem with potentially dependent causes of failure (Bhattacharjee and Sengupta, 1994).

In Chapter 3 (Bhattacharjee, 2007a), we extend the notions of partial orders in two samples to a continuous covariate setup. Specifically, these notions posit relative ageing of the lifetime distribution conditional on a higher value of the covariate, as compared with a lower covariate value. Tests for the proportional hazards assumption, with respect to a continuous covariate, against such partial orders are developed. The tests are powerful not only against the above ordered alternatives, but also changepoint situations where the direction of ordering is potentially different in various regions of the sample space. Sequential and joint testing is discussed when there are multiple covariates, and the tests are extended to models with shared frailty or univariate frailty from a one parameter family.

Based on Bhattacharjee (2003), we draw a link between the above ordered alternatives and a hazard regression model with time varying coefficients in Chapter 4. After motivating the relevance of the time varying coefficients model with monotone covariate effect in this situation, we follow Bhattacharjee (2004) in discussing estimation under the hypothesized order restrictions on covariate effects, using biased bootstrap methods like data tilting and local adaptive bandwidths.

In Chapter 5 (Bhattacharjee, 2007b), we extend the tests developed in Chapter 3 to the case when there is univariate frailty with arbitrary and unrestricted distribution. We show that, in this case, the testing problem is closely related to the problem of testing for absence of covariate dependence. We develop inference procedures for both these testing problems, and highlight issues relating to their practical implementation.

Following Bhattacharjee and Bhattacharjee (2007), Chapter 6 develops Bayesian inference on covariate dependence when the covariate effects are potentially order restricted. This approach has some attractive features, including the incorporation of prior beliefs on order restrictions in the specification of the prior distributions, in joint inference under order restrictions both on covariate effects and on the shape of the baseline hazard function (ageing), in accommodating potentially unrestricted frailty, and in addressing parameter uncertainty in an appropriate way.

Finally, in Chapter 7, we illustrate the use of the proposed methods in deriving useful economic inferences in empirical applications specific to the area to firm dynamics (industrial organisation). Following Bhattacharjee *et al.* (2008a), we study the effect of macroeconomic instability on survival of quoted firms in the UK. We observe evidence of monotone covariate effects and ageing patterns in the shape of the baseline hazard function. Similar evidence is observed for US firms (Bhattacharjee *et al.*, 2008a), but the somewhat weaker covariate effects in this case can be related to changes in the legal regulatory environment. Following Bhattacharjee (2007c), we also highlight the fact that frailty is likely in these models, and should be appropriately modelled before robust inferences on order restricted covariate effects can be drawn.

Overall, the framework developed in the thesis for modelling the nature of potentially non-proportional or order restricted covariate effects, and the methods of testing and estimation, are useful in a wide range of applications. The usefulness of the analytical methods developed in our work go well beyond the scope of the current applications and models studied here. Standard counting process approaches (see, for example, Andersen *et al.*, 1993) are useful for deriving asymptotic results for statistics of the form $\sum_{i=1}^{n} \int K_i(t).dM_i(t)$,where $M_i(.)$ are martingales and $K_i(.)$ are predictable processes. By contrast, we derive inferences for statistics like $\sum_{i=1}^{n} \int K_i(t).M_i(t).dt$,where the $K_i(.)$ and $M_i(.)$ $(i = 1, \ldots, n)$ are iid copies of stochastic processes, and for statistics $\sum_{i=1}^{n} \int K_i(t).H(t).dt$,where $H(.)$ involves data from all the n observations. Some of the proposed methods have been used in other contexts, for example in the work of Dauxois and Kirmani (2003, 2004, 2005), Kirmani and Dauxois (2003), Sun *et al.* (2007) and Alvarez-Andrade *et al.* (2007a). Combined with the work in Lin *et al.* (2000) and Lin and Ying (2001), our work, extends inference methods to a wide range of statistics.

8.2 Limitations and future work

In our view, the main issue in the practical applicability of the work developed in this thesis lies not in the methods proposed here, but in the lack of appropriate inference procedures for hazard regression models with unrestricted frailty. This is currently an area of active research, and new and useful methods will be developed in the future.

Other potentially useful approaches for testing the proportional hazards assumption against ordered alternatives have been discussed in the thesis. In the two sample case (Chapter 2), further work along the lines of Breslow (1974) and Dabrowska *et al.* (1989, 1992) may be useful. Similarly, in the continuous covariate case, further research may focus on inferences in changepoint problems and on joint testing and modelling for multiple covariates along the lines discussed in Chapter 3. In Chapter 4, we consider the biased bootstrap approach for estimation of our models under order restrictions. Alternative promising approaches based, for example, on taut strings or density regression approach have considerable potential. The issue of joint inference under order restrictions on both covariate dependence and ageing in a classical setup also requires further attention. In addition to new work on inference in frailty models with unrestricted univariate frailty (Chapter 5), further research on calibration of the Horowitz (1996, 1999) estimator using backfitting or bootstrap based approaches may be useful. Finally, our proposed Bayesian methods (Chapter 6) can be further developed along the lines developed in recent research on Bayesian inference under order restrictions (Dunson and Waddala, 2007; Gunn and Dunson, 2007; Wang and Dunson, 2007).

In summary, this thesis develops a new framework for inference under order restrictions on covariate effects in hazard regression models and proposes several inference procedures. The methods developed are useful in many applications. However, our work also highlights the need for further research along several directions.

Bibliography

[1] Aalen, O.O. (1975). *Statistical inference for a family of counting processes.* PhD Thesis, University of California, Berkeley, CA.

[2] Aalen, O.O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics* **6**, 701–726.

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

[4] Aalen, O.O. (1994). Effects of frailty in survival analysis. *Statistical Methods in Medical Research* **3**, 227–243.

[5] Abbring, J.H. and van den Berg, G.J. (2007). The unobserved heterogeneity distribution in duration analysis. *Biometrika* **94**(1), 87–99.

[6] Abrahamowicz, M., MacKenzie, T. and Esdaile, J.M. (1996). Time-dependent hazard ratio: modelling and hypothesis testing with application in Lupus Nephritis. *Journal of the American Statistical Association* **91**, 1432–1439.

[7] Ahmad, I.A. (2001). Moments inequalities of aging families of distributions with hypotheses testing applications. *Journal of Statistical Planning and Inference* **92**, 121–132.

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

[9] Alvarez-Andrade, S., Balakrishnan, N. and Bordes, L. (2007a). Homogeneity tests based on several progressively Type-II censored samples. *Journal of Multivariate Analysis* **98**, 1195–1213.

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

[11] Andersen, P.K. (1982). Testing for the goodness-of-fit of Cox's regression model. *Biometrics* **38**, 67–77.

