AgEcon SEARCH
RESEARCH IN AGRICULTURAL & APPLIED ECONOMICS

The World's Largest Open Access Agricultural & Applied Economics Digital Library

This document is discoverable and free to researchers across the globe due to the work of AgEcon Search.

Help ensure our sustainability.

Give to AgEcon Search

AgEcon Search
http://ageconsearch.umn.edu
aesearch@umn.edu

*Papers downloaded from **AgEcon Search** may be used for non-commercial purposes and personal study only. No other use, including posting to another Internet site, is permitted without permission from the copyright owner (not AgEcon Search), or as allowed under the provisions of Fair Use, U.S. Copyright Act, Title 17 U.S.C.*

No endorsement of AgEcon Search or its fundraising activities by the author(s) of the following work or their employer(s) is intended or implied.

AMSTER

Rept 17/84

FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

A & E REPORT

REPORT AE 17/84

ON THE ESTIMATION OF THE PROPORTIONAL HAZARDS MODEL IN THE PRESENCE OF UNOBSERVED HETEROGENEITY

Geert Ridder

Wim Verbakel

GIANNINI FOUNDATION OF AGRICULTURAL ECONOMICS LIBRARY
OCT 15 1985

University of Amsterdam.

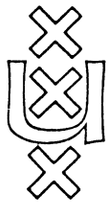
FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

REPORT AE 17/84

ON THE ESTIMATION OF THE PROPORTIONAL HAZARDS MODEL IN THE PRESENCE OF UNOBSERVED HETEROGENEITY

Geert Ridder and Wim Verbakel

Faculty of Actuarial Science and Econometrics
University of Amsterdam
Jodenbreestraat 23
1011 NH Amsterdam The Netherland



University of Amsterdam

Title: On the estimation of the proportional hazards model in the presence of unobserved heterogeneity

Running title: Unobserved heterogeneity

Geert Ridder and Wim Verbakel *)

ABSTRACT

In this paper we discuss the consequences of omitted covariates for the estimates of the regression coefficients and the base-line hazard of the proportional hazards model. We find that the regression coefficients are underestimated and that the estimated base-line hazard is less increasing or more decreasing than the true base-line hazard.

Key-words: Survival data, Proportional hazards, Unobserved heterogeneity, Cox-estimator

*) Geert Ridder is assistant professor at the Institute of Actuarial Science and Econometrics, Jodenbreestraat 23, 1011 NH Amsterdam; Wim Verbakel is at the Rijksgebouwendienst. We are indebted to Chris Elbers, J.S. Cramer, two referees and the associate editor for helpful comments.

1. INTRODUCTION

Survival analysis is a flourishing branch of statistics. Applications range from reliability testing to clinical trials, life tables and unemployment durations. The popularity of the subject is partly due to the availibility of flexible models and simple estimation methods for these models. Especially, the contribution of Cox (1972) who introduced the proportional hazards specification and a semi-parametric estimation method for this model, was important.

The proportional hazards model has been applied in experimental and non-experimental situations. In non-experimental situations (and even in some experimental situations e.g. clinical trials) the control of the data generating mechanism is imperfect. For that reason analysts usually include a number of covariates in the model in order to measure the effects of interest with more precision. Or they randomize the treatment allocation, so that the omitted covariates are not correlated with the treatment.

For comparison we give a short review of regression methods for the analysis of experimental data. If we consider a simple randomized experiment then we can analyse the data from the experiment with a simple analysis of variance model

$$(1.1) \qquad \underline{y} = \beta_o + \beta_1 d + \underline{\varepsilon}$$

with $\underline{y}$ the response variable, $\underline{\varepsilon}$ the error and d the indicator of treatment. (We underline random variables). Of course, the randomized allocation of treatments implies that d is random as well. However, because d is ancillary for the parameters of interest we can consider d as a given constant. Because the treatment allocation is random, we have

$$(1.2) \qquad E(\underline{\varepsilon}|d=1) = E(\underline{\varepsilon}|d=0) = 0,$$

a sufficient condition for unbiased estimation of the treatment effect. If

$$(1.3) \qquad E(\underline{\varepsilon}|d=1) > E(\underline{\varepsilon}|d=0)$$

i.e. if $\underline{\varepsilon}$ and d are positively correlated, then the treatment effect will be overestimated. If d is replaced by a vector x of covariates then a sufficient condition for unbiased estimation is that $\underline{\varepsilon}$ and x are not correlated (if $\underline{x}$ is ancillary for the parameters of interest, then $\underline{x}$ can be considered as a random vector or as a vector of constants). If $\underline{x}$ and $\underline{\varepsilon}$ are correlated then the ordinary least-squares estimates of the regression parameters are biased (for the direction of the bias, see Theil (1971), pp. 549-550).

In this paper we show that imperfect control of experimental conditions has different consequences for the estimates of treatment effects in survival analysis. Neglecting omitted covariates (even if they are not correlated with the treatment allocations) leads away to underestimation of the treatment effect. We assess this bias by a combination of a Monte Carlo analysis and an analytical derivation.

