

Does Repeated Measurement Improve Income Data Quality?*

PAUL FISHER†

†*Institute for Social and Economic Research, University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK (e-mail: pfisher@essex.ac.uk)*

Abstract

This paper exploits a natural experiment created by a survey design to show that the quality of income data systematically changes across waves of a panel. We estimate that the effect of being interviewed for a second time, relative to the first, is to increase mean monthly income by 8%. Dependent interviewing – a recall device commonly used in panel surveys – explains one third of the observed increase. The remaining share is attributed to changes in respondent behaviour (panel conditioning). We review the evidence for and against a reporting improvement vs. a behavioural response by survey participants.

I. Introduction

An extensive body of work in economics is underpinned by household survey data on incomes. Recent examples include: Chetty *et al.* (2017) [Current Population Survey (CPS)], Hoynes, Schanzenbach and Almond (2016) (Panel Survey of Income Dynamics), Yang (2017) (Survey of Income and Program Participation), and Blundell *et al.* (2016) (British Household Panel Survey). Distinctly, smaller scale household panel surveys are regularly used to analyse randomized trials and relevant income variables. Recent examples include: Haushofer and Shapiro (2016), Banerjee *et al.* (2015) and Karlan and Zinman (2011). Government statistics on the income distribution are commonly based on large household surveys, for example, the official US poverty rate is based upon the CPS. Official income statistics may also rely on panel data, where they monitor changes in income over a lifetime. Examples include official statistics of the European Union (European Union Income and Living Conditions survey); and the UK Government (UK Household Longitudinal Study).

JEL Classification numbers: C83, D31, I32

*Thanks to Tom Crossley for reading drafts of this article and providing insightful feedback. I thank Jonathan Burton; Mike Brewer; Nick Buck; Brian Bucks; Emanuele Ciani; Annette Jackle; Peter Lynn and Steve Pudney for their useful comments. I also thank seminar participants at the ‘Improving the measurement of household finances’ workshop, University of Essex; and the International Association for Research in Income and Wealth conference 2016, Dresden, Germany. Financial support received from the ESRC is also gratefully acknowledged, award numbers ES/K005146/1, ES/N00812X/1 and ES/N006534/1. Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service.

This paper exploits a natural experiment created by a survey design to study how measurement error in income evolves with repeated interviewing in a panel survey. Previous research has shown that state transfers and self-employment income are under-reported in household surveys (Meyer and Sullivan, 2003, 2011; Lynn *et al.*, 2012; Hurst, Li and Pugsley, 2014; Meyer and Mittag, 2015; Brewer, Etheridge and O’Dea, 2017), but there is little evidence on whether measurement error is on average stable across waves of a panel. We find that the quality of measured income systematically changes across the early waves of a leading panel survey. If the changes represent reporting improvements, then this suggests a major benefit of repeated interviewing. Irrespectively, estimates of distributional change based on the early waves of the panel will confound true changes with data quality changes and be biased.

Panel experience may affect a given respondent’s income report, for a given year, for two reasons. First, panel conditioning (PC) effects may operate where panel participants change their behaviour (reporting or economic) as a result of being part of the panel. PC improves data quality if it reflects a respondents improved understanding of the questionnaire content or a growing trust in the interviewer or data holders. PC reduces data quality if respondents learn to strategically answer questions with the aim of reducing the interview length. Related, the data will become unrepresentative if survey participation leads to changes in real behaviour (Das and Leino, 2011; Zwane *et al.*, 2011; Crossley *et al.*, 2017; Bach and Eckman, 2018). Second, dependent interviewing (DI) – a tool that survey respondents of their reports at the previous interview – will lead to differences in data quality between the baseline and subsequent interviews where it takes effect.

Few studies have examined the stability of measurement error in income across waves of a panel. David and Bollinger (2005) find that false negative reporting of US food stamps is highly correlated across wave one and two of the Survey of Income and Program Participation. Das, Toepoel and van Soest (2011) propose a methodology to quantify PC effects. They compare responses of first-time responders in refreshment samples to responses from experienced panel members and make assumptions on attrition.¹ In this spirit, Halpern-Manners, Warren and Torche (2014) note that experienced panel members in the US General Social Survey are less likely to refuse questions about their income. Similarly, Frick *et al.* (2006) find that experienced panel members report higher income in the German Socio-Economic Panel. Nevertheless, despite the large interest of economists in living standards, we know of no study that has performed a systematic analysis of how measurement error in income and its components evolves across waves of a panel.

In this study, we provide evidence on the comparability of reported income across the waves of a large general purpose panel survey: the UK Household Longitudinal Study (UKHLS). Our approach is novel in that it exploits two features of the survey design to estimate the causal effect of panel experience on reported income (and does not require data linkage or refreshment samples). First, the fieldwork period for adjacent waves overlaps by one year. This generates a natural experiment in which randomly selected samples of individuals are interviewed at different waves of the panel but in the same calendar year. Second, UKHLS uses ‘reactive’ DI. This means that individuals are prompted with previous

¹Taking this approach, Van Landeghem (2014) finds a drop in a stated utility measure across the first-rounds of interviews in two panel surveys.

wave responses only after their initial response. Both their initial and final responses are recorded meaning that we can observe exactly which individuals would have failed to report an income source in the absence of DI. This allows us to decompose observed data quality changes into shares due to PC and to DI.

Our approach and data offer several further advantages over other studies exploring measurement error in income data that have used small scale administrative data linkage or refreshment samples (e.g. David and Bollinger, 2005; Lynn *et al.*, 2012). First, the large sample sizes available mean we can precisely estimate the effects of interest. Second, our data are representative of the Great Britain population (and not subsamples such as the poor or individuals covered by tax records) meaning that we can study how effects vary by representative subgroups of interest such as pensioners, working age groups and families. Third, our analysis covers a comprehensive set of income sources including earnings, investment income, and a total of 39 unearned sources. This enables us to identify precisely which income sources are most sensitive to prior panel participation.

Our main finding is that repeated interviewing changes the quality of income data (in particular for pensions and state transfers) and this occurs across the initial waves of the panel. A second interview, relative to the first, increases reported gross monthly income by £142 or 8% and about one third of the difference can be explained by DI, with the other two thirds due to PC. As to whether survey participants changed their reporting or economic behaviours, the evidence we present is inconclusive. However, it points to a large share (83%) of the total PC effect representing a reporting improvement. Indeed, we present evidence showing an upgrade in respondent confidence in the confidentiality of their sensitive data and falls in income refusal rates following a first interview. Separately, we present evidence of similar effects in another leading panel survey (British Household Panel Survey). More generally, the effects might be expected to extend to other sensitive areas of data collection.

Importantly, the effects we identify have substantive consequences for analysis of the income distribution. They lead to lower estimates of inequality and poverty but also income mobility.

The paper proceeds as follows: the next section describes the data. Section III discusses identification and the results are shown in section IV. Section V explores the mechanisms behind the results and section VI concludes. A supplementary appendix includes additional materials.

II. Data

UKHLS data

This paper uses data from the UK Household Longitudinal Study (UKHLS) that began in 2009. UKHLS is a large general-purpose social survey. It replaces the former BHPS as the data source for official UK Government statistics on poverty dynamics. We work with the main ‘General Population Sample Great Britain’ sub-sample, which is representative of the Great Britain household population at wave one.

The large UKHLS sample requires the fieldwork for each wave to be spread over 24 months. Households selected to take part in the panel were randomly allocated across the

24-month interview period of wave one. All adult members (aged 16 plus) of the households are interviewed every 12 months. These two features imply an overlap of one year in the fieldwork period of consecutive waves. This overlap is an issue we exploit in identification.

At wave one, interviews were conducted in 24,797 households with 41,586 individuals receiving an interview. As with all household panel surveys, there is an initial drop-off in individual response rates and 75.4% of wave one respondents completed an interview at wave two with a further 1.9% completing a proxy interview.²

Income variables and dependent interviewing

UKHLS includes a detailed set of questions on income. These are collected for each sample member in a face-to-face interview that is conducted by computer-assisted personal interviewing. A list of the income questions used in analysis is included in Appendix B.

Data collection of the income components occurs in different modules. An ‘employee’s’ module asks for gross (and net) pay at last payment and the usual pay if ‘last’ and ‘usual’ differed. A ‘self-employment’ module asks self-employees for their share of the profit or loss on their most recent accounts or, where not available, an estimate of their usual monthly or weekly self-employment income. All respondents receive the ‘second jobs’ and ‘household finances’ modules. The first asks about gross income from any second or odd jobs. The second records the amount received in interest and dividends in the last 12 months. None of the above modules make use of DI.

