Estimation in Hazard Regression Models under Ordered Departures from Proportionality

Bhattacharjee, A.

Publication date:
2003

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

Citation for published version (APA):
Bhattacharjee, A. (2003). Estimation in Hazard Regression Models under Ordered Departures from Proportionality. (Cambridge Working Papers in Economics; No. 0337). Cambridge: Faculty of Economics, University of Cambridge.

Estimation in Hazard Regression Models under Ordered Departures from Proportionality

Arnab Bhattacharjee

July 2003

CWPE 0337

Not to be quoted without permission

Estimation in Hazard Regression Models under Ordered Departures from Proportionality *

Arnab Bhattacharjee †

June 3, 2003

Abstract

The Cox proportional hazards model is widely used for semiparametric analysis of lifetime/ duration data. It is, however, well acknowledged that the proportionality assumption underlying this model is strong; the assumption is often unreasonable from a theoretical point of view, and found to be empirically violated in many cases. In the two-sample setup, it is of interest to test proportionality against ordered/ monotone alternatives, where the ratio of hazard functions/ cumulative hazard functions increase or decrease in duration. Several tests for proportionality against these kinds of non-parametrically specified ordered alternatives exist in the literature (Gill and Schumacher, 1987; Sengupta *et. al.*, 1998). Recently, a natural extension of such monotone ordering to the case of continuous covariates has been discussed, and tests for the proportional hazards model against such alternatives developed (Bhattacharjee and Das, 2002). It is observed that such monotone/ ordered departures are common in applications, and provide useful additional information about the nature of covariate dependence.

In this paper, we describe methods for estimating hazard regression models when such monotone departures are known to hold. In particular, it is shown how the histogram sieve estimators (Murphy and Sen, 1991) in this setup can be smoothed and order restricted estimation performed using biased bootstrap techniques like adaptive bandwidth kernel estimators (Brockmann *et. al.*, 1993; Schucany, 1995) or data tilting (Hall and Huang, 2001). The performance of the methods is compared using simulated data, and their use is illustrated with applications from biomedicine and economic duration data.

Key words: Proportional hazards; Ordered restricted inference; Age-varying covariate effects; Biased bootstrap; Data tilting; Adaptive bandwidth selection; Histogram sieve estimator

---

*This work was supported by KPMG. The author thanks Sean Holly, Hashem Pesaran, Elvezio Ronchetti and Alastair Young for useful comments. Thanks are also due to the organisers and participants at the International workshop on Computational Management Science, Economics, Finance and Engineering at Limassol, Cyprus, 28-30 March 2003, where this work was presented.

†Research Associate, Department of Applied Economics, University of Cambridge, and Lecturer and Director of Studies (Economics), Robinson College, University of Cambridge, Sidgwick Avenue, Cambridge CB3 9DE, UK (email: A.Bhattacharjee@econ.cam.ac.uk).

1 Introduction

The proportional hazards (PH) model, and more specifically the Cox regression model (Cox, 1972), has played an important role in the theory and practice of lifetime and duration data analysis over the past few decades. 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 PH model restricts the coefficients of the regressors in the logarithm of the hazard function to be constant over the lifetime. This may not hold in many situations, or may even be unreasonable from the point of view of relevant theory. Since, such and other kinds of misspecifications often leads to misleading inferences about the shape of the baseline hazard and the effects of explanatory variables, testing the PH model, particularly against the omnibus alternative, has been an area of active research.

As opposed to such broad alternatives, 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). This kind of situation could arise, for example, if the coefficient of the covariate is not constant over time, or is dependent on some other (possibly unobserved) covariate. 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', which is equivalent to convex ordering of the lifetime distribution in one sample with respect to the other (Throughout this article, the word 'increasing' would mean 'non-decreasing', and 'decreasing' would mean 'non-increasing'). Under the same setup, Sengupta *et. al.* (1998) have proposed a test of the PH model against the weaker alternative hypothesis of 'increasing ratio of cumulative hazards' (star ordering of the two samples). The above alternative hypotheses ('increasing hazard ratio' and 'increasing ratio of cumulative hazards') often provide an explanation for the phenomenon of 'crossing hazards' frequently 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), have come to be accepted as natural ordered alternatives to the proportional hazards model. Empirical evidence of such ordering are abundant in the literature on empirical survival analysis, demography and economic duration models. Recently, Bhattacharjee and Das (2002) have discussed 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 and Das, 2002; Scheike, 2002), and provide useful information about the nature of covariate dependence.

Another popular approach, in this literature, is on interpretation of violations of the PH model in terms of age-varying covariate effects (for a review, see Scheike, 2002). [1] Several contributions have suggested testing the PH model through detection of

[1] Most of the literature refers to such effects as time-varying effects. We prefer to use the ter-

age-varying covariate effects (see, for example, Grambsch and Therneau, 1994; Scheike and Martinussen, 2002), and several methods for estimation of age-varying covariate effects have been proposed (see, for example, Zucker and Karr, 1990; Murphy and Sen, 1991; Martinussen *et. al.*, 2002).

In this paper, we build on the notion of ordered departures from proportionality introduced in Bhattacharjee and Das (2002), and propose estimation methods for hazard regression models under such order restrictions. Building on a natural interpretation of these alternatives in terms of monotonicity of age-varying covariate effects, we use biased bootstrap methods to estimate the covariate effects when such monotone departures are known to hold. The small sample properties of the estimators are explored using simulated data. The methods are shown to be conveniently implemented in applications, and conducive to useful inference.

