ISER Working Paper Series

Does repeated measurement improve income data quality?

Paul Fisher
Institute for Social and Economic Research
University of Essex

No. 2016-11
October 2016

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

University of Essex

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

www.iser.essex.ac.uk

Non-Technical Summary

Social scientists studying material living standards commonly analyse data from surveys that measure the income of the same individual at multiple points in time (panel data) e.g. large-scale household panel surveys or panel data collected as part of field experiments. It is well known that respondents to a survey tend to under-report their income, but researchers commonly assume that the amount of under-reporting is stable across waves of a panel. This assumption is critical, as if reporting behaviour changes, then estimates of how material living standards evolve over time would be confounded with the changes in reporting behaviour and give a misleading picture of changes in real living standards. In this paper we provide evidence on how the underreporting of income in a large scale household panel survey (the Understanding Society Survey) changes across waves.

Our main finding is that repeated interviewing results in better quality income data. The improvements in quality are large and are strongest for unearned income sources (largely pensions and state benefits) and take effect in the initial waves of the panel. Our estimates indicate that the effect of being interviewed for a second time is to increase the mean of reported monthly income by a substantial £142 (8 percent). We find evidence of similar effects in another leading panel survey (British Household Panel Survey) confirming that the findings generalise to at least one other survey.

Using Understanding Society, we show that the reporting improvements have important implications for analysis of the income distribution. They lead to lower estimates of inequality and poverty but also income mobility.

As to why the reporting improvements occur, dependent interviewing (DI) – a common recall device used in household panel surveys that reminds respondents of their reports at an earlier interview – takes effect only after a first survey interview. We find it can explain approximately one third of the observed increase in reported income. On the remaining two thirds, given the use of the same interviewers, infrastructure and questionnaires at both waves, this points to changes in the reporting behaviour of survey respondents. Indeed, we present evidence suggestive of a reduction in respondent confidentiality concerns following the first interview. This is backed up with an examination of refusal rates on the income variables which fall off most sharply between waves one and two.

If our evidence is interpreted as reducing under-reporting of income, it suggests that income data provided as part of panel surveys offers some quality advantages over that collected from cross-sectional surveys deriving both from the use of DI and also through being able to improve respondent cooperation through repeated measurement. Separately, the reporting improvements found for income might be expected to extend to other sensitive areas of data collection, although this is left for future research.

Does repeated measurement improve income data quality?*

Paul Fisher†

Institute for Social and Economic Research, University of Essex

November 2017

Abstract

This paper exploits a natural experiment created by a survey design to show that survey data on household incomes systematically improves across waves of a panel. Our estimates indicate that the effect of being interviewed for a second time is to increase the mean of reported monthly income by £142 (8 percent). Dependent interviewing - a recall device commonly used in panel surveys - takes effect only after a first interview. It explains approximately one third of the observed increase. The remaining share is attributed to changes in respondent reporting behaviour (panel conditioning). We show that the reporting improvements coincide with improving respondent confidence in the confidentiality of their sensitive data.

JEL Classification: C83, D31, I32

Keywords: personal income distribution, poverty measurement, survey error

*Thanks to Tom Crossley for reading drafts of this article and providing insightful feedback. I would also like to 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. Data from the Understanding Society Survey has been made available through the UK Data Archive and has been used with permission.

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

1 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 randomised 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).

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); Meyer and Mittag (2015); Lynn et al. (2012); Brewer, Etheridge, and O'Dea (2017); Hurst, Li, and Pugsley (2014)), 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 improves across the early waves of a leading panel survey. This suggests a major benefit of repeated interviewing. On the other hand, 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 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 (see Crossley et al. (2017)). Second, dependent interviewing (DI) - a tool that reminds 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 separate changes in reporting behaviour from true income changes (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 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.

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

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 (eg. Lynn et al. (2012); David and Bollinger (2005)). 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 improves the quality of income data (in particular unearned income 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 percent and about one third of the difference can be explained by DI, with the other two thirds due to PC. As to why these effects occur, we are able to rule out interviewer and fieldwork agency learning as explanations. Given the stability of the survey infrastructure and questionnaire content, this points to changes in the reporting behaviour of survey respondents. 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.

On the wider implications of our results, we show that the reporting changes 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 3 discusses identification and the results are shown in section 4. Section 5 explores the mechanisms behind the results and section 6 concludes. A supplementary appendix includes additional materials.

2 Data

2.1 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.[2] 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 percent of wave one respondents completed an interview at wave two with a further 1.9 percent completing a proxy interview.