An ‘unearned income and state benefits’ module does use DI. It first asks respondents to examine a list of 9 broad types of payment and indicate which they are currently receiving. Respondents are then filtered to lists of specific sources (up to 39) where they indicate those that they receive.³ DI is used to check whether any sources not reported but reported at wave $t - 1$ are currently received. A final stage asks for the amounts for each reported source, the period it covered and whether the income was received solely or jointly.

One dimension of data quality is the extent to which respondents refuse to answer a question. Figure 1 plots trends in refusal rates (refusing to provide an amount), for respondents who completed a full-interview at each of waves one (2009–10) to five (2014–15).⁴ The figure is consistent with improving data quality as the panel ages, as all the refusal rates fall with the length of time the individuals have been in the panel. The biggest drop in the refusal rate occurs for self-employment profit which starts at 42.0% in wave 1 and reaches a minimum over the five waves of 34.2% at wave 4. The drop-off in refusal rates is notably sharper for all sources between waves 1 and waves 2. For example, refusal rates for self-employment profit fall from 42.0% to 37.1% and from 12.0% to 10.3% for earnings.⁵

² Reinterview rates for comparable panels are: BHPS: 86.0 (1991); HILDA: 87.0 (2002); GSOEP A: 88.1 (1985); GSOEP B: 81.0 (1985) and PSID: 92.7 (1969). The year each survey started is listed in parenthesis. Full details can be found in Schoeni *et al.* (2013).

³ Respondents meeting certain criteria are automatically prompted with the specific benefit showcards, e.g. those of retirement age are shown the pensions showcard; the long-term sick are shown the ‘disability benefits’ showcard.

⁴ Refusals are counted as ‘refusal’ + ‘don’t know’ where a ‘don’t know’ could be a polite refusal.

⁵ The refusal rates for ‘unearned income and state benefits’ are low. For example, refusal rates on the pension showcard for our balanced sample are: 0.007, 0.003, 0.004, 0.003 and 0.004.

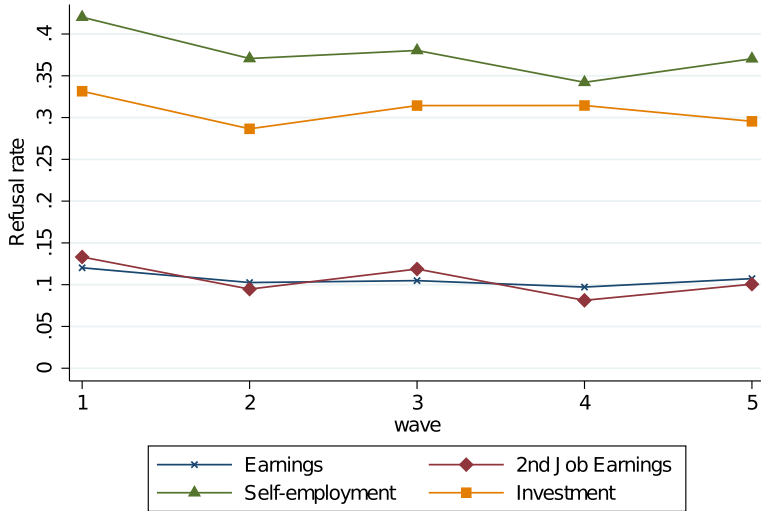


Figure 1. Income refusal rates by wave

Notes: Sample of individuals responding at each of wave 1–5

In our analysis, we replace item missing values with the standard longitudinal imputes of the data producers. There can be two other types of missing data: missing an individual interview (unit non-response) and missing an individual interview but a proxy answers a shorter questionnaire. The latter two are addressed in the methodology section.

Comparison to a cross-sectional gold standard

Before moving to our main method and results, we look for evidence of data quality changes by comparing the UKHLS income distribution to that from a cross-sectional gold standard. The gold standard we use is the ‘Households Below Average Income’ (HBAI) series, which are the data source for official UK statistics on the income distribution. The HBAI data sets are built from a mixture of survey data, from a specialist income survey (the Family Resources Survey), and administrative tax records. The HBAI data undergoes extensive editing and imputation, by the UK Department for Work and Pensions, which is based on their access to administrative records and knowledge of the tax and benefit system.⁶ As the HBAI specializes in income measurement and incorporates administrative data and UKHLS is a general purpose survey, HBAI can be considered as a ‘gold standard’.

Estimates of selected quantiles of the income distribution from the two sources are shown in Figure 2. Focusing on the lower half of the distribution, whilst there is a clear similarity between the estimates the difference between them diminishes across the early waves of the panel. At wave 1 HBAI gives higher estimates of incomes for the 1st, 5th, 10th and 25th percentiles but by wave 3 the differences have notably reduced (in fact UKHLS gives slightly higher estimates by wave 3). Thereafter, the differences are small and stable. As the UKHLS estimates gets closer to the HBAI ones across the early waves of the panel, the figure is consistent with improvements in UKHLS data quality as the panel ages. The

⁶The HBAI data-set includes income variables that have been adjusted to better measure top incomes. In order that our data sources are comparable, we use the unadjusted HBAI variables.

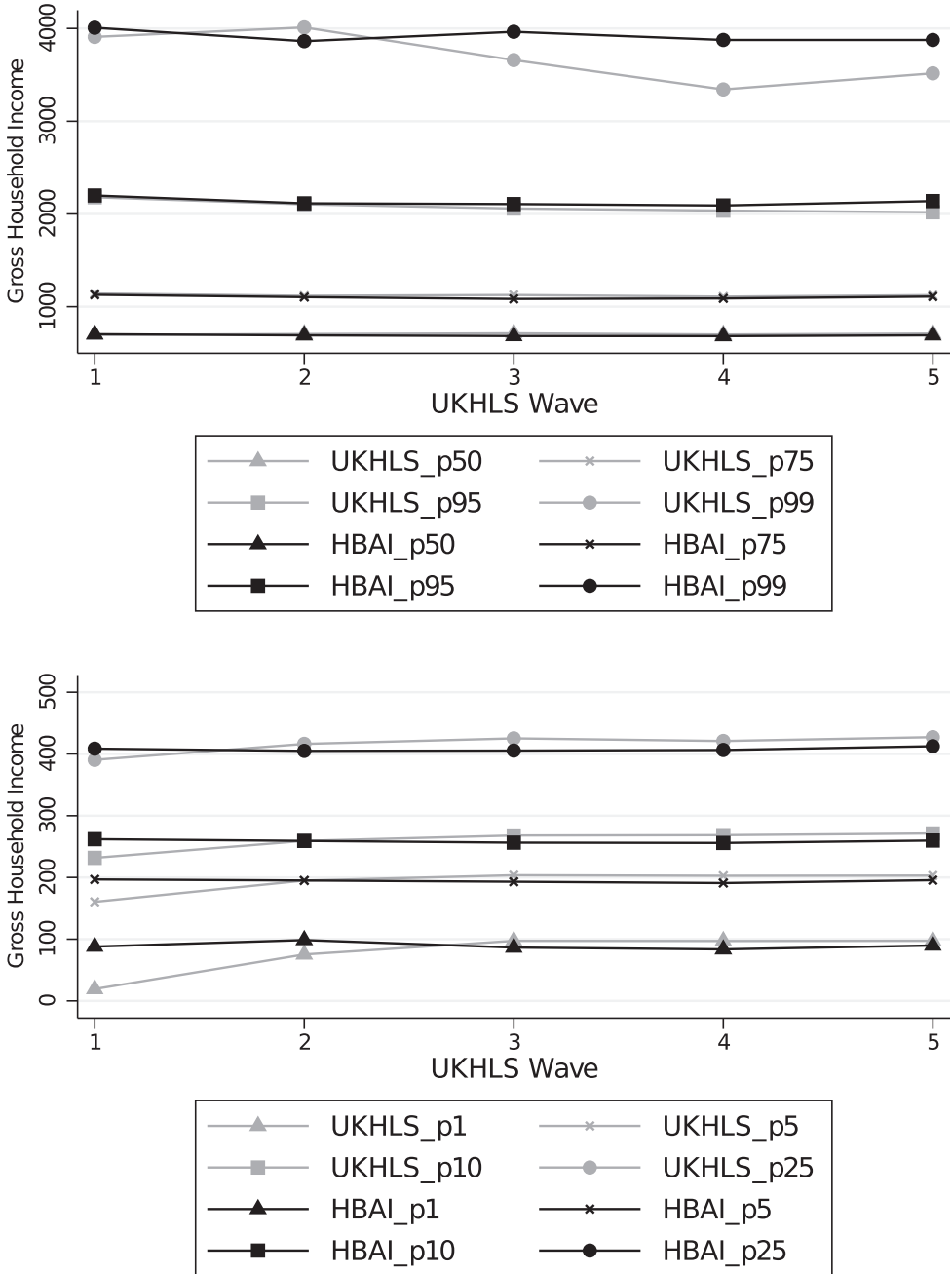


Figure 2. Selected quantiles of UKHLS/HBAI (2009/10–2013/14)

Notes: The HBAI corresponds to a financial year (April to March) and a UKHLS wave to two calendar years. To account for differences in the fieldwork period of the two sources, we pool two consecutive HBAI data sets when comparing to a single UKHLS wave. All figures are expressed in 2014–15 prices using the bespoke monthly CPI price index used in the official UK income statistics and produced by the Office for National Statistics.

top half of the panel shows the median, 75th, 95th and 99th percentiles. The estimates line-up remarkably closely.

