

Changing from PAPI to CAPI - A longitudinal study of Mode-Effects based on an Experimental Design in the SOEP

*Jörg-Peter Schräpler**, *Jürgen Schupp***
*and Gert G. Wagner****

* Ruhr-University Bochum and DIW Berlin, ** DIW Berlin, FU Berlin and HWK

*** DIW Berlin and Berlin University of Technology

European Conference on Quality in Survey Statistics,
24th - 26th April 2006

Outline

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP
 - ★ The experimental design of Sample E in the SOEP

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP
 - ★ The experimental design of Sample E in the SOEP
- Mode effects on data quality - Hypotheses

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP
 - ★ The experimental design of Sample E in the SOEP
- Mode effects on data quality - Hypotheses
- Empirical Analysis

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP
 - ★ The experimental design of Sample E in the SOEP
- Mode effects on data quality - Hypotheses
- Empirical Analysis
 - ★ Unit-Nonresponse in the following wave

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP
 - ★ The experimental design of Sample E in the SOEP
- Mode effects on data quality - Hypotheses
- Empirical Analysis
 - ★ Unit-Nonresponse in the following wave
 - ★ Item-Nonresponse

Outline

- Introduction: Computer assisted Personal interviewing (CAPI)
- Previous experiences on CAPI in surveys
- Data collection methods in the SOEP
 - ★ The experimental design of Sample E in the SOEP
- Mode effects on data quality - Hypotheses
- Empirical Analysis
 - ★ Unit-Nonresponse in the following wave
 - ★ Item-Nonresponse

Introduction: Computer assisted Personal interviewing (CAPI)

Introduction: Computer assisted Personal interviewing (CAPI)

- CAPI is an increasingly viable alternative for data collection in survey research

Introduction: Computer assisted Personal interviewing (CAPI)

- CAPI is an increasingly viable alternative for data collection in survey research
- Interviewer visit respondents with a portable computer and conduct a face-to-face interview using the computer

Introduction: Computer assisted Personal interviewing (CAPI)

- CAPI is an increasingly viable alternative for data collection in survey research
- Interviewer visit respondents with a portable computer and conduct a face-to-face interview using the computer
- after the interview the data are sent to a central computer

Introduction: Computer assisted Personal interviewing (CAPI)

- CAPI is an increasingly viable alternative for data collection in survey research
 - Interviewer visit respondents with a portable computer and conduct a face-to-face interview using the computer
 - after the interview the data are sent to a central computer
- ⇒ First tested in Europe by Statistics Sweden (Damielsson/Marstad 1982) in 1982 and by Statistics Netherland in 1984 (Bemelmans-Spork/Sikkel 1985)

Introduction: Computer assisted Personal interviewing (CAPI)

- CAPI is an increasingly viable alternative for data collection in survey research
 - Interviewer visit respondents with a portable computer and conduct a face-to-face interview using the computer
 - after the interview the data are sent to a central computer
- ⇒ First tested in Europe by Statistics Sweden (Damielsson/Marstad 1982) in 1982 and by Statistics Netherland in 1984 (Bemelmans-Spork/Sikkel 1985)

⇒ The first national household survey which used CAPI for all its data collection was the Netherlands Labor Force Survey 1987 (van Bastelaer/Kerssemakers/Sikkel 1988).

- ⇒ The first national household survey which used CAPI for all its data collection was the Netherlands Labor Force Survey 1987 (van Bastelaer/Kerssemakers/Sikkel 1988).
- Since that time the use of CAPI grew rapidly

- ⇒ The first national household survey which used CAPI for all its data collection was the Netherlands Labor Force Survey 1987 (van Bastelaer/Kerssemakers/Sikkel 1988).
- Since that time the use of CAPI grew rapidly
 - An important challenge is the step from PAPI to CAPI with an ongoing panel study (for example the Socio-economic Panel SOEP).

Reasons for CAPI

Reasons for CAPI

- Reduction in the time elapsed between fieldwork and the availability of the data for analysis (Martin et al. 1993).

Reasons for CAPI

- Reduction in the time elapsed between fieldwork and the availability of the data for analysis (Martin et al. 1993).
⇒ Most frequently given reason for adopting CAPI mentioned in the literature (Nicholls and De Leeuw 1996).

Reasons for CAPI

- Reduction in the time elapsed between fieldwork and the availability of the data for analysis (Martin et al. 1993).
 - ⇒ Most frequently given reason for adopting CAPI mentioned in the literature (Nicholls and De Leeuw 1996).
- Improvement in data quality:

Reasons for CAPI

- Reduction in the time elapsed between fieldwork and the availability of the data for analysis (Martin et al. 1993).
 - ⇒ Most frequently given reason for adopting CAPI mentioned in the literature (Nicholls and De Leeuw 1996).
- Improvement in data quality:
 - ⇒ Fewer instances of missing data, because interviewer cannot make routing errors (Sebestik et al. 1988, Olsen 1992)

Reasons for CAPI

- Reduction in the time elapsed between fieldwork and the availability of the data for analysis (Martin et al. 1993).
 - ⇒ Most frequently given reason for adopting CAPI mentioned in the literature (Nicholls and De Leeuw 1996).
- Improvement in data quality:
 - ⇒ Fewer instances of missing data, because interviewer cannot make routing errors (Sebestik et al. 1988, Olsen 1992)
- Cost saving: Not a common outcome (sizeable investment in hardware and development work) (Couper/Nicholls 1998)

Reasons for CAPI

- Reduction in the time elapsed between fieldwork and the availability of the data for analysis (Martin et al. 1993).
 - ⇒ Most frequently given reason for adopting CAPI mentioned in the literature (Nicholls and De Leeuw 1996).
- Improvement in data quality:
 - ⇒ Fewer instances of missing data, because interviewer cannot make routing errors (Sebestik et al. 1988, Olsen 1992)
- Cost saving: Not a common outcome (sizeable investment in hardware and development work) (Couper/Nicholls 1998)

Previous experiences with CAPI

Previous experiences with CAPI

- Concern about Unit-Nonresponse:

Previous experiences with CAPI

- Concern about Unit-Nonresponse:
 - ⇒ Some concern that CAPI respondent would object to having their information stored on a computer.

Previous experiences with CAPI

- Concern about Unit-Nonresponse:
 - ⇒ Some concern that CAPI respondent would object to having their information stored on a computer.
 - ⇒ Studies found no significant differences in refusal rates (Baker et al. 1995).

