Introduction to the Socio-economic Panel study
Overview

1. **What is Panel Data??**

2. **What is the SOEP?**
   a) Topics
   b) Questionnaire
   c) Structure

3. **Enhancements**
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. **Data Access**
What is Panel Data?

• Panel Data: Repeated measurements of many units at fixed time points
## A) Cross-Section Data

<table>
<thead>
<tr>
<th>Unit 1</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></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit 2</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></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Unit 3</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></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></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>
<tr>
<td>Unit 2</td>
<td></td>
<td></td>
<td></td>
<td></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>
</tbody>
</table>
### 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><strong>Unit 1</strong></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><strong>Unit 2</strong></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><strong>Unit 3</strong></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>
# Types of Data

## 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>X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Advantages of Panel Data

⇒ Match information from different phases of the life course

⇒ Control for time-constant characteristics (-> validate causal interpretations)

⇒ (Model sequentiality)

⇒ (Measure trends)

⇒ Decomposition of trends

Paper:
Giesselmann, Schröder, Giesecke, Haisken-DeNew, Rasner & Specht (2015)
Advantages of Panel Data, spec. SOEP

Advantages of Panel Data: Illustration

The consequences of life-events:
What is the influence of socio-demographic changes on life-satisfaction? (Diener et al. 2006)
Figure 2
Adaptation to Good and Bad Events

Life Satisfaction

Year

Year of Event

- Widowhood
- Divorce
- Unemployment
- Marriage

Diener et al. 2006, American Psychologist, 61(4), 310
Advantages of Panel Data: Illustration

Distributional Analyses: The distribution of incomes in Germany over time
Einkommensschichtung in Deutschland¹ 1984 bis 2006
In Prozent, Einkommensposition auf Basis des jahresspezifischen Median

<table>
<thead>
<tr>
<th>Jahr</th>
<th>&lt;50%</th>
<th>50 bis 70%</th>
<th>70 bis 90%</th>
<th>90 bis 110%</th>
<th>110 bis 130%</th>
<th>130 bis 150%</th>
<th>150% und mehr</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>11,4</td>
<td>14,0</td>
<td>17,6</td>
<td>14,1</td>
<td>13,4</td>
<td>9,0</td>
<td>11,3</td>
</tr>
<tr>
<td>2004</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2002</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1998</td>
<td>7,3</td>
<td>13,4</td>
<td>19,0</td>
<td>18,9</td>
<td>13,4</td>
<td>10,0</td>
<td>11,6</td>
</tr>
<tr>
<td>1996</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1994</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1992</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1990</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1988</td>
<td>6,3</td>
<td>14,4</td>
<td>20,0</td>
<td>17,9</td>
<td>15,3</td>
<td>9,9</td>
<td>11,2</td>
</tr>
<tr>
<td>1986</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1984</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

¹ Nennt die Prozentsätze der Bevölkerung, die in der jeweiligen Einkommensschicht leben.
Disadvantages of Panel Data?

⇒ Attrition Bias
⇒ Panel Effects
⇒ Structural Disproportions
# Overview

1. **What is Panel Data??**

2. **What is the SOEP?**
   - a) Topics
   - b) Questionnaire
   - c) Structure

3. **Enhancements**
   - a) Spatial Data
   - b) SOEP-IS
   - c) Migrants & Refugees
   - d) Mother/Child Questionnaire

4. **Data Access**
What is the SOEP?

- Long running, representative yearly panel survey (since 1984)
- Large sample (in 2015 ~16,000 HH)
- Wide range of objective and subjective measures
- Part of social science research infrastructure

Paper:
Who is interviewed?

- Interviews with all household members >17yrs. In sampled Households
- (Proxy-)Information on children
- Persons moving into SOEP-Household
- Persons leaving a SOEP-Household
Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
Objective and subjective measures

Multi-disciplinary commitment
Key Questions:

- Demography
- Education & Qualification
- Labour Market Participation
- Income / Wages
- Housing
- Health
- Basic Orientations: Goals and Attitudes
- Life Satisfaction
SOEP Topics: what is measured?