2.2 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

[2] It is the issue date which is fixed at 12 monthly intervals and not the actual date of interview. The two may differ by some months in order to maximise the chances of an individual response.

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.

An 'unearned income and state benefits' module uses 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 of 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 per cent in wave 1 and reaches a minimum over the five waves of 34.2 per cent 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 per cent to 37.1 per cent and from 12.0 per cent to 10.3 per cent for earnings.[5]

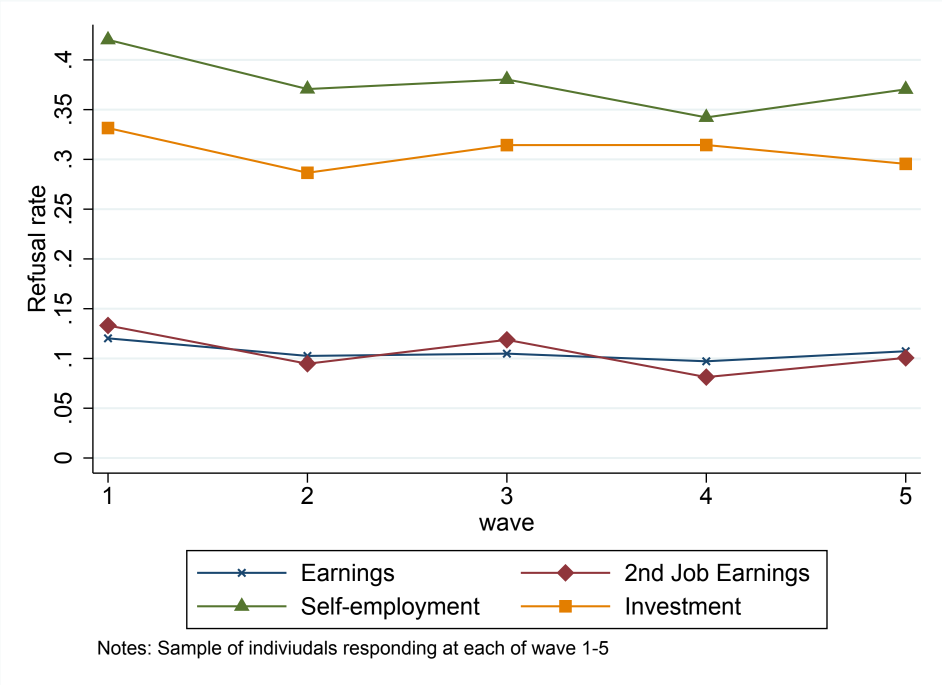
In our analysis, we replace item missing values with the standard longitudinal imputes of the data producers. There can be two other types of missing data: missing an individual interview (unit non-response) and missing an individual interview but a proxy answers a shorter questionnaire. The later two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

[3]Respondents meeting certain criteria are automatically prompted with the specific benefit showcards eg. 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 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



2.3 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 specialises 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

[6]The 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.

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 top half of the panel shows the median, 75th, 95th and 99th percentiles. The estimates line-up remarkably closely.[7]

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 improvements observed above.

