J Econ Inequal
https://doi.org/10.1007/s10888-018-9378-x

Survey mode effects on measured income inequality

Pirmin Fessler[1] · Maximilian Kasy[2] · Peter Lindner[1]

Received: 28 January 2016 / Accepted: 22 March 2018

Abstract We study the effect of interview modes on estimates of economic inequality which are based on survey data. We exploit variation in interview modes in the Austrian EU-SILC panel, where between 2007 and 2008 the interview mode was switched from personal interviews to telephone interviews for some but not all participants. We combine methods from the program evaluation literature with methods from the distributional decomposition literature to obtain causal estimates of the effect of interview mode on estimated inequality. We find that the interview mode has a large effect on estimated inequality, where telephone interviews lead to a larger downward bias. The effect of the mode is much smaller for robust inequality measures such as interquantile ranges, as these are not sensitive to the tails of the distribution. The magnitude of effects we find are of a similar order as the differences in many international and intertemporal comparisons of inequality.

Keywords Income inequality · Survey methodology · Survey modes · Distributional decompositions

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10888-018-9378-x) contains supplementary material, which is available to authorized users.

✉ Pirmin Fessler
pirmin.fessler@oenb.at

Maximilian Kasy
maximiliankasy@fas.harvard.edu

Peter Lindner
peter.lindner@oenb.at

1 Economic Analysis Division, Oesterreichische Nationalbank, Postbox 61, 1011 Vienna, Austria

2 Department of Economics, Harvard University, 1805 Cambridge Street, Cambridge, MA 02138, USA

Published online: 11 April 2018

1 Introduction

Intertemporal and international comparisons of economic inequality are typically based on inequality measures calculated from survey data; see for instance a recent report by the OECD (2011). The surveys most widely used for such calculations include the Survey of Consumer Finances (SCF), the Panel Study of Income Dynamics (PSID), and the EU - Statistics on Income and Living Conditions (EU-SILC). Across different surveys and across different waves of the same survey different interview modes are used. In particular, some of these surveys are conducted in person (computer assisted personal interview, CAPI), while others are conducted via telephone (computer assisted telephone interview, CATI). For a comparison of the different EU-SILC surveys see Table 1.

Existing evidence (e.g. de Leeuw 1992; Lohmann 2011) suggests that in such surveys non-response and misreporting of income is a concern. This is of particular importance in the upper and lower tails of the distribution, and the literature suggests that both of these problems might be even more severe in telephone interviews. Non-response and misreporting are important problems in the context of inequality measurement because many common inequality measures, such as the Gini coefficient, are not robust in the sense of Huber (2003); that is, they do not have a bounded influence function (c.f. Cowell and Victoria-Feser 1996). This implies that minor data contaminations can have a large impact on measured inequality. The possibly large influence of minor mismeasurement stands in contrast to comparably small differences in inequality measures across countries and time. The fact that differences are small and of a similar order of magnitude as possible effects of mismeasurement suggests that comparisons (rankings) across different countries based on surveys might be quite problematic.

In this paper, we use variation of interview methodology in the Austrian EU-SILC 2008 survey to provide causal estimates of the effect of the interview mode (CAPI vs. CATI) on measures of income inequality, such as the Gini coefficient and the 90/10 percentile ratio. This causal effect of interview modes on measured inequality is our primary object of interest. To shed further light on the mechanisms driving this causal effect, we additionally discuss the effect of interview modes on the response behavior of survey participants. Section 2 below reviews earlier literature aiming to estimate the effect of interview modes in various context; to the best of our knowledge, our paper is the first considering the effect of interview modes on the measurement of income inequality, which poses special challenges due to difficulties in the tails.

The EU-SILC is a dataset widely used by policymakers and statistical offices in the European Union. Our estimates exploit the panel structure of the EU-SILC survey conducted in Austria in 2007 and 2008 and control for a rich set of covariates from the baseline survey. The Austrian EU-SILC survey was based exclusively on CAPI in 2007, and used a mixed-mode design (CAPI and CATI) in 2008. The concurrent mixed-mode design aimed at interviewing as many panel households as possible via CATI. Due to accessibility problems via phone and the possibility to opt for the CAPI mode, roughly 40% of households were interviewed via CATI and 60% via CAPI.

We find that a switch from CAPI to CATI leads to major changes in response behaviour (unit and item non-response) and answering behaviour (potential misreporting), leading to large differences of estimated inequality measures. Selective item non-response in the tails is significantly higher for CATI (on average by about 20%–30%), while incomes for CATI are less dispersed around mean income. A switch from CAPI to CATI in particular decreases the Gini coefficient of household income by roughly 10%, and has a statistically insignificant effect on the 90/10 percentile ratio. The smaller effect on the latter statistic is likely

Table 1 Overview of EU-SILC surveys and their mode of data collection

	PAPI	CAPI	CATI	Self-administrated	Gini 2007	adjusted Gini
Belgium	0.0	100.0	0.0	0.0	26.3	–
Czech Republic	99.7	0.0	0.0	0.3	25.3	–
Denmark	0.0	0.0	94.2	5.8	25.2	27.6
Germany	0.0	0.0	0.0	100.0	30.4	–
Estonia	2.2	97.6	0.2	0.0	33.4	33.4
Ireland	0.0	100.0	0.0	0.0	31.3	–
Greece	80.8	14.9	2.1	2.3	34.3	34.4
Spain	0.0	92.9	7.1	0.0	31.2	31.4
France	0.0	100.0	0.0	0.0	26.6	–
Italy	100.0	0.0	0.0	0.0	32.3	–
Cyprus	0.0	100.0	0.0	0.0	29.8	–
Latvia	11.3	81.2	7.5	0.1	35.7	36.0
Lithuania	95.3	0.0	3.8	0.9	33.8	33.9
Luxembourg	100.0	0.0	0.0	0.0	27.4	–
Hungary	100.0	0.0	0.0	0.0	25.6	–
Malta	0.0	100.0	0.0	0.0	26.3	–
The Netherlands	0.0	0.0	100.0	0.0	27.6	30.4
Austria	0.0	94.0	6.0	0.0	26.2	26.4
Poland	100.0	0.0	0.0	0.0	32.2	–
Portugal	8.0	92.0	0.0	0.0	36.8	–
Slovenia	0.0	44.5	55.5	0.0	23.2	24.5
Slovakia	99.4	0.0	0.0	0.7	24.5	–
Finland	0.0	3.4	96.6	0.0	26.2	28.7
Sweden	0.0	0.0	100.0	0.0	23.4	25.7
United Kingdom	0.0	100.0	0.0	0.0	32.6	–
Iceland	0.0	0.0	100.0	0.0	28.0	30.8
Norway	0.0	0.6	99.4	0.0	23.7	26.1

Notes:
(i) This table shows percent shares of EU-SILC surveys in 2007 conducted by paper assisted personal interview (PAPI), computer assisted personal interview (CAPI), computer assisted telephone interview (CATI), and self administered
(ii) Gini-Coefficients are based on household disposable equivalence income
(iii) The adjusted Gini-Coefficients is the "Back of the Envelope" calculation accounting for the effect (decrease of 10%) of the CATI interviewing technique from the RIF-regression
(iv) *Source:* Eurostat: Comparative Intermediate EU Quality Report 2007. Version 5, and Eurostat website for Gini Coefficients

due to the fact that it is robust, whereas the Gini coefficient is not. These findings imply that international and intertemporal comparisons of inequality are quite sensitive to mode choices, as we discuss below.