**Rotating Questions:**

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
Measuring Grip Strength in SOEP

Results of the SOEP-Pretest
Grip Strength by Age and Sex
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
# Measuring Adults’ Cognitive Abilities

<table>
<thead>
<tr>
<th>Measures of cognitive ability:</th>
<th>2 ultrashort tests (Lang, 2005)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Symbol Correspondence test</strong> (fluid mechanics):</td>
<td><strong>Word fluency test</strong> (crystallized pragmatics):</td>
</tr>
<tr>
<td>developed after the Symbol-digit-modalities-test (SDMT) by Smith (1973, 1995)</td>
<td>developed after the Animal-naming-task (Lindenberger and Baltes, 1995)</td>
</tr>
<tr>
<td>Respondents have 90 seconds to assign as many correct signs as possible to the consecutively displayed digits, while the appropriate assignment code is visible to them.</td>
<td>Respondents have 90 seconds to name as many different animals as possible.</td>
</tr>
<tr>
<td>This test measures fluid mechanics, which refers to general and largely innate ability.</td>
<td>This test measures crystallized pragmatics, which includes all cognitive abilities that consist of accumulated knowledge and skills acquired in the past.</td>
</tr>
</tbody>
</table>
Objectives
- Gain insight in skill formation processes among young people (skill acquisition)
- Contribution to interdisciplinary research efforts on social disparities in educational attainment and labor market success

Intelligence-Structure-Test
- I-S-T 2000 R (Amthauer et al. 2001)

Dimensions
- Verbal potential (analogies)
- Numerical potentials (number sequences)
- Figural potentials (matrices/pictorial material)
- Reasoning (= sum score)

Survey
- Test take about 30 minutes
- Participation rate about 90%
- Additional incentive: stop-watch

Paper:
Personal Traits

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

SOEP Respondents
What about international Comparisons?
### Cross-National Equivalent File (CNEF)

<table>
<thead>
<tr>
<th>Country</th>
<th>Study Abbreviation</th>
<th>Study Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>USA</td>
<td>PSID</td>
<td>Panel Study of Income Dynamics</td>
</tr>
<tr>
<td>Germany</td>
<td>SOEP</td>
<td>German Socio-Economic Panel Study</td>
</tr>
<tr>
<td>Great Britain</td>
<td>BHPS</td>
<td>British Household Panel Study</td>
</tr>
<tr>
<td>Canada</td>
<td>SLID</td>
<td>Survey of Labour and Income Dynamics</td>
</tr>
<tr>
<td>Australia</td>
<td>HILDA</td>
<td>Household, Income and Labour Dynamics in Australia</td>
</tr>
<tr>
<td>Switzerland</td>
<td>SHP</td>
<td>Swiss Household Panel</td>
</tr>
<tr>
<td>Korea</td>
<td>KLIPS</td>
<td>Korea Labor and Income Panel Study</td>
</tr>
</tbody>
</table>
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
- Switzerland  SHP   - Swiss Household Panel
- Korea  KLIPS    - Korea Labor and Income Panel Study

Paper:
Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire & Sample
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
Standard Questionnaires

- Household Questionnaire (→ HI – Files)
- Individual Questionnaire (→ PI - Files)

Biographic Interviews

- New adult sample members (→ Biosoc, Bio...)
- Grown-up children in sample HH (→ Bioage 17, $page17)

Additional Survey Instruments

- Mother & Child Questionnaires (→ Bioagel)
Developement of the Sample

- **Exit from the survey**
  - Death
  - Exmigration
  - Unit-Nonresponse

- **Entry to the survey:**
  - Drawn for original or refreshment sample
  - Birth in a SOEP Household
  - Moving into SOEP-Household
  - Household formation with a member from a former SOEP-Household
Developement of Sample Size

Paper:
Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
## Structure (SOEPlong)

<table>
<thead>
<tr>
<th>Basic Information</th>
<th>Individual Level</th>
<th>Generated Panel Variables</th>
<th>Generated Episode Data</th>
<th>Biography Interviews</th>
<th>Additional Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>PPFADI (PBRUTTO) (PPHRF)</td>
<td>Individual Questionnaire</td>
<td>Individual Questionnaire</td>
<td>Generated Panel Variables</td>
<td>Generated Episode Data</td>
<td>Biography Interviews</td>
</tr>
<tr>
<td>(HBRUTTO) (HHRF)</td>
<td>PI</td>
<td>PKAL</td>
<td>PEQUIV</td>
<td>PGEN</td>
<td>PBIOSPE</td>
</tr>
<tr>
<td></td>
<td>KIND</td>
<td>$_mig$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Basic Information**
- **Individual Level**
  - Individual Questionnaire
  - Individual Level Variables
  - Kind
  - $\_mig$

