gesis

Leibniz Institute
for the Social Sciences

What's my wage again?

Comparing survey and administrative data to validate earnings measures.

J. Daikeler , B. Gauly, T. Gummer, and B. Rammstedt

2022 PIAAC Conference – 24/03/2022

Session "Linking PIAAC data to administrative data and other large-scale assessments "

Leibniz Association

In a Nutshell

Motivation

- Earning measures from surveys widely used as a basis for research and political recommendations *although* measures are error-prone.
- Administrative data as a "true measure" to validate survey information.

Research Question

- Do we find bias in regression results when using survey versus administrative data on earnings?

Main Results

- Bias in regression coefficients which highlights the importance to validate survey measures and findings with external data.

Importance of Survey Data

- Surveys are one of the most important data source for social science researchers.
- Earnings as an important and frequently used variable.
- Based on earnings information, research on, e.g.,
  - Social (in)equality (Atkinson, 2003),
  - Human capital models (Becker, 1962), and
  - Labor supply (Bound & Krueger, 1989).

- Policy recommendations and evaluations are based on this research

  → high quality standards for survey data are necessary.

Measurement Error in Survey Data

- Measurement error (ME) = difference between true and observed values.
- ME can arise due to respondents' characteristics, question and survey design, or interviewer presence.
- Earnings question are particularly error-prone:
  - Sensitive question.
  - Difficult task to recall detailed information.
  - Unclear earnings concept (gross/net, …).

- ME in earnings might obscure true economic relationships, bias surveys findings, and, thus, impact substantive conclusions.

Benefits of Data Linkage

- Data linkage: Extending survey data with administrative data.
- Two main benefits of data linkage (e.g., Calderwood & Lessof, 2009):
  - Extend survey data without increasing respondents‘ burden.
  - Assessing validity of survey measures and estimates.
- (German) administrative data considered as a less-error prone data source (e.g. Antoni & Bethmann, 2018; Sakshaug & Kreuter, 2012) :
  - Basis for calculating amount of social insurance and pension entitlement.

- Present study: Does the comparison of survey and administrative measures reveal a bias in estimation results?

Data I

- **Survey data:** Programme for the International Assessment of Adult Competencies (PIAAC) 2012 (OECD, 2013; Zabal, 2014).
- Focus: Adults cognitive skills and labor market participation.
- Representative sample of working-age population (16-65 years).
- German sample N = 5,465.

Data II

- **Administrative data:** Integrated Employment Biographies (IEB); provided by German Federal Employment Agency (see, e.g., Antoni et al., 2016).
- Individuals between 13 and 75 years.
- One of the following employment status since 1975:
  - Employment subject to social security,
  - Marginal part-time employment,
  - Benefit recipient (SGB II or III),
  - Officially registered as job-seeking, or
  - Participation in active labour market policies.
- Basis for calculating amount of social insurance and pensions.

Data III

- Both data sets contain information on individual earnings.
  - PIAAC: Hourly/weekly/monthly/annual/per piece
  - IEB: Daily
  - IEB * 30.5 to obtain monthly earnings in both data sets
- Challenge: multiple spells in IEB data match “PIAAC period”.

	ID	PIAAC_Earn
1	1111	3500
2	1112	2750

	ID	IEB_Earn	start	end	spell_nr
1	1111	3450	01.10.2011	31.01.2012	1
2	1111	2100	01.02.2012	31.03.2012	2
3	1112	2750	01.09.2011	30.09.2011	1
4	1112	3660	01.10.2011	31.12.2011	2
5	1112	.	01.01.2012	31.09.2012	3

- Strategy to identify correct period:
  1. Matching according to occupational group.
  2. Choose spell with the smallest wage difference.

→ 1,323 matched individuals.