In Section 2, we follow Bhattacharjee and Das (2002) and present concepts of ordered alternatives to the PH model, with respect to continuous covariates. Estimation methods are proposed in Section 3. In Section 4, we illustrate the use of the estimators using simulations and two real life applications. Section 4 collects the concluding remarks.

2 Partial ordering with respect to continuous covariates

The concept of partial ordering of lifetime distributions have been quite useful in applications. The most popular of the available notions of partial ordering, namely convex ordering and star ordering (Kalashnikov and Rachev, 1986; Sengupta and Deshpande, 1994), can both be conveniently interpreted in terms of monotonicity of ratios of hazards/ cumulative hazards over time. These constitute intuitive and meaningful departures from the PH model in two samples and in the competing risks framework. Gill and Schumacher (1987), Deshpande and Sengupta (1995), and Sengupta *et. al.* (1998) consider several empirical applications where the departure from the PH model in two samples is evident from the fact that the ratio of the hazard rates is not constant over the lifetime.

The following definition (Bhattacharjee and Das, 2002) extends, to the continuous covariate setup, the notion of monotone ordering in two samples discussed in Gill and Schumacher (1987) (monotone hazard ratio). 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 $T = t$. Then, T is defined to be *increasing (decreasing) hazard ratio for continuous covariate (IHRCC (DHRCC))* with respect to X if, whenever $x_1 > x_2$, $\lambda(t|x_1)/\lambda(t|x_2) \uparrow t$ $(\downarrow t)$.

IHRCC describes a notion of positive ageing with respect to the continuous covariate X; the higher the covariate, the faster the ageing of the individual – a situation common in empirical applications (Bhattacharjee and Das, 2002; Scheike, 2002). In biomedical applications, such monotonically age-dependant covariate effects have been noted in the literature, both under additive hazard models and multiplicative hazard models. For survival with malignant melanoma, for example, Andersen *et. al.* (1992) observe that, while “hazard seems to increase with tumor thickness” (pp. 389), the plot of estimated cumulative baseline hazards for patients with ‘2mm $\leq$

minology age-varying so as to distinguish this from any time-series effect that varies with calendar time.

tumor thickness $<$ 5mm' and 'tumor thickness $\geq$ 5mm' against that of patients with 'tumor thickness $<$ 2mm' reveal "concave looking curves indicating that the hazard ratios decrease with time" (pp. 544–545) (see also Martinussen *et. al.*, 2002). Similar monotonicity has been noted in the empirical literature on economic duration models. Jayet and Moreau (1991), using French data on employment durations, found evidence of non-monotone departures from the PH model, in that the ratio of hazard function for individuals in the age groups 24–28 years to that for 37–40 years was increasing upto a duration of approximately 120 days. Bhattacharjee and Das (2002) observe evidence of monotone ordering (with respect to "production index") in data on strike durations, Bhalotra and Bhattacharjee (2001) and Bhattacharjee *et. al.* (2002) observe similar evidence in data on child mortality (with respect to mother's age at childbirth) and firm exits (with respect to various measures of macroeconomic instability) respectively.

These definitions are also closely related to time-varying covariate effects. For example, within the class of hazard regression models with age-varying covariate effects (i.e., where $\lambda(t|x) = \lambda_0(t).\exp(\beta(t).x)$), IHRCC holds if and only if the prognostic effect $\beta(t)$ increases in age.

Bhattacharjee and Das (2002) proposed tests of the proportional hazards model against such monotone alternatives, by repeated application of the usual two-sample tests (Gill and Schumacher, 1987) for different pairs of covariate values, and then taking supremum, infemum or average of these individual test statistics. While such monotone structures are observed in many empirical applications, sometimes neither proportionality nor such monotone alternatives may hold, in the sense that $\beta(t)$ may increase in age over one range of the covariate space, and decrease over another; we call these models non-monotonic alternatives to the proportionality assumption. The supremum/ infemum tests can detect such situations, and can be used to estimate the change-point (for details, see Bhattacharjee and Das, 2002). The extension of these tests to the situation where other covariates are present is straightforward.

Once the nature of non-proportionality is established, whether monotonic or non-monotonic, using the above tests, Bhattacharjee and Das (2002) suggests using the iusual estimates of age-varying covariate effects, like the histogram sieve estimator of Murphy and Sen (1991), for further inference. One would hope that these estimates would capture the nature of covariate dependence indicated by the tests. It would, however, be advantageous to construct estimators that impose the ordering implied by these relationships. We consider such ordered restricted estimation methods in Section 3.

3 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 may either be the results of tests of proportionality against monotone alternatives (Bhattacharjee and Das, 2002), or theoretical considerations, or prior knowledge. A non-monotonic regression structure can often also be expressed in this form, in terms of auxilliary covariates. For example, if $\beta(t) \uparrow t$ over one range of the covariate space, say $x \leq x_0$, and $\beta(t) \downarrow t$ otherwise, we can write the regression model as $\lambda(t|x) = \lambda_0(t).\exp(\beta_1(t).x_1 + \beta_2(t).x_2), x_1 = x.I(x \leq x_0), x_2 = -x.I(x > x_0)$, where $\beta_1(t)$ and $\beta_2(t)$ both increase in t.

In the following three subsections, we discuss estimation of $\beta(t)$ under such models. We monotonise usual kernel regression or sieve estimates by reweighting the original data, so as to impose the order restrictions. These methods are inspired by the biased bootstrap techniques, and in particular data tilting (Hall and Presnell, 1999; Hall and Huang, 2001) and adaptive bandwidth kernel estimators (see, for example, Brockmann *et. al.* (1993) or Schucany (1995))[2].