- **Household Level**
  - Household Questionnaire
  - Household Level Variables

- **Generated Panel Variables**
  - PEQUIV
  - PGEN
  - BIOAGE17
  - BIOSOC
  - BIOCEDU
  - BIOPAREN
  - BIORESID
  - BIOJOB
  - BIOIMMIG
  - MIGSPELL

- **Generated Episode Data**
- **Biography Interviews**
- **Additional Data**

---

38

DIW BERLIN
### Missing Conventions

<table>
<thead>
<tr>
<th>Code</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1</td>
<td>no answer / don’t know</td>
</tr>
<tr>
<td>-2</td>
<td>does not apply</td>
</tr>
<tr>
<td>-3</td>
<td>implausible value</td>
</tr>
<tr>
<td>-4</td>
<td>Inadmissible multiple response</td>
</tr>
<tr>
<td>-5</td>
<td>Not included in this version of the questionnaire</td>
</tr>
<tr>
<td>-6</td>
<td>Version of questionnaire with modified filtering</td>
</tr>
<tr>
<td>-8</td>
<td><em>Question not part of the survey program this year</em></td>
</tr>
</tbody>
</table>

*Only applicable for datasets in long format.*
Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
### Regional Data

<table>
<thead>
<tr>
<th>Level</th>
<th>Available Since</th>
<th>Data Access</th>
<th>Data Protection</th>
</tr>
</thead>
<tbody>
<tr>
<td>States (Bundesländer)</td>
<td>1984</td>
<td>Standard SOEP dataset (Scientific Use File)</td>
<td>Data distribution contract</td>
</tr>
<tr>
<td>Municipal size classes (e.g., Boustedt)</td>
<td>1984</td>
<td>Standard SOEP dataset with special password</td>
<td>Expanded data distribution contract on the use of municipal size classes &amp; data protection concept</td>
</tr>
<tr>
<td>Spatial planning regions (geocodes)</td>
<td>1985</td>
<td>Standard SOEP dataset plus SOEP geocode disk</td>
<td>Expanded data distribution contract on the use of geocodes &amp; expanded data protection concept</td>
</tr>
<tr>
<td>Official county codes (KKZ)</td>
<td>1985</td>
<td>SOEPROMote (online access to county-level regional data) or at the SOEP Research Data Center at DIW Berlin</td>
<td>Expanded data distribution contract on the use of SOEPROMote &amp; SOEPROMote access form</td>
</tr>
<tr>
<td>Official municipality key, postal codes, Microm neighborhood data</td>
<td>2000 1993 2000</td>
<td>Use of data only at the SOEP Research Data Center at DIW Berlin</td>
<td>Only by personal arrangements in the framework of our SOEP in residence program</td>
</tr>
</tbody>
</table>

### Verfügbarkeit Regional-Codes

<table>
<thead>
<tr>
<th>2001</th>
<th>ROR</th>
<th>Kreise</th>
<th>Gemeinden</th>
<th>PLZ</th>
</tr>
</thead>
<tbody>
<tr>
<td># der Regionen</td>
<td>97</td>
<td>439</td>
<td>12141</td>
<td>17201</td>
</tr>
<tr>
<td>Anzahl (%) der Regionen mit SOEP Befragungspersonen</td>
<td>97 (100%)</td>
<td>439 (100%)</td>
<td>2561 (21,1%)</td>
<td>2420 (14,1%)</td>
</tr>
<tr>
<td>Durchschn. Anzahl von SOEP-HH (Personen)</td>
<td>113 (206)</td>
<td>25 (45)</td>
<td>4 (8)</td>
<td>4 (8)</td>
</tr>
<tr>
<td>Verfügbar für SOEP-Wellen seit...</td>
<td>1985</td>
<td>1985</td>
<td>2000</td>
<td>1993</td>
</tr>
<tr>
<td>Zugang</td>
<td>Zusatzvertrag (nur EU)</td>
<td>SOEPremote am DIW</td>
<td>Nur am DIW</td>
<td>Nur am DIW</td>
</tr>
</tbody>
</table>
Kooperation mit microm

