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

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
### B) Time-Series Data

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Unit 2

Unit 3
C) Panel Data

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Unit 2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Unit 3</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>
### D) Pooled Cross-Section Data

<table>
<thead>
<tr>
<th></th>
<th>2000</th>
<th>2001</th>
<th>2002</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit 1</td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unit 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
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
**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

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

- Personality „Big Five“
- Reciprocity
- Locus of Control
- Trust
- Risk Aversion
- Time Preference, Impulsiveness, Patience
- Cognitive Abilities

SOEP Respondents
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)

The Cross-National Equivalent File (CNEF) is a resource that standardizes various measures across different countries. Its focus includes income, demographics, employment, and health. The CNEF consists of several country-specific household panel studies:

- **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)

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*

- **Conception Phase**
  - characteristics of mother related to pregnancy

- **Fetal Phase**
  - physical indicators, health, personality, care

- **Pre-School**
  - physical indicators, health, personality, care, hobbies, school outcomes

- **Primary School**
  - type of school, grades, hobbies, health, personality

- **Secondary School**
  - subjective and objective living conditions, personal traits, values, physical and cognitive test measures

- **Adult Life**
  - memories & widow(er) Pensions

- **Death**

Information on parents, household context, partner, regional context
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

Diener et al. 2006, American Psychologist, 61(4), 310
Critical life events and poverty dynamics

Giesselmann/Goebel 2013, Analyse und Kritik
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):
  - \textit{SOEPremote} 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

- SOEP hotline
  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