Methods

1. Correlation me and "typical" covariates.

2. Correlation me and true earnings y^*.

3. Separate OLS models using survey and administrative earnings as a dependent variable.

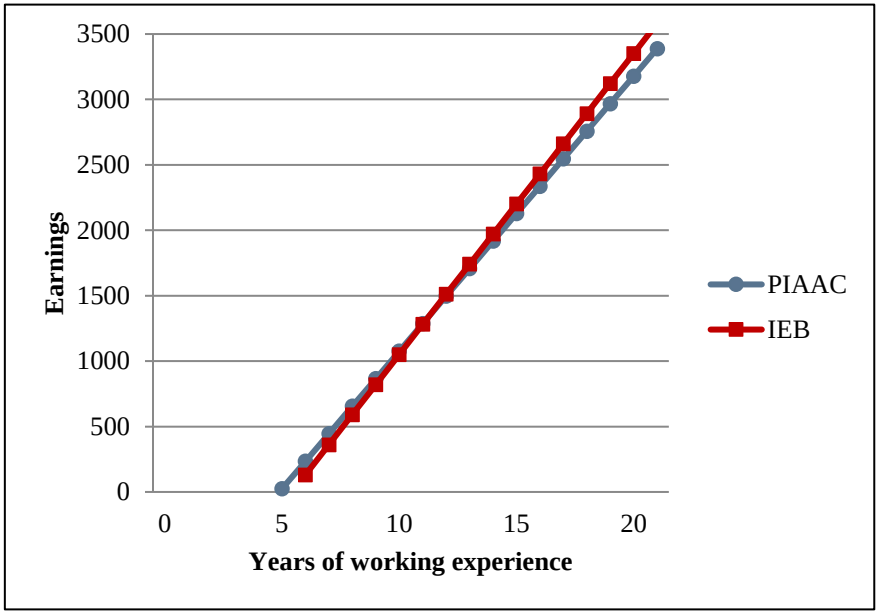
Results: ME correlates with…

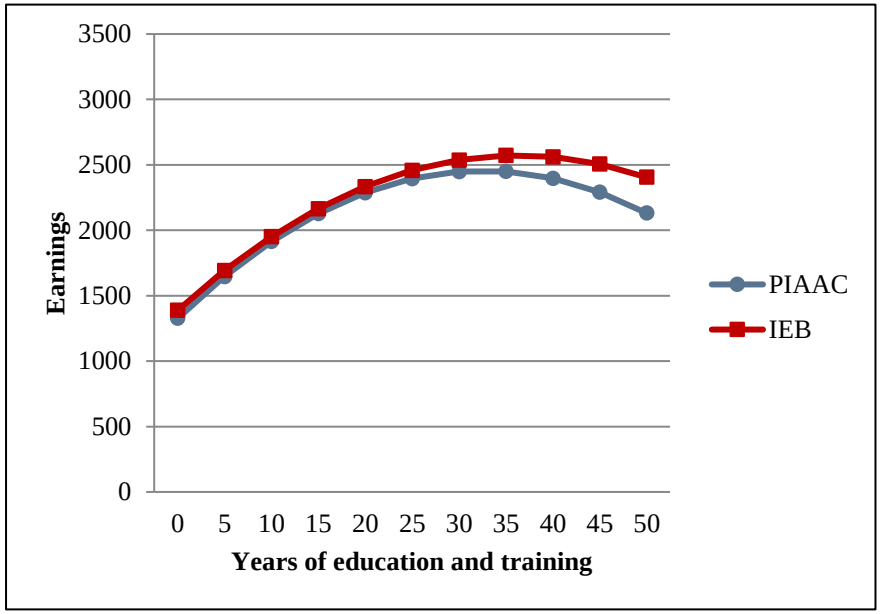
Covariate	Correlation
Female	-0.002
Age	-0.083**
Years of education and training	-0.077**
Working experience	-0.067*
Literacy	-0.056*
Numeracy	-0.066*
y*	-0.420***
N	1,232

Results: Biased Regressions

	Dependent variable: Monthly earnings		
Covariate	**PIAAC**	**IEB**	**Chow Test**
Female	-694.717***	-692.153***	0.01
	(70.777)	(78.695)	
Age	25.152***	28.914***	11.53***
	(2.716)	(2.936)	
Years of education and training	238.316***	257.274***	6.15*
	(14.437)	(15.563)	
Working experience	29.601***	32.628***	6.54*
	(2.724)	(3.002)	
Literacy	5.163***	5.938***	4.50*
	(0.884)	(0.970)	
Numeracy	7.001***	7.836***	5.22*
	(0.808)	(0.885)	
N	1,232	1,232	1,232