- Zuspielung der microm Daten seit 2005
- Hohe Datenschutzanforderungen (4-Augen Prinzip mit Infratest)
- Eigentliche Zuspielung der Daten rückwirkend bis 2000
- Fokus der ersten Jahre auf Zuspielung der Daten möglichst kleinräumig
- Später auch "weitere" Umgebung der Haushalte
- Umstellung Marktzelle auf PLZ-8 Gebiete, mit Überlappung in 2010 und 2011
<table>
<thead>
<tr>
<th>Soziodemographie</th>
<th>Sozioökonomie</th>
<th>Soziokultur</th>
<th>Wohnumfeld</th>
</tr>
</thead>
<tbody>
<tr>
<td>microm Sozio</td>
<td>microm Sozio</td>
<td>Sinus-Milieu²</td>
<td>microm Typologie</td>
</tr>
<tr>
<td>Status (Bildung, Beruf)</td>
<td>Status (Einkommensbestandteil)</td>
<td>Milieuwahrscheinlichkeiten</td>
<td>38 Typen</td>
</tr>
<tr>
<td>Alter (Δ des HHV)</td>
<td>microm Kaufkraft</td>
<td>microm Geo Milieu²</td>
<td>10 Gruppen</td>
</tr>
<tr>
<td>Familienstruktur</td>
<td>microm Zahlungsindex</td>
<td>Migranterhaushalte</td>
<td></td>
</tr>
<tr>
<td>Ausländeranteil</td>
<td></td>
<td>Milieuwahrscheinlichkeiten</td>
<td></td>
</tr>
<tr>
<td>microm Ethno</td>
<td></td>
<td>Lebensstilinformationen und Konsumgewohnheiten</td>
<td></td>
</tr>
<tr>
<td>microm Lebensphasen</td>
<td></td>
<td>microm B4P (Best for Planning)</td>
<td></td>
</tr>
<tr>
<td>microm Einwohner</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gesamtzahl</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nach Geschlecht</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nach 17 amtlichen Altersklassen</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nach Geschlecht und 17 amtlichen Altersklassen</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Sozialraum</th>
</tr>
</thead>
<tbody>
<tr>
<td>microm Basis</td>
</tr>
<tr>
<td>Privathaushalte</td>
</tr>
<tr>
<td>Gebäude</td>
</tr>
<tr>
<td>microm Bebauung</td>
</tr>
<tr>
<td>Haustyp</td>
</tr>
<tr>
<td>microm Wohnen</td>
</tr>
<tr>
<td>Mieter</td>
</tr>
<tr>
<td>microm Mobilität</td>
</tr>
<tr>
<td>Umzugsvolumen</td>
</tr>
<tr>
<td>Fluktuation</td>
</tr>
<tr>
<td>microm Arbeitslosenquote</td>
</tr>
</tbody>
</table>
Beispiel: Mikroanalysen mit Raumindikatoren

Quelle: SOEPv30-microm-Datensatz Welle 25-30
Nachbarschaftsradius 0,5km
Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
SOEP-IS (Innovationspanel)

- Seit Herbst 2011: längsschnittliche Innovations-Stichprobe (ca. 5.000 HH)
- Methodische und thematische Innovationen:
  - Verhaltensexperimente und Interventionsstudien
  - ausführliche Kompetenzmessungen
  - implizite Assoziations-Tests (IAT)
  - komplexere Verfahren der Messung von Zeitverwendung (DRM)
  - „invasive“ Verfahren, genetische Studien
- Ansonsten ein zu großes Risiko hoher Ausfallraten für die Kern-SOEP-Stichprobe
- Externe Wissenschaftler können und sollen sich um Befragungs- und Experimentier-Platz im SOEP-Innovationssample bewerben
Inclusion of questions

- **Application requirements:**
  - 2-5-page scientific description in English
  - Description of the scientific rationale, type of sample, mode of data collection, repetition of the survey
  - Information on the total number of minutes required

