ICPSR 2015 “Advanced Maximum Likelihood”: Survival Analysis

Day Five

August 7, 2015

Proportional Hazards

For two individuals A and B, their relative hazards will be:

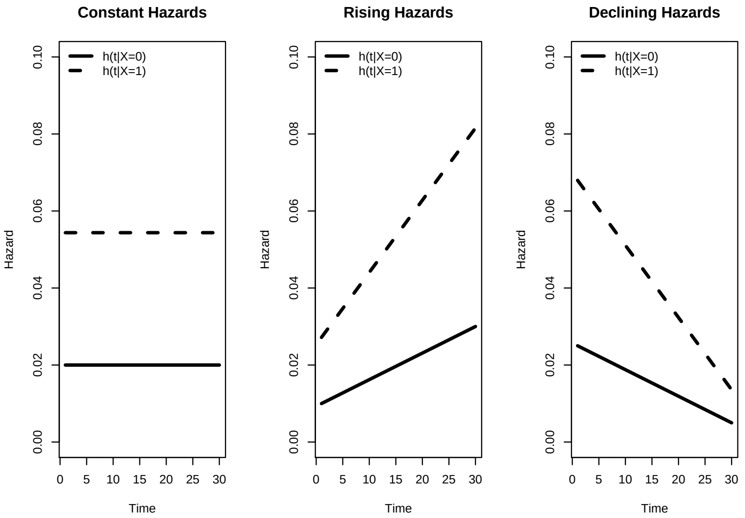
$$h_A(t) = Ch_B(t)$$

where C is the hazard ratio between A and B.

Proportionality:

- "Flat" hazards $\rightarrow$ parallel
- Rising hazards $\rightarrow$ diverging
- Falling hazards $\rightarrow$ converging

Proportional Hazards, Illustrated



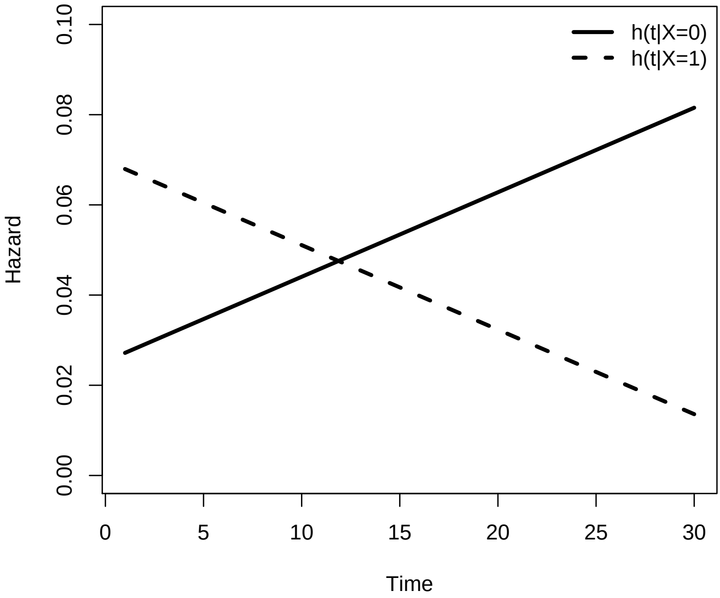
Proportional Hazards, continued

Why might hazards not be proportional?

- Resistance ($\rightarrow$ converging hazards)
- Learning ($\rightarrow$ converging hazards)
- Reinforcement ($\rightarrow$ diverging hazards)

Also, crossing hazards (always non-proportional)

Crossing Hazards



What Proportional Hazards Mean

Covariate influence over time

- PH assumes that the (proportional) influence of covariates $\mathbf{X}$ on the hazard will be the same at any point in the duration.
- Suggests how to think about it:

Conventional model:

$$h(t|\mathbf{X}_i) = h_0(t)\exp(\mathbf{X}_i\beta)$$

Generalized model:

$$h(t|\mathbf{X}_i) = h_0(t)\exp[\mathbf{X}_i\beta + \mathbf{X}_i g(t)\gamma]$$

Tests

Three kinds of tests for nonproportionality:

1. Tests for *changes in parameter values for coefficients estimated on a subsample of the data* defined by t,
2. Tests based on *plots of survival estimates and regression residuals against time*, and
3. Explicit tests of *interactions of covariates and time*.

