Supporting information for "Semiparametric modeling of grouped current duration data with preferential reporting"

Alexander C. McLain[1]

Department of Epidemiology and Biostatistics,

University of South Carolina

Rajeshwari Sundaram, Marie Thoma, and Germaine M. Buck Louis

Division of Intramural Population Health Research,

Eunice Kennedy Shriver National Institute of Child Health and Human Development

A Nonparametric inference for discrete backwards recurrence times

Jankowski and Wellner (2009) investigated nonparametric methods of estimating a discrete monotone non-increasing probability mass function. In this section we discuss their methods, and how they can be used to estimate $\bar{F}$ based on discrete current duration data (see Jankowski and Wellner, 2009, for the code to implement this method).

Let $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ denote the observed discrete current durations. The empirical estimator of g is $\hat{g}_E(y) = n^{-1}\sum_{i=1}^{n} I(Y_i = y)$, with empirical distribution function $\hat{G}_E(y) = \sum_{j=0}^{y} g_E(j)$. To estimate g restricted maximum likelihood estimation is used. The restricted likelihood is $L(g) = \prod_{i=1}^{n} g(Y_i)$ such that $g \in \mathcal{P}_0$, where $\mathcal{P}_0$ is the space of all probability mass

[1]Corresponding to: Department of Epidemiology and Biostatistics, University of South Carolina, 915 Greene Street, Columbia, SC 29208, USA. E-mail:mclaina@mailbox.sc.edu.

functions such that $g(j) \geq g(k)$ for all $j < k$. The g that maximizes $L(g)$ such that $g \in \mathcal{P}_0$, denoted by $\hat{g}_{NP}$, is known as the Grenander estimator, and is defined as the left derivative at y of the least concave majorant (LCM) of the empirical distribution function $\hat{G}_E(y)$. One can find $\hat{g}_N$ by plotting $\hat{G}_E$, and identifying the M points that outline the LCM of $\hat{G}_E$ denoted by $\{y_j^L, \hat{G}_E(y_j^L)\}$ for $j = 0, 1, 2, \ldots, M$, where $\{y_0^L, \hat{G}_E(y_0^L)\} \equiv \{-1, 0\}$ is included to find the left derivative at y = 0. For $y_j^L < y \leq y_{j+1}^L$ we have

$$\hat{g}_{NP}(y) = \frac{G_E(y_{j+1}^L) - G_E(y_j^L)}{y_{j+1}^L - y_j^L},$$

and $\hat{g}_{NP}(y) = 0$ for all $y > \max_{i=1,\ldots,n}(Y_i)$. The LCM of $\hat{G}_E$ can be found using the `chull` function in `R` (R Core Team, 2013). The survivor function is then estimated as $\hat{\bar{\bar{F}}}_{NP}(y) = \hat{g}_{NP}(y)/\hat{g}_{NP}(0)$. Type I censoring, does not change $\hat{g}_{NP}(y)$ for all $y < \tau$. As a result, it is not necessary to develop a separate estimator to account for censoring.

The Grenander estimator $\hat{g}_{NP}$ has appealing asymptotic properties for discrete Y, such as uniform consistency and weak convergence to a Gaussian process (see Theorem 2.4 and 3.8 of Jankowski and Wellner, 2009). The asymptotic uniform consistency of $\hat{g}$ is a particularly attractive feature. It doesn't hold for the Grenander estimator when Y is continuous, and adaptive methods are required. It appears that the continuous mapping theorem can be applied to state that uniform consistency and weak convergence of $\hat{g}_{NP}$ will hold for $\hat{\bar{\bar{F}}}_{NP}(y)$ for discrete T, though further work is needed to formally investigate the these properties.

