WZB

Social Science Research Center
Berlin

Sequence Analysis with Stata

Potential & Limits

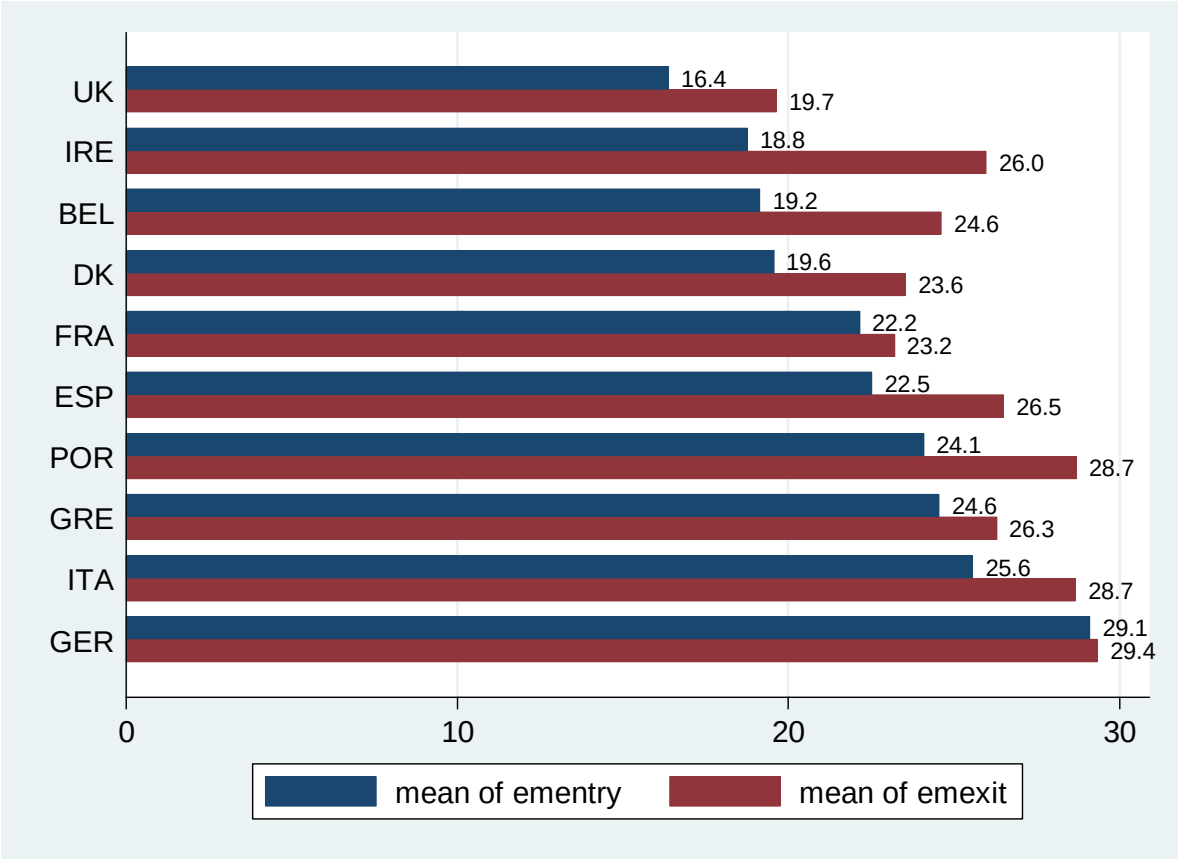
Christian Brzinsky–Fay

Lausanne Conference on Sequence Analysis"
6th/8th June 2012

Overview

- Plugin (B. Halpin, see http://teaching.sociology.ul.ie/seqanal/):
  - optimal matching (Needleman-Wunsch), duration-compensated Needleman-Wunsch & Hamming distance
  - utilities
- Plugin/ado (L. Lesnard, see http://laurent.lesnard.free.fr):
  - dynamic Hamming distance
- sq-ado's (U. Kohler/C. Brzinsky-Fay/M. Luniak)
  - optimal matching
  - egen-functions for sequence description
  - different kind of graphical displays
  - utilities (also for cluster analysis & MDS)

egen-functions: employment entry & exit

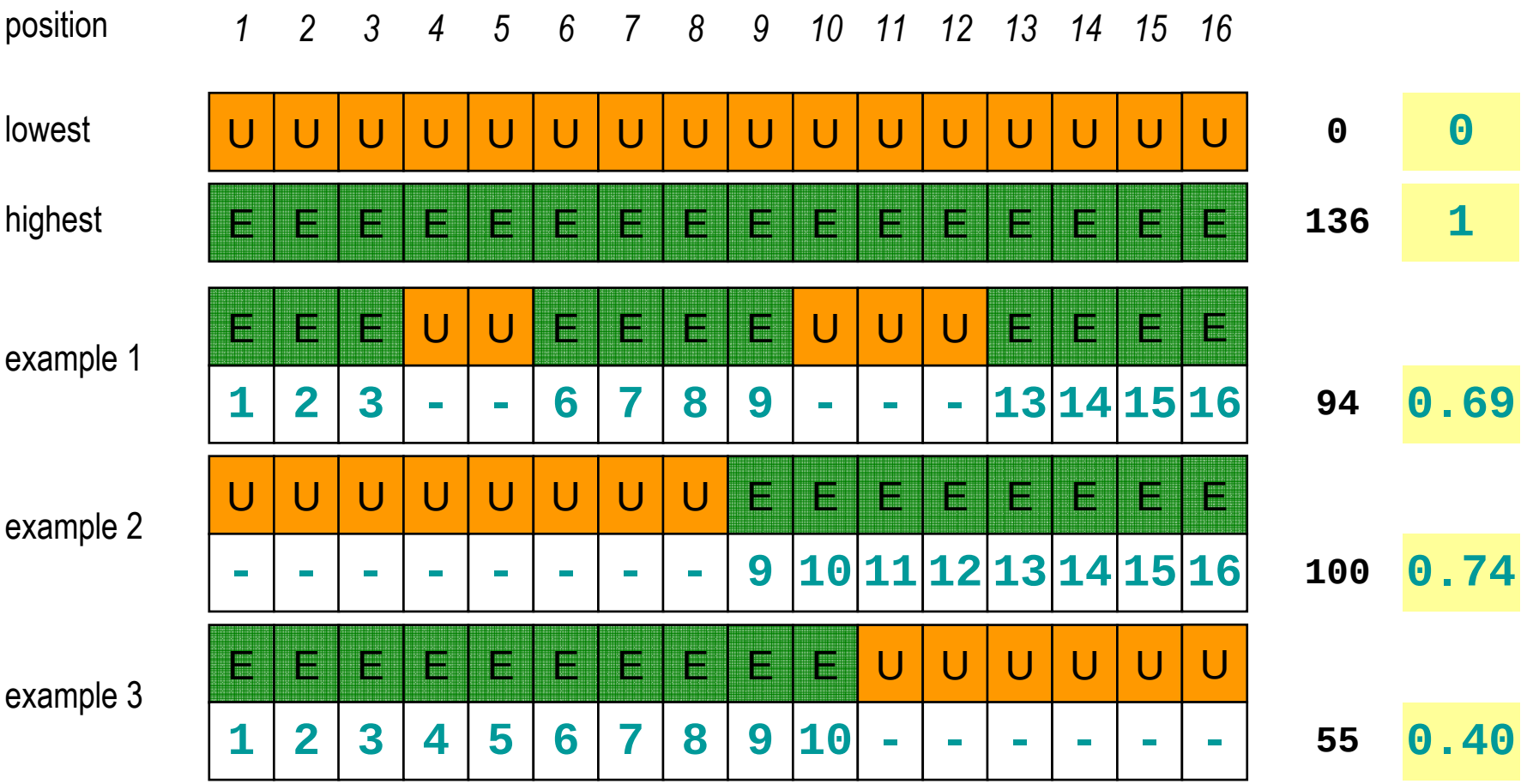


egen-functions: volatility & **integration**

- main ideas:
  - the first transition into employment is not meaningful
  - the longer / more employment episodes, the higher the quality of integration process
  - later employment episodes are better regarding the quality of integration

- calculation:
  - integ = sum of number of sequence positions when status is ‚employed'
  - divided by $\Sigma\{1, ..., 60\} = 1830$
  - range: $0 \leq \text{integ} \leq 1$