The differing distribution of measured outcomes between the CAPI and the CATI sample can be decomposed into a measurement effect and a selection effect. The measurement effect is due to potential differences in answering behavior. The selection effect is due to potential selection (i) into unit non-response and (ii) between the two interview modes.

We are interested in the causal effect of interview modes on the distribution of measured outcomes, that is, in the measurement effect. We use panel data and control for a rich set of baseline covariates to eliminate the selection effect from our estimates.

The rest of this paper is structured as follows. Section 2 shortly reviews relevant earlier literature on mode effects. The data used as well as the source of variation which we exploit are discussed in Section 3. After introducing the data (3.1), we describe the switch of the interviewing mode (3.2) and the outcome measures that are used (3.3). Section 4 discusses identification and estimation. Section 5 discusses our results on the effect of the interviewing mode on item non-response, on reported income, on the measured distribution of income and possible mechanisms leading to these effects. Finally, Section 6 summarizes our results and concludes. Additional results and robustness checks are provided in a supplementary online Appendix.

2 Earlier literature

There are three mechanisms contributing to the differing distribution of measured outcomes between CAPI and CATI interview modes. The first of these is unit non-response, which contributes to selection bias. The second and third are item non-response and misreporting, which (in combination with imputation) drive the measurement effect that we aim to isolate in this paper. These mechanisms have been discussed in the prior literature. The key difference between our paper and earlier studies is that we focus on the impact of survey modes on measured (income) inequality. This is both a substantively important area, and it raises distinct methodological challenges, as will be discussed in greater detail in Section 3.

2.1 Methodological contributions

A new textbook by Dillman et al. (2014) provides a comprehensive introduction to the design of surveys using various modes. Vannieuwenhuyze and Loosveldt (2013) and Klausch et al. (2015a) present methods to evaluate mixed-mode effects and disentangle selection and measurement effects. Schouten et al. (2013) uses a design and estimation method similar to ours to disentangle measurement effects in social surveys. This literature discusses a set of methods to decompose total mode effects into measurement and selection effects. These methods are mostly based on controlling for covariates which are mode invariant to balance respondent compositions across modes and filter out selection effects, and in some instances use instrumental variable techniques. Our paper similarly relies on controlling for covariates. However, our data resembles a quasi-experiment as we have information on all households based on the same interview mode (CAPI) for 2007. These data also include our outcome variables in 2008, namely income and item non-response with regard to income. Then in 2008 some but not all households were interviewed using a new mode (CATI). Therefore, we are in the fortunate situation of having a rich set of lagged covariates (including the outcome), all measured using the same survey mode.

2.2 Empirical findings in the literature

Several empirical studies have considered the effect of survey modes on response behavior (though not on measured income inequality). Vannieuwenhuyze et al. (2014) statistically model both selection and measurement effects in the context of mixed mode surveys. Their application focuses on the comparison between self-administered postal surveys and CAPI. They find only small effects of interview modes, with qualitative conclusions changing depending on modeling assumptions.

A series of papers by Klausch et al. (2013a, b, 2015b) investigates the effect of survey modes in the context of the Dutch Crime Victimisation Survey. In this survey, modes were experimentally (randomly) assigned. Four different modes are considered: CAPI, CATI, self-administered paper-based, and web-based. Focusing on attitudinal questions, they find that item non-response is significantly higher for self-administered interview modes and lowest for CAPI. Additionally, the sample of respondents is most representative of the population for CAPI. The respondents for the web-based mode are almost as representative. Their results suggest that sequential (mixed mode) designs might improve response rates and the representativity of respondents.

Atkeson et al. (2014) compare CATI and self-administered (web- and paper-based) interviews in a survey on voting in New Mexico. They use matching on observable sociodemographic variables to compare modes. They find that both modes predict the election outcome equally well, while self-administered surveys are cheaper by 30% relative to CATI.

Jäckle et al. (2010) similarly study the effect of interview modes using experimental mode assignment. They consider a survey conducted in 2003/2005 in Hungary. They argue "[...] that in order to evaluate whether mode affects data comparability it is necessary to move away from an assessment of means and marginal distributions toward an assessment of the effect of mode on relevant estimates" (p. 12). This is one of the key motivations for the analysis in the present paper, where we study the effect of interview modes on measured income inequality. The online Appendix discusses some additional relevant contributions to this literature.

3 Data and identifying variation

The analysis in this paper is based on a panel of households participating in the two consecutive EU-SILC survey waves for the years 2007 and 2008 in Austria. In 2007, all households in our sample were interviewed via CAPI; in 2008 42.6% of these households switched to CATI. Our main results are based on variation of interview modes in 2008, conditioning on the rich set of baseline covariates from the year 2007, which include in particular lagged income. The panel structure of our data and the richness of available covariates make a causal interpretation of the resulting estimates very plausible, as we shall argue below in greater detail.

3.1 The EU-SILC

Our empirical analysis uses the Austrian part of the European Union Statistics on Income and Living Conditions (EU-SILC) data of the waves 2007 and 2008. Documentation for these data is provided by Statistik Austria, see Statistik Austria (2010) and http://ec.europa.eu/eurostat/web/microdata/european-union-statistics-on-income-and-living-conditions [accessed on January 19th 2018] for further background material. EU-SILC is an annual survey collecting information at the micro level on income, poverty and living conditions.