[12] Andersen, P.K. (1983). Comparing survival distributions via hazard ratio estimates. *Scandinavian Journal of Statistics* **10**, 77–85.

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

[14] Andersen, P.K., Bentzon, M.W. and Klein, J.P. (1996). Estimating the survival function in the proportional hazards regression model: a study of the small sample size properties. *Scandinavian Journal of Statistics* **23**, 1–12.

[15] Andersen, P.K., Borgan, Ø., Gill, R.D. and Keiding, N. (1982). Linear nonparametric tests for comparison of counting processes, with application to censored data. *International Statistical Review* **50**(2), 219–258.

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

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

[18] Andersen, P.K., Klein, J.P. and Zhang, M.J. (1999). Testing for centre effects in multi-centre survival studies: a monte carlo comparison of fixed and random effects tests. *Statistics in Medicine* **18**, 1489–1500.

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

[20] Arjas, E. (1988). A graphical method for assessing goodness of fit of Cox's proportional hazard model. *Journal of the American Statistical Association* **83**, 204–212.

[21] Arjas, E. and Bhattacharjee, M. (2003). Modelling heterogeneity: Hierarchical Bayesian approach. In Mazzuchi, T.A., Singpurwalla, N.D. and Soyer, R. (Eds.), *Mathematical Reliability: An Expository Perspective*, Kluwer Academic Publishers.

[22] Arjas, E. and Gasbarra, D. (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica*, **4**, 505–524.

[23] Arjas, E. and Gasbarra, D. (1996). Bayesian inference of survival probabilities, under stochastic ordering constraints. *Journal of the American Statistical Association* **91**, 1101–1109.

[24] Arjas, E. and Haara, P. (1987). A logistic model for hazard: asymptotic results. *Scandinavian Journal of Statistics* **14**, 1–18.

[25] Armour, J., Cheffins, B.R., Skeel, D.A. Jr. (2002). Corporate ownership structure and the evolution of bankruptcy law: lessons from the UK. *Vanderbilt Law Review* **55**, 1699–1785.

[26] Asplund, M. and Nocke, V. (2003). Firm turnover in imperfectly competitive markets. *PIER Working Paper* **03-010**, University of Pennsylvania.

[27] Atkinson, A., Gomulka, J., Mickelwright, J. and Rau, N. (1984). Unemployment benefits, duration, and incentives in Britain: how robust is the evidence? *Journal of Public Economics* **23**, 3–26.

[28] Bagai, I., Deshpande, J.V. and Kochar, S.C. (1989a). A distribution-free test for equality of failure rates due to two competing risks. *Communications in Statistics - Theory and Methods* **18**, 107–120.

[29] Bagai, I., Deshpande, J.V. and Kochar, S.C. (1989b). Distribution-free tests for stochastic ordering in the competing risks model. *Biometrika* **76**, 775–781.

[30] Bagdonavičius, V.B. (1978). Testing the hypothesis of the additive accumulation of damages. *Probability Theory and its Applications* **23**(2), 403–408.

[31] Bagdonavičius, V.B. and Nikulin, M.S. (1999). Generalized proportional hazards models based on modified partial likelihood. *Lifetime Data Analysis* **5**, 329–350.

[32] Bagdonavičius, V.B. and Nikulin, M.S. (2004). Statistical modeling in survival analysis and its influence on the duration analysis. In: Balakrishnan, N. and Rao, C.R. (Eds.) *Handbook of Statistics 23: Advances in Survival Analysis*, North-Holland: Amsterdam, 411–429.

[33] Baird D.G. and Rasmussen, R.K. (2003). Chapter 11 at twilight. *Stanford Law Review* **56**, 673–699.

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

[35] Baltazar-Aban, I. and Peña, E. (1995). Properties of hazard-based residuals and implications in model diagnostics. *Journal of the American Statistical Association* **90**, 185–197.

[36] Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). *Inference under order restrictions*. Wiley: London.

[37] Barlow, R.E. and Proschan, F. (1975). *Theory of Reliability and Life Testing*. Holt, Rinehart and Winston Inc.: New York.

[38] Bebchuk, L.A. (2002). Ex ante costs of violating absolute priority in bankruptcy. *Journal of Finance* **57**, 445–460.

[39] Begg, C.B., McGlave, P.B., Bennet, J.M., Cassileth, P.A. and Oken, M.M. (1984). A critical comparison of allogenic bone-marrow transplantation and conventional chemotherapy as treatment for acute non-lymphomytic leukemia. *Journal of Clinical Oncology* **2**, 369–378.

[40] Begun, J.M. and Reid, N. (1983). Estimating the relative risk with censored data. *Journal of the American Statistical Association* **78**, 337–341.

[41] Bergin, J. and Bernhardt, D. (2004). Comparative learning dynamics. *International Economic Review* **45**, 431–465.

[42] Bergin, J. and Bernhardt, D. (2006). Industry dynamics with stochastic demand. Queen's Economics Department Working Paper No. **1043**, Queen's University, Canada.

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

[44] Bernanke, B. and Gertler, M. (1989). Agency costs, net worth and business fluctuations. *American Economic Review* **79**, 14–31.

[45] Bernanke, B., Gertler, M. and Gilchrist, S. (1996). The financial accelerator and the flight to quality. *Review of Economics and Statistics* **78**, 1–15.

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

[47] Bhattacharjee, A. (2003). Monotone departures from proportional hazards with respect to continuous covariates: Inference procedures and applications. Presented at the *54^{th} Biennial Session of the International Statistical Institute*, Berlin.

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

[49] Bhattacharjee, A. (2007a). Testing the Proportional Hazards Model with Continuous Covariates in Duration Models against Monotone Ordering. *Mimeo.* Previous version circulated as: Bhattacharjee, A. and Das, S. (2001), Paper presented at the 56^{th} *European Meeting of the Econometric Society*, Lausanne (Switzerland) and DAE Working Paper No. **0220**, Department of Applied Economics, University of Cambridge, UK.

[50] Bhattacharjee, A. (2007b). A simple test for the absence of covariate dependence in hazard regression models. Presented at the *5^{th} International Mathematical Methods in Reliability Conference (MMR2007)*, Glasgow. Previous version circulated as: Bhattacharjee, A. (2004b). A simple test for the absence of covariate dependence in duration models. Cambridge Working Papers in Economics **0401**, Faculty of Economics, University of Cambridge, UK.

[51] Bhattacharjee, A. (2007c). Models of firm dynamics and the hazard rate of exits: is unobserved heterogeneity really that important? *Mimeo.* Previous version circulated as:

Bhattacharjee, A. (2005). Models of firm dynamics and the hazard rate of exits: reconciling theory and evidence using regression models. CRIEFF Discussion Paper No. **0502**, Centre for Research in Industry, Enterprise, Finance and the Firm, University of St. Andrews, UK.