The plan of the paper is as follows. In section 2 we discuss a model for survival data in the presence of omitted covariates. In section 3 we consider the effects of omitted covariates on the popular semi-parametric estimation method for the proportional hazards model. Section 4 contains some conclusions and suggestions.

2. THE PROPORTIONAL HAZARDS MODEL AND THE MIXED PROPORTIONAL HAZARDS MODEL

We consider the well-known Proportional Hazards (PH) specification introduced by Cox (1972) and discussed in e.g. Kalbfleisch and Prentice (1980). Under this specification the hazard function is given by

$$(2.1) \qquad \lambda(t,x;\beta) = \phi(x;\beta)\psi(t)$$

where $\psi(t)$ is the base-line hazard function and $\phi(x;\beta)$ describes how this base-line hazard varies with the covariates $x;\beta$ is a vector of regression parameters. It is common practice to assume that

$$(2.1a) \qquad \phi(x;\beta) = \exp(x'\beta) \ .$$

Further we assume that the regressors x are constant over time. Most results obtained below also apply if there are time-varying regressors.

In most applications the number of covariates in the model is small. There are several reasons for this. In non-experimental situations one is usually limited by the number of available covariates, but even in experimental situations it may be impossible to obtain observations on all (potentially) relevant explanatory variables (if one knew this set in the first place). Moreover, it is common practice (see e.g. Kalbfleisch and Prentice (1980) pp.89-98) to identify relevant covariates by including them in the model one at a time. These observations indicate that in most applications the heterogeneity in the sample will only be partially described by the included covariates.

This fact has been noted by several authors (e.g. Vaupel, Manton and Stallard (1979), Lancaster (1979), Lancaster and Nickell (1980)). They note that as a consequence of unobserved heterogeneity observation units with identical x and t will have different hazard rates. More specifically they assume that the additional heterogeneity can be introduced in the hazard as follows

$$(2.2) \qquad \lambda(t,x,\underline{v};\beta) = \phi(x;\beta)\psi(t)\underline{v}$$

where the positive random variable $\underline{v}$ has a distribution with d.f. G. We assume that $E(\underline{v}) < \infty$. The random variable $\underline{v}$ is analogous to the disturbance term in a regression model and therefore it is natural to assume that the distribution of $\underline{v}$ does not depend on x. In experimental studies x will contain the treatment variables and therefore a randomized allocation of treatments will ensure the independence of $\underline{v}$ and x. Of course, chosing $\phi(x;\beta) = \exp(x'\beta)$ leads to a multiplicative specification of the disturbance term $\underline{v}$. In the following we will restrict attention to this form of ϕ.

Two of the three factors in the PH-specification (2.2) have to be normalized. We will chose

(2.3a) $\quad E(\underline{v}) = 1$

and

(2.3b) $\quad \psi(1) = 1$.

However, for reasons indicated below we will use another normalization in our Monte-Carlo study.

The survivor function corresponding to (2.2) is

$$(2.4) \qquad \bar{F}(t|x;\beta) = \int_0^\infty \exp\{-\phi(x;\beta)\, z(t)\, v\}\, dG(v)$$

with

$$(2.4a) \qquad z(t) = \int_0^t \psi(s)\, ds.$$

For obvious reasons we refer to the specification (2.2) as the Mixed Proportional Hazards (MPH) specification. The MPH-model has been considered by several authors, in most cases with a specific choice of G. (Vaupel, Manton and Stallard (1979), Lancaster (1979), Lancaster and Nickell (1980), Hougaard (1983)).

The MPH-model is characterized by a function ϕ which describes the observed heterogeneity, a time-dependence function ψ and a d.f. G of the distribution of the unobserved heterogeneity. One can ask whether this characterization is unique. Elbers and Ridder (1982) have shown that if ϕ takes on at least two distinct values and $E(\underline{v}) < \infty$, then to every MPH-model corresponds just one triple $\{\phi,\psi,G\}$. If ϕ is a constant different combinations of $\{\psi,G\}$ may lead to the same MPH-model. In

particular, every MPH-model has an equivalent PH-specification, so that the unobserved heterogeneity can be completely absorbed into the time-dependence ψ. For if $\bar{F}(t)$ is the survivor function of an arbitratry MPH-model with constant ϕ, then

(2.4b) $\quad \tilde{z}(t) = -\ln\bar{F}(t)$

where $\tilde{z}(t)$ is differentiable on $(0,\infty)$ with $\tilde{z}'(t) > 0$, $\tilde{z}(0) = 0$ and $\tilde{z}(\infty) = \infty$. Therefore

(2.4c) $\quad \bar{F}(t) = \exp\{-\tilde{z}(t)\}$

where the right-hand side is a survivor function of a proper PH-model.

If $\phi(x;\beta)$ varies with x, misspecification of G or ψ can result in biases in the estimates of the regression coefficients or the base-line hazard. Specifically, application of the semi-parametric estimation method suggested by Cox with the implicit assumption that $\underline{v}$ has a degenerate distribution can result in biased estimates.

We conclude this section with a closer examination of the MPH-model and a comparison of this model with other models that have been proposed for the analysis of failure-time data. For this purpose we prove the following theorem (the symbol $\stackrel{d}{=}$ indicates that two random variables have the same distribution).