These data include imputed observations to correct for item non-response. Unit non-response is corrected for using a non-response adjustment reweighting procedure. This procedure is based on a logit model for response propensities conditional on external available information about the households (see Statistik Austria 2010). Depending on

Table 2 Number of households in different modes and waves

	Number	%-Share
Total number of households 2008	5,711	100%
CATI-interview-mode	1,710	29.94%
CAPI-interview-mode	4,001	70.06%
Panel households 2007/2008	3,772	66.05%
CATI-test households 2007	395	6.92%
Effective sample size	3,377	59,13%
From the effective sample mode was:		
CATI	1,438	42.58%
CAPI	1,939	57.42%

Notes:
(i) This table reports the household sample size for the 2007/2008 waves of EU-SILC
(ii) Effective sample size is "Panel households" minus "Test households 2007"
(iii) *Source:* EU-SILC 07/08

the availability of information various forms of single imputation methods are deployed to address item non-response. We use the imputations and weights provided by the data producer in our analysis; these are used in almost all studies based on EU-SILC.

3.2 The interview mode

In addition to the data available in the standard user dataset we obtained an indicator for the years 2007 and 2008 describing whether a specific household was interviewed using the CAPI or the CATI interviewing mode from Statistik Austria. All respondents were asked an identical set of questions concerning income independent of the interview mode. We use both waves; the CATI option was first introduced in 2008, after a test period for 2007.

Table 2 reports how many households can be used for the identification of the effect of the interview mode. There are 5,711 households interviewed in 2008, about 30% (i.e. 1,710 households) of which were surveyed via a CATI interview. Only two thirds are panel households in the sense that they were also interviewed in the previous wave. Close to 400 households are part of a CATI test that was run in 2007 and hence are excluded from the analysis because these households are already interviewed using the CATI mode in 2007. Thus the effective sample size is reduced to 3,377 households which could be (self-) selected to one of the two interview modes. All households were interviewed in 2007 via CAPI; the response rate was 78.8% in 2007. All of those who responded were targeted to be re-interviewed via CATI in 2008. Due to accessibility problems via phone and the possibility to opt for the CAPI mode 42.6% of them were interviewed via CATI and 57.4% via CAPI. The attrition rate between 2007 and 2008 was about 20%.

To identify the effect of interview modes on the measured distribution of income and on response behavior, we control for a host of variables measured in 2007 for all households with the same (CAPI) mode, including income and item non-response information. By construction, these variables could not have been causally impacted by interview modes in 2008; they are predetermined relative to our treatment variable.

The lower part of Table 3 shows the means of all household-level and personal-level control variables from the EU-SILC wave of 2007, when all households where interviewed via

Table 3 Means of dependent and explanatory variables

	Panel	CAPI	CATI
Dependent variables: 2008			
Log disposable household income (mean)	10.24	10.16	10.37
	(0.009)	(0.015)	(0.015)
Log gross household income (mean)	10.51	10.42	10.66
	(0.010)	(0.016)	(0.017)
Item non-response (mean)	21.84	18.08	27.53
	(0.711)	(0.874)	(1.178)
Share of interviews below 10k	16.52		
	(0.640)		
Interviewed by		74.80	25.20
		(1.944)	(1.944)
Share of interviews above 200k	0.07		
	(0.045)		
Interviewed by		100.00	0.00
		(.)	(.)
Control variables at the household level: 2007			
Household size (mean)	2.35	2.28	2.45
Share of households with kids	31.06	29.27	33.65
Share of single member households	39.34	37.16	42.48
Share of home-owners	53.51	49.40	59.43
Size of flat in sqm (mean)	97.96	92.60	105.70
Land line coverage (share)	66.16	58.14	77.74
Mobile phone coverage (share)	87.30	85.18	90.36
Log disposable household income (mean)	10.18	10.09	10.32
Log gross household income (mean)	10.42	10.32	10.58
Share of households in city	38.44	38.18	38.81
Share of households in urban areas	26.17	26.95	25.04
Share of households in rural areas	36.03	35.73	36.46
Control variables at the level of the household head: 2007			
Female (share)	45.87	44.03	48.52
Employment: Blue collar (share)	43.44	42.46	44.85
Employment: Self-employment (share)	6.21	5.77	6.83
Employment: Jobless (share)	50.36	51.77	48.32
Weekly working hours (mean)	19.91	19.84	20.02
Married (share)	49.69	45.52	55.71
Education: secondary school (share)	21.35	25.91	14.76
Education: apprenticeship (share)	52.54	52.68	52.34
Education: higher secondary school (share)	16.04	13.13	20.25

Table 3 (continued)

	Panel	CAPI	CATI
Education: university (share)	10.07	8.28	12.66
Age (mean)	52.39	51.86	53.14

Notes:
(i) This table shows the means of the outcome variables we consider, as well as of the full set of control variables
(ii) Each statistic is provided for the panel component of the 2008 wave, and the CAPI and CATI sub-samples of the panel component
(iii) All income variables are reported after taking the natural logarithm
(iv) The percentage of households in each region is left out due to space constraints, but the partitioning is approximately as follows: Burgenland 4%; Kaernten 8%; Niederoesterreich 20%; Oberoesterreich 20%; Salzburg 7%; Steiermark 14%; Tirol 8%; Vorarlberg 5%; and Wien 17%
(vi) *Source:* EU-SILC 07/08

CAPI. The distribution of households across regions (not shown in the table, but used as controls) is approximately as follows: Burgenland 4%; Carinthia 8%; Lower Austria 20%; Upper Austria 20%; Salzburg 7%; Styria 14%; Tirol 8%; Vorarlberg 5%; and Vienna 17%. There are some minor differences in the 2007 averages between the two groups defined by the interview mode in 2008. Higher income and wealthier households seem to be (self-) selected to CATI. Additionally, phone coverage (see land line and mobile phone coverage) and education seem to be lower for households with a CAPI-interview. Women answer the questionnaire more often over the telephone. These differences indicate the possibility of households with certain characteristics to be (self-) selected into one or the other interviewing mode. We model this process and control for it in the estimates reported below.

3.3 Outcome measures

Our analysis mainly focuses on disposable (i.e., after-tax) income as reported in the EU-SILC data. Household disposable income is constructed by summing up all of the household's income sources. We construct a household income item non-response dummy (HINR) as 1 if the household disposable income flag indicates that household disposable income is "partly imputed" or "completely imputed" and 0 otherwise. The dummy therefore indicates if there was missing information with respect to the overall household income variable.