[52] Bhattacharjee, A. and Bhattacharjee, M. (2007). Bayesian analysis of hazard regression models under order restrictions on covariate effects and ageing. Presented at the *5th International Mathematical Methods in Reliability Conference (MMR2007)*, Glasgow. *Mimeo.*

[53] Bhattacharjee, A., Bonnet, J., Le Pape, N. and Renault, R. (2006). Inferring the unobserved human capital of entrepreneurs. Working Paper **200603**, Center for Research in Economics and Management (CREM), University of Rennes 1, University of Caen and CNRS, France.

[54] Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2008a). Macroeconomic conditions and business exit: determinants of failures and acquisitions of UK firms. *Economica* (forthcoming). Previous version circulated as: DAE Working Paper No. **0206**, Department of Applied Economics, University of Cambridge, UK, 2002.

[55] Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2008b). Macroeconomic instability and corporate failure. *Review of Law and Economics* (forthcoming). Previous version circulated as: Business failure in US and UK quoted firms: impact of macroeconomic instability and the role of legal institutions. Cambridge Working Papers in Economics No. **0420**, Faculty of Economics, University of Cambridge, UK, 2004.

[56] Bhattacharjee, M., Arjas, E. and Pulkkinen, U. (2003). Modelling heterogeneity in nuclear power plant valve failure data. In Lindqvist, B. and Doksum, K. (Eds.), *Mathematical and Statistical Methods in Reliability*, World Scientific Publishing.

[57] Bickel, P.J. (2007). Discussion of Zeng, D. and Lin, D.Y. (2007), "Maximum likelihood estimation in semiparametric regression models", *Journal of the Royal Statistical Society*, Series B **69**, 546–547.

[58] 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.

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

[60] Bowman, A.W., Jones, M.C. and Gijbels, I. (1998). Testing monotonicity of regression. *Journal of Computational and Graphical Statistics* **7**, 489–500.

[61] Breiman, L. and Friedman, J.H. (1985). Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association* **80**, 580–598.

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

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

[64] Breslow, N.E., Edler, L. and Berger, J. (1984). A two-sample censored data rank test for acceleration. *Biometrics* **40**, 1049–1062.

[65] 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.

[66] Brockmann, M., Gasser, T. and Herrmann, E. (1993). Locally adaptive bandwidth choice for kernel regression estimators. *Journal of the American Statistical Association* **88**, 1302–1309.

[67] 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 Proschan, F. and Serfling, R.J. (Eds.) *Reliability and Biometry, Statistical Analysis of Lifelength*, Society for Industrial and Applied Mathematics: Philadelphia, 327–354.

[68] Brunk, H.D. (1955). Maximum likelihood estimates of monotone parameters. *Annals of Mathematical Statistics* **26**, 607–616.

[69] Burdett, K. (1979). Unemployment insurance payments as a search subsidy: a theoretical analysis. *Economic Inquiry* **17**, 333–343.

[70] Caballero, R. and Hammour, M. (1994). The cleansing effect of recessions. *American Economic Review* **84**, 1350–1368.

[71] Cabral, L.M.B. (1993). Experience advantages and entry dynamics. *Journal of Economic Theory* **59** (2), 403–416.

[72] Cai, Z. and Sun, Y. (2003). Local linear estimation for time-dependent coefficients in Cox's regression models. *Scandinavian Journal of Statistics* **30**, 93–111.

[73] Calvo, G.A. (2000). Capital-markets crises and economic collapse in emerging markets: an informational-frictions approach. *American Economic Review* **90**, 59–64.

[74] Campbell, J. (1998). Entry, exit, embodied technology, and business cycles. *Review of Economic Dynamics* **1**, 371–408.

[75] Campolieti, M. (2001). Bayesian semiparametric estimation of discrete duration models: An application of the Dirichlet process prior. *Journal of Applied Econometrics*, **16**, 1–22.

[76] Caplehorn, J. and Bell, J. (1991). Methadone dosage and retention of patients in methadone maintenance. *Medical Journal of Australia* **154**, 195–199.

[77] Card, D. and Olson, C.A. (1992). Bargaining power, strike duration, and wage outcomes: An analysis of strikes in the 1880s. Working Paper No. **4075**, National Bureau of Economic Research, Cambridge, MA.

[78] Caves, R.E. (1998). Industrial organization and new findings on the turnover and mobility of firms. *Journal of Economic Literature* **36**, 1947–1982.

[79] Champlin, R., Mitsuyasu, R., Elashoff, R. and Gale, R.P. (1983). Recent advances in bone marrow transplantation. In: Gale, R.P. (Ed.) *UCLA Symposia on Molecular and Cellular Biology* **7**, Alan R. Liss: New York, 141–158.

[80] Chang, M.N. and Chung, D. (1998). Isotonic window estimators of the baseline hazard function in Cox's regression model under order restriction. *Scandinavian Journal of Statistics* **25**(1), 151–161.

[81] Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94**, 807–823.

[82] Chaudhuri, P. and Marron, J.S. (2000). Scale space view of curve estimation. *Annals of Statistics* **28**, 408–428.

[83] Chen, S. and Manatunga, A.K. (2007). A note on proportional hazards and proportional odds models. *Statistics and Probability Letters* **77**(10), 981–988.

[84] Cheng, S.C., Wei, L.J. and Ying, Z. (1995). Analysis of transformation models with censored data. *Biometrika* **82**, 835–845.

[85] Choi, E. and Hall, P. (1999). Data sharpening as a prelude to density estimation. *Biometrika* **86**, 941–947.

[86] Choi, E., Hall, P. and Roussen, V. (2000). Data sharpening methods for bias reduction in nonparametric regression. *Annals of Statistics* **28**, 1339–1355.

[87] Ciampi, A. and Etezadi-Amoli, J. (1985). A general model for testing the proportional hazards and accelerated failure time hypothesis in the analysis of censored survival data with covariates. *Communications in Statistics - Theory and Methods* **14**, 651–676.

[88] Claeskens, Gerda and Hall, Peter (2002). Data Sharpening for Hazard Rate Estimation. *Australian and New Zealand Journal of Statistics* **44**, 277–283.

[89] Clayton, D. G. (1991). A Monte Carlo method for Bayesian inference in frailty models. *Biometrics*, **64**, 141–151.

[90] Cooley, T. and Quadrini, V. (2001). Financial markets and firm dynamics. *American Economic Review* **91**, 1286–1310.

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

[92] Cox, D.R. (1975). Partial likelihood. *Biometrika* **62**, 269–276.

[93] Cox, D.R. and Oakes, D. (1984). *Analysis of Survival Data*. Chapman and Hall: New York.

[94] Cressie, N.A.C. and Reid, T.R.C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society* Series B **46**, 440–464.