Separately, in the main results, we also compare estimates of income sub-components to another gold-standard. We compare UKHLS estimates of benefit receipt to known administrative totals.

In the next section, we use a quasi-experiment induced by the design of UKHLS to understand if PC and DI are responsible for the data quality changes observed above.

III. Methodology

We have two randomly allocated groups $G = 1, 2$. At time t group 2 (the treatment group) is in wave $S(t)$ while group 1, which begins the survey one year later, is in wave $S(t) - 1$. Let $R(S(t)) = 1$ indicate that an individual remained in the survey (did not attrit) up to wave $S(t)$, and $R(S(t)) = 0$ otherwise. Groups 1 and 2 are random samples of the population. However, group 1 was selected one year later. If the age structure and other characteristics of the population are stable, group 1 will be on average the same as group 2, except that they will be one year younger at any t . Our strategy is to compare income reports of the two groups at time t , conditional on $R(S(t)) = 1$ (in both groups) and age. So, for example, in $t = 2010$, group 2 is in wave 2 and group 1 in wave 1. We compare the income reports of group 2 respondents of a given age to the wave 1 income reports of those group 1 respondents who are of the same age, and who also responded to wave 2 (in 2011).

In our example, this comparison is as follows:

$$E [y_{2igt} | t = 2010, g = 2, R(S = 2) = 1, A] - E [y_{1igt} | t = 2010, g = 1, R(S = 2) = 1, A] \quad (1)$$

where y_{2igt} and y_{1igt} denote reported income at a second and first interview, respectively, and A denotes age. Note that this is equal to:

$$\underbrace{E [y_{2igt} | t = 2010, g = 2, R(S = 2) = 1, A] - E [y_{1igt} | t = 2010, g = 2, R(S = 2) = 1, A]}_{\text{casual effect}} + \underbrace{E [y_{1igt} | t = 2010, g = 2, R(S = 2) = 1, A] - E [y_{1igt} | t = 2010, g = 1, R(S = 2) = 1, A]}_{\text{selection bias}} \quad (2)$$

The initial random selection of groups 1 and 2 from the population should ensure that the second term is zero at each age. Threats to the internal validity of our design are as follows. First, as in any random experiment, it could be that the randomization was not implemented correctly. Second, if the macroeconomy affects the opportunity cost of people joining the survey, there could be compositional differences between the groups as they joined the panel one year a part. Third, it is possible that the process of panel attrition was different for group 1 and group 2, so that conditioning on $R(S = 2) = 1$ is differentially selective in the two groups. This could happen, for example, because group 1 started a year later, when the survey field work agency had acquired additional experience.

We address these possibilities, in the usual way, by checking for covariate balance across the two groups. We also performed further checks on the compositional stability

of people joining the panel and on whether attrition was differentially selective across the two groups. Further details are given below.

We implement equation (1), for a given t , through linear regression. The estimating equation is:

$$Y_i = \alpha + \beta I(g=2)_i + \text{age}'_i \gamma + \epsilon_i \quad (3)$$

where Y_i is reported income of individual i (earnings, benefits and unearned income, or investment income), $I(g=2)_i$ an indicator variable taking the value 1 for a group 2 respondent, age'_i a vector of 1×1 age dummies and ϵ_i an error term with $E[\epsilon_i | I(g=2)_i, \text{age}'_i] = 0$. β is the causal effect of interest. As covariates are balanced across our groups, including controls in the model is unnecessary, but we nevertheless include them to improve precision.⁷ We also estimated models without controls (Appendix A1) and the estimates line-up closely with our main results. We report standard errors robust to heteroskedasticity.⁸

To estimate the reporting effect net of DI, we set to zero any income source for which an individual received a DI reminder and then re-estimate our main coefficient of interest. This works as the DI reminders were triggered only after a respondent did not report a source they received at the previous wave.

Checks on internal validity confirmed covariate balance at wave one.⁹ Full results are included in Table A1, Appendix A.

We also performed three checks on the stability with which respondents joined the panel. First, we confirmed that the share of people agreeing to take part in the survey did not systematically differ by calendar year (see Figure A1). Second, a balance check confirmed that respondents that joined the panel in 2009 were observationally similar to those joining in 2010 (see Table A2).¹⁰ Third, our main findings are robust to dropping months from the sample that showed differences in the joining rate across the calendar years of wave one (see Figure A13).

A further check on attrition confirmed that it is unrelated to the initial group allocation. The check involved estimating an attrition model including a wide range of controls and checking the significance of their interaction with a dummy for being in the treated group.¹¹ None of the interaction terms was statistically significant and neither were they jointly significant.¹²

⁷ Given that the control variables are potentially also subject to PC, we focus on controls with low item non-response rates and that we judge are unlikely to be sensitive areas of questioning. The full list controls is given in the footnote to Figure 3.

⁸ We also tried clustering standard errors at the level of the Primary Sampling Unit but it made little difference to the estimated standard errors.

⁹ The samples were balanced in terms of sex; age; education; marital status; and health; small but statistically significant differences were observed for: the share Indian or Chinese (0.5 percentage points) but not for other ethnic groups; living in social housing (1.6 percentage points) although not for those renting, with a mortgage or owning their home; and the number of people in the household (0.04 of a person).

¹⁰ Small but statistically significant differences were observed for the share White (0.9 percentage points), Indian or Chinese (0.5 percentage points) and living in social housing (1.2 percentage points).

¹¹ The controls are: sex, age, ethnicity, education, relationship status, economic status, health, housing tenure, household size, number of children.

¹² The F -stat = 1.10, $P = 0.33$, critical value = 1.65.

IV. Results

Differences in reporting between waves 1 and 2

Figure 3 presents estimates of β from equation (3). The figure shows results from models estimated separately for total: income; benefits and unearned income (social security benefits; pensions; and other unearned income); earnings; and investment income. For details of the main pension types and social security benefits see Table 1.

The causal effect of being interviewed for a second time, relative to a first, is to increase total monthly income by £141.71 or 8%. The effect is driven solely by changes in the benefits and unearned income sub-component of income. In particular, the strongest effects occur for social security benefits and pensions [£24.85 (10.9%) and £81.24 (28.2%)]. These numbers imply substantial differences in the quality of reported data across the first two waves of the panel. Failure to account for this survey effect would give a highly misleading picture of population income changes over the period.

Figure 4 examines the extent to which the survey effects are due to PC. It presents results from re-estimating equation (3) but setting a reported wave 2 amount to zero where a DI prompt was triggered. The estimates have fallen in size but surprisingly they remain large and statistically significant. For example, the PC effect on total income is to increase reporting by 5.5%. Overall the results show that a large share of the reporting changes seen at the second interview is due to PC and not DI.