3 Methodology

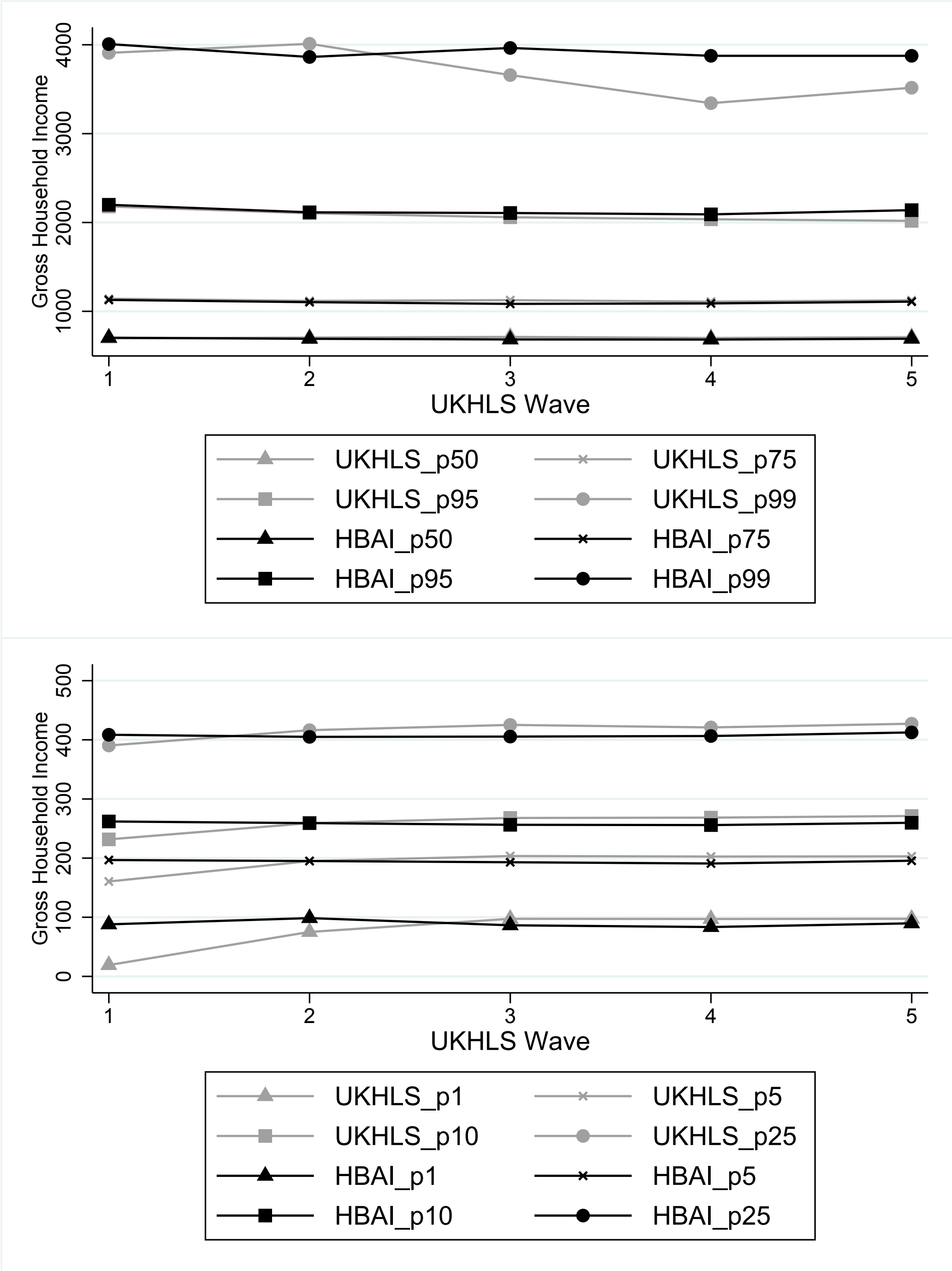
We have two randomly allocated groups $G = 1, 2$. At time t group 2 (the treatment group) is in wave $S(t)$ while group 1, which begins the survey one year later, is in wave $S(t) - 1$. Let $R(S(t)) = 1$ indicate that an individual remained in the survey (did not attrit) up to wave $S(t)$, and $R(S(t)) = 0$ 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:

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

[7]Estimates of the 99th percentiles are diverging from wave 3 onwards but measurement error is known to be large for the very highest incomes in household surveys (see for example Bricker et al. (2015)).

Figure 2: Selected quantiles of UKHLS/FRS (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.

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:

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

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

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, 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. In addition we further test for differential attrition processes across the two groups by estimating a statistical model of attrition. Further details are given below.

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

$$Y_i = \alpha + \beta I(g=2)_i + age_i'\gamma + \epsilon_i \quad (3)$$

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 1x1 age dummies and ϵ_i an error term with $E[\epsilon_i|I(g=2)_i, age_i]=0$. 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.[8] We also estimated models without controls (appendix A1) and the estimates line-up closely with our

[8]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.

main results. We report standard errors robust to heteroskedasticity.[9]

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.[10] Full results are included in table A1, appendix A. A further check confirmed that attrition 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 were statistically significant.

4 Results

4.1 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.

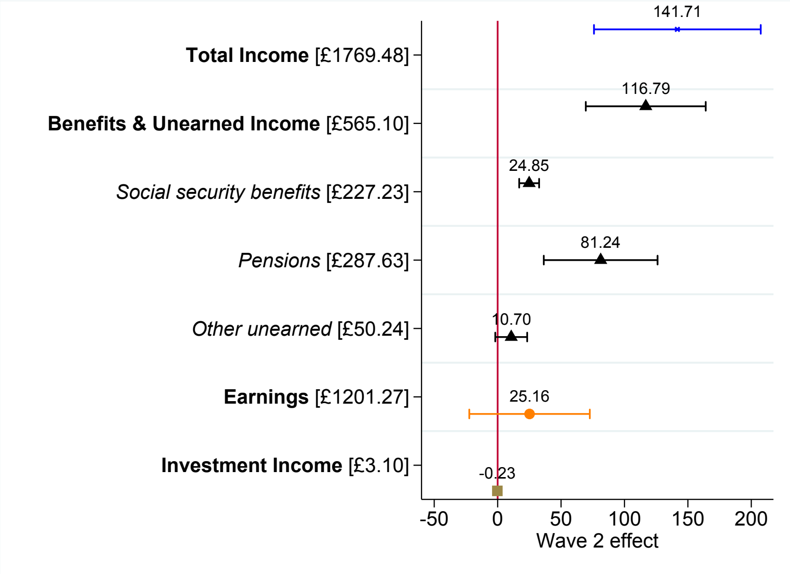
The causal effect of being interviewed for a second time, relative to a first, is to increase total monthly income by £141.71 or 8 percent. The effect is driven solely by changes in the benefits and unearned income sub-component of income. In particular, the strongest effects occur for social security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

