Survey mode effects on income inequality measurement*

Pirmin Fessler[†], Maximilian Kasy[‡], and Peter Lindner[§]

December 18, 2013

Abstract

In this paper, we study the effect of interview modes on estimates of economic inequality which are based on survey data. We exploit quasi-experimental 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 (exploiting the panel structure and the rich set of covariates from the 2007 survey) with methods from the distributional decomposition literature (reweighting and influence function regression) 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.

JEL Classification: C14, C21, C81, C83, D31

Key Words: Income Inequality, Survey Methodology, Survey Modes, Distributional Decompositions

*We'd like to thank Markus Knell 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.

[†]Economic Analysis Division, Oesterreichische Nationalbank; Address: Postbox 61, 1011 Vienna, Austria; e-Mail: pirmin.fessler@oenb.at; Telephone: (+43-1) 404 20-7414 ; Fax: (+43-1) 404 20-7499.

[‡]Assistant Professor, Department of Economics, Harvard University; Address: Littauer Center 121, Cambridge, MA 02138; e-Mail: maximiliankasy@fas.harvard.edu; Telephone: 617-495-2896; Fax: 617-495-7730.

[§]Economic Analysis Division, Oesterreichische Nationalbank; Address: Postbox 61, 1011 Vienna, Austria; e-Mail: peter.lindner@oenb.at; Telephone: (+43-1) 404 20-7425 ; Fax: (+43-1) 404 20-7499.

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 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 nonresponse and misreporting of income data is a concern.[1] This is of particular importance in the upper and lower tails of the distribution, and both of these problems might be even more severe in telephone interviews. Nonresponse and misreporting are important problems in the context of inequality measurement because many common inequality measures, such as the Gini coefficient, are not robust[2] (c.f. Cowell and Victoria-Feser, 1996), which 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, see Table 1, which suggests that comparisons across different surveys might be quite problematic.

In this paper, we use quasi-experimental 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 the estimated income distribution measured in several ways as e.g. the Gini coefficient and the 90/10 percentile ratio. 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. We find that a switch from CAPI to CATI leads to major changes in response behaviour, which imply large differences of estimated inequality measures. Selective item-nonresponse in the tails is significantly higher for CATI (on average by about 20%-30%), and incomes for CATI tend to be closer to the 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 (see Table 2). The smaller effect on the latter statistic is likely due to the fact that it is robust, whereas the Gini coefficient is not. A "Back of the Envelope" calculation in the last column of Table 1 illustrates these possible effects and leads to severe re-ranking of several countries. These calculations suggest that countries with considerable proportions of CATI interviews would

[1] Savage and Waldman (2008), studying different outcomes but motivated by similar concerns, investigate the effect of survey mode on respondent learning and fatigue during repeated choice experiments. They find that lower cost survey modes are associated with larger measurement errors.

[2] A statistic is called robust if it has a bounded influence function, see Huber (2003).

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 with 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 as coverage effects of the interview mode cannot be accounted for.

In our analysis of the effects of modes on the distribution of measured incomes we are using reweighting and regression methods developed in the literature on distributional decompositions, in particular by DiNardo et al. (1996) and Firpo et al. (2009). Our paper deviates from this literature in studying effects on measured, rather than actual, incomes. Survey methodology is not part of the standard economics curriculum, but economists become increasingly aware of the importance of careful data collection. This is partly due to the rise of experimental methods, and the consequent collection of primary data by development economists (see Duflo et al., 2008) and labor economists (see List and Rasul, 2011), but also true for economists using "subjective" data from large pre-existing surveys, see for instance Gabriella and Pudney (2011). This paper is intended to contribute to the rising awareness of data issues among economists.

The rest of this paper is structured as follows. In the next section we introduce the hypotheses in the background of the analysis and some more literature for it. It helps to clarify the mechanism that are at play in this specific part of the data production process and embeds the analysis in a wider context. The underlying data as well as the quasi-experimental set-up are laid out in the following part of the paper. It gives a detailed account on the switch of the interviewing mode and the measures that are used. Section 4 is dedicated to giving a short overview of the identification strategy. Additionally, we introduce the specific implementation of the estimation strategy. The main part of the paper, section 5, discusses the results reached from the investigation. We split it into the discussion of the effect of the interviewing mode on item-nonresponse, reported income level, and the distribution of income. Finally, section 6 sums up our results and concludes.

2 Interview Modes

Evidence on the impact of the interview mode is mixed. In a meta-study de Leeuw (1992) compared face-to-face interviews, telephone surveys, and self-administered mail questionnaires. By employing 67 mode-comparison-papers she showed that the main difference lies between self-administrated and interviewer-based survey modes, the differences between interviewer-based modes, and therefore also between CATI and CAPI

seem less pronounced. While there was no significant difference in response validity and social desirability bias, Face-to-Face interviews lead to slightly less item non-response. CATI is the oldest of the computer assisted interviewing methods and is heavily used especially in market research. CAPI is widely used especially for complex surveys of governmental statistic agencies and universities (de Leeuw (2008)).

> "Because of the increased availability of other survey modes, face-to-face interviews are typically reserved for the most difficult and longest surveys that place the greatest burden on respondents. These are all kinds of surveys for which the other modes are not so likely to perform well. Face-to-face surveys also tend to be reserved for surveys that are most important to society, for which sponsors are willing to pay the cost." (de Leeuw et al. (2008), p. 164)

The main advantage of face-to-face interviews is that they are more flexible than telephone interviews. The interviewer can use response cards, visual scales etc. but also explain things better by being physically present which allows for a broader range of communication and interaction between the interviewer and the respondent. Via telephone the respondent can only rely on his/her memory when answering questions with multiple answer possibilities.
Also interview length is an important determinant of the quality of the interviews given a certain mode. While CAPI and face to face in general is usually used for longer interviews (>30min) telephone interviewing is less suitable for longer interviews. Major data collection organizations in the US are refusing to conduct telephone interviews which are expected to last longer than 18 minutes (de Leeuw et al. (2008)).