Piecewise Regression

Step function:

$$
\begin{aligned}
g(t) &= 0 \;\forall\; t \leq \tau \\
&= 1 \;\forall\; t > \tau
\end{aligned}
$$

Implies:

$$h_i(t) = f\{X_i\beta_1 + [g(t)_i]\beta_2 + X_i[g(t)_i]\beta_3\}$$

Things to think about:

- Abrupt change?
- Choice of t in $g(t)$
- Multiple "steps"?

log-log-Survival Plots

Kalbfleisch and Prentice (1980) note that in the Cox model:

$$S(t) = \exp\left[-\exp(\mathbf{X}_i\beta)\int_0^t h_0(t)\,dt\right]$$

which means

$$\ln\{-\ln[S(t)]\} = H_0(t) \times \mathbf{X}_i\beta.$$

Implies that plots of $\ln\{\widehat{-\ln[S(t)]}\}$ vs. $\ln(T)$ for different values of $\mathbf{X}$ should be parallel to one another.

Residual-Based Methods

Recall:

$$\hat{M}_i(t) = C_i(t) - \hat{H}_i(t)$$

where $C_i(t) \equiv N_i(t)$ is the censoring indicator at t and $\hat{H}_i(t)$ is the integrated hazard.

Proportional hazards implies:

$$\hat{M}_i(t) = C_i(t) - \underbrace{\exp(\mathbf{X}_{it}\hat{\beta})\hat{H}_0(t)}_{\text{(“Cox-Snell” residual)}}$$

Martingale Residuals

Under the usual assumptions:

- $\mathrm{E}(M_i) = 0$ and
- $\mathrm{Cov}(M_i, M_j) = 0$ asymptotically.

If data are time-varying, then $M_i(t)$ is the "partial" martingale residual, and

$$M_i = M_i(\infty) = \sum_{t=1}^{t_i} M_i(t)$$

Schoenfeld Residuals

$$\begin{aligned}
\frac{\partial \ln L(\beta)}{\partial \beta_k} &= \sum_{i=1}^{N} C_i \left\{ X_{ik} - \frac{\sum_{j \in R(t)} X_{jk} \exp(X_j \beta)}{\sum_{j \in R(t)} \exp(X_j \beta)} \right\} \\
&= \sum_{i=1}^{N} C_i (X_{ik} - \bar{X}_{w_i k}).
\end{aligned}$$

$$\hat{r}_{ik} = C_i \left[X_{ik} - \frac{\sum_{j \in R(t)} X_{jk} \exp(X_j \hat{\beta})}{\sum_{j \in R(t)} \exp(X_j \hat{\beta})} \right]$$

Schoenfeld Residuals

Intuition:

> *"(Schoenfeld residuals) ...can essentially be thought of as the observed minus the expected values of the covariate at each failure time."*

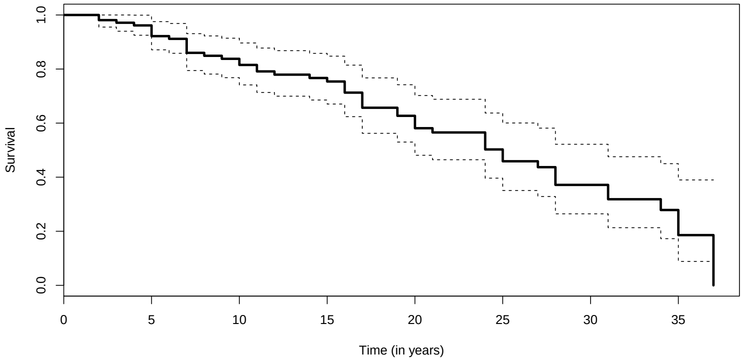
– Box-Steffensmeier and Jones (2004, 121)

Properties:

- Are defined only at event times, for non-censored observations,
- $\sum_{i=1}^{N} \hat{r}_{ik} = 0$
- $\text{Cov}(\hat{r}_{ik}, T) = 0$ if X_k's effect is proportional
- Tend to be skewed; in practice, *scaled* Schoenfeld residuals are used (see Grambsch and Therneau 1994).