Previous experiences with CAPI

- Concern about Unit-Nonresponse:
 - ⇒ Some concern that CAPI respondent would object to having their information stored on a computer.
 - ⇒ Studies found no significant differences in refusal rates (Baker et al. 1995).
- Broad respondent acceptance (Baker 1992, De Leeuw 1995): most reactions are neutral or positive.

Previous experiences with CAPI

- Concern about Unit-Nonresponse:
 - ⇒ Some concern that CAPI respondent would object to having their information stored on a computer.
 - ⇒ Studies found no significant differences in refusal rates (Baker et al. 1995).
- Broad respondent acceptance (Baker 1992, De Leeuw 1995): most reactions are neutral or positive.
- Generally broad interviewer acceptance of CAPI (De Leeuw 1995, Martin et al. 1992) (But difficulty of grasping the overall structure of the questionnaire, weight of the computer)

Data collection methods in the SOEP

- Face-to-Face (PAPI): Paper-and-Pencil based face-to-face Interview (main data collection method in the first 15 years).

Data collection methods in the SOEP

- Face-to-Face (PAPI): Paper-and-Pencil based face-to-face Interview (main data collection method in the first 15 years).
- Self Completion mode (SELF): Respondents complete the questionnaire by themselves in front of the interviewer or with interviewer assistance (is NOT a pure self administered mode)

Data collection methods in the SOEP

- Face-to-Face (PAPI): Paper-and-Pencil based face-to-face Interview (main data collection method in the first 15 years).
- Self Completion mode (SELF): Respondents complete the questionnaire by themselves in front of the interviewer or with interviewer assistance (is NOT a pure self administered mode)
- Mixed mode: Combination of PAPI and SELF

Data collection methods in the SOEP

- Face-to-Face (PAPI): Paper-and-Pencil based face-to-face Interview (main data collection method in the first 15 years).
- Self Completion mode (SELF): Respondents complete the questionnaire by themselves in front of the interviewer or with interviewer assistance (is NOT a pure self administered mode)
- Mixed mode: Combination of PAPI and SELF
- CAPI: Computer-Assisted Personal Interview. Interviewers visit respondents and conduct a face-to-face interview using the computer

Data collection methods in the SOEP

- Face-to-Face (PAPI): Paper-and-Pencil based face-to-face Interview (main data collection method in the first 15 years).
- Self Completion mode (SELF): Respondents complete the questionnaire by themselves in front of the interviewer or with interviewer assistance (is NOT a pure self administered mode)
- Mixed mode: Combination of PAPI and SELF
- CAPI: Computer-Assisted Personal Interview. Interviewers visit respondents and conduct a face-to-face interview using the computer
- Mail: Some responses are elicited by mail when respondents would otherwise fail to cooperate.

Designing a mode comparison study is difficult

- Pure mode effect: measurement bias that is specially attributed to the mode

Designing a mode comparison study is difficult

- Pure mode effect: measurement bias that is specially attributed to the mode
 - ★ one has to compare the results of collections where the only difference between the collections is the mode

Designing a mode comparison study is difficult

- Pure mode effect: measurement bias that is specially attributed to the mode
 - ★ one has to compare the results of collections where the only difference between the collections is the mode
- ⇒ unrealistic assumption, situation has to rely on a laboratory experiment

Designing a mode comparison study is difficult

- Pure mode effect: measurement bias that is specially attributed to the mode
 - ★ one has to compare the results of collections where the only difference between the collections is the mode
- ⇒ unrealistic assumption, situation has to rely on a laboratory experiment
- ★ In the complex reality of large-scale data collection we measure only the combined effects of mode and other design factors.

Designing a mode comparison study is difficult

- Pure mode effect: measurement bias that is specially attributed to the mode
 - ★ one has to compare the results of collections where the only difference between the collections is the mode
 - ⇒ unrealistic assumption, situation has to rely on a laboratory experiment
 - ★ In the complex reality of large-scale data collection we measure only the combined effects of mode and other design factors.
- ⇒ Experimental design in Sample E, wave 1

Experimental longitudinal design - Sample E in the SOEP

- Sample E contains 2000 German households

Experimental longitudinal design - Sample E in the SOEP

- Sample E contains 2000 German households
- Sample E, wave 1, is split into two additional subsamples E1 and E2 with the same structure using twin sample points

Experimental longitudinal design - Sample E in the SOEP

- Sample E contains 2000 German households
- Sample E, wave 1, is split into two additional subsamples E1 and E2 with the same structure using twin sample points
- Each of the 125 sample points contains 16 addresses (8 for E1 and 8 for E2) and had to be realized in the first wave alternately with PAPI and CAPI mode per interviewer

Experimental longitudinal design - Sample E in the SOEP

- Sample E contains 2000 German households
 - Sample E, wave 1, is split into two additional subsamples E1 and E2 with the same structure using twin sample points
 - Each of the 125 sample points contains 16 addresses (8 for E1 and 8 for E2) and had to be realized in the first wave alternately with PAPI and CAPI mode per interviewer
- ⇒ control for possible interviewer effects in the analysis of mode effects