Given the potential importance of these results for the measurement of poverty and inequality, it is important to establish whether the survey effects are constant across the

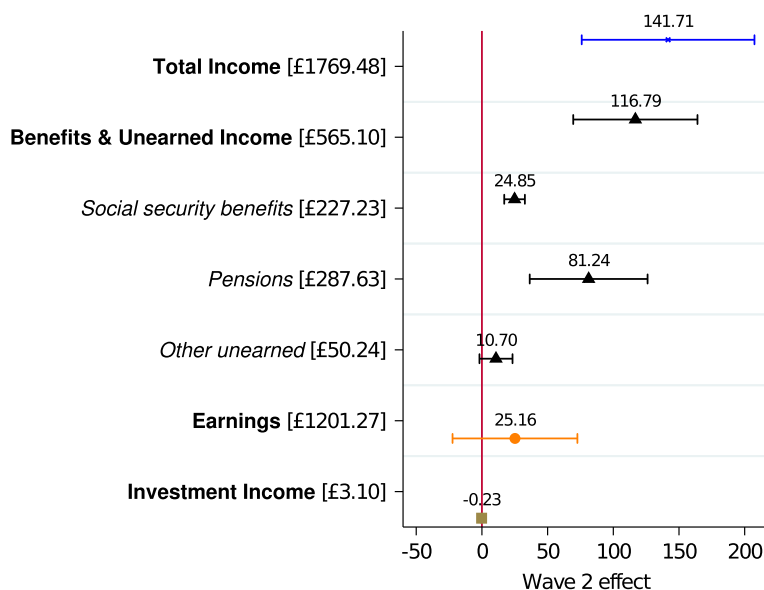


Figure 3. Effect of second interview on reported income

Notes: The point estimates presented correspond to β from equation (3) with the full-set of controls. Means from the baseline (wave 1) interview are reported in square brackets. Confidence intervals are calculated using robust standard errors. The controls are dummy variables for (number of categories in parenthesis): sex, age in one year bands, ethnicity (6), highest qualification (6), retired, student, relationship status (5), housing type (4), long-standing illness, household size (16), number of children (11), region (12) and interview month (12).

TABLE 1
Effect of wave 2 interview on reporting of selected unearned sources

	<i>Receipt</i>		<i>Amount</i>	
	DI+PC	PC	DI+PC	PC
<i>Pensions</i>				
State pension† [23.62, £528.24]	1.327*** (0.204)	0.477* (0.214)	6.289 (4.091)	6.696 (4.127)
Private pension [5.00, £455.50]	1.932*** (0.264)	0.883*** (0.255)	90.497 (61.799)	109.403 (70.711)
Employer pension [14.84, £822.87]	1.497*** (0.337)	0.381 (0.335)	279.624* (121.585)	274.081* (126.274)
Spouse employer pension [2.76, £649.47]	0.666*** (0.175)	0.316 (0.171)	-2.274 (221.224)	36.725 (238.808)
Social security benefits:				
<i>Family benefits</i>				
Working Tax Credit [6.33, £254.43]	1.154*** (0.278)	0.598* (0.274)	-0.105 (8.499)	-3.016 (8.616)
Child benefit [24.04, £145.75]	0.535 (0.312)	0.078 (0.312)	-0.563 (1.020)	-0.664 (1.014)
Child tax credit [17.68, £297.48]	1.108*** (0.337)	0.295 (0.336)	-3.796 (5.870)	-2.424 (5.921)
<i>Disability benefits</i>				
Incapacity benefit [2.71, £409.44]	0.805*** (0.193)	0.453* (0.188)	5.199 (10.554)	3.205 (10.786)
Disability living allowance [4.96, £305.65]	1.497*** (0.256)	0.853*** (0.249)	-12.521 (8.455)	-16.577 (8.539)
<i>Low income benefits</i>				
Income support [4.81, £355.62]	0.334 (0.231)	0.078 (0.228)	-1.840 (12.003)	-1.548 (12.167)
Housing benefit [10.57, £372.02]	1.058*** (0.288)	-0.119 (0.286)	-4.354 (6.238)	-3.370 (6.455)
Council tax benefit [12.10, £101.06]	1.969*** (0.332)	0.356 (0.327)	3.173 (4.542)	2.424 (4.720)

$N = 29,528$

Notes: see Figure 3 notes. Sample means for $g = 1$ are reported in square brackets.

†The state pension is a part contribution based benefit. It is paid from age 65 for men. For women, it is in the process of being increased from 60 to 65.

Significance levels indicated as $*P < 0.05$, $***P < 0.001$.

distribution of income. We extended the above analysis by estimating quantile regression models (9 quantiles) for total income and its logarithm. To summarize the results, the level effects are fairly constant across the distribution but the log effects are large for those at the bottom. For example, quantiles 0.1 and 0.7 see similar increases in reported income (around £90 per month), but this amounts to a substantial 17% increase for quantile 0.1 but only 4% for quantile 0.7. Full figures of results are provided in the supplementary materials (Figures A2 and A3). We explore in detail the effects of the survey effects on inequality measurement in a separate subsection below.

We also explored the possibility of heterogenous treatment effects by estimating models for subsamples of: pensioners, working age with children and working age without

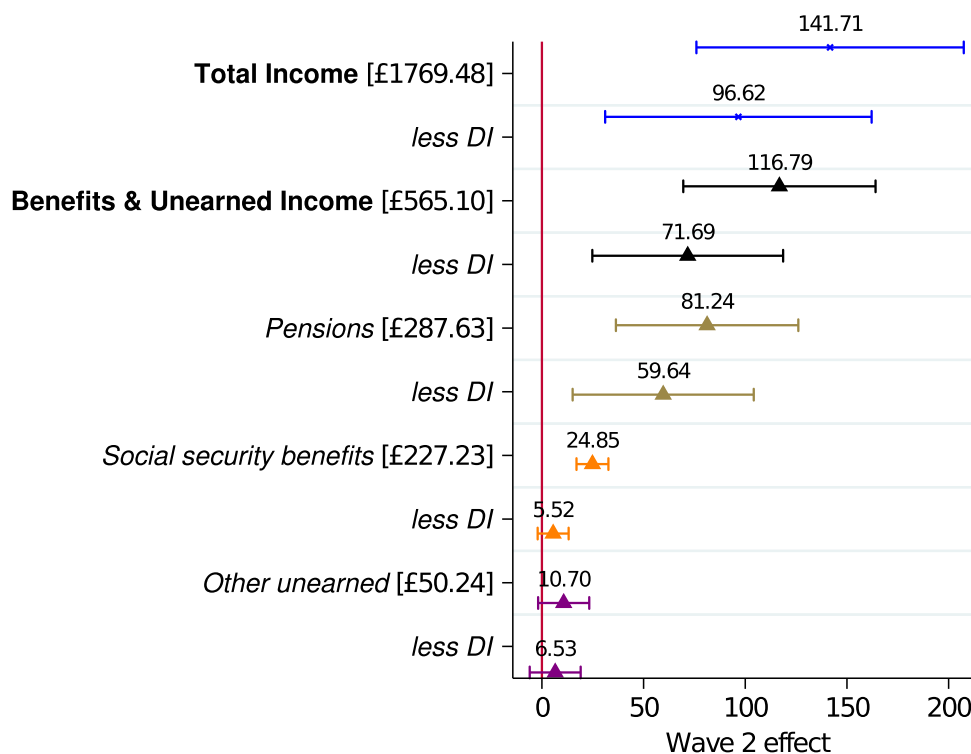


Figure 4. Panel Conditioning effect of second interview on reported income
 Notes: see Figure 3 notes.

children.¹³ In the interests of space, we only briefly review the results here. The effects are strongest for the pensioner subsample and are concentrated in the ‘benefits and unearned income’ component of income. The wave 2 effect is to increase reporting of this category by a large 24.1%. Moreover, 73.8% of this reported increase is due to PC and not DI. For the ‘working age without children’ subsample, the effects are weaker in absolute value but are proportionally large. Benefits and unearned income increase by 35.9% of the wave 1 mean and 57.5% is due to PC and not DI. Finally, for the ‘working age with children’ subsample, the effects are smaller and statistically significant only for the total effect in ‘benefits and unearned income’ (7.4% of the wave 1 mean). The interested reader can find the full set of results in Figures A4–A9, Appendix A.

We also explored the extent to which imputation can reduce the data quality differences across waves. Details are provided in section A.2 of Appendix A.

An important question is whether the effects presented for benefits and unearned income are due to changes in reported receipt or the amounts. Table 1 explores this matter by presenting estimates of equation (3) separately for 12 of the most widely received pensions and social security benefits in the data. Columns 1 and 2 refer to receipt, and 3 and 4 to the (conditional) amounts. The odd columns show the total effect (PC + DI), and the even the PC effects only.

¹³ We define ‘pensioners’ as those of UK state pension age (60 for women and 65 for men) and ‘Working age’ those below it.

