A Joint Model for Longitudinal and Event History Data in Life Course Research: Occupational Trajectories and Time to Retirement

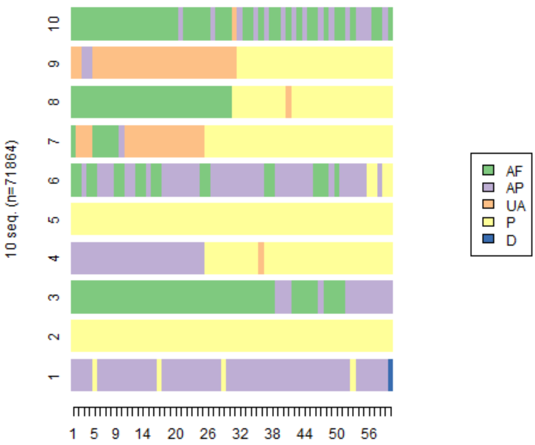
Jolien Cremers

15 Jan 2020

Labour market registry (2008 - 2017)

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	unemployed	74	16-01-2011	31-01-2011
3	Febuary	M	02-11-1950	unemployed	37	01-02-2011	07-02-2011
3	Febuary	M	02-11-1950	pension	109	08-02-2011	28-02-2011
4	March	F	07-04-1950	pension	148	01-03-2011	31-03-2011
5	January	M	08-12-1950	employed	148	01-01-2011	31-01-2011

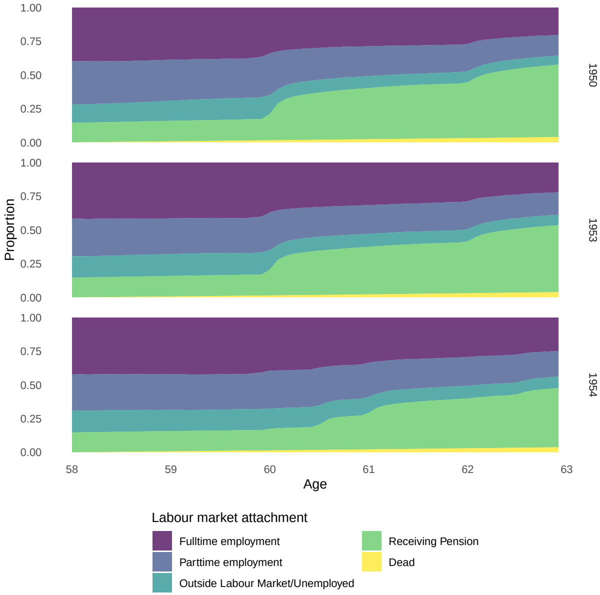
Occupational Trajectories for 10 individuals



Early Retirement Pension

Date of Birth	ERP age
< 1954	60
$\geq$ 01-01-1954	60.5
$\geq$ 01-07-1954	61
$\geq$ 01-01-1955	61.5
$\geq$ 01-07-1955	62
$\geq$ 01-01-1956	62.5
$\geq$ 01-07-1956	63
$\geq$ 01-01-1959	63.5
$\geq$ 01-07-1959	64
> 1963	computed in relation to life expectancy

Occupational Trajectories for 3 Cohorts



Joint Models

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

Longitudinal submodel:

$$y_{it} = \boldsymbol{x}_{it}^{(1)t} \boldsymbol{\beta}^{(1)} + \boldsymbol{z}_{it}^{t} \boldsymbol{b}_i + \epsilon_{ij}$$

Survival submodel:

$$\lambda_i(t) = \lambda_0(t) \exp(\boldsymbol{x}_i^{(2)^t} \boldsymbol{\beta}^{(2)} + w_i(t))$$

Joint Models

Parameterization $w_i(t)$:

- current value: $w_i(t) = \alpha \boldsymbol{z}_{it}^t \boldsymbol{b}_i$, implemented in stan_jm/JM/JMbayes
- correlated random-effects: $w_i(t) = \alpha\theta_i$, where $\begin{bmatrix} \boldsymbol{b}_i \\ \theta_i \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{bb}\Sigma_{b\theta} \\ \Sigma_{b\theta}\sigma_\theta^2 \end{bmatrix}\right)$

See Hickey et.al. (2018)[1] for more parameterizations.