The upper part of Table 3 reports the mean of these variables in the panel households and the CAPI and CATI sub-samples. These estimates are not weighted; we are not interested in population but only sample averages at this stage of the analysis. We see that the average log income is lower for households surveyed in a CAPI interview. The item non-response is about 18% for CAPI households but 27.5% for CATI households. Furthermore, we can see that most of the income-poor households, i.e. 75% of the households below an income of € 10,000 , and all of the income rich households (above € 200,000) are interviewed via CAPI. This last result illustrates the problems originating from the telephone interview mode in terms of a possible compression of the income distribution via unit and item non-response. CATI seems to decrease participation rates especially at the tails of the distribution. Apparently none of the very high as well as very few low income households could be reached via CATI. Average disposable and gross income increased from 2007 to 2008 by 6% and 9%, respectively.

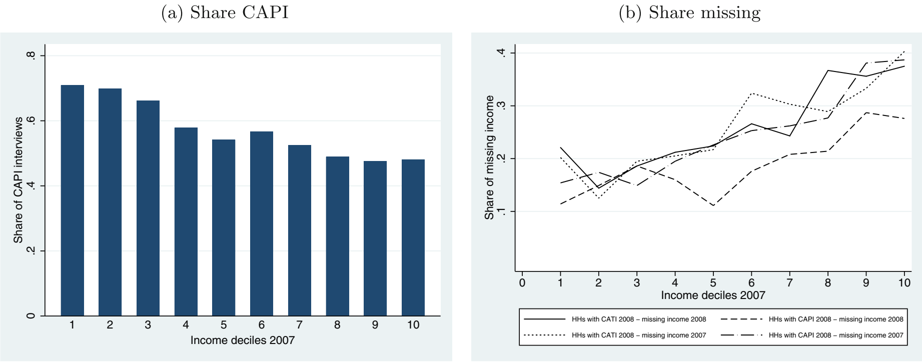

Fig. 1 Share of observations with missing income across deciles, conditional on interview mode. Notes: (i) Graph **a** shows share of households interviewed by CAPI in 2008 over 2007 income deciles. (ii) Graph **b** shows the share of observations with missing incomes separately by treatment group and income decile. (iii) Source: EU-SILC 07/08

Further evidence on selection into the different interviewing modes is provided by Fig. 1a and b. These figures show the share of CAPI interviews (Fig. 1a) and item non-response rates for our treatment (CATI08/CAPI07) and control groups (CAPI08/CAPI07) over deciles of income in 2007 (Fig. 1b). The participation rate in a CAPI interview is decreasing with income. Additionally item non-response is higher for higher income households. In 2008 there is a large difference of more than 10 percentage points between CAPI and CATI in the two top deciles. At the lowest end, the first decile, the difference is almost as high, and for households with intermediate incomes, CATI-interviews generally have higher item non-response, although the difference is not stable over the deciles. Suggestive evidence that this is not caused by selection is provided by the fact that these patterns do not prevail for the response rates in 2007, when everyone was interviewed by CAPI. As can be seen, the percentage of households with missing income data does not follow the same pattern in 2007. Here the two groups are defined according to the 2008 interview mode choice and we report item non-response in 2007 when both groups (CAPI as well as CATI) were still interviewed via CAPI. Again the share of households with missing income data increases with income. The difference between the groups, however, is not stable. This suggests the presence of an interview mode effect.

4 Identification and estimation

Having introduced our data and key identifying variation, we now turn to a discussion of the identification and estimation approaches used in this paper. Our approach is based on controlling for a rich set of lagged covariates, and using distributional decomposition techniques to estimate the effect on measures of inequality. Controlling for covariates is standard in the literature on mode effects; focusing on distributional statistics rather than averages is not. We therefore discuss the latter in greater detail below. We will consider alternative estimation techniques to explore the robustness of our findings: (linear) regression and coarsened exact matching to control for covariates, and reweighting and RIF regression to estimate the effect on distributional statistics.

4.1 Controlling for baseline covariates

Our goal is to estimate the effect of the interview mode (CAPI versus CATI) on the measured income distribution, as well as on item non-response. Since interview mode was not randomly assigned in the EU-SILC 2008 we need to control for selection, exploiting the panel structure of the data and the rich set of baseline covariates, in order to justify a causal interpretation of the estimated effects. Textbook introductions to the causal literature can be found in Angrist and Pischke (2008) or Imbens and Rubin (2015); a good review of treatment effect estimation under conditional independence (missing at random) assumptions is provided by Imbens (2004), see also Rosenbaum and Rubin (1983).

Even small effects of the interview mode on some parts of the distribution of reported incomes might translate into economically significant effects on measures of income inequality. We employ methods of the distributional decomposition literature including reweighting similar to DiNardo et al. (1996), as well recentered influence function (RIF) regression as in Firpo et al. (2009) to estimate the effect of the mode (CAPI vs. CATI) on various distributional statistics $\nu(P(Y))$, where $P(Y)$ is the unconditional distribution of household income.

4.2 Reweighting

Suppose we observe a cross-section with independent and identically distributed (i.i.d.) draws from the distribution P of the variables (Y, M, X), where X denotes a rich set of covariates, including lagged Y from the 2007 survey. The variable Y denotes reported income in 2008 and M is the mode the respondent is confronted with. We are interested in isolating the effect of a change of the interview mode on the distribution of reported incomes Y, $P(Y)$, or statistics thereof, $\nu(P(Y))$. Possible choices for ν include the mean, the share below the poverty line, quantiles, quantile-ratios, and the Gini coefficient.

Let $P^m(Y|X)$ denote the conditional distribution of Y given X and $M = m$. Define

$$P^m(Y) := \int_X P^m(Y|X)dP(X). \tag{1}$$

This distribution is given by the conditional distribution of Y given X for the *subpopulation* where $M = m$, averaged over the *full population* distribution of X. This counterfactual distribution is constructed similarly to the counterfactual changes in the wage distribution of the United States, ascribed to changes in unionization and the minimum wage etc., which were analyzed in DiNardo et al. (1996).

This counterfactual distribution can be interpreted causally under the conditional independence assumption (CIA). Denote Y^m the income of an individual that she would report if interviewed using mode $M = m$, $m \in \{0, 1\}$. If

$$(Y^1, Y^0) \perp M|X, \tag{2}$$

then $P^m(Y|X) = P(Y|M = m, X) = P(Y^m|X)$, and $P^m(Y) = P(Y^m)$. This assumption states that there is no self-selection into interview modes correlated with potential reported income, conditional on the covariates X. This assumption is reasonably credible with a rich set of covariates (including lagged Y from the 2007 survey), as we have at our disposition. Under the CIA, we can interpret the average partial effect of M on Y, $E[E[Y|X, M = 1] - E[Y|X, M = 0]]$ as an average treatment effect (ATE), $E[Y^1 - Y^0]$; see for instance Imbens and Rubin (2015).