egen-functions: volatility & **integration**

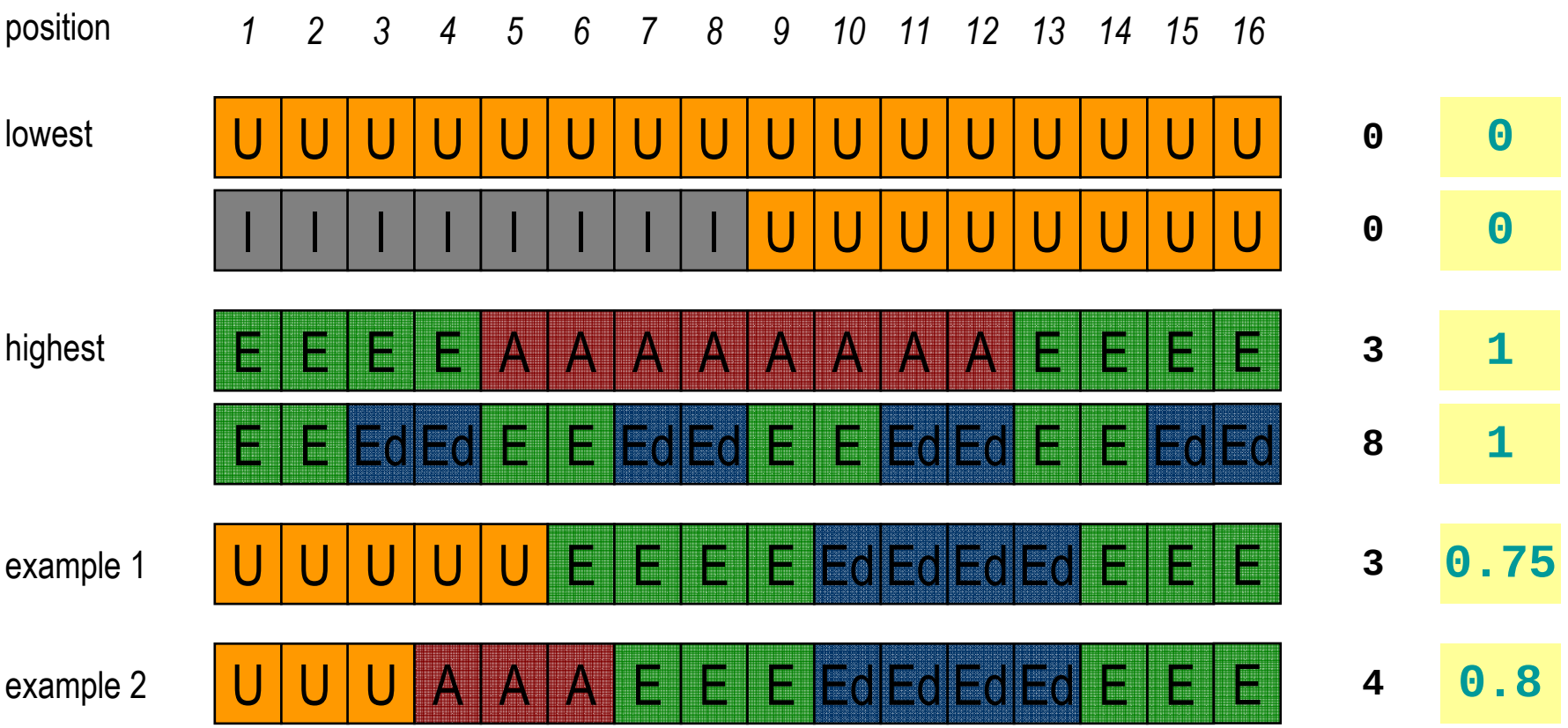


egen-functions: **volatility** & integration

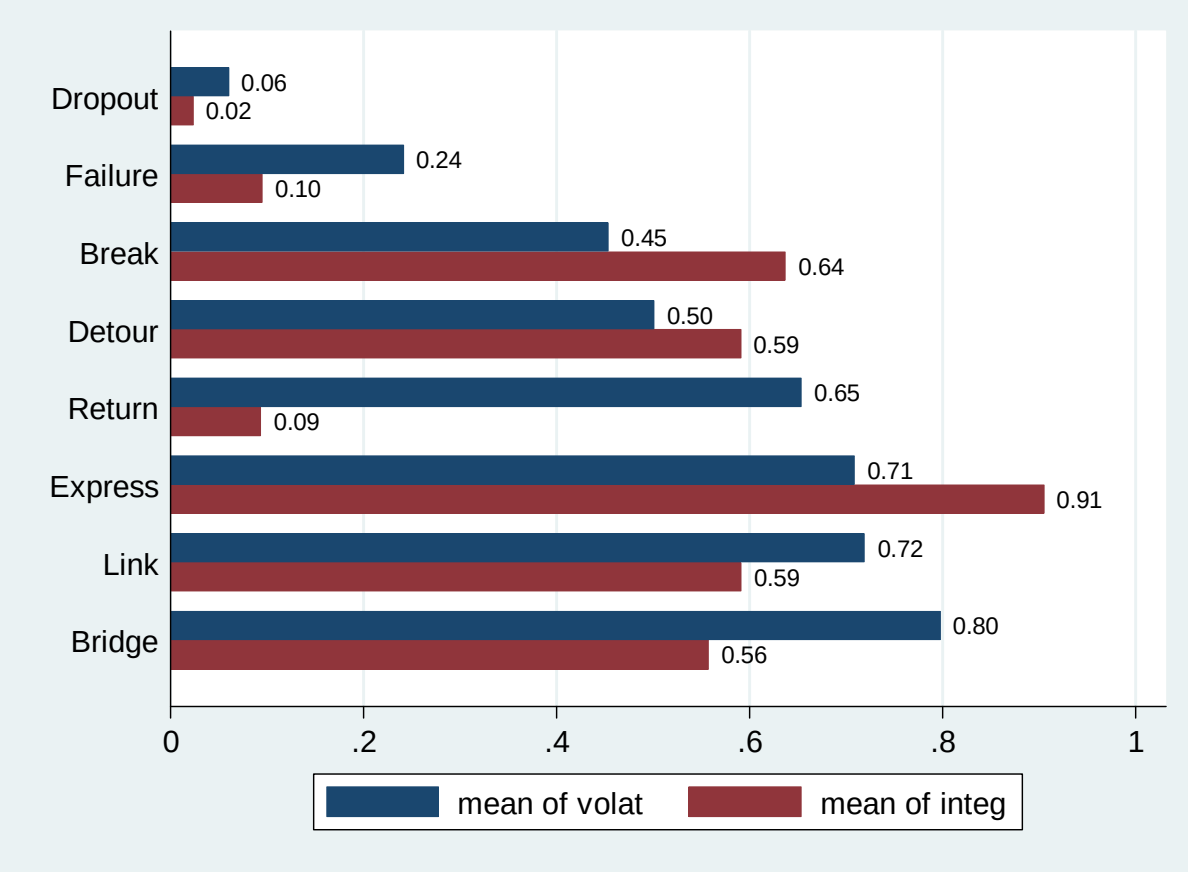
- main idea:
  - not only (full-time) ‚employment' reflects positive statuses, but also ‚education' and ‚apprenticeship'
  - frequent changes between ‚employment' and ‚education' and ‚apprenticeship' = high quality of transition

- calculation:
  - volat = number of episodes with status ‚education', ‚apprenticeship' or ‚employed'
  - divided by number of all episodes
  - range: $0 \leq \text{volat} \leq 1$

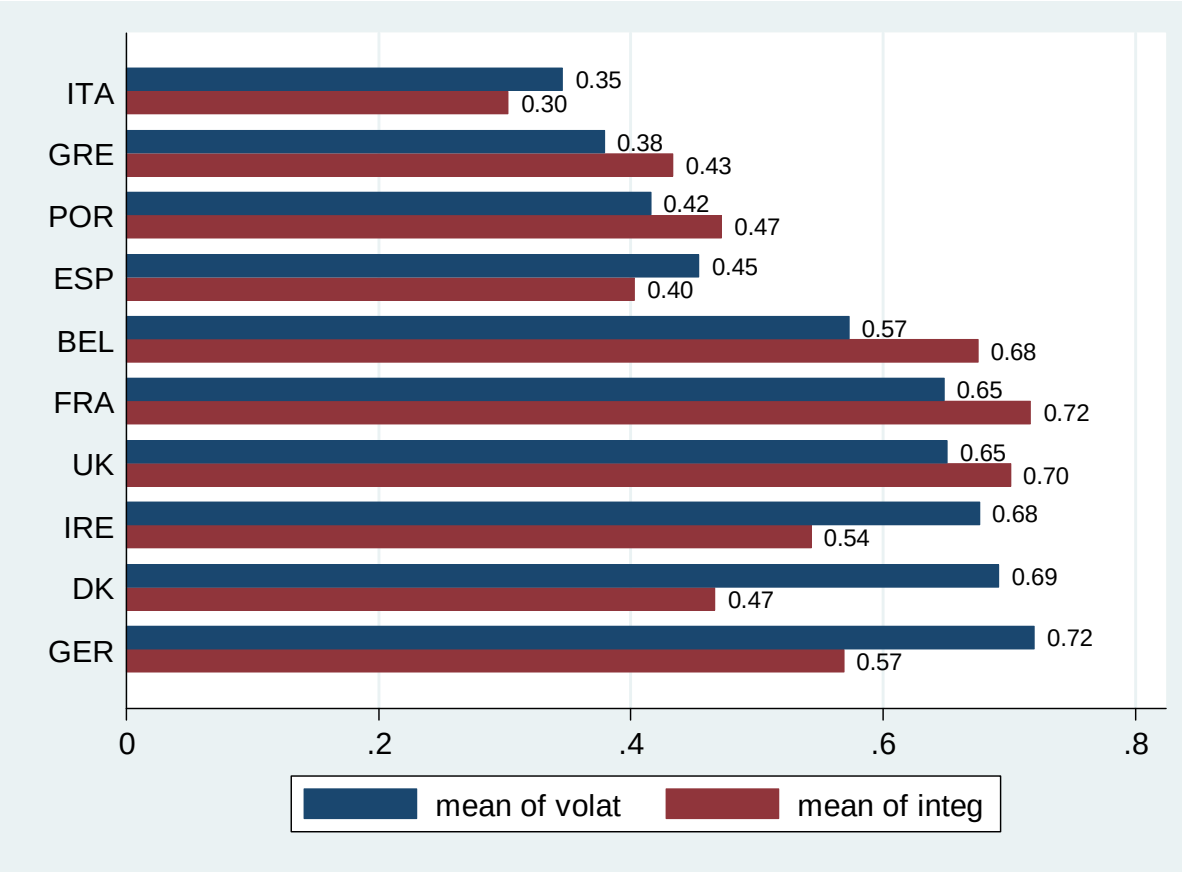
egen-functions: **volatility** & integration



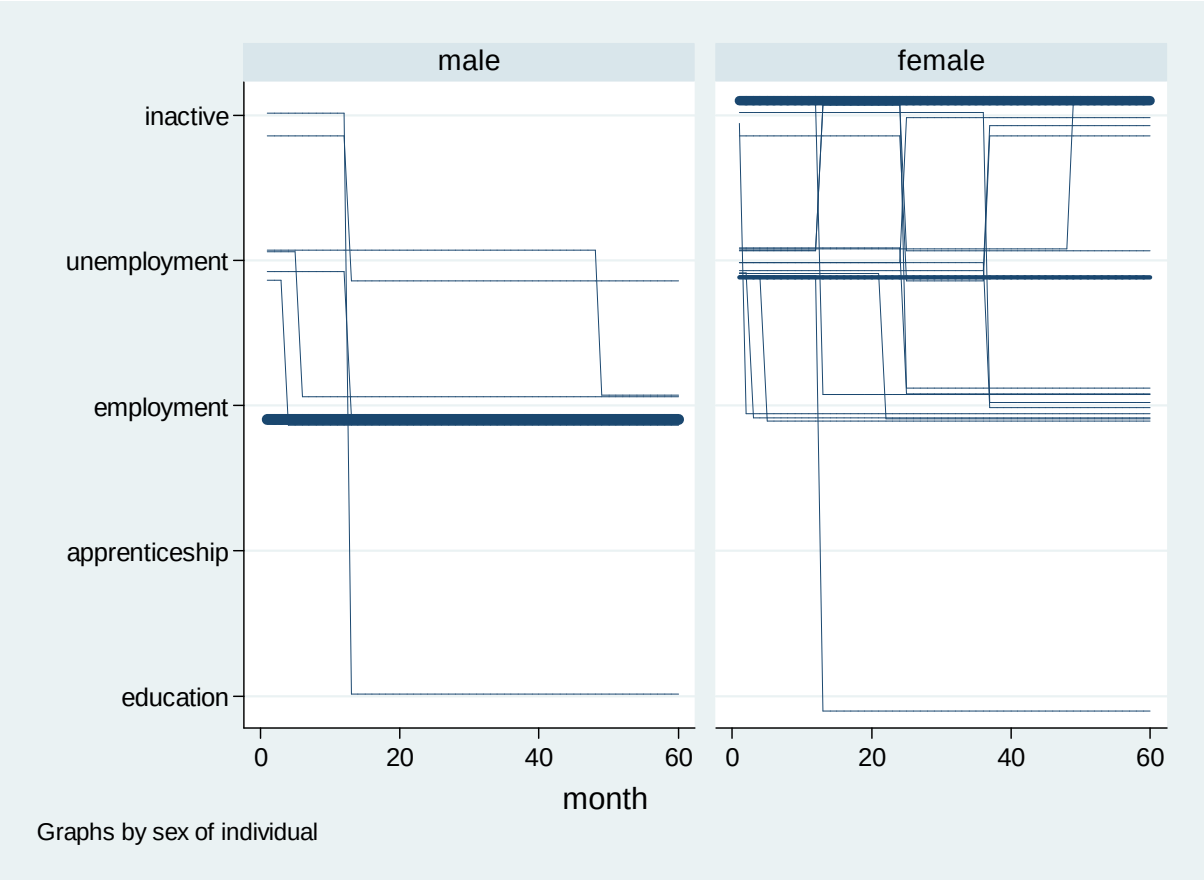
egen-functions:
volatility & integration



