

MINIMUM QUADRATIC DISTANCE ESTIMATION
FOR THE PROPORTIONAL HAZARDS REGRESSION MODEL
WITH GROUPED DATA.

JACQUES RIOUX AND ANDREW LUONG

Université Laval, Québec, Canada

ABSTRACT. In this paper, we study a survival regression model known as Cox's proportional hazards model. We assume that the data available are already grouped and develop a minimum distance estimation procedure that is also applicable when the data are also left or right censored. Asymptotic properties of the estimators are established and the numerical implementation of the method based on an iterated reweighted least squares algorithm is discussed.

1. INTRODUCTION

If one wants to estimate the survival function from a sample, one could suppose that all individuals under observation have the same lifetime distribution. However, in practice, most samples are somewhat heterogeneous. Consequently, one must take into account a number of factors having a direct influence on each individual's survival time.

One of the most popular model used to deal with this heterogeneity is Cox's proportional hazards regression model. According to it, any two individuals **A** and **B** with vectors of concomitant variables \mathbf{x}_A and \mathbf{x}_B respectively have hazard functions $h(t | \mathbf{x}_A)$ and $h(t | \mathbf{x}_B)$ such that $h(t | \mathbf{x}_A)/h(t | \mathbf{x}_B)$ is a constant function of t . One immediate implication of the model is that

$$h(t | \mathbf{x}) = h_0(t) \cdot g(\mathbf{x})$$

Here $h_0(t)$ is called the baseline hazard function and in general we restrict ourselves to those models for which $g(\mathbf{x})$ has a particular form, namely

$$g(\mathbf{x}) = e^{\mathbf{x}'\boldsymbol{\beta}}$$

Key words and phrases. Proportional hazard; Censored data; Quadratic distance estimation; Weighted least squares; Grouped data; Life tables.

The second author wishes to acknowledge support for part of this research from the National Science and Engineering Research Council of Canada

so that

$$h(t | \mathbf{x}) = h_0(t)e^{\mathbf{x}'\boldsymbol{\beta}}$$

where $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ is any vector of regressors and $\boldsymbol{\beta}' = (\beta_1, \dots, \beta_p)$ is the unknown vector of parameters of interest.

For this model, the survival time function can be written as

$$S(t | \mathbf{x}) = S_0(t)^{\exp(\mathbf{x}'\boldsymbol{\beta})}$$

where $S_0(t)$ is called the baseline survival function, see Cox (1972), Kalbfleisch and Prentice (1980), Lawless (1982) and Amemiya (1988) for example, for more discussion on hazard functions and proportional hazard model.

The proportional hazards model includes the Weibull and exponential models as particular cases. It has been used to model duration time until a special event occurs. Among many successful applications beside modelling survival times or remission time in cancer study, we mention modelling of compensation time in actuarial science, duration of working time in pension study and duration of unemployment time in economics see Lancaster (1979), Butler and Worrall (1985) or Johnson and Ondrich (1990) for example.

Estimation of the proportional hazards model has been studied essentially with maximum likelihood methods, see Lawless (1982), Kalbfleisch and Prentice (1980). Looking at one of these references, one can quickly see that although the asymptotic properties of those estimators are quite good in general, the calculations are fairly complicated and can get a lot more involved when censored data are present. Our minimum quadratic distance approach is very simple to compute and as we'll indicate, can be easily adapted to handle left or right censored data. In fact, anybody familiar with any linear regression package can use it to find our estimator after a little manipulation of the data. Ryu (1994) gets essentially the same estimator as we do here, however, our quadratic distance approach allows us to unify the problems of estimating the model's parameters and testing the goodness of fit. This is what we propose to show in subsequent paper where we will extend our method to other models of regression survival models.

The situation we study is the following. We have a sample of independent observations which is supposed to follow a proportional hazards survival model. The concomitant or covariate characterization vector $\mathbf{x} = (x_1, \dots, x_p)$ of each individual is known. Moreover, we suppose that $\forall i$ in $\{1, \dots, p\}$, x_i can only take a finite number $n(i)$ of values $\{x_i^1, x_i^2, \dots, x_i^{n(i)}\}$. Note that this setup is particularly suitable when we have indicator variables taking on value 0 or 1 for example.