[1] Hickey, G.L., Philipson, P., Jorgensen, A. & Kolamunnage-Dona, R. (2018). A comparison of joint models for longitudinal and competing risks data, with application to a epilepsy drug randomized controlled trial. Journal of the Royal Statistical Society A, 181(4), p. 1105-1123.

Joint Model: Longitudinal Submodel

Hierarchical Multinomial:

Probability state
$k \in \{1, \ldots, K\} = \{\text{Fulltime}, \text{Parttime}, \text{Outside the labour market}\}$:

$$\pi_{itk} = P(Y_{it} = k \mid \boldsymbol{x}_{it}, b_{ih}) = \begin{array}{l} \dfrac{1}{1 + \sum_{h=1}^{K-1} \exp(\boldsymbol{x}_{it}\boldsymbol{\beta}_h + b_{ih})} \text{ if } k = K \\ \dfrac{\exp(\boldsymbol{x}_{it}\boldsymbol{\beta}_h + b_{ih})}{1 + \sum_{h=1}^{K-1} \exp(\boldsymbol{x}_{it}\boldsymbol{\beta}_h + b_{ih})} \text{ if } k = 1, \ldots, K-1, \end{array}$$

$\boldsymbol{x}_{it}\boldsymbol{\beta}_h = \beta_{0h} + \beta_{1h} * \text{Sex}_i + \beta_{2h} * \text{Education}_i$
$b_{ih} = \text{random intercept}$

Joint Model: Survival Submodel

Hierarchical Poisson log-linear model (proportional hazards model with piece-wise constant baseline hazard):

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

μ_{it} = hazard
$\alpha_t = \log \lambda_t = \mu_\lambda + N(\log \lambda_{t-1}, \sigma_\lambda)$ (random walk baseline hazard)
$\boldsymbol{x}_{it}\boldsymbol{\eta} = \eta_1 * \text{Sex}_i + \eta_2 * \text{Education}_i$
u_i = frailty
t_{it} = offset

Joint Model: Link

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

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

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

Joint Model: Estimation

Bayesian Model in Stan on a subset of the data ($\pm 10\%$ stratified sample)

Priors:

- regression coefficients ($\boldsymbol{\beta}$, $\boldsymbol{\eta}$):
  - $N(\mu_0 = 0, \sigma_0 = 10000)$
- baseline hazards:
  - $\mu_\lambda \sim N(0,1)$
  - $\sigma_\lambda \sim N(0,1)$
  - $\log(\lambda_0) \sim N(0,1)$
- random effects:
  - $(\boldsymbol{b}_i, u_i) \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$
- variance-covariance random effects ($\boldsymbol{\Sigma}$):
  - $LKJ(2)$ prior for correlations ($LKJ(1) = U(-1,1)$)
  - $\exp(\lambda = 0.5)$ prior for standard deviation

Random Effects/Frailties

Instead of looking at the variance-covariance matrix we look at a matrix with correlations on the off-diagnonals and standard deviations on the diagonals:

$$\boldsymbol{R} = \begin{bmatrix} \sigma_{\boldsymbol{b}_1} & r_{(\boldsymbol{b}_1,\boldsymbol{b}_2)} & r_{(\boldsymbol{b}_1,\boldsymbol{u})} \\ r_{(\boldsymbol{b}_1,\boldsymbol{b}_2)} & \sigma_{\boldsymbol{b}_2} & r_{(\boldsymbol{b}_2,\boldsymbol{u})} \\ r_{(\boldsymbol{b}_1,\boldsymbol{u})} & r_{(\boldsymbol{b}_2,\boldsymbol{u})} & \sigma_{\boldsymbol{u}} \end{bmatrix} \tag{1}$$

Random Effects/Frailties

	1950 cohort		1953 cohort		1954 cohort	
	mean	CI	mean	CI	mean	CI
$r_{(\boldsymbol{b}_1,\boldsymbol{b}_2)}$	0.61	(0.53, 0.69)	0.64	(0.56, 0.71)	0.64	(0.56, 0.72)
$r_{(\boldsymbol{b}_1,\boldsymbol{u})}$	-0.79	(-0.95, -0.58)	-0.82	(-0.96, -0.63)	-0.77	(-0.95, -0.53)
$r_{(\boldsymbol{b}_2,\boldsymbol{u})}$	-0.35	(-0.59, -0.11)	-0.53	(-0.73, -0.31)	-0.49	(-0.76, -0.20)
$\sigma_{\boldsymbol{b}_1}$	5.67	(5.16, 6.25)	5.97	(5.41, 6.59)	6.31	(5.73, 6.98)
$\sigma_{\boldsymbol{b}_2}$	3.90	(3.55, 4.28)	4.90	(4.45, 5.42)	5.21	(4.70, 5.79)
$\sigma_{\boldsymbol{u}}$	0.50	(0.32, 0.70)	0.57	(0.38, 0.79)	0.44	(0.25, 0.65)