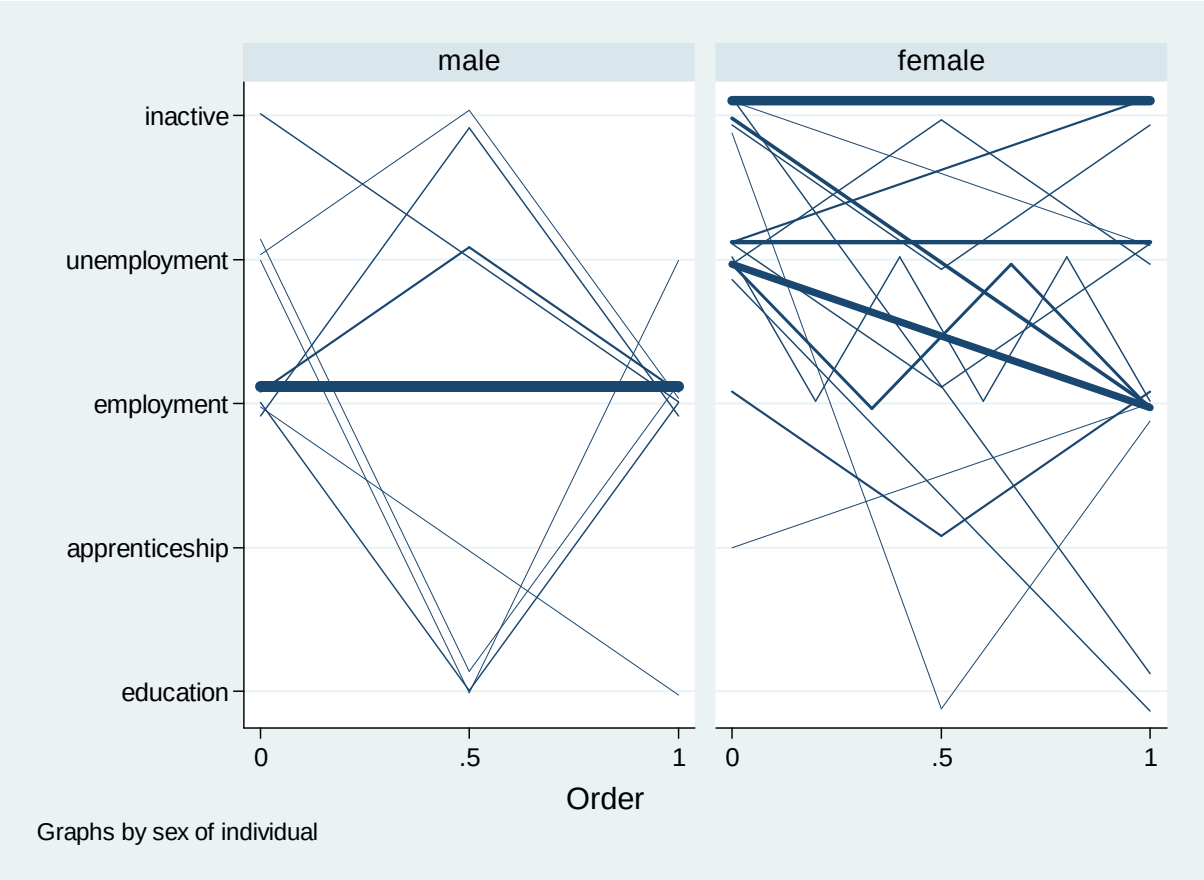
egen-functions:
volatility & integration



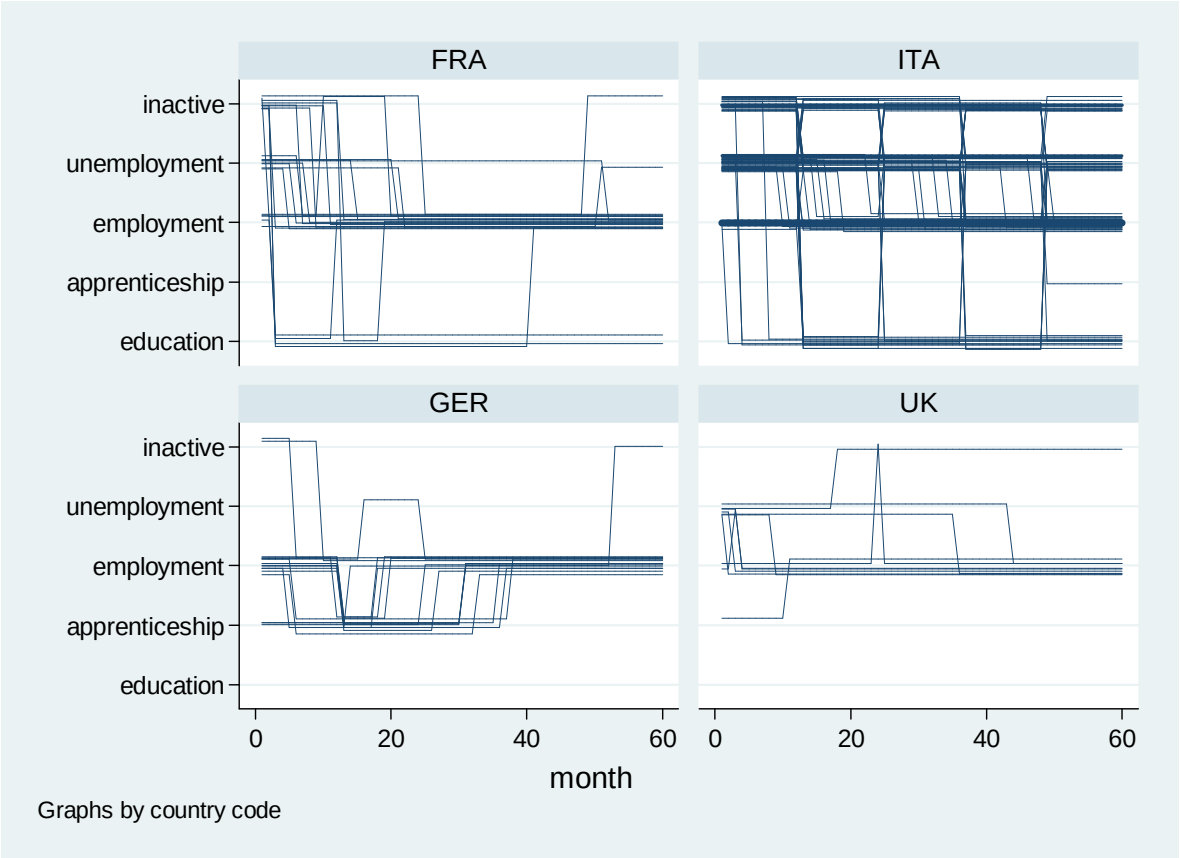
graphical display: parallel coordinates plots



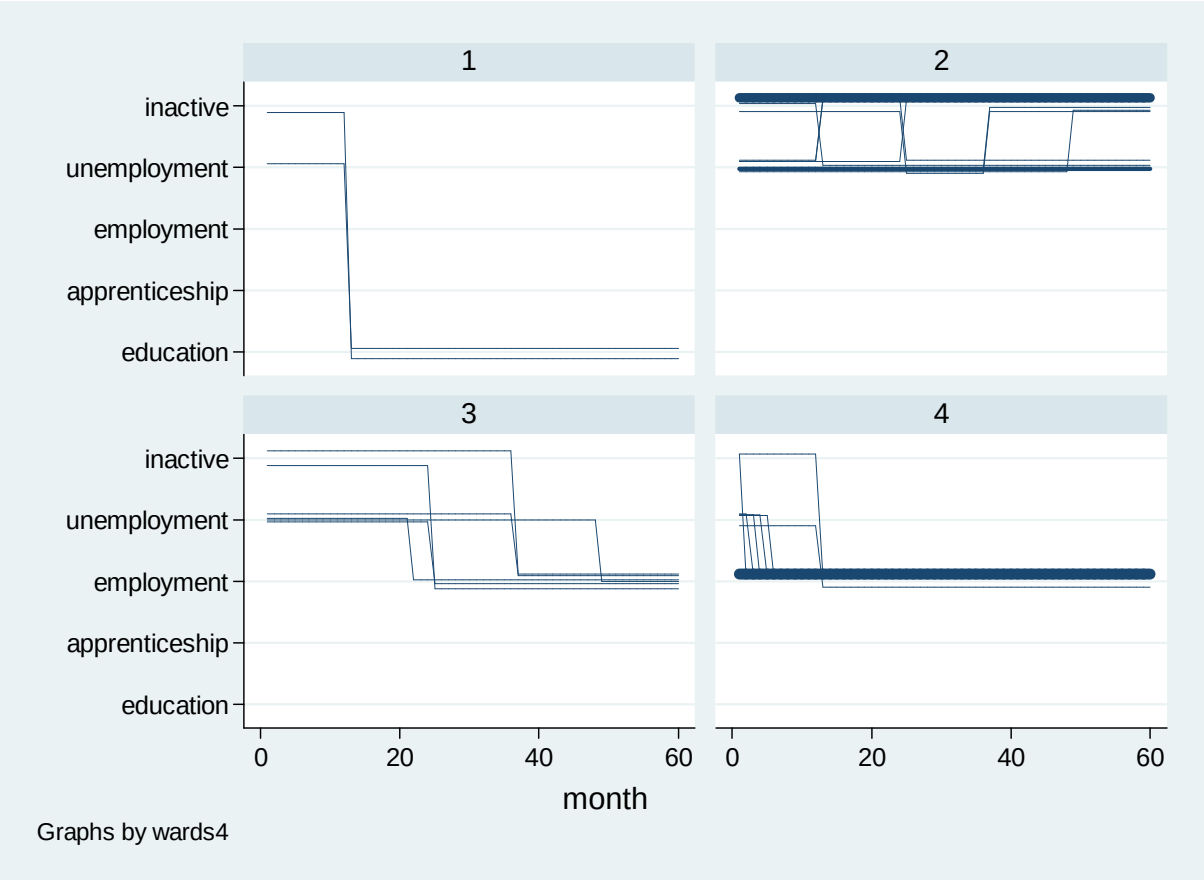
graphical display: parallel coordinates plots