3.1 Data tilting

We begin with a suitable estimator of age-varying covariate effects at r distinct ordered lifetimes/ durations $t^{(1)} < t^{(2)} < \ldots < t^{(r)}$. 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 $p_{n\times 1} \geq 0, \sum p_i = 1$ 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 age-varying covariate effects, 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 $D_\rho(p) = \{n - \sum_{i=1}^n (np_i)^\rho\} / \{\rho(1-\rho)\}, \rho \neq 0, 1,\ D_0(p) = -\sum_{i=1}^n \log(np_i)$ and $D_1(p) = -\sum_{i=1}^n p_i \log(np_i)$. The estimator then is

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

It is reasonably straightforward to abstract to an estimator over a continuous range of 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

$$I(p; t, x, Y, \delta) = \int_0^T \hat{\beta}'(s; t, x, Y, \delta, p).I\left(\hat{\beta}'(s; t, x, Y, \delta, p) < 0\right) ds = 0.$$

[2]SiZer maps (Chaudhuri and Marron, 1999, 2000) are also closely related

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. While this is theoretically an appealing estimation procedure, there are some problems with its implementation in the general form. Firstly, the likelihood function is complicated, and influence functions are not available in closed form. Hence, computation of the influence of each datapoint would require either some sort of Taylor series expansion of the partial derivatives of the objective function (which is problematic since this function is highly nonlinear), or jacknife-type computation of row-deletion influence (which is also obviously supoptimal). Second, the estimation procedure itself involves optimisation in high dimensions, and this dimension increases directly with the sample size.

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. Further, we fix $n_0 \geq 1$ and at each iteration of the estimation process, increase the weight of the n_0 observations with the highest influence on $I(\tilde{p}^{r-1})$ by $1/n$ each, and correspondingly reduce the weight of the lowest influence observations. This iterative procedure is continued till we achieve $I(p) = 0$.

Thus, the algorithm is as follows:

Step 1. **Initialize**: Fix $p = p_{unif}$, n_0 (the number of p_j's reduced at each iteration), $r = 1$ (first iteration), and R (the maximum number of iterations). Compute age-varying coefficients and $I(p_{unif})$.

Step 2. **Loop**: Do while $r \leq R$ and $I(\tilde{p}^r) < 0$

a) **Computation of influence functions** (for each observation for which $\tilde{p}_j^{r-1} > 0$): This can be done either by computing partial derivatives of the Taylor series expansion of $I(p)$ with respect to each data point, or by actual row deletion (jacknife) followed by usual age-varying coefficient estimation.

b) **Compute** $\tilde{p}_r$: Increase p_j by $1/n$ for the n_0 datapoints with highest influence, and correspondingly reduce p_j for the n_0 datapoints with lowest influence.

c) **Compute age-varying coefficients and** $I(\tilde{p}^{r+1})$.

d) $r = r + 1$.

Endo.

Step 3. **If-Else**: If $I(\tilde{p}^r) = 0$, return p^* and age-varying coefficient estimates. Else, achieve monotonicity by minimum degree of interpolation and/or extrapolation. Such interpolation/ extrapolation are likely to be required, particularly close to the boundaries of the sample space (where data are sparse).

An attractive feature of this algorithm is that the most computation-intensive step (Step 2a) is amenable to parallel computation. On the other hand, the algorithm can rapidly reduce effective sample size, by reducing a sizeable number of the p_j's to nil. The algorithm can be useful in applications, if such sample size reduction is matched with fast convergence towards monotonicity. How far the algorithm is effective in applications is thus largely an empirical issue; we address this issue in the next Section.

3.2 Local adaptive bandwidth

Adaptive bandwidth selection has a long and fairly established tradition in nonparametric regression.[3] In addition to the ability to adapt to the density of design points, and to 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 final advantage immediately suggests the usefulness of adaptive bandwidth estimators for order restricted inference in regression models, including hazard regressions. Since, 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 etc.

In the sense that this method involves rewighting of the original data in a particular way, it is similar to biased bootstrap methods (Hall and Presnell, 1999). However, this estimation procedure is richer than the data tilting method since it offers the possibility of choosing different bandwidths at different age levels, instead of choosing a general overall reweighting of the whole data.

Such adaptive bandwidth estimation is also similar in spirit to the way in which the location and scale[4] view (SiZer maps) has been proposed by Chaudhuri and Marron (1999, 2000) as an attractive way for exploring structures in curves. However, while Chaudhuri and Marron (1999, 2000) focus on identifying features of a non-parametric curve that are relatively more robust to changes in bandwidth (in other words, their focus is on testing), we propose to use adaptive bandwidths to perform kernel regression estimation subject to some maintained monotone structure.

We begin with a global bandwidth h_0 which provides an initial kernel estimator that is reasonably smooth. Now, for each lifetime/ duration t and local bandwidth $h(t)$ we can estimate a local kernel regression age-varying covariate effect $\hat{\beta}\left(t, h(t)\right)$ in the neighbourhood of t.[5] Then, our adaptive bandwidth kernel estimator will be given by $\hat{\beta}\left(t, h^*(t)\right)$, where $h^*(t)$ minimises $\int_0^T ||h(t) - h_0||.dt$ within the class of $h(t)$'s that satisfy

$$I\left(h(.); t, x, Y, \delta\right) = \int_0^T \hat{\beta}'\left(s, h(.)\right).I\left(\hat{\beta}'\left(s, h(.)\right) < 0\right) ds = 0.$$

Given our earlier intuition regarding the fact that a higher bandwidth would flatten out the kernel estimates, we expect the adaptive bandwidth method to give higher bandwidths to points on the lifetime/ duration scale that are either peaky in terms of the age-varying covariate effects, or the data at or around these points are sparse.