We hypothesize that there are three different mechanisms through which the mode influences measured income inequality.
First, on the level of unit non-response. We find that even though higher income households tend to be selected more often towards a CATI interview the highest values are found for households interviewed via CAPI. The same is also true for the lower end of the distribution. Therefore it seems that via CATI the households at the tails of the income distribution can not be reached as well as via CAPI (see Table 4).
Second, on the level of item nonresponse. In addition to the effect of leading to generally higher item nonresponse the effect of CATI is particularly pronounced at the tails of the distribution. Especially at the top this is very worrisome as the share of total income held by the top income earners is much larger than their population share.
Third, on the level of the income values. We find that CATI leads to a positive effect on income values reported, where the effect is larger for lower income groups than for very high income groups, and leads to significantly lower measures of 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 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.

We can only quantify the third effect in terms of measures of income inequality 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 exist and is non-negligible. Therefore it is very likely that we underestimate the total effect of the interview mode on observed inequality measures as both the first and second effect also seem to compress the observed 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 combination of all three sources of measurement error seems to be larger when CATI instead of CAPI is used as an interview mode. The decision to not report at all (unit nonresponse), 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.

3 Quasi-Experiment and Data

For this empirical analysis we use Austrian EU-SILC data[3] of the waves 2007 and 2008. The documentation on the data is provided by Statistik Austria.[4] In addition to the data available in the standard user dataset we obtained an indicator for both years describing whether a specific household was interviewed using the CAPI or the CATI interviewing mode from Statistik Austria. We use both waves, as the CATI option was first introduced in 2008, after a test period for 2007 with households that are excluded as is described below. Only households that already were interviewed in 2007 could choose between a personal or telephone interview. The preference of the data producers was to use the cheaper CATI option and first contact was - wherever possible - established via phone. We model the selection (see Section 4.2) and control for it when estimating the effect of interview modes, as discussed below.[5] Table 3 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 of which about 30% (i.e. 1,710 households) are surveyed via a CATI interview. Only two thirds are panel households in the sense that they were also interviewed in the previous wave. Thus the effective sample size is reduced to 3,377 households which could be (self-) selected to one of both interview modes. 42.6% of these actually were interviewed via telephone while the rest

[3]See `http://epp.eurostat.ec.europa.eu/portal/page/portal/microdata/eu_silc` for further informations [accessed on 13th of November 2012].

[4]See Statistik Austria (2010) for a version in German.

[5]According to the handbook (see Statistik Austria (2010)) there were about 160 and 40 interviewer respectively for the CAPI and CATI part of the survey. Information on which household was interviewed by which interviewer, to control for possible interviewer effects, however, is unfortunately not part of the EU-SILC data. There seem to be enough interviewers, however, for the possible interviewer bias being rather small.

opted for a personal survey.[6]
In EU-SILC 2008 selection towards CAPI is characterized by the following rules.[7] Households that re-participate, i.e. already took part in a previous wave, in the 2008 EU-SILC-Survey were supposed to answer the questionnaire via the telephone. This preference was implemented through a first contact (after the sending of an information letter) by calling the respondent and asking the household to participate via telephone. The CAPI-mode was available for households refused to take part in the survey over the phone, could not respond on the phone due to illness or age related reasons, or could either not be reached over the phone (i.e. wrong number, not available on the phone, etc.) or set dates were postponed by the respondent for several times.[8] All CATI-interviews were conducted during weekdays (Monday to Friday) from 4pm to 8pm and additionally on Tuesdays between 9am and 1pm.[9] In general there seems to be a tendency to push households to conduct interviews on the phone while no unit-non-response is allowed due to the wish of a respondent to conduct a CAPI-interview.

To sum up: All households were interviewed in 2007 via CAPI. All of them were targeted to be reinterviewed via CATI in 2008. Due to accessibility problems via phone and the possibility to opt for the CAPI mode roughly 40% of them were interviewed via CATI and 60% via CAPI. We are able to control for the non-random part in treatment assignment using the information which was gathered in 2007 for all households with the same (CAPI) mode, including income and item nonresponse information, to estimate the effects of the treatment - the interview mode - on the observed income distribution in 2008.

Data from the 2007 wave (instead of the 2008 wave) are used as controls because they are pre-determined and hence by construction not exposed to a mode effect. Thus the controls are no outcome of the selected mode and can be considered strictly exogenous to the mode. Descriptive statistics of the controls by sub-samples are given in Table 4.

3.1 Measures

Our analysis mainly focuses on disposable income. Household disposable income is constructed by summing up all of the households income sources. EU-SILC provides

[6]There is a minor technical issue with 15 households that were not part of the 2007 wave, were, however, interviewed over the telephone in 2008. From the information provided by Statistik Austria in a personal correspondence this is due to the fact the households that were already part of the panel component in 2006, but could not be reached in 2007 (wave non-response), were possibly interviewed in the CAPI mode. Since this is only the case for the 2008-wave and was changed in subsequent waves to completely drop those households, we leave out these 15 households in the analysis. Furthermore as there exists no 2007 information on these 15 households we can not use them in our empirical exercise.

[7]This information is based both on the official documentation (see Statistik Austria (2010)) and personally transmitted material from Statistik Austria.

[8]Information from Statistik Austria: "CAPI interview mode was offered only if there is a refusal of the phone interview, there are health related difficulties, or age related reasons that make a telephone interview impractical."

[9]No general rules were set for interviewing time of CAPI-interviews.

flag variables to distinguish three categories with regard to necessary imputation of this variable, (i) not imputed, (ii) partly imputed, (iii) completely imputed. Whereas "not imputed" implies that all items the aggregate household level variable consists of, were provided by the respondents, "partly imputed" means that one or more items where missing and had to be imputed and "completely imputed" means that all items where missing and had to be imputed.