Theorem 2.1

If the distribution of $\underline{t}$ is of the MPH-type with base-line hazard ψ, observed heterogeneity $\phi(\underline{x};\beta)$ and unobserved heterogeneity $\underline{v}$ with d.f. G then

(2.5a) $\quad \ln z(\underline{t}) \stackrel{d}{=} -\ln \phi(\underline{x};\beta) + \underline{w} - \ln \underline{v}$ [2)]

with z(t) given by (2.4a) and $\underline{w}$ a random variable with a Type-1 Extreme Value distribution (Johnson and Kotz (1970) p.272). $\underline{w}$ is stochastically independent of $(\underline{x},\underline{v})$. Conversely, if $\underline{t}$ has a distribution such that (2.5a) holds for a differentiable and monotonically increasing function z (the distribution of $\underline{w}$ is as before), then this distribution is of the MPH-type with time-dependence $\psi = z'$.

Proof

We first prove the second assertion.
Define $\underline{r} \overset{d}{=} \underline{w} - \ln\phi(\underline{x};\beta) - \ln \underline{v}$. Because $\underline{w}$ has a Type-1 Extreme Value distribution and is independent of $(\underline{x},\underline{v})$ it follows that

$$\Pr(\underline{r} > r \mid v,x) = \Pr(\underline{w} > \ln\phi(x;\beta) + \ln v + r \mid v,x)$$

(2.5b)

$$= \exp\{-\phi(x;\beta)v e^{r}\} \ .$$

From (2.5a) $\underline{t} \overset{d}{=} z^{-1}(e^{\underline{r}})$. Then it is easily seen that $\underline{t}$ has a MPH-distribution with heterogeneity $\phi(\underline{x};\beta)$, unobserved heterogeneity $\underline{v}$ and time-dependence $z'(t)$.

Next, we prove the first assertion.
Define $\underline{w} \overset{d}{=} \ln z(\underline{t}) + \ln\phi(\underline{x};\beta) + \ln \underline{v}$. Then

$$\text{(2.6)} \quad \Pr(\underline{w} > w \mid v,x) = \Pr(\ln z(\underline{t}) > w - \ln\phi(x;\beta) - \ln v \mid v,x)$$

$$= \exp\{-e^{w}\} \ .$$

Thus $\underline{w}$ has a Type-1 Extreme Value distribution and is independent of $(\underline{x},\underline{v})$. The conditional distribution of $\underline{r}$ which is defined above given v, x is given by (2.5b). It is easily seen that

$$\text{(2.7)} \quad \Pr(\ln z(\underline{t}) > r \mid v,x) = \exp\{-\phi(x;\beta)ve^{r}\}$$

Therefore the random variables on the left-hand and right-hand side of (2.5a) have identical distributions. □

Note that the representation in (2.5a) is unique (Elbers and Ridder (1982)). Note also that in (2.5a) $\underline{x}$ is a random variable. If, as usual, $\underline{x}$ is ancillary for the parameters of interest, there is no difference between the fixed and random covariate case.

If $\phi(x;\beta) = \exp\{x'\beta\}$ it follows that the MPH-model is equivalent to a linear regression model with an unknown form of the dependent variable and a variance components error term. The PH-model is the special case in which $\underline{v}$ is concentrated in 1. From (2.5) we see that there is a relation between the MPH-model and the Accelerated Failure-Time (AFT) model. The AFT model is specified as

(2.9) $\ln \underline{t} = \beta'x + \underline{e}$

where $\underline{e}$ is an error term with an unspecified distribution. All AFT-models in which $\underline{e}$ is a convolution of a Type-1 Extreme Value distribution and a distribution of a random variable $\underline{v}$ normalized such that $E(e^{-\underline{v}})=1$ belong to the class of MPH-models. As is well-known the only PH-model that leads to a log-linear specification is the Weibull-model (Kalbfleisch and Prentice (1980)).

Equation (2.5a) is useful in generating observations in a Monte-Carlo study. We find

(2.10) $\underline{t} \overset{d}{=} z^{-1}[\exp\{-\beta'x - \ln \underline{v} + \underline{w}\}]$

Another application of (2.5a) is to a decomposition of the total variation in $\ln z$ ($\underline{t}$) into an observed and an unobserved part. The explained fraction of the heterogeneity is

(2.11) $$R^2=\frac{\text{var}(\underline{x}'\beta)}{\text{var}(\underline{x}'\beta) + \text{var}(\ln \underline{v})}$$

which as the notation suggests can be interpreted analogously to the coefficient of determination in a regression model. This measure will be used below.

