25) Cox's Proportional Hazard Model

Vitor Kamada

July 2018

Reference

Tables, Graphics, and Figures from:

1) https://lifelines.readthedocs.io/

Survival Regression

2) Cameron and Trivedi (2005)

Ch 17 - Transition Data: Survival Analysis

Standard Parametric Models

Parametric Model	Hazard Function	Survivor Function
Exponential	γ	$\exp(-\gamma t)$
Weibull	$\gamma\alpha t^{\alpha-1}$	$\exp(-\gamma t^{\alpha})$
Generalized Weibull	$\gamma\alpha t^{\alpha-1} S(t)^{-\mu}$	$[1-\mu\gamma t^{\alpha}]^{1/\mu}$
Gompertz	$\gamma \exp(\alpha t)$	$\exp(-(\gamma/\alpha)(e^{\alpha t}-1))$
Log-normal	$\frac{\exp\left(-(\ln t-\mu)^2/2\sigma^2\right)}{t\sigma\sqrt{2\pi}[1-\Phi((\ln t-\mu)/\sigma)]}$	$1-\Phi\left((\ln t-\mu)/\sigma\right)$
Log-logistic	$\alpha\gamma^{\alpha} t^{\alpha-1}/\left[(1+(\gamma t)^{\alpha})\right]$	$1/\left[1+(\gamma t)^{\alpha}\right]$
Gamma	$\frac{\gamma(\gamma t)^{\alpha-1}\exp[-(\gamma t)]}{\Gamma(\alpha)[1-I(\alpha,\gamma t)]}$	$1-I(\alpha,\gamma t)$

Fully Parametric Analysis: MLE

$$Pr[T > t] = \int_t^\infty f(u|x,\theta)du = 1 - F(t|x,\theta) = S(t|x,\theta)$$

$$f(t_i|x_i,\theta)^{\delta_i} S(t_i|x_i,\theta)^{1-\delta_i}$$

$$\delta_i \begin{cases} 1 & no & censoring \\ 0 & right & censoring \end{cases}$$

$$Ln(\theta) = \sum_{i=1}^{N} [\delta_i lnf(t_i|x_i,\theta) + (1-\delta_i)] lnS(t_i|x_i,\theta)$$

$$Ln(\theta) = \sum_{i=1}^{N} [\delta_i ln\lambda(t_i|x_i,\theta) + \Lambda(t_i|x_i,\theta)]$$

Cox's Proportional Hazard Model

$$\lambda(t|x,\beta) = \lambda_0(t)\phi(x,\beta)$$

$$\lambda(t|x,\beta) = \lambda_0(t)exp(\beta_1 x_1 + ... + \beta_k x_k)$$

$$\eta_i = \beta_1 x_{i1} + ... + \beta_k x_{ik}$$

$$\eta_j = \beta_1 x_{j1} + ... + \beta_k x_{jk}$$

$$\frac{\lambda_i(t|x,\beta)}{\lambda_j(t|x,\beta)} = \frac{\lambda_0(t)e^{\eta_i}}{\lambda_0(t)e^{\eta_j}} = \frac{e^{\eta_i}}{e^{\eta_j}}$$

McCall (1996)

Current Population Survey's Displaced Workers Supplements (DWS) for the years 1986, 1988, 1990, and 1992

spell: if person is re-employed at a full-time job

UI: the subject filed an unemployment claim or not

Replacement: weekly benefit amount divided by the amount of weekly earnings in the lost job

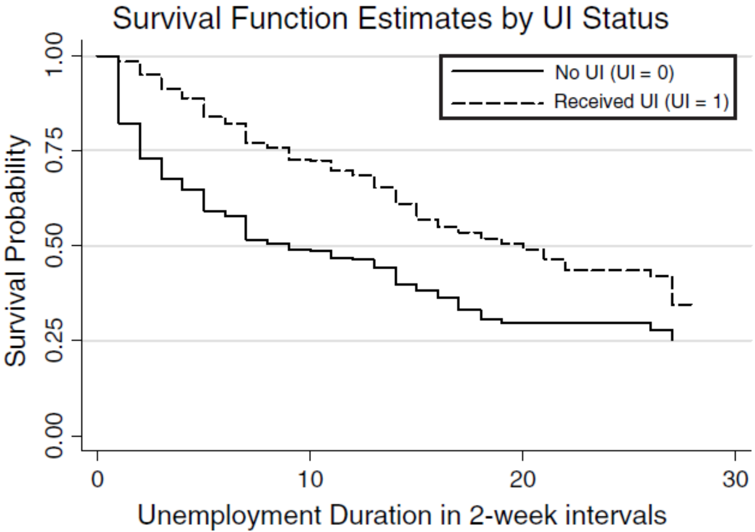
Disregard: the threshold amount up to which recipients of unemployment insurance who accept part-time work can earn without any reduction in unemployment benefits

