Institute for Employment Research

The Research Institute of the Federal Employment Agency

Item Sum: A New Technique for Asking Quantitative Sensitive Questions

AAPOR 67th Annual Conference

Orlando, May 18th, 2012

Questions on Sensitive Topics and Social Desirability Bias

Antje Kirchner

Mark Trappmann (IAB)

Ivar Krumpal (University of Leipzig)

Ben Jann (University of Bern)

Research interests

1. Develop a privacy preserving survey-based technique to measure continuous sensitive characteristics

2. Derive estimators to compare with standard (direct) questioning (‘more-is-better assumption’) and to estimate individual expectations

3. Application in the context of a CATI study on undeclared work

Problems associated with sensitive topics

"A question is sensitive when it asks for a socially undesirable answer, when it asks in effect, that the respondent admits he or she has violated a social norm."

(Tourangeau and Yan 2007:860)

- Biased estimates of undeclared work via direct questioning as a consequence of non-random (Tourangeau and Smith 1996: 276)
  - partial nonresponse (break-offs),
  - (item-)nonresponse (refusal),
  - misreporting (here: underreporting)

Measurement of undeclared work, a brief review

- Wording
  - forgiving wording (Mummert and Schneider 2001),
  - loading of questions (Mummert and Schneider 2001),
  - asking about long periods/distant past (Lamnek et al. 2000),
  - paraphrasing (Wolff 1991; Eurobarometer 2007),

  examples: the “Casual Approach”, “Everybody Approach” or the “Other People Approach” etc. (Barton 1958)

- Mode change
  - self-administration methods

→ Variation of estimates from surveys:

12.5 hours (Eurobarometer 2007) to 7.3 hours (Feld and Larsen 2008)

→ Recommendation to “consider alternatives to standard questioning […]”

(Bradburn et al. 2004: 81; cf. also Boockmann et al. 2010: 100)

The idea: Increase perceived privacy protection

- indirect survey-based estimation techniques that minimize respondent's feeling of jeopardy

- by 'scrambling' the individual response in such a way, that it is impossible for the interviewer or the researcher to know the true answer, i.e. introducing random 'noise'

- examples for binary items comprise the randomized response technique, the item count technique, etc.

(Warner 1965; Droitcour et al. 1991)

Implementation of the item sum technique (IST)

Group LL (Long List)	Group SL (Short List)
C1: How many hours did you watch TV last week? S1: How many hours per week do you usually engage in undeclared work?	C1: How many hours did you watch TV last week?
Please sum up the answer to both questions, please, do not report individual answers.	
Group LL (Long List)	**Group SL (Short List)**
C2: How high are your monthly costs for your apartment respectively your house? S2: How high are your usual earnings per month engaging in undeclared work?	C2: How high are your monthly costs for your apartment respectively your house?
Please sum up the answer to both questions, please, do not report individual answers.	

Preceded by brief definitions of undeclared work and instructions regarding the technique if in treatment group. Embedded in items on employment, predictors of undeclared work and socio-demographics.

Project outline & study design

Institute for Employment Research
The Research Institute of the Federal Employment Agency

Mode — **CATI Study**

Randomization — split-ballot experiment

Sample
(n = 3,211)

Structure
(Hours & Earnings)

Direct Questioning

Item Sum Technique

Short List

Long List

Two random samples from federal employment agency registers (RR1: 17.5 %):

- register sample of employees (18-70) and
- register sample of basic income support recipients (18-64)

Estimators: Aggregate estimates

- Let S be the sensitive item of interest and C be the non-sensitive control item. Observed is:

$$Y_i = \begin{cases} S_i + C_i & \text{if } i \in LL \text{ (Long List)} \\ C_i & \text{if } i \in SL \text{ (Short List)} \end{cases}$$