Experimental longitudinal design - Sample E in the SOEP

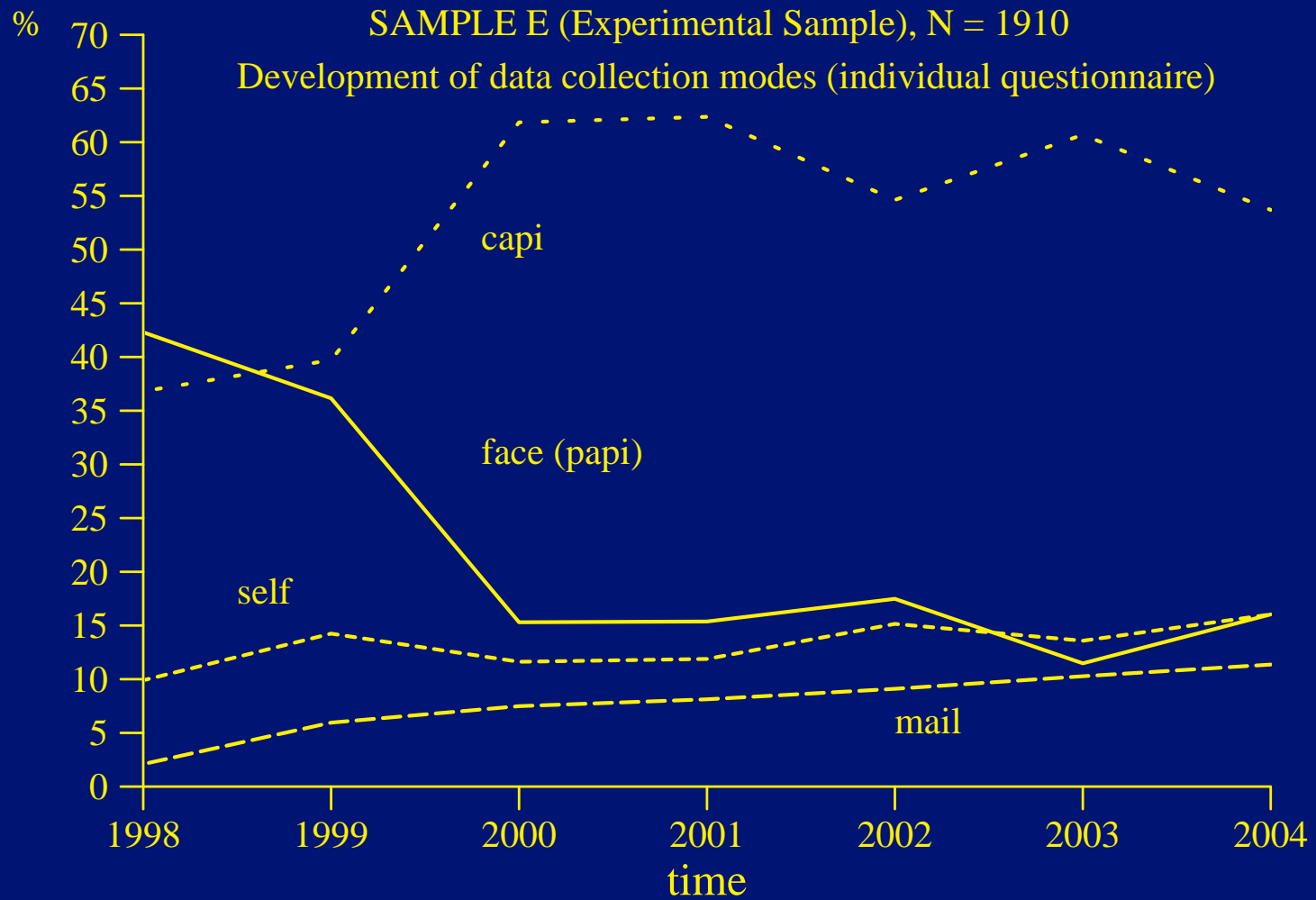
- Sample E contains 2000 German households
 - Sample E, wave 1, is split into two additional subsamples E1 and E2 with the same structure using twin sample points
 - Each of the 125 sample points contains 16 addresses (8 for E1 and 8 for E2) and had to be realized in the first wave alternately with PAPI and CAPI mode per interviewer
- ⇒ control for possible interviewer effects in the analysis of mode effects
- ⇒ after wave 1 the split was cancelled

<i>Method</i>	SAMPLE E, wave 1					
	<i>E1 PAPI</i>		<i>E2 CAPI</i>		<i>total</i>	
	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>	<i>N</i>	<i>%</i>
Households						
PAPI (face-to-face)	432	80.3	90	17.4	522	49.4
CAPI	16	3.0	398	76.8	414	39.2
Self completed	41	7.6	6	1.2	47	4.5
Mixed	25	4.6	6	1.2	31	2.9
Mail	22	4.1	16	3.1	38	3.6
d.k.	2	0.4	2	0.4	4	0.4
Total	538	100	518	100	1056	100
Persons						
PAPI (face-to-face)	650	65.4	158	17.2	808	42.3
CAPI	24	2.4	679	74.1	703	36.8
Proxy	2	0.2	-	-	2	0.1
Self completed	146	14.7	43	4.7	189	9.9
Mixed	127	12.7	15	1.7	142	7.4
Mail	21	2.1	19	2.1	40	2.1
d.k.	24	2.4	2	0.2	26	1.4
Total	994	100	916	100	1910	100

Source: SOEP 1998, Sample E, wave 1

Development of the data collection methods in Sample E (first wave 1998)

Development of the data collection methods in Sample E (first wave 1998)



Mode-Effects on data quality - Hypotheses

Mode-Effects on data quality - Hypotheses

1. **Respondent's acceptance:** There aren't significant differences between PAPI and CAPI for the probability of non-participation in the following wave (Baker 1992).

Mode-Effects on data quality - Hypotheses

1. **Respondent's acceptance:** There aren't significant differences between PAPI and CAPI for the probability of non-participation in the following wave (Baker 1992).
2. **Implausible values:** CAPI interviews have a fewer number of implausible values than PAPI interviews because of plausibility tests.

Mode-Effects on data quality - Hypotheses

1. **Respondent's acceptance:** There aren't significant differences between PAPI and CAPI for the probability of non-participation in the following wave (Baker 1992).
2. **Implausible values:** CAPI interviews have a fewer number of implausible values than PAPI interviews because of plausibility tests.
3. **Willingness to disclose sensitive information:** CAPI respondents haven't more reservation for giving sensitive statements like the gross-income than respondents in traditionally PAPI mode. (Baker 1995; de Leeuw 1995)

