INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

22.04 – 10.05 2024 (3 ECTS)

Linus Andersson[12]

[1]University of Turku, [2]Stockholm University

linus.andersson@utu.fi

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

lecture IV, part I – Parametric models II:

Thursday 02.05

Linus Andersson[1,2]

[1]*University of Turku*, [2]*Stockholm University*

linus.andersson@utu.fi

Parametric models I

- We covered:
  - interpretation of the estimate
  - Assumptions about S(t) and h(t)
  - Proportional hazards assumption

- Now
  - Discrete and continuous time data structure for EHA models
  - Data structure with time-varying variables

Scheduling

Activity	Lesson topic	Keywords	Homework
Lesson 1	Introduction Key Concepts	Process time, Cencoring, Time-to-Event, Continous and discrete time	✓
Lesson 2	Key estimates Descriptive models	Kaplan-meyer, Density, Cum. Distribution function, Survival and Hazard funciton, Kaplan-Meyer, Life tables	✓
Lab 1. Non-Parametric models	Descriptive analysis		
Lesson 3	Key concepts, estimates for Parametric models	Exponential and Piece-wise exponential models, Shape parameter, the proportional hazard assumption, hazard ratios	✓
Lesson 4	Discrete and Continous models, Data structure	Time-varying variables, Cox,	
Lab 2. Parametric models I	Model fit		
Lesson 5	Piecing things together + Extensions	Competing risk models, Causality, heterogeneity	
Lab 3. Parametric models II	Diagnostics		
Lesson 6.	Discussion. Presentations.		Presentations

Key estimates in EHA

1. Continous and discrete event history models
2. Logic of time varying variables
3. Data structure for continous and discrete event history analysis with and without time varying variables

RECAP LESSON 2 - When to use event history analysis

- EHA appropriate when
  1. When Y is qualitative event, not quantitative
  2. We are interested in an time-dependent process leading up to Y
  3. We care about when Y occur as well as whether Y occur

RECAP LESSON i and II - Continous and discrete time

Continous time: The event of conception

- Can occur any time - From year, month, infinitesimal – a continous process
- We have monthly level data, and consider it adequate.
- We fit a continous EHA model, obtaining monthly Survival, failure, hazard rate

RECAP LESSON 2 - Continous and discrete time

”As if discrete time”: The event of conception with less precise data

- Can occur any time - From year, month, to infinitesimal – a continous process
- But we just have data on a yearly basis
- We might want to fit a discrete EHA model, estimating a probability of conception during a yearly intervall spell.

RECAP LESSON 2 - Continous and discrete time

Discrete time: The event of entry into medical school

- Can occur only at a given time – but not any given months (only in september)
- We reason that data in years is adequate
- We fit a discrete EHA model, estimating a probability of enrolling in medical school by time t

Parametric models II

	Continuous time	Discrete time
Key estimate	The hazard rate	Approximates hazard rates of cont. time models using, e.g., logistic, poisson regression
Property	A rate	A probability
Gives	Hazard ratios	Odds ratios (logistic regression)
Use when	time-units can be considered continous in relation to the time process of the studied phenomenon	Time-units are forced into ”large” intervals or if the time process is truly discrete
Data structure	“Wide” 1 row per person with columns for time process.	“Long” 1 row per person-time-interval.

Parametric models II

	b
Intercept	−3.374
Age = 50	0.166
Grandparent? (ref = no)	0.128
Age = 55	1.075
. . . × Grandparent	
Age = 60	1.527
. . . × Grandparent	
Age = 65	1.838
Age when starting first job	−0.026
Standard retirement age reached	0.463
. . . × Grandparent	−0.519
Level of education (ref = low)	
Medium	−0.193
High	−0.368
Gender = male (ref = female)	−0.371
Variance of country effect	0.450
N person years	98,806
N countries	22

Source: Own calculations based on ESS3.
°<0.1; *<0.05; **<0.005; ***<0.001.