B Expanded simulation studies

In this section we present expanded simulation studies of the proposed methods. In Section B.1, we present results from simulation studies that used an asymmetric rounding mechanism where 'rounding down' is more prevalent then 'rounding up'. In Section B.2, we compare the proposed method to a Weibull AFT model. For both simulations, the current duration for the ith subject was simulated by generating the unobserved total durations as $T_{ij} \sim F$ for $j = 1, 2, \ldots, K$, where $K = \min(k; \sum_{j=1}^{k} T_{ij} > M)$ and M is a fixed large integer, replicating a renewal process in equilibrium with renewal distribution (see Feller, 1966, for details). For the continuous F scenario the backward recurrence times were grouped with $Y_i = \lfloor M - T_{iK-1} \rfloor$. For the discrete $\bar{F}$ scenario the continuous T_{ij} were grouped with $T^*_{ij} = \lceil T_{ij} \rceil$ and $Y_i = M - T^*_{iK^*-1}$ where $K^* = \min(k; \sum_{j=1}^{k} T^*_{ij} > M)$. Here, F had hazard function $\lambda(t|\boldsymbol{z}_i) = \theta\gamma t^{\gamma-1} \exp(\boldsymbol{\beta}^\top \boldsymbol{z}_i)$ with $\theta = 0.3$ and $\alpha = 0.75$, and $\boldsymbol{z}_i = (z_{i1}, z_{i2})$ were independently generated as Bernoulli(0.5) and $\mathrm{N}(0, 0.5^2)$, respectively.

B.1 Properties Under Asymmetric Rounding

As mentioned in Section 2.3 of the text, the proposed piecewise constant model will account for digit preference when rounding in the data is at random and people are equally likely to round up as they are to round down (corresponding to the coarsening at random assumption, see Heitjan and Rubin, 1991; Gill et al., 1997; Gill and Robins, 1997). This can also be seen as symmetric rounding, and in this case the observations are essentially "unbiased" values. However, if people round down more often then they round up the values for Y are negatively biased which would result in negatively biased estimates of $\bar{F}$. We investigated the degree of bias via simulation studies. The setting of the simulation was similar to that in Section 4.1 of the main text. We induced asymmetric rounding by making rounding down more likely than rounding up. Specifically, if $4 \leq Y_i \leq 5$ then Y_i was rounded to 6 with probability 0.2, if $7 \leq Y_i \leq 9$ then Y_i was rounded to 6 with probability 0.4, if $10 \leq Y_i \leq 11$ then Y_i was rounded to 12 with probability 0.4, if $13 \leq Y_i \leq 18$ then Y_i was rounded to 12 with probability 0.6, if $Y_i > 18$ then Y_i was rounded up to the nearest multiple of 12 with probability 0.6, and rounded down to nearest multiple of 12 with probability 0.8. As a result, for $4 \leq Y_i \leq 9$ people were twice as likely to round down to 6 as they were to round up to 6.

The results of the asymmetric simulation are contained in Table 1. The bias in the estimates of the survivor function did not appear to be affected by the true value of $\boldsymbol{\beta}$. Further, the estimates of $\boldsymbol{\beta}$ were relatively unbiased for both settings. It is evident from the simulation results that there is negative bias in the estimates of the survivor function at the points of digit preference when 'rounding down' is more prevalent than 'round up.' The degree of bias is relatively small, however, the simulation results for the symmetric rounding showed a small

Table 1: Summary of 1,000 simulated samples with $n = 250$ for the piecewise constant models when the digit preference is subjected to asymmetric rounding. The piecwise model was fit with knot location $\{1, 2, 5, 8, 11, 18, Y_{(m)}\}$. Displayed is the empirical bias (BIAS), empirical standard deviation (SD), and $||\bar{F} - \tilde{\bar{F}}||_2$ (l_2).

	TRUE	BIAS (SD)	TRUE	BIAS (SD)
β_1	-0.5	0.020 (0.088)	0.0	-0.001 (0.083)
β_2	-0.5	-0.025 (0.213)	0.0	-0.000 (0.207)
$\bar{F}(6)$	0.317	-0.016 (0.049)	0.317	-0.010 (0.052)
$\bar{F}(12)$	0.145	-0.020 (0.022)	0.145	-0.025 (0.021)
$\bar{F}(24)$	0.039	-0.006 (0.014)	0.039	-0.002 (0.003)
l_2		0.194		0.150

amount of positive bias. As a result, these results could be optimistic.