The algorithm is as follows:
Step 1. **Fix** h_0: Obtain kernel estimators for several candidate (fixed) bandwidths, and choose the one for which the age-varying coefficient estimates are reasonably

[3]Some recent contributions to this literature are Brockmann *et. al.* (1993), Schucany (1995), Hermann *et. al.* (1995), Hermann (1997) and Fan and Jiang (2000).

[4]Scale is interpreted here as the "level of resolution" or "bandwidth".

[5]$\hat{\beta}\left(t, h(t)\right)$ can be estimated by a simple modification of the usual estimates of age-varying covariate effects, like those proposed by Murphy and Sen (1991) or Martinussen *et. al.* (2002). A simple reweighting of the data using the weights suggested by the kernel, and then usual estimation with the reweighted data suffices.

smooth.
Step 2. **Explore a wide range of bandwidths**: Obtain kernel estimators for a wide range of bandwidths $h_1 < \ldots < h_0 < \ldots < h_r$, both above and below h_0.
Step 3. **Find candidate h's**: Find $h(.)$ within the class of combinations of bandwidths $h_1 < \ldots < h_0 < \ldots < h_r$ for which $I\left(h(.); t, x, Y, \delta\right) = 0$. Iterate between Step 2 (choosing finer grids and/ or widening the range of candidate bandwidths) and Step 3 until a satisfactory class of candidate $h(.)$'s is obtained, or if maxima of 0 is not achievable at all with reasonable bandwidths.
Step 4. **Optimal h^***: If several $h(.)$'s achieve a maxima of 0, choose the one that minimises $\int_0^T ||h(t) - h_0||.dt$. If none achieve 0, choose the one with the highest $I\left(h(.); t, x, Y, \delta\right)$, and then attain monotonicity by minimum degree of interpolation and/or extrapolation. As in the case of data tilting, interpolation/ extrapolation may often be required, particularly close to the boundaries of the sample space.

While this method is obviously less parsimonious than data tilting, it offers more degrees of choice, and therefore makes it easier for monotonicity to be imposed. In spite of its lower parsimony, however, this method is easier to implement, being less computation intensive than the elaborate jacknife procedures required in the previous method.

Further, we find adaptive bandwidth estimators easier to interpret than data tilting. Given the optimal bandwidths at the different ages, the user can also infer about the strength of the monotonicity, much in the same spirit as SiZer maps (Chaudhuri and Marron, 1999, 2000). Similarly, usual confidence intervals are easier to construct, and provide useful inference about the strength of the maintained order restriction at different ages.[6] Like data tilting, however, adaptive bandwidth estimators may require some interpolation/ extrapolation, which is obviously undesirable. The degree to which such measures are necessary is an empirical issue and we shall come back to it in the context of the simulation exercise presented later in the paper.

3.3 Interpolation/ extrapolation of usual age-varying coefficient estimates

The fact that both the two methods suggested so far may require further refining through interpolation/ extrapolation, particularly towards the boundary of the sample space, raises another interesting possibility. One could achieve monotonicity by directly applying such interpolation/ extrapolation methods on the usual estimates of time-varying covariate effects.

We explored this possibility in the empirical exercise described in Section 4; however, these estimates required a substantially greater degree of interpolation, and particularly, extrapolation. We therefore used the (unadjusted) Murphy-Sen estimator to benchmark the findings with the biased bootstrap techniques.

[6]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 construction will, however, be computation intensive.

4 Applications and simulations

In this section, we analyse issues relating to performance of and implementation of the above methods by way of a small simulation study, and two applications – one each in biomedicine and economic duration models.

4.1 Simulation study

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*$[-1,1]$, and the censoring random variable C has distribution function $F(c)=(c-0.005)^3, c\in[0.005,1.005]$.

20 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 Section 3 were applied to each. An Epanechnikov kernel was used, and the histogram sieve estimators of Murphy and Sen (1991) were used for estimation of age-varying covariate effects. For data tilting, R and n_0 were taken as 12 and 10 respectively. The methods were evaluated at 10 equidistant lifetimes 0.06 to 0.60 with increments of 0.06.

The performance of the data tilting method was the worst of the three. None of the 20 samples converged to a monotone model in 12 iterations, and a high degree of interpolation/ extrapolation had to be employed, particularly close to the boundaries. As expected, the effective sample size was reduced to about 430 in the 12 iterations. Though the average decrease in $|I(p)|$ was about 83 per cent, the estimates were wayward; for $t=0.06$ and $t=0.60$, for example, the average estimates were -0.594 and 3.282 respectively, as compared to values of $\beta(0.06)=0.06$ and $\beta(0.60)=0.60$ respectively under the model.

Adaptive bandwidth estimator performed the best of the three, the average estimates for $t=0.06$ and $t=0.60$ were 0.022 and 0.662 respectively. However, only 4 of the 20 samples converged to monotone estimates (i.e., $I\left(h^*(.);t,x,Y,\delta\right)=0$), and monotonicity for the others had to be achieved by interpolation/ extrapolation. The Murphy-Sen estimator by contrast had average estimates for $t=0.06$ and $t=0.60$ of 0.033 and 0.534 respectively. However, the adaptive bandwidth estimator appears to be more efficient, with the estimates for the different samples tightly clustered together, as seen from the box and whiskers plot for these two estimators in Figure 1 (adaptive bandwidth) and Figure 2 (Murphy-Sen histogram sieve estimator). The average absolute deviation of the estimates from actual values for these 10 points was 0.020 for adaptive bandwidth and 0.042 for Murphy-Sen, while this measure was as high as 0.567 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.