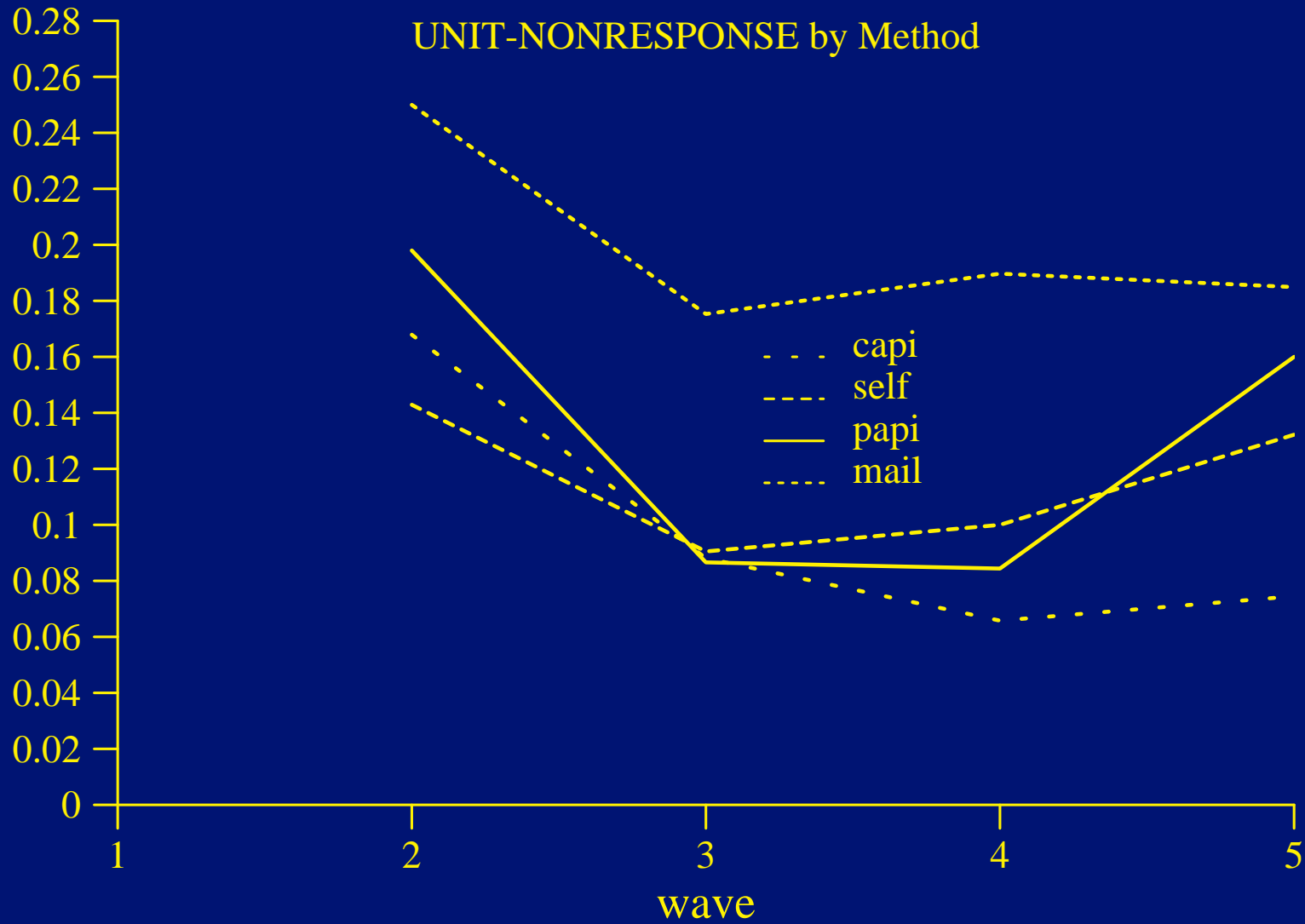
Unit-Nonresponse Reasons in Sample E

Unit-Nonresponse Reasons in Sample E

	Wave								total N
	2 N	%	3 N	%	4 N	%	5 N	%	
unsuccessful at the time (e.g. sick)	13	3.4	9	3.6	13	5.9	5	2.1	40
unwilling	92	24.2	72	28.5	74	33.8	96	39.7	334
final refusal	271	71.1	146	57.7	110	50.2	116	47.9	643
dead			15	5.9	13	5.9	20	8.3	48
HH not found	5	1.3	11	4.4	9	4.1	5	2.1	30
total	381	100	253	100	219	100	242	100	1095

Source: SOEP, Sample E, 1998 - 2002

Non-Participation in the next wave



Modelling Non-Participation in the next wave

Modelling Non-Participation in the next wave

- Binary response character:

Modelling Non-Participation in the next wave

- Binary response character:

$$y_{ij} = \begin{cases} 1, & \text{if } y_{ij}^* > 0, \text{ non-participation (next wave)} \\ 0, & \text{otherwise} \end{cases}$$

Modelling Non-Participation in the next wave

- Binary response character:

$$y_{ij} = \begin{cases} 1, & \text{if } y_{ij}^* > 0, \text{ non-participation (next wave)} \\ 0, & \text{otherwise} \end{cases}$$

⇒ **Multilevel logit models**: Level 1 - respondents, Level 2 - interviewer

Modelling Non-Participation in the next wave

- Binary response character:

$$y_{ij} = \begin{cases} 1, & \text{if } y_{ij}^* > 0, \text{ non-participation (next wave)} \\ 0, & \text{otherwise} \end{cases}$$

- ⇒ **Multilevel logit models**: Level 1 - respondents, Level 2 - interviewer
- ⇒ **Two models**: Model 1 - intercept is specified as random; Model 2 - random coefficient for CAPI

Modelling Unit-Nonresponse - Model 1

Model 1: Two-level random intercept model, the probability for π_{ij} for each response is estimated from:

Modelling Unit-Nonresponse - Model 1

Model 1: Two-level random intercept model, the probability for π_{ij} for each response is estimated from:

$$y_{ij} = \pi_{ij} + u_{ij} \quad (1)$$

Modelling Unit-Nonresponse - Model 1

Model 1: Two-level random intercept model, the probability for π_{ij} for each response is estimated from:

$$y_{ij} = \pi_{ij} + u_{ij} \quad (1)$$

$$\pi_{ij} = \left[1 + \exp(-(\beta_{0j} + \sum_{h=1}^H \beta_{h,ij} x_{h,ij} + v_{0j})) \right]^{-1} \quad (2)$$

Modelling Unit-Nonresponse - Model 1

Model 1: Two-level random intercept model, the probability for π_{ij} for each response is estimated from:

$$y_{ij} = \pi_{ij} + u_{ij} \quad (1)$$

$$\pi_{ij} = \left[1 + \exp(-(\beta_{0j} + \sum_{h=1}^H \beta_{h,ij} x_{h,ij} + v_{0j})) \right]^{-1} \quad (2)$$

where $x_{h,ij}$ represents values for covariates x_h ($h = 1, \dots, H$) of respondent i and interviewer j .

The intercept β_{0j} is specified as random on level 2 (interviewer level) and the variance is estimated as v_{0j} . The random variation among the respondents on level 1 is estimated as the variance u_{ij} .