B.2 Comparison to Weibull AFT model

In this section, we present an expanded simulation study of the proposed methods, along with a comparison to a Weibull accelerated failure time (AFT) model. In order to compare the results of the proposed proportional hazards models to the AFT model some transformations of the parameters is required. The AFT model and the proportional hazards model coincide under a Weibull AFT and a PH model with a Weibull baseline hazard (the setting used in Section 4 of the manuscript). That is, if we have the Weibull AFT

$$\log(T) = -(\mu + \boldsymbol{\eta}^\top \boldsymbol{Z}) + \epsilon/\gamma, \tag{1}$$

where ϵ has extreme value distribution, then the hazard function of T is

$$\lambda(t|\boldsymbol{Z}) = \theta\gamma t^{\gamma-1} \exp(\boldsymbol{\beta}^\top \boldsymbol{Z}),$$

which satisfies the proportional hazards assumption where $\log(\theta) = \mu\gamma$ and $\gamma\eta_j = \beta_j$. The AFT model gives estimates of μ, γ, and $\boldsymbol{\eta}$, thus the estimate of $\boldsymbol{\eta}$ can be compared to the estimate of $\boldsymbol{\beta}$ using the relationship $\gamma\eta_j = \beta_j$. Our proposed model estimates $\boldsymbol{\beta}$, so we could theoretically compare $\hat{\gamma}\hat{\eta}_j$ with $\hat{\beta}_j$. However, when we fit the AFT model to Y the estimate of

γ does not coincide with the γ in (1). To see this notice that under (1) the survivor function of T has the form

$$\bar{F}(t|\boldsymbol{Z}) = \exp\left[-\exp\left\{\gamma(\mu+\boldsymbol{\eta}^\top \boldsymbol{Z})\right\} t^\gamma\right], \tag{2}$$

and $E(T) = \exp\{\gamma(\mu+\boldsymbol{\eta}^\top \boldsymbol{Z})\}^{-1/\gamma}\Gamma(1+1/\gamma)$. Note that

$$g(t|\boldsymbol{Z}) = \frac{\bar{F}(t|\boldsymbol{Z})}{E(T)} = \frac{\exp\left[-\exp\left\{\gamma(\mu+\boldsymbol{\eta}^\top \boldsymbol{Z})\right\} t^\gamma\right]}{\exp\{\gamma(\mu+\boldsymbol{\eta}^\top \boldsymbol{Z})\}^{-1/\gamma}\Gamma(1+1/\gamma)}$$

will not be a Weibull density. Similarly, if Y is distributed according to a Weibull AFT it can be shown that $\bar{F}(y|\boldsymbol{Z}) = g(y|\boldsymbol{Z})/g(0|\boldsymbol{Z})$ is not a Weibull survivor function. The estimates of $\boldsymbol{\eta}$ based on T or Y are unbiased (as shown by Yamaguchi, 2003), but the estimates of μ and $\boldsymbol{\eta}$ will vary based on which outcome is used. Using the notation of Section 2 in the manuscript, $\mathcal{P}_0$ for the Weibull AFT corresponds to $\mu \in \mathbb{R}$, $\boldsymbol{\eta} \in \mathbb{R}^p$, and $\gamma \geq 1$. Most AFT approaches do not restrict the parameter space. Thus, we can have $\hat{\gamma} < 1$ and $\hat{g} \notin \mathcal{P}_0$.

