Validating `coxtdw` function coded in R

R.A. Ali, M.A. Ali and Z. Wei

April 8, 2012

1 Introduction

The function `coxtdw` implements a Cox model with time-dependent weights. This algorithm is analogous to that of Binder (1992), where weights are used to adjust for sampling bias, and is appropriate for fitting marginal structural Cox models.

1.1 Notation

Suppose a sample of size n is drawn from a finite population of N units with unequal probability π_i for each subject i. Let t_i be the observed failure time of subject i, $i = 1, ..., n$; assume that subjects are ordered such that $t_1 < t_2 < \ldots < t_n$; and let δ_i be the censoring (or failure) indicator such that $\delta_i = 1$ if subject i is an observed failure, and $\delta_i = 0$ if subject i is censored.

For each subject, $Y_i(t)$ is a survival indicator such that $Y_i(t) = 1$ if $t \leq t_i$; $Y_i(t) = 0$ otherwise. Let $X_{ik}(t)$ be the k^{th} covariate of subject i, where $i = 1, ..., n$ and $k = 1, ..., p$; $X_{ik}(t)$ could be either time-fixed or time-dependent. Then, $\vec{X}_i(t) = [X_{i1}(t), X_{i2}(t), \ldots, X_{ip}(t)]'$ is the covariate vector for the i^{th} subject at time t. Each subject in the population is then represented by $(Y_i(t), \vec{X}_i(t))$, for $0 \leq t \leq t_i$, $i = 1, ..., N$. Note that, as is done in the `coxph` function, the covariates in $X(t)$ are mean-centred in order to make the fitting algorithm more stable.

Let $w_i(t), i = 1, ..., n$, represent time-dependent weights such that the contribution to the partial likelihood of subject i at risk at its failure time t_i is weighted by $w_i(t_i)$ while $w_j(t_i)$ represents the contribution of subject j at the time of the subject i's failure. We assume that these weights are known and that the sum of the weights at any given time point is one, i.e. $\sum_{i=1}^{n} w_i(t) = 1$. If the data is distributed with hazard function $h(t|\vec{x}(t)) = h_0(t)\exp(\beta'\vec{x}(t))$, then the corresponding weighted partial likelihood function is,

$$L = \prod_{i=1}^{n} \left\{ \frac{\exp(\hat{\beta}'\vec{X}_i(t_i))}{\sum_{j=1}^{n} Y_j(t_i) w_j(t_i) \exp(\hat{\beta}'\vec{X}_j(t_i))} \right\}^{\delta_i w_i(t_i)}.$$

and weighted log-likelihood function is given by,

$$LL = \sum_{i=1}^{n} \delta_i w_i(t_i) \left\{ \hat{\beta}'\vec{X}_i(t_i) - \ln \left(\sum_{j=1}^{n} Y_j(t_i) w_j(t_i) \exp(\hat{\beta}'\vec{X}_k(t_i)) \right) \right\}. \quad (1)$$

The basic idea is that the above likelihood is an approximation to the unweighted likelihood that would have been obtained had the subjects been sampled without sampling bias.

1.2 Fitting a Cox model with time-dependent weights

The partial likelihood function is maximized by solving its associated estimating equation. However, because of the sampling bias inherent in the data, we use the following (weighted) plug-in estimator for the score, which is given by:

$$\tilde{\mathbf{U}}(\hat{\beta}) = \sum_{i=1}^{n} \delta_i w_i(t_i) \left\{ \vec{X}_i(t_i) - \frac{\tilde{S}^{(1)}(\hat{\beta}', t_i)}{\tilde{S}^{(0)}(\hat{\beta}', t_i)} \right\}, \quad (2)$$

where

$$\tilde{S}^{(0)}(\hat{\beta}', t) = \sum_{j=1}^{n} w_j(t) Y_j(t) \exp(\hat{\beta}'\vec{X}_j(t)), \quad (3)$$

and

$$\tilde{S}^{(1)}(\hat{\beta}', t) = \sum_{j=1}^{n} w_j(t) Y_j(t) \vec{X}_j(t) \exp(\hat{\beta}'\vec{X}_j(t)). \quad (4)$$

The negative partial derivative of $\tilde{\mathbf{U}}(\hat{\beta})$ with respect to $\hat{\beta}$ is:

$$\tilde{\mathbf{I}}(\hat{\beta}) = \sum_{i=1}^{n} \delta_i w_i(t_i) \left\{ \frac{\tilde{S}^{(2)}(\hat{\beta}', t_i)\tilde{S}^{(0)}(\hat{\beta}', t_i) - [\tilde{S}^{(1)}(\hat{\beta}', t_i)]^{\otimes 2}}{\left(\tilde{S}^{(0)}(\hat{\beta}', t_i)\right)^2} \right\}, \tag{5}$$

where

$$\tilde{S}^{(2)}(\hat{\beta}', t) = \sum_{j=1}^{n} w_j(t) Y_j(t) (\vec{X}_j(t))^{\otimes 2} \exp(\hat{\beta}' \vec{X}_j(t)). \tag{6}$$

and for any column vector $\vec{\alpha}$, $\vec{\alpha}^{\otimes 2} = \vec{\alpha}\vec{\alpha}'$. To obtain the finite population pseudo-maximum lilkelihood estimator $\hat{\beta}$, the Newton-Raphson algorithm is used to solve the non-linear system of equations $\tilde{\mathbf{U}}(\hat{\beta}) = \vec{0}$ with update at the t^{th} iteration given by:

$$\hat{\beta}^{(t)} = \hat{\beta}^{(t-1)} + \tilde{\mathbf{I}}^{-1(t-1)} \tilde{\mathbf{U}}(\hat{\beta}^{(t-1)}).$$

Further, the design-based variance estimator of $(\hat{\beta} - \beta)$ takes the form $\tilde{\mathbf{I}}^{-1}\mathbf{V}_U\tilde{\mathbf{I}}^{-1}$ where $\mathbf{V}_U$ is the variance of the $n \times 1$ vector $\tilde{U}(\hat{\beta})$. In practice, we estimate $\mathbf{V}_U$ by $\tilde{U}(\hat{\beta})'\tilde{U}(\hat{\beta})$ where we re-write $\tilde{U}(\hat{\beta})$ as follows,

$$\tilde{\mathbf{U}}(\hat{\beta}) = \sum_{i=1}^{n} w_i(t_i) u_i(\hat{\beta}), \tag{7}$$

where

$$\begin{aligned} u_i(\hat{\beta}) = \delta_i &\left\{ \vec{X}_i(t_i) - \frac{\tilde{S}^{(1)}(\hat{\beta}, t_i)}{\tilde{S}^{(0)}(\hat{\beta}, t_i)} \right\} - \\ &\frac{1}{w_i(t_i)} \sum_{j=1}^{n} \delta_j w_j(t_j) \left\{ \frac{w_i(t_j) Y_i(t_j) \vec{X}_i(t_j) \exp(\beta' \vec{X}_i(t_j))}{\tilde{S}^{(0)}(\hat{\beta}, t_j)} \right\} + \\ &\frac{1}{w_i(t_i)} \sum_{j=1}^{n} \delta_j w_j(t_j) \left\{ \frac{w_i(t_j) Y_i(t_j) \tilde{S}^{(1)}(\hat{\beta}, t_j) \exp(\beta' \vec{X}_i(t_j))}{\left(\tilde{S}^{(0)}(\hat{\beta}, t_j)\right)^2} \right\}. \end{aligned} \tag{8}$$

Finally, the current implementation of `coxtdw` in R uses the Breslow approximation (Breslow (1974)) for handling ties. Note that the Breslow method produces conservative bias if the data set has very heavy ties (Cox and Oakes (1984)), though the Efron method is also a popular method for tie handling. Further, we assume that the response used in the call to `coxtdw` is a `Surv` object with both `start` and `stop` times specified.

1.3 Parameter Estimation in `coxtdw` with Breslow Approximation

There is a term in the (partial) likelihood function for every event. When there are multiple subjects who have an event at the same time, i.e. event times are tied, the Breslow approximation does not assume that the exact time of any death is unique. Hence the contribution to the likelihood is simply the ratio of each subject's score to the sum of scores for all subjects at risk just before the event time (i.e. any subject for which $Y(t_i) = 1$ for event time t_i).

In `coxtdw`, the log-likelihood is computed as follows (comparing to Equation (1)):

$$LL = \sum_{i}^{n} W_i A_i,$$

where $W_i = \delta_i w_i(t_i)$, and $A_i = \left(B_i - \ln\left(\sum_j^n C_j(t_i)\right)\right)$ with

$$\begin{aligned} B_i &= \beta' \vec{X}_i(t_i), \text{and} \\ C_j(t) &= w_j(t) Y_j(t) \exp(\beta' \vec{X}_j(t)). \end{aligned}$$

Define

$$\bar{X}(t) = \frac{\sum_{j=1}^{n} X_j(t) C_j(t)}{\sum_{j=1}^{n} C_j(t)}.$$

It is easy to verify that $\bar{X}(t) = \tilde{S}^{(1)}(\hat{\beta}', t)/\tilde{S}^{(0)}(\hat{\beta}', t)$. Further, we can rewrite the score vector in Equation (2) and information matrix in Equation (5) as follows:

$$
\begin{aligned}
U &= \sum_{i=1}^{n} W_i\{X_i(t_i) - \bar{X}_i(t_i)\}, \text{and} \\
I &= \sum_{i=1}^{n} W_i \left\{ \frac{\sum_{j=1}^{n} (X_j(t_i))^{\bigotimes 2} C_j(t_i)}{\sum_{j=1}^{n} C_j(t)} - \left(\bar{X}(t_i)\right)^{\bigotimes 2} \right\}
\end{aligned}
$$

In fact, we can simplify the information matrix even further as follows:

$$
I_{kl} = \sum_{i=1}^{n} W_i \left\{ \frac{\sum_{j=1}^{n} X_{kj}(t_i) X_{lj}(t_i) C_j(t_i)}{\sum_{j=1}^{n} C_j(t)} - \bar{X}_k(t_i)\bar{X}_l(t_i) \right\},
$$

for $k, l = 1, \ldots, p$ where $X_{kj}(t_i)$ is subject j's value of the k^{th} covariate. Similarly, $\bar{X}_k(t_i)$ is the k^{th} component of $\bar{X}(t_i)$. In the special case that $k = l$ we get,

$$
I_{kk} = \sum_{i=1}^{n} W_i \left\{ \frac{\sum_{j=1}^{n} X_{kj}^2(t_i) C_j(t_i)}{\sum_{j=1}^{n} C_j(t_i)} - \bar{X}_k^2(t_i) \right\}.
$$

Finally, for variance estimation of the model coefficients, we can re-write Equation (7) as:

$$
\begin{aligned}
\tilde{U}(\hat{\beta}) &= \sum_{i=1}^{n} W_i \left\{X_i(t_i) - \bar{X}(t_i)\right\} - \sum_{i=1}^{n}\sum_{j=1}^{n} W_j \frac{X_i(t_j)C_i(t_j)}{\sum_{k=1}^{n} C_k(t_j)} + \sum_{i=1}^{n}\sum_{j=1}^{n} W_j \frac{\bar{X}(t_j)C_i(t_j)}{\sum_{k=1}^{n} C_k(t_j)} \\
&= \sum_{i=1}^{n} \left[W_i \left\{X_i(t_i) - \bar{X}(t_i)\right\} - \sum_{j=1}^{n} W_j \left\{X_i(t_i) - \bar{X}(t_i)\right\} \frac{C_i(t_j)}{\sum_{k=1}^{n} C_k(t_j)} \right]
\end{aligned}
$$

Let $\tilde{U}$ be a $n \times 1$ vector containing the i^{th} contribution to $\tilde{U}(\hat{\beta})$, for $i = 1, \ldots, n$. Then we have,

$$
\tilde{U}_i = W_i \left\{X_i(t_i) - \bar{X}(t_i)\right\} - \sum_{j=1}^{n} W_j \left\{X_i(t_j) - \bar{X}(t_j)\right\} \frac{C_i(t_j)}{\sum_{k=1}^{n} C_k(t_j)}.
$$

The final sandwich estimator for producing robust variance estimates is given by $V = (I^{-1} V_{\tilde{U}} I^{-1}) = (I^{-1}\tilde{U}')(\tilde{U} I^{-1})$. Note that this variance estimator can be shown to be an approximation of the jackknife estimate of variance, and

Table 1: Test Data Mini

id	(start, stop]	event	x	B_i	C_j	$\bar{X}$
1	(1,2]	0	0			
1	(2,3]	1	0	0	1	$\frac{2e^\beta}{1+2e^\beta}$
2	(1,2]	0	0			
2	(2,3]	1	1	β	e^β	$\frac{2e^\beta}{1+2e^\beta}$
3	(1,2]	0	1			
3	(2,3]	0	1		e^β	
				$\sum C_j =$	$1 + 2e^\beta$	

is a natural extension of the analogous quantity in the case weight setting. Further, Therneau and Grambsch (2001) (pg. 159) explain that the actual jackknife estimate tends to overestimate the variance for small n, whereas $D'D$ performs quite well. In the case weight or unweighted settings, our variance estimator reduces to the same robust variance estimators given for `coxph` in R.

Using the equations detailed in this section, in the examples that follow we will need to compute W_i, B_i, $C_j(t_i)$, $\sum_{j=1}^{n} C_j(t_i)$ as well as $\bar{X}(t_i)$.

2 Test-mini

The first data set we use to test our program contains only 3 subjects observed over 2 time intervals, comprising 6 observations. This data set does not contain weights, and the covariate x is binary; there is one tied death time. The data is provided in Table 1.

For the data in '`test-mini`', we then have 2 terms in the log-likelihood, and the corrseponding the score and negative-information are scalars. Since there are no weights, and there is only one covariate and it is binary, we have:

$$
\begin{aligned}
U &= \sum_{i=1}^{n} \delta_i \left\{ X_i(t_i) - \bar{X}(t_i) \right\} \\
I &= \sum_{i=1}^{n} \delta_i \left\{ \bar{X}(t_i) - \bar{X}^2(t_i) \right\}
\end{aligned}
$$

Hence the log-likelihood, score and information for `test-mini` are as follows.

$$\begin{aligned}
\bar{X}(3) &= \frac{2e^{\beta}}{1+2e^{\beta}} \\
LL(\beta) &= \beta - 2\ln(1+2e^{\beta}) \\
U &= \{0-\bar{X}(3)\} + \{1-\bar{X}(3)\} = 1-2\bar{X}(3) \\
I &= 2\left\{\bar{X}(3) - \bar{X}^2(3)\right\}
\end{aligned}$$

Since $U = 0$ at the maximum, we find $1 - 2e^{\beta} = 0$ which implies $\hat{\beta} = \ln 2 = -0.6931472$. So the values of LL, U and I at the intial estimate of $\beta = 0$ and at the MLE $\hat{\beta} = -\ln 2$ are:

Table 2: Update statistics for Test Mini Data

	$\beta = 0$	$\beta = -0.75$	$\hat{\beta} = -\ln(2)$
$\bar{X}(3)$	$\frac{2}{3} = 0.66667$	0.48579	$\frac{1}{2} = 0.5$
LL	$-2\ln 2 - \ln 4 = -2.19773$	-2.08025	$-3\ln 2 = -2.07944$
U	$1 - \frac{4}{3} = -0.33333$	0.02842	0
I	$2\{\frac{2}{3} - \frac{4}{9}\} = \frac{4}{9} = 0.44444$	0.49960	$2\{\frac{1}{2} - \frac{1}{4}\} = \frac{1}{2} = 0.5$

We now turn focus to robust variance estimation. Because there is only one model coefficient to be estimated, the variance of $\tilde{U}$ is given by $V_{\tilde{U}} = \sum_{i=1}^{n} \tilde{U}_i^2$. Further, since only subjects 1 and 2 failed, and both at time 3, we have that,

$$\begin{aligned}
\tilde{U}_i &= \delta_i\left\{X_i(3) - \bar{X}(3)\right\} - \sum_{j=1}^{2}\left\{X_i(3) - \bar{X}(3)\right\}\frac{C_i(3)}{\sum_{k=1}^{3} C_k(3)} \\
&= \delta_i\left\{X_i(3) - \frac{1}{2}\right\} - 2\left\{X_i(3) - \frac{1}{2}\right\}\frac{C_i(3)}{2} \\
&= \{\delta_i - C_i(3)\}\left\{X_i(3) - \frac{1}{2}\right\}.
\end{aligned}$$

Recall that at the MLE of β we have $\bar{X}(3) = \frac{1}{2}$. Hence, the variance based on the score residuals has terms given in Table 3. Consequently, $V_{\tilde{U}} = 0 + \frac{1}{16} + \frac{1}{16} = \frac{1}{8}$. The standard error of $\hat{\beta}$ is given by $\sqrt{I^{-1}V_{\tilde{U}}I^{-1}} = \sqrt{\frac{4}{8}} = \frac{1}{\sqrt{2}} = 0.7071$.

Table 3: Score residuals for Test Mini Data

id	$\tilde{U}_i$
1	$(1-1)(0-\frac{1}{2}) = 0$
2	$(1-\frac{1}{2})(1-\frac{1}{2}) = \frac{1}{4}$
3	$-\frac{1}{2}(1-\frac{1}{2}) = -\frac{1}{4}$

Table 4: Test-mini3

id	(start, stop]	event	x	B_i	C_j	$\sum C_j$	$\bar{X}$
1	(1,2]	0	1				
1	(2,3]	1	1	β	e^β	$2+2e^\beta$	$\frac{2e^\beta}{2+2e^\beta} = \frac{e^\beta}{1+e^\beta}$
2	(1,2]	0	0				
2	(2,3]	0	1		e^β		
3	(1,2]	0	0				
3	(2,3]	0	0		1		
3	(3,4]	0	0				
4	(2,3]	0	0		1		
4	(3,4]	0	1				
4	(4,5]	1	1	β	e^β	$1+e^\beta$	$\frac{e^\beta}{1+e^\beta}$
5	(4,5]	1	0	0	1	$1+e^\beta$	$\frac{e^\beta}{1+e^\beta}$

3 Test-mini3

The second data set contains 5 subjects observed over 1 to 4 time intervals, comprising 11 observations. Again, covariate x is binary and there is one tied death time. The data is provided in Table 4.

The general form of the score and information take the same form as for `test-mini`. We now have 3 terms in the log-likelihood, and the corrseponding the score and negative-information are as follows:

$$
\begin{aligned}
\bar{X} &= \bar{X}(3) = \bar{X}(5) = \frac{e^{\beta}}{1+e^{\beta}} \\
LL &= \left\{\beta - \ln(2+2e^{\beta})\right\} + \left\{\beta - \ln(1+e^{\beta})\right\} + \left\{0 - \ln(1+e^{\beta})\right\} \\
&= 2\left\{\beta - \ln(1+e^{\beta})\right\} - \ln(2+2e^{\beta}) \\
U &= \left[2 - 3\bar{X}\right] \\
I &= 3\left\{\bar{X} - \bar{X}^2\right\}
\end{aligned}
$$

Setting $U = 0$ and solving for β we find that $\hat{\beta} = \ln 2 = 0.6931472$. So the values of LL, U and I at the intial estimate of $\beta = 0$ and at the MLE $\hat{\beta} = \ln 2$ are as shown in Table 5.

Table 5: Update statistics for Test Mini Data

β:	0	0.66667	0.69304	$\hat{\beta} = \ln 2$
$\bar{X}$	$\frac{1}{2} = 0.5$	0.66076	0.66664	0.66667
LL	$-4\ln 2 = -2.77259$	-2.6029	-2.60269	-2.60269
U	$2 - \frac{3}{2} = 0.5$	0.01773	0.00008	0
I	$3(0.5 - 0.25) = 0.75$	0.67247	0.66597	0.66667
UI^{-1}	0.66667	0.02637	0.00012	

We now compute the score residuals for variance estimation at the MLE $\hat{\beta} = \ln 2$, exploiting that $\bar{X}(3) = \bar{X}(5) = \frac{2}{3}$.

$$
\tilde{U}_i = \delta_i \left\{X_i(t_i) - \frac{2}{3}\right\} - \sum_{j=1}^{5} \delta_j \left\{X_i(t_j) - \frac{2}{3}\right\} \frac{C_i(t_j)}{\sum_{k=1}^{5} C_k(t_j)}
$$

Table 6 gives the residual values for each subject. Hence,

$$
V_{\tilde{U}} = \frac{1}{81} + \frac{4}{81} + \frac{1}{81} + \frac{0}{81} + \frac{4}{81} = \frac{10}{81}.
$$

Finally, the robust standard error is given by $\sqrt{I^{-1}V_{\tilde{U}}I^{-1}} = \sqrt{\frac{9}{4}\left(\frac{10}{81}\right)} = \sqrt{\frac{5}{18}} = 0.52705$.

Table 6: Score residuals for Test Mini Data

id	$\tilde{U}_i$
1	$\left\{1-\frac{2}{3}\right\}-\left\{1-\frac{2}{3}\right\}\frac{1}{3}=\frac{1}{9}$
2	$0-\left\{1-\frac{2}{3}\right\}\frac{2}{3}=-\frac{2}{9}$
3	$0-\left\{0-\frac{2}{3}\right\}\frac{1}{6}=\frac{1}{9}$
4	$\left\{1-\frac{2}{3}\right\}-\left\{0-\bar{X}\right\}\frac{1}{6}-2\left\{1-\frac{2}{3}\right\}\frac{2}{3}=0$
5	$\left\{0-\frac{2}{3}\right\}-2\left\{0-\frac{2}{3}\right\}\frac{1}{3}=-\frac{2}{9}$

4 Test3cw

The third data set contains 8 subjects observed over 1 to 6 time intervals, comprising 19 observations. Unlike the previous examples, this one has 2 covariates, x_1 is a binary baseline variable, while x_2 is binary but time-dependent. Further, this exmaple has case weights. There are four failures in test3cw but no ties. See Table 7 for the full data. Note that, as with previous examples, the subjects are ordered based on their respective failure times.

Let $d_1 = 11r_1r_2 + 7r_1 + 9$, $d_3 = 6r_1r_2 + 4r_1 + 9$, $d_5 = 4r_1r_2 + 2r_2 + 5$ and $d_6 = 4r_1r_2 + 5$. The corresponding log-likelihood, score vector and negative-information matrix are as follows:

$$
\begin{aligned}
LL &= 3(\beta_1 - \ln d_1) + 6(\beta_1 + \beta_2 - \ln d_3) + 2(\beta_2 - \ln d_5) + 4(\beta_1 + \beta_2 - \ln d_6) \\
&= 13\beta_1 + 12\beta_2 - 3\ln d_1 - 6\ln d_3 - 2\ln d_5 - 4\ln d_6
\end{aligned}
$$

$$
\begin{aligned}
U_1 &= 3\left(1 - \bar{X}_1(1)\right) + 6\left(1 - \bar{X}_1(3)\right) + 2\left(0 - \bar{X}_1(5)\right) + 4\left(1 - \bar{X}_1(6)\right) \\
U_2 &= 3\left(0 - \bar{X}_2(1)\right) + 6\left(1 - \bar{X}_2(3)\right) + 2\left(1 - \bar{X}_2(5)\right) + 4\left(1 - \bar{X}_2(6)\right)
\end{aligned}
$$

$$
\begin{aligned}
I_{11} &= 3\left(\bar{X}_1(1) - \bar{X}_1^2(1)\right) + 6\left(\bar{X}_1(2) - \bar{X}_1^2(2)\right) + 2\left(\bar{X}_1(4) - \bar{X}_1^2(4)\right) + 4\left(\bar{X}_1(5) - \bar{X}_1^2(5)\right) \\
I_{22} &= 3\left(\bar{X}_2(1) - \bar{X}_2^2(1)\right) + 6\left(\bar{X}_2(2) - \bar{X}_2^2(2)\right) + 2\left(\bar{X}_2(4) - \bar{X}_2^2(4)\right) + 4\left(\bar{X}_2(5) - \bar{X}_2^2(5)\right) \\
I_{12} &= 3\bar{X}_2(1)\left(1 - \bar{X}_1(1)\right) + 6\bar{X}_2(2)\left(1 - \bar{X}_1(2)\right) + 2\bar{X}_1(4)\left(1 - \bar{X}_2(4)\right) + 4\bar{X}_1(5)\left(1 - \bar{X}_2(5)\right)
\end{aligned}
$$

Setting $U = 0$ and solving for β we find that $\hat{\beta}_1 = 0.9289$ and $\hat{\beta}_2 = 1.2260$. Before computing LL, U and I we compute the preliminary calculations in

Table 7: Test3-cw

id	(start, stop]	event	x_1	x_2	cw	B_i	C_j	$\sum C_j$	$\bar{X}_1$	$\bar{X}_2$
1	(0,1]	1	1	0	3	β_1	$3r_1$	d_1	$\frac{d_1-9}{d_1}$	$\frac{11r_1r_2}{d_1}$
2	(0,1]	0	1	1	5		$5r_1r_2$			
3	(0,1]	0	1	1	6		$6r_1r_2$			
3	(1,2]	1	1	1	6	$\beta_1+\beta_2$	$6r_1r_2$	d_3	$\frac{d_3-9}{d_3}$	$\frac{6r_1r_2}{d_3}$
4	(0,2]	0	0	0	2		2			
4	(2,3]	0	0	1	2					
5	(0,2]	0	0	0	2		2			
5	(2,3]	0	0	1	2					
5	(3,4]	1	0	1	2	β_2	$2r_2$	d_5	$\frac{4r_1r_2}{d_5}$	$\frac{d_5-5}{d_5}$
6	(0,1]	0	1	0	4		$4r_1$			
6	(1,2]	0	1	0	4		$4r_1$			
6	(2,4]	0	1	1	4		$4r_1r_2$			
6	(4,5]	1	1	1	4	$\beta_1+\beta_2$	$4r_1r_2$	d_6	$\frac{4r_1r_2}{d_6}$	$\frac{4r_1r_2}{d_6}$
7	(0,1]	0	0	0	2		2			
7	(1,3]	0	0	0	2		2			
7	(3,4]	0	0	0	2		2			
7	(4,5]	0	0	0	2		2			
8	(0,3]	0	0	0	3		3			
8	(3,6]	0	0	0	3		3			

Table 8 as shown in Table 9. Since there are 4 failures, there are only 4 times in which we are interested.

Table 8: Preliminary calculations for `test3cw` data

	i	t_i	W_i	d_i	$\bar{X}_1$	$\bar{X}_2$
$\beta = 0$	1	1	3	27	18/27	11/27
	3	2	6	19	10/19	6/19
	5	4	2	11	4/11	6/11
	6	5	4	9	4/9	4/9
$\beta = \hat{\beta}$	1	1	3	121.61936	0.925999	0.780281
	3	2	6	70.88906	0.873041	0.730186
	5	4	2	46.32325	0.744941	0.892063
	6	5	4	39.50811	0.873443	0.873443

Table 9: Update statistics for `test3cw` data

	$\beta = 0$	$\hat{\beta}$
LL	-41.13883	-35.558997
U_1	5.33705	0.0
U_2	6.01435	0.0
I_{11}	3.61298	1.6927
I_{22}	3.50420	2.3311
I_{12}	2.62315	1.3324

Table 10 gives the residual values for each subject. Finally, we can compute the variance-covariance matrix for the betas using

$$V = I^{-1}\tilde{U}'\tilde{U}I^{-1} = \begin{pmatrix} 5.590205 & -5.178767 \\ -5.178767 & 5.302033 \end{pmatrix}$$

which gives the standard errors of $\hat{\beta}_1$ and $\hat{\beta}_2$ as 2.3643 and 2.3026 respectively. All of the test data sets considered thus far are ones that `coxph` is capable of handling, and the results herein can be verified by the `coxph` function. For example, we can fit the data `test3cw` in both functions as follows:

Table 10: Score residuals for `test3cw` data

id	$\tilde{U}_1$	$\tilde{U}_2$
1	0.208139	-2.292100
2	-0.078739	0.233790
3	0.111046	0.156260
4	0.193471	0.162099
5	-1.077722	0.346210
6	-0.443261	0.724030
7	0.434660	0.415990
8	0.651990	0.623990

```
> cph <- coxph ( Surv(start,stopp,event)   x1 + x2 + cluster(id),
test3cw, weights=cw, robust=T, method=‘‘breslow"))

> ali <- coxtdw (Surv(start,stopp,event)   x1 + x2 + cluster(id),
test3cw, weights=cw, robust=T, method=‘‘breslow", init=c(0,0) ))
```

The quantities evaluated in this section can be compared/mapped to the `coxph` output as follows:

$$
\begin{aligned}
\bar{X} &= \texttt{coxph.detail(cph)\$means} \\
LL &= \texttt{cph\$loglik} \\
U &= \texttt{sum(coxph.detail(cph)\$score)} \\
I &= \texttt{cph\$var} \\
\hat{\beta}_1, \hat{\beta}_2 &= \texttt{summary(cph)\$coeff} \\
\tilde{U} &= \texttt{residuals(cph, collapse = test3cw\$id, weighted = T, type = "score")} \\
D = I^{-1}\tilde{U} &= \texttt{residuals(cph, collapse = test3cw\$id, weighted = T, type = "dfbeta")}
\end{aligned}
$$

5 Test3tdw

The last data set is essentially the same as the previous, except that the weights are now time-dependent. However, there are now 23 instead of only

Table 11: Test3-tdw

id	(start, stop]	event	x_1	x_2	cw	B_i	C_j	$\sum C_j$	$\bar{X}_1$	$\bar{X}_2$
1	(0,1]	1	1	0	3	β_1	$3r_1$	d_1	$\frac{d_1-11}{d_1}$	$\frac{11r_1r_2}{d_1}$
2	(0,1]	0	1	1	5		$5r_1r_2$			
3	(0,1]	0	1	1	6		$6r_1r_2$			
3	(1,2]	1	1	1	8	$\beta_1+\beta_2$	$8r_1r_2$	d_3	$\frac{d_3-12}{d_3}$	$\frac{8r_1r_2}{d_3}$
4	(0,1]	0	0	0	2		2			
4	(1,2]	0	0	1	2		2			
4	(2,3]	0	0	1	4					
5	(0,1]	0	0	0	2		2			
5	(1,2]	0	0	0	2		2			
5	(2,3]	0	0	1	2					
5	(3,4]	1	0	1	4	β_2	$4r_2$	d_5	$\frac{3r_1}{d_5}$	$\frac{12r_2}{d_5}$
6	(0,1]	0	0	0	4		4			
6	(1,2]	0	0	0	5		5			
6	(2,4]	0	0	1	8		$8r_2$			
6	(4,5]	1	0	1	8	β_2	$8r_2$	d_6	$\frac{4r_1}{d_6}$	$\frac{8r_2}{d_6}$
7	(0,1]	0	1	0	2		$2r_1$			
7	(1,3]	0	1	0	2		$2r_1$			
7	(3,4]	0	1	0	3		$3r_1$			
7	(4,5]	0	1	0	4		$4r_1$			
8	(0,1]	0	0	0	3		3			
8	(1,3]	0	0	0	3		3			
8	(3,4]	0	0	0	6		6			
8	(4,6]	0	0	0	6		6			

19 observations because some of the subjects' profiles have been expanded, in some cases to reflect times at which the time-dependent weights change. The data are given below in Table 11.

Let $d_1 = 11r_1r_2 + 5r_1 + 11$, $d_3 = 8r_1r_2 + 2r_1 + 12$, $d_5 = 3r_1 + 12r_2 + 6$ and $d_6 = 4r_1 + 8r_2 + 6$. The corresponding log-likelihood, score vector and negative-information matrix are as follows:

$$\begin{aligned} LL &= 3(\beta_1 - \ln d_1) + 8(\beta_1 + \beta_2 - \ln d_3) + 4(\beta_2 - \ln d_5) + 8(\beta_2 - \ln d_6) \\ &= 11\beta_1 + 20\beta_2 - 3\ln d_1 - 8\ln d_3 - 4\ln d_5 - 8\ln d_6 \end{aligned}$$

$$\begin{aligned} U_1 &= 3\left(1 - \bar{X}_1(1)\right) + 8\left(1 - \bar{X}_1(2)\right) + 4\left(0 - \bar{X}_1(4)\right) + 8\left(0 - \bar{X}_1(5)\right) \\ U_2 &= 3\left(0 - \bar{X}_2(1)\right) + 8\left(1 - \bar{X}_2(2)\right) + 4\left(1 - \bar{X}_2(4)\right) + 8\left(1 - \bar{X}_2(5)\right) \end{aligned}$$

$$\begin{aligned} I_{11} &= 3\left(\bar{X}_1(1) - \bar{X}_1^2(1)\right) + 8\left(\bar{X}_1(2) - \bar{X}_1^2(2)\right) + 4\left(\bar{X}_1(4) - \bar{X}_1^2(4)\right) + 8\left(\bar{X}_1(5) - \bar{X}_1^2(5)\right) \\ I_{22} &= 3\left(\bar{X}_2(1) - \bar{X}_2^2(1)\right) + 8\left(\bar{X}_2(2) - \bar{X}_2^2(2)\right) + 4\left(\bar{X}_2(4) - \bar{X}_2^2(4)\right) + 8\left(\bar{X}_2(5) - \bar{X}_2^2(5)\right) \\ I_{12} &= 3\bar{X}_2(1)\left(1 - \bar{X}_1(1)\right) + 8\bar{X}_2(2)\left(1 - \bar{X}_1(2)\right) - 4\bar{X}_1(4)\bar{X}_2(4) - 8\bar{X}_1(5)\bar{X}_2(5) \end{aligned}$$

Table 12: Preliminary calculations for `test3tdw` data

	i	t_i	W_i	d_i	$\bar{X}_1$	$\bar{X}_2$
$\beta = 0$	1	1	3	27	2/3	11/27
	3	2	8	22	10/22	20/22
	5	4	4	21	1/7	12/21
	6	5	8	18	4/18	8/18
$\beta = \hat{\beta}$	1	1	3	180.56757	0.939081	0.890089
	3	2	8	132.42668	0.909384	0.882663
	5	4	4	110.40566	0.048076	0.897579
	6	5	8	79.14234	0.089423	0.834764

Setting $U = 0$ and solving for β we find that $\hat{\beta}_1 = 0.5705749$ and $\hat{\beta}_2 = 2.1112007$. Before computing LL, U and I we compute the preliminary calculations in Table 12 as shown in Table 13. Since there are 4 failures, there are only 4 times in which we are interested.

Table 13: Update statistics for `test3tdw` data

	$\beta = 0$	$\hat{\beta}$
LL	-69.91691	-59.96284
U_1	3.01443	0.0
U_2	9.30014	0.0
I_{11}	4.52260	1.66533
I_{22}	5.66265	2.59323
I_{12}	1.27422	0.03275

Table 14 gives the residual values for each subject. Finally, we can compute

Table 14: Score residuals for `test3tdw` data

id	$\tilde{U}_1$	$\tilde{U}_2$
1	0.177385	-2.591770
2	-0.073941	-0.133406
3	-0.003668	-0.109349
4	0.141077	0.136221
5	-0.051226	0.423330
6	0.232819	0.299038
7	0.200866	1.010800
8	0.276302	0.905734

the variance-covariance matrix for the betas using

$$V = I^{-1}\tilde{U}'\tilde{U}I^{-1} = \begin{pmatrix} 1.66533 & 0.03275 \\ 0.03275 & 2.59323 \end{pmatrix}$$

which gives the standard errors of $\hat{\beta}_1$ and $\hat{\beta}_2$ as 0.082976 and 1.320177 respectively.

As expected, the MLE matches that of `coxph`, but the standard errors do not. Hence, this document also proves that the R code for computing the standard errors in `coxph` defintiely do not accommodate time-dependent weights.

However, how far off the standard errors are, in general, would depend on the nature of the weighting. A simulation study should be performed to shed further light on this particular topic.

References

Binder, D. A. (1992). Fitting Cox's proportional hazards models from survey data. *Biometrika 79*(1), 139–147.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics 30*(1), 89–99.

Cox, D. and D. Oakes (1984). *Analysis of Survival Data*. London: Chapman & Hall.

Therneau, T. M. and P. M. Grambsch (2001). *Modeling Survival Data: Extending the Cox Model*. New York, NY: Springer.