Using this information we construct a household income item non-response dummy (HINR) being 1 if the household disposable income flag is indicating 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 main household income variable.[10]

Disposable income[11] is taken as it is provided in the data. Table 4[12] reports in the upper part the mean of the variables in the panel households and the CAPI and CATI subsamples. We see that the average income is lower for households surveyed in a CAPI interview and the item non-response is 20%-30% lower for these 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 illustrates the problems originating from the telephone interview mode in terms of a possible compression of the income distribution via unit- and item nonresponse. CATI seems to increase coverage problems espcially at the tails of the distribution. Obviously none of the very high as well as very few low income households could be reached via CATI. Furthermore, it can be seen that both disposable and gross income increased from 2007 to 2008 on average 6% and 9% respectively. The difference in the change between the two groups, however, seems small compared to the difference of 2008 income, i.e. 2 percentage points relative to about 20% (non log).

Table 4 also provides the mean for all household level[13] and personal level control variables from the EU-SILC wave of 2007. There are some albeit generally small differences in the averages between the two groups. Most noticeable, more affluent households in terms of income (see the difference of log income in 2007) as well as wealth (see as two indicators the percentage of owning the primary residence and the size of the primary residence) seem to be (self-) selected to CATI. Additionally, phone coverage (see land

[10]Furthermore, as a robustness check, we repeated the analysis (results can be provided upon request) using a differently constructed measure of item non-response. The construction is based on the individual income components and is calculated as the percentage of components imputed (i.e. missing in the first place) for the total household income. The results are qualitatively very similar and left out due to space constraints.

[11]As a robustness check we also worked with gross income.

[12]These estimates are not weighted since we are not interested in population but only sample averages at this stage of the analysis.

[13]We also included regional indicators for of households as controls, but do not report them in Table 4 due to space constraints. 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%.

line and mobile phone coverage) and education seem to be lower for households with a CAPI-interview and 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 estimations reported below; see Section 4. Another insight into how strong the selection into the different interviewing modes is can be gained from the statistics in Table 5. According to income deciles in 2008 we see in the second column that the participation rate in a CAPI interview is decreasing with income. Stated otherwise households with higher income tend to opt at a higher rate for CATI-interviews. Also it is reported that item nonresponse is higher for higher income households. However, there is a large difference of more than 10 percentage points in the two top deciles. At the lowest end, in the first decile the difference is almost as high, and in between income from households with CATI-interviews always have higher item nonresponse, although the difference is not stable over the deciles. As can be seen in the last two columns, the percentage of households with missing income data does not follow the same trend in 2007. Here the two groups are defined according to the 2008 interview mode choice and we report item nonresponse in 2007 when both groups (CAPI as well as CATI) were still interviewed via CAPI. Again the percent of households with missing income data increases with income level. The difference between the groups however is not stable already suggesting that there might be indeed a interview mode effect at play. In the top decile the income of households that were interviewed via CAPI in 2008 was missing more often than for the other group.

4 Identification and Estimation Strategy

We estimate the effect of the interview mode (CAPI versus CATI) on the income distribution by combining methods from the literature on causal inference and program evaluation with methods from the distributional decomposition literature. We also estimate the effect of the interview mode on item non-response. As 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. In order to check the robustness of our results we employ various alternative estimation methods (OLS, OLS with controls, fully interacted model, propensity score matching, and coarsened exact matching techniques) to control for the selection into interview modes when estimating the effect of modes on item nonresponse and the distribution of observed incomes.

Even small effects of the interview mode on some parts of the observed income distribution might translate into economically significant effects on measures of income inequality. After a naive comparison of various distributional statistics across interview modes, we follow the approach suggested by Firpo et al. (2009) and use recentered influence functions (RIF) to estimate the effect of the mode (i.e. CAPI vs. CATI) on various distributional statistics $\nu(F(Y))$, where $F(Y)$ is the unconditional distribution

of household income. For all distributional effect estimates we weight observations to obtain a sample representative of the Austrian population. The next subsection provides a review of reweighting and RIF regression methods, and describes how we relate them to program evaluation methods. We then discuss various alternative estimation methods which we use in order to confirm the robustness of our results.

4.1 A review of reweighting and RIF regression

Suppose we observe a cross-section with 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 again 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 variance, the share below the poverty line, quantiles or the Gini coefficient.

Let $P^d(Y|X)$ 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 *sub-population* 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 unionisation and the minimum wage etc., which were analyzed in DiNardo et al. (1996). This counterfactual distribution can be interpreted causally under an assumption of conditional independence. Denote Y^m the potential reported income of an individual that she would report if interviewed using mode M. 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 not unproblematic, but reasonably credible with a rich set of covariates (including lagged Y from the 2007 survey), as we have at our disposition. The remaining question is how to select the vector of covariates X. This set of controls needs to be chosen in a certain way to reasonably ensure the CIA but without introducing bad control bias, which is another form of selection bias (see e.g. Angrist and Pischke (2008)). Good controls are in our case variables which are themselves not a (possible) outcome of the mode. A safe way to prevent the danger of using controls which could induce selection bias is to use controls which are themselves already fixed at the time of mode selection or strictly exogenous variables in relation to the mode (Imbens (2004)). In our case, given the panel structure of the data at hand, all characteristics of a certain household i that

were gathered already before the interview and thus before the possibility of the two different interview mode (in 2008) are natural candidates (2007 wave variables). The panel data allow us in particular to control for household income in 2007, i.e. before the introduction of the CATI interview mode.

We can rewrite the distribution P^m as

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

where

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

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).