Fixed Effects: Survival Submodel

Hazard Ratios:

	1950 cohort		1953 cohort		1954 cohort	
	mean	CI	mean	CI	mean	CI
sex	1.47	(1.14, 1.87)	1.51	(1.17, 1.91)	1.92	(1.45, 2.49)
education	0.78	(0.59, 1.01)	0.78	(0.59, 1.01)	0.64	(0.46, 0.85)

Fixed Effects: Longitudinal Submodel

Coefficients and odds ratios:

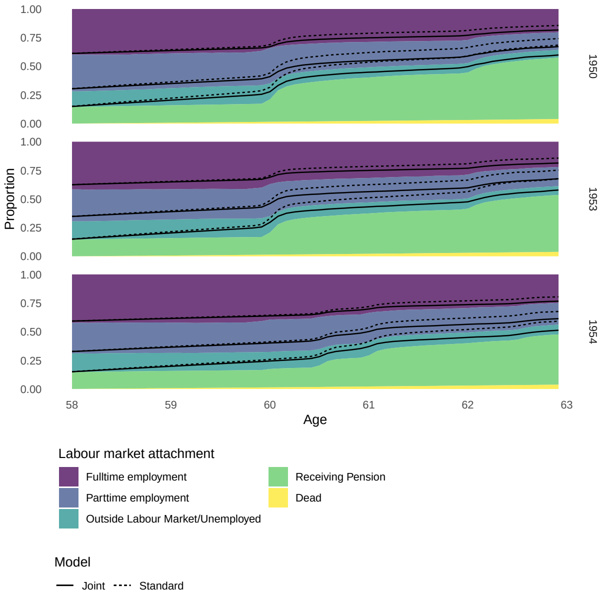
		1950 cohort			1954 cohort		
		mean		CI	mean		CI
		β	$\exp(\beta)$		β	$\exp(\beta)$	
fulltime employment	intercept	2.65	14.15	(1.91, 3.37)	3.00	20.09	(2.12, 3.90)
	sex	-2.23	0.11	(-3.28, -1.23)	-2.06	0.13	(-3.25, -0.88)
	education	1.81	6.11	(0.69, 2.97)	2.29	9.87	(1.05, 3.52)
parttime employment	intercept	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)

Fixed Effects: Longitudinal Submodel

Probabilities:

		1950 cohort		1954 cohort	
		mean	CI	mean	CI
mean	CI				
fulltime	male + low	0.71	(0.59, 0.82)	0.82	(0.71, 0.90)
	male + high	0.90	(0.81, 0.96)	0.90	(0.81, 0.96)
	female + low	0.14	(0.21, 0.57)	0.32	(0.18, 0.48)
	female + high	0.37	(0.10, 0.39)	0.49	(0.29, 0.70)
parttime	male + low	0.23	(0.14, 0.34)	0.14	(0.08, 0.22)
	male + high	0.09	(0.04, 0.17)	0.09	(0.03, 0.18)
	female + low	0.77	(0.67, 0.85)	0.55	(0.40, 0.70)
	female + high	0.59	(0.40, 0.74)	0.49	(0.29, 0.69)
outside	male + low	0.05	(0.03, 0.09)	0.04	(0.02, 0.09)
	male + high	0.01	(0.00, 0.03)	0.01	(0.00, 0.01)
	female + low	0.09	(0.05, 0.15)	0.13	(0.06, 0.23)
	female + high	0.04	(0.02, 0.08)	0.02	(0.01, 0.05)

Comparison to standard survival model



Extensions

- Increase computational speed
  - Replace multinomial setup with multiple logistic models
  - Parallell MCMC computation
  - Different parameterization (random effects + covariances)
- Include death in a competing risks setup
- Inclusion of health related variables (e.g. disease diagnosis, medicine usage) as additional longitudinal or survival outcomes