Modelling Unit-Nonresponse - Model 2

Modelling Unit-Nonresponse - Model 2

Model 2: Two-level random coefficient model:

Modelling Unit-Nonresponse - Model 2

Model 2: Two-level random coefficient model:

$$\beta_{(capi)j} = \beta_{capi} + v_{capi,j} \quad (3)$$

Modelling Unit-Nonresponse - Model 2

Model 2: Two-level random coefficient model:

$$\beta_{(capi)j} = \beta_{capi} + v_{capi,j} \quad (3)$$

$$\pi_{ij} = \left[1 + \exp(-(\beta_{0j} + \sum_{h=1}^H \beta_{h,ij} x_{h,ij} + x_{capi} v_{capi,j} + v_{0j})) \right]^{-1} \quad (4)$$

Modelling Unit-Nonresponse - Model 2

Model 2: Two-level random coefficient model:

$$\beta_{(capi)j} = \beta_{capi} + v_{capi,j} \quad (3)$$

$$\pi_{ij} = \left[1 + \exp\left(-\left(\beta_{0j} + \sum_{h=1}^H \beta_{h,ij} x_{h,ij} + x_{capi} v_{capi,j} + v_{0j}\right)\right) \right]^{-1} \quad (4)$$

$v_{capi,j}$ is a normally distributed random effect with mean zero and variance $\sigma_{v,capi}^2$

$$\begin{bmatrix} v_{0,j} \\ v_{capi,j} \end{bmatrix} \sim N(0, \Sigma_R) : \Sigma_R = \begin{bmatrix} \sigma_{v,0}^2 & \\ \sigma_{v,capi0} & \sigma_{v,capi}^2 \end{bmatrix}$$

Allowing the coefficient of CAPI to vary across interviewers has also introduced the parameter

$\sigma_{v,capi0}$ which is the covariance between v_{0j} and $v_{capi,j}$.

April, 2006

	Estimates FIXED PART							
	wave 1				wave 2			
	model 1		model 2		model 1		model 2	
	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.
Intercept	-2.099***	0.63	-2.012***	0.67	1.215	1.43	-1.125	1.52
<i>respondent</i>								
sex (1 - men)	-0.044	0.15	-0.061	0.16	-0.013	0.22	0.001	0.24
age (year)	0.005	0.01	-0.004	0.01	-0.022**	0.01	-0.020**	0.01
not empl. (ref)								
low occup.	-0.364	0.56	-0.343	0.6	-1.859*	1.05	-1.930*	1.09
med occup.	-0.278	0.51	-0.299	0.53	-0.902	0.87	-0.998	0.92
high occup.	-0.167	0.55	-0.102	0.57	-1.86	0.97	-1.899*	1.02
Trainees	0.189	0.66	0.032	0.69	-0.198	0.96	-0.399	1.02
self-empl.	-0.27	0.59	-0.138	0.62	-0.822	0.93	-0.823	0.97
Milit./civi. serv.	-0.327	0.51	-0.297	0.54	-1.061	0.89	-1.167	0.93
size of HH	-0.077	0.07	-0.088	0.07	-0.291**	0.11	-0.267**	0.13
Move (t + 1)	1.290***	0.43	1.468***	0.44	1.933***	0.5	2.154***	0.56
<i>Interviewer</i>								
isex (1 - men)	0.946***	0.27	1.032***	0.28	-0.31	0.36	-0.35	0.39
<i>Situation</i>								
change of int. (t + 1)	0.468*	0.29	0.438	0.3	2.318***	0.36	2.266***	0.39
sum of part.					0.573	0.52	0.36	0.56
Papi (ref)								
Capi	-0.244	0.17	-0.341	0.27	-0.022	0.25	0.182	0.41
self completed	-0.037	0.32	-0.092	0.34	-0.03	0.39	-0.279	0.47
Mixed	-0.145	0.34	-0.294	0.36	0.508	0.6	0.345	0.66

Estimates - Random Part

	wave 1				wave 2			
	model 1		model 2		model 1		model 2	
	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.	$\hat{\beta}$	s.e.
Random								
<i>Respond. level</i>								
σ_u^2	1		1		1		1	
<i>Interv. Level</i>								
σ_v^2	0.913***	0.22	1.293***	0.35	1.433***	0.41	3.075***	0.96
$\sigma_{v(capi)}^2$			2.799***	0.88			5.572***	1.93
$\sigma_{v,capi0}$			-1.265**	0.47			-3.396***	1.22
Interviewer	110		110		115		115	
Persons	1583		1583		1477		1477	

Source: SOEP Sample E, individual questionnaire, 1998 - 1999; significance: *10%; ** 5%; ***1%

Results for Non-Participation in the next wave

Results for Non-Participation in the next wave

- Strong positive effects for moving and for change of the interviewer

Results for Non-Participation in the next wave

- Strong positive effects for moving and for change of the interviewer
- male interviewers have lost more respondents than female

Results for Non-Participation in the next wave

- Strong positive effects for moving and for change of the interviewer
- male interviewers have lost more respondents than female
- no significant differences in CAPI and PAPI

Results for Non-Participation in the next wave

- Strong positive effects for moving and for change of the interviewer
- male interviewers have lost more respondents than female
- no significant differences in CAPI and PAPI
- but the CAPI coefficient varies significantly between the interviewers in model 2 in wave 1 and 2.

Results for Non-Participation in the next wave

- Strong positive effects for moving and for change of the interviewer
 - male interviewers have lost more respondents than female
 - no significant differences in CAPI and PAPI
 - but the CAPI coefficient varies significantly between the interviewers in model 2 in wave 1 and 2.
- ⇒ The impact of the CAPI mode at time t on participation in $t + 1$ depends on interviewer's performance

Results for Non-Participation in the next wave

- Strong positive effects for moving and for change of the interviewer
 - male interviewers have lost more respondents than female
 - no significant differences in CAPI and PAPI
 - but the CAPI coefficient varies significantly between the interviewers in model 2 in wave 1 and 2.
- ⇒ The impact of the CAPI mode at time t on participation in $t + 1$ depends on interviewer's performance
- ⇒ Caused by interviewer's skill to manage the new data collection method

Hypothesis 2 and 3: Missing values and Implausible Values

- Hypothesis 2: Compare the average number of missing values and implausible values by mode

Hypothesis 2 and 3: Missing values and Implausible Values

- Hypothesis 2: Compare the average number of missing values and implausible values by mode
- Hypothesis 3: Willingness to disclose sensitive information - perform a gross income-nonresponse analysis

Average number of missings

Average number of missings

missing values	wave 1		wave 2		wave 3		wave 4		wave 5	
	all	empl.	all	empl.	all	empl.	all	empl.	all	empl.
face-to-face	5.28	5.88	1.97	2.81	1.66	2.14	2.60	3.14	2.52	2.97
self	6.76	7.44	2.89	2.94	2.53	3.13	3.44	3.71	3.41	3.69
mail	6.48	6.10	3.66	3.91	5.72	6.17	4.91	5.38	5.22	5.44
capi	5.46	5.87	2.84	3.37	1.99	2.38	2.57	3.22	4.15	4.87
total	5.68	6.25	2.61	3.22	2.38	2.97	2.97	3.58	3.82	4.37

Source: SOEP, Sample E, individual questionnaire, 1998 - 2002 (own calc.)