Example: Supreme Court Departures

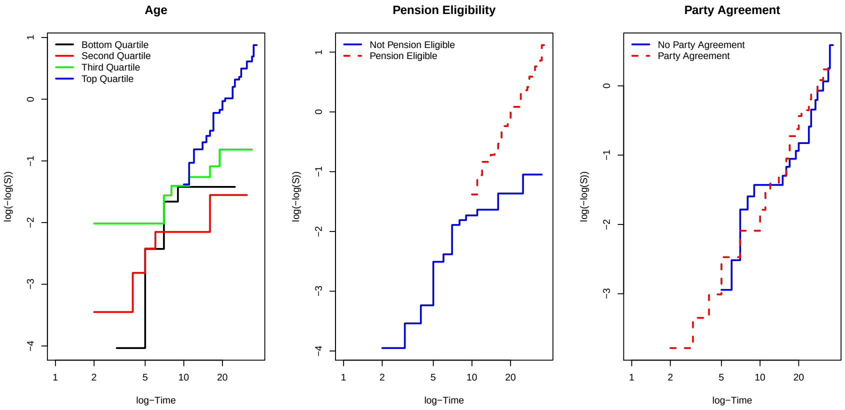
```
> summary(scotus)
    justice          service          retire             age           pension           pagree      
 Min.   :  1.00   Min.   : 1.00   Min.   :0.0000   Min.   :32.0   Min.   :0.0000   Min.   :0.0000  
 1st Qu.: 26.00   1st Qu.: 5.00   1st Qu.:0.0000   1st Qu.:56.0   1st Qu.:0.0000   1st Qu.:0.0000  
 Median : 51.00   Median :10.00   Median :0.0000   Median :62.0   Median :0.0000   Median :1.0000  
 Mean   : 52.13   Mean   :11.74   Mean   :0.0289   Mean   :62.1   Mean   :0.1989   Mean   :0.6164  
 3rd Qu.: 78.00   3rd Qu.:17.00   3rd Qu.:0.0000   3rd Qu.:69.0   3rd Qu.:0.0000   3rd Qu.:1.0000  
 Max.   :107.00   Max.   :37.00   Max.   :1.0000   Max.   :91.0   Max.   :1.0000   Max.   :1.0000  
```



SCOTUS Departures: Cox Regression

```
> scotus.Cox<-coxph(scotus.S~age+pension+pagree,data=scotus,ties="efron")

> summary(scotus.Cox)
Call:
coxph(formula = scotus.S ~ age + pension + pagree, data = scotus,
    ties = "efron")

  n= 1765, number of events= 51

           coef exp(coef) se(coef)     z Pr(>|z|)
age     0.06395   1.06604  0.02731 2.341 0.019216 *
pension 2.05136   7.77847  0.55040 3.727 0.000194 ***
pagree  0.13748   1.14738  0.29831 0.461 0.644898
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

        exp(coef) exp(-coef) lower .95 upper .95
age         1.066     0.9381    1.0105     1.125
pension     7.778     0.1286    2.6448    22.877
pagree      1.147     0.8716    0.6394     2.059

Concordance= 0.647  (se = 0.049 )
Rsquare= 0.022   (max possible= 0.194 )
Likelihood ratio test= 38.82  on 3 df,   p=1.898e-08
Wald test            = 26.82  on 3 df,   p=6.426e-06
Score (logrank) test = 35.27  on 3 df,   p=1.068e-07
```