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 obtained from

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

In general, ν will not have this linear form but can be approximated by a linear first order expansion around 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 θ^m 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. Ignoring the remainder, this representation of ν has the linear form required for the use of the representation (5), i.e.,

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

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).

For the first approach, we need to estimate the conditional probability $P(M = m|X)$

We use a logit model with flexibly specified regressors for the distribution of M given X:

$$P(M = 1|X) = \frac{\exp(X \cdot \beta^M)}{1 + \exp(X \cdot \beta^M)}. \tag{8}$$

Based on estimates of the parameter vector β^M, we can calculate the weights θ^m, as

$$\theta^1 = \frac{\mathbf{1}(M = 1)}{P(M = 1|X)} = \mathbf{1}(M = 1) \cdot (1 + \exp(-X \cdot \beta^M)), \tag{9}$$

and

$$\theta^0 = \frac{\mathbf{1}(M = 0)}{P(M = 0|X)} = \mathbf{1}(M = 0) \cdot (1 + \exp(X \cdot \beta^M)). \tag{10}$$

For the influence-function regression approach, we need estimates of $E[IF|X, M]$. We run, in particular, the following regression, with full interactions between X and M,

$$IF = (X \times M) \cdot \beta^{IF} + \epsilon, \tag{11}$$

and assume $E[IF|X, M] = (X \times M) \cdot \beta^{IF}$.

We also use various alternative estimation methods popular in the program evaluation literature (in particular propensity score matching and coarsened exact matching) in order to check the robustness of our results, see section 4.3 below.

Note that, while the above approach is parametric, identification does not rely on the parametric choices: both the logit specification for $P(M|X)$ and the linear specification for $E[IF|X, M]$ are in fact "nonparametric" if we allow for sufficiently rich interactions and powers between the components of X and M. Furthermore, following the arguments of Newey (1994), the choice of nonparametric estimator for this "first stage," if it is consistent, does not affect the asymptotic variance of root-n estimable parameters. This covers all our examples for ν, except for the counterfactual densities. Confidence sets for all estimators are obtained by the delta method, but bootstrapping the entire procedure yields very similar standard errors. Additionally, we also include the estimates for the nonparametric coarsened exact matching approach.

Influence functions of measures of inequality

We estimate the effect on several popular measures of income inequality, in particular the Gini-Coefficient and the poverty rate (the share of the household population with less than 60% of the median income), as well as the $P90/P10$-percentile ratio. For the RIF regression approach, we need to derive the influence functions of these measures.

The Recentered Influence Function (RIF) for the Gini-Coefficient (see e.g. Firpo et al. (2007) page 24f.) is given by

$$RIF(y, \nu^{GI}, F^0) = 1 + \frac{2}{\mu^2} R(F_y) y - \frac{2}{\mu} \left[y(1 - p(y)) + GL(p(y), F_y) \right] \tag{12}$$

where $R(F_y)$ is the integral of the Generalized Lorenz curve with the ordinates $GL(\bullet)$, and $p(y)$ is the density.

The poverty rate is defined as $\nu^{PR}(F) = F\left(0.6 \cdot F^{-1}(0.5)\right)$, where $med := F^{-1}(0.5)$ is the median for the distribution given by F and $0.6 \cdot F^{-1}(0.5) = 0.6 \cdot med$ is the poverty ceiling. The RIF for the poverty rate is equal to

$$RIF(y; \nu^{PR}, F^0) = \mathbf{1}(y \leq 0.6 \cdot med) - \frac{0.6 \cdot f(0.6 \cdot med)}{f(med)} \cdot \mathbf{1}(y \leq med). \tag{13}$$

The $P90/P10$-percentile ratio is defined as $\nu^{RA} = \frac{Q^{90}}{Q^{10}}$ where Q^i is the i's percentile of the income distribution. The RIF of the $P90/P10$-percentile ratio equals

$$RIF(y; \nu^{RA}, F^0) = \nu + \nu * \left[\frac{\mathbb{1}(y \leq Q^{10})}{Q^{10} * f(Q^{10})} - \frac{\mathbb{1}(y \leq Q^{90})}{Q^{90} * f(Q^{90})} \right]. \tag{14}$$

4.2 Preliminaries: (Self-) Selection into interview modes, and non-response

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 4.[14] Table 6 shows the average marginal effects calculated from this model. Additionally, Figure 1 displays the support of the propensity scores as implied by the logit model. As can be seen in Figure 1 the support of propensity scores for CATI- and CAPI-households is nearly identical. However, the chance of (self-) selection to CAPI decreases (statistically) significant 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.

We next estimate logit regressions of HINR (household item non-response) on interview mode and household-level as well as person-level controls, and similar regression using household income as dependent variable.

4.3 Alternative estimation methods

In this section, we discuss various alternative estimation approaches implementing the ideas reviewed in section 4.1. We first estimate a fully interacted linear model (FILM),[15] which allows the effect of the mode to vary over all controls. We then estimate treatment effects based on propensity score matching (PSM).[16] While FILM allows the mode

[14] We also control for regional dummies; the coefficients on these are not significantly different from 0.

[15] We use *film*, a STATA program provided by Edwin Leuven and Barbara Sianesi, see `http://www.ifs.org.uk/publications/2712` [accessed on 19th of November 2012].

[16] We use *psmatch2*, a STATA program provided by Edwin Leuven and Barbara Sianesi, see `http://www.ifs.org.uk/publications/2684` [accessed on 19th of November 2012].

effects to vary over all controls, propensity score matching additionally allows (i) to impose common support based on the overlapping regions of the propensity scores (see Figure 1) and (ii) does - due to its semi-parametric nature - not impose as strong linearity assumptions as logit and FILM are imposing. On the imposed common support we match the nearest - in terms of the propensity score - CATI-neighbour to every CAPI-household (1 to 1 matching) in order to balance the joint distribution of the covariates (controls).[17]

We use coarsened exact matching (CEM) as a further robustness check. 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.[18]

While PSM still uses a parametric model to match CAPI and CATI observations and therefore extrapolates outside the common support[19] in order to calculate treatment effects, 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 but of course comes with a decrease in sample-size.