There are two ways we can compare the estimates from the AFT and proportional hazard models. First, we can use the true value of γ and compare $\gamma\hat{\eta}_j$ with $\hat{\beta}_j$. This is an artificial comparison since the the true value of γ will be unknown in an actual analysis. Second, if the parameter vector $\boldsymbol{\beta} = \mathbf{0}$ then $\boldsymbol{\eta} = \mathbf{0}$ and the output of the models can be compared. In Table 2, we present simulation results comparing our proposed piecewise and semiparametric approaches to the AFT approach. For the AFT approach we present the summarized results for the actual $\hat{\eta}$ estimates (Standard AFT), and for the corrected $\gamma\hat{\eta}$ estimates (Corrected AFT). The settings for the simulations, including the values of the parameters, were identical for those in Section 4.2 of the manuscript. The Table contains the average estimated coefficient, empirical standard deviation, and the l_2 norm as given in Section 4 of the manuscript. Each setting for the AFT model had at least one iteration where $\hat{\gamma} < 1$ and $\hat{g} \notin \mathcal{P}_0$. In such iterations $\hat{F}(\cdot|\boldsymbol{Z})$ could not be estimated, so the l_2 norm could not be reported. We report the number of iterations (out of 1,000) that where $\hat{g} \notin \mathcal{P}_0$.

For the settings with $\boldsymbol{\beta} = \mathbf{0}$ the estimates between the proposed approaches and the standard AFT have relatively similar behavior. In the discrete scenario, the piecewise and semiparametric approaches have slightly less variability and coverage probabilities that are

Table 2: Summary of 1,000 simulated samples with $n = 250$, and 500 for the piecewise constant, semiparametric and Weibull AFT models under the discrete (true F is discrete) and continuous (true F is continuous) scenarios. For the 'Corrected AFT' model the estimates of the coefficients are multiplied by α to coincide with estimates from the proportional hazards model. Displayed is the average coefficient (MEAN), empirical standard deviation (SD), empirical coverage probabilities (ECP), and l_2 norm (l_2). For the AFT model we report the number of estimates where $\bar{F}$ could not be estimated, i.e., $\hat{g} \notin \mathcal{P}_0$.

		Discrete F							