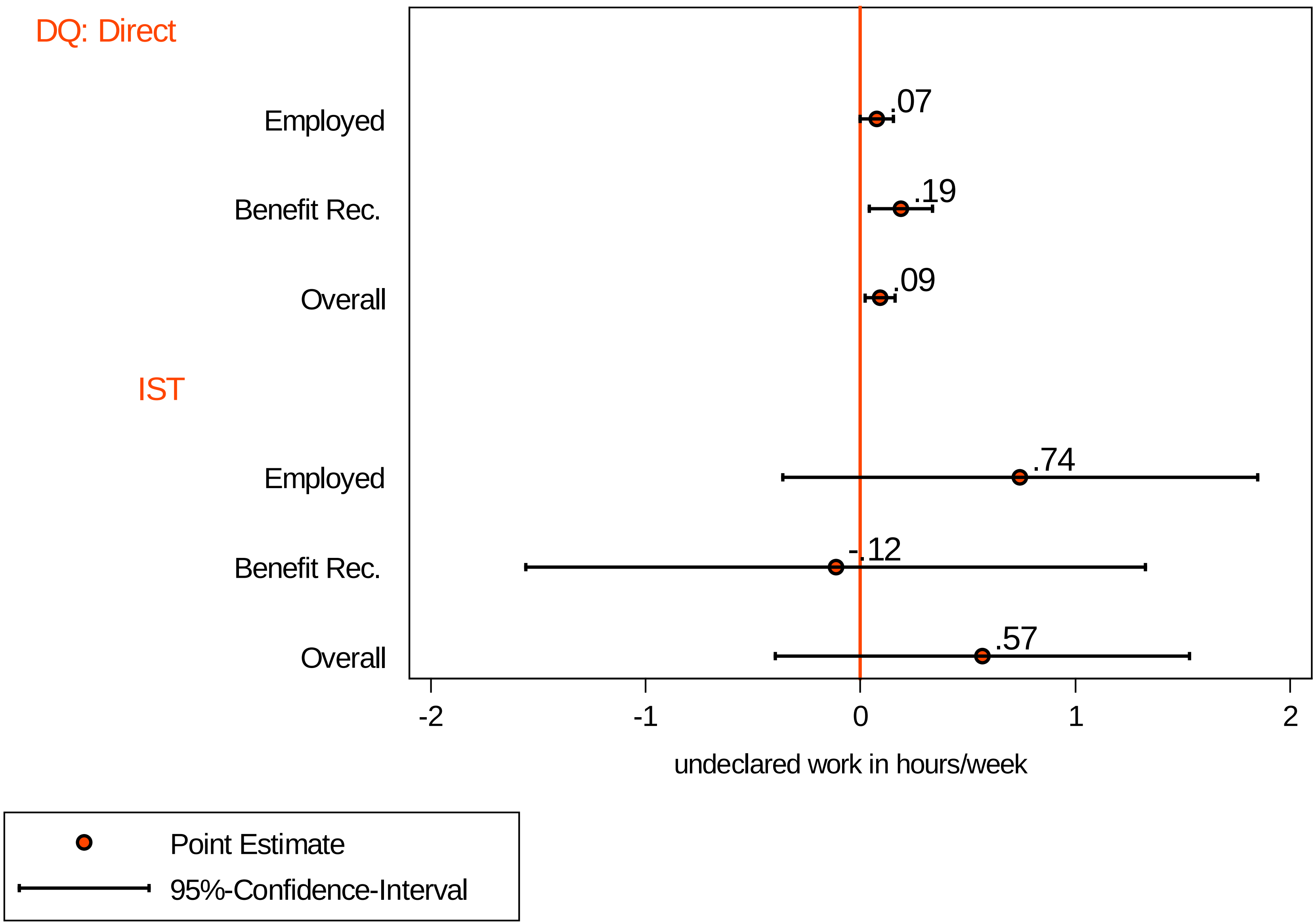
- The mean difference of Y between the two groups is an unbiased estimate of the population mean of S:

$$\hat{E}(S) = \hat{E}_{LL}(Y) - \hat{E}_{SL}(Y)$$

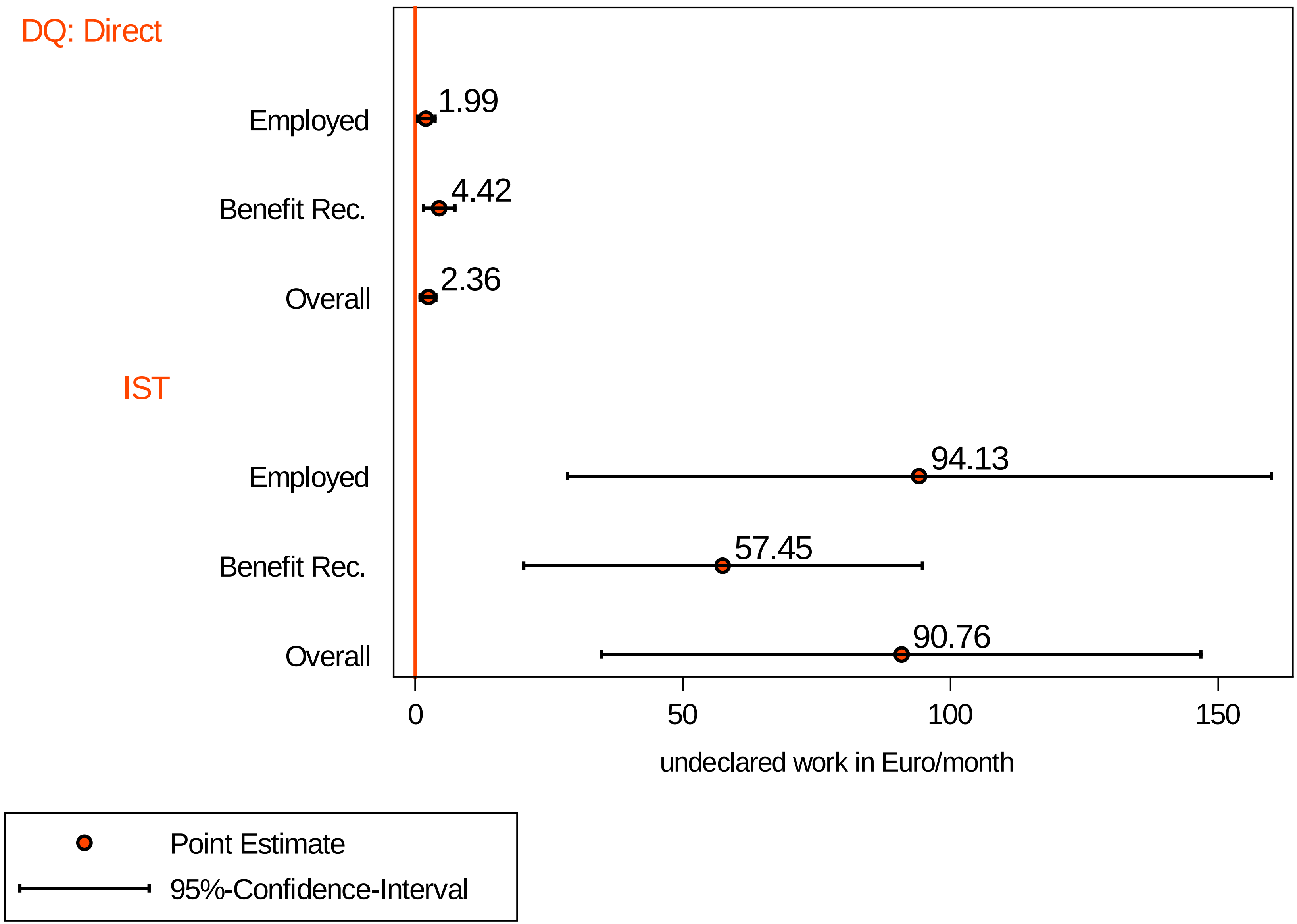
- The sampling variance of the mean estimate of S is given as:

$$\hat{V}\left[\hat{E}(S)\right] = \hat{V}\left[\hat{E}_{LL}(Y)\right] + \hat{V}\left[\hat{E}_{SL}(Y)\right]$$

Aggregate estimates: Hours per week



Aggregate estimates: Earnings per month



Estimators: Least squares IST

- Let $T_i = \begin{cases} 1 & \text{if } i \in LL \\ 0 & \text{if } i \in SL \end{cases}$ so that $Y_i = T_i S_i + C_i$ (1)

- Suppose that *S* and *C* both depend linearly on a vector of covariates X and Z (including a constant):

$$S_i = X_i'\beta + \upsilon_i, \; E(\upsilon_i) = 0 \tag{2a}$$

$$C_i = Z_i'\gamma + \nu_i, \; E(\nu_i) = 0 \tag{2b}$$

Estimators: Least squares IST

- then we can model Y as:

$$\begin{aligned} Y_i &= T_i(X_i'\beta + \upsilon_i) + (Z_i'\gamma + \nu_i) \\ &= T_i \cdot X'\beta + Z_i'\gamma + \varepsilon_i \end{aligned} \qquad (3)$$

$$\begin{aligned} \text{with } \varepsilon_i &= T_i \cdot \upsilon_i + \nu_i \\ \text{hence } E(\varepsilon_i) &= 0 \quad \text{and} \quad \sigma_\varepsilon = \sigma_\upsilon^2 + \sigma_\nu^2 + 2\rho\sigma_\upsilon\sigma_\nu \end{aligned}$$