Average number of missings

missing values	wave 1		wave 2		wave 3		wave 4		wave 5	
	all	empl.	all	empl.	all	empl.	all	empl.	all	empl.
face-to-face	5.28	5.88	1.97	2.81	1.66	2.14	2.60	3.14	2.52	2.97
self	6.76	7.44	2.89	2.94	2.53	3.13	3.44	3.71	3.41	3.69
mail	6.48	6.10	3.66	3.91	5.72	6.17	4.91	5.38	5.22	5.44
capi	5.46	5.87	2.84	3.37	1.99	2.38	2.57	3.22	4.15	4.87
total	5.68	6.25	2.61	3.22	2.38	2.97	2.97	3.58	3.82	4.37

Source: SOEP, Sample E, individual questionnaire, 1998 - 2002 (own calc.)

- highest number in the case of employed respondents and mail and self completed questionnaires

Average number of missings

missing values	wave 1		wave 2		wave 3		wave 4		wave 5	
	all	empl.	all	empl.	all	empl.	all	empl.	all	empl.
face-to-face	5.28	5.88	1.97	2.81	1.66	2.14	2.60	3.14	2.52	2.97
self	6.76	7.44	2.89	2.94	2.53	3.13	3.44	3.71	3.41	3.69
mail	6.48	6.10	3.66	3.91	5.72	6.17	4.91	5.38	5.22	5.44
capi	5.46	5.87	2.84	3.37	1.99	2.38	2.57	3.22	4.15	4.87
total	5.68	6.25	2.61	3.22	2.38	2.97	2.97	3.58	3.82	4.37

Source: SOEP, Sample E, individual questionnaire, 1998 - 2002 (own calc.)

- highest number in the case of employed respondents and mail and self completed questionnaires
- lowest number in the case of face-to-face interviews

Average number of implausible values

Average number of implausible values

	wave 1		wave 2		wave 3		wave 4		wave 5	
	all	empl.	all	empl.	all	empl.	all	empl.	all	empl.
face-to-face	0.07	0.09	0.12	0.12	0.11	0.11	0.04	0.04	0.04	0.05
self	0.06	0.05	0.13	0.11	0.11	0.12	0.15	0.17	0.08	0.09
mail	0.03	0.03	0.09	0.07	0.06	0.06	0.13	0.04	0.09	0.1
capi	0.11	0.12	0.06	0.06	0.04	0.05	0.03	0.04	0.03	0.05
total	0.09	0.1	0.1	0.09	0.07	0.08	0.06	0.06	0.05	0.07

Source: SOEP, Sample E, individual questionnaire, 1998 - 2002 (own calc.)

Average number of implausible values

	wave 1		wave 2		wave 3		wave 4		wave 5	
	all	empl.	all	empl.	all	empl.	all	empl.	all	empl.
face-to-face	0.07	0.09	0.12	0.12	0.11	0.11	0.04	0.04	0.04	0.05
self	0.06	0.05	0.13	0.11	0.11	0.12	0.15	0.17	0.08	0.09
mail	0.03	0.03	0.09	0.07	0.06	0.06	0.13	0.04	0.09	0.1
capi	0.11	0.12	0.06	0.06	0.04	0.05	0.03	0.04	0.03	0.05
total	0.09	0.1	0.1	0.09	0.07	0.08	0.06	0.06	0.05	0.07

Source: SOEP, Sample E, individual questionnaire, 1998 - 2002 (own calc.)

- CAPI reduces the number of implausible values in the data set (except wave 1 because of transposition problems)

Mode and Income-Nonresponse

Mode and Income-Nonresponse

- Has CAPI a significant effect on respondent's decision to reveal his earnings?

Mode and Income-Nonresponse

- Has CAPI a significant effect on respondent's decision to reveal his earnings?
- Previous results suggest to distinguish between 'refusals' and 'don't know' (Schräpler 2004; Schräpler 2006)

Mode and Income-Nonresponse

- Has CAPI a significant effect on respondent's decision to reveal his earnings?
 - Previous results suggest to distinguish between 'refusals' and 'don't know' (Schräpler 2004; Schräpler 2006)
1. Description: development of gross-income nonresponse in sample E

Mode and Income-Nonresponse

- Has CAPI a significant effect on respondent's decision to reveal his earnings?
 - Previous results suggest to distinguish between 'refusals' and 'don't know' (Schräpler 2004; Schräpler 2006)
1. Description: development of gross-income nonresponse in sample E
 2. Modelling: Multilevel multivariate model with three response variables, 'refusal', 'don't know' and 'participation in the next wave'

**Gross-Income-Nonresponse-rate from employed persons in
Sample E (in per cent)**

Gross-Income-Nonresponse-rate from employed persons in Sample E (in per cent)

<i>wave</i>	<i>including self-employed and trainees employed</i>			<i>excluding self-employed and trainees selected</i>		
	<i>respondents</i>	<i>missing</i>	<i>%</i>	<i>respondents</i>	<i>missing</i>	<i>%</i>
1	1032	272	26.4	870	206	23.7
2	886	167	18.8	736	113	15.4
3	858	151	17.6	716	106	14.8
4	805	153	19	658	95	14.4
5	746	131	17.6	613	89	14.5
total	4327	874	20.2	3593	609	16.9

Source: SOEP, Sample E, 1998 - 2002 (own calc.)

**Gross Income-Nonresponse rate by data collection method in
Sample E, employed persons (in per cent)**

Gross Income-Nonresponse rate by data collection method in Sample E, employed persons (in per cent)

<i>method</i>	<i>Wave</i>					<i>N</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
Face-to-face	21.4	13.0	12.9	4.1	8.3	864
Mixed	14.6	12.5	14.8	28.6	23.0	173
Self completed	22.8	12.9	8.9	12.1	1.6	546
Mail	22.2	15.0	23.3	16.2	22.1	305
Capi	27.2	18.6	15.5	16.9	16.0	1676
N	870	736	716	658	613	3593

Source: SOEP, Sample E, 1998 - 2002 (own calc.)

