

Introduction to the German Socio-Economic Panel (SOEP)

Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

What is Panel Data?

- Panel Data: Repeated measurements of many units at fixed time points

Types of Data

A) Cross-Section Data

	2000	2001	2002	2003	2004	2005	2006	2007
Unit 1				X				
Unit 2				X				
Unit 3				X				

B) Time-Series Data

	2000	2001	2002	2003	2004	2005	2006	2007
Unit 1	X	X	X	X	X	X	X	X
Unit 2								
Unit 3								

C) Panel Data

	2000	2001	2002	2003	2004	2005	2006	2007
Unit 1	X	X	X	X	X	X	X	X
Unit 2	X	X	X	X	X	X	X	X
Unit 3	X	X	X	X	X	X	X	X

D) Pooled Cross-Section Data

	2000	2001	2002	2003	2004	2005	2006	2007
Unit 1				X				
Unit 2					X			
Unit 3						X		

Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

Advantages of panel data

- ⇒ Match information from different phases of the life course
- ⇒ Control for time-constant characteristics (-> validate causal interpretations)
- ⇒ Model sequentiality (-> validate causal interpretations)
- ⇒ (Measure trends)
- ⇒ Decomposition of trends

Potential Drawbacks

- ⇒ Bias due to selective attrition
- ⇒ Panel effects
- ⇒ Continuity of questioning
(phrasing of questions, modes, ...)

Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

SOEP Basics

What is the "SOEP" ?

- ⇒ The SOEP was started in 1984: Now 29 Waves available!
- ⇒ Longest-running longitudinal survey of private households and persons in Germany
- ⇒ Started with 6,000 Households in 1984, in 2012 approx. 12,000
- ⇒ Representative micro-data on persons, households, families
- ⇒ Objective (e.g. income) and Subjective (e.g. satisfaction) indicators
- ⇒ Topics in Economics, Sociology, Political Science, Psychology, Geography
- ⇒ Collects information about all individuals within a given household at the time of the interview
- ⇒ Follows individuals over time (even if they move; form their own household)

2. Content

Core Questions

- ⇒ Population and demography
- ⇒ Education, training, and qualification
- ⇒ Labor market and occupational dynamics
- ⇒ Earnings, income and social security
- ⇒ Housing
- ⇒ Health
- ⇒ Basic orientation (preferences, values, etc.)
- ⇒ Satisfaction with life in general and various aspects

Topic Modules → Replication every ~ 5-10 years

1984	Employment biography	2000	Further education or training
1985	Marriage and family biography	2001	Social networks, working cond.
1986	Social origins (Bio), neighborh.	2002	<i>Social Security, assets</i>
1987	<i>Social security, early retirement</i>	2003	Ecology and envir. behavior
1988	<i>Assets</i>	2004	Further education or training
1989	Further education or training	2005	Time use and preferences
1990	Time use and preferences	2006	Family and social networks
1991	Family and social services	2007	<i>Social Security, assets</i>
1992	<i>Social security and poverty</i>	2008	Further education or training
1993	Further education or training	2009	Social networks, working cond.
1994	Neighborhood, values	2010	Consumption, savings, health
1995	Time use and preferences	2011	Family and social networks
1996	Family and social network	2012	<i>Social Security, assets</i>
1997	<i>Social security and poverty</i>		
1998	Ecology and envir. behavior		
1999	Expectations, Use of time		

Challenge: Observing the „*unobservable*“

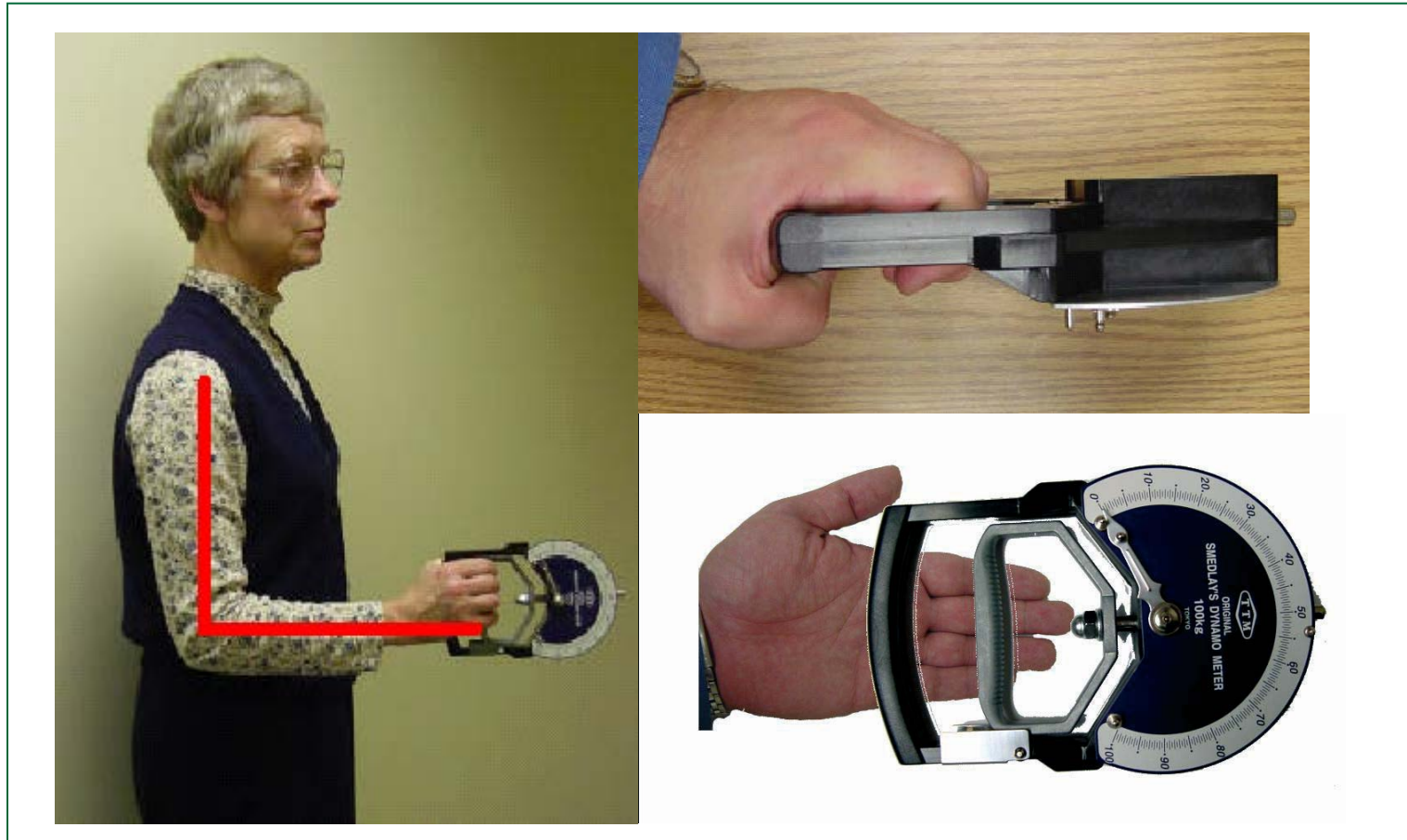
... and going beyond standard interview techniques

- Evaluation of health
(by taking grip strength)
- Evaluation of cognitive abilities
- Personal traits
(Big Five, risk aversion, trust, fairness, ...)

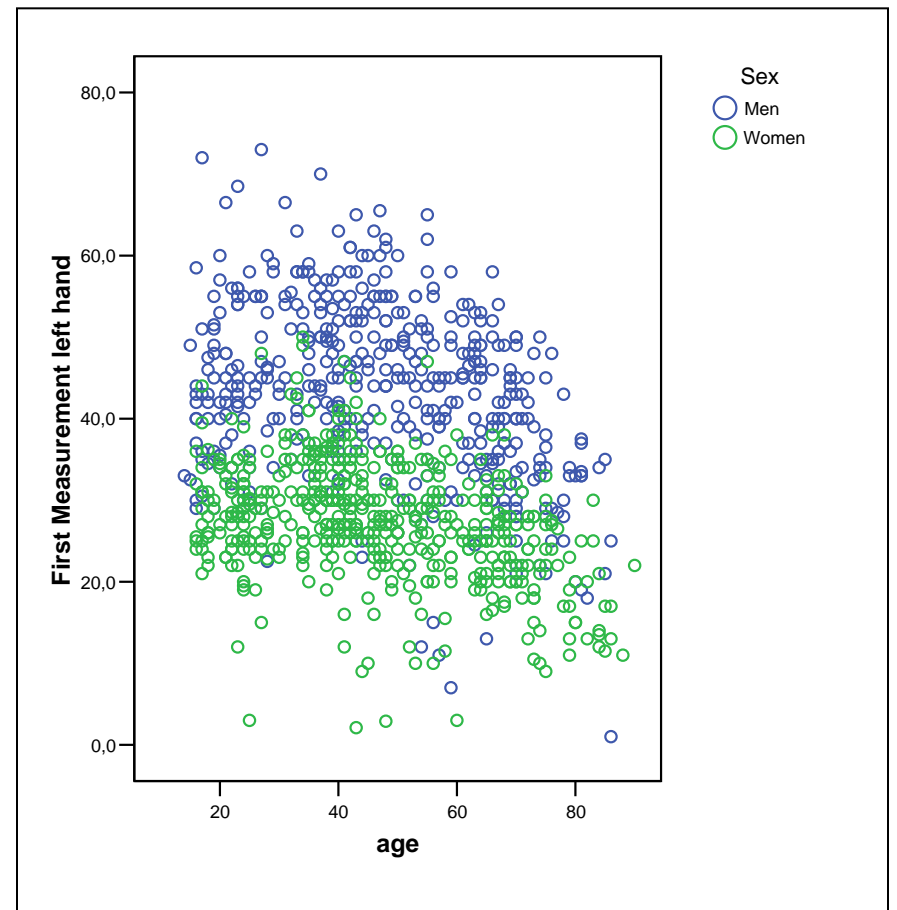
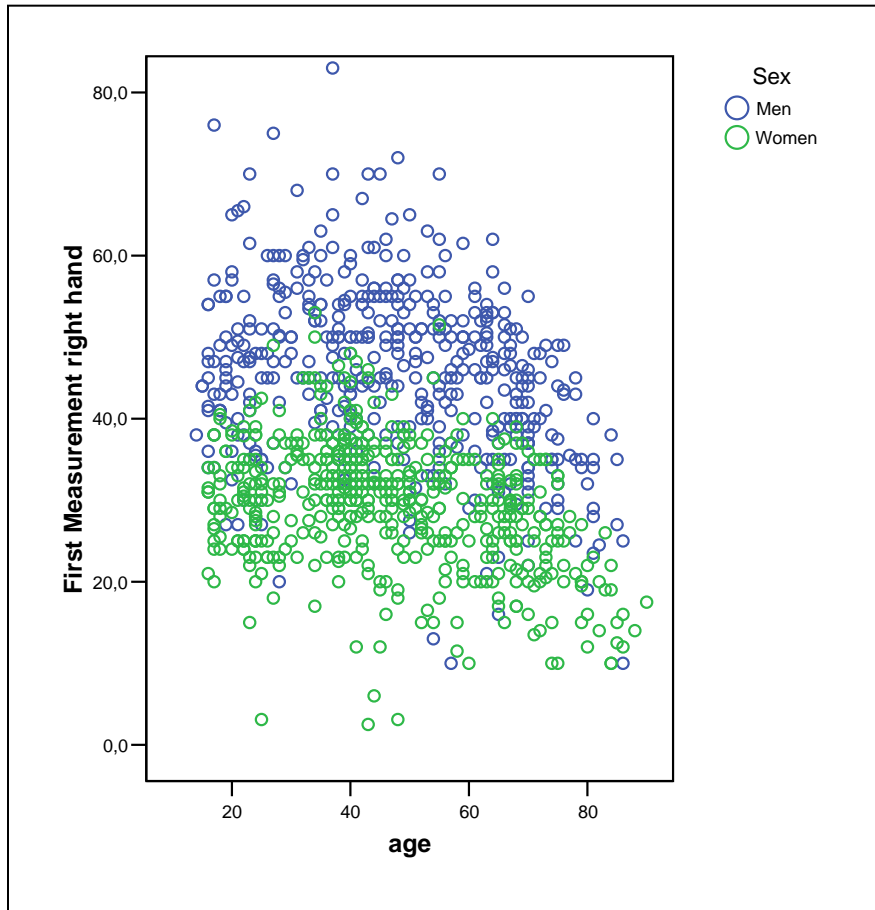
Health (innovations since 2002)

- ✓ Standard questions since 1984 (self-rated health status, disability status, sick days, # nights in hospital, ...)
- ✓ Improved health measures since 2002
 - ✓ SOEP-version of SF12 (physical and mental health status),
 - ✓ Self assessed Height, Weight, BMI in panel perspective (2004, 2006, 2008, 2010, 2012)
 - ✓ Smoking behavior (since 2002, 2004, 2006, 2008, 2010, 2012)
 - ✓ Alcohol consumption (2006, 2008, 2010, 2012)
 - ✓ Grip strength (2006, 2008, 2010, 2012) – see also SHARE / HRS

Measuring Grip Strength in SOEP



Results of the SOEP-Pretest Grip Strength by Age and Sex



Challenge: Observing the „*unobservable*“

... and going beyond standard interview techniques

- Evaluation of health
(by taking grip strength)
- Evaluation of cognitive abilities
- Personal traits
(Big Five, risk aversion, trust, fairness, ...)

Cognitive Abilities (since 2006)

- 2 Ultra short tests (90 sec) for adult respondents
- Full ability measures (about 30 min) for teenagers, only

Paper:

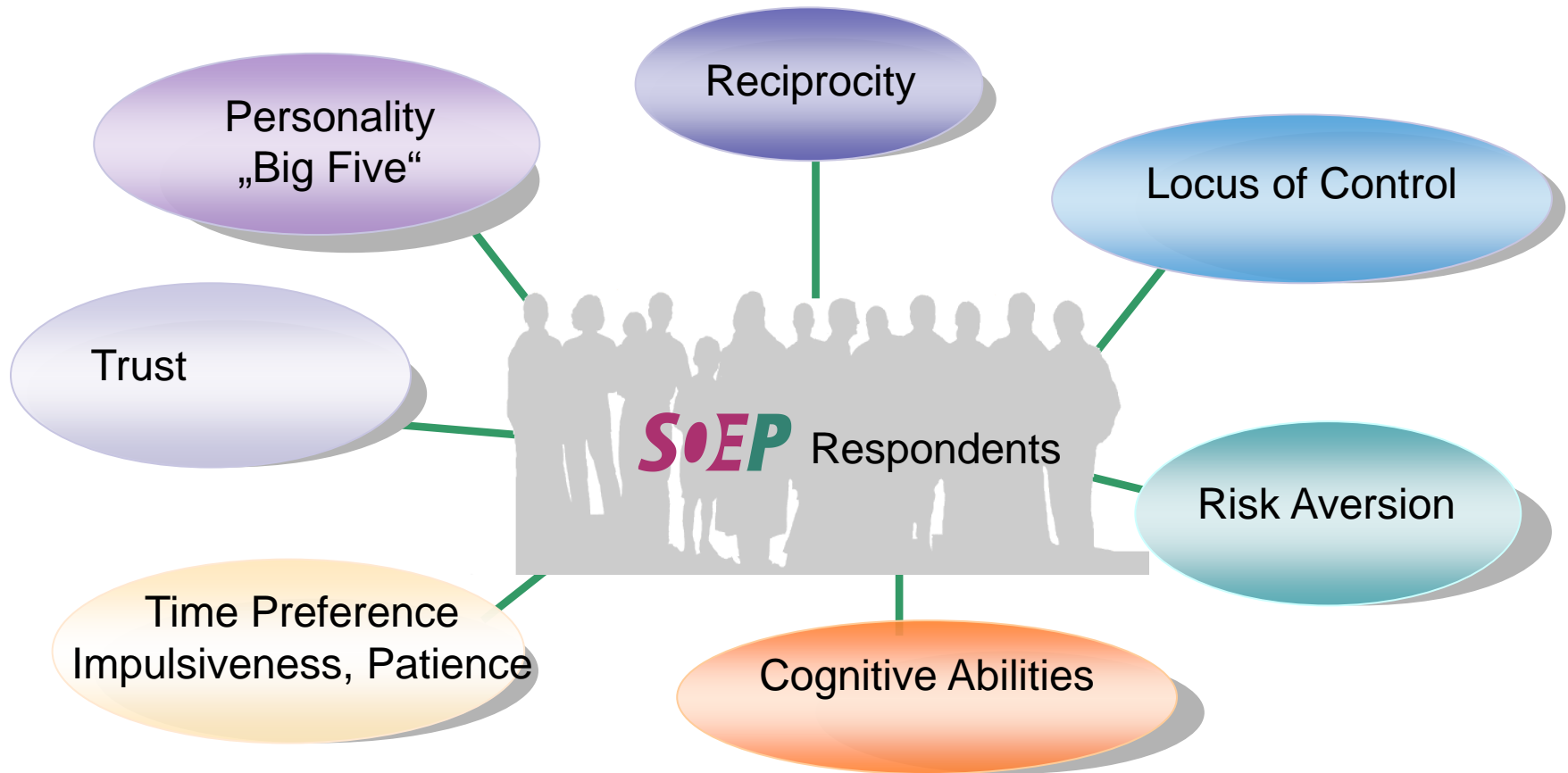
Anger & Heineck (2009): Do Smart Parents Raise Smart Children?: The Intergenerational Transmission of Cognitive Abilities, SOEPpapers #156

Challenge: Observing the „*unobservable*“

... and going beyond standard interview techniques

- Evaluation of health
(by taking grip strength)
- Evaluation of cognitive abilities
- Personal traits
(Big Five, risk aversion, trust, fairness, ...)

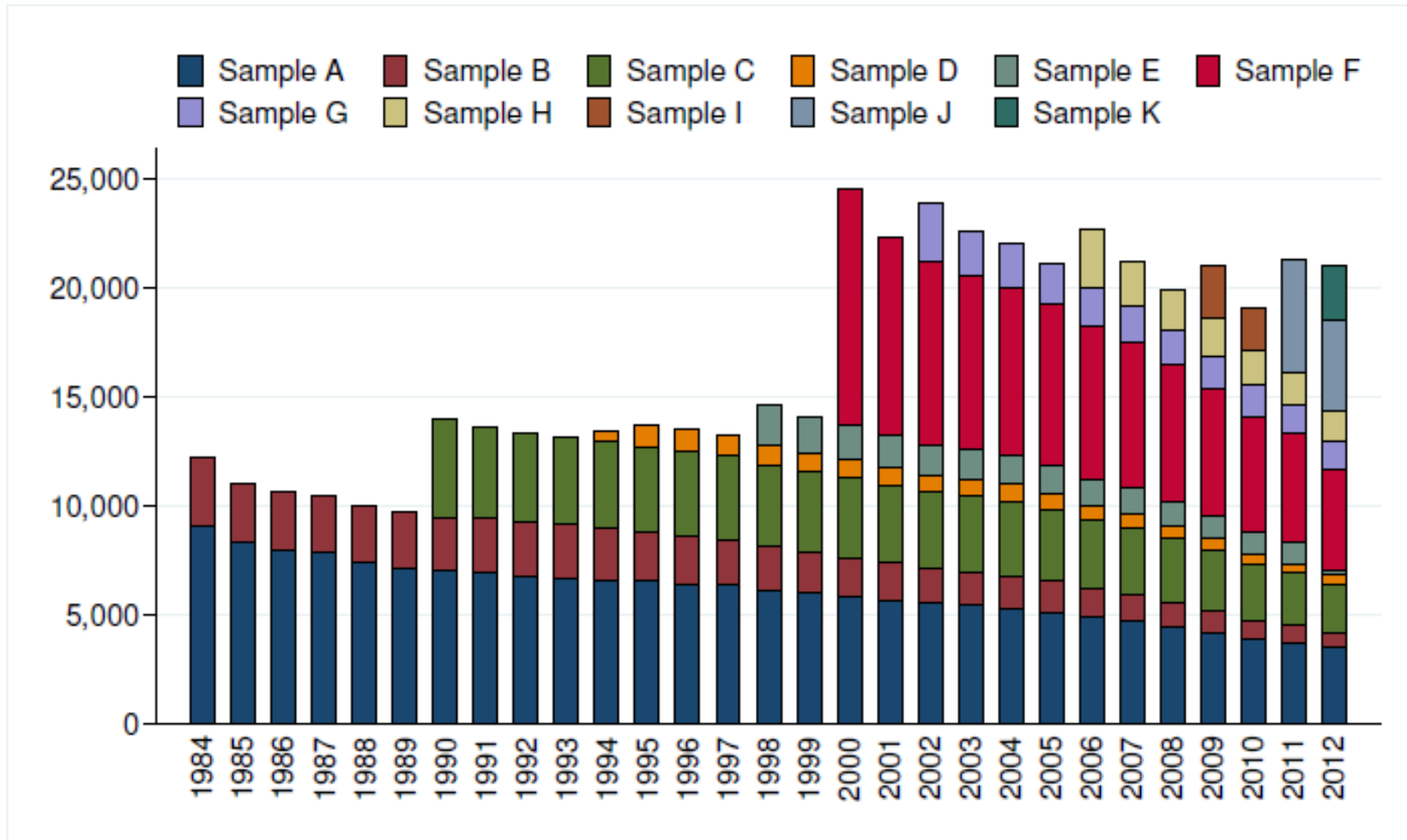
Personal Traits



Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
 - Development
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

The SOEP sample 1984 – 2010 (persons)



Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
 - Questionnaires
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

Survey Instruments: Questionnaires (\$P/\$H-files)

⇒ Yearly Standard Instruments

- Household questionnaire
- Individual questionnaire for each HH member aged 17 and over
- Individual questionnaire „Gap“ (Temporary Drop-outs)

⇒ Biography Questionnaire „Life History“

- Covering life-time information up until 1st interview (education and labor market, marriage and fertility, socialization and parental background, immigration)
- Until 2000 to be answered by first time respondents aged 17+
- Since 2001: Different biography questionnaires for a) *adult first time respondents*, and b) *young first time respondents*

Modes of entry:

- Being sampled in original- or refreshment samples
 - Moving into a SOEP-Household
 - Birth/Growing up in a SOEP-Household
- } **Standard biography interview**
- } **Youth-specific biography interview**

Capturing the full life course

Improving measures about the early phase of life

→ „mother & child questionnaires“

Fostering longitudinal, intergenerational & life course research

- ✓ Since 2003: data on new born children (about 0-15 mths)
- ✓ Since 2005: data on infants (about 3 yrs)
- ✓ Since 2008: data on first-graders (about 6 yrs)
- ✓ Since 2010: data on school-children (about 8 yrs)
- Since 2013: data on secondary level students (about 11 yrs)

... what about international comparisons?

Cross-National Equivalent File (CNEF)



- ⇒ Standardization of Various Measures
(Focus: Income, Demographics, Employment, Health)
- ⇒ USA PSID - Panel Study of Income Dynamics
- ⇒ Germany SOEP - German Socio-Economic Panel Study
- ⇒ Great Britain BHPS - British Household Panel Study
- ⇒ Canada SLID - Survey of Labour and Income Dynamics
- ⇒ Australia HILDA - Household, Income and Labour
 Dynamics in Australia
- ⇒ Switzerland SHP - Swiss Household Panel
- ⇒ Korea KLIPS - Korea Labor and Income Panel Study

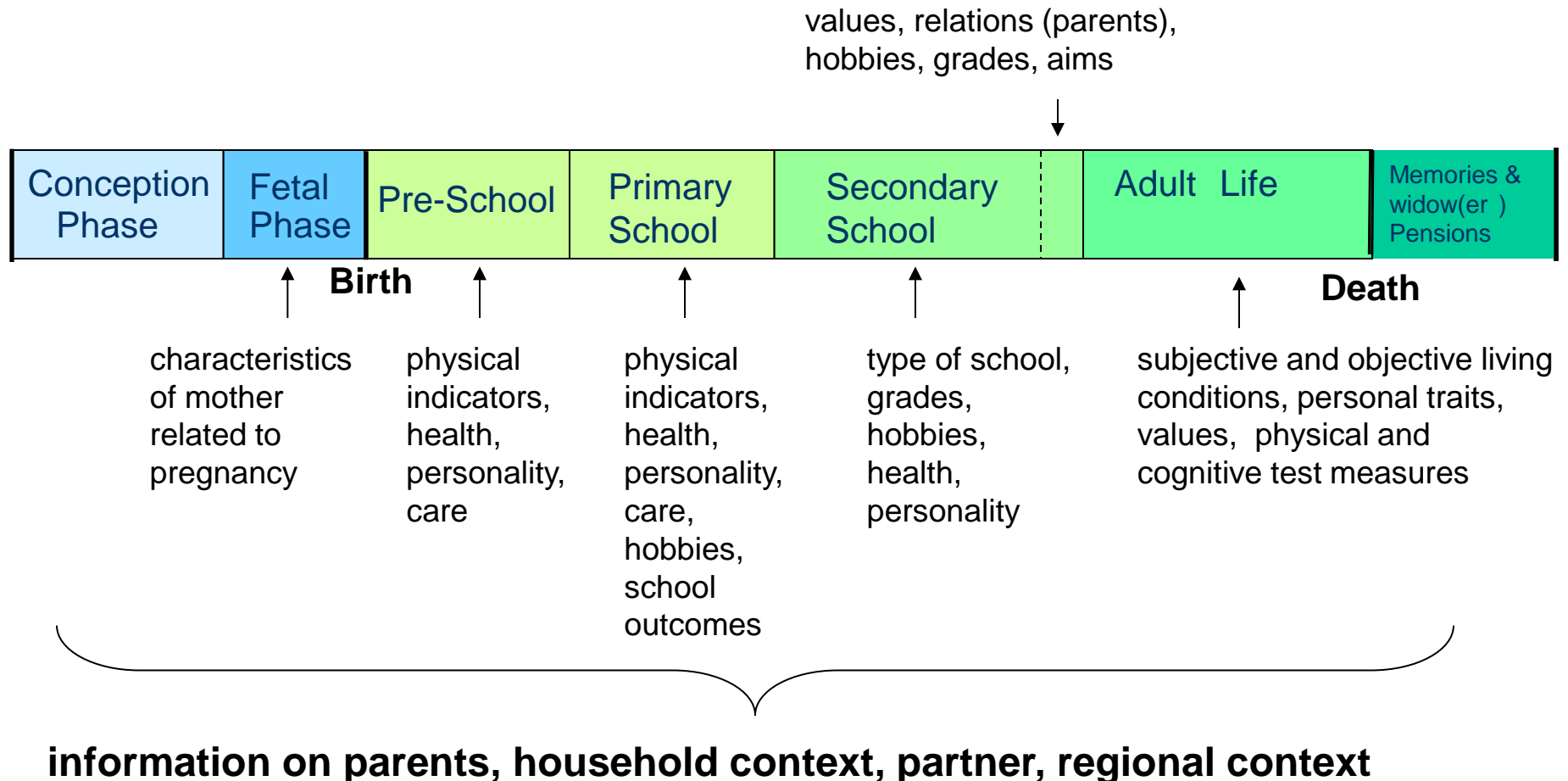
Frick, Jenkins, Lillard, Lipps and Wooden (2007): The Cross-National Equivalent File (CNEF) and its Member Country Household Panel Studies. *Schmoller's Jahrbuch - Journal of Applied Social Science Studies*. 127 (4): 627-654.

Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
 - Summary
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

Topics in the SOEP

life-course and available information

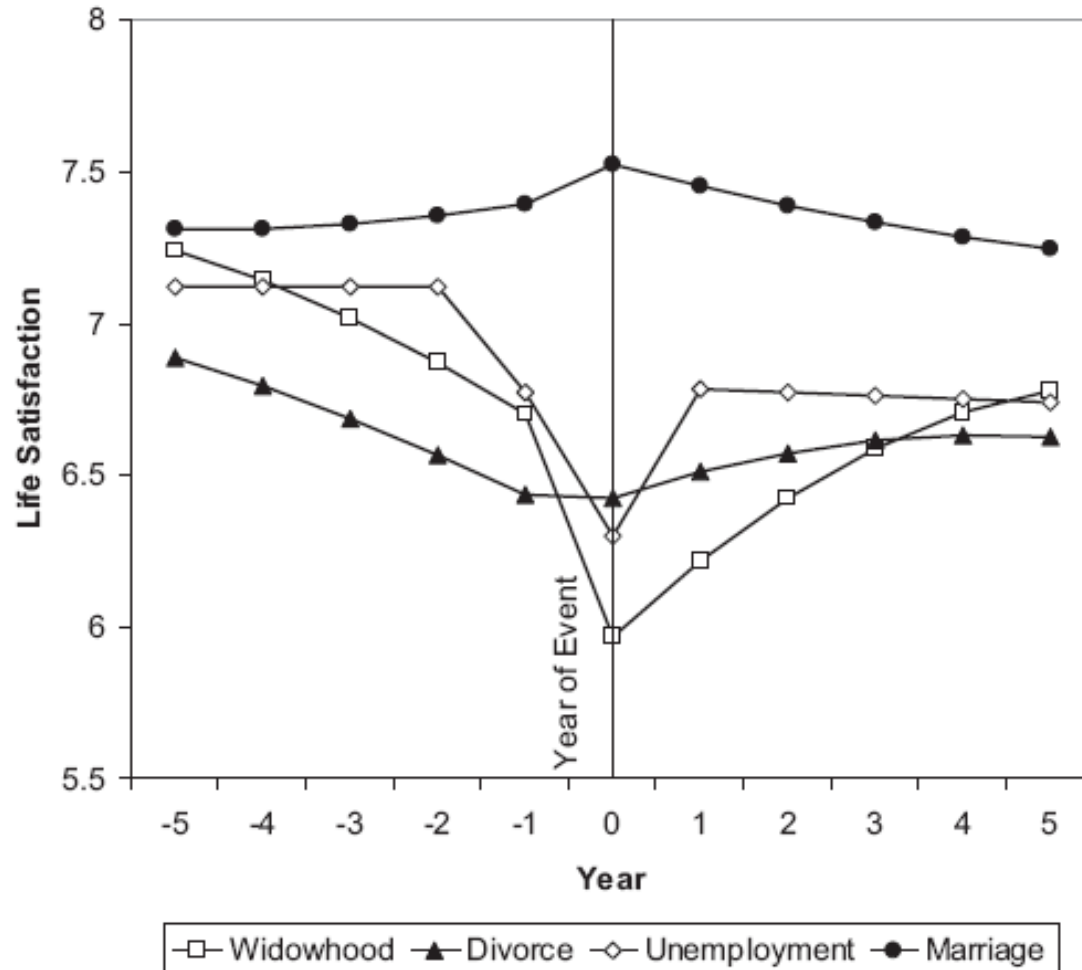


Overview

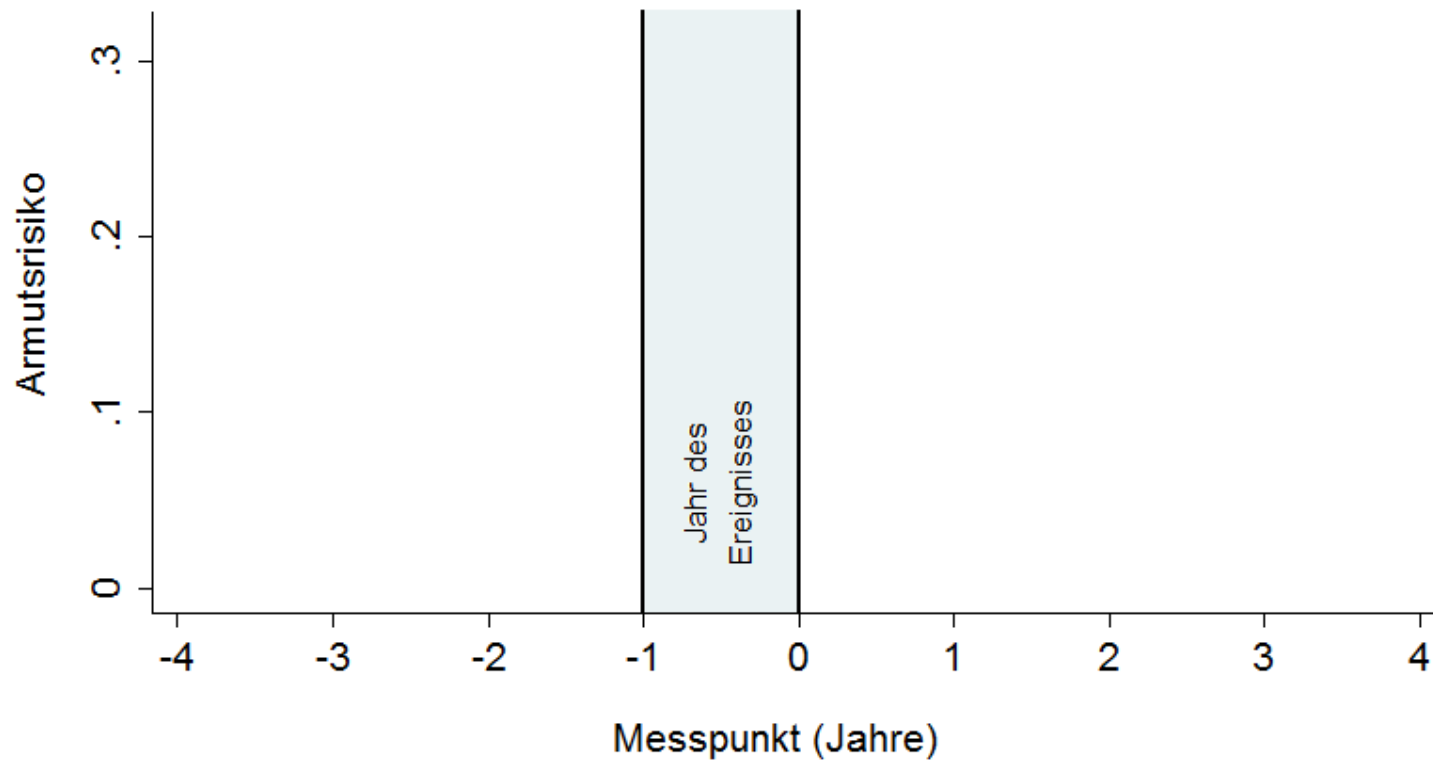
1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
4. **Research Examples**
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

Potential of the SOEP for Analyses: Examples

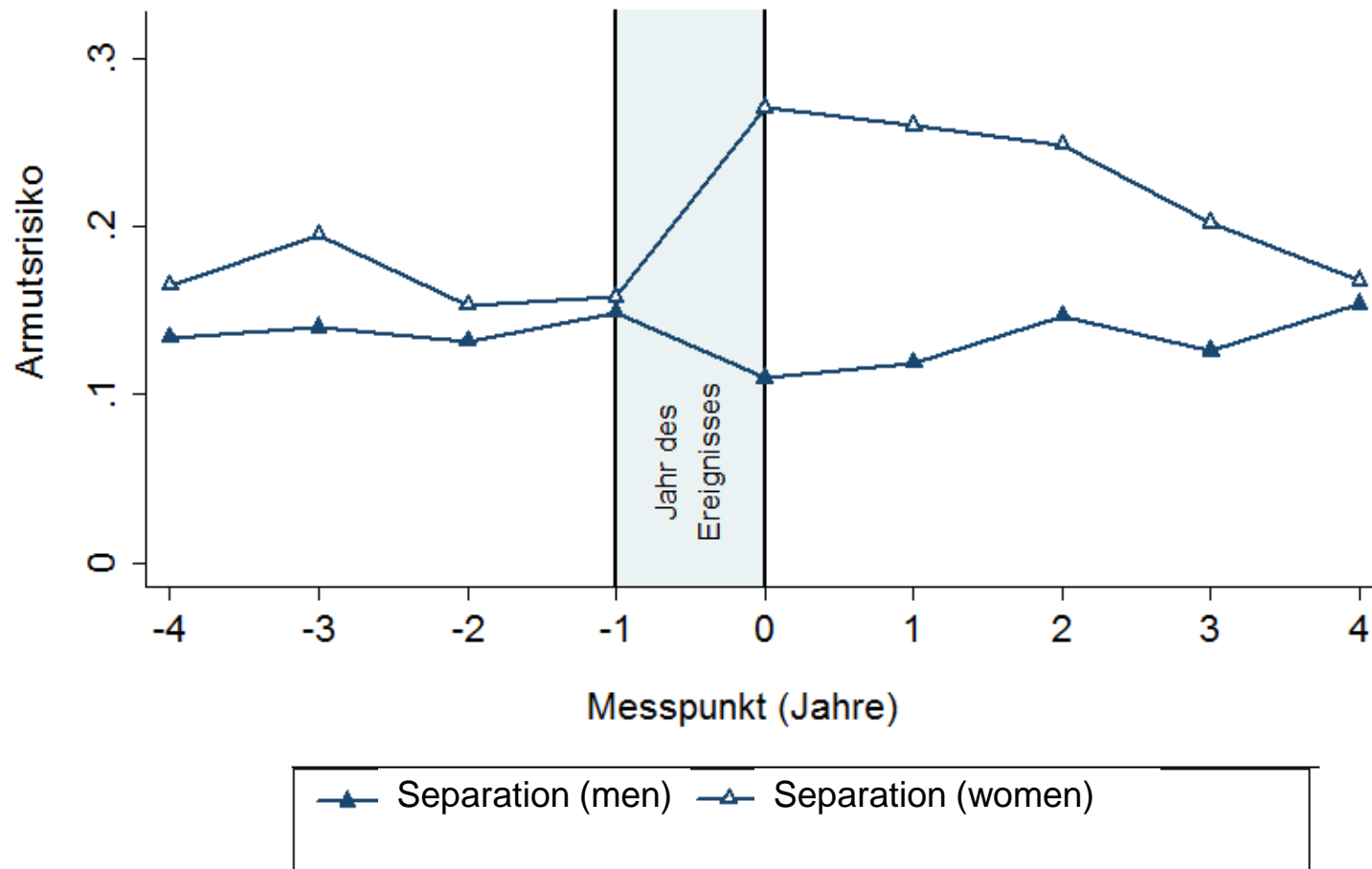
Figure 2
Adaptation to Good and Bad Events



Critical life events and poverty dynamics

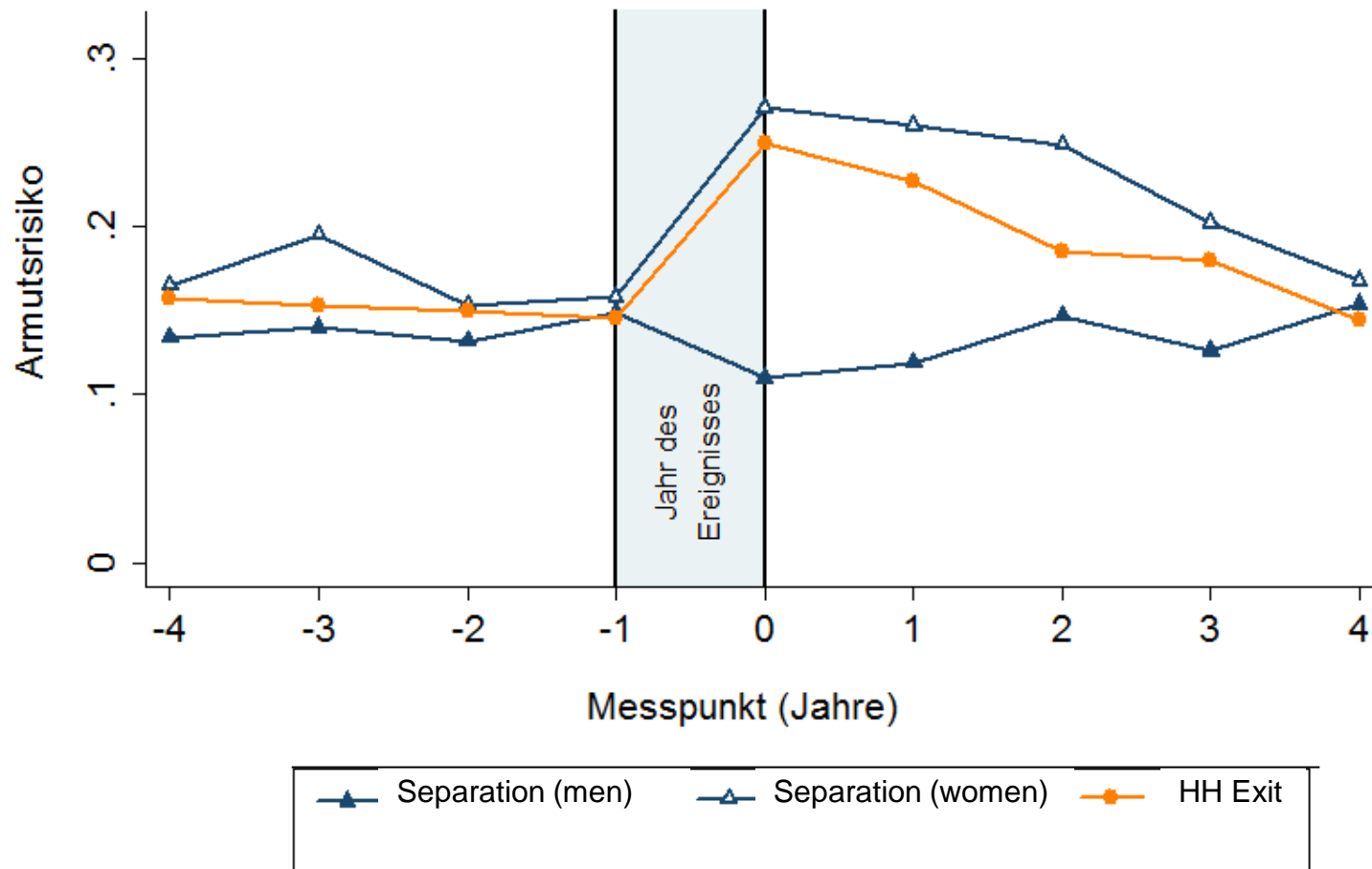


Critical life events and poverty dynamics



Giesselmann/Goebel 2013, *Analyse und Kritik*

Critical life events and poverty dynamics



Giesselmann/Goebel 2013, *Analyse und Kritik*

Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
4. Research Examples
5. **Data Distribution**
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal

Data Distribution

- ⇒ Strict data protection legislation requires ...
 - Data user contract with DIW Berlin
 - Intl. Users: Co-operation with Cornell Univ., Ithaca/NY

- ⇒ Data dissemination on DVD only
 - detailed written documentation
 - STATA, SPSS, SAS
 - full German and English labeling (variables and values)

- ⇒ Low nominal data fee (30 EUR/year; 125 USD one-time)
 - yearly updates on CD-Rom / DVD

- ⇒ Access to sensitive information (e.g. regional data):
 - *SOEPrmote* allows processing of sensitive data from outside DIW Berlin

- SOEP Newsletter: Information on distribution, updates, news and bugs

Data Distribution

- data user contract can be downloaded:
http://www.diw.de/de/diw_02.c.222836.de/datenzugang.html
- **SOEPhotline**
Michaela Engelmann
Tel.: +49 30 89789-292, Fax.: +49 30 89789-109
soepmail@diw.de

Overview

1. What is Panel Data?
2. Advantages / drawbacks of panel data
3. What is the SOEP?
4. Research Examples
5. Data Distribution
6. Data Structure
7. Preparing Data for Analysis
 - Cross-Section
 - Longitudinal