3. THE CONSEQUENCES OF UNOBSERVED HETEROGENEITY FOR THE SEMI-PARAMETRIC ESTIMATION OF THE PH-MODEL

3.1. Preliminaries

We next consider the consequences of neglected heterogeneity for the estimates of the regression coefficients and the time-dependence function (the base-line hazard) of the PH-model. Throughout we consider situations in which there is no censoring and there are no ties in the data. Therefore if the data are (t_i, x_i) $i=1,\ldots,N$ (N is the number of observations) and we order the observations according to $t_1 < t_2 < \ldots < t_N$, then the estimate of β is obtained by maximizing

$$(3.1) \qquad L(\beta) = \frac{e^{\beta' \sum_{i=1}^{N} x_i}}{\prod_{i=1}^{N} \left\{ \sum_{k=i}^{N} e^{\beta' x_k} \right\}}$$

Under the assumptions made above this likelihood can alternatively be considered as the marginal likelihood based on the rank statistic (Kalbfleisch and Prentice (1973)) and as a partially maximized (with respect to the base-line hazard) likelihood of a multivariate counting process (Johansen (1983)). In both cases $L(\beta)$ can be treated as an ordinary likelihood. Given ML-estimates of β, we can estimate ψ. Instead of using a discontinuous estimator of the survivor function, we follow the suggestion of Kalbfleisch and Prentice (1973) and approximate $\psi(t)$ by a piecewise constant function i.e. we assume

$$\begin{aligned} \psi(t) &= \psi_1 \qquad 0 \leq t < b_1 \quad ; \quad t \in I_1 \\ (3.2) \qquad &= \psi_2 \qquad b_1 \leq t < b_1 + b_2 \quad ; \quad t \in I_2 \\ &\vdots \qquad\qquad \vdots \\ &= \psi_r \qquad b_1 + \ldots + b_{r-1} \leq t < \infty \quad ; \quad t \in I_r \,. \end{aligned}$$

The estimates of $\psi_1,\ldots,\psi_r$ are (Kalbfleisch and Prentice (1973) give an erroneous expression for $\hat{\psi}_r$)

$$\begin{aligned} \hat{\psi}_i &= \frac{d_i}{D_i + b_i C_i} \qquad i=1,\ldots,r-1 \\ (3.3) \qquad &= \frac{d_i}{D_i} \qquad i=r \,, \end{aligned}$$

with d_i the number of failures in I_i and

$$(3.4a) \qquad C_i = \sum_{\ell=i+1}^{r} \sum_{j \in J_\ell} \exp\{\hat{\beta}' x_j\}$$

$$(3.4b) \qquad D_i = \sum_{j \in J_i} (t_j - b_1 - \ldots - b_{i-1}) \exp\{\hat{\beta}' x_j\}$$

where J_i is the index set of observations with failure time in I_i. The reason for using the step-function approach is that it allows for a clearer comparison of estimates under different assumptions.

We use a Monte Carlo approach to study the behaviour of the estimates of β and ψ_i. However, before we describe the design of the experiments and the results, it is useful to reflect on the likely outcome of the experiments. In their paper Lancaster and Nickell (1980) study the hazard of the MPH-model. They find that the observed time-dependence function of the MPH-model i.e. the function

$$(3.5) \qquad \hat{\psi}(t) = \frac{f(t|x;\beta)}{\phi(x;\beta)\bar{F}(t|x;\beta)}$$

with $\bar{F}$ given by (2.4) and f the corresponding density, is always smaller than the underlying time-dependence function $\psi(t)$ and that the difference increases with t. Furthermore, they conclude that the effect of x on the hazard of the MPH-model is smaller (in absolute value) than the true effect which is given by $\frac{\partial \phi}{\partial x}$.

In advance, it is not clear how this will affect the estimates of β and the ψ_i. After all, the semi-parametric estimates make use of only a part of the information (the rank statistic). The above suggestions contradict the conjecture of Andersen (1983) that neglecting unobserved heterogeneity will only lead to underestimation of the variances of the regression parameters.

3.2. The Monte Carlo design and the simulation results

Throughout we generate durations according to

$$(3.6) \qquad \underline{t}_i = z^{-1}\left(\frac{\underline{a}_i}{e^{\underline{x}_i'\beta}\,\underline{v}_i}\right)$$

The components in (3.6) are specified as follows

a. Random variables

$$\underline{a} \overset{d}{=} \text{exponential ; mean 1}$$

(3.7) $$\underline{v} \overset{d}{=} \text{lognormal ; mean 1, variance differs between experiments}$$

$$\underline{x} \overset{d}{=} \text{multinormal ; diagonal covariance matrix (independent regressors), means (3.5, 1.0, -.5), standard deviations (.35, .90, .50).}$$

b. Parameters

$$\beta_1 = -.80$$
$$\beta_2 = -.40$$
$$\beta_3 = -.50$$

c. Time-dependence function

$$\psi(t) = 1 \qquad 0 \leq t < 10$$
$$= 1.25 \qquad 10 \leq t < 40$$
$$= .75 \qquad t \geq 40 \ .$$

The values of the parameters and the means and variances of the covariates are inspired by the study of Lancaster (1979). Note that we do not specify a value for β_0, the constant in the regression part of the model. The reason for this is that we varied the value of β_0 from experiment to experiment.