Van Bavel (2013) European Sociological Review 29(6): 2013 1295–1308 1295 DOI:10.1093/esr/jct005

B: risk of retirement

βGrandparent = exp(0.128) = 1.14

The odds of retirement are estimated to increase with 14 per cent on becoming a grandparent

RECAP LESSON 2 - Continous and discrete time

- What is special about discrete time models?
  - Discrete time models approximate continuous time models using e.g. logistic regression
  - Gives odds ratios not hazard ratios
  - Use discrete time models only if time-units are forced into large intervals or if the time process is truly discrete
  - Data structure differ

Key estimates in EHA

1. Continous and discrete event history models ✓
2. Logic of time varying variables
3. Data structure for continous and discrete event history analysis with and without time varying variables

RECAP LESSON 2 - When to use event history analysis

- EHA appropriate when
  1. When Y is qualitative event, not quantitative
  2. We are interested in an time-dependent process leading up to Y
  3. We care about when Y occur as well as whether Y occur

Parametric models II – time-varying variables – why?

- Why time-varying covariates?
  - X sometimes change over time – can be time-varying
    - House ownership status
    - Parity progression
    - Seasonal variation
    - Even having had contact with GP a (general practitioner)

- Use to model:
  - A way to study interdependent processess
  - Gets at causality (sometimes, maybe)

Parametric models II – time-varying variables – why?

- Example: Getting a Laugh: Gender, Status and Humor in Task Discussions (Robinson and Smith–Lovin, Social forces 2001)
    - RQ: how does jokes and humor influence directions of a conversations?
    - Group discussion data.
    - Time process: time to ’joke’

Parametric models II – time-varying variables – why?

- Example: Getting a Laugh: Gender, Status and Humor in Task Discussions (Robinson and Smith–Lovin, Social forces 2001)
  - RQ: how does jokes and humor influence directions of a conversations?
  - Group discussion data.
  - Time process: time to ’joke’
  - Findings:
    - Generally: The risk of a joke decreases with time after the first joke.
    - Men joke more
    - Individuals that talk much early in conversions (a counter variable, i.a. a time-varying variable) have elevated risk of making jokes later on in conversation

Key estimates in EHA

1. Continous and discrete event history models ✓
2. Logic of time varying variables ✓
3. Data structure for continous and discrete event history analysis with and without time varying variables

Parametric models II

- Data structure for EHA is
  - Tideous data managment
  - Error prone data managment

- Data structure for EHA differ depending on
  - continuous and discrete time data in Stata
  - Uses time-varying covariates or not

Parametric models II – continuous time

- **Continous** time data structure **without** time-varying variables
  - Remember: one row per individual. Time-info in colulmns
  - Set EHA format using: `stset studytime, failure(died) id(id)`

	id	died	studytime	transplant	wait_to_transplant
1	2	1	6	0	0
2	38	1	5	1	3

Parametric models I – continuous time

- **Continous** time data structure **without** time-varying variables
  - _st = observation affected by stset command (0/1)
  - _d = cencoring variable (0 right-censored /1 died)
  - _t = duration variable. _t = studytime
  - _t0 = time of first observation

Parametric models I – continuous time

- **Continous** time data structure **without** time-varying variables
  - person 2 [id=2] : Died [_d=1] after 6 months [_t=6] without recieving a transplant [transplant=0]
  - person 38 [id=38] : Died [_d=1] after 5 months [_t=5] and recieved a transplant [transplant=1] at 5 months [wait_to_transplant=5] after signing for transplant [_t0=0]

id	died	studytime	transplant	wait_to_transplant	_t0	_t	_d
2	1	6	0	0	0	6	1
38	1	5	1	3	0	5	1

Parametric models I – continuous time

- **Continous** time data structure **witht** time-varying variables
  - person 38 `[id=38]` : Died `[_d=1]` after 5 months `[_t=5]` and recieved a transplant `[transplant=1]` at 3 months `[wait_to_transplant=3]` after signing for transplant `[_t0=0]`.