As matching variables we use a subset of our household and personal level covariates. For categorical variables (household size and being female as the household head) we impose an exact matching strategy and for the two continuous variables we use a coarsening strategy for matching on certain parts of the distributions by imposing cut-points (household level: household disposable income 20000, 30000, 40000, 50000, 60000; and personal level: age 30, 40, 50, 60, 70). This matching strategy leads to a perfectly balanced dataset in terms of the joint distribution of the categorical variables. As household disposable income is allowed to be matched in approximately 10,000 Euro brackets (and below 20,000 Euro as well as over 70,000 Euro) and age is allowed to be matched in about 10 year age brackets (and below 30 as well as older than 70) some imbalances remain with respect to those two variables. Out of the 343 covariate combinations defined we find 224 which define the common support, i.e. where at least one CAPI and one CATI observation can be found. In terms of the sample the total of 3,377 observations collapses to 3,190 observations which lie inside the common support. 63 CATI- and 124 CAPI-observations lie outside the common support. The weights for the matched observations are chosen in a way that CAPI and CATI observations are balanced for each covariate combination. To control for the remaining imbalances in

[17]Note that propensity score matching is not exact matching. That means, that the common support imposed by the propensity score differs from the common support in the strict sense which would include only cells of covariate combinations which include both CAPI- and CATI-households (the later is a subset of the former.). As exact matching is not feasible given finite data and many (and including continous) covariates, the resulting model still extrapolates also for cells which do not include both CAPI- and CATI-households.

[18]We use *cem*, a STATA program provided by Matthew Blackwell, Stefano Iacus, Gary King and Giuseppe Porro, see `http://ideas.repec.org/c/boc/bocode/s457127.html` [accessed on 19th of November 2012].

[19]With relation to realizations of CAPI- and CATI-observations in cells defined by covariate combinations and not with relation to the common support of the propensity scores.

the matched dataset we still use age and household disposable income as controls when we calculate the treatment effects.

5 Results

5.1 Mode effects on income item non-response

Table 7 shows the estimated effects of CAPI on item non-response. All estimated effects resulting from the logit regressions are significant at a 1% significance level, but are not significantly different from each other indicating again that selection bias is rather small. The estimate using the largest set of controls, i.e. all household- and personal-level controls given in Table 4, is −0.071 which indicates 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.

The estimates resulting from the FILM and PSM estimations closely resemble the logit estimates. The FILM and PSM estimates are −0.076 and −0.072 respectively and significant at the 5% level. 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.

5.2 Mode effects on the income level

Analogous to the estimates of the mode effects on item nonresponse we estimate the effect of the mode on household income again with increasing flexibility and less restrictive assumptions. All estimates shown in Table 8 are produced analogously to the estimates in Table 7, but with the logarithm of household income of 2008 being the dependent variable instead of HINR. Overall we find a significant negative effect of the CAPI mode on average (log) household income. The estimate ranges between −0.02 (propensity score matching) and −0.2 (OLS without controls). This estimate (taking as an example 0.04) implies that interviewing via CAPI leads on average about 1,000 Euro lower reported income values than interviewing 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. Again the more flexible methods FILM, PSM and CEM confirm the negative effect, however the PSM estimate is not significant.
The average effect, however, is of no help if one wants to analyse the impact on the entire distribution and resulting inequality measures. In fact it might be quite misleading since the effect on inequality measures is positive instead of negative. Furthermore,

even many small and insignificant effects over the distribution could accumulate to relevant effects on inequality measures.

5.3 Mode effects on the income distribution

Differences

Starting from a simple idea in order to compare the whole distribution in more detail we calculate the difference of a given statistic for both sub-samples,[20] i.e. for households interviewed with CAPI and CATI respectively, and provide standard errors of this difference using a bootstrap with 500 replicates.[21] We apply this procedure to the GINI-coefficient; the poverty rate, i.e. the proportion of population with lower income than 60% of the median; and the 90/10 percentile ratio. Additionally, we provide the estimates of the difference between the CAPI- and the CATI sample statistics (i) without any adjustments and (ii) on the common support resulting from the CEM-Procedure, which controls partly for selection bias but at the cost of a reduction of the sample.

Table 9 reports first the measures of the Gini-Coefficient, the Poverty Rate, and the 90/10 percentile ratio for both the CAPI and CATI sub-samples in EU-SILC 2008. We find that the difference of the Gini-Coefficient is 0.026 for the full sample and 0.024 using only the common support sample implying a 8.1% and 7.2% lower Gini-coefficient using CATI instead of CAPI. While the estimate of the whole sample is significant at the 5% level, the difference for the balanced sample is only significant at the 10% level.[22] Table 9 additionally reports a significantly higher poverty rate for the CAPI sample (i.e. 3.9 percentage points in the whole and 3.1 percentage points in the balanced sample) and an (statistically) insignificant difference of the 90/10-percentile ratio between the two sub-samples. To control rigorously for selection we employ RIF-regressions to estimate the effect of interview mode on the distributional statistics of income.

Overall Effect

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 households characteristics (including income). In both cases also all interactions of the control variables with the interview

[20] Appropriate weights are used in the calculation of the statistics, since we are now interested in the effect on the population and not, as before, in the sample.

[21] For the estimations the whole procedure is bootstrapped in the sense that a bootstrap sub-sample is drawn, and then the difference of the statistic calculated. With these 500 estimates of the differences we are able to estimate standard errors.