[95] Cucker, F. and Smale, S., (2002). On the mathematical foundations of learning. *Bulletin of the American Mathematical Society* **39**, 1–49.

[96] Cuthbertson, K. and Hudson, J. (1996). The determinants of compulsory liquidation in the UK. *Manchester School of Economic and Social Studies*, **64**, 298–308.

[97] Dabrowska, D.M., Doksum, K.A. and Song, J.-K. (1989) Graphical comparison of cumulative hazards for two populations. *Biometrika* **76**, 763–773.

[98] Dabrowska, D.M., Doksum, K.A., Feduska, N.J., Husing, R. and Neville, P. (1992). Methods for comparing cumulative hazard functions in a semi-proportional hazard model. *Statistics in Medicine* **11**, 1465–1476.

[99] Damien, P. (2005). Some Bayesian nonparametric model. In: Dey, D.K. and Rao, C.R. (Eds.) *Handbook of Statistics, volume 25: Bayesian Thinking, Modeling and Computation*, Elsevier: Amsterdam, 279–314.

[100] Damien, P., Laud, P. and Smith, A. (1996). Implementation of Bayesian nonparametric inference using Beta processes. *Scandinavian Journal of Statistics* **23**, 27–36.

[101] Dauxois, J.-Y. and Kirmani, S.N.U.A. (2003). Testing the proportional odds model under random censoring. *Biometrika* **90**(4), 913–922.

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

[103] Dauxois, J.-Y. and Kirmani, S.N.U.A. (2005). On testing for a survival ratio under random right censoring. *Journal of Nonparametric Statistics* **17**(8), 949–955.

[104] Davies, P.L. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution (with discussion). *Annals of Statistics* **29**, 1–65.

[105] Delaigle, A. and Gijbels, I. (2003). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics and Data Analysis* (forthcoming).

[106] Delli Gatti, D., Gallegati, M., Guilioni, G. and Palestrini, A. (2001). Financial fragility, pattern of firms' entry and exit and aggregate dynamics. *Journal of Economic Behavior and Organization* **51**, 79–97.

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

[108] Detre, K.M., Hultgren, H. and Takaro, T. (1977). Veterans' Administration cooperative study of surgery for coronary arterial occlusive disease, III: methods and baseline characteristics including experience with medical treatment. *American Journal of Cardiology* **40**, 212–225.

[109] Dette, H., Neumeyer, N. and Pilz, K.F. (2006). A simple nonparametric estimator of a monotone regression function. *Bernoulli* **12**, 469–490.

[110] Devine, T.J. and Kiefer, N.M. (1991). *Empirical Labor Economics*. Oxford University Press: Oxford.

[111] Dixit, A. (1989). Entry and exit under uncertainty. *Journal of Political Economy*, **97**, 620–638.

[112] Doksum, K.A. (1974). Tailfree and neutral random probabilities and their posterior distributions. *Annals of Probability* **2**, 183–201.

[113] Doksum, K.A. and Yandell, B.S. (1984). Testing for exponentiality. In: Krishnaiah, P.R. and Sen, P.K. (Eds.) *Handbook of Statistics 4: Nonparametric Methods*, North-Holland: Amsterdam, 579–611.

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

[115] Dümbgen, L. (1998). New goodness-of-fit tests and their application to nonparametric confidence sets. *Annals of Statistics* **26**, 288–314.

[116] Dunne, T., Roberts, M.J. and Samuelson, L. (1989). The growth and failure of U.S. manufacturing plants. *Quarterly Journal of Economics* **104**, 671–698.

[117] Dunson, D.B. and Herring, A.H. (2003). Bayesian inferences in the Cox model for order restricted alternatives. *Biometrics* **59**, 918–925.

[118] Dunson, D.B. and Peddada, S.D. (2007). Bayesian nonparametric inference on stochastic ordering. *Biometrika*, forthcoming.

[119] Dykstra, R.L. (1982). Maximum likelihood estimation of the survival functions of stochastically ordered random variables. *Journal of the American Statistical Association* **77**, 621–628.

[120] Dykstra, R.L., Kochar, S. and Robertson, T. (1991). Statistical inference for uniform stochastic ordering in several populations. *Annals of Statistics* **19**, 870–888.

[121] Dykstra, R. L. and Laud, P. (1981). A Bayesian nonparametric approach to reliability. *Annals of Statistics* **9**, 356–367.

[122] Elbers, C. and Ridder, G. (1982). True and spurious duration dependence: the identifiability of the proportional hazard model. *Review of Economic Studies* **49**, 403–410.

[123] Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *Review of Economic Studies* **62**, 53–82.

[124] Evans, D.S. (1987). The relationship between firm growth, size, and age: estimates for 100 manufacturing industries. *Journal of Industrial Economics* **35**, 567–581.

[125] Farmen, M. and Marron, J.S. (1999). An assessment of finite sample performance of adaptive methods in density estimation. *Computational Statistics and Data Analysis* **30**, 143–168.

[126] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Annals of Statistics* **1**, 209–230.

[127] Ferguson, T. S. and Phadia, E. G. (1979). Bayesian nonparametric estimation based on censored data. *Annals of Statistics* **7**, 163–186.

[128] Fisher, N.I., Mammen, E. and Marron, J.S. (1994). Testing for multimodality. *Computational Statistics and Data Analysis* **18**, 499–512.

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

[130] Fleming, T.R. and Lin, D.Y. (2000). Survival analysis in clinical trials: past developments and future directions. *Biometrics* **56**, 971–983.

[131] Fleming, T.R., O'Fallon, J.R., O'Brien, P.C. and Harrington, D.P. (1980). Modified Kolmogorov-Smirnov test procedures with application to arbitrarily censored data. *Biometrics* **36**, 607–625.

[132] Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Annals of Statistics* **22**, 732–746.

[133] Gamerman, D. (1991). Dynamic bayesian models for survival data. *Applied Statistics* **40**, 63–79.

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

[135] Gelfand, A.E. and Kottas, A. (2001). Nonparametric Bayesian modeling for stochastic ordering. *Annals of the Institute of Statistical Mathematics* **53**, 865–876.

[136] Gelfand, A.E., Smith, A.F.M. and Lee, T.-M. (1992). Bayesian analysis of constrained parameter and truncated data problems using the Gibbs sampler. *Journal of the American Statistical Association* **87**, 523–532.

[137] Gill, R.D. and Schumacher, M. (1985). A simple test of the proportional hazards assumption. Tech. Rep. No. **MS R8504**, Department of Mathematics and Statistics, Centrum voor Wiskunde en Informatica, Amsterdam.

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

[139] Gore, S.M., Pocock, S.J. and Kerr,G.R. (1984). Regression models and non-proportional hazards in the analysis of breast cancer survival. *Applied Statistics* **33**, 176–195.