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

[10] Small but statistically significant differences were observed for the share White (0.9 percentage points), Indian (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.

Figure 3: Effect of second interview on reported income



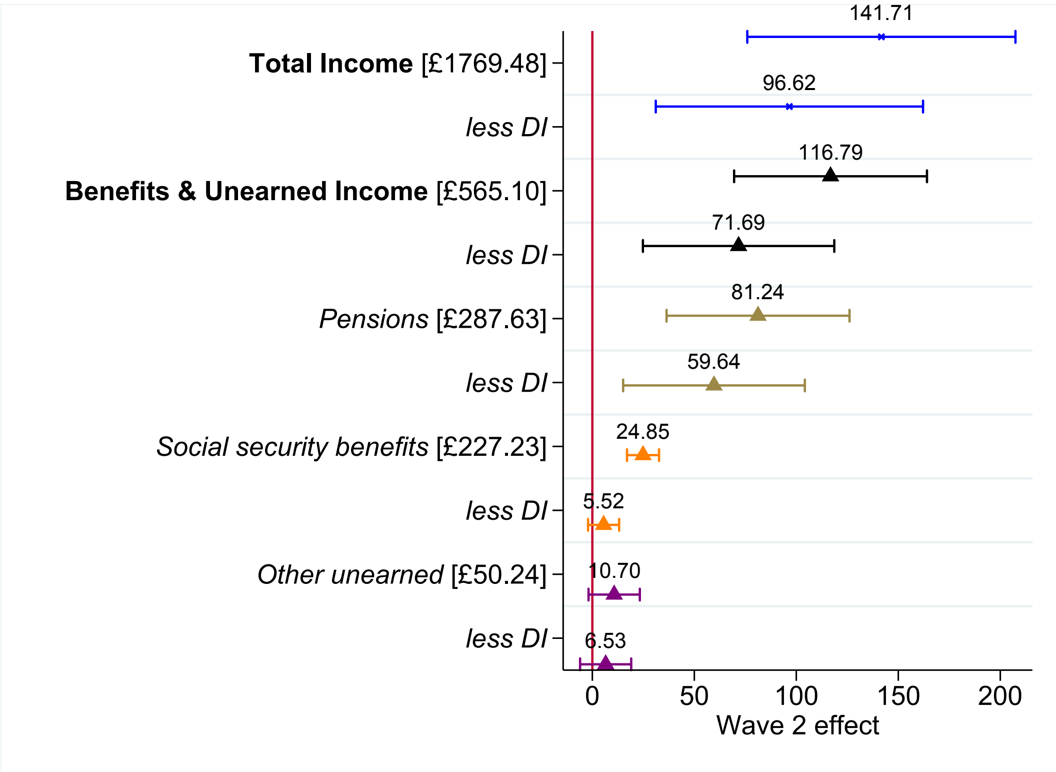
Notes: The point estimates presented correspond to β from equation (3) with the full-set of controls. Means from the baseline (wave 1) interview are reported in square brackets. Confidence intervals are calculated using robust standard errors. The controls are dummy variables for (number of categories 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).

Figure 4 examines the extent to which the reporting changes 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 percent. Overall the results show that a large share of the reporting improvements seen at the second interview are due to PC and not DI.

We also explored the possibility of heterogenous treatment effects by estimating models for subsamples of: pensioners, working age with children and working age without children.[12] 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 percent. Moreover, 73.8 percent 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

[12]We define 'pensioners' as those of UK state pension age (60 for women and 65 for men) and 'Working age' those below it.

Figure 4: Panel Conditioning effect of second interview on reported income



Notes: see figure 3 notes.

large. Benefits and unearned income increase by 35.9 percent of the wave 1 mean and 57.5 percent 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 percent of the wave 1 mean). The interested reader can find the full set of results in figures A1-A6, 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 reporting effects 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.