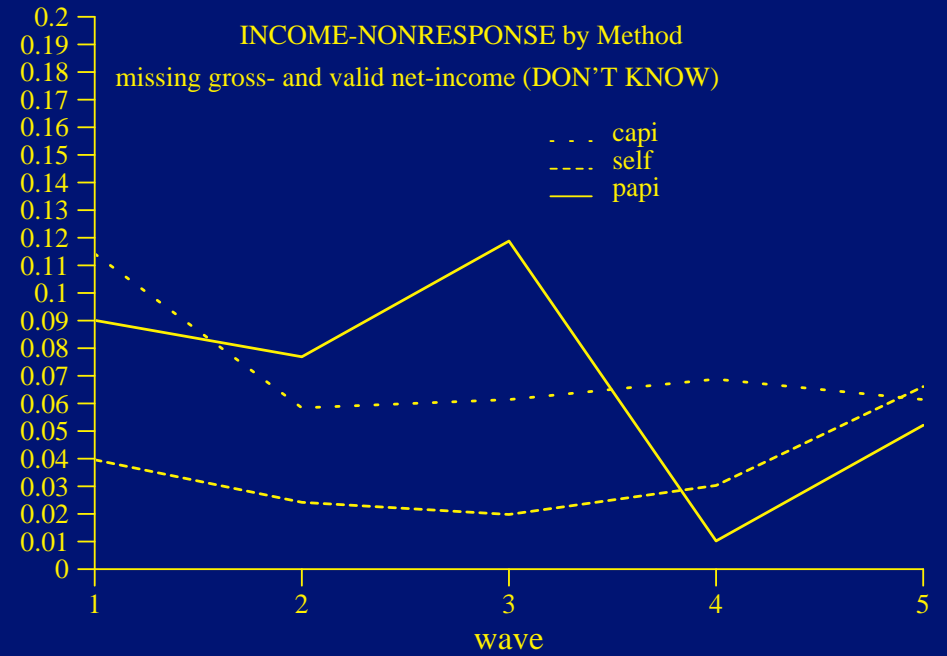
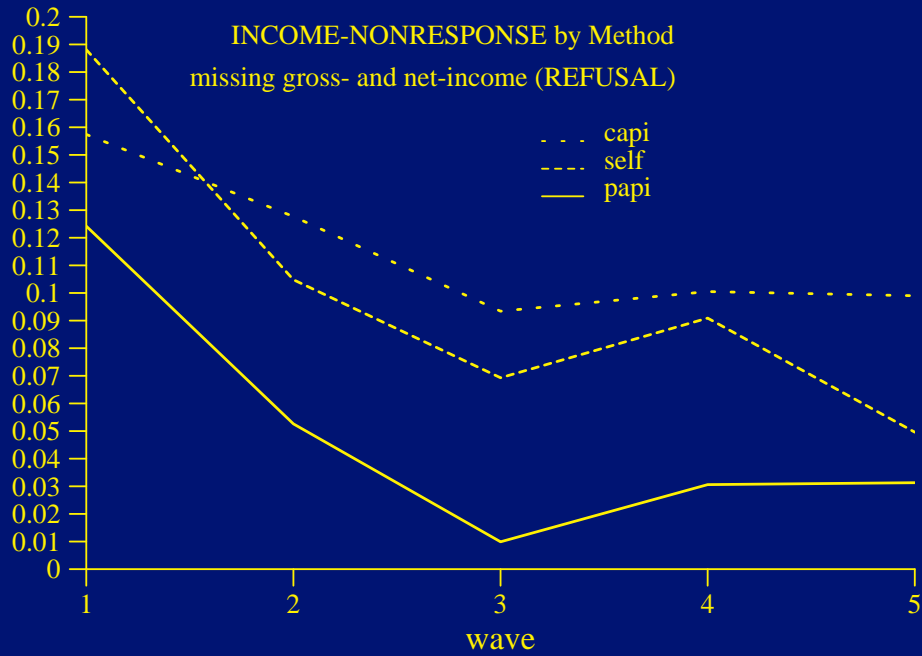
Missing gross- and net-income in sample E, wave 1 - 5

Missing gross- and net-income in sample E, wave 1 - 5

<i>gross-income</i>	<i>net-income</i>					
	valid		missing		total	%
	N	%	N	%		
valid	2831	78.8	149	4.2	2980	83.0
missing	249	7.0	360	10.0	609	17.0
total	3080	85.8	509	14.2	3589	100.0

Source: SOEP, Sample E, 1998 - 2002 (own calc.)

Development: Refusals and Don't Know by mode



Modelling Income-Nonresponse

Modelling Income-Nonresponse

For respondent j and interviewer k one dichotomous variable y_{ijk} is observed:

$$y_{ijk} = \pi_{ijk} + u_{ijk}$$

$$y_{1jk} = \begin{cases} 1, & \text{if } y_{1jk}^* > 0, \text{ refuse} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$y_{2jk} = \begin{cases} 1, & \text{if } y_{2jk}^* > 0, \text{ don't know} \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

$$y_{3jk} = \begin{cases} 1, & \text{if } y_{3jk}^* > 0, \text{ unit-response (next wave)} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Modelling Income-Nonresponse - 2

Modelling Income-Nonresponse - 2

The probability π_{ijk} for each response variable i estimated from:

Modelling Income-Nonresponse - 2

The probability π_{ijk} for each response variable i estimated from:

$$\pi_{ijk} = \left[1 + \exp\left(-\left(\beta_{0k} + \sum_{h=1}^H \beta_{h,ijk} x_{h,ijk} + v_{0k}\right)\right) \right]^{-1} \quad (8)$$

Modelling Income-Nonresponse - 2

The probability π_{ijk} for each response variable i estimated from:

$$\pi_{ijk} = \left[1 + \exp\left(-\left(\beta_{0k} + \sum_{h=1}^H \beta_{h,ijk} x_{h,ijk} + v_{0k}\right)\right) \right]^{-1} \quad (8)$$

where $x_{h,ijk}$ represents values for covariates x_h ($h = 1, \dots, H$) of respondent j and interviewer k . The intercept is specified as random on level 3 (interviewer level) and the variance is estimated as v_{0k} . The random variation among the respondents on level 2 is estimated as the variance/covariance u_{ijk} .