[140] 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.

[141] Goudie, A.W. and Meeks, G. (1991). The exchange rate and company failure in a macro-micro model of the UK company sector. *Economic Journal*, **101**, 444–457.

[142] Gozalo, P.L. and Linton, O.B. (2001). A nonparametric test of additivity in generalized nonparametric models regression with estimated parameters. *Journal of Econometrics* **104**, 1–48.

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

[144] Grambsch, P.M., Therneau, T.M. and Fleming, T.R. (1995). Diagnostic plots to reveal functional form for covariates in multiplicative intensity models. *Biometrics* **51**(4), 1469–1482.

[145] Gray, R.J. (1992). Flexible methods for analyzing survival data, using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* **87**, 942–951.

[146] Gray, R. J. (1994). A Bayesian analysis of institutional effects in a multicenter cancer clinical trial. *Biometrics*, **50**, 244–253.

[147] Greenwald, B. and Stiglitz, J. (1990). Asymmetric information and the new theory of the firm: financial constraints and risk behavior. *American Economic Review* **80**, 160–165.

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

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

[150] Gunn, L.H. and Dunson, D.B. (2007). Bayesian methods for assessing ordering in hazard functions. *Biostatistics*, under revision.

[151] Guo, G. and Rodríguez, G. (1992). Estimating a multivariate proportional hazards model for clustered data using the EM algorithm, with an application to child survival in Guatemala. *Journal of the American Statistical Association* **87**, 969–976.

[152] Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *Annals of Statistics* **29**, 624–647.

[153] Hall, P. and Presnell, B. (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society* Series B **61**, 143–158.

[154] Hall, P. and Turlach, B.A. (1999). Reducing bias in curve estimation by use of weights. *Computional Statistics and Data Analysis* **30**, 67–86.

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

[156] Han, A. and Hausman, J.A. (1990). Flexible parametric estimation of duration and competing risk models. *Journal of Applied Econometrics* **5**, 1–28.

[157] Hanson, D.L., Pledger, G. and Wright, F.T. (1973). On consistency in monotonic regression. *Annals of Statistics* **1**, 401–421.

[158] Härdle, W. and Stoker, T.M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American Statistical Association* **84**, 986–995.

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

[160] Hastie,T. and Tibshirani, R. (1993). Varying-coefficient models (with discussion). *Journal of the Royal Statistical Society* Series B **55**, 757–796.

[161] Hausman, J.A. (1978). Specification tests in econometrics, *Econometrica* **46**, 1251–1271.

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

[163] 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.

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

[165] Henderson, R. and Oman, P. (1999). Effect of frailty on marginal regression estimates in survival analysis. *Journal of the Royal Statistical Society* - Series B **61**(2), 367–379.

[166] Herrmann, E. (1997). Local bandwidth choice in kernel regression estimation. *Journal of Computional and Graphical Statistics* **6**, 35–54.

[167] Hess, K.R. (1994). Assessing time-by-covariate interactions in proportional hazards regression models using cubic spline functions. *Statistics in Medicine* **13**, 1045–1062.

[168] Higson, C., Holly, S. and Kattuman, P. (2002). The cross sectional dynamics of the US business cycle: 1950-1999. *Journal of Economic Dynamics and Control*, **26**, 1539–1555.

[169] Hildreth, C. (1954). Point estimate of ordinates of concave functions. *Journal of the Americal Statistical Association* **49**, 598–619.

[170] Hjort, N.L. (1985). Discussion of Andersen, P.K. and Borgan, Ø. (1985), "Counting process models for life history data: a review", *Scandinavian Journal of Statistics* **12**, 141–150.

[171] Hjort, N. L. (1990). Nonparametric Bayes estimators based on beta processes in models for life history data. *Annals of Statistics* **18**, 1259–1294.

[172] Hoel, D.G. (1972). A representation of mortality data by competing risks. *Biometrics* **28**, 475–488.

[173] Honoré, B.E. (1990). Simple estimation of a duration model with unobserved heterogeneity. *Econometrica* **58**, 453–473.

[174] Hopenhayn, H.A. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* **60**, 1127–1150.

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

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

[177] Horowitz, J. (2007). Discussion of Zeng, D. and Lin, D.Y. (2007), "Maximum likelihood estimation in semiparametric regression models", *Journal of the Royal Statistical Society*, Series B **69**, 546–547.

[178] Horowitz, J.L. and Neumann, G.R. (1992). A generalized moments specification test of the proportional hazards model. *Journal of the American Statistical Association* **87**(417), 234–240.

[179] Hougaard, P. (1991). Modelling heterogeneity in survival data. *Journal of Applied Probability* **28**, 695–701.

[180] Hougaard, P. (2000). *Analysis of Multivariate Survival Data*. Springer-Verlag: New York.

[181] Hougaard, P., Myglegaard, P. and Borch-Johnsen, K. (1994). Heterogeneity models of disease susceptibility, with an application to diabetic nephropathy. *Biometrics* **50**, 1178–1188.

[182] Hsieh, F. (1996). A transformation model for two survival curves: an empirical process approach. *Biometrika* **83**, 519– 528.

[183] Ibrahim, J.G., Chen, M.-H. and Sinha, D. (2001). *Bayesian Survival Analysis*. Springer-Verlag: New York.

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

[185] Jacod, J. and Shiryayev, A.N. (1980). *Limit theorems for stochastic processes*. Springer-Verlag: New York.

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

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

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

[189] Jin, Z., Lin, D.Y., Wei, L.J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341–353.

[190] Johnson, M.E., Tolley, H.D., Bryson, M.C. and Goldman, A.S. (1982). Covariate analysis of survival data: a small-sample study of Cox's model. *Biometrics* **38**, 685–698.

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

[192] 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.

[193] Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica* **50**. 649–670.

[194] Jovanovic, B. (1984). Wages and Turnover: A Parametrization of the Job-Matching Model. In Neumann, G.R. and Westergard-Nielsen, N. (Eds.) *Studies in Labor Market Dynamics*, Springer-Verlag: Heidelberg.

[195] Jovanovic, B. and Rousseau, P.L. (2001). Mergers and technological change: 1885–1998. Working Paper No. **01-W16**, Department of Economics, Vanderbilt University.

[196] Jovanovic, B. and Rousseau, P.L. (2002). The Q-Theory of Mergers. *American Economic Review - Papers and Proceedings*, **92**, 198–204.

[197] Jung, S.-H. and Jeong, J.-H. (2003). Rank tests for clustered survival data. *Lifetime Data Analysis* **9**, 21–33.

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

[199] Kalbfleisch, J. D. (1978). Nonparametric Bayesian analysis of survival time data. *Journal of the Royal Statistical Society* Series B **40**, 214–221.