		Piecewise		Semiparametric		Standard AFT		Corrected AFT	
	TRUE	MEAN(SD)	ECP	MEAN(SD)	ECP	MEAN(SD)	ECP	MEAN(SD)	ECP
				n=250					
β_1	-0.5	-0.505 (0.148)	0.964	-0.510 (0.162)	0.950	-0.685 (0.157)	0.739	-0.514 (0.118)	0.922
β_2	-0.5	-0.507 (0.354)	0.951	-0.509 (0.368)	0.945	-0.690 (0.376)	0.895	-0.518 (0.282)	0.935
l_2		0.068		0.053		32/1000 with $\hat{g} \notin \mathcal{P}_0$			
β_1	0.0	-0.005 (0.143)	0.948	-0.002 (0.145)	0.950	0.004 (0.151)	0.937	0.003 (0.113)	0.937
β_2	0.0	0.005 (0.349)	0.955	-0.003 (0.356)	0.947	0.006 (0.382)	0.935	0.005 (0.286)	0.935
l_2		0.074		0.053		25/1000 with $\hat{g} \notin \mathcal{P}_0$			
				n=500					
β_1	-0.5	-0.506 (0.105)	0.952	-0.507 (0.104)	0.956	-0.677 (0.106)	0.566	-0.508 (0.080)	0.934
β_2	-0.5	-0.506 (0.252)	0.949	-0.505 (0.245)	0.965	-0.678 (0.262)	0.884	-0.508 (0.196)	0.944
l_2		0.050		0.029		3/1000 with $\hat{g} \notin \mathcal{P}_0$			
β_1	0.0	0.002 (0.095)	0.948	0.002 (0.097)	0.950	0.001 (0.110)	0.926	0.001 (0.083)	0.926
β_2	0.0	0.007 (0.230)	0.965	-0.002 (0.241)	0.950	0.005 (0.256)	0.942	0.004 (0.192)	0.942
l_2		0.050		0.034		1/1000 with $\hat{g} \notin \mathcal{P}_0$			
		Continuous F							
		Piecewise		Semiparametric		Standard AFT		Corrected AFT	
	TRUE	MEAN(SD)	ECP	MEAN(SD)	ECP	MEAN(SD)	ECP	MEAN(SD)	ECP
				n=250					
β_1	-0.5	-0.517 (0.129)	0.965	-0.511 (0.167)	0.949	-0.661 (0.146)	0.768	-0.496 (0.109)	0.952
β_2	-0.5	-0.515 (0.318)	0.938	-0.510 (0.389)	0.952	-0.662 (0.356)	0.924	-0.496 (0.267)	0.947
l_2		0.050		0.056		61/1000 with $\hat{g} \notin \mathcal{P}_0$			
β_1	0.0	0.002 (0.124)	0.942	-0.003 (0.143)	0.953	0.004 (0.107)	0.943	0.006 (0.142)	0.943
β_2	0.0	0.005 (0.305)	0.951	0.008 (0.375)	0.944	0.005 (0.279)	0.929	0.007 (0.372)	0.929
l_2		0.064		0.049		67/1000 with $\hat{g} \notin \mathcal{P}_0$			
				n=500					
β_1	-0.5	-0.513 (0.094)	0.948	-0.509 (0.115)	0.944	-0.664 (0.103)	0.608	-0.498 (0.077)	0.947
β_2	-0.5	-0.511 (0.214)	0.948	-0.507 (0.259)	0.954	-0.679 (0.271)	0.859	-0.509 (0.203)	0.954
l_2		0.032		0.026		13/1000 with $\hat{g} \notin \mathcal{P}_0$			
β_1	0.0	0.004 (0.083)	0.946	-0.002 (0.100)	0.947	0.003 (0.105)	0.941	0.002 (0.079)	0.941
β_2	0.0	-0.014 (0.201)	0.941	-0.017 (0.251)	0.957	0.002 (0.268)	0.931	0.002 (0.201)	0.951
l_2		0.028		0.031		15/1000 with $\hat{g} \notin \mathcal{P}_0$			