The effect of being interviewed for a second time is to increase the receipt of each of the 12 sources in the table. The effects are statistically significant (excluding income support)

Table 1: Effect of wave 2 interview on reporting of selected unearned sources

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

N = 29528

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.

and the magnitudes are non-trivial. For example, the effect on the state pension[13] is 1.33 percentage points or 5.6 percent of the wave 1 mean. Column (2) shows that much of the reporting 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 percent of the wave 1 mean.[14] We find no evidence of reporting 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.[15]

The above increases in income reporting are consistent with improvements in data quality, as state transfers are known to be under-reported in household surveys. Table A3 of appendix A presents evidence to support this claim by showing that the wave 2 estimates of benefit receipt are generally closer to known administrative totals, relative to estimates from wave 1.

To look for reporting changes at later interviews, we re-estimate (3) 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 reporting effect is always small and statistically insignificant. Therefore, the biggest change in response behaviour 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 reporting improvements are confined to the early waves of the panel (appendix A.3).

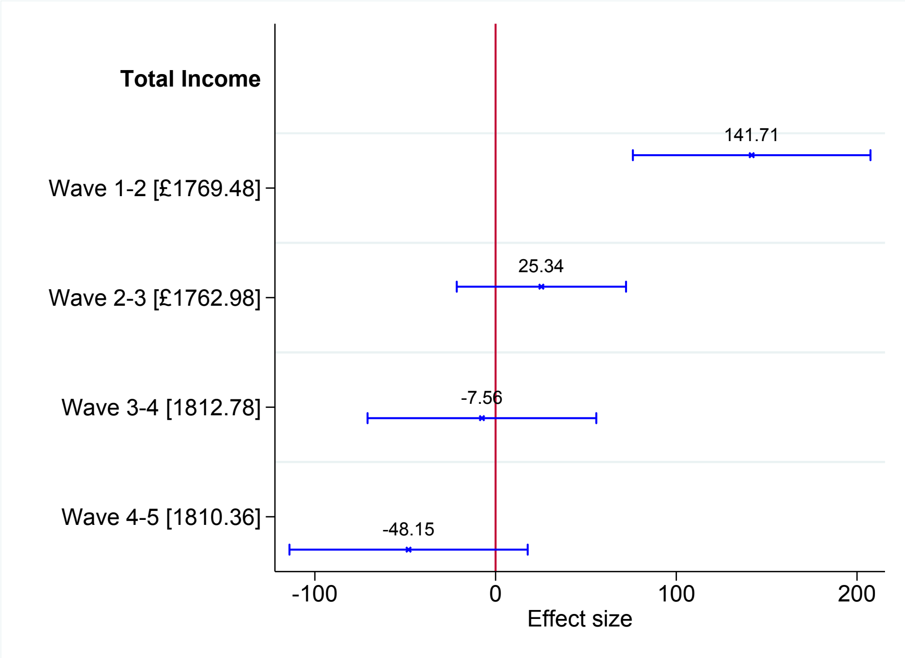
To comment on the implications of our findings, analysis of a short-panel of the early waves will suffer from bias. In contrast, estimates of change based on the full panel will benefit from the quality improvements that derive from repeated measurement. Distributional estimates based on the wave one cross-section will also suffer from under-reporting problems (relative to later cross-sections), but insofar as wave one of a panel is a cross-section, analysis from a similar cross-sectional survey could be expected to suffer from comparable under-reporting.

[13]See table 1 notes.

[14]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.

[15]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.

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.