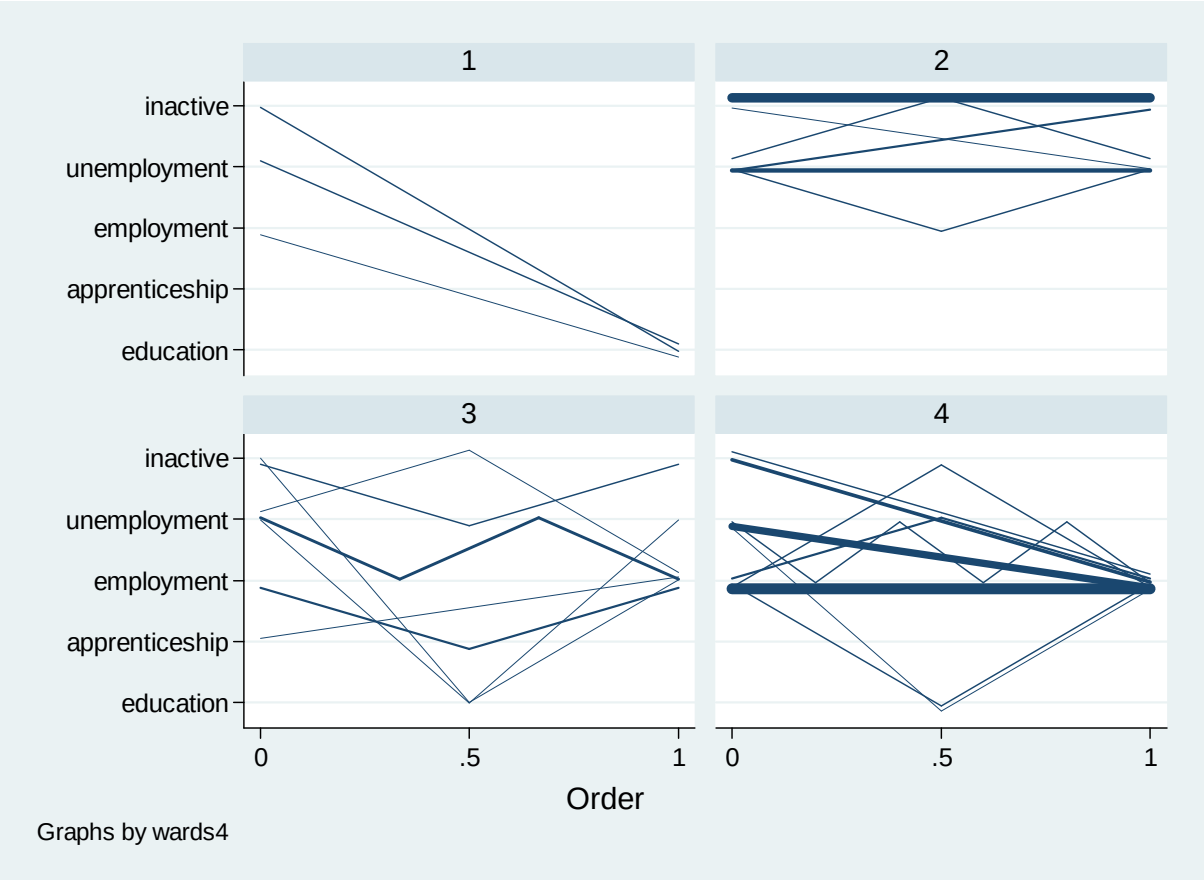
graphical display: parallel coordinates plots



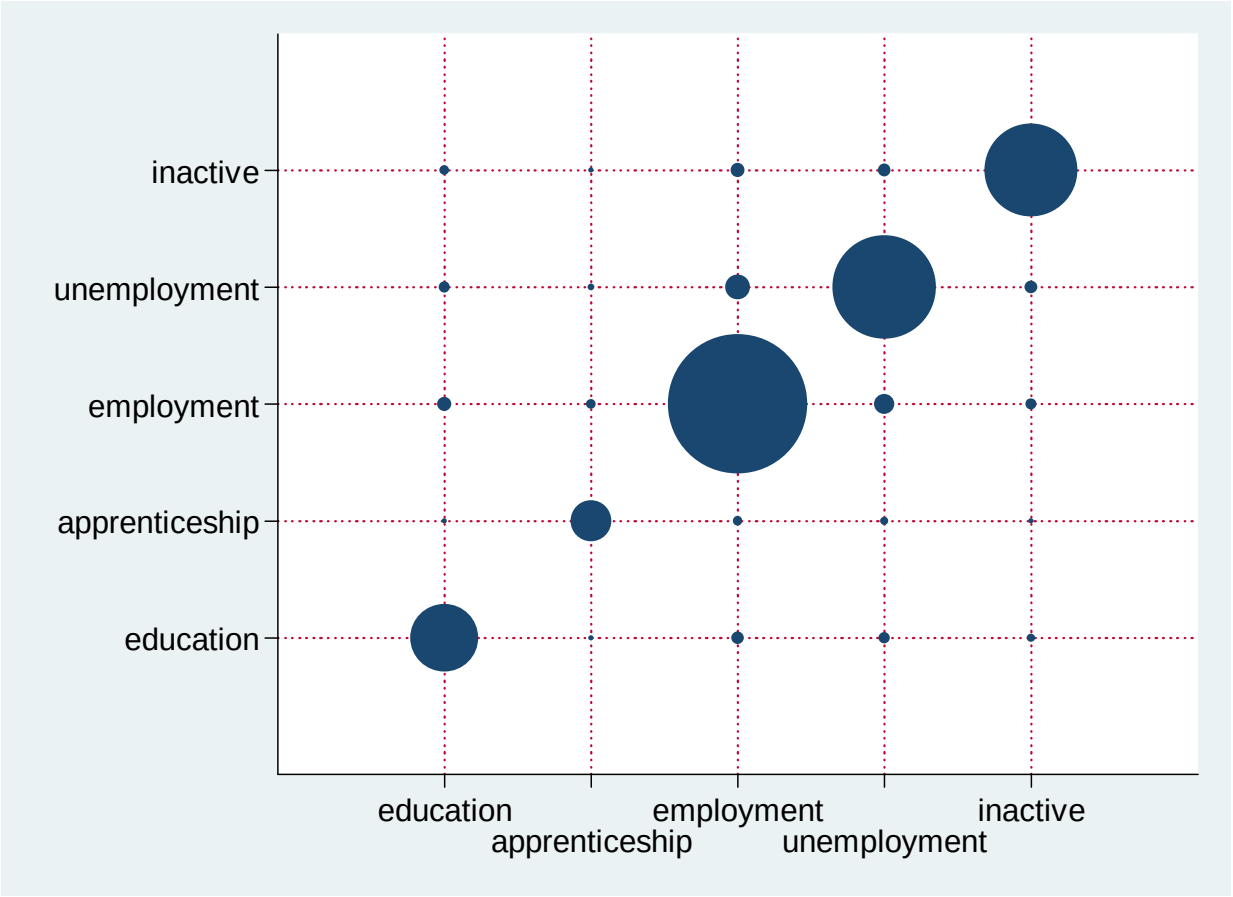
graphical display: parallel coordinates plots



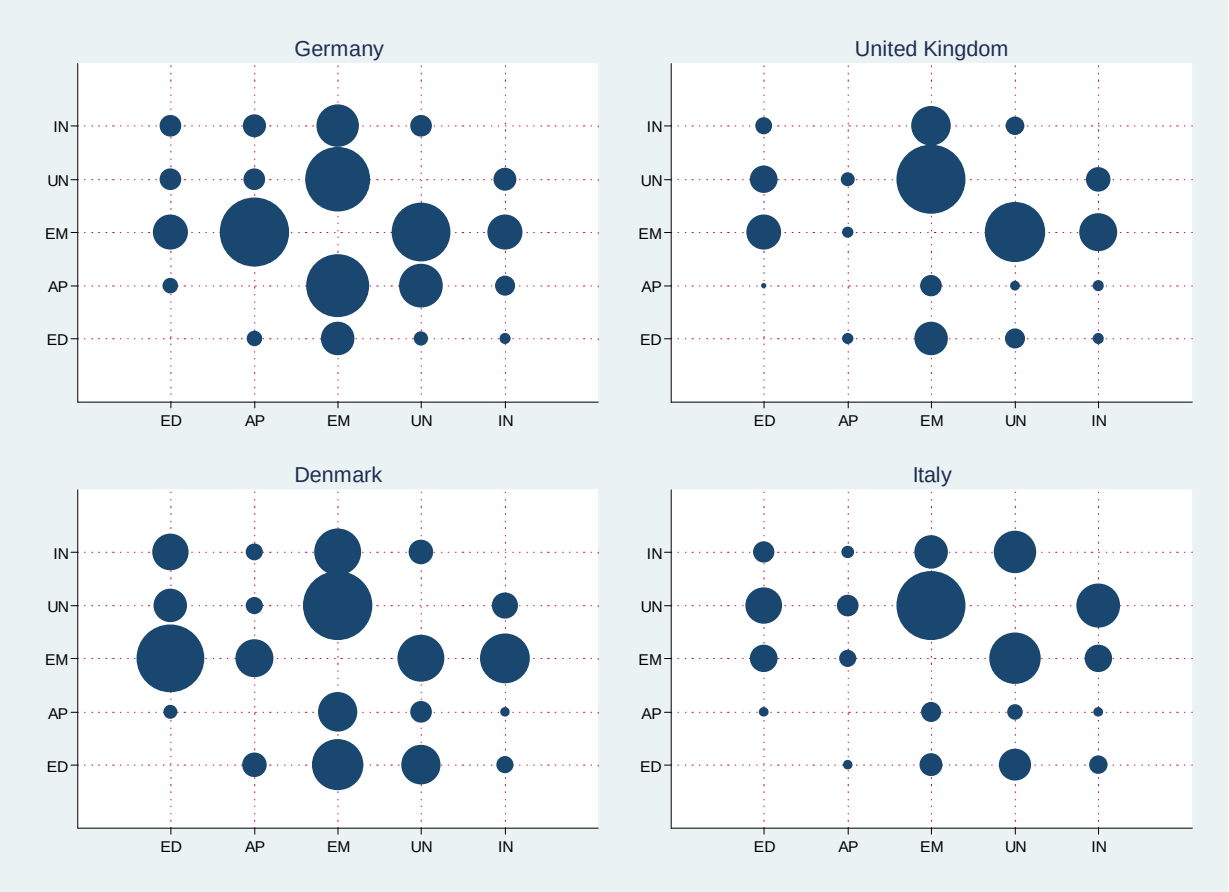
graphical display: parallel coordinates plots