log-log-Survival Plots



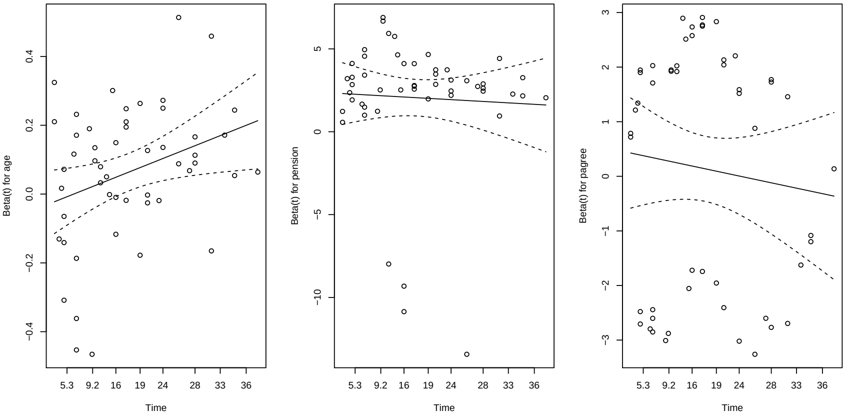
Martingale Residuals

```
>scotus$mgres<-residuals(scotus.Cox,type="martingale")

> # William Howard Taft...
> print(scotus[scotus$justice==69,])
     justice service retire age pension pagree       mgres
1173      69       1      0  63       0      1  0.00000000
1174      69       2      0  64       0      1 -0.03510077
1175      69       3      0  65       0      1 -0.01816026
1176      69       4      0  66       0      1 -0.01899776
1177      69       5      0  67       0      1 -0.07903096
1178      69       6      0  68       0      1 -0.02063125
1179      69       7      0  69       0      1 -0.11090925
1180      69       8      0  70       0      1 -0.02384340
1181      69       9      0  71       0      1 -0.02117129
1182      69      10      1  72       1      1  0.87052892

> L.Q.C. Lamar:
> print(scotus[scotus$justice==49,])
    justice service retire age pension pagree       mgres
851      49       1      0  62       0      1  0.00000000
852      49       2      0  63       0      0 -0.02869710
853      49       3      0  64       0      0 -0.01484716
854      49       4      0  65       0      0 -0.01553187
855      49       5      0  66       0      0 -0.06461280
856      49       6      0  67       0      1 -0.01935322
```

Schoenfeld Residuals / Tests

```
> # scotus.schres<-residuals(scotus.Cox,type="schoenfeld")
> # scotus.scares<-residuals(scotus.Cox,type="scaledsch")

> PHtest<-cox.zph(scotus.Cox)
> PHtest
            rho chisq      p
age      0.3160 5.359 0.0206
pension -0.0471 0.113 0.7370
pagree  -0.0962 0.504 0.4779
GLOBAL       NA 5.824 0.1205
```

Plots of Schoenfeld Residuals



log-Time Interactions

Model becomes:

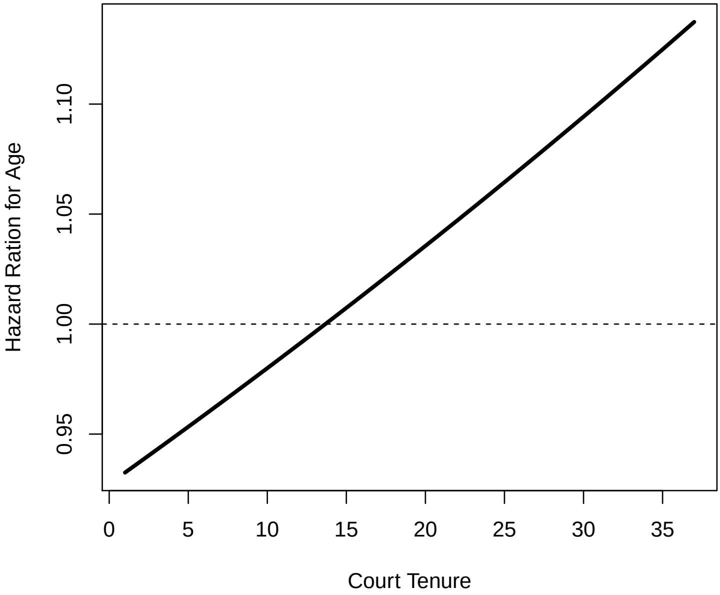
$$h_i(t) = h_0(t)\exp[X_i\beta + X_i \ln(T_i)\gamma + ...]$$