The location of the distribution of $\underline{t}$ is very sensitive to changes in e.g. the variance of the distribution of $\underline{v}$. To control for this, we adjusted β_0 between experiments, so that the median of $\underline{t}$ in every case was approximately equal to 20. In estimating the base-line hazard we used the following five-step approximation

$$\psi(t) = 1 \qquad 0 \leq t < 10$$
$$= \psi_2 \qquad 10 \leq t < 20$$
(3.10) $$= \psi_3 \qquad 20 \leq t < 40$$
$$= \psi_4 \qquad 40 \leq t < 80$$
$$= \psi_5 \qquad 80 \leq t$$

i.e. we estimate the normalized time-dependence function characterized by $\psi_2,\ldots,\psi_5$.

The results of the Monte Carlo experiments are reported in Table 3.1-3.2. In these tables R^2 refers to the measure defined in (2.11). We first consider the results for the correctly specified model ($R^2=1$).

Here Table 3.1

Here Table 3.2

The biases in the regression coefficients are small. The standard errors of the regression coefficients are correctly estimated. The asymptotic normal distribution of the regression estimates applies. The biases in the estimates of base-line hazard funtion are small.

We also considered samples of size 50. For that sample size we obtained the same results, except that the step sizes of the time dependence functions are clearly overestimated. For samples of size 50 we also studied the quality of the normal approximation to the distribution of $\hat{\psi}_i$'s. We concluded that the construction of e.g. confidence intervals for $\hat{\psi}_i$ should be based on the aymptotic normal distribution of $\ln \hat{\psi}_i$. Moreover we found that the estimated standard errors of $\ln \hat{\psi}_i$ only slightly underestimated the sample standard deviations.

If there is neglected heterogeneity ($R^2<1$), we find that the regression coefficients are biased towards 0 . The size of the bias increases as the importance of the neglected heterogeneity increases (i.e. if R^2 decreases). Note that the relative bias is substantial only if $R^2<.5$. It is interesting to note that the estimated standard errors of the regression coefficients are approximately correct. Therefore, tests of no effect of a particular covariate have correct size but lower power. If R^2 decreases the sample variance of the estimates does not increase. Note the difference with the result that would obtain in a linear regression model. The normal approximation to the asymptotic distribution of the (biased) estimates is quite good. Unobserved heterogeneity affects the location of the likelihood function but not its curvature.

Neglected heterogeneity also leads to underestimation of the ψ_i's. This implies that unobserved heterogeneity gives an estimate of the base-

line hzard which is less increasing or more decreasing than the true base-line hazard. The bias is considerable and increases if R^2 decreases. The sample variances of the estimates (and also the estimated variances) decrease if R^2 decreases.

For samples of size 50 we obtained the same results. However, the overestimation of the ψ_i's if $R^2=1$, led to smaller biases if $R^2<1$. In a preliminary study (Verbakel (1983)) it was found that if $R^2=1$ the discrete approximation to a continuous time dependence function is accurate. Moreover, it was found that the conclusions in the text are not altered if we choose (in the generating model) another form of the base-line hazard or another distribution of the unobserved heterogeneity.

From these results it can be concluded that the semi-parametric estimates share the properties of the corresponding components of the MPH-model. In the next section we give an analytic derivation of the biases which also allows for an interpretation of the results obtained in the Monte-Carlo study.

3.3. Interpretation of the bias

In this section we present a heuristic analysis of the results of section 3.2. We focus on the bias in the regression coefficients. We will show that the bias arises because of a dynamic selection process which causes a correlation between the covariates and the unobserved heterogeneity. This correlation biases the estimates (see section 1).

We assume that we have observations (t_i, x_i); $i = 1,\ldots,N$ with $t_1<t_2<\ldots<t_N$. Let the event H_i denote the failure history of the sample over the time interval $[0,t_i)$. Thus H_i records the failure times during $[0,t_i)$ and the failure of a subject with an as yet unknown identity at t_i. For simplicity of exposition we assume that there is no censoring. The population hazard is given by

$$\lambda_0(t, x, \underline{v}; \beta_0) = e^{x'\beta_0}\psi_0(t)\underline{v} \quad . \tag{3.11}$$

The assumed hazard in the construction of (3.1) is

$$\lambda(t, x; \beta) = e^{x'\beta}\psi(t) \quad . \tag{3.12}$$

Let $\underline{\ell}$ denote the random index of the subject failing at t_i. Of course, the sample space of $\underline{\ell}$ is the risk set at t_i. The contribution of the i-th (ordered) observation to the likelihood (3.1) is

$$L_i(\beta) = \frac{e^{x_i'\beta}}{\sum_{k=i}^{N} e^{x_k'\beta}} \quad . \tag{3.13}$$

If (3.12) were correct, (3.13) would be $\Pr(\underline{\ell}=i \mid H_i)$. However, the true probability of this event is

$$\Pr(\underline{\ell}=i \mid H_i) = E_v\left[\frac{e^{x_i'\beta}\underline{v}_i}{\sum_{k=i}^{N} e^{x_k'\beta}\underline{v}_k} \mid \underline{t} \geq t_i\right] \tag{3.14}$$

where the expectation is taken with respect to the joint distribution of $\underline{v}_i,\ldots,\underline{v}_N$ given $\underline{t} \geq t_i$.