[200] Kalbfleisch, J.D. and Prentice, R.L. (1973). Marginal likelihoods based on Cox's regression and life model. *Biometrika* **60**, 267–278.

[201] Kalbfleisch, J.D. and Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data.* Wiley: New York.

[202] Kay, R. (1977). Proportional hazards regression models and the analysis of censored survival data. *Applied Statistics* **26**, 227-237.

[203] Keiding, N. (1998). Selection effects and nonproportional hazards in survival models and models for repeated events. Working paper, University of Copenhagen, Copenhagen.

[204] 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.

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

[206] Kennan, J.F. and Wilson, R. (1989). Strategic bargaining models and interpretation of strike data. *Journal of Applied Econometrics* **4**, S87–S130.

[207] Kiefer, N.M (1988). Economic duration data and hazard functions. *Journal of Economic Literature* **26**, 646–679.

[208] Kirmani, S.N.U.A. and Dauxois, J.-Y. (2003). Testing relative risk under random censoring. *Statistics and Probability Letters* **62**, 1–7.

[209] Kirmani, S.N.U.A. and Dauxois, J.-Y. (2004). Testing the Koziol-Green model against monotone conditional odds for censoring. *Statistics and Probability Letters* **66**, 327–334.

[210] Kiyotaki, N. and Moore, J. (1997). Credit cycles. *Journal of Political Economy*, **105**, 211–248.

[211] Koziol, J.A. and Green, S.B. (1976). A Cramér-von Mises statistic for randomly censored data. *Biometrika* **63**, 465–474.

[212] Klefsjö, B. (1983). Testing exponentiality against HNBUE. *Scandinavian Journal of Statistics* **10**, 65–75.

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

[214] 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.

[215] Koul, H., Susarla, V. and Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Annals of Statistics* **9**, 1276–1288.

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

[217] Krogen, B.L. and Magel, R.C. (2000). Proposal of k sample tests for bivariate censored data for nondecreasing ordered alternatives. *Biometrical Journal* **42**, 435-455.

[218] Kvaløy, J.T. and Neef, L.R. (2004). Tests for the proportional intensity assumption based on the score process. *Lifetime Data Analysis* **10**, 139–157.

[219] Lagakos, S. (1988). The loss in efficiency from misspecifying covariates in proportional hazards regression models. *Biometrika* **75**, 156–160.

[220] Lagakos, S. and Schoenfeld, D. (1984). Properties of proportional hazards score tests under misspecified regression models. *Biometrics* **40**, 1037–1048.

[221] Lai, T.L. and Ying, Z. (1991). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Annals of Statistics* **19**, 1370–1402.

[222] Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association* **73**, 805–811.

[223] Lambson, V. (1991). Industry evolution with sunk costs and uncertain market conditions. *International Journal of Industrial Organization* **9**, 171–196.

[224] Lancaster, T. (1979). Econometric methods for the duration of unemployment. *Econometrica* **47**, 939–956.

[225] Lancaster T. (1985). Generalized residuals and heterogeneous duration models with applications to the Weibull model. *Journal of Econometrics* **28**, 113–126.

[226] Lancaster, T. (1990). *The Econometric Analysis of Transition Data*. Cambridge University Press: Cambridge.

[227] Laud, P., Damien, P. and Smith, A. F. M. (1998). Bayesian nonparametric and covariate analysis of failure time data. In Dey, D., Müller, P. and Sinha, D. (Eds.) *Practical Nonparametric and Semiparametric Bayesian Statistics*, Springer-Verlag, 213–225.

[228] Lausen, B. and Schumacher, M. (1996). Evaluating the effect of optimized cutoff values in the assessment of prognostic factors. *Computational Statistics and Data Analysis* **21**, 307–326.

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

[230] Lee, T.C.M. (2003). Smoothing parameter selection for smoothing splines: A simulation study. *Computational Statistics and Data Analysis* **42**, 139–148.

[231] Lee, T.C.M. and Solo, V. (1999). Bandwidth selection for local linear regression: A simulation study. *Computational Statistics* **14**, 515–532.

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

[233] Leurgans, S. (1982). Asymptotic distributions of slope-of-greatest-convex-minorant estimators. *Annals of Statistics* **10**, 287–296.

[234] 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.

[235] Liang, K.Y., Self, S.G. and Liu, X.H. (1990). The Cox proportional hazards model with change point: an epidemiologic application. *Biometrics* **46**(3), 783–793.

[236] Lin, D.Y. (1991). Goodness-of-fit analysis for the Cox regression model based on a class of parameter estimators. *Journal of the American Statistical Association* **86**, 725–728.

[237] Lin, D.Y. (1994). Cox regression analysis of multivariate failure time: the marginal approach. *Statistics in Medicine* **13**, 2233–2247.

[238] Lin, D.Y. and Wei, L.J. (1989). The robust inference for the Cox proportional hazards model. *Journal of the American Statistical Association*, **84**, 1074–1078.

[239] Lin, D.Y. and Wei, L.J. (1991). Goodness-of-fit tests for the general Cox regression model. *Statistica Sinica* **1**, 1–17.

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

[241] Lin, D.Y., Wei, L.J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika* **80**(3), 557–572.

[242] Lin, D.Y., Wei, L.J. and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**(3), 605–618.

[243] Lin, D.Y. and Ying, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61–71.

[244] Lin, D.Y. and Ying, Z. (1996). Semi-parametric analysis of the general additive-multiplicative hazards model for counting processes. *Annals of Statistics* **23**(5), 1712–1734.

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

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

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

[248] Lippman, S. and Rumelt, R. (1982). Uncertain immitability: an analysis of interfirm differences in efficiency under competition. *Bell Journal of Economics* **13**, 418–438.

[249] Little, R.J.A. and Rubin, D.B. (1987). *Statistical Analysis with Missing Data*. John Wiley: New York.

[250] Liu, J. (2004). Macroeconomic determinants of corporate failures: evidence from the UK. *Applied Economics*, **36**, 939–945.

[251] Liu, K., Qiu, P. and Sheng, J. (2007). Comparing two crossing hazard rates by Cox proportional hazards modelling. *Statistics in Medicine* **26**(2), 375–391

[252] 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.

[253] 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.

[254] Liu, R.Y. and Singh, K. (1997). Notions of limiting p values based on data depth and bootstrap. *Journal of the American Statistical Association* **92**, 266–277.

[255] LoPucki, L.M. and Whitford, W.C. (1993). Bankruptcy reorganization of large, publicly held firms. *Cornell Law Review*, 597–618.

[256] Mammen, E. (1991a). Estimating a smooth monotone regression function. *Annals of Statistics* **19**, 724–740.