- Implies that the effect of the covariate on $h(t)$ varies linearly in T
- No T term is included
- Interpretation is standard

log-Time Interactions

```
> scotus$lnT<-log(scotus$service)
> scotus$ageLnT<-scotus$age*(scotus$lnT)
> scotus.NPH<-coxph(scotus.S~age+pension+pagree+ageLnT,
          data=scotus,ties="efron")
> summary(scotus.NPH)

  n= 1765, number of events= 51

            coef exp(coef) se(coef)      z Pr(>|z|)
age     -0.06988   0.93251  0.07729 -0.904 0.365933
pension  1.99866   7.37915  0.55167  3.623 0.000291 ***
pagree   0.09501   1.09966  0.30298  0.314 0.753849
ageLnT   0.05499   1.05653  0.03062  1.796 0.072552 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Concordance= 0.605  (se = 0.049 )
Rsquare= 0.023   (max possible= 0.194 )
Likelihood ratio test= 41.88  on 4 df,   p=1.768e-08
Wald test            = 28.84  on 4 df,   p=8.429e-06
Score (logrank) test = 36.55  on 4 df,   p=2.232e-07
```

Hazard Ratio Changes Over Time



More Proportional Hazards Tests

```
> PHtest2<-cox.zph(scotus.NPH)

> PHtest2
            rho   chisq     p
age     -0.1388 1.02621 0.311
pension  0.0126 0.00814 0.928
pagree  -0.1086 0.66902 0.413
ageLnT   0.1878 1.66856 0.196
GLOBAL       NA 2.58245 0.630
```

Additional Considerations

- Licht (2012): Inclusion of $\ln(T)$ interactions alters the substantive interpretation of the regression results.

- Keele (2010): Residual-based tests for nonproportionality can also be detecting model misspecification (specifically, unmodeled nonlinearity).

Duration Dependence

1. *State Dependence*
   - E.g., Institutionalization / Degradation

Positive State Dependence $\longrightarrow$ Negative Duration Dependence

Negative State Dependence $\longrightarrow$ Positive Duration Dependence

Duration Dependence

2. *Unobserved / Unmodeled Heterogeneity*

- $h(t|\mathbf{X}_i) \neq h(t|\mathbf{X}_j)$ for $\mathbf{X}_i = \mathbf{X}_j$
- Adverse selection in the sample / data
- Result: "Spurious" duration dependence

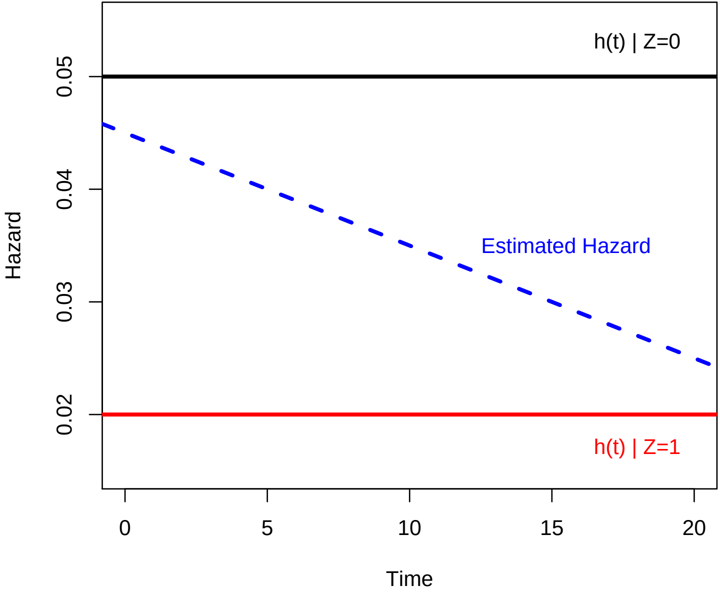
Suppose we have an unobserved Z, with

$$h_i(t|\mathbf{X}_i, Z_i = 0) = 0.05$$

and

$$h_i(t|\mathbf{X}_i, Z_i = 1) = 0.02.$$

Unobserved Heterogeneity Illustrated



Unobserved Heterogeneity: A Simulation

```
> set.seed(7222009)
> W<-rnorm(500)
> X<-rnorm(500)
> Z<-rnorm(500)
> T<-rexp(500,rate=(exp(0+0.5*W+0.5*X-0.6*Z))) # exponential hazard
> C<-rep(1,times=500)
> S<-Surv(T,C)
> summary(survreg(S~W,dist="weibull"))

Call:
survreg(formula = S ~ W, dist = "weibull")
              Value Std. Error      z        p
(Intercept) -0.0101     0.0629  -0.16 8.73e-01
W           -0.6339     0.0610 -10.40 2.47e-25
Log(scale)   0.2833     0.0333   8.52 1.62e-17

Scale= 1.33 # implies p = 1/Scale = 0.753

Weibull distribution
Loglik(model)= -568.1   Loglik(intercept only)= -615.3
Chisq= 94.47 on 1 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

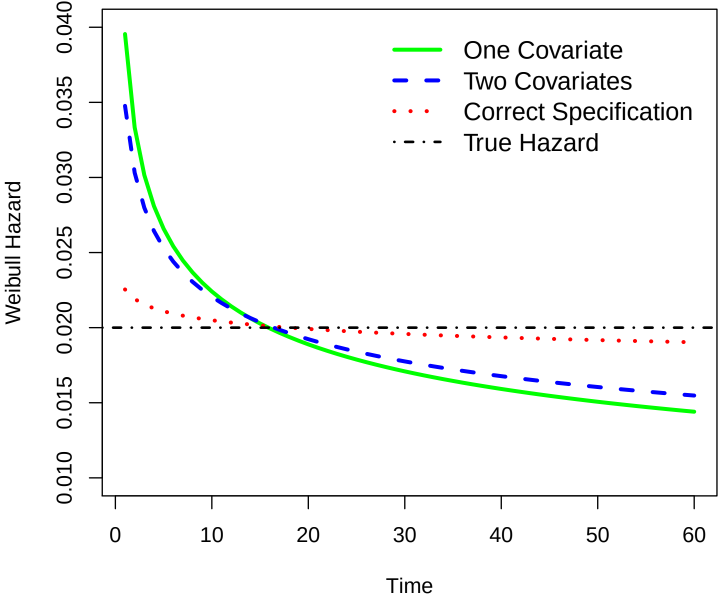
Unobserved Heterogeneity: A Simulation

```
> summary(survreg(S~W+X,dist="weibull"))

Call:
survreg(formula = S ~ W + X, dist = "weibull")
              Value Std. Error       z        p
(Intercept) -0.0511     0.0591  -0.865 3.87e-01
W           -0.5907     0.0581 -10.160 2.98e-24
X           -0.4750     0.0556  -8.549 1.24e-17
Log(scale)   0.2202     0.0329   6.689 2.24e-11

Scale= 1.25 # implies p = 1/Scale = 0.802

Weibull distribution
Loglik(model)= -534.5   Loglik(intercept only)= -615.3
Chisq= 161.6 on 2 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

Unobserved Heterogeneity: A Simulation

```
> summary(survreg(S~W+X+Z,dist="weibull"))

Call:
survreg(formula = S ~ W + X + Z, dist = "weibull")
              Value Std. Error      z        p
(Intercept) -0.0777     0.0494  -1.57 1.16e-01
W           -0.5665     0.0468 -12.11 9.17e-34
X           -0.5041     0.0473 -10.66 1.58e-26
Z            0.5923     0.0446  13.29 2.73e-40
Log(scale)   0.0423     0.0345   1.22 2.21e-01

Scale= 1.04 # implies p = 1/Scale = 0.959

Weibull distribution
Loglik(model)= -464.3   Loglik(intercept only)= -615.3
Chisq= 302.01 on 3 degrees of freedom, p= 0
Number of Newton-Raphson Iterations: 5
n= 500
```

Unobserved Heterogeneity: A Simulation



Duration Dependence: What To Do?

(At least) Three Options:

1. Model Specification
2. Unit-Level Effects
3. <u>Model</u> the Duration Dependence

Modeling Duration Dependence

Weibull with:

$$p = \exp(\mathbf{Z}_i\gamma)$$

Gives:

$$h_i(t) = \exp(\mathbf{X}_i\beta)\exp(\mathbf{Z}_i\gamma)[\exp(\mathbf{X}_i\beta)t]^{[\exp(\mathbf{Z}_i\gamma)]-1}$$

and (more usefully):

$$S(t) = \exp(-\exp(\mathbf{X}_i\beta)t)^{\exp(\mathbf{Z}_i\gamma)}$$

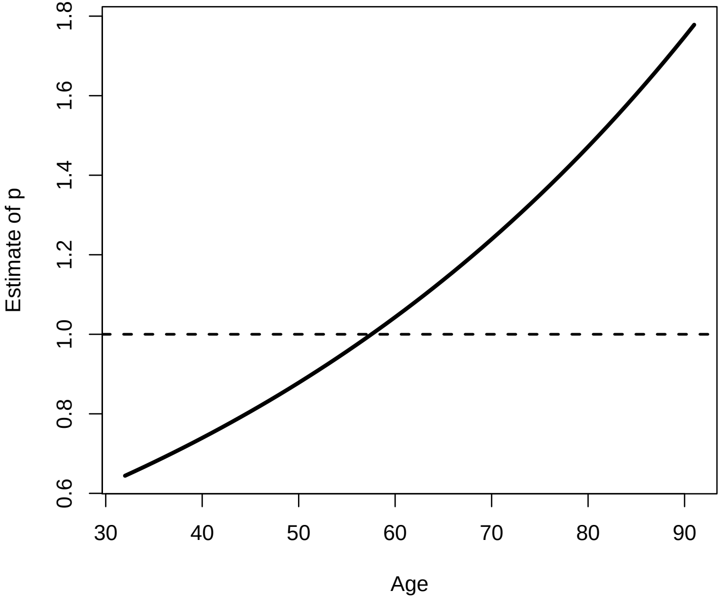
Example: SCOTUS Departures

```
> library(flexsurv)
> ct.weib<-flexsurvreg(scotus.S~age+pension+pagree,
                    data=scotus,dist="weibull")
> ct.weib

Estimates:
        data mean  est         L95%       U95%       exp(est)
shape          NA      0.999      0.637      1.570         NA
scale          NA    942.000     13.700  64800.000         NA
age        62.100     -0.041     -0.102      0.019      0.959
pension     0.199     -1.310     -2.360     -0.265      0.269
pagree      0.616     -0.113     -0.673      0.447      0.893
        L95%       U95%
shape          NA         NA
scale          NA         NA
age         0.903      1.020
pension     0.095      0.767
pagree      0.510      1.560

N = 1765,  Events: 51,  Censored: 1714
Total time at risk: 1765
Log-likelihood = -209, df = 5
AIC = 429
```

Example: SCOTUS Departures

```
> ct.weib.DD<-flexsurvreg(scotus.S~age+pension+pagree+shape(age),
                        data=scotus,dist="weibull")
> ct.weib.DD

Estimates:
           data mean   est         L95%        U95%
shape             NA      0.3710      0.1260      1.0900
scale             NA    491.0000     16.7000  14500.0000
age          62.1000     -0.0307     -0.0779      0.0164
pension       0.1990     -1.0900     -1.9700     -0.2190
pagree        0.6160     -0.0328     -0.4840      0.4180
shape(age)   62.1000      0.0172     -0.0011      0.0356
           exp(est)    L95%        U95%
shape             NA          NA          NA
scale             NA          NA          NA
age           0.9700      0.9250      1.0200
pension       0.3350      0.1400      0.8030
pagree        0.9680      0.6160      1.5200
shape(age)    1.0200      0.9990      1.0400

N = 1765,  Events: 51,  Censored: 1714
Total time at risk: 1765
Log-likelihood = -208, df = 6
AIC = 427
```

$\hat{p}$ by Age



$\widehat{h(t)}$s by Age and Model