We can represent the distribution P^m as

$$P^m(Y \leq y) = E\left[\mathbf{1}(Y \leq y) \cdot \theta^m(X)\right], \tag{3}$$

where

$$\theta^m(X) := \frac{\mathbf{1}(M = m)}{P(M = m|X)} \tag{4}$$

and $\mathbf{1}(.)$ denotes the indicator function. Equation 3 states that P^m is a re-weighted version of the distribution P. Any counterfactual distributional characteristic ν of P^m can be estimated based on estimates of P^m, as in DiNardo et al. (1996). This requires estimation of the ratio (4). We estimate this ratio using a coarsened exact matching procedure. Iacus et al. (2008) developed a method to temporarily coarse data based on ex-ante user choice and then run the analysis on the common support of the uncoarsened data. A detailed description as well as robustness checks using propensity score matching and fully integrated linear models can be found in the online Appendix.

4.3 RIF-regression representation of counterfactual distributions

Alternatively, assume for a moment that ν can be written as the expectation of a function f of Y, $\nu = E[f(Y)]$. Then the effect of treatment on ν can be represented as

$$\nu^1 - \nu^0 = \int (E[f(Y)|X, M = 1] - E[f(Y)|X, M = 0])\, dP(X). \tag{5}$$

In general, ν will not have this linear form but can be approximated by a linear first order expansion around some baseline distribution P^*. The baseline distribution is chosen arbitrarily; when ν is non-linear, the first order approximation will be better for distributions close to P^*. This idea underlies the influence-function regression approach proposed in Firpo et al. (2009). It requires estimation of the regression $E[f(Y)|X, M]$.

Corresponding to these two representations of the counterfactual ν^m, we consider two estimation approaches; reweighting observations and influence-function regression. The reweighting approach estimates the weight $\theta^m(X)$ and calculates counterfactual ν from the reweighted distribution P^m.

The influence-function regression approach is based on the first order approximation of ν, as a function of P, around P^*:

$$\nu(P) = \nu(P^*) + \int IF(y; \nu, P^*) d(P - P^*)(y) + R^*, \tag{6}$$

where IF is the influence function of the parameter ν at P^* and R^* is a second order remainder term. This equation can be thought of as the definition of the influence function. To gain intuition for this definition, assume for a moment that P has finite support, so that we can think of P as a k vector. Then ν is a differentiable function on $\mathbb{R}^k$, and the derivative of ν is equal to the influence function, $\partial\nu/\partial P = IF$, where we can think of the influence function as a vector in $\mathbb{R}^k$, as well. Equation 6 is the natural generalization of this definition to measures P with arbitrary, possibly infinite, support.

Ignoring the remainder R^* in Eq. 6, this representation of ν has the linear form required for the use of the representation (5), that is,

$$\nu(P) \approx E[\nu(P^*) + IF(Y; \nu, P^*)]. \tag{7}$$

We can hence calculate first order approximations to the counterfactual ν based on estimates of $E[IF|X, M = m]$. For details, the reader is referred to Firpo et al. (2009).

In the online Appendix we describe details of how to estimate the weights for reweighting and we also use various alternative estimation methods popular in the program evaluation literature in order to check the robustness of our results.

5 Results

This section presents and discusses our empirical findings. Recall that our main goal is to estimate the causal effect of survey modes on measured inequality, i.e., the measurement effect. A naive comparison of measured inequality is contaminated by the selection effect (non-response, choice of interview mode). The Gini coefficient in our CAPI sample is 0.34, in the CATI sample 0.31, so that a naive comparison suggests a reduction of the Gini coefficient by 0.03 when switching to CATI. The 90/10 percentile ratio in our CAPI sample is 4.88, in the CATI sample 4.57, suggesting a reduction of 0.31 when switching to CATI.

Our goal is to present estimates of the causal effect of modes on measured inequality purged of the selection effect, and to shed light on the mechanisms driving the measurement effect. To do so, we first discuss some preliminary estimates demonstrating that selection into interview modes is indeed non-random. Hence selection is a potential concern. We next present estimates of the causal effect of modes on item non-response and on average (log) income. The effects on item non-response show that CATI reduces response rates relative to CAPI, which contributes to distortions in measured inequality. The effects on (log) income show that indeed reported income changes depending on survey modes. We finally get to our core results, estimating the mode effect on measured income inequality in Section 5.2. The construction of these estimates is similar to those of the previous subsection, but we now focus on distributional statistics rather than averages. These estimates show that inequality as measured by the Gini coefficient or the 90/10 percentile ratio is lower when using CATI rather than CAPI. Our preferred estimates show a reduction of the Gini of 0.027

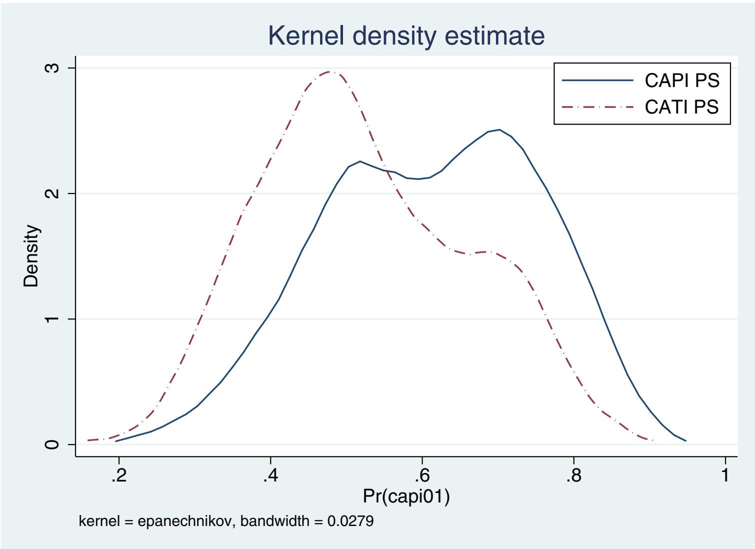

Fig. 2 Propensity score density of CAPI and CATI. Notes: (i) This graph shows estimated propensity score densities resulting from the logit model presented in the supplementary Appendix. (ii) Source: EU-SILC 07/08

Table 4 Effect of interview mode on income

	Logit			OLS			Reweighting
Specification	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Income Item Non-Response	–0.088	–0.075	–0.071				–0.068
	(0.014)	(0.014)	(0.015)				(0.014)
Log Household Income				–0.205	–0.052	–0.040	–0.060
				(0.022)	(0.016)	(0.015)	(0.017)
Controlling for							
household characteristics		yes	yes		yes	yes	yes
personal characteristics			yes			yes	yes

Notes:
(i) This table shows average partial effects (APE) of being interviewed by CAPI on household income item non-response and household income. Results are reported from a logistic (using an item non-response dummy for household income [at least one item non-response in an income question] and log household income as dependent variables). Reweighting based on coarsened exact matching is used as our preferred method
(ii) Standard errors are given in parentheses
(iii) *Source:* EU-SILC 07/08

when switching to CATI, and an effect on the 90/10 percentile ratio that is statistically indistinguishable from 0. The naive comparison of the Gini between modes thus reflects primarily a true measurement effect. The comparison of the 90/10 percentile ratio, on the other hand, reveals no true measurement effects, presumably due to the greater robustness of this statistic.