[257] Mammen, E. (1991b). Nonparametric regression under qualitative smoothness assumptions. *Annals of Statistics* **19**, 741–759.

[258] Mammen, E., Linton, O. and Nielsen, J. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *Annals of Statistics* **27**, 1443–1490.

[259] Mammen, E., Marron, J S., Turlach, B.A. and Wand, M.P. (2001). A general projection framework for constrained smoothing. *Statistical Science* **16**, 232–248.

[260] Mammen, E. and Thomas-Agnan, C. (1999). Smoothing splines and shape restrictions. *Scandinavian Journal of Statistics* **26**, 239–252.

[261] Mammen, E. and van de Geer, S. (1997). Locally adaptive regression splines. *Annals of Statistics* **25**, 387–413.

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

[263] Manton, K.G., Stallard, E. and Wing, S. (1991). Analyses of black and white differentials in the age trajectory of mortality in two closed cohort studies. *Statistics in Medicine* **10**, 1043–1059.

[264] Martin, L.G., Trussell, J., Salvail, F.R. and Shah, N. (1983). Covariates of child mortality in the Philippines, Indonesia, and Pakistan: An analysis based on hazard models. *Population Studies* **37**, 417–432.

[265] Martinussen, T. and Scheike, T.H. (2002). A flexible additive multiplicative hazard model. *Biometrika* **89**, 283–298.

[266] 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.

[267] Marzec, L. and Marzec, P. (1997). On fitting Cox's regression model with time-dependent coefficients. *Biometrika* **84**, 901–908.

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

[269] 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.

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

[271] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models* (2nd Edition). Chapman and Hall: London.

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

[273] Metcalf, D., Wadsworth, J. and Ingram, P. (1992). Do strikes pay? Centre for Economic Performance, Discussion Paper No. **92**, ESRC Research Centre, London School of Economics, UK.

[274] Mooradian, R.M. (1994). The effect of bankruptcy protection on investment: Chapter 11 as a screening device. *Journal of Finance* **49**, 1403–1430.

[275] Moreau, T., O'Quigley, J. and Mesbah, M. (1985). A global goodness-of-fit statistic for the proportional hazards model. *Applied Statistics* **34**, 212–218.

[276] Mortensen, D.T. (1977). Unemployment insurance and job search decisions. *Industrial and Labor Relations Review* **30**, 505–517.

[277] Mukerjee, H. (1988). Monotone nonparametric regression. *Annals of Statistics* **16**, 741–750.

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

[279] 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.

[280] Nagelkerke, N.J.D., Oosting, J. and Hart, A.A.M. (1984). A simple test for goodness of fit of Cox's proportional hazards model. *Biometrics* **40**(2), 483–486.

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

[282] Nelson, W. (1969). Hazard plotting for incomplete failure data. *Journal of Quality Technology* **1**, 25–27.

[283] Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics* **14**, 945–965.

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

[285] Nickell, S.J. (1979). Estimating the probability of leaving unemployment. *Econometrica* **47**, 1249–1266.

[286] Nieto-Barajas, L. and Walker, S.G. (2002a). Bayesian nonparametric survival analysis via Lévy driven Markov processes. Technical report, Department of Mathematical Sciences, University of Bath, UK.

[287] Nieto-Barajas, L. and Walker, S.G. (2002b). Markov beta and gamma processes for modelling hazard rates. *Scandinavian Journal of Statistics* **29**, 413–424.

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

[289] O'Quigley, J. (2003). Khmaladze-type graphical evaluation of the proportional hazards assumption. *Biometrika* **90**(3), 577–584.

[290] O'Quigley, J. and Natarajan, L. (2004). Erosion of regression effect in a survival study. *Biometrics* **60**(2), 344–351.

[291] O'Quigley, J. and Pessione, F. (1991). The problem of a covariate-time qualitative interaction in a survival study. *Biometrics* **47**, 101–115.

[292] Pakes, A. and Ericson, R. (1998). Empirical implications of alternative models of firm dynamics. *Journal of Economic Theory* **79**, 1–46.

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

[294] Pebley, A.R. and Stupp, P.W. (1987). Reproductive patterns and child mortality in Guatemala. *Demography* **24**, 43–60.

[295] Peña, E.A. (1998). Smooth goodness-of-fit tests for the baseline hazard in Cox's proportional hazards model. *Journal of the American Statistical Association* **93**, 442, 673–692.

[296] 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.

[297] Pettitt, A.N. and Bin Daud, I. (1990). Investigating time dependence in Cox's proportional hazards model. *Applied Statistics* **39**(3), 313–329.

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

[299] Pollard, D. (1990). *Empirical Processes: theory and applications.* Institute of Mathematical Statistics: Hayward.

[300] Polzehl, J. and Spokoiny, V. (2003). Image denoising: pointwise adaptive approach. *Annals of Statistics* **31**, 30–57.

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

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

[303] Prewitt, K.A. (2003). Efficient bandwidth selection in non-parametric regression. *Scandinavian Journal of Statistics* **30**, 75–92.

[304] Qian, S. (1994). Generalization of least-square isotonic regression. *Journal of Statistical Planning and Inference* **38**, 389–397.

[305] Qiou, Z., Ravishanker, N. and Dey, D. K. (1999). Multivariate survival analysis with positive stable frailties. *Biometrics* **55**, 637–644.

[306] Ramsay, J.O. (1988). Monotone regression splines in action (with discussion). *Statistical Science* **3**, 425–461.

[307] Ren, J.J. (2003). Goodness of fit tests with interval censored data. *Scandinavian Journal of Statistics* **30**, 211–226.

[308] Ritov, Y. (1990). Estimation in linear regression model with censored data. *Annals of Statistics* **18**, 354–372.

[309] Robertson, T., Wright, F.T. and Dykstra, R.L. (1988). *Order restricted statistical inference*. Wiley: New York.

[310] Robins, J.M. and Tsiatis, A.A. (1992). Semiparametric estimation of an accelerated failure time model with time dependent covariates. *Biometrika* **79**, 311–319.

[311] Sargent, D.J. (1997). A flexible approach to time-varying coefficients in the Cox regression setting. *Lifetime Data Analysis* **3**, 13–25.

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

[313] Sasieni, P.D. (1996). Proportional excess hazards. *Biometrika* **83**, 127–141.

[314] Sather, H., Coccia, P., Nesbit, M., Level, C. and Hammond, D. (1981). Disappearance of the predictive value of prognostic factors for childhood acute lymphoblastic leukemia. *Cancer* **48**, 370–376.

[315] Scheike, T.H. (2004). Time varying effects in survival analysis. In: Balakrishnan, N. and Rao, C.R. (Eds.) *Handbook of Statistics 23: Advances in Survival Analysis*, North-Holland: Amsterdam, 61–85.

[316] 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.