- **Evaluation criteria:**
  - Scientific quality and significance
  - Suitability to the SOEP Innovation Sample
  - Feasibility of implementation
SOEP-IS (Datenweitergabe)

Data distribution of SOEP-IS

- All SOEP data are freely available to all researchers
  (scientific open access *after one year of exclusive use*)

- This also applies to:
  - all of the different sets of questions and experiments submitted by external applicants
  - data collected with funding from external sources

Paper:

Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
Total Number of Foreigners who immigrated to Germany, 1984-2015

Source: Statistical Office
Existing Migrant Samples in the SOEP

- Existing Migration Samples in the SOEP (1984, 1994)
  - Immigration up to 1983 (“Guest Workers”, Sample B)
  - Immigration from 1984-94 (“Ethnic Germans”, Sample D)
  - General population refreshers (Samples E, H, K), with oversampling of migrants (Samples F and J)
- IAB-SOEP Migration Samples (2013, 2015)
  - Immigration between 1995-2012 and second-generation migrants
    \((Adult\ Anchor\ Person\), majority EU-25, Sample M1)
  - Immigration between 2009-14 (often EU-27, Sample M2)
- IAB-BAMF-SOEP Refugee Samples (both 2016)
  - Refugees between 2013-16 \((Adult\ Anchor\ Person\), Sample M3)
  - Refugees between 2013-16 \((Underage\ Anchor\ Persons\), M4)
Existing Migrant Samples in the SOEP

- Existing Migration Samples in the SOEP (1984, 1994)
  - Immigration up to 1983 ("Guest Workers", Sample B)
  - Immigration from 1984-94 ("Ethnic Germans", Sample D)
  - General population refreshers (Samples E, H, K), with oversampling of migrants (Samples F and J)
- IAB-SOEP Migration Samples (2013, 2015)
  - Immigration between 1995-2012 and second-generation migrants (Adult Anchor Person, majority EU-25, Sample M1)
  - Immigration between 2009-14 (often EU-27, Sample M2)
- IAB-BAMF-SOEP Refugee Samples (both 2016)
  - Refugees between 2013-16 (Underage Anchor Persons, Sample M4)
  - Refugees between 2013-16 (Adult Anchor Person, Sample M3)

Paper:
### Number of Migrants in SOEP-Core in 2013

<table>
<thead>
<tr>
<th>Adults (Children)</th>
<th>Samples A-L</th>
<th>Sample M1</th>
<th>Sample M2</th>
<th>Samples M3-M4</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Generation Migrants</td>
<td>2,678 (94)</td>
<td>3,709 (261)</td>
<td></td>
<td></td>
<td>6,387 (355)</td>
</tr>
<tr>
<td>Second Generation Migrants</td>
<td>2,722 (2,457)</td>
<td>945 (1,933)</td>
<td></td>
<td></td>
<td>3,667 (4,990)</td>
</tr>
<tr>
<td>Subtotal</td>
<td>5400 (3,151)</td>
<td>4,654 (2,194)</td>
<td></td>
<td></td>
<td>10,054 (5,345)</td>
</tr>
<tr>
<td>No Migration Background</td>
<td>21,126 (7,570)</td>
<td>310 (92)</td>
<td></td>
<td></td>
<td>21,426 (7,662)</td>
</tr>
<tr>
<td>Total</td>
<td>26,516 (10,721)</td>
<td>4,964 (2,286)</td>
<td></td>
<td></td>
<td>31,480 (13,007)</td>
</tr>
</tbody>
</table>
### Number of foreign born migrants in SOEP-Core in 2014 by country of origin

<table>
<thead>
<tr>
<th>Country Of Birth</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Germany</td>
<td>13,833</td>
</tr>
<tr>
<td>Turkey</td>
<td>2,584</td>
</tr>
<tr>
<td>Russia</td>
<td>1,499</td>
</tr>
<tr>
<td>Poland</td>
<td>1,386</td>
</tr>
<tr>
<td>Kazakhstan</td>
<td>1,239</td>
</tr>
<tr>
<td>Italy</td>
<td>1,042</td>
</tr>
<tr>
<td>Ex-Yugoslavia</td>
<td>969</td>
</tr>
<tr>
<td>Greece</td>
<td>727</td>
</tr>
<tr>
<td>Romania</td>
<td>674</td>
</tr>
<tr>
<td>Spain</td>
<td>534</td>
</tr>
<tr>
<td>Kosovo-Albania</td>
<td>305</td>
</tr>
<tr>
<td>Ukraine</td>
<td>289</td>
</tr>
<tr>
<td>Eastern Europe</td>
<td>163</td>
</tr>
<tr>
<td>Austria</td>
<td>161</td>
</tr>
<tr>
<td>Iraq</td>
<td>145</td>
</tr>
<tr>
<td>Kyrgyzstan</td>
<td>132</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>124</td>
</tr>
<tr>
<td>Bosnia-Herzegovina</td>
<td>124</td>
</tr>
<tr>
<td>France</td>
<td>105</td>
</tr>
<tr>
<td>USA</td>
<td>105</td>
</tr>
<tr>
<td>Serbia</td>
<td>103</td>
</tr>
</tbody>
</table>
Questionnaire for new Immigrant Samples

• Individual biography questionnaire:
  • Country of origin
  • How respondent came to Germany
  • Status of asylum procedure
  • Arrival in Germany
  • Intention to stay
  • Language knowledge
  • Support / networks
  • Employment (prior to coming to Germany and now), income
  • Education, training, and qualifications (in country of origin, recognition of qualifications, qualifications already obtained in Germany, currently ongoing training, numbers and figures test)
Questionnaire for new Immigrant Samples

• Individual biography questionnaire:
  • Health (SF12, loneliness, resilience, fear and depression, self-esteem)
  • Personality (reciprocity, control beliefs, risk-taking)
  • Attitudes and opinions (connection to country of origin, feeling welcome in Germany, perceptions of discrimination, religion, worries, political interest in country of origin and in Germany, conception of democracy, gender role images)
  • Expectations for the future in relation to education and labor market participation
  • Family situation
  • Payments made
• Household questionnaire:
  • Current housing / apartment
  • Transfers received
  • Provision of daily necessities
  • Participation
  • Children (attendance of daycare, preschool, school)

All Immigrant-Samples are integrated in the regular SOEP-datasets!

Informational overlap is organized in additional datasets!
Paper:
Goebel, J. und Strauch, K. (2016)
BIOIMMIG: Generated and Status Variables from SOEP for Foreigners and Migrants. In: Goebel, J. (Hrsg) SOEP 2014 – Documentation on Biography and Life History Data for SOEP v31 and v31.1
• PPFAD contains the following user-friendly variables:
  • Germborn – Respondents born in Germany
  • Immiyear – Year of immigration (for those not born in Germany)
  • Corigin – Country of origin (for those not born in Germany)
  • Migback – Migration background (direct or indirect migration background)
• PGEN contains nation$$ - citizenship

→ Other migration and integration variables mentioned on
  https://paneldata.org/topics/365
Overview

1. What is Panel Data??

2. What is the SOEP?
   a) Topics
   b) Questionnaire
   c) Structure

3. Enhancements
   a) Spatial Data
   b) SOEP-IS
   c) Migrants & Refugees
   d) Mother/Child Questionnaire

4. Data Access
Seit 2003 werden detaillierte Informationen zu Kindern in den Befragungshaushalten erhoben.
Elterninterviews zu Kindern im SOEP

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Start</th>
<th>Topics covered</th>
<th>N (SOEP v30, 2003-2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Kinderinfos im SOEP

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
<th>Start</th>
<th>Topics covered</th>
<th>N (SOEP v30, 2003-2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>bioage01</td>
<td>0-1</td>
<td>2003</td>
<td>birth information, health, temperament, care situation</td>
<td>2,307</td>
</tr>
<tr>
<td>(mothers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bioage03</td>
<td>2-3</td>
<td>2005</td>
<td>+ activities with the child, personality</td>
<td>1,879</td>
</tr>
<tr>
<td>(mothers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bioage06</td>
<td>5-6</td>
<td>2008</td>
<td>+ socio-emotional behavior</td>
<td>1,260</td>
</tr>
<tr>
<td>(mothers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bioage08</td>
<td>7-8</td>
<td>2010</td>
<td>+ school situation, parental role</td>
<td>847 / 562</td>
</tr>
<tr>
<td>(parents)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bioage10</td>
<td>9-10</td>
<td>2012</td>
<td>+ grades, educational aspirations and expectation</td>
<td>449</td>
</tr>
<tr>
<td>(mothers)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 3: Number of respondents to the “Mother & Child” and “Parent” questionnaires

<table>
<thead>
<tr>
<th>Bioage File</th>
<th>2003</th>
<th>2004</th>
<th>2005</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bioage01</td>
<td>318</td>
<td>247</td>
<td>246</td>
<td>234</td>
<td>205</td>
<td>185</td>
<td>196</td>
<td>182</td>
<td>146</td>
<td>166</td>
<td>182</td>
<td>2307</td>
</tr>
<tr>
<td>Bioage03</td>
<td></td>
<td>257</td>
<td>222</td>
<td>237</td>
<td>246</td>
<td>186</td>
<td>190</td>
<td>173</td>
<td>190</td>
<td>178</td>
<td></td>
<td>1879</td>
</tr>
<tr>
<td>Bioage06</td>
<td></td>
<td></td>
<td>237</td>
<td>210</td>
<td>214</td>
<td>210</td>
<td>187</td>
<td>202</td>
<td></td>
<td></td>
<td></td>
<td>1260</td>
</tr>
<tr>
<td>Bioage08a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>221</td>
<td>174</td>
<td>214</td>
<td>238</td>
<td></td>
<td></td>
<td>847</td>
</tr>
<tr>
<td>Bioage08b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>152</td>
<td>117</td>
<td>152</td>
<td>141</td>
<td></td>
<td>562</td>
</tr>
<tr>
<td>Bioage10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>222</td>
<td>227</td>
<td></td>
<td>449</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>318</td>
<td>247</td>
<td>503</td>
<td>456</td>
<td>442</td>
<td>668</td>
<td>592</td>
<td>959</td>
<td>820</td>
<td>1131</td>
<td>1168</td>
<td>7304</td>
</tr>
</tbody>
</table>
The Family of SOEP Studies 1984-

SOEP-Core (1984-)
- Oversampling of immigrants (A/B)
- Since 1990 East Germany (C)
- 1994/95 Immigration sample (D)
  1998 Refresher I (CAPI) (E)
  2000 Refresher II (F)
  2002 High-Income Refresher (G)
  2006 Refresher III (H)
  2009 Innovation Sample (Incentives) (I)
  2011 Refresher IV (J)
  2012 Refresher V (K)
  2013 Immigrant Sample II (M)
  2014 Integration of FiD

SOEP-IS
- Start in 2012
- Innovation sample 2009 (incentives)
- Integration of sample E
  Refresher 2012
  Refresher 2013
  Refresher 2014

SOEP-Related Studies
- Families in Germany (FiD) 2010-2013
  Bonn Intervention Study (BIP) (2014-2017)
Total no. of realized households in 2014 – about 20,000
## Overview

1. **What is Panel Data??**

2. **What is the SOEP?**
   - a) Topics
   - b) Questionnaire
   - c) Structure

3. **Enhancements**
   - a) Spatial Data
   - b) SOEP-IS
   - c) Migrants & Refugees
   - d) Mother/Child Questionnaire

4. **Data Access**
Der anonymisierte Mikrodatensatz des SOEP wird Universitäten und anderen Forschungseinrichtungen für Forschung und Lehre kostenfrei zur Verfügung gestellt.

Für die Datennutzung ist aus Gründen des Datenschutzes der Abschluss eines Datenweitergabevertrages mit dem DIW Berlin notwendig.

Nach Abschluss können die Daten über einen personalisierten Link mit individuellem Passwort verschlüsselt herunter geladen werden.
Der Antrag auf Datennutzung kann unter: 
http://www.diw.de/de/diw_02.c.222836.de/Data Access.html
heruntergeladen werden.

SOEPhotline
Ansprechpartnerin: Michaela Engelmann
Tel.: +49 30 89789-292, Fax.: +49 30 89789-109
soepmail@diw.de
Entwicklung der Erhebungsmethodik

- PAPI
- CAPI
- SAQ (with Iwer)
- SAQ (w/o Iwer)
- Mail
- CAWI