Key Economic Covariates to Explain Joblessness Duration

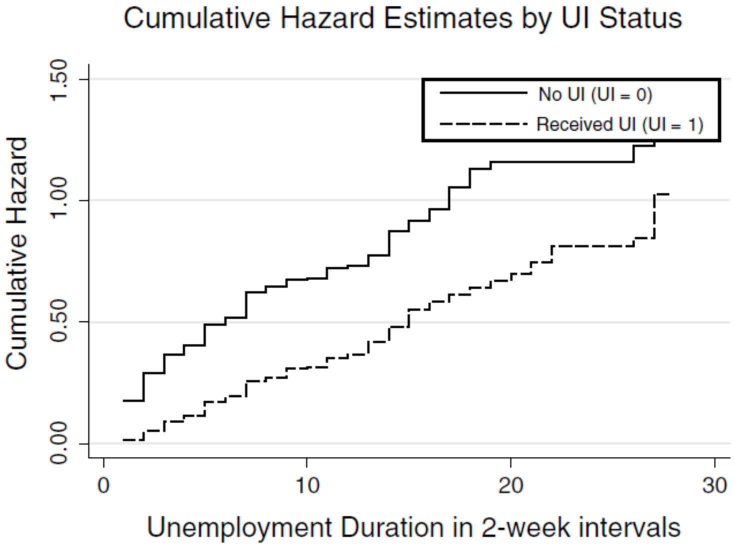
Variable Name	Variable Label	Mean
spell	periods jobless: two-week interval	6.248
CENSOR1	1 if reemployed at full-time job	0.321
CENSOR2	1 if reemployed at part-time job	0.102
CENSOR3	1 if reemployed but left job: pt–ft status unknown	0.172
CENSOR4	1 if still jobless	0.375
UI	1 if filed UI claim	0.553
RR	eligible replacement rate	0.454
DR	eligible disregard rate	0.109
TENURE	tenure years in lost job	4.114
LOGWAGE	log weekly earnings	5.693

+ large # of covariates

Survival Function: Unemployment Duration by Unemployment Insurance



Cumulative Hazard Functions: Unemployment Duration by Unemployment Insurance



Estimated Parameters

	Exponential		Weibull		Gompertz		Cox PH	
Var	**coeff.**	***t***	**coeff.**	***t***	**coeff.**	***t***	**coeff.**	***t***
RR	0.472	0.79	0.448	0.70	0.472	0.78	0.522	0.91
DR	−0.576	−0.75	−0.427	−0.53	−0.563	−0.74	−0.753	−1.04
UI	−1.425	−5.71	−1.496	−5.67	−1.428	−5.69	−1.317	−5.55
RRUI	0.966	0.92	1.105	1.57	0.969	1.58	0.882	1.52
DRUI	−0.199	−0.20	−0.299	−0.28	−0.211	−0.21	−0.095	−0.10
LOGWAGE	0.35	3.03	0.37	2.99	0.35	3.03	0.34	3.03
CONS	−4.079	−4.65	−4.358	−4.74	−4.097	−4.65	–	–
α			1.129					
−ln L	2700.7		2687.6		2700.6		–	

Positive state dependence ($\alpha = 1.129 > 1$): the probability of the spell terminating increases as the spell lengthens

Hazard Ratios

	Exponential		Weibull		Gompertz		Cox PH	
Var	β	t	β	t	β	t	β	t
RR	1.603	0.63	1.565	0.57	1.604	0.62	1.686	0.71
DR	0.562	−1.02	0.653	−0.66	0.570	−0.99	0.471	−1.55
UI	0.241	−12.65	0.224	−13.12	0.240	−12.65	0.268	−11.53
RRUI	2.626	1.01	2.760	0.99	2.635	1.01	2.416	1.01
DRUI	0.819	−0.22	0.742	−0.33	0.810	−0.23	0.909	−0.10
LOGWAGE	1.420	2.56	1.441	0.08	1.42	2.55	1.40	2.57
α			1.129					
$-\ln L$	2700.7		2687.6		2700.6		–	

Cox PH model has no intercept and makes no assumption about the shape of the baseline hazard

Experimental Study of Recidivism

week: week of first arrest after release, or censoring time

arrest: 1 for those arrested during the period of the study

fin: financial aid was a randomly assigned

race: 1 for black

wexp: full-time work experience prior to incarceration

mar: individual was married

paro: individual was released on parole

prio: # of prior convictions

from lifelines.datasets import load_rossi

rossi_dataset = load_rossi()

rossi_dataset.sample(9)

	week	arrest	fin	age	race	wexp	mar	paro	prio
308	52	0	0	19	1	1	0	1	3
129	52	0	0	22	0	0	0	1	4
93	52	0	1	24	1	0	0	0	1
261	52	0	0	22	1	0	0	0	1
103	49	1	1	19	1	0	0	1	1
243	52	0	0	20	1	1	0	1	1
211	52	0	1	30	1	0	0	1	1
410	27	1	0	20	0	1	0	0	1
40	52	0	1	33	1	1	1	0	9

from lifelines import CoxPHFitter

cph = CoxPHFitter()

cph.fit(rossi_dataset, duration_col='week', event_col='arrest')

cph.print_summary()

```
n=432, number of events=114

        coef  exp(coef)  se(coef)        z      p  lower 0.95  upper 0.95
fin  -0.3794     0.6843    0.1914  -1.9826 0.0474     -0.7545     -0.0043   *
age  -0.0574     0.9442    0.0220  -2.6109 0.0090     -0.1006     -0.0143  **
race  0.3139     1.3688    0.3080   1.0192 0.3081     -0.2898      0.9176
wexp -0.1498     0.8609    0.2122  -0.7058 0.4803     -0.5657      0.2662
mar  -0.4337     0.6481    0.3819  -1.1357 0.2561     -1.1822      0.3147
paro -0.0849     0.9186    0.1958  -0.4336 0.6646     -0.4685      0.2988
prio  0.0915     1.0958    0.0286   3.1938 0.0014      0.0353      0.1476  **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Concordance = 0.640
Likelihood ratio test = 33.266 on 7 df, p=0.00002
```

cph.plot_covariate_groups('prio', [0, 5, 10, 15])