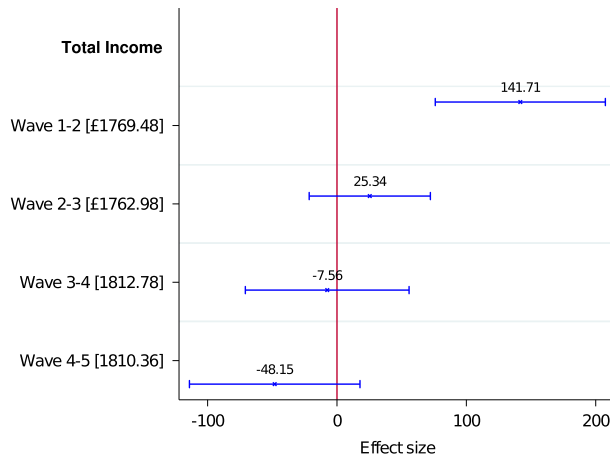


Figure 5. Reporting differences at later waves

Notes: For a list of control variables, see the notes to Figure 3. Analysis samples correspond: to 2010 (wave 1–2), 2011 (wave 2–3), 2012 (wave 3–4) and 2013 (wave 4–5) and are restricted to respondents who had a full interview at both corresponding waves.

The effect of being interviewed for a second time is to increase the reported receipt of each of the 12 sources in the table. The effects are statistically significant (excluding income support) and the magnitudes are non-trivial. For example, the effect on the state pension¹⁴ is 1.33 percentage points or 5.6% of the wave 1 mean. Column (2) shows that much of the increases are attributable to PC (statistically significant PC effects are seen for both disability benefits, two types of pension, and Working Tax Credit). For example, PC increased state pension receipt by nearly 0.4 percentage points or 2.0% of the wave 1 mean.¹⁵ We find no evidence of changes in the (conditional) amounts in columns 3. 11/12 of the estimated coefficients are statistically insignificant. Column 4 confirms this finding when estimating the PC effect only.¹⁶

To look for survey effects at later interviews, we re-estimate equation (2) for different t [$t = 2010$ (waves 1 and 2); $t = 2011$ (waves 2 and 3), $t = 2012$ (waves 3 and 4) and $t = 2013$ (waves 4 and 5)]. Figure 5 presents the results. After $t = 2010$, the estimated effect is always small and statistically insignificant. Therefore, the biggest change in behaviour (reporting or economic) occurs between the first and second waves of the panel and not at later waves. A comparison of income quantiles by wave and group confirmed that the effects are confined to the early waves of the panel (Appendix A3).

In the next two subsections, we ask if the survey effects documented here (i) might generalize to other surveys and (ii) whether they have meaningful economic consequences.

¹⁴ See Table 1 notes.

¹⁵ The conditional pension amounts are the largest in the table. This explains why even small differences in the reporting of pension receipt can have sizable effects on the income distribution.

¹⁶ Our results are in contrast to David and Bollinger (2005) who found that false negative reporting of food stamp receipt in the US Survey of Income and Program Participation was stable across waves. A lack of statistical power in their study could plausibly explain the difference.

Evidence from the British Household Panel Survey

It is important to know whether the pattern established above is a peculiarity of UKHLS or a more general feature of income data collection. We turn to the predecessor of UKHLS – the British Household Panel Survey (BHPS) – that added refreshment samples to the original (1991) sample in 1999 and 2000. We can therefore compare first-time responders in the refreshment samples to experienced panel members and see how differences in income change across waves. DI was not introduced in BHPS until 2006 and so our results relate to the effects of PC only.

Figure 6 plots selected income quantiles for a sample of respondents interviewed in each of waves 9–13 of the BHPS and living in Scotland or Wales, separately by whether they form part of the refreshment sample or were an original sample member. At wave 9, we observe that the refreshment sample gives lower values of percentiles 1, 5, 10 and 25 but that the differences notably decrease at wave 10. Thereafter, the gaps remain relatively stable.

We find similar reporting changes for the Northern Ireland refreshment sample. For percentiles 1 and 5, we again observe that, relative to the original sample, the refreshment sample provides lower estimates in wave 11 but by an amount that noticeably decreases at wave 12. Thereafter, the gap between the two estimates is relatively stable. For higher quantiles, there was no obvious reporting change in either of the refreshment samples. We also compared trends in BHPS and UKHLS item non-response rates. Both panels show the same pattern of falling refusal rates, the longer the respondents are in the panel. Full details are included in Appendix A4.

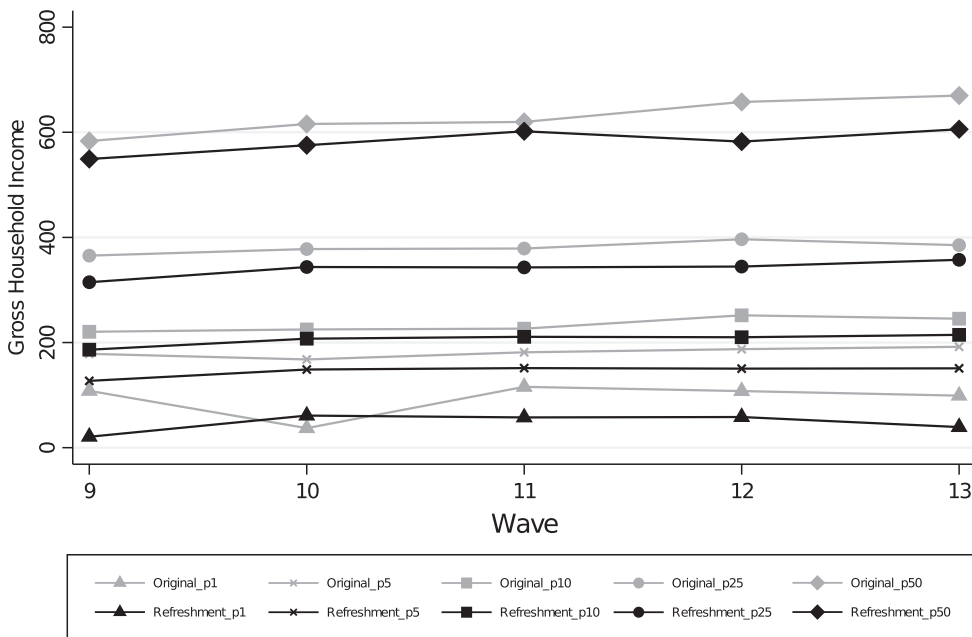


Figure 6. BHPS Scotland and Wales refreshment samples (wave 9)
 Notes: Sample of respondents interviewed in each of waves 9–13. Original sample members are restricted to living in Scotland or Wales. Scotland and Wales refreshment samples entered at wave 9.

Implications for inequality and mobility estimates

This section explores the consequences of the survey effects for one important economic analysis: inequality and poverty measurement. Analysis of inequality and poverty is typically based on net income rather than gross income. The following is therefore based on the UKHLS-derived household net income variable. The variable has been produced by the UKHLS data providers, has been considered reliable and aims to replicate the official HBAI measure.

Figure 7 compares the distribution in the gold standard and each of the treated and control groups by means of a pair of quantile–quantile plots. The control group (first-time reporters) has lower income relative to the gold standard. The missing income is concentrated at the bottom of the distribution and steadily decreases across the distribution. Conversely the treated group (second-time reporters) has a distribution that matches very closely indeed to the gold standard.

The first three columns of Table 2 show 2010 estimates of seven inequality measures for our treated and control groups. The measures are commonly used in inequality analysis and are as follows: the Gini coefficient, Standard deviation of logs, Atkinson index with aversion parameter 2, percentile ratios (90–10, 90–50, 10–50) and the share below the poverty line. Each captures inequality at different parts of the distribution. For example, the Gini is sensitive to changes in the centre of the distribution, whereas, the Aitkinson measure is sensitive to changes in the tails. Columns 1 and 2 show the estimates for the control and treated group, respectively, while column 3 shows the difference.

We observe statistically significant differences for five of the inequality measures (standard deviation of logs, percentile ratios and the share below the poverty line). Overall, the effect of treatment is to compress the income distribution, relative to the control. We see that treatment reduces the standard deviation of the logs, two of the percentile ratios and the share below the poverty line. However, the 90–50 ratio increases showing the median rising relative to the mean. Overall, the results show that the survey effects documented in the previous sections lead to meaningful differences in cross-sectional estimates of inequality.