closer to the nominal 0.95 level than the standard AFT estimates. For the settings with $\boldsymbol{\beta} = \mathbf{-0.5}$ the standard AFT estimates are larger in absolute value. Bias in these estimates is expected since the true $\boldsymbol{\eta} \neq \mathbf{-0.5}$. In the discrete scenario, the corrected AFT estimates have less bias, more variability, and coverage probabilities that are further from the nominal 0.95 level than the proposed approaches. In the continuous scenario, however, the corrected AFT has less bias and variability with similar coverage probabilities than the proposed approaches. A decrease in variability is expected for the AFT model since it is fully parametric.

C Semiparametric, and piecewise constant `R` code

In this section we give the `R` code (R Core Team, 2013) to run the semiparametric, and piecewise constant backward recurrent Cox models (Cox, 1972) given in Sections 2 and 3 of the main text. To obtain `R` code for a nonparametric estimator of $g(\cdot)$ see Jankowski and Wellner (2009). In Section D we give code to implement and display the results of the semiparametric, and piecewise constant backward recurrent Cox models from simulated data.

C.1 Log-likelihood for the semiparametric model

```
SP_BR_GC_like <- function(par,X,T,CEN,w=1){
## USAGE
 # SP_BR_GC_like(par,X,T,CEN,w) in conjunction with an
 # optimization program, such as optim or nls.
## ARGUMENTS
 # par: The values of alpha, and beta to evaluate the log-
 # likelihood.  The length of alpha should be equal to the
 # number of distinctly observed observations +1 when
 # the largest censored observation is equal to the
 # largest observation and number of distinctly observed
 # observations otherwise.
 # X: covariate matrix
```

```
 # T: vector current durations
 # CEN: vector of indicators that a subject was not
 # censored.
 # w: weight function (optional).
## OUTPUT
 # -sum(log(likelihood))
## DETAILS
 # This program evaluates the semiparametric backward
 # recurrent discrete Cox likelihood with censoring,
 # corresponding to the negative log of equation (3.1).

if(length(w)==1){w <- rep(1,length(T))}
if(max(T[CEN==0]) == max(T)){
T_vals<- c(sort(unique(T[CEN==1])),max(T)+1)
}
if(max(T[CEN==0]) < max(T)){
T_vals<- sort(unique(T[CEN==1]))
}

len_p <- length(T_vals)-1
T_seq <- 1:max(T_vals)

alpha <- exp(par[1:len_p])
beta  <- par[(len_p + 1):length(par)]
if(length(beta)==1){eta <- X*beta}
if(length(beta) >1){eta <- X%*%beta}

ful_alpha <- T_seq*0
ful_alpha[T_vals[T_vals>0]] <- alpha
```

```
gamma       <- cumsum(ful_alpha)
GAM_T       <- T
GAM_T[T> 0] <- gamma[T[T>0]]

i_const <- numeric(0)
i_surv  <- numeric(0)
for(i in 1:length(T)){
temp_vals<- c(1,exp(-gamma[-length(gamma)]*exp(eta[i])),
exp(-(gamma[length(gamma)-1]+c(1:1000)*alpha[length
(alpha)])*exp(eta[i])))
t_pmf    <- temp_vals/sum(temp_vals)
t_surv   <- 1 - sum(t_pmf[1:c(T[i]+1)])
i_surv   <- c(i_surv,t_surv)
i_const  <- c(i_const,sum(temp_vals))
}

like      <- (CEN)*exp(-GAM_T*exp(eta))/i_const +
(1-CEN)*i_surv
llike     <- w*log(like)
-sum(llike)
}
```

C.2 Log-likelihood for the piecewise constant model

```
SP_BR_Piece_like <- function(par,X,T,CEN,knots,w=1){
## USAGE
 # SP_BR_GC_like(par,X,T,CEN,w) in conjunction with an
```

```
 # optimization program, such as optim or nls.
## ARGUMENTS
 # par: The values c(gamma, beta) to evaluate the log-
 # likelihood, resulting from equation (2.7).  The length
 # of gamma should be equal to the number of knots +1.
 # X: covariate matrix
 # T: vector current durations
 # CEN: vector of indicators that a subject was not
 # censored.
 # knots: vector of knot locations.
 # w: weight function (optional).
## OUTPUT
 # -sum(log(likelihood))

## DETAILS
 # This program evaluates the backward recurrent Cox
 # likelihood with a piecewise constant specification
 # baseline grouped hazard and censoring.



if(length(w)==1){w <- rep(1,length(T))}
T_vals<- c(sort(unique(T)),max(T))
len_p <- length(knots)+1
T_seq <- 1:max(T_vals)

alpha <- exp(par[1:len_p])
ex_parm <- alpha[length(alpha)]
beta <- par[(len_p + 1):length(par)]
if(length(beta)==1){eta <- X*beta}
```

```
if(length(beta)>1){eta <- X%*%beta}

alph_vec <- numeric(0)
knots2 <- c(1,knots,max(T_vals))
for(k in 1:len_p){
alph_vec <- c(alph_vec,rep(alpha[k],(knots2[k+1]-
knots2[k])+1*I(k==1)))
}
ful_alpha <- alph_vec

gamma       <- cumsum(ful_alpha)
GAM_T       <- T
GAM_T[T> 0] <- gamma[T[T>0]]

i_const  <- numeric(0)
i_surv    <- numeric(0)
for(i in 1:length(T)){
temp_vals<- c(1,exp(-gamma*exp(eta[i])),exp(-
(gamma[length(gamma)]+c(1:1000)*ex_parm*exp(eta[i]))))
t_pmf    <- temp_vals/sum(temp_vals)
t_surv   <- 1 - sum(t_pmf[1:c(T[i]+1)])
i_surv   <- c(i_surv,t_surv)
i_const <- c(i_const,sum(temp_vals))
}

like       <- (CEN)*exp(-GAM_T*exp(eta))/i_const +
(1-CEN)*i_surv
like[!(like>0)] <- 0
llike      <- w*log(like)
```

```
-sum(llike)
}
```