Results: Mincer Equation





Summary of Results

- Overall, more similarities than disparities:
  - General direction (positive/negative) of coefficients remains the same.
  - SD and significance level remain the same.
- But: Size of the (absolute value of the) coefficients is larger when using wages from the administrative data as DV.
- Chow test suggests statistically significant differences.
- Several robustness checks due to “spell challenges”.
  - IEB spell with smallest difference to PIAAC earnings for all respondents.
  - Only respondents who indicated monthly income in PIAAC.

    → Confirm our main results.

Conclusion I

- High similarity probably due to high accuracy/complexity in (earnings) data collection and high interview standards in PIAAC.
- Nevertheless some differences → shows that relying on error prone survey measure could lead to misjudging economic relationships, erroneous recommendations, and ineffective political implications.
- Data linkage as one opportunity to validate survey measures (and findings).
- Alternatively: Cross-checking results with further survey data.

Conclusion II

Of course, some limitations and room for further research:

- Linking survey and administrative data requires respondents' consent (*Daikeler et al. 2020*):
  - Our results cannot be generalized to population at large.
  - Evaluate strategies to increase linkage consent or develop correction methods.
- Matching "earnings period" between administrative data and survey data might be challenging.
- Lacking cross-country generalizability:
  - Quality of administrative data differs across countries.
  - Cross-national perspective as a next step in this line of research.