[317] Schemper, M. (1992). Cox analysis of survival data with non-proportional hazard functions. *The Statistician* **41**(4), 455–465.

[318] Schoenfeld, D. (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika* **67**(1), 145–153.

[319] Schucany, W.R. (1995). Adaptive bandwidth choice for kernel regression. *Journal of the American Statistical Association* **90**, 535–540.

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

[321] Sen, A.K. (1998). Mortality as an indicator of economic success and failure. *Economic Journal* **108**, 1–25.

[322] Sengupta, D. (1995). Graphical tools for censored survival data. In Koul, H.L. and Deshpande, J.V. (Eds.) *Analysis of censored data*, Institute of Mathematical Statistics: Hayward, CA, 193–217.

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

[324] 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**(4), 637–647.

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

[326] Sengupta, D. and Jammalamadaka, S.R. (1993). Inference from discrete life history data: a counting process approach. *Scandinavian Journal of Statistics* **20**(1), 51–61.

[327] Shapiro, S.S. and Francia, R.S. (1972). An approximate analysis of variance test for normality. *Journal of the American Statistical Association* **67**, 215–216.

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

[329] Shleifer, A. and Vishny, R. (2003). Stock-market driven acquisitions. *Journal of Financial Economics* **70**, 295–311.

[330] Siegfried, J.J. and Evans, L.B. (1994). Empirical studies of entry and exit: a survey of the evidence. *Review of Industrial Organization*, **9**, 121–155.

[331] Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association* **88**, 979–983.

[332] Sinha, D., Chen, M.-H. and Ghosh, S.K. (1999). Bayesian analysis and model selection for interval-censored survival data. *Biometrics* **55**(2), 585–590.

[333] Sinha, D. and Dey, D.K. (1997). Semiparametric Bayesian analysis of survival data. *Journal of the American Statistical Association* **92**, 1195–1212.

[334] Solomon, P.J. (1984). Effects of misspecification of regression models in the analysis of survival data. *Biometrika* **71**, 291–298. Correction: (1986), **73**, pp. 245.

[335] Spiegelhalter, D.J., Thomas, A. and Best, N.G. (1999). WinBUGS Version 1.2 User Manual. MRC Biostatistics Unit, Institute of Public Health, Cambridge, UK.

[336] 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.

[337] Stablein, D.M., Carter, W.H. and Novak, J.W. (1981). Analysis of survival data with nonproportional hazard functions. *Controlled Clinical Trials* **2**, 149–159.

[338] Stablein, D.M. and Koutrouvelis, I.A. (1985). A two-sample test sensitive to crossing hazards in uncensored and singly censored data. *Biometrics* **41**, 643–652.

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

[340] Sueyoshi, G.T. (1995). A class of binary response models for grouped duration data. *Journal of Applied Econometrics* **10**(4), 411–431.

[341] Sun, J., Sun, L. and Zhu, C. (2007). Testing the proportional odds model for interval-censored data. *Lifetime Data Analysis* **13**, 37–50.

[342] Tantiyaswasdikul, C. and Woodroofe, M. (1994). Isotonic smoothing splines under sequential designs. *Journal of Statistical Planning and Inference* **38**, 75–87.

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

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

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

[346] Thompson, P. (2005). Selection and firm survival: evidence from the shipbuilding industry, 1825-1914. *Review of Economics and Statistics* **87** (1), 26–36.

[347] Tian, L., Zucker, D.M. and Wei, L.J. (2005). On the Cox model with time varying regression coefficients. *Journal of the American Statistical Association* **100**(469), 172–183.

[348] Trussell, J. and Hammerslough, C. (1983). A hazard-model analysis of the covariates of infant and child mortality in Sri Lanka. *Demography* **20**, 1–26.

[349] Trussell, J. and Richards, T. (1985). Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. In: Tuma, N. (Ed.) *Sociological Methodology*, Jossey-Bass: San Francisco, 242–276.

[350] Tsiatis, A.A. (1981). The asymptotic joint distribution of the efficient scores test for the proportional hazards model calculated over time. *Biometrika* **68**, 311–315.

[351] Tsiatis, A.A. (1990). Estimating regression parameters using linear rank tests for censored data. *Annals of Statistics* **18**, 354–372.

[352] Tubert-Bitter, P., Kramar, A., Chalé, J.J. and Moreau, T. (1994). Linear rank tests for comparing survival in two groups with crossing hazards. *Computational Statistics and Data Analysis* **18**, 547– 559.

[353] Valsecchi, M.G., Silvestri, D. and Sasieni, P. (1996). Evaluation of long-term survival: use of diagnostics and robust estimators with Cox's proportional hazards model. *Statistics in Medicine* **15**, 2763–2780.

[354] van den Berg, G.J. (1992). Nonparametric tests for unobserved heterogeneity in duration models. Working paper, Free University, Amsterdam.

[355] van den Berg, G.J. (2001). Duration Models: Specification, Identification, and Multiple Durations. In Heckman, J.J. and Leamer. E. (Eds.) *Handbook of Econometrics* Volume V, North Holland: Amsterdam.

[356] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer: New York.

[357] Vaupel, J.W., Manton, K.G. and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* **16**, 439–454.

[358] Verweij, J.M. and van Houwelingen, H.C. (1995). Time-dependent effects of fixed covariates in Cox regression. *Biometrics* **51**, 1550–1556.

[359] Wadhwani, S.B. (1986) Inflation, bankruptcy, default premia and the stock market, *Economic Journal*, **96**, 120-138.

[360] Walker, S.G. and Mallick, B.K. (1997). Hierarchical generalized linear models and frailty models with Bayesian nonparametric mixing. *Journal of the Royal Statistical Society* Series B **59**, 845–860.

[361] Wang, L. and Dunson, D.B. (2007). Bayesian isotonic density regression. *Mimeo.*

[362] Wei, L.J. (1984). Testing goodness of fit for proportional hazards model with censored observations. *Journal of the American Statistical Association* **79**(387), 649–652.

[363] Wei, L.J., Lin, D.Y. and Weissfield, L. (1989). Regression analysis of multivariate incomplete failure time data. *Journal of the American Statistical Association*, **84**, 1065–1073.

[364] Wei, L.J., Ying, Z. and Lin, D.Y. (1990). Linear regression analysis of censored survival data based on rank tests. *Biometrika* **77**, 845–851.

[365] Winnett, A. and Sasieni, P. (2003). Iterated residuals and time-varying covariate effects in Cox regression. *Journal of the Royal Statistical Society* Series B **65**(2), 473–488.

[366] Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Annals of Statistics* **21**, 76–99.

[367] Zeng, D. and Lin, D.Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data (with discussion and a rejoinder). *Journal of the Royal Statistical Society*, Series B **69**, 507–564.

[368] 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.

———x———