[22] To check for the differences at the top we also used a General Entropy Class index with $\alpha = 2$ which shows huge differences (0.18 for the CATI versus 0.49 for the CAPI common support samples). This is due to the fact that it is very sensitive to top income and as we saw the highest income observations are all from CAPI interviews. This sensitivity, however, also renders a high variability of replicates in this bootstrapping procedure and thus generates high standard errors yielding insignificant results.

modes are included in the model. For the standard errors of the reported ATE we use the delta method.[23] Using CATI instead of CAPI as interview modes reduces the observed Gini-Coefficients significantly (see Table 2). According to our estimates of 0.033 with the limited set of controls and 0.027 using the full set of controls, inequality measured by the Gini Coefficient drops by around 10% if the interview mode is switched from CAPI to CATI. This is a severe effect on the most prevalent inequality measure. Furthermore, we see that the estimate on the poverty rate decreases greatly and looses its significance and the 90/10-percentile ratio switches signs but remains insignificant. The result suggests that using the latter two measures when comparing inequality over time or between countries might be more robust.

Heterogeneity of the effect

The estimates for the difference in percentiles are reported in Figure 2. Panel (a) to (c) show that the effect of the interview mode on the percentiles[24] 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 before) in Panel c). The u-shape means that the percentiles are lower for households interviewed with CAPI but less so (and in the more controlled version even higher) at the extremes of the distribution. This implies a higher spread of income in households interviewed with CAPI. Our understanding is, that there are two mechanism that play an important role here. First, it is easier to lie over the phone, hence biasing reported income at the extremes of the distribution towards the mean; and second, households at the tail of the distribution are especially hard to interview, and thus are mostly interviewed via CAPI (whilst in a CATI survey these households, however important they are for a correct estimate, do not take part at all).

Panel d) of Figure 2 shows the 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, i.e. that households at the extremes of the distribution are only covered with CAPI interviews, is not accounted for in this estimation.

[23] Thus the standard errors are based on the whole sample of the 2008 wave of EU-SILC and not only on the panel component. Bootstrapped standard errors using 500 replicates were also calculated as a robustness check and yield very similar results.

[24] The effect is estimated for 20 quantiles.

6 Conclusion

In this paper we exploited the availability of panel data where the interview mode changed during the panel, which occurred in the 2008 wave of the Austrian EU-SILC data for some - but not all - panel households. The quasi-experimental nature of these data allows us to estimate causal effects of this change of the interview mode. 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.

First, we find descriptive evidence that CATI compresses the income distribution by leading to less coverage in the final sample via higher unit-nonresponse 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 statistical as well as real terms. This result is robust over all the parametric, semi-parametric and non-parametric methods - which allow for a great degree of flexibility - we applied. 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. Again CATI tends to compress the observed income distribution - in this case via item nonresponse especially at the tails. Third, we find that households which are interviewed by CATI on average report higher income values. 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. Compared to CAPI the effect of CATI therefore is a mean-reverting one, again compressing the income distribution. Ultimately we find that this level effect leads to a severe bias with regard to income inequality 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%.

Given these results in terms of the effects of the mode of on income distribution we advocate the careful use of survey data taking into account different interview modes and being aware of the possibility that differences between countries and within countries over time might be due to yet another candidate, namely the interview mode. Commonly used rankings of countries by Gini coefficients (see OECD (2011) or Table 1) of the income distribution might largely be an artifact of different survey techniques such as the interview mode rather than true differences in income inequality.

Appendix A Figures and tables

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.

Table 2: Effect of Interview Mode an Inequality

	Gini Coeff.		Poverty rate		90/10 Percentile Ratio	
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	X	X	X	X	X	X
other covariates		X		X		X

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 sub-samples and the effect of the interview mode using RIF-regressions.
(ii) Standard errors are reported using delta methods.
(iii) *Source:* EU-SILC 07/08.

Table 3: 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 Household 2007".
(iii)*Source:* EU-SILC 07/08.

Table 4: Mean of Dependent and Explanatory Variables

	Panel	CAPI	CATI
Dependent variables: 2008			
Mean household disposable income	10.24	10.16	10.37
Mean household gross income	10.51	10.42	10.66
Household Item nonresponse	21.84	18.08	27.53
Share of interviews below 10k	16.52		
Interviewed by		74.80	25.20
Share of interviews above 200k	0.07		
Interviewed by		100.00	0.00
Mean difference disposable income to 2007	0.06	0.07	0.05
Mean difference gross income to 2007	0.09	0.10	0.08
Control variables at the household level: 2007			
Household size	2.35	2.28	2.45
Households with kids	31.06	29.27	33.65
Single family home	39.34	37.16	42.48
Home-owners	53.51	49.40	59.43
Size of Flat in sqm	97.96	92.60	105.70
Land line coverage	66.16	58.14	77.74
Mobile phone coverage	87.30	85.18	90.36
Mean household disposable income	10.18	10.09	10.32
Mean household gross income	10.42	10.32	10.58
City	38.44	38.18	38.81
Urban	26.17	26.95	25.04
Rural	36.03	35.73	36.46
Control variables at the level of the household head: 2007			
Female	45.87	44.03	48.52
Blue collar	43.44	42.46	44.85
Self-employment	6.21	5.77	6.83
Jobless	50.36	51.77	48.32
Weekly working hours	19.91	19.84	20.02
Married	49.69	45.52	55.71
Education: secondary school	21.35	25.91	14.76
Education: apprenticeship	52.54	52.68	52.34
Education: higher secondary school	16.04	13.13	20.25
Education: university	10.07	8.28	12.66
Age	52.39	51.86	53.14

Notes:
(i) This table shows the mean of the variables on the consideration as well as 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.