If we define

$$(3.15a) \qquad U_i(\beta) = \frac{\partial \ln L_i(\beta)}{\partial \beta}$$

then the expected (the expectation is with respect to the distribution of $\underline{\ell}$) conditional score is given by

$$(3.15b) \qquad E(U_{\underline{\ell}}(\beta) \mid H_i) = \sum_{j=i}^{N} U_j(\beta) \Pr(\underline{\ell} = j \mid H_i) \quad .$$

where $U_j(\beta)$ is the score that would obtain if j fails at t_i (i.e. the risk set is the same). Because

$$(3.16) \qquad U_j(\beta) = x_j - \sum_{k=i}^{N} \frac{e^{x_k'\beta}}{\sum_{m=i}^{N} e^{x_m'\beta}} x_k$$

we find

$$(3.17) \qquad E(U_{\underline{\ell}}(\beta) \mid H_i) = E_v\left[\frac{\sum_{k=i}^{N} x_k v_k e^{\beta_0' x_k}}{\sum_{k=i}^{N} v_k e^{\beta_0' x_k}} \mid \underline{t} \geq t_i \right] - \frac{\sum_{k=i}^{N} e^{\beta' x_k} x_k}{\sum_{k=i}^{N} e^{\beta' x_k}}$$

Let $\underline{x}_k$; k = 1, 2, ... be a sequence of i.i.d. random variables (or vectors). This covers the cases of randomized experiments and behavioral studies. We make this assumption for expositional reasons; the non-random covariates case can be analyzed in a similar fashion. Then for k = i, i+1, ... the vectors $(\underline{x}_k, \underline{v}_k)$ in (3.17) are i.i.d. random vectors with a common distribution identical to that of $(\underline{x}, \underline{v})$ given $\underline{t} \geq t_i$. Therefore if $N \to \infty$,

$$(3.18) \qquad E(U_{\underline{\ell}}(\beta) \mid H_i) \to \frac{E_{v,x}(\underline{x}\underline{v}e^{\beta_0'\underline{x}} \mid \underline{t} \geq t_i)}{E_{v,x}(\underline{v}e^{\beta_0'\underline{x}} \mid \underline{t} \geq t_i)} - \frac{E_x(\underline{x}e^{\beta'\underline{x}} \mid \underline{t} \geq t_i)}{E_x(e^{\beta'\underline{x}} \mid \underline{t} \geq t_i)}$$

It is easily seen that if $\underline{v} \equiv 1$ and $\beta = \beta_0$, the expected conditional score vanishes. This ensures the consistency of the Cox-estimator under these assumptions. The same conclusion can be drawn if $\underline{x}$ and $\underline{v}$ are independent given $\underline{t} \geq t_i$ (i.e. among the survivors at t_i). However, $\underline{x}$ and $\underline{v}$ are not independent given $\underline{t} \geq t_i$. To show this we consider the joint distribution of $\underline{v}$ and $\underline{x}$ given $\underline{t} \geq t$ with density (we omit the normalizing

constant)

$$(3.19) \qquad k(v,x \mid \underline{t} \geq t) \propto g(v)h(x)\exp\{-z(t)e^{x'\beta_0}v\}$$

Next we show that $\underline{x}$ and $\underline{v}$ are correlated among the survivors at t. For ease of exposition we consider the case with one regressor. Some calculations give that if $H(x) = E(\underline{v}|x,\underline{t}\geq t)$ then (see Appendix)

$$(3.20) \qquad H'(x) = -\beta_0 z(t)e^{x\beta_0}\,\mathrm{var}(\underline{v}|x,\ \underline{t}\geq t) \ .$$

Therefore

$$(3.21) \qquad H'(x) \gtrless 0 \iff \beta_0 \lessgtr 0 \ .$$

From (3.21) we conclude that

$$(3.22) \qquad \mathrm{cov}(\underline{x},\underline{v} \mid \underline{t}\geq t) \gtrless 0 \iff \beta_0 \lessgtr 0 \ .$$

Returning to (3.18), denote this limit by $K(\beta)$. Differentiation gives

$$(3.23) \qquad K'(\beta) = -\frac{E(\underline{x}^2 e^{\beta\underline{x}}|\underline{t}\geq t_i)}{E(e^{\beta\underline{x}}|\underline{t}\geq t_i)} + \left[\frac{E(\underline{x}\, e^{\beta\underline{x}}|\underline{t}\geq t_i)}{E(e^{\beta\underline{x}}|\underline{t}\geq t_i)}\right]^2$$

An application of the Cauchy-Schwarz inequality for integrals gives for all β

$$(3.24) \qquad K'(\beta) \leq 0 \ .$$

Because (we omit the normalizing constant)

$$(3.25) \qquad k(v,x \mid \underline{t}=t) \propto ve^{\beta_0 x}g(v)h(x)e^{-z(t)\exp(\beta_0 x)v}$$

it follows that

$$(3.26) \qquad K(\beta_0) = \frac{E(\underline{x}|\underline{t}=t_i)E(\frac{1}{\underline{v}}|\underline{t}=t_i) - E(\frac{\underline{x}}{\underline{v}}|\underline{t}=t_i)}{E(\frac{1}{\underline{v}}|\underline{t}=t_i)} \ .$$