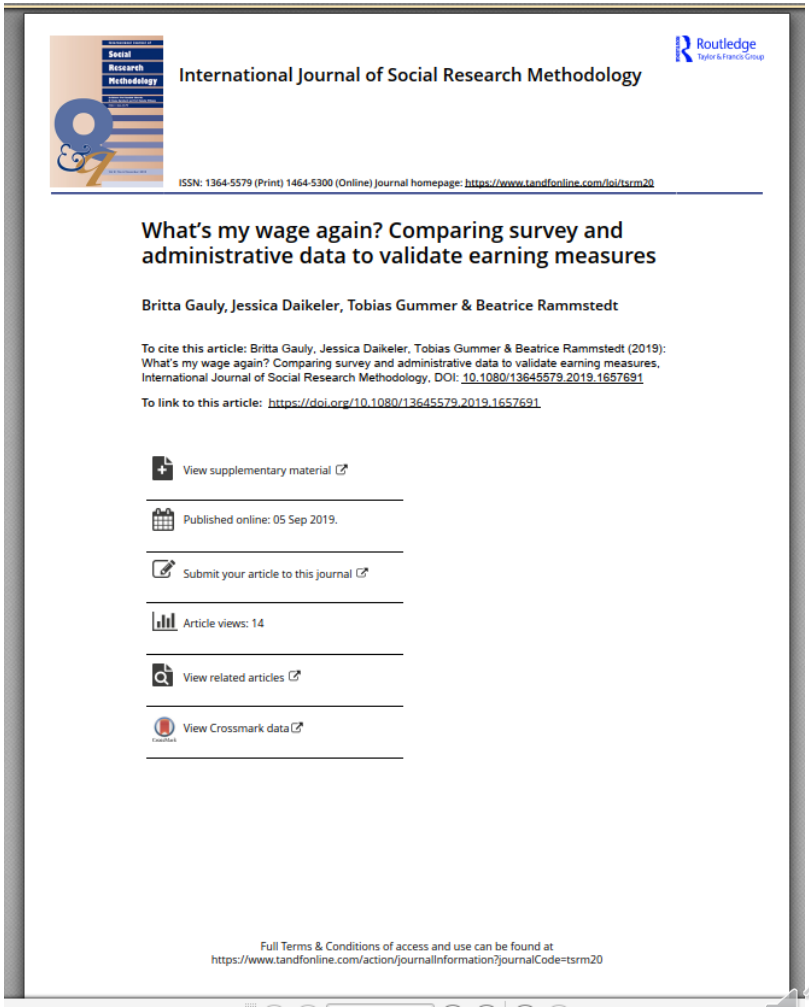
References

- Antoni, M., & Bethmann, A. (2019). PASS-ADIAB – linked survey and administrative data for research on unemployment and poverty. Jahrbücher Für Nationalökonomie Und Statistik, 239(4), 747–756.
- Antoni, M., Ganzer, A., & Vom Berge, P. (2016). Sample of Integrated Labor Market Biographies (SIAB) 1975 Data Report No. 4/2016. Nuremberg: Research Data Centre of the Federal Employment Agency in the Institute for Employment Research. Retrieved from http://www.iab.de/389/section.aspx/Publikation/k160729303
- Atkinson, A. C. (2003). Income inequality in OECD countries: Data and explanations. CESifo Working Paper 881. Munich, Germany: CESifo Group. Retrieved from http://hdl.handle.net/10419/76290
- Becker, G. S. (1962). Investment in human capital: A theoretical analysis. Journal of Political Economy, 70(5, Part 2), 9–49.
- Bound, J., & Krueger, A. B. (1989). The extent of measurement error in longitudinal earnings data: do two wrongs make a right?: NBER working paper series. Cambridge, MA: National Bureau of Economic Research. Retrieved from http://www.nber.org/papers/w2885
- Daikeler, Jessica, Britta Gauly, and Matthias Rosenthal. 2020. "Linking PIAAC data to individual administrative data: Insights from a German pilot project." In *Large-scale cognitive assessment: Analyzing PIAAC data*, edited by Débora B. Maehler, and Beatrice Rammstedt, Methodology of Educational Measurement and Assessment, 271-290 . Springer International Publishing. doi: https://doi.org/10.1007/978-3-030-47515-4_11. https://link.springer.com/chapter/10.1007/978-3-030-47515-4_11.
- OECD. (2013). OECD skills outlook 2013: First results from the survey of adult skills. Paris: Author.
- Sakshaug, J. W., & Kreuter, F. (2012). Assessing the magnitude of non-consent biases in linked survey and administrative data. Survey Research Methods, 6(2), 113–122.
- Zabal, A., Martin, S., Massing, N., Ackermann, D., Helmschrott, S., Barkow, I., & Rammstedt, B. (2014). PIAAC Germany 2012: Technical report. Muenster, Germany: Waxmann.

Would you like to learn more?

Gauly, B., Daikeler, J., Gummer, T., & Rammstedt, B. (2020). What's my wage again? Comparing survey and administrative data to validate earning measures. *International Journal of Social Research Methodology*, *23*(2), 215-228.

DOI: 10.1080/13645579.2019.1657691



International Journal of Social Research Methodology

ISSN: 1364-5579 (Print) 1464-5300 (Online) Journal homepage: https://www.tandfonline.com/loi/tsrm20

What's my wage again? Comparing survey and administrative data to validate earning measures

Britta Gauly, Jessica Daikeler, Tobias Gummer & Beatrice Rammstedt

To cite this article: Britta Gauly, Jessica Daikeler, Tobias Gummer & Beatrice Rammstedt (2019): What's my wage again? Comparing survey and administrative data to validate earning measures, International Journal of Social Research Methodology, DOI: 10.1080/13645579.2019.1657691

To link to this article: https://doi.org/10.1080/13645579.2019.1657691

View supplementary material

Published online: 05 Sep 2019.