Table 5: Structure of Missingness of Income over Deciles Conditional on Interview Mode

		2008		2007	
Interview mode		CAPI or CATI		CAPI	
Interview mode in 2008	% CAPI Int	CATI	CAPI	CATI	CAPI
Total	.	0.271	0.181	0.271	0.234
First Decile	0.691	0.233	0.135	0.223	0.157
Second Decile	0.715	0.124	0.108	0.124	0.188
Thrid Decile	0.629	0.174	0.141	0.215	0.161
Fourth Decile	0.596	0.241	0.143	0.158	0.173
Fifth Decile	0.557	0.203	0.118	0.250	0.237
Sixth Decile	0.592	0.223	0.214	0.304	0.260
Seventh Decile	0.556	0.278	0.190	0.311	0.243
Eigth Decile	0.520	0.295	0.267	0.301	0.233
Ninth Decile	0.467	0.341	0.236	0.358	0.344
Tenth Decile	0.441	0.435	0.316	0.330	0.424

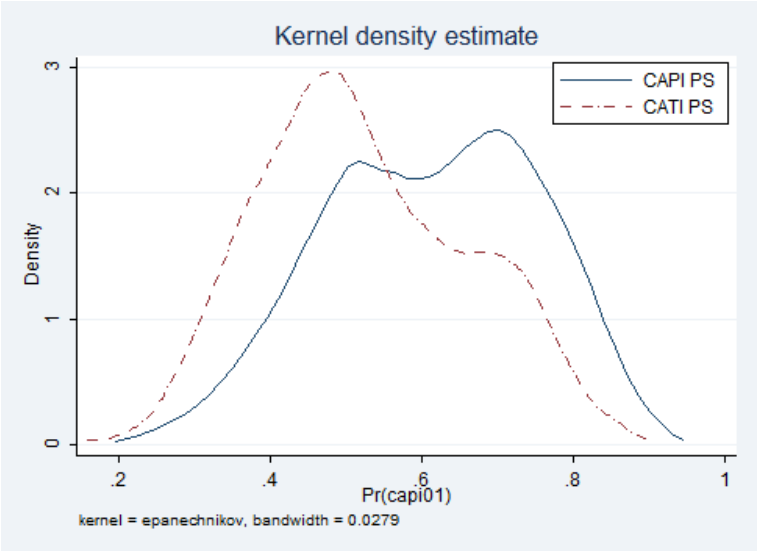
Notes:
(i) This table shows the average use of the CAPI interviewing mode over deciles (column 2).
(ii) Additionally, one can see the average rate of missing data on the income variable for 2008 and 2007 over the deciles (column 3 and 4, as well as 5 and 6 respectively).
(iii) *Source:* EU-SILC 07/08.

Table 6: Logit-regression of the (Self-)Selection of the Mode on Control Variables

Selection towards CAPI	
	Mode Selection
household characteristics	
Household size	0.031
	(0.012)
Household with kids	-0.041
	(0.028)
Single family home	0.023
	(0.022)
Owner occupier	-0.028
	(0.021)
Living space in sqm	-0.001
	(0.000)
Land line	-0.178
	(0.020)
Mobile phone	-0.097
	(0.029)
Log-disposable income	-0.052
	(0.018)
Personal characteristics of household head	
Female	-0.072
	(0.018)
Self-employed	0.009
	(0.037)
Jobless	0.079
	(0.037)
Weekly working hours	0.002
	(0.001)
Married and living together	-0.068
	(0.022)
Education: apprenticeship	-0.090
	(0.023)
Education: higher sec. school	-0.191
	(0.030)
Education: university	-0.131
	(0.036)
Age	-0.001
	(0.001)
N	3376

Notes:
(i)This table shows average marginal effects (AME) of household characteristics of the 2007 EU-SILC wave on being interviewed by CAPI in the 2008 EU-SILC wave. All average marginal effects are calculated from a logistic regression using an CATI(0)/CAPI(1) as dependent variable. Furthermore we controlled for federal states and regional population density. The coefficents of both are not significant.
(ii) Standard errors calculated by the delta method are given in parentheses.
(iii) *Source:* EU-SILC 07/08.

Figure 1: Propensity score density of CAPI and CATI



Notes:
(i) This graph shows estimated propensity score densities resulting from the logit model presented in table 6.
(ii) *Source:* EU-SILC 07/08.

Table 7: Interview Mode Effect on Income Item Non-Response

	I	*II*	*III*
Logit Model	-0.088	-0.075	-0.071
	(0.014)	(0.014)	(0.015)
Fully Interacted Model			-0.076
			(0.016)
Propensity Score Matching			-0.072
			(0.022)
Coarsened Exact Matching			-0.068
			(0.014)
N	3291	3290	
Household Controls		X	X
Personal Controls			X

Notes:
(i)This table shows average partial effects (APE) of being interviewed by CAPI on household and personal income item non-response. 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] as dependent variable) as well as a fully interacted model and various matching techniques.
(ii) Standard errors are given in parentheses (for the standard errors of the marginal effects that delta method is applied).
(iii) *Source:* EU-SILC 07/08.

Table 8: Interview Mode Effect on Household Income

	I	*II*	*III*
Average effect of CAPI on log household income			
OLS-Regression	-0.205	-0.052	-0.040
	(0.022)	(0.016)	(0.015)
Fully Interacted Model			-0.043
			(0.016)
Propensity Score Matching			-0.017
			(0.032)
Coarsened Exact Matching			-0.060
			(0.017)
N	3377	3376	
Household Controls		X	X
Personal Controls			X

Notes:
(i) This table shows the effect (regression coefficient, as well as matching estimators) of being interviewed by CAPI on the logarithm of household disposable income.
(ii) Standard are given in parentheses.
(iii) *Source:* EU-SILC 07/08.