C.3 Data Generation and Survival Estimation Programs

```
#Function for generating current duration data.
data_gen_cont_T <- function(eta,alpha,theta,Cen_V=1e100){
tau <- 11000
t   <- 10000
T   <- numeric(0)
for(i in 1:n){
S_t <- 0
S_vec <- 0
lambda_i <- theta*exp(eta[i])
t_len <- ceiling(1.1*tau/((1/lambda_i)^(1/alpha)*
gamma(1+1/alpha)))
while(S_t < tau){
r_u <- runif(t_len)
r_w <- ((-log(1-r_u)/lambda_i)^(1/alpha))
t_l  <- cumsum(r_w)
S_t <- S_t + max(t_l)
S_vec <- c(S_vec,t_l)
}
A_t <-floor(t - max(S_vec[S_vec<=t]))
T <- c(T,A_t)
}

CEN <- 1*I(T<=Cen_V)
```

```
T[T>Cen_V] <- Cen_V
return(list(T=T,CEN=CEN))
}



#Function for estimating the survival function
#of the total durations based on a fitted semi-
#parametric model.
est_surv_SP <- function(par_est,T,CEN){
if(max(T[CEN==0]) == max(T)){
T_vals<- c(sort(unique(T[CEN==1])),max(T)+1)
}
if(max(T[CEN==0]) < max(T)){
T_vals<- sort(unique(T[CEN==1]))
}
len_p   <- length(T_vals)-1
T_seq   <- 1:max(T_vals)
alpha.h <- exp(par_est[1:len_p])
beta.h <- par_est[(len_p + 1):length(par_est)]
ful_alpha <- T_seq*0
ful_alpha[T_vals[-1]] <- alpha.h
gamma    <- cumsum(ful_alpha)
temp_vals<- c(1,exp(-gamma[-length(gamma)]),exp(-
(gamma[length(gamma)-1]+c(1:1000)*alpha.h[
length(alpha.h)])))
f_est    <- temp_vals/sum(temp_vals)
surv_est <- f_est/f_est[1]
step_est_SP <- stepfun(1:max(T_vals),c(1,surv_est[1:
```

```
max(T_vals)]),right=FALSE)
return(step_est_SP)
}


#Function for estimating the survival function
#of the total durations based on a fitted piece-
#wise constant model.
est_surv_PC <- function(par_est,T,CEN,knots){
T_vals    <- sort(unique(T))
len_p     <- length(knots)+1
T_seq     <- 1:max(T_vals)
alpha.h   <- exp(par_est[1:len_p])
ex_parm.h <- exp(par_est[len_p+1])
beta.h    <- par_est[(len_p + 2):length(par_est)]
ex_parm   <- par_est[len_p+1]
alph_vec  <- numeric(0)
knots2 <- c(1,knots,max(T_vals))
for(k in 1:len_p){
alph_vec <- c(alph_vec,rep(alpha.h[k],(knots2[k+1]-
knots2[k])+1*I(k==1)))
}
ful_alpha <- alph_vec
gamma    <- cumsum(ful_alpha)
i_const  <- sum(c(1,exp(-gamma),exp(-(gamma[
length(gamma)]+c(1:1000)*ex_parm.h))))
f_est    <- c(1/i_const,exp(-gamma)/i_const)
surv_est <- f_est/f_est[1]
step_est_PC <- stepfun(c(0,T_seq,max(T)+1),
```

```
c(1,surv_est,0),right=FALSE)
return(step_est_PC)
}
```