Finally, we suppose that lifetimes are broken into J intervals $I_j = [a_{j-1}, a_j)$ $j = 1, \dots, J$ with $0 = a_0 < a_1 < \dots < a_{J-1} < a_J$.

We observe for each regressor vector \mathbf{x} , the number $n(\mathbf{x}, j)$ of individuals of type \mathbf{x} at the beginning of I_j and $c(\mathbf{x}, j)$ the number of those that survived until the end

of I_j .

From the situation above, we want to get estimators $\hat{\beta}$ of β as well as $\hat{S}_0(t)$ of $S_0(t)$ simultaneously. Consequently, our estimators are semi-parametric estimators since we do not require $S_0(t)$ to be specified. Furthermore, we shall see our estimators are fairly easy to compute and will be based on grouped data. It happens often that data is already grouped, such as in a life table study, our method is thus particularly suitable.

We introduce our quadratic distance estimators in section 2 and derive its asymptotic properties. Procedures to estimate the survival function based on quadratic distance estimators are given here. In section 3, we show the methods can also handle censored data. Adjustments are easy and outlined here. In section 4, we show the procedure how to obtain our quadratic distance estimator numerically based on an iterated reweighted least squares algorithm and finally a numerical example is included at the very end.

2. MINIMUM QUADRATIC DISTANCE ESTIMATION

Now that the problem is stated, let's list all that we know. Since each x_i takes only a finite number of different values, there are only a finite number l of different vectors \mathbf{x} . We chose an ordering of those l vectors $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(l)}$ as in Amemiya (1988, p.275) and we let this ordering be fixed from now on. This way, one can classify all the covariates into only one of the l classes represented by $\{\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \dots, \mathbf{x}_{(l)}\}$. Moreover, we assume that there are replications of observations from each $\mathbf{x}_{(i)}$ covariate for large samples.

To simplify notation, we write $n_{ij} = n(\mathbf{x}_{(i)}, j)$ for the number of individuals with concomitant variables vector $\mathbf{x}_{(i)}$ present in the study at the beginning of interval I_j and $c_{ij} = c(\mathbf{x}_{(i)}, j)$ for the number of individuals with regression vector $\mathbf{x}_{(i)}$ still alive at the end of interval I_j .

We then write $\hat{p}_{ij} = c_{ij}/n_{ij}$. As mentioned in London (1988) or Lawless (1982), conditioning on the set $\{n_{ij}\}$, the distribution of each c_{ij} is binomial. We compute the mean and variance of \hat{p}_{ij} as follows. First, let's define

$$\begin{aligned} P_j(\mathbf{x}) &= S(a_j | \mathbf{x}) \\ P_0(\mathbf{x}) &= 1 \quad \forall \mathbf{x} \end{aligned}$$

and

$$p_j(\mathbf{x}) = \frac{P_j(\mathbf{x})}{P_{j-1}(\mathbf{x})} \quad j = 1, 2, \dots, J.$$

If we also let $P_j = P_j(0)$ and $p_j = p_j(0)$ then one can easily verify that

$$p_j(\mathbf{x}) = p_j^{\exp(\mathbf{x}'\beta)}$$

is the probability of survival in interval j for an individual with covariate \mathbf{x} . Therefore

$$E[\hat{p}_{ij}] = p_j^{\exp(\mathbf{x}_{(i)}'\boldsymbol{\beta})} = p_{ij}, \quad i = 1, \dots, l$$

and

$$\text{Var}[\hat{p}_{ij}] = \frac{p_{ij}(1 - p_{ij})}{n_{ij}}, \quad i = 1, \dots, l.$$

We will be using minimum quadratic distance methods to find estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{p}} = (\hat{p}_1, \dots, \hat{p}_J)$. So we will minimize an expression of the following form

$$\sum_{i=1}^l \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{W_{ij}}.$$

This can also be viewed as a form of minimum chi square method.