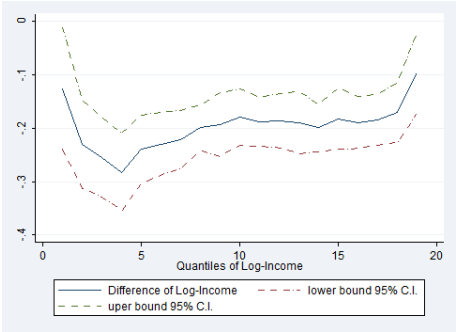
Table 9: Differences of Inequality Measures between Interview Modes

	Gini Coeff.		Poverty rate		90/10 Percentile Ratio	
	Ia	*IIa*	*Ib*	*IIb*	*Ic*	*IIc*
Inequality Measure CAPI	0.3341	0.3332	0.2280	0.2226	4.801	4.791
Inequality Measure CATI	0.3076	0.3096	0.1885	0.1920	4.523	4.527
Difference (Bootstrap)	0.026	0.024	0.039	0.031	0.278	0.263
Bootstrapped Std. Err.	(0.012)	(0.013)	(0.014)	(0.013)	(0.218)	(0.267)

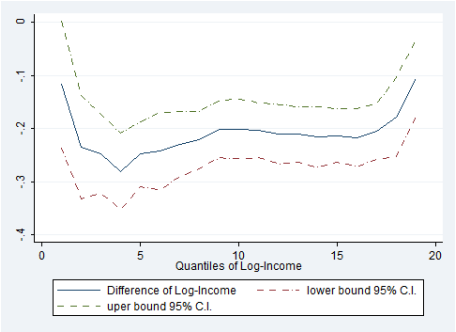
Notes:
(i) This table shows the effect of the interview mode on aggregate measures of inequality. We report the inequality statistic (Gini Coefficient, the Poverty Rate, and the Percentile Ratio), and the difference between the sub-samples.
(ii) Standard errors are reported using bootstrapping method.
(iii) Columns *Ia-c* show the results using the full (panel) sample, i.e. 3.377 observation; and columns *IIa-c* only use the matched observation from the above coarsened exact matching procedure.
(iv) *Source:* EU-SILC 07/08.

Figure 2: Average treatment effect on quantiles

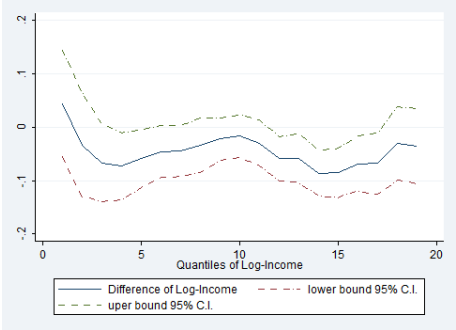
(a) Difference in percentiles



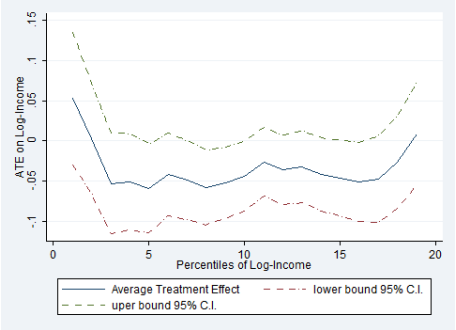
(b) Common support



(c) Reweighting



(d) RIF-regression on percentiles



Notes:
(i) This graphs show the effect of the interview mode on the percentiles (20 percentiles were used) over the whole income distribution. Panel a) displays the simple difference of the percentiles, 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.

References

Angrist J, Pischke JS. 2008. *Mostly harmless econometrics*. Princeton University Press, Princeton New Jersey.

Cowell FA, Victoria-Feser MP. 1996. Robustness properties of inequality measures. *Econometrica* **64**: 77–101.

de Leeuw ED. 1992. *Data Quality in Mail, Telephone and Face to Face Surveys*. Amsterdam: TT-Publikaties, Amsterdam.

de Leeuw ED. 2008. The effect of computer-assisted interviewing on data quality: A review of the evidence. Methodika/department of methodology and statistics, utrecht university.

de Leeuw ED, Hox JJ, Dillman DA. 2008. *International Handbook of Survey Methodology*. European Association of Methodology: Psychology Press, Taylor & Francis, New York.

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

Duflo E, Glennerster R, Kremer M. 2008. Using randomization in development economics research: A toolkit. *Handbook of Development Economics* **4**: 3895–3962.

Firpo S, Fortin N, Lemieux T. 2007. Decomposing wage distributions using recentered influence function regressions. Unpublished working paper, University of British Columbia.

Firpo S, Fortin NM, Lemieux T. 2009. Unconditional quantile regression. *Econometrica* **77**: 953–973.

Gabriella C, Pudney S. 2011. Survey design and the analysis of satisfaction. *Review of Economics and Statistics* **93**: 1087–1093.

Huber PJ. 2003. *Robust Statistics*. Wiley Series in Probability and Statistics. John Wiley & Sons. ISBN 9780471650720.

Iacus SM, King G, Porro G. 2008. Matching for causal inference without balance checking. Working paper series, Harvard.

Imbens GW. 2004. Nonparametric estimation of average treatment effects under exogeneity: A review. *The Review of Economics and Statistics* **86**: 4–29.

List JA, Rasul I. 2011. Field experiments in labor economics. *Handbook of Labor Economics* **4**: 103–228.

Lohmann H. 2011. Comparability of eu-silc survey and register data: The relationship among employment, earnings and poverty. *Journal of European Social Policy* **21**: 1–18.

Newey WK. 1994. The asymptotic variance of semiparametric estimators. *Econometrica* **62**: 1349–1382.

OECD. 2011. *Divided We Stand - Why Inequality Keeps Rising.*

Savage SJ, Waldman DM. 2008. Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics* **23**: 351–371.

Statistik Austria. 2010. *Standard Dokumentation Metainformationen - Definitionen, Erlaeuterungen, Methoden, Qualitaet.*