

Interviewers' influence on bias in reported income

4th Baltic-Nordic Conference on Survey Statistics (BaNoCoSS 2015)
25 August 2015

Manfred Antoni
Basha Vicari
Daniel Bela

- Interaction between respondent and interviewer influences response behavior
- Answers to sensitive questions often affected by social desirability bias
- Income questions have very high sensitivity, with non-response rates ranging from 20-27% (Krumpal 2013)
- Growing literature on item non-response with income questions

⇔ So far little known about accuracy of reported income

⇔ Linked survey and administrative data enable us to validate responses

- Interaction between respondent and interviewer influences response behavior
 - Answers to sensitive questions often affected by social desirability bias
 - Income questions have very high sensitivity, with non-response rates ranging from 20-27% (Krumpal 2013)
 - Growing literature on item non-response with income questions
- ⇔ So far little known about accuracy of reported income
- ⇔ Linked survey and administrative data enable us to validate responses

- Interaction between respondent and interviewer influences response behavior
 - Answers to sensitive questions often affected by social desirability bias
 - Income questions have very high sensitivity, with non-response rates ranging from 20-27% (Krumpal 2013)
 - Growing literature on item non-response with income questions
- ⇔ So far little known about accuracy of reported income
- ⇔ Linked survey and administrative data enable us to validate responses

Research questions



- 1) What is the extent of income misreporting?
- 2) How do respondent characteristics influence the report behavior?
- 3) How do interviewer characteristics influence the report behavior?

- Hypotheses on influence of respondent characteristics:

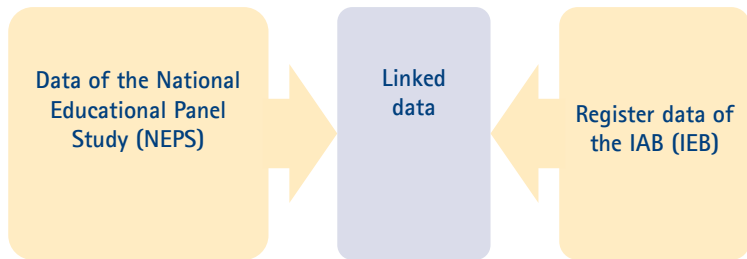
H1: Female respondents report more accurately

H2: More highly educated respondents report more accurately

- Hypotheses on influence of interviewer characteristics:

H3: More experienced interviewers produce more accurate reports

H4: Similarity between interviewer and respondent reduces misreporting



Data of the National Educational Panel Study (NEPS)



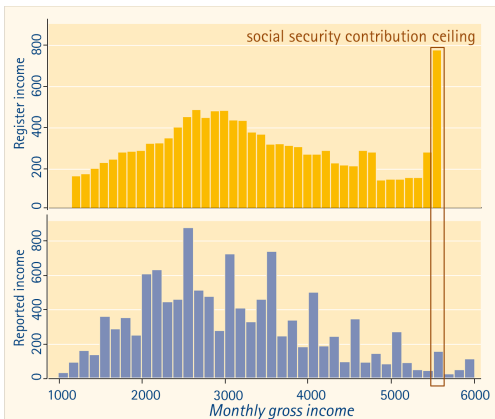
- NEPS Starting Cohort 6 (adults), waves 2 through 5, birth cohorts 1944-1986 (doi:10.5157/NEPS:SC6:5.1.0)
- N: 17,140
- CATI/CAPI with focus on educational history, also covering (un)employment, social background etc.
- Information on net and gross income for current job episodes
- Some paradata on interviewers and interview situation

- Daily longitudinal data on:
 - employment (since 1975)
 - registered unemployment (since 1975)
 - participation in labor market programs (since 2000)
 - registered job search activities (since 2000)
- Covering over 85% of the German labor force
- Mandatory social security notifications by employers on their dependent employees
 - ⇒ highly reliable information on gross income
- Consistent person identifier
 - ⇒ once a survey respondent is identified in the administrative data, complete employment history is available

- Record linkage of survey and administrative data using name, address, birth date, and sex of respondents
- Combination of deterministic and probabilistic linkage methods
- Informed consent to linkage from about 90% of respondents
- Linkage success rate: 91%

Comparison of frequency distributions

- Administrative income evenly distributed, only heap at the cut-off point of social security contribution ceiling
- Heaping across whole distribution of reported income



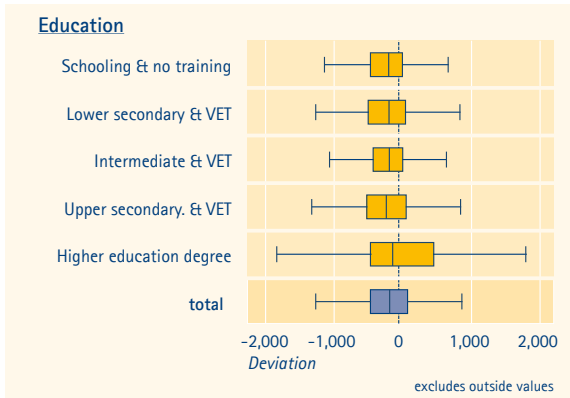
- Only episodes of dependent, full-time employment
- Only employment episodes that are ongoing at or have ended shortly before the time of the interview
- No spells with implausible or censored income

Table: Comparison of register and reported income

(N= 12,486)	median	s.d.	min	max
Register income	3,228	1,191	1,217	5,598
Reported income	3,000	1,784	980	20,000

Bivariate results: respondents

- Respondents with higher education degree show highest deviation in both directions
- Below that level of education very similar deviations