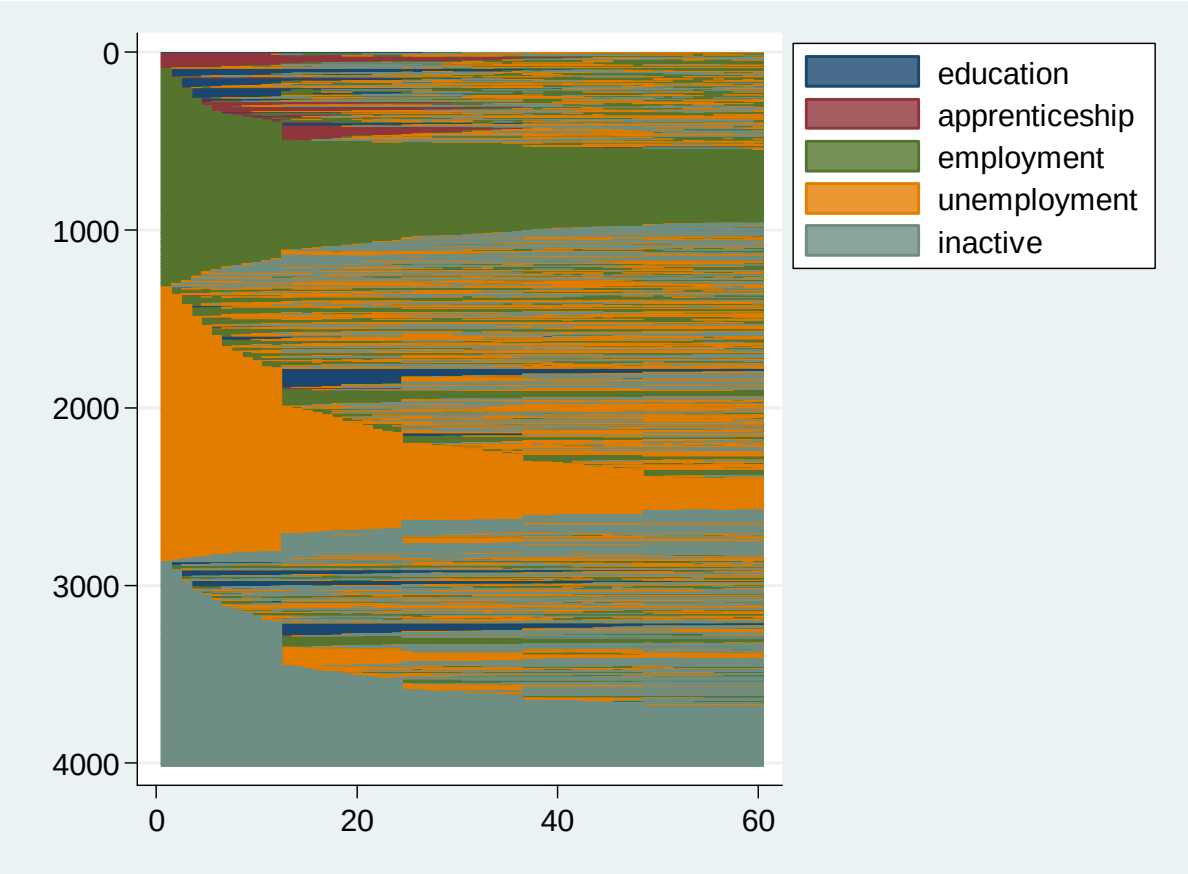
graphical display: transition plots



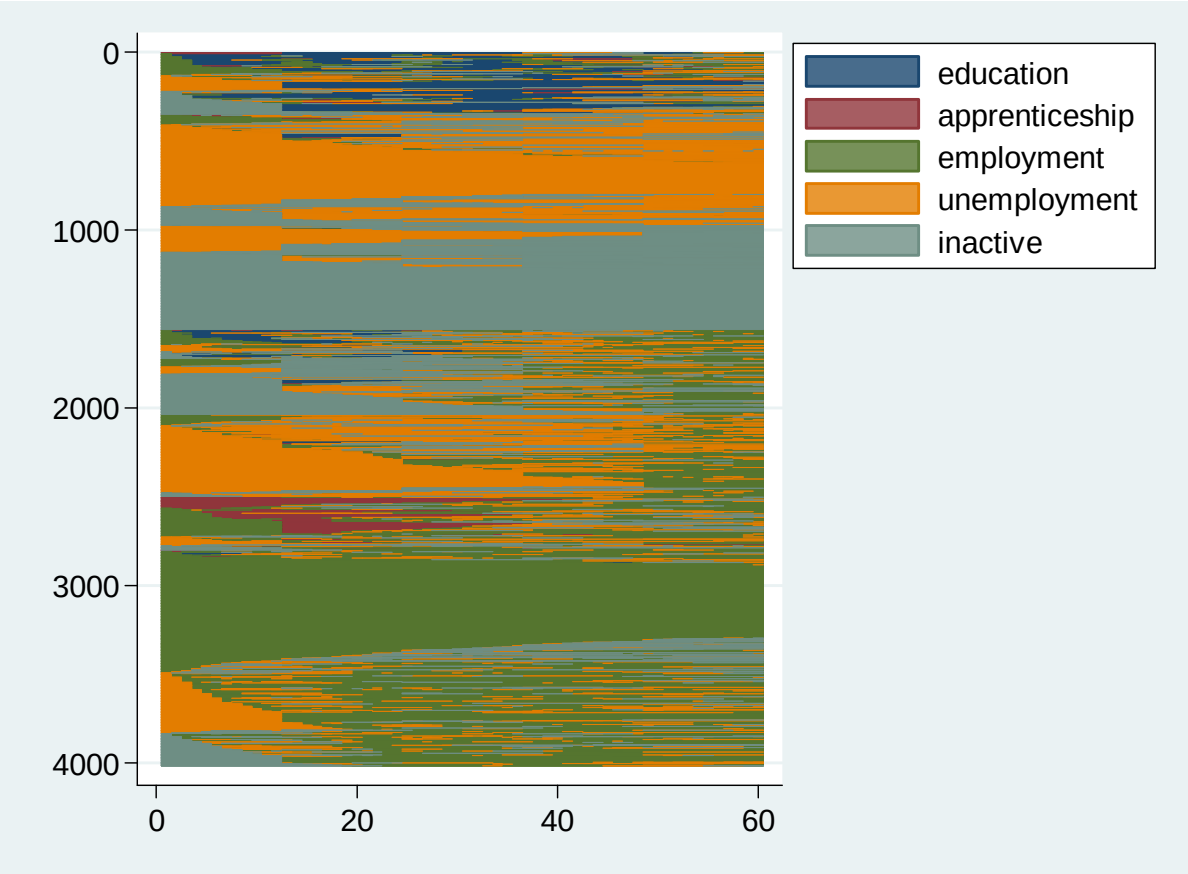
graphical display: transition plots



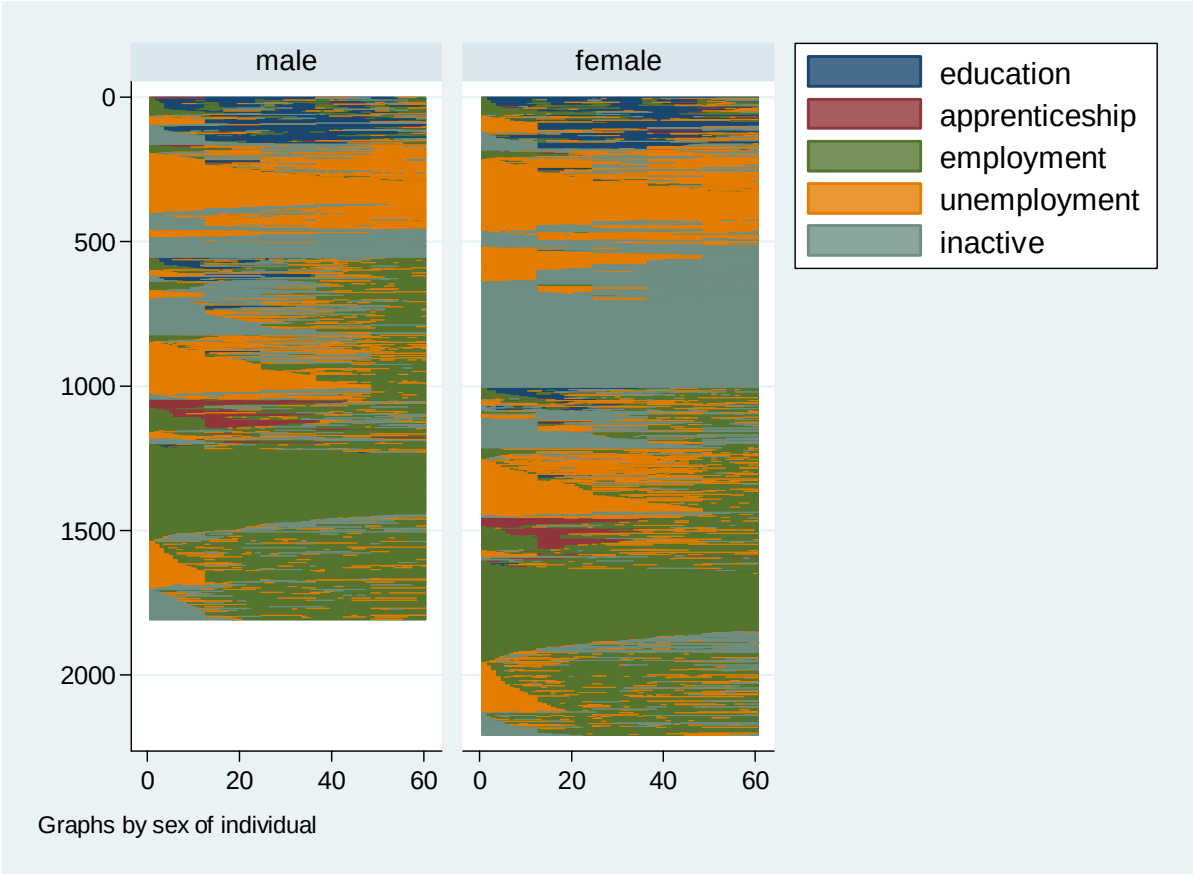
graphical display: sequence index plots – order