By explicitly considering the heterogeneity of this effect across quantiles of the income distribution, we see that this effect is primarily driven by higher incomes in the middle of the income distribution when using CATI. The section concludes with a discussion of mechanisms contributing to this measurement effect based on our findings, considering in particular unit non-response, item non-response, and misreporting.

5.1 Preliminaries

In order to model selection into interview modes, we run a logit regression of interview mode (CATI=0, CAPI=1) on the set of available controls, as listed in Table 3. Figure 2 displays the support of the propensity scores as implied by the logit model. As can be seen in Fig. 2, the support of propensity scores for CATI- and CAPI-households is nearly identical. However, the chance of (self-) selection to CAPI decreases (statistically) significantly on a 5% level with household size, household disposable income, the availability of a telephone line, and a mobile phone in the household as well as the main income earner being female, being married and living together as well as having higher educational attainment.[1]

Our controls are highly predictive of the outcomes of interest (reported income and item non-response); the R^2 of a linear regression of log income 2008 on all controls including lagged income is 0.58, and without lagged income it is 0.47. Table 4 shows the estimated effects of CAPI on item non-response and income. All estimated effects resulting from the logit regressions are significant at the 1% level, but are not significantly different from

[1] See the online Appendix for a table showing the average marginal effects calculated from this model.

each other. This is despite the fact that an increasingly rich set of controls is used across specifications (starting with no controls, then using household-level controls, finally using both household-level and personal-level controls), which suggests again that selection bias is rather small. The estimate using the largest set of controls, namely all household- and personal-level controls, is −0.071, indicating that the probability of item non-response when interviewed via CAPI is 7.1 percentage points (the average item non-response is 21.8%) lower than in the case of a CATI interview. This implies a reduction of item non-response of about 30% given the EU-SILC estimates.

As our preferred method to effectively filter out selection bias, we use the reweighting procedure as described in Section 4. To calculate weights we use coarsened exact matching (CEM). CEM imposes a user input based non-parametric matching strategy to balance the joint distributions of covariates among CAPI- and CATI-observations. This reduces the necessary extrapolation outside the common support and of course comes with a decrease in sample-size. Out of the 343 covariate combinations defined, we find 224 which define the common support, that is where at least one CAPI and one CATI observation can be found. The sample collapses to 3,190 households which lie inside the common support. Details on CEM as well as a version using classical propensity score reweighting as well as other robustness checks can be found in the online Appendix. The estimate using 3,190 CAPI households after the coarsened exact matching and reweighting to balance the covariate distributions is −0.068 and significant at the 5% level. Thus these more flexible estimation approaches confirm the previous finding of the interview mode on the item non-response using the somewhat naive linear approach.

Analogous to the estimates of the mode effects on item non-response, we estimate the effect of the mode on household income again with increasing flexibility and less restrictive assumptions. Overall we find a significant negative effect of the CAPI mode on average (log) household income. The estimate ranges between −0.2 (OLS without controls) and −0.04 (OLS with all controls). An estimate of −0.04 implies that interviewing via CAPI leads to reported incomes that are on average about € 1,000 lower than those obtained via CATI. As the income distribution is very skewed and most households have income below the mean income, this result already implies that CATI leads on average to income values closer to the mean income. The more flexible method CEM confirms the negative effect.

5.2 Main results: Mode effects on the income distribution

While it is interesting to estimate the effect of modes on average reported (log) income, this provides only a partial picture and does not allow for the inference of the effect of modes on measures of inequality. We thus turn next to our main results, estimating the effect of modes on measures including the Gini, the poverty rate, and the 90/10 percentile ratio of reported incomes.

We regress the RIF of the Gini, the poverty rate, and the 90/10 percentile ratio of the income distribution of 2008 on the interview mode and (i) linear, squared, and cubed income from 2007 as well as (ii) all our controls for household and personal characteristics. In both cases all interactions of the control variables with the interview modes are included in the model, as well. For the standard errors of the reported average treatment effect (ATE) we use the delta method. Bootstrapped standard errors using 500 replicates were also calculated as a robustness check and yield very similar results.

Using CATI instead of CAPI as interview modes reduces the observed Gini-coefficients significantly (see Table 5). We estimate a reduction of 0.033 with the limited set of controls,

Table 5 Effect of interview mode on inequality

	Gini coefficient		Poverty rate		90/10 Percentile ratio	
Specification	(1)	(2)	(3)	(4)	(5)	(6)
RIF-regression	0.0329	0.0274	0.0068	0.0019	–0.2344	–0.1829
	(0.0107)	(0.0101)	(0.0127)	(0.0134)	(0.2469)	(0.2571)
Controlling for						
past income	yes	yes	yes	yes	yes	yes
other covariates		yes		yes		yes

Notes:
(i) This table shows the effect of the interview mode on aggregate measures of inequality. We report the inequality statistic (Gini coefficient, poverty rate, and the percentile ratio), the difference between the subsamples and the effect of the interview mode using RIF-regressions
(ii) Standard errors are reported using delta methods
(iii) *Source:* EU-SILC 07/08

and a reduction of 0.027 using the full set of controls. Inequality as measured by the Gini Coefficient thus drops by around 10% if the interview mode is switched from CAPI to CATI. By contrast, the estimated effect on the poverty rate is economically small and statistically insignificant. The same holds for the 90/10-percentile ratio. These findings suggest that using the latter two measures when comparing inequality over time or between countries might be a more robust choice than the Gini-coefficient.

Figure 3 shows estimates of the effect of the interview mode on observed percentiles of the income distribution. To interpret this figure, note that an estimate of -.2 for the 50th percentile would imply that the median of reported incomes is lower by about 20% under the CAPI interview mode relative to the CATI interview mode, for instance.