4.2 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 it. One of the strongest prognostic factors in malignant melanoma identified in the literature is tumor thickness. As discussed in Section 2, 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.

Our earlier work (Bhattacharjee and Das, 2002) showed that the null hypothesis of proportional hazard is rejected 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 $IHRCC$ pattern (this was also confirmed by the Murphy-Sen histogram sieve estimators). Figures 3 and 4 show kernel estimators of the age-varying covariate effects 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 come out prominently in these plots, and that constrained estimation using adaptive bandwidth selection can be used to obtain order-restricted covariate effects for tumor thickness.

4.3 Example: Macroeconomic instability and business failure

Bhattacharjee *et. al.* (2002) have analysed firm exits in the UK through bankruptcy over the period 1965 to 1998. The data pertain to around 4300 listed manufacturing companies, covering approximately 49,000 company years and including 166 exits due to bankruptcy. The data are right censored (by the competing risks of acquisitions, delisting etc.), left truncated in 1965, and contain delayed entries. 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.

The tests of proportional hazards against monotone departures proposed in Bhattacharjee and Das (2002) indicate monotone departures in both cases, and this is also confirmed by the Murphy-Sen estimates, after conditioning on industry dummies and firm level factors like size, profitability and cash flow.

The kernel estimates of age-varying covariate effects (Figures 5 and 6) 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 7 and 8), and provide usable 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; they depend strongly on the magnitude of the bandwidth, and hence not only on the peakedness feature of the kernel estimates at different points, but also on the density of data around these durations.

In summary, the adaptive bandwidth estimators appear to be an useful 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 difficult to obtain and often to interpret, and their performance in the simulation study was poor.

5 Concluding remarks

This paper proposes estimation methods for hazard regression models under order restrictions, where the age-varying covariate effects are known to be monotonically increasing or decreasing. Such situations occur frequently in applications, and encompass a wide range of regression models for survival and duration data. Two versions of biased bootstrap methods are considered, and usual estimates of age-varying covariate effects, suitably monotonised by interpolation/ extrapolation, are used for benchmarking the performance of these estimators.

The adaptive bandwidth estimator proposed in the paper performed very well with simulated data, and are very useful in applications. Though the estimator was not able to produce a monotone feature at all points on the duration scale (and particularly near the boundary of the sample space), interpolation/ extrapolation in these regions restored monotonicity and improved the estimator.

Together with our earlier work (Bhattacharjee and Das, 2002) on testing proportionality against monotone alternatives in hazard regression models, these inference techniques provide a new and useful way of analysing lifetime / duration data regression models in non-proportional hazards situations.

References

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

Bhalotra, S.R. and Bhattacharjee, A. (2001). Understanding regional variations in child mortality in India. Presented at the Workshop on *'Welfare, Demography and Development'*, University of Cambridge, Cambridge, UK, September 2001.

Bhattacharjee, A. and Das, S. (2002). Testing proportionality in duration models with respect to continuous covariates. DAE Working Paper 0220, University of Cambridge. Submitted.

Bhattacharjee, A., Higson, C., Holly, S. and Kattuman, P. (2002). Macro economic instability and business exit: Determinants of failures and acquisitions of large UK firms. DAE Working Paper 0206, University of Cambridge. Submitted.

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.

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

Chaudhuri, P. and Marron, J.S. (2000). Scale space view of curve estimation. *The Annals of Statistics* **28**, 408–428.

Chaudhuri, P. and Marron, J.S. (1999). SiZer for exploration of structures in curves. *Journal of the American Statistical Association* **94**, 807-823.

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.

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

Fan,J. and Jiang, J. (2000). Variable bandwidth and one-step local M-estimator. *Scientia Sinica* Series A **43**, 65–81.

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

Grambsch, P.M. and Therneau, T.M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* **81**, pp. 515–526. Correction: 95v82 pp. 668.

Hall, P. and Huang, L.-S. (2001). Nonparametric kernel regression subject to monotonicity constraints. *The Annals of Statistics* **29**, 624–647.

Hall, P. and Presnell, B. (1999). Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society Series B* **61**, 143–158.

Herrmann, E. (1997). Local bandwidth choice in kernel regression estimation. *Journal of Computational and Graphical Statistics* **6**, 35–54.

Herrmann, E., Wand, M.P., Engel, J. and Gasser, T. (1995). A bandwidth selector for bivariate kernel regression. *Journal of the Royal Statistical Society* Series B **57**, 171–180.

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

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

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.

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.

Scheike, T.H. (2002). Time varying effects in survival analysis. Research Report **2002-06**, Department of Biostatistics, University of Copenhagen.

Scheike, T.H. and Martinussen, T. (2002). On estimation and tests of time-varying effects in the proportional hazards model. *Scandinavian Journal of Statistics.* To appear.

Schucany, W.R. (1995). Adaptive bandwidth choice for kernel regression. *Journal of the American Statistical Association* **90**, 535–540.

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

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

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.