Are the differences in inequality consistent with higher data quality for the treated group? To explore this, we compare our treatment and control group estimates of inequality to those from a gold standard (HBAI). We should expect the estimates of higher quality

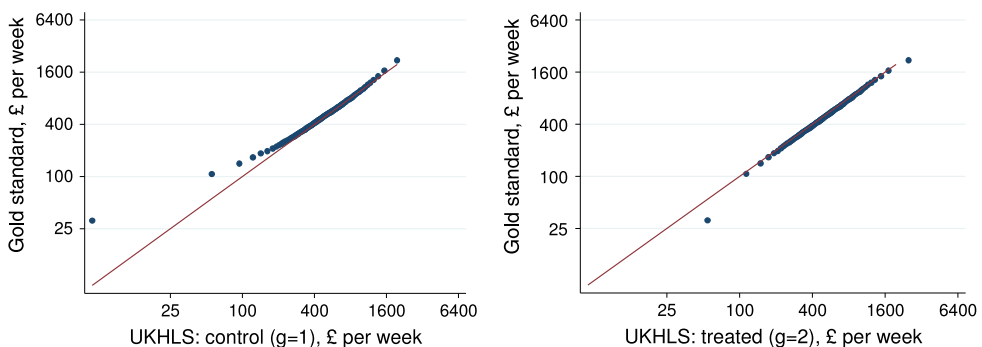


Figure 7. Quantile–quantile plot of the Gold Standard vs. Treated and Control

TABLE 2
The effect of reporting changes on inequality estimates

	Cross-sectional (t = 2010)			Change (2011 minus 2010)		
	Control (g = 1)	Treated (g = 2)	Difference	Control (g = 1)	Treated (g = 2)	Difference
Gini	0.346	0.364	0.018 (0.00941)	-0.0154	-0.0301	-0.0147 (0.0155)
Standard deviation of logs	0.693	0.637	-0.055* (0.0216)	-0.0660	-0.0189	0.0472* (0.0196)
Atkinson	0.833	0.719	-0.114 (0.228)	-0.226	-0.0452	0.181* (0.0821)
<i>Ratios:</i>						
90-10	4.488	3.996	-0.493*** (0.0794)	-0.476	-0.0414	0.435*** (0.121)
90-50	2.046	1.976	-0.071** (0.0251)	-0.0151	-0.0207	-0.0056 (0.0437)
10-50	0.456	0.494	0.039*** (0.00688)	0.0522	0.0000	-0.0522*** (0.0107)
Share below 60% median hh income	0.192	0.173	-0.019*** (0.00446)	-0.0287	0.0004	0.0291*** (0.00722)
Observations	14,731	14,911		11,823	12,350	

Notes: Household income is net of taxes, deflated and equivalized using the modified OECD equivalence scale. The UKHLS household net income measure aims to replicate the HBAI definition and the two differ only in minor deductions.

‡Households Below Average Income. See the data section and Figure 2 notes.

†The poverty line follows the standard UK definition and is 60% of median income. Bootstrapped standard errors shown in parentheses (1,000 replications). Significance levels indicated as * $P < 0.1$, ** $P < 0.05$, *** $P < 0.01$.

TABLE 3
The effect of reporting changes on mobility estimates

	Control ($g = 1$)	Treated ($g = 2$)	Difference
<i>Transitions (between 2010 and 2011)</i>			
Exit poverty (share)	0.483	0.427	−0.056*** (0.015)
Enter poverty (share)	0.076	0.083	0.007 (0.004)
Mean income change (£ per week)	5.441	−29.680	−35.121** (12.112)
Mean income change (log)	0.034	−0.004	−0.038*** (0.007)
Changed decile (share)	0.654	0.642	−0.012 (0.006)
Observations	11,823	12,350	

Note: See Table 2 notes.

to more closely match the HBAI ones. The table confirms that when there are significant differences between the estimates from our treatment and control groups, it is the treated group who is closer to the gold standard estimates (column 4). The observed reporting increases of the previous sections are therefore consistent with reporting improvements.¹⁷

We also compared estimates of distributional change between 2010 and 2011 (columns 5–8). The treated group is being measured for the second and third time, while the control for the first and second time. We find that the survey effects have notable consequences for the measures of change. We find statistically significant differences in estimates of change for 5 out of 7 of the inequality measures. As we expect, it is the control group that overestimates distributional change. For example, the change in the 90–10 ratio is estimated to be over 11 times larger for the control group compared to the treated group. Moreover, when a statistically significant difference occurs, it is the treated group who is now closer to the gold-standard estimate (column 8).

Finally, we ask whether the observed survey effects lead to bias in estimates of income mobility. This is important as while inequality can be monitored using repeated cross-sections, estimates of mobility require repeated measurement. Table 3 explores this by comparing estimates of income mobility between 2010 and 2011 for the treated and control groups. We find that the treated group had lower exit rates from poverty and lower changes in income (both in levels and in logs) but not in poverty entry rates or in the share that changed income decile. The magnitudes are non-trivial. For example, the treated group shows an exit rate from poverty of 5.6 percentage points (11.59%) below the control group. Put together, this section has shown that our findings from the previous sections have substantive consequences for an important type of economic analysis.

To confirm the robustness of our results, we performed a placebo check for the $t = 2012$ year. As reporting changes were confined to the start of the panel, we would not expect to

¹⁷ The Atkinson measure can be sensitive to extreme values in the lower tail of the distribution (see Cowell and Flachaire, 2007). However, when we drop the top and bottom 1% of the sample, the Atkinson estimates line up very closely to the gold standard (see Table A5).

find significant inequality and mobility differences for this placebo check. Results support this and are provided in Appendix Tables A7 and A8.

V. Explaining the panel conditioning effect

This section discusses the potential mechanisms through which PC operates. It first considers whether the observed effects reflect a reporting change or an economic change. It reviews the existing literature and summarizes the empirical evidence for and against each. Overall, the evidence is inconclusive. However, it does indicate that a substantial share of the main findings (for pensions) arises through a reporting effect. The latter subsections then present evidence as to why this might have occurred.

Changes in reporting or in take-up behaviour?

Zwane *et al.* (2011) and Das and Leino (2011) find survey participation increases the uptake of a health insurance in a developing country setting. Related, Bach and Eckman (2018) find that repeated interviewing increases participation in Active Labour Market Programs in Germany (job application training, continuing education courses). In contrast to the present paper, they find that each additional exposure to the treatment, i.e. each additional wave, intensifies this effect. Could a similar change in take-up behaviour explain our PC effects for ‘Benefits and Unearned Income’?

We begin by comparing survey estimates of benefit receipt to known administrative totals. If our main results correspond to reporting improvements, then we expect survey estimates for wave two to be closer to the known totals relative to the wave one estimates. On the other hand, with a behavioural response we might expect survey respondents to have a higher take-up rate than the population at large. Table A4 of Appendix A presents the comparison. The wave two estimates of benefit receipt are generally closer to known administrative totals, relative to the wave one estimates. Moreover, the survey never overestimates benefit receipt relative to the administrative sources. This provides at least suggestive, but far from conclusive, evidence in favour of a reporting effect.

Distinctly, one indicator of reporting quality is the extent to which survey participants refuse to answer questions about income. Figure 1 showed clear evidence of reporting improvements, as the longer participants were in the panel, the lower were refusal rates. Indeed, as with the main results of the paper, the most dramatic changes in refusal rates occurred between waves one and two. Also, suggestive of data quality improvements was the analysis in section II that showed that the survey estimates of inequality were closer to a gold standard at wave 2 relative to wave 1.

Eighty three per cent of our PC effect occurs in pensions (Figure 4) and is concentrated on the extensive margin (column 2, Table 1). In particular, our pensions category consists of: state pensions; private pensions; and employer pensions (Table 1). If take-up of pensions is already high; then it would rule out a behavioural response in take-up explaining our pension findings. An official parliamentary report indicates that take-up of the state pension ‘is close to 100%’ (HC, 2005).¹⁸ Employer and private pensions could potentially go

¹⁸ All eligible citizens automatically receive a letter from the relevant Government department four months before retirement.

unclaimed, but there are no official estimates of this number. In principle, it should be rare as the schemes are regulated by public bodies (The Pensions Regulator and The Financial Conduct Authority).¹⁹ For example, trustees of workplace pensions are legally required to take regular steps to trace missing members.²⁰ Put together, it seems unlikely that increases in take-up could be of sufficient magnitude to explain our main pension results (ie. our effect (in percentage terms) in Table 1 would require that take-up of state pensions be no more than 74%, private pensions no more than 69% and employer pensions no more than 80%).