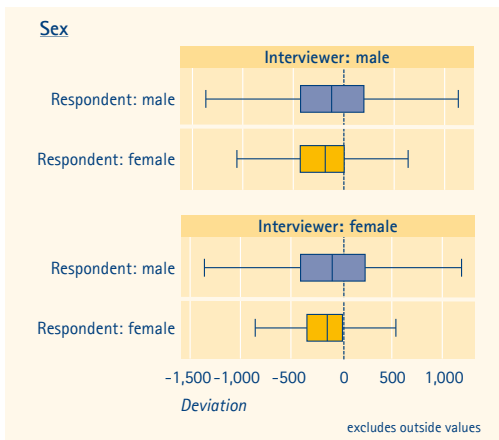
Bivariate results: interviewers

- Interviewers' experience only weakly affects report accuracy
- Least experienced interviewers produce highest deviation



Bivariate results: interaction of characteristics

- Interviewers' sex does not produce differences in report accuracy
- Male respondents vary more in report accuracy



- Tables show results from logistic regression.
- Dependent variables: indicator whether absolute difference is...
 - Model 1: larger than one standard deviation of administrative income (share: 8%)
 - Model 2: more than 20% larger than administrative income (share: 21%)

Results of multivariate regression II

Respondent	Model 1		Model 2	
	coef	z	coef	z
Female	-0.291*	-1.97	-0.079	-0.93
Lower second., voc. training (ref.:no train.)	-0.647*	-2.34	-0.444**	-3.11
Intermediate, voc. training	-0.279	-1.03	-0.365**	-2.62
Upper secondary, voc. training	-0.001	0.00	-0.316*	-2.04
Higher education degree	-0.018	-0.07	-0.473**	-3.25
Inc.: <2000 EUR (ref.:3000 - <4000 EUR)	-0.896***	-3.88	-0.200	-1.53
Inc.: 2000 - <3000 EUR	-1.364***	-5.9	-0.331***	-3.70
Inc.: 4000 - <5000 EUR	0.576***	4.38	0.183*	2.05
Inc.: 5000 EUR and above	1.647***	11.74	0.687***	6.79
Big 5: Extraversion	-0.011	-0.22	0.024	0.73
Big 5: Agreeableness	-0.345***	-4.23	-0.118*	-2.07
Big 5: Conscientiousness	0.345***	4.66	0.190***	4.08
Big 5: Neuroticism	-0.105	-1.46	-0.065	-1.42
Big 5: Openness to experience	0.002	0.04	-0.019	-0.59

Results of multivariate regression III

Interviewer	Model 1		Model 2	
	coef	z	coef	z
I: aged 30-49 (ref.:below 30)	0.138	0.85	0.018	0.15
I: aged 50-65	0.132	0.77	0.095	0.77
I: aged older than 65	0.215	0.92	-0.002	-0.01
I: intermediate (ref.:lower secondary)	0.066	0.45	0.004	0.03
I: upper secondary	-0.109	-0.95	-0.094	-1.02
I: exp. 2-3 years (ref.:<2 years)	0.000	0.00	0.050	0.59
I: exp. 4-5 years	0.055	0.41	0.050	0.51
I: exp. 6 or more years	-0.09	-0.62	-0.033	-0.33
I: running no. of interviews per wave	0.002	1.91	0.001	1.24
R: Male # I: female (ref.:R: male#I: male)	0.024	0.23	-0.018	-0.23
R: Female # I: female	-0.322	-1.62	-0.101	-0.91

Results of multivariate regression IV

Interview situation	Model 1		Model 2	
	coef	z	coef	z
CATI (ref.:CAPI)	-0.291	-1.56	-0.200	-1.67
2010/2011 (3rd wave)	0.295	1.36	0.046	0.35
2011/2012 (4th wave)	0.157	1.06	0.133	1.41
2012/2013 (5th wave)	0.303	1.53	0.144	1.20
Constant	-2.935***	-4.89	-1.367***	-3.38
Observations	8060		8060	
pseudo R ²	0.17		0.028	
BIC	4078.305		8065.351	

Source: NEPS-SC6-ADIAB; omitted respondent characteristics: age classes, country of birth, paid overtime; ***/**/* indicates significance at the 0.1/1/5% level; robust standard errors based on 542 interviewers as clusters; z-statistics in parentheses.

- Average deviation of reported income from administrative income: underestimation of about 200 EUR (< 10% of median administrative income)
- Descriptive evidence shows only small variation of deviation across subgroups
- Higher female report accuracy corroborates H1
- Least qualified show highest likelihood of deviation, supporting H2
- Results on H1 and H2 strongly depend on measurement of deviation / specification of dependent variable
- Multivariate results hint at negligible influence of interviewer characteristics
⇒ no support for H3 or H4

Further steps:

- Consider direction of misreporting in a multinomial model
- Include other interaction terms between characteristics of respondents and interviewers to measure similarity
- Run separate analyses by mode
- Experiment with specification of dependent variable
- Quantile regression that considers income quantiles
- Face facts and change title of paper

Further steps:

- Consider direction of misreporting in a multinomial model
- Include other interaction terms between characteristics of respondents and interviewers to measure similarity
- Run separate analyses by mode
- Experiment with specification of dependent variable
- Quantile regression that considers income quantiles
- Face facts and change title of paper

Thank you for your attention!

Manfred Antoni
manfred.antoni@iab.de

Basha Vicari
basha.vicari@iab.de

Daniel Bela
daniel.bela@lifbi.de

Overview of IAB data