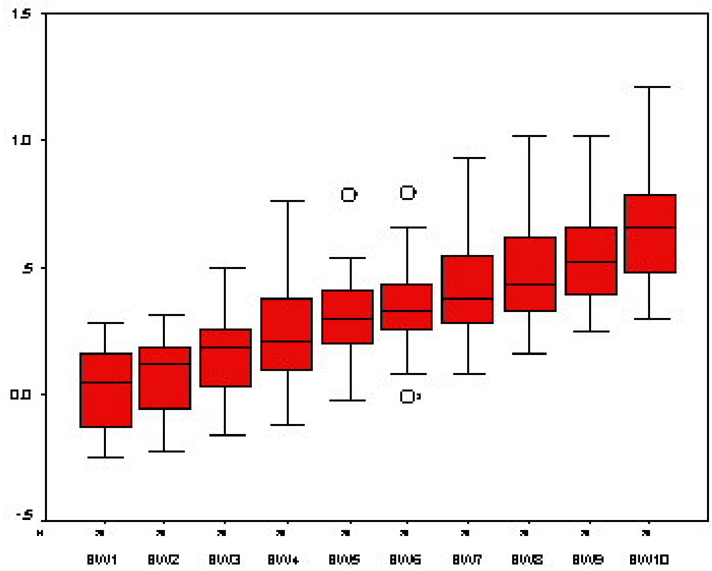

Figure 1:

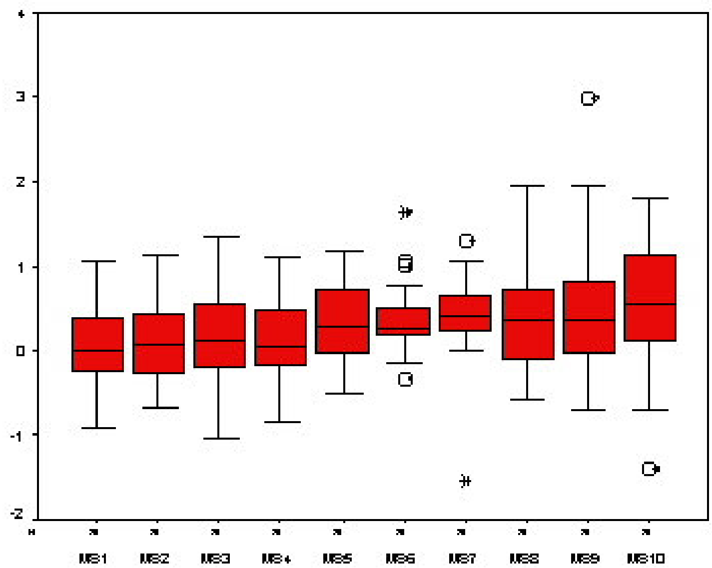

Figure 2:

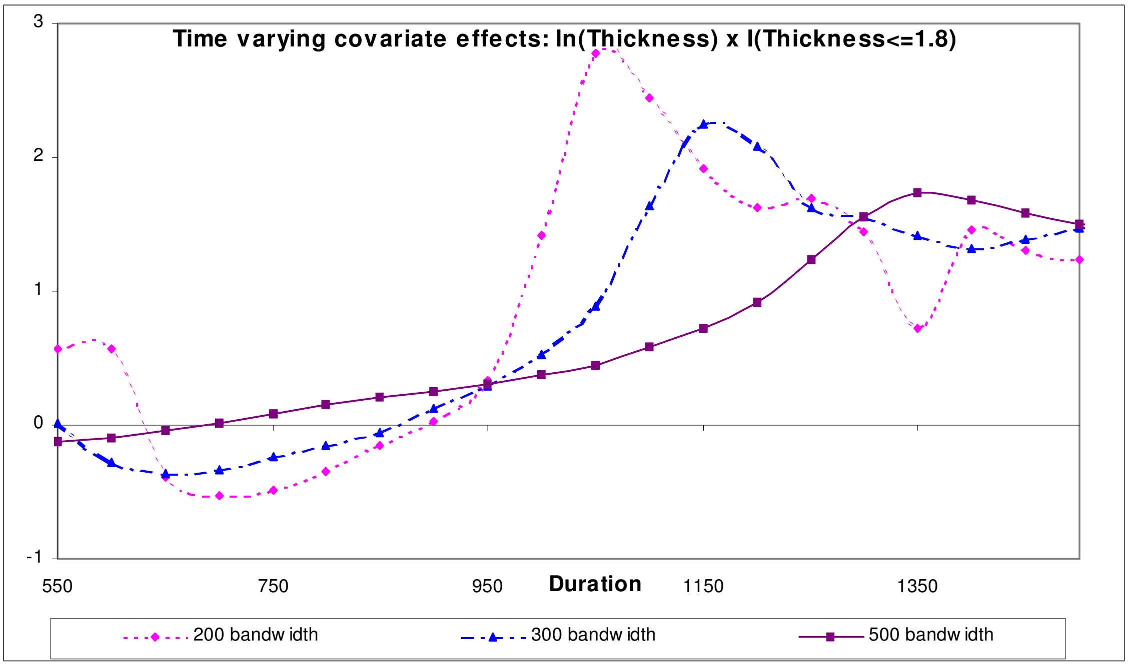

Figure 3:

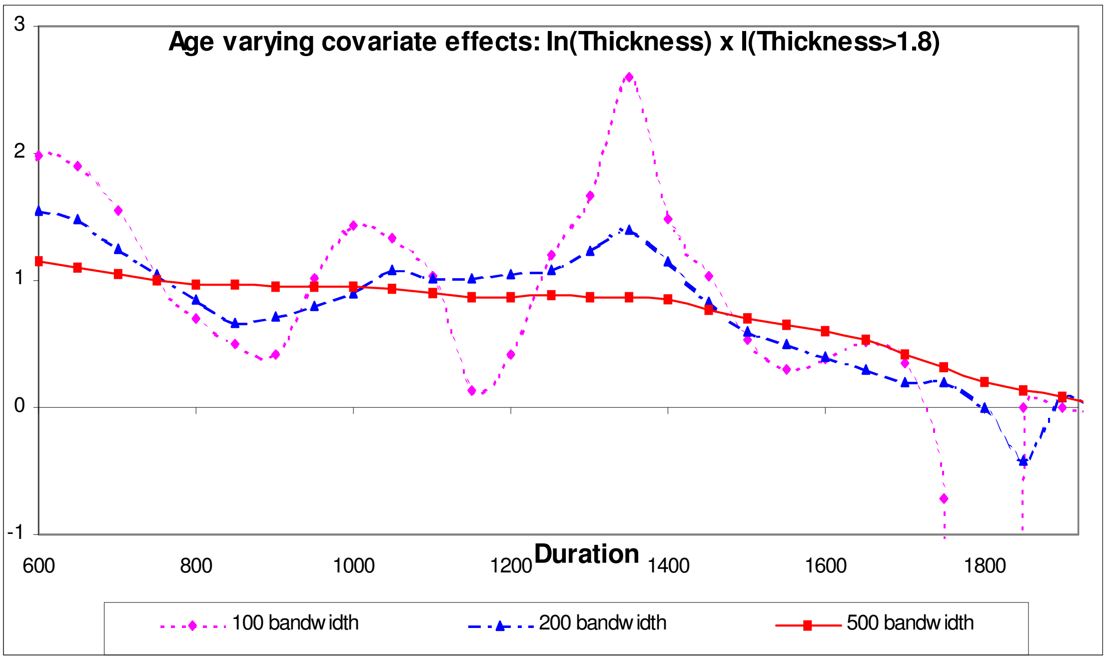

Figure 4:

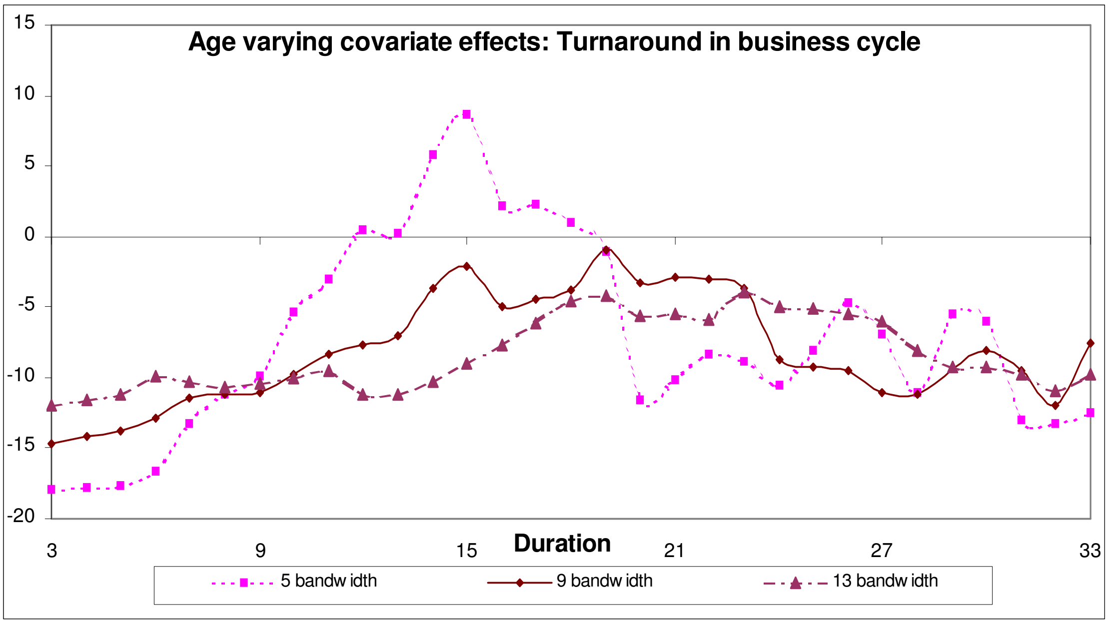

Figure 5:

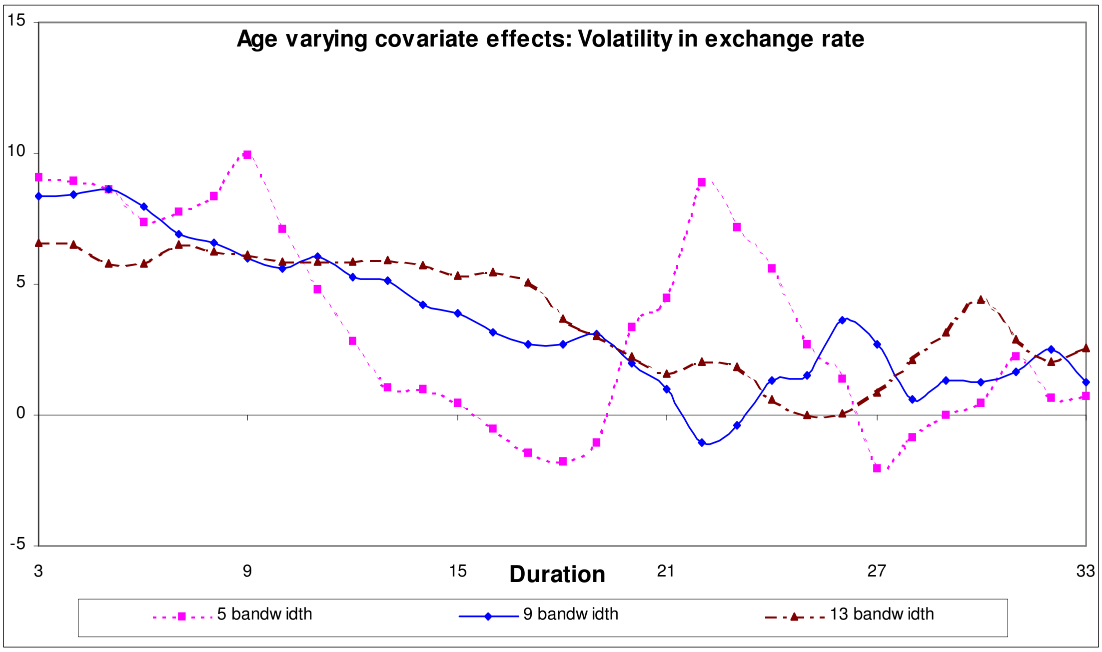

Figure 6:

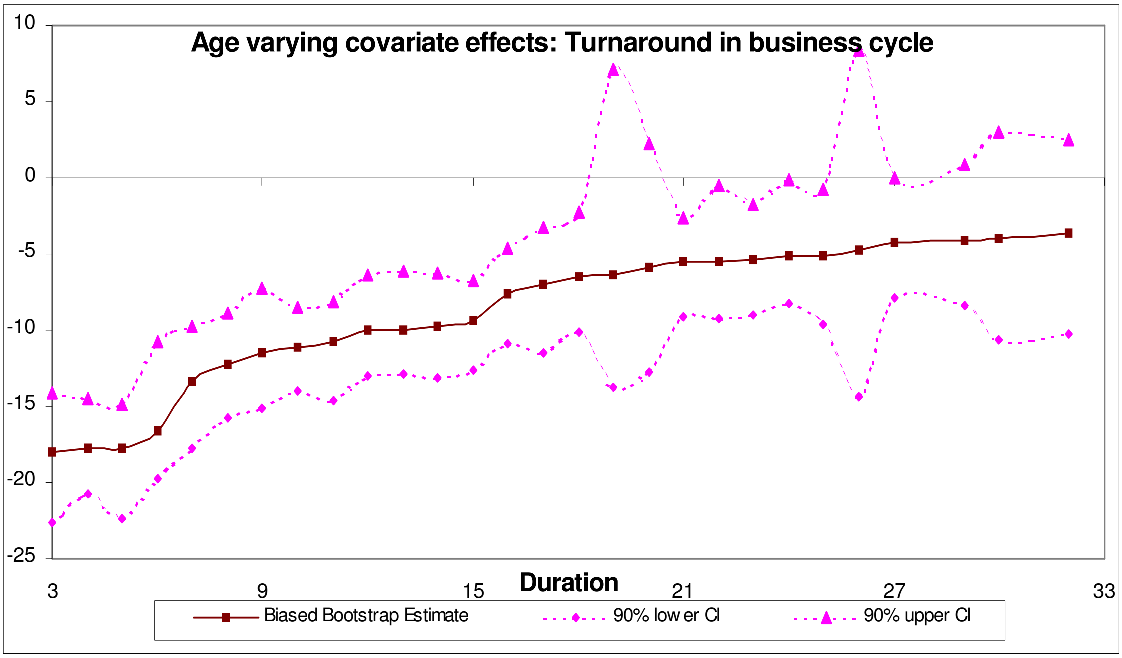

Figure 7:

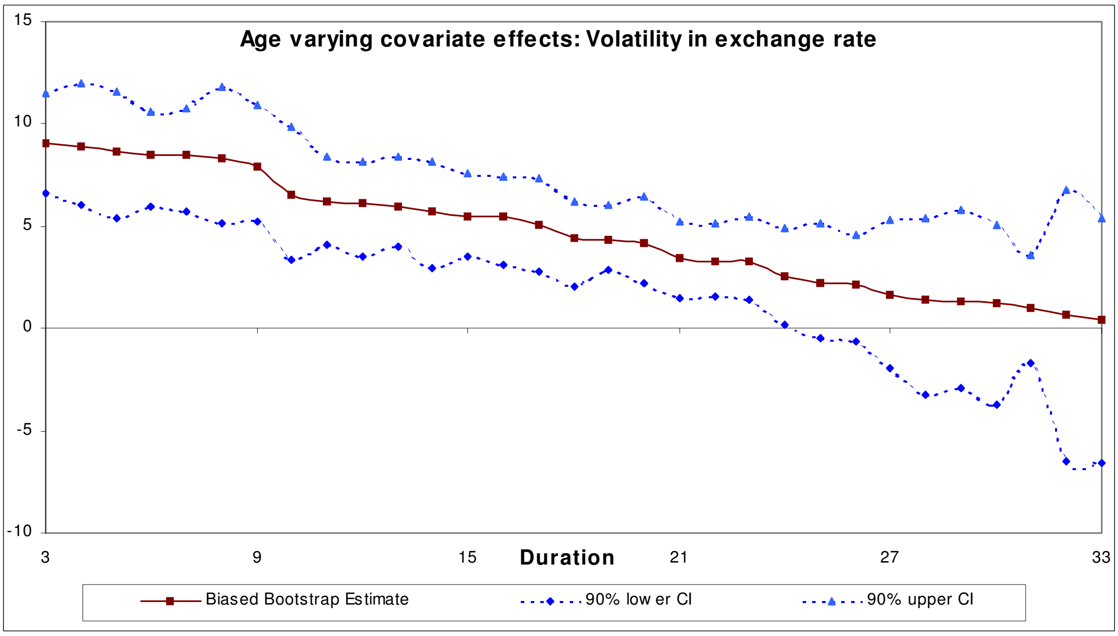

Figure 8: