ALWA-Befragungsdaten verknüpft mit administrativen Daten des IAB (ALWA-ADIAB)

Erster ALWA-User-Workshop
Nürnberg
6. Dezember 2011

Manfred Antoni
Outline

- Data
- Process of record linkage
- References
- Outlook
Administrative data of the German Federal Employment Agency

- Sources of data of the German Federal Employment Agency (BA):
  - Social security notifications by employers
  - Data from internal processes of the local employment agencies and the BA
- Used to compute social security contributions and benefit entitlements
- High accuracy of income information
- Recall error or sample selectivity impossible
- Consistent Social Security Number ⇒ once a survey respondent is identified in the register data, the complete employment history is available.
Administrative labor market data on individuals

- Day-to-day information on histories of dependent employment and registered unemployment since 1975 (over 80% of the German labor force)
- Longitudinal information on earnings and benefit receipt
- Since 2000: detailed information on the participation in active labor market policy measures and job seeking
- Detailed social security benefit information since 2005 (household-level)
Administrative labor market data on establishments

- Daily establishment-level employment and wage bill from 1975-present
- Establishment History Panel (BHP) includes yearly measures on:
  - economic sector
  - qualification and age structure
  - wage distribution inside the firm
  - worker flows for different subgroups of employees
  - founding and closing of firms
- Variables computed for every firm based on 30th of June of each year
Overview of IAB data

Social security notifications

BA/IAB data

Employment History

Process-generated data

Benefit Recipient History

Participation in labor market programmes

Jobseeker History

Integrated Employment Biographies

FDZ data

Establishment History Panel (BHP)

Sample of Integrated Labor Market Biographies (SIAB)

Linked Employer-Employee Data (LIAB)

IAB Establishment Panel (IABB)

Working and Learning in a Changing World (ALWA)

Panel Study ‘Labor Market and Social Security’ (PASS)

Surveys
Structure of ALWA–ADIAB

- (un-)employment history
- formal and non-formal educational history
- residential history
- partner, child-rearing and parental leave history
- military and alternative services etc.

ALWA survey data
- cross-sectional data
- numeracy and literacy test scores
- paradata
- linkage-related data

ALWA survey data
- (un-)employment history
- formal educational history
- residential history
- job-seeking history
- history of participation in active labour market policy measures

Administrative data
- yearly establishment information
Consent to record linkage in Germany

- Survey institute or interviewers are legally bound to inform respondents about the nature of information that is going to be matched.
- The interviewer has to ask for approval to that procedure explicitly.
- The respondent has to be able to make an informed decision.
- During the ALWA survey, consent was given by 9,531 (92%) of the respondents.
Identifiers of respondents

- ALWA lacks unique identifiers for a direct link to administrative records.
- Identifiers for matching:
  - first and last name
  - gender
  - day, month and year of birth
  - postal code, place name, street name and house number
- Sources of identifiers:
  - Field information (infas Institute for Applied Social Sciences)
  - Personal register data (IAB department IT Services and Information Management)
- Reference year for both sources: 2007
- Extensive standardization of identifiers
Subsequent steps of the linkage process

- **Exact matching**
  - Records from both data sources are directly matched based on all available fields.
  - Even smallest variations in spelling lead to a rejection of a potential match.

- **Probabilistic record linkage**
  - Comparison with software Merge ToolBox (MTB, v0.7). (cf. Schnell, Bachteler, and Reiher, 2005)
  - String similarity functions penalize deviations between identifiers.
  - Classification of record pairs into links, possible links and non-links (cf. Fellegi and Sunter, 1969)

- **Manual matching**
  - Independent manual review and classification of possible links by two staff members
  - Supervision and final decision in case of contradicting classifications by third person

⇒ Several variables indicating quality of link
Subsequent steps of the linkage process

- **Exact matching**
  - Records from both data sources are directly matched based on all available fields.
  - Even smallest variations in spelling lead to a rejection of a potential match.

- **Probabilistic record linkage**
  - Comparison with software Merge ToolBox (MTB, v0.7). (cf. Schnell, Bachteler, and Reiher, 2005)
  - String similarity functions penalize deviations between identifiers.
  - Classification of record pairs into links, possible links and non-links (cf. Fellegi and Sunter, 1969)

- **Manual matching**
  - Independent manual review and classification of possible links by two staff members
  - Supervision and final decision in case of contradicting classifications by third person

⇒ Several variables indicating quality of link
Subsequent steps of the linkage process

- Exact matching
  - Records from both data sources are directly matched based on all available fields.
  - Even smallest variations in spelling lead to a rejection of a potential match.
- Probabilistic record linkage
  - Comparison with software Merge ToolBox (MTB, v0.7). (cf. Schnell, Bachteler, and Reiher, 2005)
  - String similarity functions penalize deviations between identifiers.
  - Classification of record pairs into links, possible links and non-links (cf. Fellegi and Sunter, 1969)
- Manual matching
  - Independent manual review and classification of possible links by two staff members
  - Supervision and final decision in case of contradicting classifications by third person

⇒ Several variables indicating quality of link
Subsequent steps of the linkage process

- **Exact matching**
  - Records from both data sources are directly matched based on all available fields.
  - Even smallest variations in spelling lead to a rejection of a potential match.

- **Probabilistic record linkage**
  - Comparison with software Merge ToolBox (MTB, v0.7). (cf. Schnell, Bachteler, and Reiher, 2005)
  - String similarity functions penalize deviations between identifiers.
  - Classification of record pairs into links, possible links and non–links (cf. Fellegi and Sunter, 1969)

- **Manual matching**
  - Independent manual review and classification of possible links by two staff members
  - Supervision and final decision in case of contradicting classifications by third person

⇒ Several variables indicating quality of link
### Number of observations over all stages of linkage

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>$N_{N_c}$</th>
<th>$N_{N_r}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CATI respondents ($N_r$)</td>
<td>10404</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Consenting CATI respondents ($N_c$)</td>
<td>9531</td>
<td>100%</td>
<td>91.61%</td>
</tr>
<tr>
<td>Exact matches</td>
<td>5035</td>
<td>52.83%</td>
<td>48.39%</td>
</tr>
<tr>
<td>Exact and probabilistic matches (Jaro-Winkler)</td>
<td>7919</td>
<td>83.09%</td>
<td>76.11%</td>
</tr>
<tr>
<td>Exact, probabilistic and manual matches</td>
<td>8243</td>
<td>86.49%</td>
<td>79.23%</td>
</tr>
<tr>
<td>Matches with administrative spell data</td>
<td>8166</td>
<td>85.68%</td>
<td>78.49%</td>
</tr>
</tbody>
</table>
References


Manfred Antoni. *Linking survey data with administrative employment data: The case of the German ALWA survey*, forthcoming

Outlook

- Update of administrative data in ALWA-ADIAB in 2012
- Further efforts to improve the potential of the data:
  - Considering administrative address data from before and after 2007
    ⇒ How much observations can be gained by the additional step?
    ⇒ Does it affect the selectivity of the linked data set?
  - Respondents willing to participate in further panel waves were included in the National Educational Panel Study (NEPS, cf. Allmendinger et al., 2011).
    ⇒ Non-consenters were asked for consent again.

- Planned methodological analyses:
  - Evaluation of success of aided recall techniques in reducing recall error
  - Validation of educational histories in administrative data and evaluation of existing correction procedures
  - Analysis of social desirability bias in respondent behavior
Thank you for your attention

Manfred Antoni
Contact: Manfred.Antoni@iab.de