The remaining 17% of our PC effect occurred in ‘social security benefits’ and in ‘Other Unearned’ income (Figure 4). In particular, Table 1 showed significant PC effects on the extensive (and not intensive) margin of one means tested benefit (Working Tax Credit) and two disability benefits (Incapacity Benefit and Disability Living Allowance). On the one hand, these differ from the health insurance considered in Das and Leino (2011) where a health insurance was newly introduced and knowledge amongst the population was imperfect. A take-up effect would also be inconsistent with the mechanism proposed in Zwane *et al.* (2011) where the survey acts to draw attention away from ‘immediate needs’ and towards ‘future contingencies’. On the other hand, unlike pensions, the non-take up of means tested benefits is a well documented problem (DWP, 2017). For Working Tax Credit in 2010–11, the take-up rate was 64% (HMRC, 2012). The take-up rate for disability benefits is harder to estimate because of the difficulty in defining the eligible population, but they share similar features with means tested benefits, where the most recent rates range from 56% to 84% (DWP, 2017). Moreover, an increase in the uptake of a Government programme following survey participation would be consistent with Bach and Eckman (2018).²¹

Put together, it looks likely that 83% of the PC effects (pensions) cannot be driven by a behavioural response in take-up, rather they are consistent with a change in the reporting behaviour of survey participants. However, we cannot make the same claim for the remaining 17% where increases in take-up would also be consistent with our findings.

Implementer learning

Here, we consider a possible explanation for the improved pension reporting documented above: implementer learning. Implementers (fieldwork agency, interviewers) accrue experience over the first wave of the panel, raising the possibility that they are better able to elicit responses at the wave 2 interviews. For implementer learning to be a convincing explanation, two conditions need to be met, and we consider both to be implausible. First, implementers must benefit from their experience at wave 2 2010 but not at wave 1 2010, even though the two were being collected at the same point in time.²² Second, im-

¹⁹ See ‘legal background’ here: <http://www.thepensionsregulator.gov.uk/guidance/guidance-record-keeping.aspx>.

²⁰ For a recent freedom of information request on the subject see: <http://www.thepensionsregulator.gov.uk/foi/percentage-of-people-who-do-not-take-up-their-pensions-march-2017.aspx>.

²¹ We found no evidence that the headline effect was stronger for low-income households (Figures A2–A3), which might be the case if there was an ‘information friction’ for low income respondents.

²² We show that there are few differences in the 2010 wave 1 and 2 interviewer experience distributions (Table A9, Appendix A).

TABLE 4

Comparison of UK Household Longitudinal Study (UKHLS) wave 1 to Households Below Average Income (HBAI) by calendar year

Percentile	UKHLS			UKHLS			Ratio of ratios (col 7/col 4)
	HBAI 2009	Wave 1 2009	Ratio (col 2/col 3)	HBAI 2010	Wave 1 2010	Ratio (col 5/col 6)	
5	199.14	161.00	1.24	194.54	152.16	1.28	1.03
10	264.05	237.21	1.11	259.56	230.77	1.12	1.01
15	317.08	297.02	1.07	309.95	286.26	1.08	1.01
20	365.61	355.30	1.03	358.02	339.49	1.05	1.02
25	412.32	411.89	1.00	404.49	395.35	1.02	1.02
30	463.11	466.50	0.99	457.18	455.04	1.00	1.01
35	518.96	530.16	0.98	513.14	509.83	1.01	1.03
40	578.80	593.04	0.98	568.75	572.68	0.99	1.02
45	644.98	654.25	0.99	627.74	634.09	0.99	1.00
50	710.95	721.08	0.99	694.89	699.68	0.99	1.01
55	779.18	794.58	0.98	761.77	779.61	0.98	1.00
60	854.60	876.58	0.97	838.50	846.41	0.99	1.02
65	942.85	961.07	0.98	920.23	933.10	0.99	1.01
70	1,034.04	1,057.79	0.98	1,008.65	1,023.25	0.99	1.01
75	1,142.53	1,172.90	0.97	1,112.50	1,144.67	0.97	1.00
80	1,273.40	1,313.70	0.97	1,236.48	1,274.61	0.97	1.00
85	1,451.13	1,487.80	0.98	1,409.49	1,453.66	0.97	0.99
90	1,702.50	1,745.69	0.98	1,666.63	1,726.21	0.97	0.99
95	2,259.56	2,193.66	1.03	2,138.24	2,170.68	0.99	0.96

Notes: Analysis is based on the 'HBAI' data sets and for household gross income before deductions. HBAI corresponds to a financial year (April to April) and to UKHLS a full calendar year.

plementer learning would have to occur beyond the first full year of data collection (when the biggest learning might be expected) as the fieldwork agency already had a full year of field experience (wave 1 2009) before the period of our analysis sample.

We compare estimates of the income distribution from UKHLS wave 1 (2009 respondents) with the HBAI 2009; and UKHLS wave 1 (2010 respondents) with the HBAI 2010. By 2010, UKHLS implementers had a full-year of experience, but the 2009 and 2010 (wave 1) respondents were new to the survey. If the problem is with implementers, rather than respondents, then the 2010 comparison should be more favourable. It should be noted that < 100% of wave 1 2010 interviewers participated at wave 1 2009. However, on average, they were still substantially more experienced in terms of having participated in the survey before (76.5% of wave 1 2010 interviewers participated at wave 1 2009).

Table 4 shows the results from this comparison. Columns 2 and 3 show estimated 2009 percentiles from the HBAI and UKHLS, respectively, and column 4 shows their ratio, where a ratio > 1 indicates that UKHLS underestimates a percentile relative to HBAI. Columns 5–7 repeat the analysis but for the 2010 (wave 1) calendar year.

The 2009 UKHLS percentiles match closely with the HBAI ones (column 2). UKHLS misses income at the bottom of the distribution and most notably for percentiles 5, 10 and 15 where the ratios are 1.24, 1.11 and 1.07, respectively. A remarkably similar pattern is seen for the 2010 (wave 1) calendar year. Column 8 presents the ratio of ratios and it is

always close to one indicating little change in the relative difference between the surveys over the period. It reaches an absolute maximum of 1.03 for the fifth percentile, which if anything, suggests that the coverage of UKHLS got worse relative to HBAI in the second year of wave one compared to the first year of wave one. We conclude that implementer learning is unlikely to be responsible for the observed improvements in data quality.²³

Respondent-interviewer rapport

Respondents and interviewers may establish a rapport during their first interview, which may result in more accurate reporting at the second. We test for this possibility by estimating the effect of having different interviewers at wave one and two, on a respondents wave 2 income report.

We augmented our main regression specification with a dummy variable for having different interviewers at wave one and two, and its interaction with the wave 2 dummy (difference-in-difference model). The ‘different interviewer’ dummy captures time invariant differences between respondents that changed interviewer or not and the coefficient on the interaction term gives the estimated causal effect of having a different interviewer on wave 2 reported income. Both the ‘different interviewer’ dummy and its interaction were highly insignificant (full results are included in Table A10). This indicates that the rapport interviewers and respondents may have built during the first interviewer played no role in increasing the reporting of income at the second interview. We conclude that the respondent-interviewer rapport cannot explain the observed reporting changes across waves.

A distinct but related possibility is that interviewers were non-randomly allocated across the treated and control group and this generated differences in income reporting. We do not have information on how the field-work agency allocated interviewers across the months of wave 1 but a balance check on some basic interviewer characteristics (sex, year of birth, year started as interviewer, ethnicity) revealed only small and statistically insignificant differences between the treated and control group (Table A12).

Respondent learning

Respondents may have an improved comprehension of the complex interview or have updated their beliefs about the trustworthiness of the data holders following a first interview. On the first, we exploit interviewer reports of how well the respondent understood the questions during the interview (on a 5 point scale). On the second, we make use of interviewer reports on whether a respondent was ‘suspicious’ about the study after the interview (3 point scale) and whether prior to the interview, the household respondent had questions about ‘confidentiality’ (binary variable). We estimate equation (3) for the three interviewer outcomes. We recoded them into binary indicators for ease of interpretation and as some of the categories contain few observations. Full details of the questions, distributions and the construction of the interviewer observation variables are provided in Appendix B6.

²³These results are also inconsistent with different refusal rates in different years of wave one as a source of the main findings of the paper.

TABLE 5

Effect of wave 2 interview on respondent behaviours

	(1)	(2)	(3)
	<i>Misunderstood questions</i>	<i>Suspicious</i>	<i>Queries confidentiality</i>
Wave 2	−0.01 (0.005)	−0.09*** (0.003)	−0.16*** (0.003)
<i>N</i>	29,502	29,502	29,365

Notes: Standard errors in parentheses.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Estimates of equation (1) with full set of controls (see Figure 3 notes).