WAVE 1 Fixed Part	Refuse		Don't Know		Uni-Response (t + 1)	
	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.
Intercept	-1.007	1.008	-3.368***	1.257	1.575*	0.899
respondent						
sex (1 - men)	-0.083	0.202	-0.496***	0.206	-0.271	0.193
age (year)	0.005	0.010	0.013	0.010	-0.002	0.009
med occup. (ref)						
low occup.	-0.171	0.319	0.854***	0.274	0.058	0.284
high occup.	0.521***	0.239	-0.602**	0.310	-0.089	0.239
size of HH	0.079	0.087	-0.047	0.095	0.190**	0.083
move	0.122	0.743	0.000	0.000	-1.722***	0.577
interviewer						
isex (1 - men)	-0.015	0.349	0.129	0.491	-1.126***	0.339
change of interviewer face (ref)	0.565	0.408	-0.480	0.452	-0.194	0.358
capi	0.463**	0.228	0.362	0.236	-0.063	0.209
self completed	0.816***	0.363	-1.054**	0.590	0.063	0.377
interviewer cluster			106			
persons			702			
-2 * LogLikelih.			-792.9			

NOTE: ^o constrained to zero; Significance: *10%; ** 5%; ***1%

Source: SOEP, Sample E, 1998, employed respondents

without self-employed and trainees, without mail interviews (own calc.)

WAVE 1 Random Part	Refuse		Don't Know		Uni-Response (t + 1)	
	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.
respondent level						
	u1		u2		u3	
u1	0.654	0.037				
u2	0.000 ^o	0.000	0.413	0.024		
u3	-0.079***	0.028	-0.019	0.022	0.716	0.041
interviewer level						
	v1		v2		v3	
v1	1.437	0.381				
v2	-0.608	0.398	2.951	0.716		
v3	0.054	0.258	0.194	0.359	1.224	0.324
interviewer cluster			106			
persons			702			
-2 * LogLikelih.			-792.9			

NOTE: ^o constrained to zero; Significance: *10%; ** 5%; ***1%

Source: SOEP, Sample E, 1998, employed respondents

without self-employed and trainees, without mail interviews (own calc.)

WAVE 2 Fixed Part	Refuse		Don't Know		Uni-Response (t + 1)	
	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.
Intercept	-2.202	1.414	-5.120**	2.090	0.178	1.218
respondent						
sex (1 - men)	0.004	0.216	-0.190	0.232	0.001	0.233
age (year)	0.001	0.010	0.048***	0.011	0.024**	0.011
med occup. (ref)						
low occup.	-0.072	0.361	-0.743***	0.336	0.715*	0.427
high occup.	0.136	0.276	-0.931***	0.332	0.777**	0.347
size of HH	-0.012	0.104	0.000	0.107	0.112	0.114
move	1.344**	0.593	0.000	0.000	-1.647***	0.564
interviewer						
isex (1 - men)	0.020	0.476	0.427	0.781	0.114	0.405
change of interviewer	-0.489	0.544	-1.360	1.232	-1.607***	0.456
face (ref)						
capi	1.339***	0.264	0.380	0.268	0.256	0.269
self completed	1.293***	0.381	-0.783*	0.425	-0.435	0.373
interviewer cluster			110			
persons			656			
-2 * LogLikelih.			-3921.9			

NOTE: ^o constrained to zero; Significance: *10%; ** 5%; ***1%

Source: SOEP, Sample E, 1999, employed respondents

without self-employed and trainees, without mail interviews (own calc.)

WAVE 2	Refuse		Don't Know		Uni-Response (t + 1)	
Random Part	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.	$\hat{\beta}_1$	s.e.
respondent level						
	u1		u2		u3	
u1	0.430	0.026				
u2	0.000 ^o	0.000	0.224	0.013		
u3	-0.036	0.020	0.008	0.014	0.516	0.030
interviewer level						
	v1		v2		v3	
v1	3.474	0.754				
v2	0.371	0.866	6.372	1.753		
v3	0.140	0.470	2.229	0.810	1.897	0.518
interviewer cluster			110			
persons			656			
-2 * LogLikelih.			-3921.9			

NOTE: ^o constrained to zero; Significance: *10%; ** 5%; ***1%

Source: SOEP, Sample E, 1999, employed respondents

without self-employed and trainees, without mail interviews (own calc.)

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

- Unit-nonresponse (in the following wave):

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

- Unit-nonresponse (in the following wave):
 - ★ no direct effects of CAPI

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

- Unit-nonresponse (in the following wave):
 - ★ no direct effects of CAPI
 - ★ in the first two waves significant interviewer variation of the CAPI coefficient

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

- Unit-nonresponse (in the following wave):
 - ★ no direct effects of CAPI
 - ★ in the first two waves significant interviewer variation of the CAPI coefficient
- ⇒ respondent's participation in the following wave depends on interviewer's skill to manage CAPI

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

- Unit-nonresponse (in the following wave):
 - ★ no direct effects of CAPI
 - ★ in the first two waves significant interviewer variation of the CAPI coefficient
- ⇒ respondent's participation in the following wave depends on interviewer's skill to manage CAPI
- CAPI reduces implausible values: in wave 2 to 5 the average numbers are lowest in the CAPI mode.

Conclusion

We have examined data collection mode effects using quality indicators like unit-nonresponse, missing values, implausible values and gross income-nonresponse

- Unit-nonresponse (in the following wave):
 - ★ no direct effects of CAPI
 - ★ in the first two waves significant interviewer variation of the CAPI coefficient
- ⇒ respondent's participation in the following wave depends on interviewer's skill to manage CAPI
- CAPI reduces implausible values: in wave 2 to 5 the average numbers are lowest in the CAPI mode.

- Gross income-nonresponse: the estimates of the multilevel multivariate logit models show that in the first waves CAPI interviews have a significant higher probability for refusals (missing gross- and net-income) than PAPI interviews.

- Gross income-nonresponse: the estimates of the multilevel multivariate logit models show that in the first waves CAPI interviews have a significant higher probability for refusals (missing gross- and net-income) than PAPI interviews.
- ★ CAPI respondents have more reservations for giving sensitive statements like gross-income than respondents in traditionally PAPI mode.

- Gross income-nonresponse: the estimates of the multilevel multivariate logit models show that in the first waves CAPI interviews have a significant higher probability for refusals (missing gross- and net-income) than PAPI interviews.
 - ★ CAPI respondents have more reservations for giving sensitive statements like gross-income than respondents in traditionally PAPI mode.
- ⇒ it might be that the use of laptops increases privacy or confidentially concerns

- Gross income-nonresponse: the estimates of the multilevel multivariate logit models show that in the first waves CAPI interviews have a significant higher probability for refusals (missing gross- and net-income) than PAPI interviews.
 - ★ CAPI respondents have more reservations for giving sensitive statements like gross-income than respondents in traditionally PAPI mode.
- ⇒ it might be that the use of laptops increases privacy or confidentiality concerns
- ⇒ important result, we can expect that in a few years the CAPI interviewing method will replace more and more the PAPI mode.

Thank You!

[back](#)