Panels (a) to (c) show that the effect of the interview mode on the percentiles (the effect is estimated for 20 quantiles) follows a u-shape. These graphs are based on the full sample in Panel (a), the balanced (using the coarsened exact matching procedure outlined above) observations in Panel (b), and the re-weighted k to k matching (same matching procedure as above) in Panel (c). The u-shape means that the percentiles are lower for households interviewed with CAPI but less so at the extremes of the distribution. This implies a higher spread of income in households interviewed with CAPI. In general, the income distribution is highly skewed, implying that most households have incomes below the mean and CATI tends to push them closer to the mean.

Panel (d) of Fig. 3 shows the average treatment effect (ATE) on the percentiles using RIF-regression estimates with the full set of control variables. We see again that it follows a u-shape. The effect of the interview mode on the percentiles is closer to zero, since the model controls for a wide range of characteristics, and is negative in most parts of the income distribution. In a distribution that is skewed to the right, this coupled with the positive effect for the highest income percentiles translates once again to a more compressed income distribution for households interviewed with CATI. Note that the coverage effect, namely that households at the extremes of the distribution are only covered with CAPI interviews, is not accounted for in this estimation.

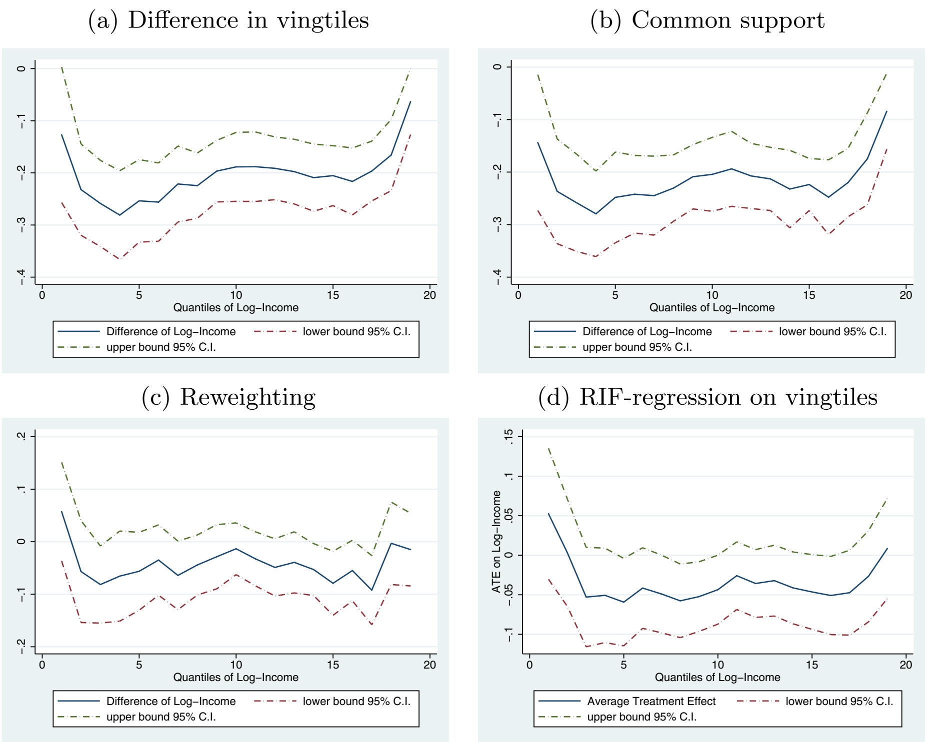

Fig. 3 Average treatment effect of CAPI vs. CATI on vingtiles. Notes: (i) These graphs show the effect of the interview mode on the vingtiles over the whole income distribution. Panel **a** displays the simple difference of the vingtiles, Panel **b** shows the differences for the balanced sample using the coarsened exact matching technique explained above, Panel **c** shows the results for the exact k to k matching within a bin of the matching procedure, and Panel **d** shows the effect on the percentiles using RIF-regressions. (ii) 95%-confidence intervals are provided using bootstrapping standard errors (Panel **a** to **c**) and the delta method (Panel **d**). (iii) Source: EU-SILC 07/08

5.3 Discussion of mechanisms

We now turn to a discussion of the mechanisms through which interview modes might affect measured income inequality. These mechanisms potentially operate on three levels.

First, we discuss the mechanisms at the level of **unit non-response**. We find that average incomes among CATI respondents are higher; cf. row 2 of Table 3. That said, households in both the upper and the lower tail of the income distribution have very low unit response rates in telephone interviews, compared to personal interviews, as can be seen in rows 5 and 7 of Table 3. This suggests that a survey based on CATI alone would not be able to adequately cover the tails of the distribution, in contrast to surveys using CAPI or a mixed-mode design in which CATI-nonrespondents are contacted for personal interviews.

Second, we discuss the mechanisms at the level of **item non-response**. In addition to the effect of leading to generally higher item non-response, the effect of CATI is particularly pronounced at the tails of the distribution. Especially at the top of the distribution, this is very worrisome, as the share of total income held by the top income earners is much larger than their population share (see Fig. 1a and b).

Third, we discuss the mechanisms at the level of **incomes reported** by respondents. We find that CATI leads to a positive effect on reported income, where the effect is larger for lower income groups than for very high income groups, and leads to significantly lower measured income inequality. This might be due to two mechanisms. On the one hand, the questionnaires in these surveys are rather complicated so that especially financially less literate (low end of the income distribution) households as well as households with very complicated income structures (high end of the income distribution) might be more likely to exhibit measurement error over the telephone than in a personal interview. On the other hand, it might be simply easier to lie over the phone than in a personal conversation.

Our setting and design allow us to estimate the magnitude of the second (item non-response) and third (income level) of these channels. We can evaluate the third effect in terms of measures of income inequality obtained, and the second in terms of a causal estimate of the interviewing mode on the item non-response. However, we also find strong evidence that the first effect (unit non-response) exists and is non-negligible. Therefore, it is very likely that we underestimate the effect that a counterfactual switch from assigning all households to CAPI to assigning all households to CATI would have on observed inequality measures; both the first and second effect also seem to compress the measured income distribution. In general people with very low income as well as people with very high income tend to report values biased towards the mean. The combined effect of all three mechanisms seems to be larger when CATI instead of CAPI is used as an interview mode. The decision to not report at all (unit non-response), selectively not report (item non-response), or report values closer to the mean might be easier via the phone than in a face to face situation.