- Being a simple mean difference, we can recast the item-sum estimator using linear regression:

$$Y_i = \gamma_0 + \beta_0 T_i + \varepsilon_i, \; E(\varepsilon_i) = 0$$

Estimators: Least squares IST

- Extension to include a third sample of respondents for which the sensitive item S was measured via direct questioning (DQ_i) yields

$$Y_i = (LL_i + DQ_i) \cdot X_i{}' \beta + SL_i \cdot Z_i' \gamma + \varepsilon_i$$

- Test of the item sum technique by including indicator LL (or DQ) in *X*.
- Inclusion of interactions between LL and other variables can be used to evaluate, whether effects of regressors differ between techniques.

So what? Item sum, a formal test

Isreg (Jann 2011)	**Hours model 1a**		**Earnings model 2a**	
Methods effect: Item sum (Ref. DQ)	.48 (.49)		**88.41**** (28.55)	
Constant	**.09*** (.04)		**2.36**** (.73)	
N	3,072		3,003	

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

Isreg (Jann 2011)	**Hours model 1a**		**Earnings model 2a**	
Methods effect: Item sum (Ref. DQ)	.48 (.49)		**88.41**** (28.55)	
Constant	**.09*** (.04)		**2.36**** (.73)	
N	3,072		3,003	

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

Isreg (Jann 2011)	**Hours model 1a**	**Hours model 1b**	**Earnings model 2a**	**Earnings model 2b**
Methods effect: Item sum (Ref. DQ)	.48 (.49)	.67 (.57)	**88.41**** (28.55)	**92.15**** (33.52)
Sample effect: Benefit recipient (Ref. employees)		.11 (.09)		2.43 (1.73)
Interaction: Item sum * benefit		-.97 (.93)		-39.12 (38.56)
Constant	**.09*** (.04)	.07 (.04)	**2.36**** (.73)	**1.99*** (.82)
N	3,072	3,072	3,003	3,003

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

Institute for Employment Research
The Research Institute of the Federal Employment Agency
IAB

Isreg (Jann 2011)	Hours model 1a	Hours model 1b	Earnings model 2a	Earnings model 2b
Methods effect: Item sum (Ref. DQ)	.48 (.49)	.67 (.57)	**88.41**** (28.55)	**92.15**** (33.52)
Sample effect: Benefit recipient (Ref. employees)		.11 (.09)		2.43 (1.73)
Interaction: Item sum * benefit		-.97 (.93)		-39.12 (38.56)
Constant	**.09*** (.04)	.07 (.04)	**2.36**** (.73)	**1.99*** (.82)
N	3,072	3,072	3,003	3,003

Standard errors in parentheses; * $p<0.05$, ** $p<0.01$, *** $p<0.001$; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

Isreg (Jann 2011)	**Hours model 1a**	**Hours model 1b**	**Earnings model 2a**	**Earnings model 2b**
Methods effect: Item sum (Ref. DQ)	.48 (.49)	.67 (.57)	**88.41**** (28.55)	**92.15**** (33.52)
Sample effect: Benefit recipient (Ref. employees)		.11 (.09)		2.43 (1.73)
Interaction: Item sum * benefit		-.97 (.93)		-39.12 (38.56)
Constant	**.09*** (.04)	.07 (.04)	**2.36**** (.73)	**1.99*** (.82)
N	3,072	3,072	3,003	3,003

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

Isreg (Jann 2011)	**Hours model 1a**	**Hours model 1b**	**Earnings model 2a**	**Earnings model 2b**
Methods effect: Item sum (Ref. DQ)	.48 (.49)	.67 (.57)	**88.41**** (28.55)	**92.15**** (33.52)
Sample effect: Benefit recipient (Ref. employees)		.11 (.09)		2.43 (1.73)
Interaction: Item sum * benefit		-.97 (.93)		-39.12 (38.56)
Constant	**.09*** (.04)	.07 (.04)	**2.36**** (.73)	**1.99*** (.82)
N	3,072	3,072	3,003	3,003

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

lsreg (Jann 2011)	Hours model 1a	Hours model 1b	Earnings model 2a	Earnings model 2b
Methods effect: Item sum (Ref. DQ)	.48 (.49)	.67 (.57)	**88.41**** (28.55)	**92.15**** (33.52)
Sample effect: Benefit recipient (Ref. employees)		.11 (.09)		2.43 (1.73)
Interaction: Item sum * benefit		-.97 (.93)		-39.12 (38.56)
Constant	**.09*** (.04)	.07 (.04)	**2.36**** (.73)	**1.99*** (.82)
N	3,072	3,072	3,003	3,003

Standard errors in parentheses; * p<0.05, ** p<0.01, *** p<0.001; design weights applied; coefficients for the C item are not reported.

