Analysis of trajectories into retirement using the Danish labour market registry

Jolien Cremers

Department of Public Health, University of Copenhagen & Statistics Denmark

December 3rd, 2020

The Danish labour market registry

id	month	sex	birthdate	status	hours	start	end
1	January	M	08-03-1950	sick leave	148	01-01-2011	31-01-2011
2	January	F	21-09-1950	employed	10	20-01-2011	31-01-2011
2	Febuary	F	21-09-1950	employed	50	01-02-2011	28-02-2011
3	January	M	02-11-1950	employed	10	01-01-2011	15-01-2011
3	January	M	02-11-1950	employed	10	01-01-2011	15-01-2011
3	January	M	02-11-1950	unemployment benefits	74	16-01-2011	31-01-2011
3	Febuary	M	02-11-1950	unemployment benefits	37	01-02-2011	07-02-2011
3	Febuary	M	02-11-1950	early retirement pension	109	08-02-2011	28-02-2011
4	March	F	07-04-1950	early retirement pension	148	01-03-2011	31-03-2011
5	January	M	08-12-1950	employed	148	01-01-2011	31-01-2011

Table 1: Simulated example of the data contained in the labour market register.

Early Retirement Pension (ERP)

Table 2: Ages at which Early Retirement Pension (ERP) is available according to date of birth.

Date of Birth	ERP age
$<$ 1954	60
$\geq$ 01-January-1954	60.5
$\geq$ 01-July-1954	61
$\geq$ 01-January-1955	61.5
$\geq$ 01-July-1955	62
$\geq$ 01-January-1956	62.5
$\geq$ 01-July-1956	63
$\geq$ 01-January-1959	63.5
$\geq$ 01-July-1959	64
$>$ 1963	computed in relation to life expectancy

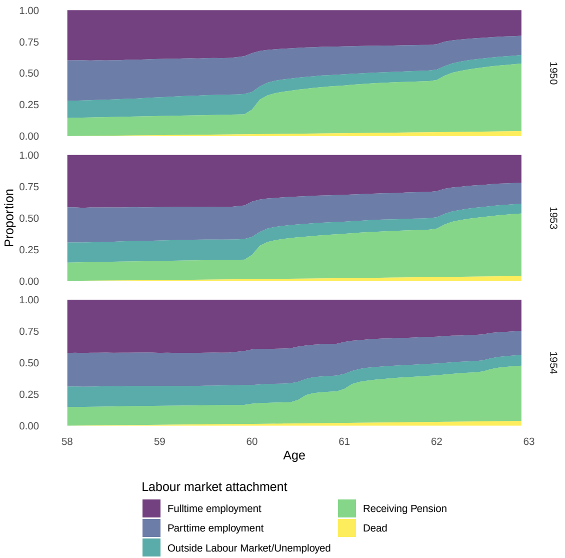

Figure 1: Attachment to the labour market over time for three cohorts.

Sequence analysis

Sequence analysis

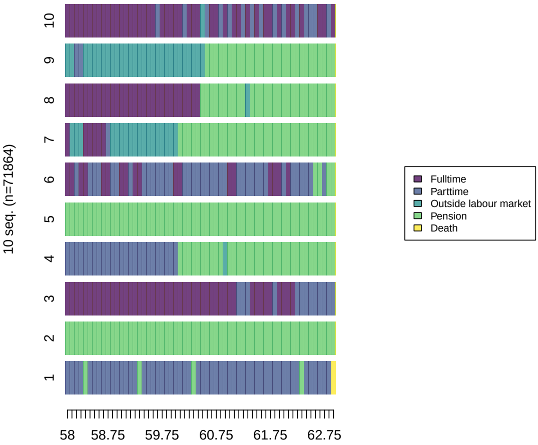

Figure 2: Ten example sequences for the 1950 cohort.

Implementation in R

R-packages:

- TraMineR
- TraMineRextras
- WeightedCluster

Clusters 1950 cohort

Clusters 1954 cohort

Crosstabulations

Cluster	1950	1954
fulltime - late pension	0.28	0.31
fulltime - early pension	0.09	0.09
parttime - late pension	0.21	0.19
parttime - early pension	-	0.09
outside labour market	0.07	0.16
mixed - early pension	0.18	-
pension	0.16	0.15

Crosstabulations II

Category	Male	Female
fulltime - late pension	14236	8143
outside labour market - late pension	2999	2929
fulltime - early pension	6566	8114
parttime - late pension	4274	7461
mixed - early pension	4765	6277
pension	3076	2908

But

What if we want to model the time to pension?