Thus t_i does not contribute to the bias if and only if $\underline{x}$ and $\frac{1}{\underline{v}}$ are uncorrelated among the failures at t_i. A sufficient condition for this is, of course, independence of $\underline{x}$ and $\underline{v}$ among the failures, which in turn is equivalent to independence of $\underline{x}$ and $\underline{v}$ among the survivors. The sign of $K(\beta_0)$ can be derived by noting that (3.22) was derived without making any assumption about the distribution of $\underline{x}$ (given $\underline{t} \geq t$). Rewriting $K(\beta_0)$ as

$$(3.27) \qquad K(\beta_0) = \frac{E(e^{\beta_0 \underline{x}} | \underline{t} > t_i)}{E(\underline{v} e^{\beta_0 \underline{x}} | \underline{t} > t_i)} \left\{ \frac{E(\underline{x}\underline{v} e^{\beta_0 \underline{x}} | \underline{t} > t_i)}{E(e^{\beta_0 \underline{x}} | \underline{t} > t_i)} - \frac{E(\underline{x} e^{\beta_0 \underline{x}} | \underline{t} > t_i) E(\underline{v} e^{\beta_0 \underline{x}} | \underline{t} > t_i)}{E(e^{\beta_0 \underline{x}} | \underline{t} > t_i)^2} \right\}$$

and replacing $h(x)$ in (3.25) by $h(x)e^{\beta_0 x}$, gives on adjusting the normalizing constants

$$(3.28) \qquad K(\beta_0) \gtrless 0 \iff \beta_0 \lesseqgtr 0 .$$

If we define β^* by

$$(3.29) \qquad K(\beta^*) = 0$$

then combining (3.28) and (3.24) gives

$$(3.30) \qquad \beta_0 \gtrless 0 \Rightarrow \beta^* \lesseqgtr \beta_0 .$$

One can ask whether there can be a change of sign. A necessary and sufficient condition for this to occur is

$$(3.31) \qquad K(0) \gtrless 0 \quad \text{and} \quad \beta_0 \gtrless 0 .$$

Now it is easily seen that

$$(3.32) \qquad K(0) = C \operatorname{cov} (\underline{x}, \underline{v} e^{\beta_0 \underline{x}} \mid \underline{t} \geq t_i)$$

with C a positive constant. Now if $\beta_0 \downarrow 0$ then the covariance in (3.32) will become negative (this follows from (3.22)). Therefore if β_0 is small, β^* may have an opposite sign.

If we use the first M observations (i.e. the sample is censored after the Mth failure) we have if $N \to \infty$ and if

$$U^M(\beta) = \sum_{i=1}^{M} U_i(\beta) \tag{3.33}$$

that

$$E(\underline{U}^M(\beta)) = \sum_{i=1}^{M} E(\underline{U}_i(\beta) \mid H_i) \ . \tag{3.34}$$

Because for all $i = 1, \ldots, M$ the solution β_i^* of $E(\underline{U}_i(\beta) \mid H_i)$ satisfies (3.30) we have that the solution $\tilde{\beta}_M$ of $E(\underline{U}^M(\beta)) = 0$ satisfies the same relation. This establishes the direction of the bias.

Note that the essential point in this proof is that there is a correlation between $\underline{x}$ and $\underline{v}$ among the survivors at t. If e.g. $\beta_0 > 0$ then subjects with a large value of x will leave first except if this large value of x is ofset by a small value of v. So among the survivors there will be a negative correlation between $\underline{x}$ and $\underline{v}$.

4. CONCLUSIONS

We have shown that application of the semi-parametric Cox-method in cases where there are omitted regressors, leads to underestimation of the regression coefficients. We have also indicated the relevance of this (asymptotic) result in a Monte Carlo study. As could be expected, the magnitude of the bias depends on the relative importance of the omitted regressors (relative to the importance of the included regressors). The relative importance of the omitted covariates has to be substantial before the bias in the regression coefficients, is sizeable. In this sense the Cox-method is robust against omitted covariates.
A consequence of the results of this paper is that estimation of treatment effects using data on heterogeneous individuals, will give underestimates of the true effects. Randomization does not alter this conclusion. The estimated standard errors of the estimates are not affected by the misspecification, so that the usual tests for $\beta = 0$ have the correct size but have less power. The proof suggests a test for omitted heterogeneity. If we order the observed durations (which may be censored) in order of increasing length and if we censor the sample at e.g. the median and estimate the regression coefficients using this sample, then we can compare these estimates with the estimates obtained from full sample. If there is omitted heterogeneity, the estimate from the artificially censored sample will be larger than the estimate from the full sample.

REFERENCES

[1] Andersen, P.K., The counting process approach to the statistical analysis of labour force dynamics; Research report Statistical Research Unit (1983).

[2] Cox , D.R., Regression models and life tables; Journal of the Roy. Stat. Soc, vol. 34, series B, pp.187-200 (1972).

[3] Cox , D.R. Partial likelihood; Biometrika, vol. 62, pp.269-274 (1975).