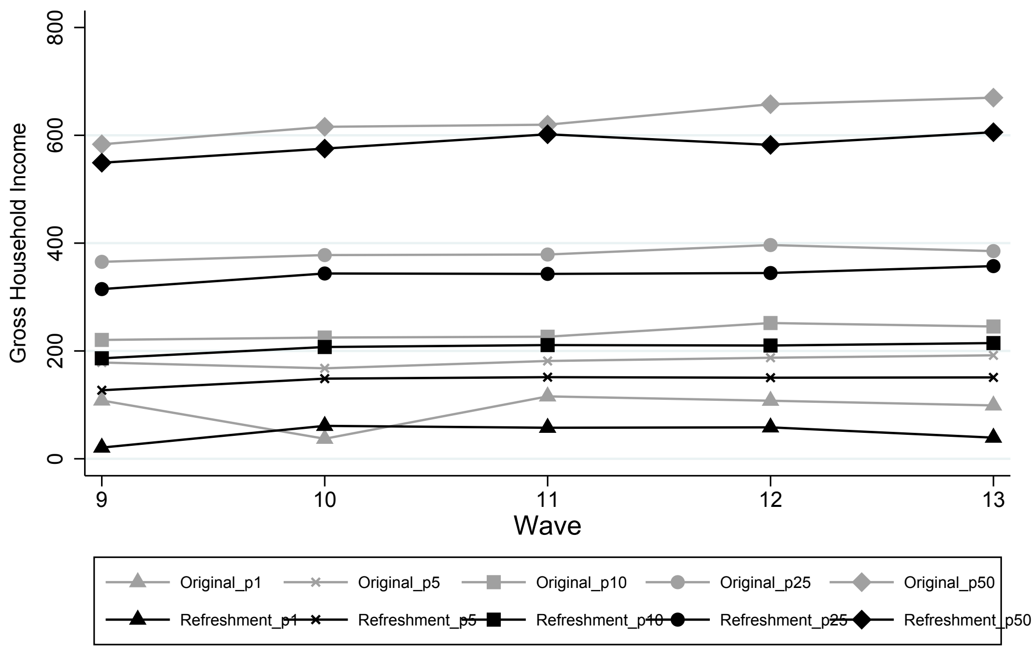
In the next two subsections we ask if the reporting changes documented here i) might generalise to other surveys and ii) whether they have meaningful economic consequences.

4.2 Evidence from the British Household Panel Survey

It is important to know whether the reporting pattern established 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

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.

notably decrease at wave 10. Thereafter, the gaps remain relatively stable.

We find similar reporting improvements 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 improvements 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 A.4.

4.3 Implications for inequality and mobility estimates

This section explores the consequences of the reporting changes 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.

The upper panel of table 2 shows 2010 estimates of seven inequality measures for our treated (second-time reporters) and control groups (first-time reporters). The measures are commonly used in inequality analysis and are: 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 center of the distribution, whereas, the Aitkinson measure is sensitive to changes in the tails. Columns 1 and 2 shows 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 reporting differences 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 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). This supports our claim that the observed reporting increases of the previous sections are consistent with reporting improvements.

We also ask whether the observed reporting improvements 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. The second half of table 2 explores this by comparing estimates of income mobility between 2010 and 2011 for the treated and control groups. 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 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 percent) 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 find significant inequality and mobility differences for this placebo check. Results support this and are provided in appendix table A4.

5 Explaining the panel conditioning effect

In this section we provide evidence on the mechanisms through which PC operates. We group the possible mechanisms as: implementer learning; respondent-interviewer rapport; and respondent learning.[16]

5.1 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 explain our results, 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.[17] Second, implementer 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.

[16] A seperate possibility is that survey participation leads to behavioural change as in Crossley et al. (2017). This looks implausible given that in the present paper the main effects are concentrated in pensions, which respondents cannot manipulate in the short-term.

[17] We show that there are few differences in the 2010 wave 1 and 2 interviewer experience distributions (table A5, appendix A).

Table 2: The effect of reporting changes on inequality and mobility estimates

	Control (g=1)	*Treated* (g=2)	*Difference*	*Gold standard*‡
Cross-section (t=2010)				
Gini	0.346	0.364	0.018	0.337
			(0.00941)	
Standard deviation of logs	0.693	0.637	-0.055*	0.612
			(0.0216)	
Atkinson	0.833	0.719	-0.114	0.389
			(0.228)	
Percentile ratios:				
90-10	4.488	3.996	-0.493***	3.972
			(0.0794)	
90-50	2.046	1.976	-0.071**	1.995
			(0.0251)	
10-50	0.456	0.494	0.039***	0.502
			(0.00688)	
Share below poverty line†	0.192	0.173	-0.019***	0.171
			(0.00446)	
Observations	14,731	14,911		
Transitions (between 2010 and 2011)				
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	11823	12350	-	-

Notes: Household income is net of taxes, deflated and equivalised 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 * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

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. Table 3 shows the results from this comparison. Columns 2 and 3 shows estimated 2009 percentiles from the HBAI and UKHLS, respectively, and column 4 shows their ratio, where a ratio greater than 1 indicates that UKHLS underestimates a percentile relative to HBAI. Columns 5-7 repeats 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 fieldwork agency learning is unlikely to be responsible for the observed improvements in data quality.

5.2 Respondent-interviewer rapport

Respondents and interviewers may establish a rapport during their first interview, which may result in more accurate reporting at the second. This can be tested 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.

