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

Erster ALWA-User-Workshop

Nürnberg

6. Dezember 2011

Manfred Antoni

Graduate Programme
Institute for Employment Research
and University Erlangen-Nuremberg
School of Business and Economics

IAB
GradAB

Outline

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 $\Rightarrow$ once a survey respondent is identified in the register data, the complete employment history is available.

Graduate Programme
Institute for Employment Research
and University Erlangen-Nuremberg
School of Business and Economics
IAB
GradAB

Graduate Programme
Institute for Employment Research
and University Erlangen-Nuremberg
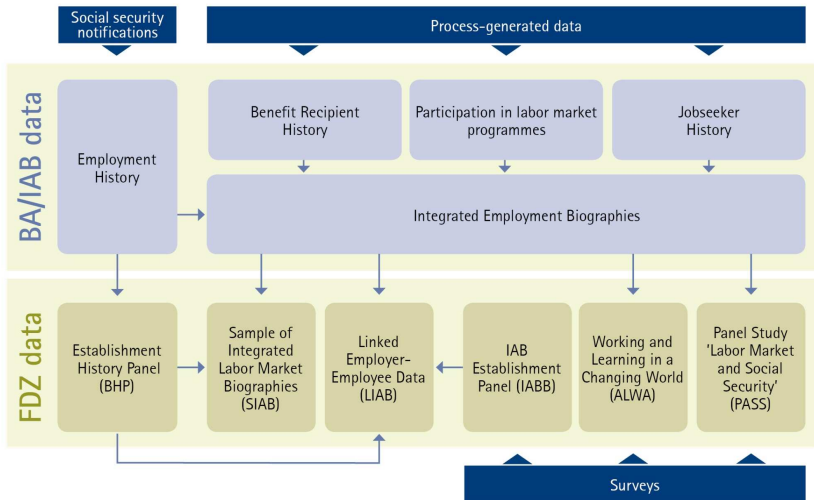School of Business and Economics

IAB
GradAB

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)
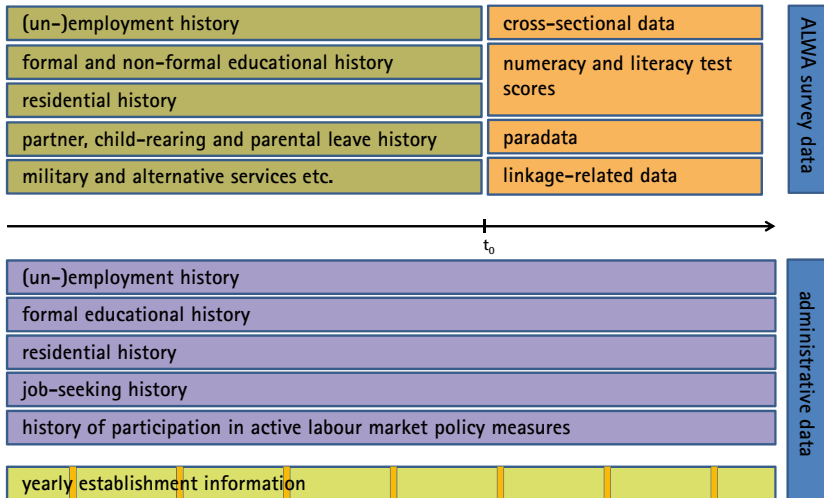
Administrative labor market data on establishments

- 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

Graduate Programme
Institute for Employment Research
and University Erlangen-Nuremberg
School of Business and Economics

IAB
GradAB

Structure of ALWA–ADIAB

(un-)employment history	cross-sectional data
formal and non-formal educational history	numeracy and literacy test scores
residential history	
partner, child-rearing and parental leave history	paradata
military and alternative services etc.	linkage-related data

ALWA survey data

t_0

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

yearly establishment information

administrative data

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

IAB
Graduate Programme
Institute for Employment Research
and University Erlangen-Nuremberg
School of Business and Economics
GradAB

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

Graduate Programme
Institute for Employment Research
and University Erlangen-Nuremberg
School of Business and Economics

IAB
GradAB

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

$\Rightarrow$ Several variables indicating quality of link

Number of observations over all stages of linkage

	N	$\frac{N}{N_c}$	$\frac{N}{N_r}$
CATI respondents (N_r)	10404		100%
Consenting CATI respondents (N_c)	9531	100%	91.61%
Exact matches	5035	52.83%	48.39%
Exact and probabilistic matches (Jaro-Winkler)	7919	83.09%	76.11%
Exact, probabilistic and manual matches	8243	86.49%	79.23%
Matches with administrative spell data	8166	85.68%	78.49%

References

Manfred Antoni, Peter Jacobebbinghaus, and Stefan Seth (2011). *ALWA-Befragungsdaten verknüpft mit administrativen Daten des IAB 1975-2009 (ALWA-ADIAB 7509)*. FDZ Datenreport 05/2011

Manfred Antoni and Stefan Seth (2012). "ALWA-ADIAB - linked individual survey and administrative data for substantive and methodological research". In: *Schmollers Jahrbuch. Journal of Applied Social Science Studies* forthcoming

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

Manfred Antoni. *Record linkage of the ALWA survey with administrative data of the Federal Employment Agency (ALWA-ADIAB): technical report*, 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

Graduate Programme
Institute for Employment Research
and University Erlangen–Nuremberg,
School of Business and Economics

Thank you for your attention

Manfred Antoni
Contact: Manfred.Antoni@iab.de