id	died	studytime	transplant	wait_to_transplant	_t0	_t	_d
2	1	6	0	0	0	6	1
38	1	5	1	3	0	5	1

  - `stsplit transpl_tv, at(0) after(wait_to_transplant)`

id	died	studytime	transpl_tv	wait_to_transplant	_t0	_t	_d
2	1	6	0	0	0	6	1
38	.	3	0	3	0	3	0
38	1	5	1	3	3	5	1

  - Person 38 has 2 rows. One ranging months 0 – 3 and one ranging month 3 – 5. Also, a dummy indicating pre/post transplant.

Parametric models I – discrete time

- **Discrete** time data structure **with and without** time-varying variables
  - If we wanted to analyse this dataset using discrete EHA
  - Every month unit needs one row
  - Last observation at the event transition (or other censoring, whichever comes first)
  - Adjust a variable indicating pre/post transplant accordingly, if required as TV variable.

id	died	studytime	transplant
2	0	1	0
2	0	2	0
2	0	3	0
2	0	4	0
2	0	5	0
2	1	6	0
38	0	1	0
38	0	2	0
38	0	3	1
38	0	4	1
38	1	5	1

Key estimates in EHA

1. Continous and discrete event history models ✓
2. Logic of time varying variables ✓
3. Data structure for continous and discrete event history analysis with and without time varying variables ✓

15 minute break

INWS0038 Longitudinal and Multilevel Modelling II - Event History Analysis

Lecture IV, part II – Parametric models II

Thursday 02.05

Linus Andersson[1,2]

[1]*University of Turku*, [2]*Stockholm University*

linus.andersson@utu.fi

Parametric models II

- We covered:
  - Interpretation of estimates
  - Model assumptions about S(t) and h(t)
  - Proportional hazards assumption
  - Discrete and continous time EHA
  - Data structure for EHA

- Now
  - The Cox model
  - Model summary

Key estimates in EHA

1. The Cox model: What's special about the Cox model?
2. Summary of models

Cox proportional hazard models

- What's so special with the Cox proportional hazard models
  - Proposed by Sir David Cox (1972) - 50,000 citations and counting
  - No assumption of the baseline hazard
  - "Semi-parametric" - does not parameterise shape, but parameterise covariate effects.
  - Does rely on the proportional hazard assumption
  - Has "own" stata command (`stcox`)

Model	Assumed shape of T
Cox	None
Exponential	Constant
Piece-wise constant exponential	Any shape across intervals
Gompertz	Increasingly increasing
Log-normal, Log-logistic	Increase AND decrease

Cox proportional hazard models

- Features
  - Covariate effects are constant over time
  - The hazard rate not meaningful interpretaiton, but covariate effects has

- Use Cox regression when
  - Interested in covariate effects rather than hazard
  - When data is continuous, few events at the same time interval (ties)
  - No theoretical reason to presume a specific distribution of T
  - Robustness checks (contrast to parametric models)

Key estimates in EHA

1. The Cox model: What's special about the Cox model? ✓
2. Summary of models

Parametric models – an overview

Model	Assumed shape of T	PH assumption	Robust vs efficient	Estimate	Used for
Kaplan-meir	none	No	Robust+++	Hazard rate	Everything w/o adjusting for covariates
Cox	None	YES	Robust++	hazard ratio	Everything
Exponential	Constant	YES	Robust+	hazard ratio	“memory-less”, Poisson-shaped things
Piece-wise constant exponential	Any shape across intervals	“NO”	Robust	hazard ratio	Everything
Gompertz	Hazard Increasing rapidly	YES	Efficient+	hazard ratio	e.g. Mortality
Log-logistic	Increase AND decrease,	YES	Efficient+	hazard ratio /AFT	Non-monotonous
Discrete time models	Various, including constant**	NO	Efficient+	Odds rate	Truly discrete time processes

Key estimates in EHA

1. The Cox model: What's special about the Cox model? ✓
2. Summary of models ✓