Table 3: Comparison of UKHLS wave 1 to HBAI by calendar year

Percentile	*FRS 2009*	*UKHLS Wave 1 2009*	*Ratio (col 2/col 3)*	*FRS 2010*	*UKHLS Wave 1 2010*	*Ratio (col 5/col 6)*	*Ratio of ratios (col 7/col 4)*
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	1034.04	1057.79	0.98	1008.65	1023.25	0.99	1.01
75	1142.53	1172.90	0.97	1112.50	1144.67	0.97	1.00
80	1273.40	1313.70	0.97	1236.48	1274.61	0.97	1.00
85	1451.13	1487.80	0.98	1409.49	1453.66	0.97	0.99
90	1702.50	1745.69	0.98	1666.63	1726.21	0.97	0.99
95	2259.56	2193.66	1.03	2138.24	2170.68	0.99	0.96

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

Both the 'different interviewer' dummy and its interaction were highly insignificant (full results are included in table A6). 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.

5.3 Respondent learning

Respondents may have an improved comprehension of the complex interview or have updated their beliefs about the trustworthness 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 3 interviewer outcomes where we recoded them

Table 4: Effect of wave 2 interview on respondent behaviours

	(1) *Misunderstood questions*	(2) *Suspicious*	(3) *Queries confidentiality*
Wave 2	-0.01	-0.09***	-0.16***
	(0.005)	(0.003)	(0.003)
N	29502	29502	29365

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.

into binary indicators.[18] Full details of the questions and the construction of the interviewer observation variables are provided in appendix B.

Table 4 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.[19]

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 organisation - a stranger to them - who may share their sensitive data with third party organisations. 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.

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 (eg. 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.

[18] Our results are insensitive to changes in the chosen thresholds.

[19] Table A7 shows that confidentiality concerns are predictors of item non-response in the wave 1 cross-section.

6 Conclusions

We find that the quality of income data collected as part of a large-scale household panel survey improves over the life-time of the panel due to changes in respondent reporting behaviour. The largest changes in reported income are 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, potentially as it is difficult to distinguish changes in reporting behaviour 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 suggest that researchers analysing data from the early waves of a panel or with short panels (such as in randomised control trials) should proceed with caution. One possibility is that researchers may want to consider adjusting data from the first waves of data collection. Our findings are also relevant for studies based on cross-sectional data, which essentially forms wave one of a panel and so are indicative of the types of income sources that may be under-reported.

Our work is suggestive that respondent confidentiality concerns play a role in the observed patterns. Addressing these during data collection may bring data quality improvements. Separately, other sensitive variables collected as part of survey data eg. voting intentions, illicit behaviours may also show similar effects. Both of the later points are left for future work.

References

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2015. "The miracle of microfinance? Evidence from a randomized evaluation." *American Economic Journal: Applied Economics* 7 (1):22–53.

Blundell, Richard, Monica Costa Dias, Costas Meghir, and Jonathan Shaw. 2016. "Female

Labor Supply, Human Capital, and Welfare Reform." *Econometrica* 84:1705–1753.

Brewer, Mike, Ben Etheridge, and Cormac O'Dea. 2017. "Why are households that report the lowest incomes so well-off?" *Economic Journal* 127 (605):F24–F49.

Bricker, Jesse, Alice M. Henriques, Jacob Krimmel, and John Sabelhaus. 2015. "Measuring Income and Wealth at the Top Using Administrative and Survey Data." Finance and Economics Discussion Series 2015-30, Board of Governors of the Federal Reserve System (U.S.).

Chetty, Raj, David Grusky, Maximilian Hell, Nathaniel Hendren, Robert Manduca, and Jimmy Narang. 2017. "The fading American dream: Trends in absolute income mobility since 1940." *Science* 356 (6336):398–406.

Crossley, Thomas, Jochem de Bresser, Liam Delaney, and Joachim Winter. 2017. "Can survey participation alter household saving behavior?" *Economic Journal* 127 (606):2332–2357.

Das, Marcel, Vera Toepoel, and Arthur van Soest. 2011. "Nonparametric Tests of Panel Conditioning and Attrition Bias in Panel Surveys." *Sociological Methods & Research* 40 (1):32–56.

David, Martin H. and Christopher R. Bollinger. 2005. "I didn't tell, and I won't tell: dynamic response error in the SIPP." *Journal of Applied Econometrics* 20 (4):563–569.

Frick, Joachim R., Jan Goebel, Edna Schechtman, Gert G. Wagner, and Shlomo Yitzhaki. 2006. "Using Analysis of Gini (ANoGi) for Detecting Whether Two Sub-Samples Represent the Same Universe: The SOEP Experience." *Sociological Methods Research* 34 (4):427–468.