[4] Elbers, C. and G. Ridder, True and spurious duration dependence: the identifiability of the proportional hazard model; Rev. of Ec. Studies, XLIX, pp.403-409 (1982).

[5] Gill, R.D., Understanding Cox's regression model: a martingale approach; Preprint Mathematisch Centrum (1982).

[6] Hougaard, P., Life table methods for heterogeneous populations: distributions describing the heterogeneity; Research report Statistical Research Unit (1982).

[7] Johansen, S., An extension of Cox's regression model; Intern. Stat. Rev., vol. 51, pp.165-174 (1983).

[8] Johnson,N.L. and S. Kotz, Continuous univariate distributions, vol. 1, Houghton Miflin, Boston (1970).

[9] Kalbfleisch, J.D. and R.L. Prentice, Marginal likelihoods based on Cox's regression and life model; Biometrika, vol. 60, pp.267-278 (1973).

[10] Kalbfleisch, J.D. and R.L. Prentice, The statistical analysis of failure time data, Wiley, New York (1980).

[11] Lancaster, T., Econometric models for the duration of unemployment; Econometrica, vol. 47, no. 4, pp.939-956 (1979).

[12] Lancaster, T. and S. Nickell, The analysis of the re-employment probabilities for the unemployed; Journal of the Roy. Stat. Soc., vol. 143, series A, pp.141-165 (1980).

[13] Theil, H., Principles of econometrics; North-Holland, Amsterdam (1971).

[14] Vaupel, J.W., K.G. Manton and E. Stallard, The impact of heterogeneity in individual frailty on the dynamics of mortality; Demography, vol. 16, no. 3, pp.439-454 (1979).

[15] Verbakel, W., Misspecification of the proportional hazards model (in Dutch), M.A. thesis, University of Amsterdam (1983).

APPENDIX. The derivation of (3.20).

From (3.19) it follows that

$$(A.1) \qquad k(v|x,\underline{t}\geq t) = \frac{g(v)\exp\{-z(t)e^{x\beta_0}v\}}{\int_0^\infty g(v)\exp\{-z(t)e^{x\beta_0}v\}dv}$$

If we define

$$(A.2) \qquad H(x) = E(\underline{v}|x,\underline{t}\geq t)$$

then

(A.3)

$$H'(x) = -z(t)e^{x\beta_0}\beta_0 \frac{\int_0^\infty g(v)\exp\{-z(t)e^{x\beta_0}v\}dv \int_0^\infty v^2 g(v)\exp\{-z(t)e^{x\beta_0}v\}dv - \left\{\int_0^\infty vg(v)\exp\{-z(t)e^{x\beta_0}v\}dv\right\}^2}{\left[\int_0^\infty g(v)\exp\{-z(t)e^{x\beta_0}v\}dv\right]^2}$$

$$= -z(t)e^{x\beta_0}\beta_0\left[E(\underline{v}^2|x,\underline{t}\geq t) - [E(\underline{v}|x,\underline{t}\geq t)]^2\right]$$

$$= -z(t)e^{x\beta_0}\beta_0 \operatorname{var}(\underline{v}|x,\underline{t}\geq t)$$

Table 3.1. Regression coefficients, semi-parametric estimation - N=500, 50 replications

R^2	Mean bias (t-ratio;df. 49) β_1	β_2	β_3	Mean standard errors; estimated (sample standard deviation of estimators) β_1	β_2	β_3	Normality of estimators; significance levels of K-S test [a] β_1	β_2	β_3
1.0	-.029	-.000	-.015	.135	.053	.093	.89	.91	.51
	(-1.59)	(-0.00)	(-1.26)	(.129)	(.067)	(.084)			
.75	0.051	0.019	.036	.134	.053	.093	1.00	.96	.88
	(2.60)	(2.42)	(2.97))	(.139)	(.056)	(.085)			
.50	.128	.059	.128	.133	.052	.092	1.00	.97	.89
	(5.74)	(7.98)	(10.25)	(.157)	(.052)	(.084)			
.10	.427	.203	.244	.131	.051	.090	.92	.97	.87
	(25.93)	(23.89)	(21.62)	(.116)	(.060)	(.080)			

a) K-S test ≡ Kolomogorov-Smirnov test for normality.

Table 3.2. Base-line hazard , step estimates
N=500, 50 replications

R^2	Mean bias (t-ratio; df.49) ψ_2	ψ_3	ψ_4	ψ_5	Sample standard deviation of estimates ψ_2	ψ_3	ψ_4	ψ_5
1.0	-.001	-.002	.003	.013	.144	.150	.049	.128
	(-.04)	(-.09)	(.24)	(.73)				
.75	-.079	-.087	-.098	-.157	.153	.145	.092	.092
	(-3.60)	(-4.22)	(-7.44)	(-11.93)				
.50	-.110	-.267	-.223	-.337	.129	.116	.078	.067
	(-6.00)	(-16.15)	(-19.91)	(-32.26)				
.10	-.610	-.828	-.578	-.696	.088	.057	.022	.014
	(-48.34)	(-102.37)	(-180.94)	(-356.53)				