D Example with Simulated Data

In this section we give an example of the semiparametric, and piecewise constant backward recurrent Cox models from simulated data. The data is simulated under the Scenario discussed in Section 2.2 of the main text. First load the log likelihoods from the previous section. Then load the data generation, and survival estimation functions in Section C.3. After that the following will estimate and display both models.

```
library(MASS)
int_vec <- 0:36
Cen_V   <- 36
alpha <- 0.75
theta <- 0.3
set.seed(97)
true_surv <- exp(-theta*(int_vec)^alpha)

knots <- c(1,2,4,9,15,27)
n   <- 250
X <- matrix(1,n,2)
X[,1] <- rbinom(n,1,0.5)
X[,2] <- rnorm(n)/5
beta <- c(-0.5,-0.5)
eta <- X%*%beta
alpha <- 0.75
```

```
theta <- 0.3

# Generating the data
dat <- data_gen_cont_T(eta,alpha,theta,Cen_V)
T <- dat$T
CEN <- dat$CEN

# Estimating the parameters of the Semiparametric model.
par <- c(rep(-2,length(unique(T[CEN==1]))),beta)
fit.res2<- optim(par,SP_BR_GC_like,X=X,T=T,CEN=CEN,
method="BFGS",hessian=TRUE)

# Estimating the survival function of the total durations,
# for the semiparametric model.
var_est <- diag(ginv(fit.res2$hessian))
beta_var<- var_est[c(length(par)-length(beta)+1,
length(par))]
par_est <- fit.res2$par
beta.h  <- par_est[c(length(par)-length(beta)+1,
length(par))]
step_est_SP <- est_surv_SP(par_est,T,CEN)
summat <- round(cbind(beta.h,sqrt(beta_var),beta.h-
1.96*sqrt(beta_var),beta.h+1.96*sqrt(beta_var)),4)
colnames(summat) <- c("EST","Std_Err","95% CI_L",
"95% CI_U")

# Estimating the parameters of the Piecewise model.
par <- c(seq(-2,-2.5,length.out=length(knots)+1),beta)
fit.res <- optim(par,SP_BR_Piece_like,X=X,T=T,CEN=CEN,
```

```
knots=knots,method="BFGS",hessian=TRUE)

# Estimating the survival function of the total durations
# for the piecewise model.
var_est <- diag(ginv(fit.res$hessian))
beta_var<- var_est[c(length(par)-length(beta)+1,
length(par))]
par_est <- fit.res$par
beta.h  <- par_est[c(length(par)-length(beta)+1,
length(par))]
step_est_PC <- est_surv_PC(par_est,T,CEN,knots)
summat <- rbind(summat,round(cbind(beta.h,sqrt(beta_var),
beta.h-1.96*sqrt(beta_var),beta.h+1.96*sqrt(beta_var)),4))
rownames(summat) <- c("SP_beta1","SP_beta2","PC_beta1",
"PC_beta2")

# Displaying the estimates of the survival functions and
# beta parameter estimates for the both models
plot(int_vec,step_est_SP(int_vec),type="l",ylim=c(0,1),
xlab="TIME",ylab="Survival Estimate",lwd=2,cex.lab=1.3)
lines(int_vec,c(true_surv),col=4,lwd=2)
lines(int_vec,step_est_PC(int_vec),col=3,lwd=2)
legend(25,0.8,legend=c("SP Estimate","PC Estimate",
"True"),col=c(1,3,4),lwd=2)
summat  #Estimates and 95% CI's for beta parameters
```

References

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

Feller, W. (1966), *An introduction to probability theory and its applications. Vol. II*, New York: John Wiley & Sons Inc.

Gill, R., Laan, M., and Robins, J. (1997), "Coarsening at Random: Characterizations, Conjectures, Counter-Examples," in *Proceedings of the First Seattle Symposium in Biostatistics*, eds. Lin, D. and Fleming, T., Springer US, vol. 123 of *Lecture Notes in Statistics*, pp. 255–294.

Gill, R. and Robins, J. (1997), "Sequential Models for Coarsening and Missingness," in *Proceedings of the First Seattle Symposium in Biostatistics*, eds. Lin, D. and Fleming, T., Springer US, vol. 123 of *Lecture Notes in Statistics*, pp. 295–305.

Heitjan, D. F. and Rubin, D. B. (1991), "Ignorability and Coarse Data," *The Annals of Statistics*, 19, 2244–2253.

Jankowski, H. K. and Wellner, J. A. (2009), "Estimation of a discrete monotone distribution," *Electron. J. Stat.*, 3, 1567–1605.

R Core Team (2013), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.

Yamaguchi, K. (2003), "Accelerated failure–time mover–stayer regression models for the analysis of last–episode data," *Sociological Methodology*, 33, 81–110.