Means of the dependent variables are: 0.31, 0.12, 0.18, respectively.

Table 5 shows the results. We find no evidence that being interviewed for a second time improved respondent understanding of the interview with the effects being small and statistically insignificant. In contrast, we observe that interviewers rated respondents as being less suspicious after the second interview and were also less likely to have confidentiality queries.²⁴ We also prepared analogous results where we recoded the interviewer reports into a series of binary variables to assess whether the effect is monotonic. We never found evidence of improvements in respondent understanding of the interview (Table A13). We did find that the treatment reduced the probability of the respondent being rated in the most ‘suspicious’ category. The effect size for the later was smaller compared to Table 5 (−0.01 vs. −0.09 above). Full results and included in Table A14.

An interpretation of these findings is that the first interview reveals information to respondents about the trustworthiness of the data holders. At the start of the panel, respondents have doubts about the survey organization – a stranger to them – who may share their sensitive data with third party organizations. But respondents learn from their first interview that the data holders are reliable and that their data does not get shared. By the second interview, respondents have updated their beliefs about the trustworthiness of the survey organisation, and are so more open in revealing details of their personal finances.²⁵

One of our findings was of PC effects for pensions, including the state pension. This raises the question of why state pension reporting is affected by PC – when on the face of it, it is not a confidential area of questioning. One explanation is that details of state pension receipt are collected alongside more sensitive pension types (e.g. private pensions) as part of a single pensions question. It may be that respondents unwilling to disclose their sensitive pension types simply choose not to engage with the pensions question at all.

VI. Conclusions

We find that the quality of income data collected as part of a large-scale household panel survey changes over the life-time of the panel. The largest changes in reported income are

²⁴ Table A11 shows that confidentiality concerns are predictors of item non-response in the wave 1 cross-section.

²⁵ The respondent learning interpretation can be reconciled with the use of the HBAI (based on survey data) as a gold standard. For example, the extensive editing that the HBAI undergoes could act as a substitute for respondent learning.

concentrated across the first waves of the panel and in unearned income sources, particularly pensions and disability benefits. The effect sizes are large and have until this point gone unnoticed, possibly as it is difficult to distinguish survey effects from real changes in living standards, without linked administrative records. The novelty of our approach is that it does not require data linkage, but makes use of unique features of the survey design of the Understanding Society survey as a quasi-experiment.

The use of income data from repeated survey measures is commonplace in economics, including the use of large scale household surveys and purpose built surveys implemented as part of field experiments. Our results indicate that researchers analysing data from the early waves of a panel or with short panels (such as in randomized control trials) should proceed with caution. One possibility is that researchers may want to consider adjusting data from the first waves of data collection.

We present evidence that at least some, but not all, of the effect we identify, is driven by a reporting improvement. We show that falls in item-non-response and higher income reports coincide with improving respondent confidence in the confidentiality of their sensitive data. This also makes our findings relevant to studies based on cross-sectional data – which essentially forms wave one of a panel – as they indicative the types of income sources that may be under-reported.

Final Manuscript Received: December 2018

References

- Bach, R. L. and Eckman, S. (2018). 'Participating in a panel survey changes respondents labour market behaviour', *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 182, 263–281.
- Banerjee, A., Duflo, E., Glennerster, R. and Kinnan, C. (2015). 'The miracle of microfinance? Evidence from a randomized evaluation', *American Economic Journal: Applied Economics*, 7, 22–53.
- Blundell, R., Dias, M. C., Meghir, C. and Shaw, J. (2016). 'Female labor supply, human capital, and welfare reform', *Econometrica*, 84, 1705–1753.
- Brewer, M., Etheridge, B. and O'Dea, C. (2017). 'Why are households that report the lowest incomes so well-off?', *Economic Journal*, 127, F24–F49.
- Chetty, R., Grusky, D., Hell, M., Hendren, N., Manduca, R. and Narang, J. (2017). 'The fading American dream: trends in absolute income mobility since 1940', *Science*, 356, 398–406.
- Cowell, F. A. and Flachaire, E. (2007). 'Income distribution and inequality measurement: the problem of extreme values', *Journal of Econometrics*, 141, 1044–1072.
- Crossley, T., de Bresser, J., Delaney, L. and Winter, J. (2017). 'Can survey participation alter household saving behavior?', *Economic Journal*, 127, 2332–2357.
- Das, J. and Leino, J. (2011). 'Evaluating the RSBY: lessons from an experimental information campaign', *Economic and Political Weekly*, 11, 85–93.
- Das, M., Toepoel, V. and van Soest, A. (2011). 'Nonparametric tests of panel conditioning and attrition bias in panel surveys', *Sociological Methods & Research*, 40, 32–56.
- David, M. H. and Bollinger, C. R. (2005). 'I didn't tell, and I won't tell: dynamic response error in the SIPP', *Journal of Applied Econometrics*, 20, 563–569.
- DWP (2017). *Income-Related Benefits: Estimates of Take-up*, Technical report, Department for Work and Pensions.
- Frick, J. R., Goebel, J., Schechtman, E., Wagner, G. G. and Yitzhaki, S. (2006). 'Using Analysis of Gini (ANoGi) for detecting whether two sub-samples represent the same universe: the SOEP experience', *Sociological Methods & Research*, 34, 427–468.

- Halpern-Manners, A., Warren, J. R. and Torche, F. (2014). 'Panel conditioning in the general social survey', *Sociological Methods & Research*, 46, 103–124.
- Haushofer, J. and Shapiro, J. (2016). 'The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya', *The Quarterly Journal of Economics*, 131, 1973.
- HC (2005). *Pension Credit: Third Report of Session 2004–05*, Technical report, House of Commons Work and Pensions Committee.
- HMRC (2012). *Child Benefit, Child Tax Credit and Working Tax Credit Take-up Rates 2010–11*, Technical report, HM Revenue and Customs.
- Hoynes, H., Schanzenbach, D. W. and Almond, D. (2016). 'Long-run impacts of childhood access to the safety net', *American Economic Review*, 106, 903–934.
- Hurst, E., Li, G. and Pugsley, B. (2014). 'Are household surveys like tax forms? Evidence from income underreporting of the self-employed', *Review of Economics and Statistics*, 96, 19–33.
- Karlan, D. and Zinman, J. (2011). 'Microcredit in theory and practice: using randomized credit scoring for impact evaluation', *Science*, 332, 1278–1284.
- Lynn, P., Jäckle, A., Jenkins, S. P. and Sala, E. (2012). 'The impact of questioning method on measurement error in panel survey measures of benefit receipt: evidence from a validation study', *Journal of the Royal Statistical Society Series A*, 175, 289–308.
- Meyer, B. D. and Mittag, N. (2015). *Using Linked Survey and Administrative Data to Better Measure Income: Implications for Poverty, Program Effectiveness and Holes in the Safety Net*, Technical report, National Bureau of Economic Research.
- Meyer, B. D. and Sullivan, J. X. (2003). *Measuring the Well-Being of the Poor Using Income and Consumption*, NBER Working Papers No. 9760, National Bureau of Economic Research, Inc.
- Meyer, B. D. and Sullivan, J. X. (2011). 'Viewpoint: further results on measuring the well-being of the poor using income and consumption', *Canadian Journal of Economics*, 44, 52–87.
- Schoeni, R. F., Stafford, F., McGonagle, K. A. and Andreski, P. (2013). 'Response rates in national panel surveys', *The Annals of the American Academy of Political and Social Science*, 645, 60–87.
- Van Landeghem, B. (2014). 'A test based on panel refreshments for panel conditioning in stated utility measures', *Economics Letters*, 124, 236–238.
- Yang, T.-T. (2017). 'Family labor supply and the timing of cash transfers: evidence from the earned income tax credit', *Journal of Human Resources*, 53, 445–473.
- Zwane, A. P., Zinman, J., Van Dusen, E., Pariente, W., Null, C., Miguel, E., Kremer, M., Karlan, D. S., Hornbeck, R., Giné, X., Duflo, E., Devoto, F., Crepon, B. and Banerjee, A. (2011). 'Being surveyed can change later behavior and related parameter estimates', *Proceedings of the National Academy of Sciences* 108, 1821–1826.

Supporting Information

Additional supporting information may be found in the online version of this article:

Appendix A. Additional tables and figures.

Appendix B. Data appendix.