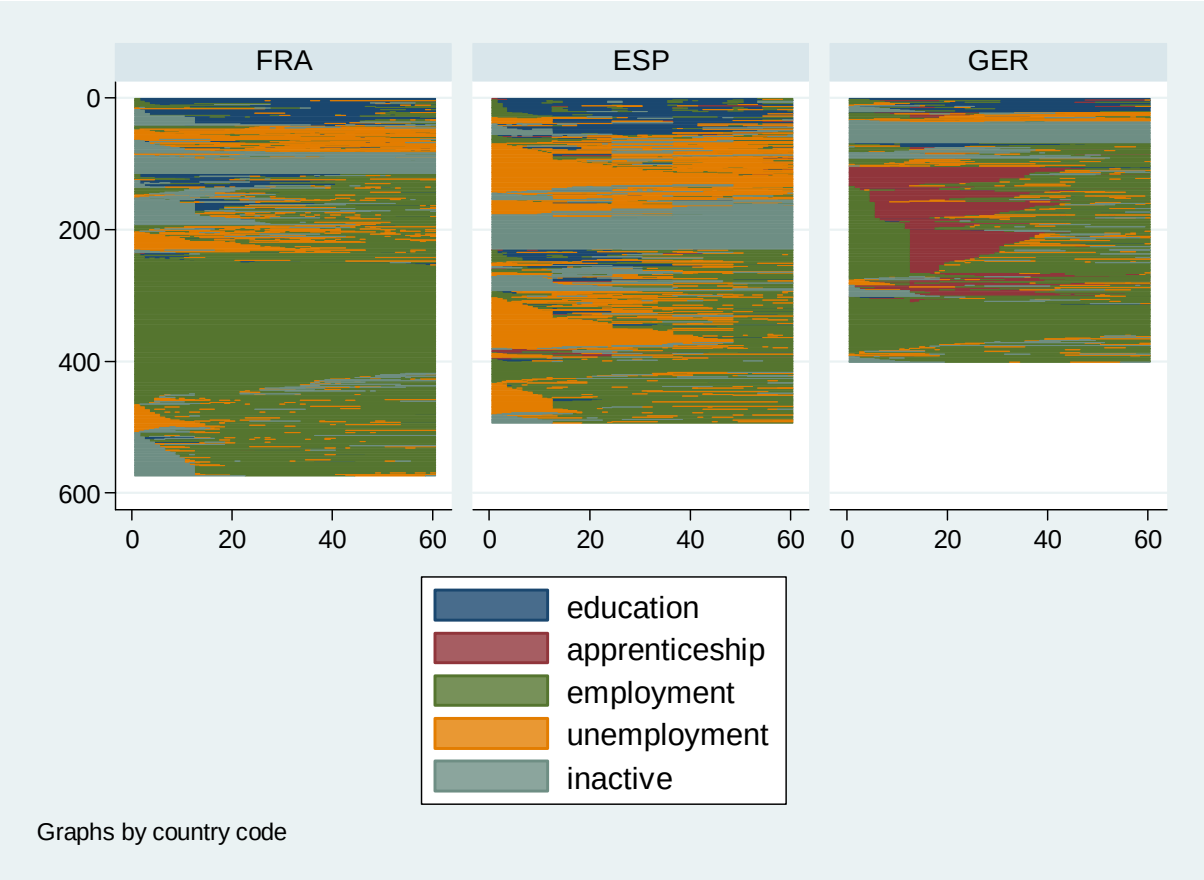
graphical display: sequence index plots – order



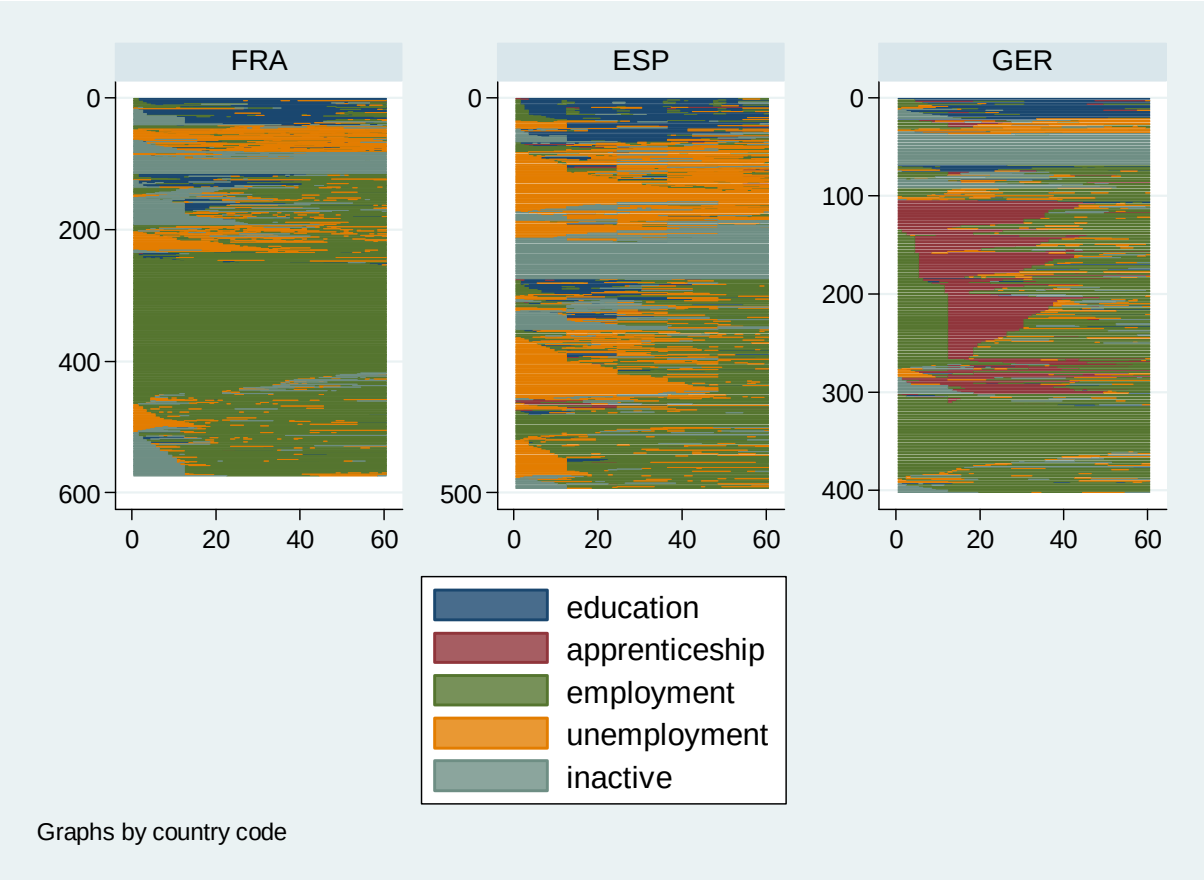
graphical display: sequence index plots – order



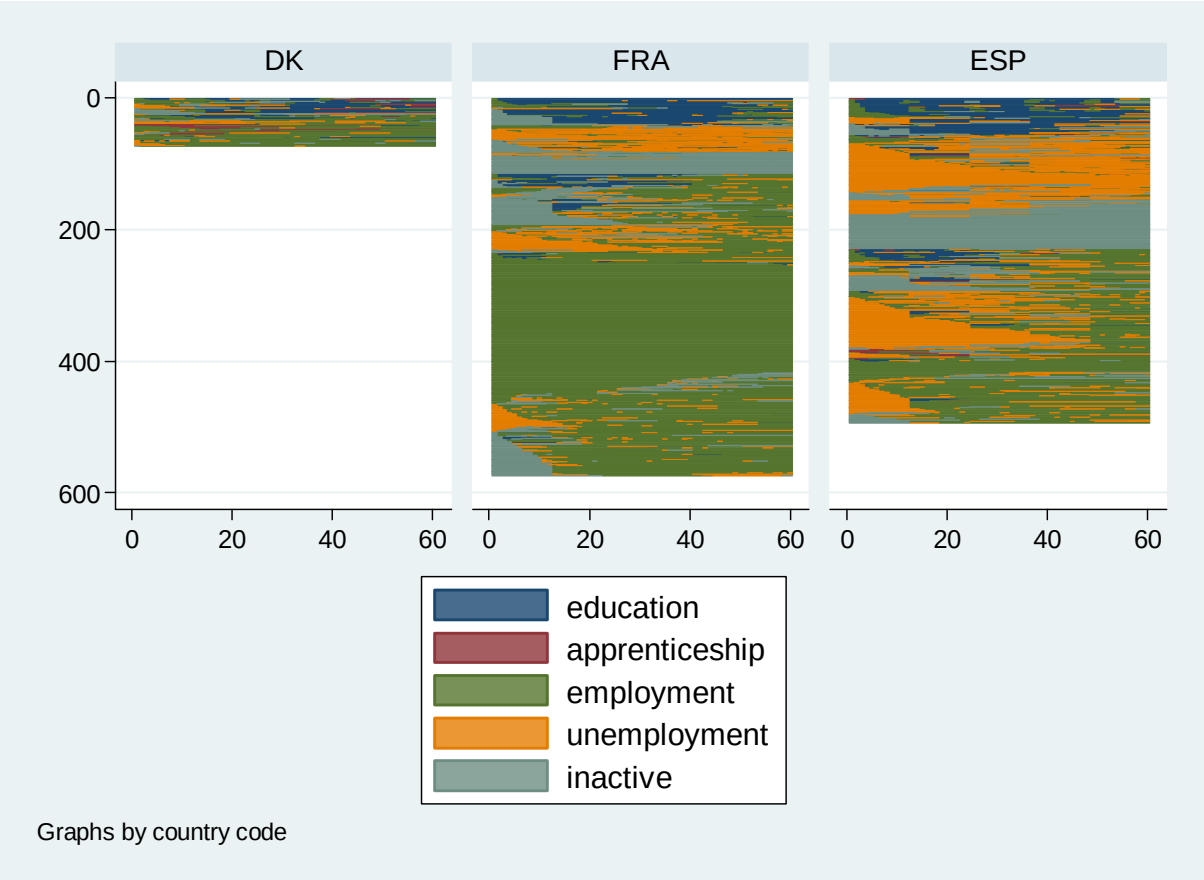
graphical display: sequence index plots – rescaling



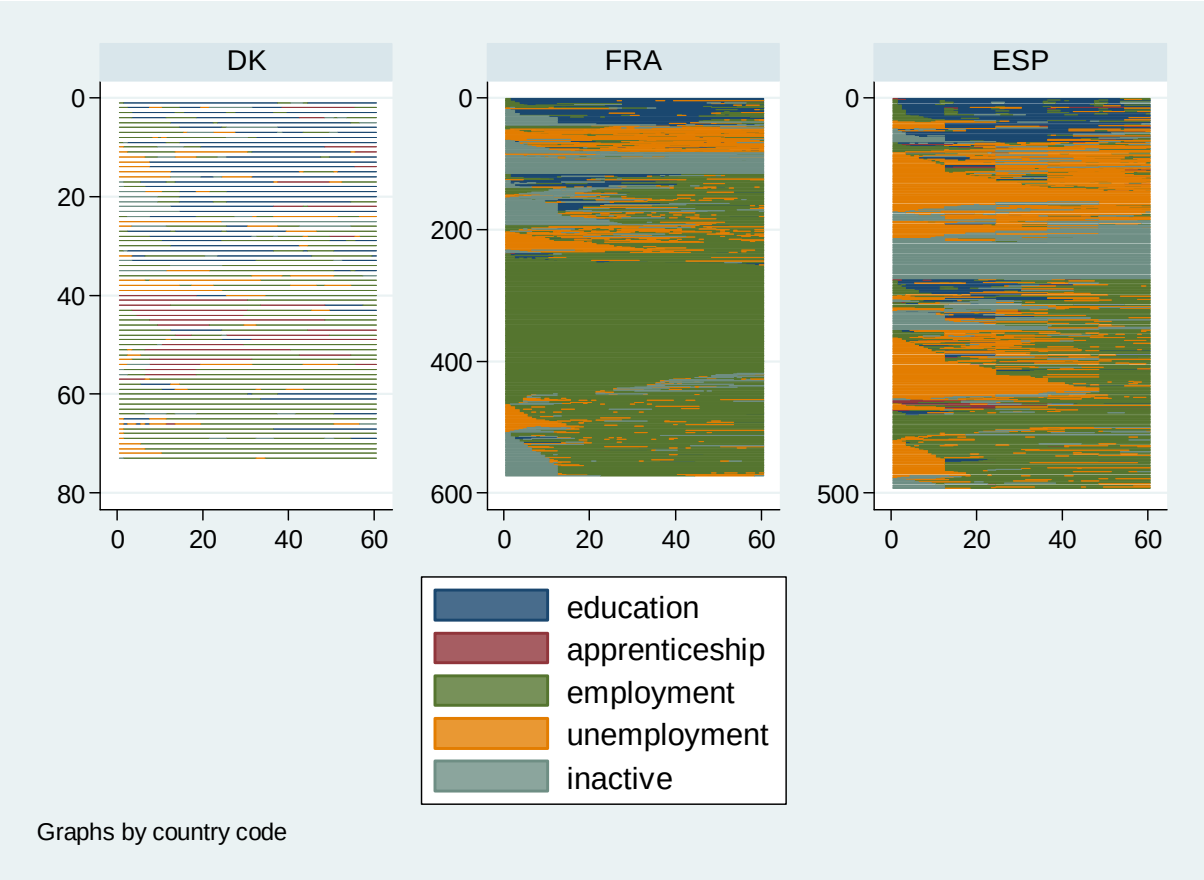
graphical display: sequence index plots – rescaling



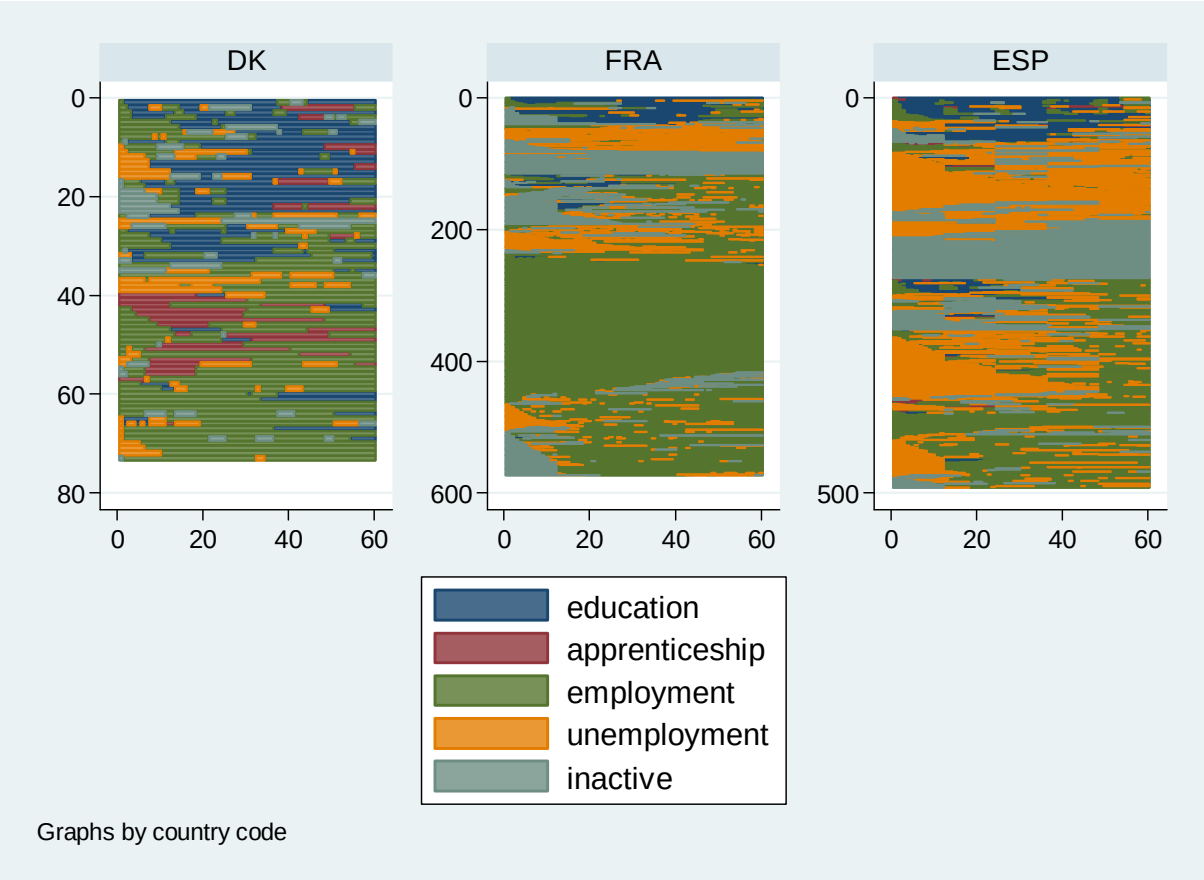
graphical display: sequence index plots - underplotting



graphical display: sequence index plots - underplotting



graphical display: sequence index plots - underplotting

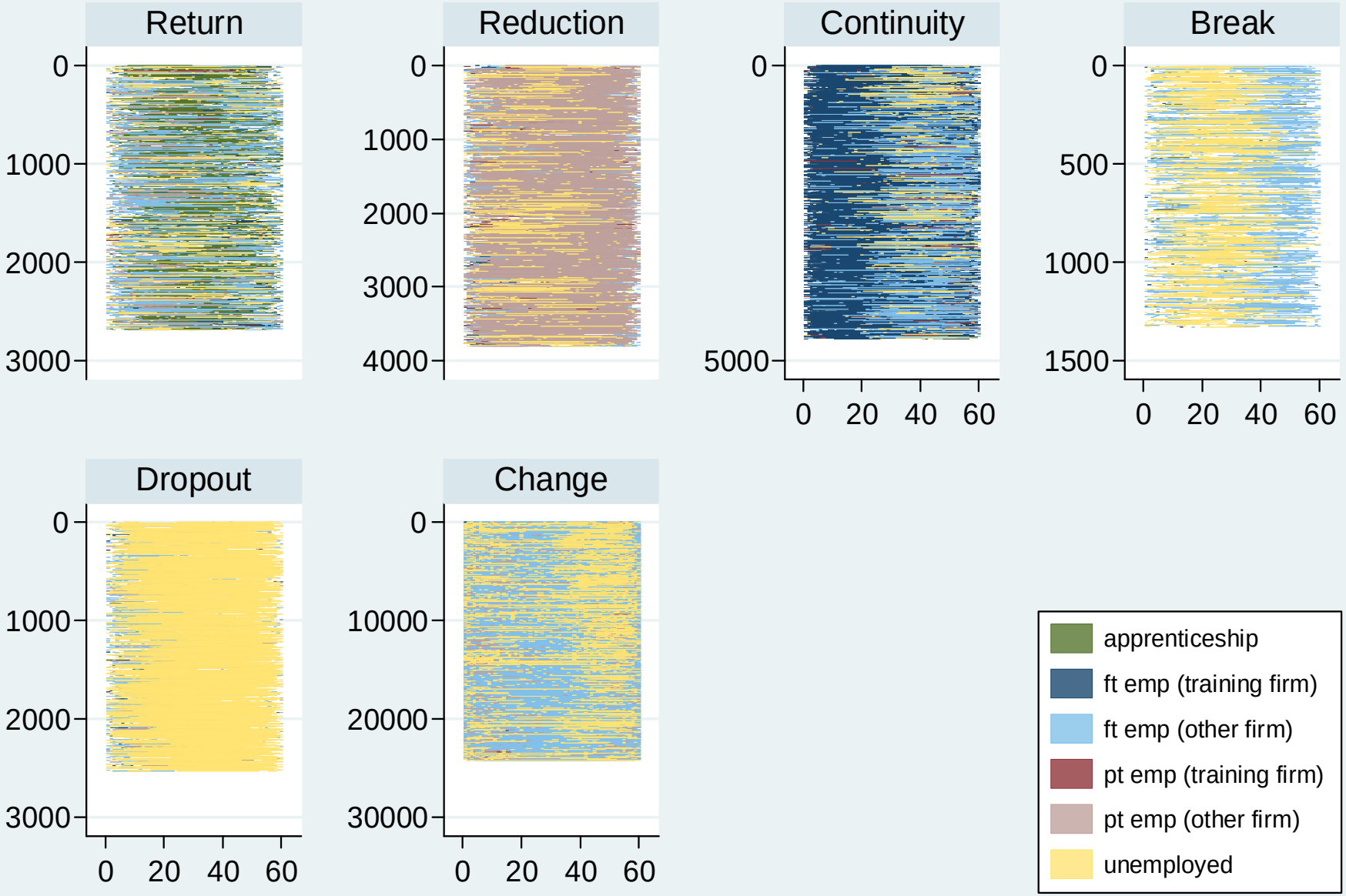


LM entry careers of apprentices

- Data: „Beschäftigten- & Leistungshilfeempfängerhistorik (BLH)“
  - dependent employees (incl. apprentices) since 1975
  - apprenticeship graduates (min. 2 years duration)
- 3 cohorts of apprentices:
  - 1979-1981 & 1989-1991 & 1999-2001
- criteria for cohort selection:
  - similar distance,
  - comparable economic situation (econ. upturn)
- case number: ~ 40.000 individuals
- limitations: only West Germany, no school-based voational education

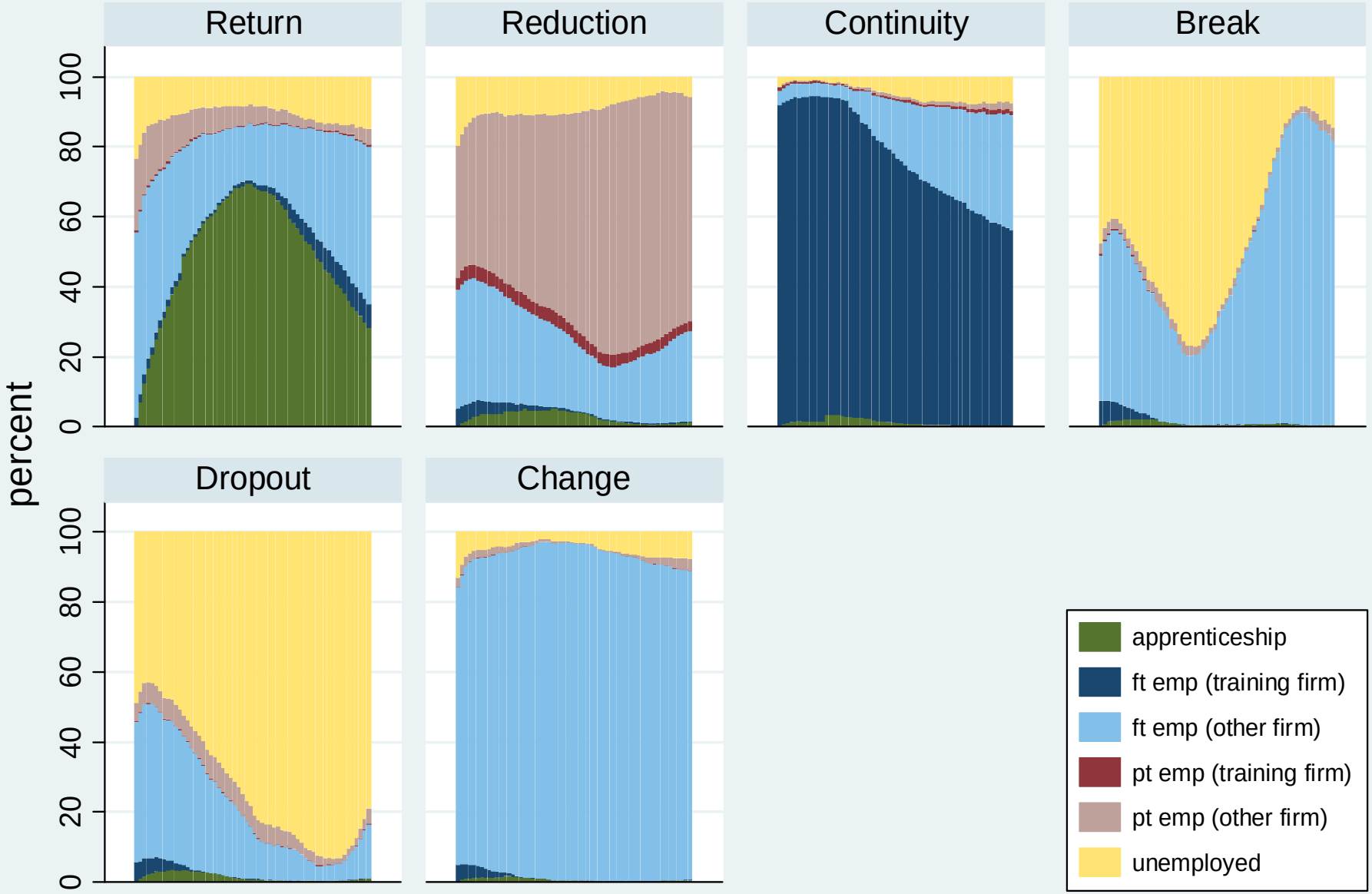
Sequence index plots by cluster

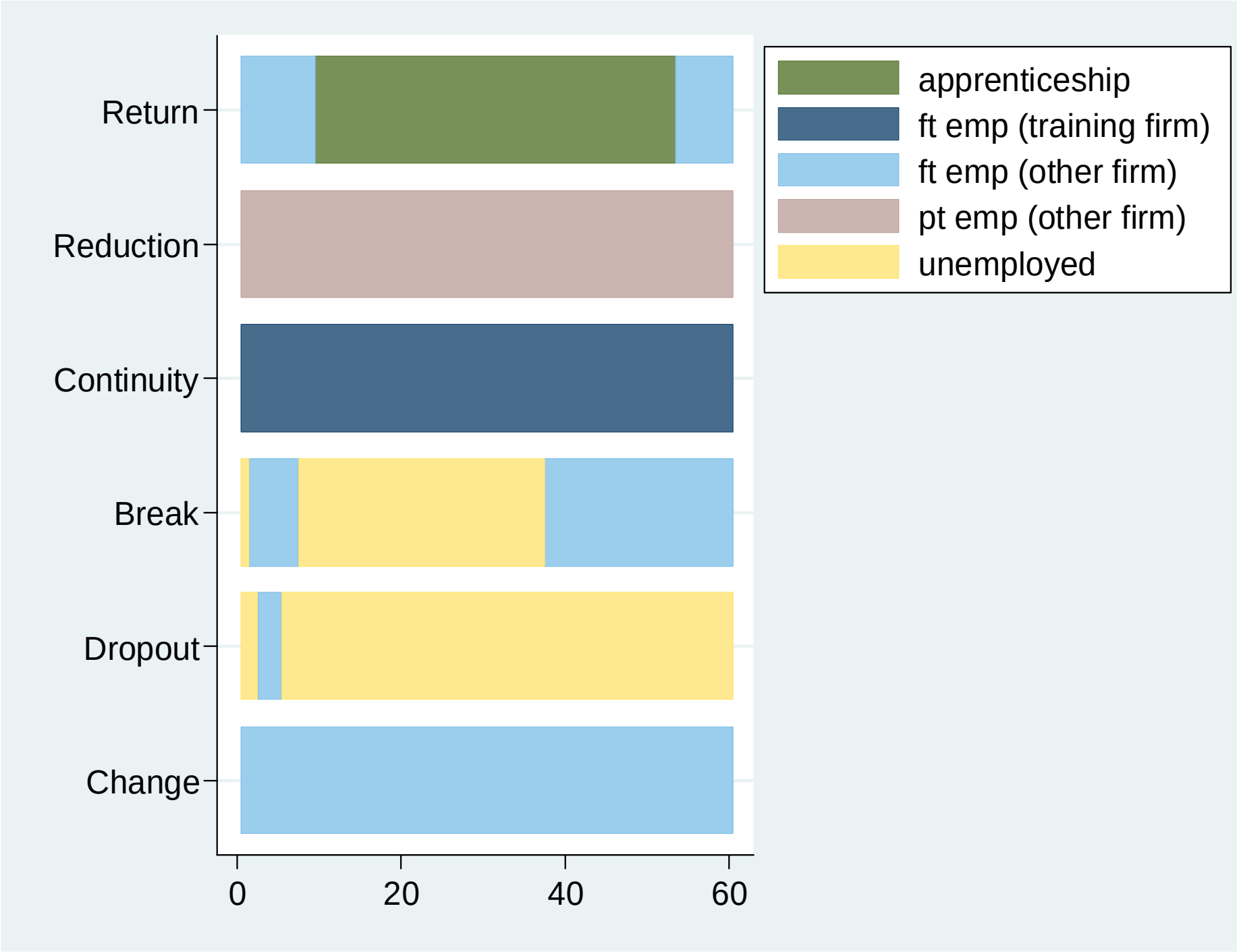
Source: BLH, own calculations



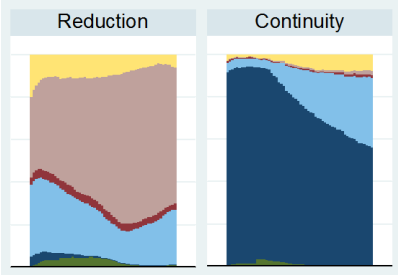
Monthly proportion of statuses by cluster

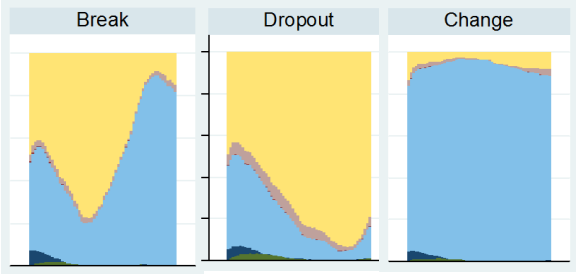
Source: BLH




Return
Reduction
Continuity
Break
Dropout
Change
0
20
40
60
apprenticeship
ft emp (training firm)
ft emp (other firm)
pt emp (other firm)
unemployed

Ergebnisse





cluster	(Ref.: Return)	Reduction	Continuity	Break	Dropout	Change
sex (Ref.: male)						
	female	1.81***	0.83**	0.51***		
nation (Ref.: Germany)						
	Turkey		0.71***			
	Ex-Yugoslavia		0.77*	0.58**		
	ITA/GRC/ESP/POR		0.79*			
	EU/USA/CAN					
	remaining Europe					
	other		0.45***			0.68***
cohort (Ref.: 1979-1981)						
	1989-1991	0.81**	0.79***	0.52***	0.78***	
	1999-2001	2.70***	0.42***	0.60***	0.54***	0.49***
occupation (Ref.: agriculture)						
	miners					17.90***
	manufacturing occupations	0.37***	2.58***		1.36***	10.16***
	technical occupations	0.62**	3.38***			10.82***
	service occupations	0.70***	2.86***	0.71***		11.00***
	other occupations	0.70***	2.55***	0.70*		9.66***
settlement (Ref.: agglomerated areas)						
	urbanised areas		1.31***		1.23***	1.30***
	rural areas			1.31*	1.37***	1.38***

Outlook

1. integration of Halpin‘s and Lesnard‘s plugins into ado command structure
2. implementation of further metrics
3. specific commands for multi-channel / multiple sequence analysis
4. depiction of alignment of single sequences
5. more ideas?

!!! Thank you very much for your attention !!!

☺