Joint Models for Longitudinal and Survival data

Joint Models for Longitudinal and Survival data

Individuals $i = 1, \ldots, N$, Timepoints $t = 1, \ldots, n_i$

Longitudinal submodel:

$$y_i(t) = \boldsymbol{x}_i(t)'\beta + \boldsymbol{z}_i(t)'\boldsymbol{b}_i + \epsilon_i(t)$$

Survival submodel:

$$\lambda_i(t) = \lambda_0(t)\exp(\boldsymbol{s}_i(t)'\boldsymbol{\gamma} + w_i(t))$$

Why?

Bias:

- Endogeneity
- non-random dropout
- measurement error

Joint Model

Longitudinal:

$$\pi_{itk} = P(Y_{it} = k) = \begin{cases} \dfrac{1}{1 + \sum_{h=1}^{K-1} \exp(\mathbf{X}_{it}\beta + b_{ih})} & \text{if } k = K \\ \dfrac{\exp(\mathbf{X}_{it}\beta + b_{ik})}{1 + \sum_{h=1}^{K-1} \exp(\mathbf{X}_{it}\beta + b_{ih})} & \text{if } k = 1, \dots, K-1, \end{cases}$$

$k \in \{1, \dots, K\} = \{\text{Fulltime, Parttime, Outside the labour market}\}$, b_{ih} = random intercepts.

Survival:

$$\log \mu_{it} = \log t_{it} + \mathbf{X}_{it}\boldsymbol{\eta} + \alpha_t + u_i,$$

μ_{it} = hazard, $\alpha_t = \log \lambda_t$ = log baseline hazard, t_{it} = time at risk (offset), u_i = frailty, $\log \lambda_t = \mu_\lambda + N(\log \lambda_{t-1}, \sigma_\lambda)$.

Joint Model

Random intercept b_{ih} and frailty u_i assumed to follow a multivariate normal distribution with the following variance-covariance matrix:

$$\boldsymbol{\Sigma} = \begin{bmatrix} \Sigma_{\boldsymbol{b}} & \Sigma'_{\boldsymbol{b}u} \\ \Sigma_{\boldsymbol{b}u} & \sigma_u^2 \end{bmatrix},$$

$$\Sigma_{\boldsymbol{b}} = \begin{bmatrix} \sigma^2_{\boldsymbol{b}_1} & \sigma_{\boldsymbol{b}_1,\boldsymbol{b}_2} \\ \sigma_{\boldsymbol{b}_2,\boldsymbol{b}_1} & \sigma^2_{\boldsymbol{b}_2} \end{bmatrix}, \;\; \Sigma_{\boldsymbol{bu}} = (\sigma_{\boldsymbol{b}_1,\boldsymbol{u}}, \sigma_{\boldsymbol{b}_2,\boldsymbol{u}})^t$$

Implementation in R

- Bayesian model
- Rstan

Results

1. Fixed Effects: Survival Submodel

	1950 cohort		1954 cohort	
	HR	CI	HR	CI
sex	1.47	(1.14, 1.87)	1.92	(1.45, 2.49)
education	0.78	(0.59, 1.01)	0.64	(0.46, 0.85)

2. Fixed Effects: Longitudinal Submodel

		1950 cohort			1954 cohort		
		β	OR	CI	β	OR	CI
fulltime employment	intercept	2.65	14.15	(1.91, 3.37)	3.00	20.09	(2.12, 3.90)
	sex	-2.23	0.11	(-3.28, -1.23)	-2.06	0.13	(-3.25, -0.88)
	education	1.81	6.11	(0.69, 2.97)	2.29	9.87	(1.05, 3.52)
parttime employment	intercept	1.51	4.53	(1.00, 2.05)	1.19	3.29	(0.43, 1.97)
	sex	0.65	1.92	(-0.07, 1.34)	0.31	1.36	(-0.68, 1.30)
	education	0.53	1.70	(-0.28, 1.39)	1.71	5.53	(0.60, 2.77)

3. Random Effects/Frailties

	1950 cohort		1954 cohort	
	mean	CI	mean	CI
$r_{(\boldsymbol{b}_1,\boldsymbol{b}_2)}$	0.61	(0.53, 0.69)	0.64	(0.56, 0.72)
$r_{(\boldsymbol{b}_1,\boldsymbol{u})}$	-0.79	(-0.95, -0.58)	-0.77	(-0.95, -0.53)
$r_{(\boldsymbol{b}_2,\boldsymbol{u})}$	-0.35	(-0.59, -0.11)	-0.49	(-0.76, -0.20)
$\sigma_{\boldsymbol{b}_1}$	5.67	(5.16, 6.25)	6.31	(5.73, 6.98)
$\sigma_{\boldsymbol{b}_2}$	3.90	(3.55, 4.28)	5.21	(4.70, 5.79)
$\sigma_{\boldsymbol{u}}$	0.50	(0.32, 0.70)	0.44	(0.25, 0.65)

Sequence analysis vs. Joint Models

- typification of trajectories
- duration and pattern
- inclusion of covariates