6 Conclusion

We have discussed the effect of the interview mode (CAPI versus CATI) on item non-response and the level as well as the distribution of household income. We exploited the availability of panel data for households whose interview mode changed. That mode change occurred in the 2008 wave of the Austrian EU-SILC data for some - but not all - panel households. This source of variation allows us to estimate causal effects of the change of the interview mode. The available panel data allow us in particular to control for household income in 2007 as measured in CAPI mode for *all* households.

Our main empirical findings are as follows. First, we find descriptive evidence that CATI compresses the income distribution by leading to less coverage in the final sample via higher unit non-response especially at the tails of the distribution. Second, controlling for a rich set of covariates from the baseline survey, we find that the change from CAPI to CATI has increased item non-response significantly in the statistical sense, and by an amount that is economically important. This result is robust over all the parametric, semi-parametric, and non-parametric methods we applied (see also online Appendix for robustness checks). Every researcher pursuing answers to economic questions with the evaluation of survey data should thus be concerned about the interview mode of data collection and the follow-up imputations. One has to keep in mind that all missing values are usually imputed in various ways or, even more severe, dropped from the analysis altogether. Third, we find that households which are interviewed by CATI on average report higher incomes. This effect is larger for lower income households than for those with higher incomes. In general, the income distribution is highly skewed, implying that most households have incomes below the mean. We find that this level effect implies a large effect on income inequality as measured by the

Gini coefficient. We conduct RIF-regressions to compare the effect of the introduction of CATI on the unconditional distribution of income and find a highly significant effect which reduces the Gini coefficient by around 10%.

Let us now discuss the extent to which the mode effects we find are quantitatively important, and what they imply for survey methodology. A "back of the envelope" calculation in the last column (denoted "adjusted Gini") of Table 1 illustrates these possible effects which lead to severe re-ranking of several countries. These calculations are based on the assumption that a switch from CAPI to CATI interviews leads to a decrease of estimated Gini-coefficients by 10%, as suggested by the estimates we discussed. These calculations imply that countries with considerable proportions of CATI interviews would typically rank higher in rankings of income inequality if they were to use CAPI (e.g. 7 positions higher in the case of Denmark and 5 positions higher in the case of Finland).

We draw the following general conclusions from these findings. First, it seems that CAPI yields more reliable measures of inequality, and should be used where possible. Second, when making international or inter-temporal comparisons of inequality, we should make sure to compare "apples to apples". Comparisons of inequality measures based on surveys using different methodologies might be quite misleading. Third, given the issues with survey data in general and surveys using CATI in particular, it might be advisable to focus on robust inequality measures such as quantile contrasts when conducting inequality comparisons. We view our estimates as the lower bound of the effect, since coverage effects of the interview mode cannot be accounted for.

Acknowledgements We would like to thank Markus Knell and Alyssa Schneebaum for valuable comments and discussion. Additional to the usual disclaimer, the opinions expressed in this work are those of the authors and do not necessarily reflect the ones of the OeNB or the Eurosystem.

References

Angrist, J., Pischke, J.-S.: Mostly Harmless Econometrics. Princeton University Press, Princeton (2008)

Atkeson, L.R., Adams, A.N., Alvarez, R.M.: Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys. Political Analysis (2014)

Cowell, F.A., Victoria-Feser, M.-P.: Robustness properties of inequality measures. Econometrica **64**(1), 77–101 (1996)

de Leeuw, E.D.: Data Quality in Mail, Telephone and Face to Face Surveys. TT-Publikaties, Amsterdam (1992)

Dillman, D.A., Smyth, J.D., Christian, L.M.: Internet, Phone, Mail, and Mixed-Mode Surveys: the Tailored Design Method. Wiley, New York (2014)

DiNardo, J., Fortin, N., Lemieux, T.: Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. Econometrica **64**, 1001–1044 (1996)

Firpo, S., Fortin, N.M., Lemieux, T.: Unconditional quantile regression. Econometrica **77**(3), 953–973 (2009)

Huber, P.J.: Robust Statistics. Wiley Series in Probability and Statistics. Wiley, New York (2003)

Iacus, S.M., King, G., Porro, G.: Matching for Causal Inference without Balance Checking. Working paper series, Harvard (2008)

Imbens, G.W.: Nonparametric estimation of average treatment effects under exogeneity: a review. Rev. Econ. Stat. **86**(1), 4–29 (2004)

Imbens, G.W., Rubin, D.B.: Causal Inference fo Statistics, Social and Biomedical Sciences an Introduction. Cambridge University Press, Cambridge (2015)

Jäckle, A., Roberts, C., Lynn, P.: Assessing the effect of data collection mode on measurement. Int. Stat. Rev. **78**(1), 3–20 (2010)

Klausch, T., Hox, J.J., Schouten, B.: Assessing the Mode-Dependency of Sample Selectivity Across the Survey Response Process. Technical Report 2013-03, Statistics Netherlands (2013a)

Klausch, T., Hox, J.J., Schouten, B.: Measurement effects of survey mode on the equivalence of attitudinal rating scale questions. Sociol. Methods Res. **42**(3), 227–263 (2013b)
Klausch, T., Hox, J.J., Schouten, B.: Evaluating bias of sequential mixed-mode designs against benchmark surveys. Sociol. Methods Res. **46**(3), 456–489 (2015a)
Klausch, T., Hox, J.J., Schouten, B.: Selection error in single- and mixed mode surveys of the Dutch general population. J. R. Stat. Soc.: Ser. A (Statistics in Society) **178**(4), 945–961 (2015b)
Lohmann, H.: Comparability of eu-silc survey and register data: the relationship among employment, earnings and poverty. J. Eur. Social Policy **21**(1), 1–18 (2011)
OECD: Divided We Stand - Why Inequality Keeps Rising (2011)
Rosenbaum, P.R., Rubin, D.B.: The central role of the propensity in observational studies for causal effects. Biometrika **70**(1), 41–55 (1983)
Schouten, B., van den Brakel, J., Buelens, B., van der Laan, J., Klausch, T.: Disentangling mode-specific selection and measurement bias in social surveys. Soc. Sci. Res. **42**, 1555–1570 (2013)
Statistik Austria: Standard Dokumentation Metainformationen - Definitionen, Erlaeuterungen, Methoden, Qualitaet (2010)
Vannieuwenhuyze, J.T., Loosveldt, G.: Evaluating relative mode effects in mixed-mode surveys: three methods to disentangle selection and measurement effects. Sociol. Methods Res. **42**(1), 82–104 (2013)
Vannieuwenhuyze, J.T., Loosveldt, G., Molenberghs, G.: Evaluating mode effects in mixed-mode survey data using covariate adjustment models. J. Off. Stat. **30**(1), 1–21 (2014)