We could use $O_{ij} = \hat{p}_{ij}$, $E_{ij} = p_j^{\exp(\mathbf{x}'_{(i)}\boldsymbol{\beta})}$, and $W_{ij} = 1$ for example but the computations would be fairly complicated and the estimators are not efficient. Instead we propose to use the standard transformation $\ln(-\ln(\cdot))$ and let

$$O_{ij} = \ln(-\ln \hat{p}_{ij})$$

and

$$\begin{aligned} E_{ij} &= \ln\left(-\ln p_j^{\exp(\mathbf{x}'_{(i)}\boldsymbol{\beta})}\right) \\ &= \ln(-\ln p_j) + \mathbf{x}'_{(i)}\boldsymbol{\beta} \\ &= \gamma_j + \mathbf{x}'_{(i)}\boldsymbol{\beta} \end{aligned}$$

if we let $\gamma_j = \ln(-\ln p_j)$.

As for W_{ij} , the efficient choice as we shall see will be given by the approximate

variance of O_{ij} that we write $\text{Var}(O_{ij})$

$$\begin{aligned}
 W_{ij} &= \text{Var}(O_{ij}) \\
 &= \left[\frac{d}{dt} (\ln(-\ln t)) \right]^2 \Bigg|_{t=p_{ij}} \times \text{Var}(\hat{p}_{ij}) \\
 &= \left(\frac{1}{t \ln(t)} \right)^2 \Bigg|_{t=p_{ij}} \times \frac{p_{ij}(1-p_{ij})}{n_{ij}} \\
 &= \frac{1}{p_{ij} \ln(p_{ij})^2} \times \frac{1-p_{ij}}{n_{ij}} \\
 &= \frac{1-p_j^{\exp(\mathbf{x}'_{(i)}\boldsymbol{\beta})}}{n_{ij} p_j^{\exp(\mathbf{x}'_{(i)}\boldsymbol{\beta})} \exp(2\mathbf{x}'_{(i)}\boldsymbol{\beta}) \ln(p_j)^2} \\
 W_{ij} &= \frac{(p_j^{-\exp(\mathbf{x}'_{(i)}\boldsymbol{\beta})} - 1) \exp(-2\mathbf{x}'_{(i)}\boldsymbol{\beta})}{n_{ij} (\ln p_j)^2} \tag{2.1}
 \end{aligned}$$

So we have transformed the problem into one where the usual linear regression methods are applicable. That is we have the analogous of a linear model

$$Y_{ij} = \gamma_j + \mathbf{x}'_{(i)}\boldsymbol{\beta} + \varepsilon_{ij}$$

where

$$Y_{ij} = \ln(-\ln(\hat{p}_{ij}))$$

Let's set $\mathbf{B}' = (\beta_1, \dots, \beta_p, \gamma_1, \dots, \gamma_J)$, the vector of parameters of interest. Therefore the same model can be written as:

$$Y_{ij} = \mathbf{Z}'_{ij} \mathbf{B} + \varepsilon_{ij}$$

where $\mathbf{Z}'_{ij} = (\mathbf{x}'_{(i)}, \boldsymbol{\delta}'_j)$ and $\boldsymbol{\delta}_j \in \mathbb{R}^J$ is a vector with a single nonzero component equal to 1 in position j . If we let

$$\begin{aligned}
 \mathbf{Y}' &= (y_{11}, y_{21}, \dots, y_{11}, y_{12}, y_{22}, \dots, y_{12}, \dots, y_{1J}, y_{2J}, \dots, y_{1J}), \\
 \boldsymbol{\varepsilon}' &= (\varepsilon_{11}, \varepsilon_{21}, \dots, \varepsilon_{11}, \varepsilon_{12}, \varepsilon_{22}, \dots, \varepsilon_{12}, \dots, \varepsilon_{1J}, \varepsilon_{2J}, \dots, \varepsilon_{1J})
 \end{aligned}$$

and \mathbf{I}_J denote the $J \times J$ identity matrix, the design matrix \mathbf{X} of the model becomes:

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}'_{(1)} \\ \vdots \\ \mathbf{x}'_{(1)} & \mathbf{I}_J \\ \mathbf{x}'_{(2)} \\ \vdots \\ \mathbf{x}'_{(2)} & \mathbf{I}_J \\ \vdots \\ \vdots \\ \vdots \\ \mathbf{x}'_{(l)} \\ \vdots \\ \mathbf{x}'_{(l)} & \mathbf{I}_J \end{pmatrix}$$

Therefore, using matrix notations, we have the following model:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\varepsilon}.$$

Conditioning on $\{n_{ij}\}$, asymptotically, $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ and

$$V(\boldsymbol{\varepsilon}) = \boldsymbol{\Sigma} = \text{Diag}(W_{11}, W_{21}, \dots, W_{l1}, \dots, W_{1J}, W_{2J}, \dots, W_{lJ}) \quad (2.2)$$

where the W_{ij} 's are defined as in (2.1).

By now, we have transformed the original problem into one for which standard weighted least squares theory applies. If we chose the weights matrix to be $I_{l \times J}$ we get the following estimator for \mathbf{B}

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

$\hat{\mathbf{B}}$ is consistent and asymptotically normal with variance covariance matrix

$$V(\hat{\mathbf{B}}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

where $\boldsymbol{\Sigma}$ is the variance-covariance matrix for \mathbf{Y} , as defined above in (2.2). Now, if we chose $\boldsymbol{\Sigma}$ as the weights matrix, we get the following estimator

$$\hat{\mathbf{B}}_w = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y},$$

which is consistent and asymptotically normal with variance-covariance matrix

$$V(\hat{\mathbf{B}}_w) = (\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X})^{-1}.$$

Theorem 1. $\hat{\mathbf{B}}_w$ is a better estimator than $\hat{\mathbf{B}}$

Proof. : We need to show that $V(\hat{\mathbf{B}}) - V(\hat{\mathbf{B}}_w)$ is non-negatively definite. But

$$\begin{aligned} V(\hat{\mathbf{B}}_w) &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Sigma^{-1})\Sigma(\Sigma^{-1}\mathbf{X})(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Sigma^{-1}\mathbf{X})(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \end{aligned}$$

So if we compute the variance-covariance matrix of the random vector

$$U = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Sigma^{-1})\mathbf{Y}$$

we get

$$\begin{aligned} V(U) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\Sigma^{-1}\mathbf{X}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &\quad - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}(\mathbf{X}'\Sigma^{-1})\Sigma(\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1} + (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\quad - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} + (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Sigma\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1} \\ V(U) &= V(\hat{\mathbf{B}}) - V(\hat{\mathbf{B}}_w). \end{aligned}$$

$V(\hat{\mathbf{B}}) - V(\hat{\mathbf{B}}_w)$ being the variance-covariance matrix of the random vector U is clearly positive semi-definite. \square

Of course from now on we will be using $\hat{\mathbf{B}}_w$ which is the efficient quadratic distance estimator and for notation sake, we write

$$\hat{\mathbf{B}}'_w = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p, \hat{\gamma}_1, \hat{\gamma}_2, \dots, \hat{\gamma}_J)$$

As a result of the above computations, we could compute the covariance matrix for the estimators of our β and consequently $S(a_i|\mathbf{x})$. Obviously, $V(\hat{\beta})$ is easy to find. By letting \mathbf{A} be the $p \times (p + J)$ matrix of the form: $\mathbf{A} = \begin{bmatrix} \mathbf{I}_p & \mathbf{0}_{p \times J} \end{bmatrix}$ then $\hat{\beta} = \mathbf{A}\hat{\mathbf{B}}_w$ and its variance-covariance matrix is $V(\hat{\beta}) = \mathbf{A}(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{A}'$.

Also, let's define \mathbf{P} such that $\mathbf{P}' = (p_1, p_2, \dots, p_J)$ as the vector of the base survival probabilities in each of the J intervals considered. So, from $\hat{\mathbf{B}}_w$, we can extract $\hat{\mathbf{P}}$ the estimator for \mathbf{P} by applying the adequate transformation to the last elements of $\hat{\mathbf{B}}_w$, as in

$$\hat{\mathbf{P}}' = (\exp(-\exp(\hat{\gamma}_1)), \exp(-\exp(\hat{\gamma}_2)), \dots, \exp(-\exp(\hat{\gamma}_J))).$$

Consequently, it is easy to estimate the survival function $S(a_i|\mathbf{x})$ and the base survival function $S_0(a_i)$. The estimators and their asymptotic properties are given in the following theorem, the proof of which consists of using a Taylor expansion type of argument as the one used in deriving (2.1).

Theorem 2. For a given value of the covariate $\mathbf{x}' = (x_1, x_2, \dots, x_p)$, we can estimate the survival function $S(a_i|\mathbf{x})$, with $\widehat{S}(a_i|\mathbf{x}) = \prod_{k=1}^i \hat{p}_k^{\exp(\mathbf{x}'\hat{\beta})}$.

In this case, the approximate variance of $\widehat{S}(a_i|\mathbf{x})$ is

$$\text{Var} \left[\widehat{S}(a_i|\mathbf{x}) \right] = [S(a_i|\mathbf{x})]^2 \exp(2\mathbf{x}'\beta) \mathbf{V}'(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{V}$$

where

$$\mathbf{V}' = \left(-x_1 \sum_{k=1}^i e^{\gamma_k}, -x_2 \sum_{k=1}^i e^{\gamma_k}, \dots, -x_p \sum_{k=1}^i e^{\gamma_k}, -e^{\gamma_1}, -e^{\gamma_2}, \dots, -e^{\gamma_i}, 0, \dots, 0 \right).$$

We can estimate the base survival function $S_0(a_i)$, with $\prod_{k=1}^i \hat{p}_k$ and in this case, the approximate variance of $\widehat{S}_0(a_i)$ is

$$\text{Var} \left[\widehat{S}_0(a_i) \right] = [S_0(a_i)]^2 \exp(2\mathbf{x}'\beta) \mathbf{V}'_0(\mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}\mathbf{V}_0$$

where

$$\mathbf{V}'_0 = \underbrace{(0, \dots, 0)}_{p \text{ times}}, -e^{\gamma_1}, -e^{\gamma_2}, \dots, -e^{\gamma_i}, 0, \dots, 0$$

Note that these variance expressions can be estimated easily in a semi-parametric way.

3. CENSORED DATA

So far, in our exposition, we assumed without stipulation that all withdrawals from a survival study were due to "death" (or due to whatever other cause whose duration time we are trying to estimate) and that all individuals under study are present at age 0 at the beginning of the observation period. However, often in such studies, there will be withdrawals due to other external factors that shouldn't be considered as "deaths", in other words, the data could be censored. Moreover, rather than tossing away those individuals retiring too early, and considering the artificially reduced population resulting, we would like to get as much information as possible from all individuals under study. We do this in the following way.

As we have seen, we are lead to use a general linear regression model to obtain our estimators for the beta's and $S(\cdot|\mathbf{x})$. Essentially, the estimators are based on the \hat{p}_{ij} 's, the probabilities of survival in interval j for individuals with characterization vector $\mathbf{x}_{(i)}$. Obviously, the closer the \hat{p}_{ij} 's reflect reality, the better our estimators should be. This leads us to the notion of exposure.

For a given vector of concomitant variables $\mathbf{x}_{(i)}$, we had defined in section 2 $\hat{p}_{ij} = c_{ij}/n_{ij}$ where n_{ij} is the number of individuals with that vector of concomitant variables present in the study at the beginning of interval I_j and c_{ij} for the number of individuals with the same vector of regression still alive at the end of

interval I_j . Suppose for simplicity that all intervals are of length 1 and that a total of K individuals with vector of concomitant variables $\mathbf{x}_{(i)}$ appear in the interval I_j for some time.

For each of these, define s_k , $k = 1, \dots, K$ as the time of entry into the study and t_k , $k = 1, \dots, K$ as the time of departure from the study for external reasons. Those are left and right censoring times. One way to take all pertinent information into account is to define

$$\hat{p}_{ij} \sim \frac{c_{ij}}{\sum_{k=1}^K (t_k - s_k)},$$

where, the denominator is the exposure measure mentioned in London (1988) and Lawless (1982). We can then apply our procedure to those \hat{p}_{ij} 's, keeping in mind that it is then an approximate one.

4. NUMERICAL IMPLEMENTATION

4.1. Algorithm. To compute the estimator $\hat{\mathbf{B}}$ is fairly straightforward with any statistical package that handles multiple linear regression. As a matter of fact, the only non-trivial work involved is in organizing the data correctly and generating the design matrix \mathbf{X} , obtaining $\hat{\mathbf{B}}$ is then simply a matter of using already available routines.

The efficient quadratic distance estimator $\hat{\mathbf{B}}_w$ though depends on the \hat{p}_{ij} 's and the $\hat{\beta}$'s. However, those quantities are not available a priori but easily estimated once one has an initial estimate of \mathbf{B} . This leads clearly to an iterative procedure where one first estimates \mathbf{B} by $\hat{\mathbf{B}}$, extracts the \hat{p}_{ij} 's and the $\hat{\beta}$'s from it and then uses those values to compute the entries of the weight matrix W . From this point on, one repeatedly computes $\hat{\mathbf{B}}_w$ and updates W until a criterion for convergence is met. So we use a series of iterated reweighted least squares procedures.

In practice, in all examples of computations we have tried, the first estimates $\hat{\mathbf{B}}$ were slightly off, though the weight matrix generated gave a preliminary $\hat{\mathbf{B}}_w$ quite different from $\hat{\mathbf{B}}$. The next couple of iterations gave it a slight correction and from that point on, any further iteration gave relative changes of far less than 0.01% for each parameter. In all examples we tried, we never needed more than 3 or 4 iterations to obtain an acceptable accuracy.

4.2. Example. We have implemented our method in S-Plus. We wrote code that would create the design matrix \mathbf{X} from the situation at hand. We haven't tested it on real life data but we did test it extensively by simulation. In the example described here, we decided to let $\beta_0 = (0.1, 0.3, 0.2)$ while the values that could be taken by \mathbf{x} were $\mathbf{x} = (x_1, x_2, x_3)$ where $x_1 \in \{0, 1\}$, $x_2 \in \{0, 0.4, 0.9\}$ and $x_3 \in \{-1, 1\}$.

We fixed the base survival function to be that of an exponential random variable with parameter $\lambda = 0.1$ so that the base expected time of survival is 10. For the 12 different vectors of covariates, we simulated corresponding times of survival for

an initial sample of 1000. We chose to split the time interval at all quarters from 0 to 3, thus generating 12 sub-intervals. With this setup, we then ran our estimation procedure on the result of the simulations. We repeated this process 200 times to have a good idea of how good our method was. The results are fairly encouraging.

In the following table, we compare the result of our two methods, namely with and without weights.

| | No weights | | With weights | |
|-----------|------------|--------------|--------------|--------------|
| | Mean | Variance | Mean | Variance |
| β_1 | 0.09369945 | 0.001677068 | 0.09302015 | 0.001587944 |
| β_2 | 0.28838041 | 0.002404128 | 0.28927907 | 0.002317426 |
| β_3 | 0.19315782 | 0.000319992 | 0.19340011 | 0.000316183 |
| p_1 | 0.97617147 | 2.64575e-006 | 0.97620276 | 2.58179e-006 |
| p_2 | 0.97623301 | 2.20511e-006 | 0.97620834 | 1.97766e-006 |
| p_3 | 0.97647781 | 2.13237e-006 | 0.97645571 | 2.06813e-006 |
| p_4 | 0.97642031 | 2.60941e-006 | 0.97639730 | 2.39343e-006 |
| p_5 | 0.97655229 | 2.34267e-006 | 0.97655295 | 2.21630e-006 |
| p_6 | 0.97648027 | 2.67874e-006 | 0.97651056 | 2.45656e-006 |
| p_7 | 0.97659894 | 2.50644e-006 | 0.97662187 | 2.52137e-006 |
| p_8 | 0.97651214 | 3.45007e-006 | 0.97650509 | 3.45923e-006 |
| p_9 | 0.97668582 | 2.76196e-006 | 0.97664985 | 2.69636e-006 |
| p_{10} | 0.97654216 | 2.81350e-006 | 0.97654069 | 2.78533e-006 |
| p_{11} | 0.97655769 | 2.09560e-006 | 0.97658714 | 2.03501e-006 |
| p_{12} | 0.97645519 | 2.09560e-006 | 0.97651284 | 2.03513e-006 |

Empirical means and variances for the two estimation procedures.

It appears that the weighted version performs better than the unweighted one, as confirmed by the theory.

4.3. Exceptional cases. Because of the transformation $\ln(-\ln)$ applied to our \hat{p}_{ij} 's, there are a few situations raising concerns. Basically there are two situations which are opposite extremes and that should be avoided.

The first one is when all individuals with the same characterisation vector die in a given interval. The problem is that this would give $\hat{p}_{ij} = 0$ and we couldn't apply the desired transformation. In other words, the observation period has gone too long and it should have been truncated before. Or another way to look at it is to say that the sample isn't large enough. Unfortunately, if the data has already been collected in this way, there is nothing that can be done to fix this.

The other problematic situation is when no individual with a given characterisation vector die in a certain interval. Here, that would mean $\hat{p}_{ij} = 1$ and the second log

can't be taken. Again, we could interpret this as too small a sample. However, this can be easily fixed by extending that interval of observation to contain also the previous or next one.

As a concluding note, we recall that

$$\sum_{i=1}^I \sum_{j=1}^J \frac{(O_{ij} - E_{ij})^2}{W_{ij}}.$$

as introduced in section 2 defines a distance. Consequently, a goodness of fit test statistic can be constructed based on that expression. This will be dealt in another paper.

REFERENCES

1. Amemiya, T. (1988), *Advanced Econometrics*, Harvard University Press, 1988.
2. Butler R. J. and Worrall J. D. (1985), *Work injury compensation and the duration of nonwork spells*, The Economic Journal, 95, 714-724.
3. Cox, D.R. (1972), *Regression models and life tables*, Journal of the Royal Statistical Society, Series B, 33, 187-202.
4. Kalbfleisch, J.D. and Prentice, R.L. (1980), *The statistical analysis of failure time data*, Wiley.
5. Johnson W. J. and Ondrich J. (1990), *The duration of post-injury absences from work*, The Review of Economics and Statistics, 72, 4.
6. Lancaster, T. (1979), *Econometric methods for the duration of unemployment*, Econometrica, 47, 939-956.
7. Lawless, J. F. (1982), *Statistical models and methods for lifetime data*, John Wiley and Sons.
8. London, D. (1988), *Survivals models and their estimation*, ACTEX publications.
9. Ryu, K. (1994), *Group duration analysis of the proportional hazard model: Minimum chi-squared estimation and specification tests*, Journal of the American Statistical Association, Vol 89, No 428, 1386-1397.

ÉCOLE D'ACTUARIAT, UNIVERSITÉ LAVAL, QUÉBEC, QUÉBEC, CANADA G1K-7P4
E-mail address: jrioux@act.ulaval.ca

ÉCOLE D'ACTUARIAT, UNIVERSITÉ LAVAL, QUÉBEC, QUÉBEC, CANADA G1K-7P4
E-mail address: aluong@act.ulaval.ca