So what? Item sum, a formal test

Isreg (Jann 2011)	Hours model 1a	Hours model 1b	Earnings model 2a	Earnings model 2b
Methods effect: Item sum (Ref. DQ)	.48 (.49)	.67 (.57)	**88.41**** (28.55)	**92.15**** (33.52)
Sample effect: Benefit recipient (Ref. employees)		.11 (.09)		2.43 (1.73)
Interaction: Item sum * benefit		-.97 (.93)		-39.12 (38.56)
Constant	**.09*** (.04)	.07 (.04)	**2.36**** (.73)	**1.99*** (.82)
N	3,072	3,072	3,003	3,003

Standard errors in parentheses; * $p<0.05$, ** $p<0.01$, *** $p<0.001$; design weights applied; coefficients for the C item are not reported.

Summary

- generalizing the item count technique we have presented a new privacy preserving technique for metric sensitive items and applied it in a study on undeclared work
- we have derived point estimators and regression estimators for IST variables
- results indicate that the item sum technique (IST) can be fruitful in yielding higher estimates of the socially undesirable behaviour than direct questioning

Limitations

- differential item nonresponse between treatment conditions contaminates randomization

- low statistical power, i.e. large standard errors of point estimates and regression coefficients: Trade-off with privacy preservation

- modelling issues:
  - natural lower bound of 0 for individual values of S not accounted for (hurdle and zero inflated models)
  - transformation of dependent variable

Implications and future work

However, due to the following reasons we suggest further inquiries to fully understand the mechanisms at work before implementing this new technique in labor market surveys:

- choice of the innocuous item
- test of assumptions required
- efficiency concerns
- comparison to other methods

References

Barton, A.H (1958): “Asking The Embarrassing Question” *Public Opinion Quarterly* 22:67-68.

Boockmann, B.; Döhrn, R.; Groneck, M. & Verbeek, H. (2010): *Abschätzung des Ausmaßes der Schwarzarbeit.* Report. Institut für Angewandte Wirtschaftsforschung e.V. (IAW) und Rheinisch-Westfälisches Institut für Wirtschaftsforschung e.V. (RWI).

Bradburn, N.; Sudman, S. & B. Wansink (2004): *Asking Questions*. Revised Edition. San Francisco:Jossey-Bass.

Droitcour, J.; Caspar, R.A.; Hubbard, M.L.; Parsley, T.L.; Visscher, W. & Ezzati, T.M. (1991): “The item count technique as a method of indirect questioning: A review of its development and a case study application” In *Measurement errors in surveys*, eds. P.P. Biemer, R.M. Groves, L.E. Lyberg, N.A. Mathiowetz & S. Sudman, p. 185-210. Wiley.

European Commission (2007): *Undeclared Work in the European Union*. Report. Special Eurobarometer 284 Wave 67.3.

Feld, L.P. & C. Larsen (2008): ““Black” Activities Low in Germany 2006.” *News from the Rockwool Foundation* Research Unit, Copenhagen.

Jann, B. (2011): “isreg: Stata module to estimate a regression for item sum data” Unpublished.

Lamnek , S.; Olbrich, G. & Schäfer, W.J. (2000): *Tatort Sozialstaat: Schwarzarbeit, Leistungsmissbrauch, Steuerhinterziehung und ihre Hintergründe*. Opladen.

Mummert , A. & Schneider, F. (2001): “German shadow economy: parted in a united Germany?” *Finanzarchiv*, 58(3):286-316.

Tourangeau, R. & Smith, T.W. (1996): “Asking sensitive questions. The impact of data collection mode, question format and question context” *Public Opinion Quarterly* 60:275-304.

Tourangeau, R. & Yan, T. (2007): “Sensitive questions in surveys.” *Psychological Bulletin* 133:859-883.

Warner, S.L. (1965): “Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias” *Journal of the American Statistical Association* 60:63-69.

Wolff, K. (1991): *Schwarzarbeit in der Bundesrepublik Deutschland. Eine mikroanalytische Untersuchung. Sonderforschungsbereich 3 der Universitäten Frankfurt und Mannheim*. Frankfurt/New York.