Submit your article to this journal

Article views: 14

View related articles

View Crossmark data

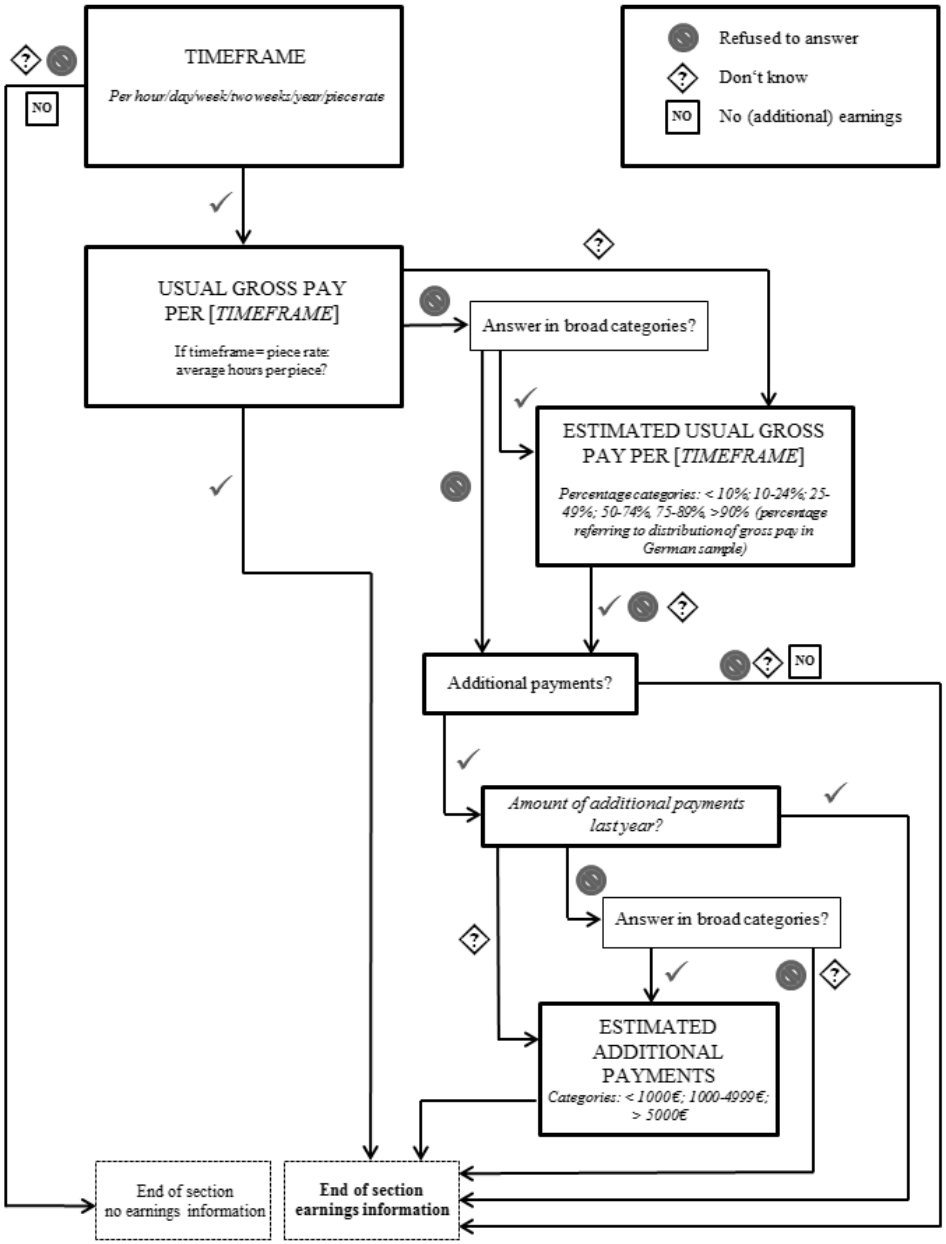
Full Terms & Conditions of access and use can be found at https://www.tandfonline.com/action/journalInformation?journalCode=tsrm20

Thank you for your attention!

gesis Leibniz Institute for the Social Sciences

Leibniz Association

@JeSs_Dalk


TIMEFRAME
Per hour/day/week/two weeks/year/piece rate
NO
USUAL GROSS PAY PER [TIMEFRAME]
If timeframe = piece rate: average hours per piece?
Answer in broad categories?
ESTIMATED USUAL GROSS PAY PER [TIMEFRAME]
Percentage categories: < 10%; 10-24%; 25-49%; 50-74%, 75-89%, >90% (percentage referring to distribution of gross pay in German sample)
Additional payments?
NO
Amount of additional payments last year?
Answer in broad categories?
ESTIMATED ADDITIONAL PAYMENTS
Categories: < 1000€; 1000-4999€; > 5000€
End of section no earnings information
End of section earnings information
Refused to answer
Don't know
NO
No (additional) earnings