Halpern-Manners, Andrew, John Robert Warren, and Florencia Torche. 2014. "Panel Conditioning in the General Social Survey." *Sociological Methods & Research* .

Haushofer, Johannes and Jeremy Shapiro. 2016. "The Short-term Impact of Unconditional Cash Transfers to the Poor: Experimental Evidence from Kenya*." *The Quarterly Journal of Economics* 131 (4):1973.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond. 2016. "Long-Run Impacts of Childhood Access to the Safety Net." *American Economic Review* 106 (4):903–934.

Hurst, Erik, Geng Li, and Benjamin Pugsley. 2014. "Are household surveys like tax forms? Evidence from income underreporting of the self-employed." *Review of economics and statistics* 96 (1):19–33.

Karlan, Dean and Jonathan Zinman. 2011. "Microcredit in theory and practice: Using randomized credit scoring for impact evaluation." *Science* 332 (6035):1278–1284.

Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 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 (1):289–308.

Meyer, Bruce D and Nikolas Mittag. 2015. "Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness and holes in the safety net." Tech. rep., National Bureau of Economic Research.

Meyer, Bruce D. and James X. Sullivan. 2003. "Measuring the Well-Being of the Poor Using Income and Consumption." NBER Working Papers 9760, National Bureau of Economic Research, Inc.

———. 2011. "Viewpoint: Further results on measuring the well-being of the poor using income and consumption." *Canadian Journal of Economics* 44 (1):52–87.

Van Landeghem, Bert. 2014. "A test based on panel refreshments for panel conditioning in stated utility measures." *Economics Letters* 124 (2):236–238.

Yang, Tzu-Ting. 2017. "Family Labor Supply and the Timing of Cash Transfers: Evidence from the Earned Income Tax Credit." *Journal of Human Resources* :0115–6857R1.

Does repeated measurement improve income data quality?

Paul Fisher

Institute for Social and Economic Research, University of Essex

November 2017

Supplementary material

Appendix A

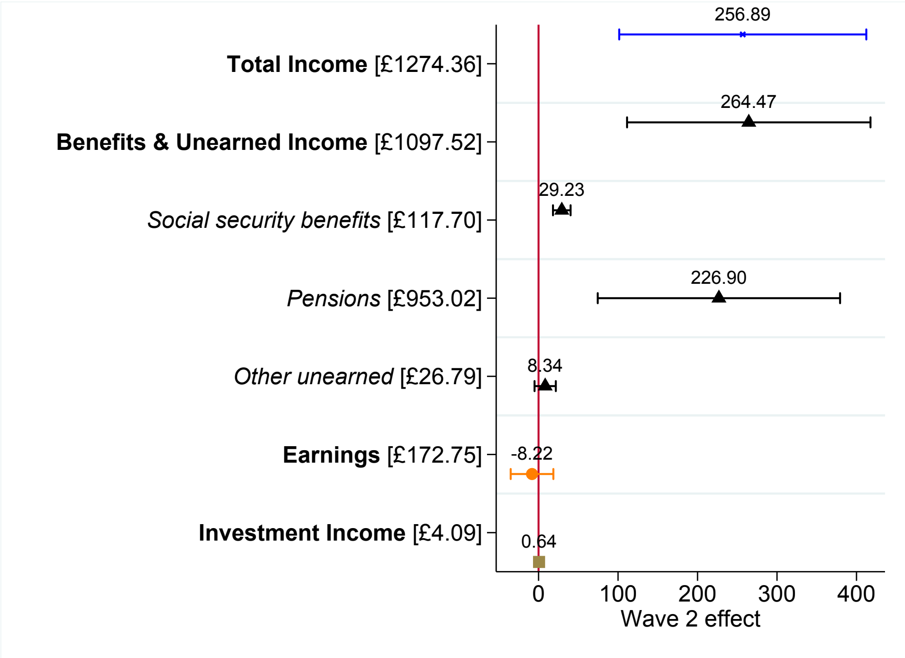
A1. Additional tables and figures

Table A1: Comparison of UKHLS wave one 2009 and 2010 responding samples

	2009	2010	Mean Diff	SE	N_Year One	N_Year Two
sex	0.4524	0.4555	-0.0031	0.0049	21447	20139
age	46.8445	47.0688	-0.2243	0.1800	21447	20139
age (bands):						
10-19 years old	0.0627	0.0603	0.0023	0.0024	21447	20139
20-29 years old	0.1470	0.1467	0.0003	0.0035	21447	20139
30-39 years old	0.1718	0.1654	0.0065	0.0037	21447	20139
40-49 years old	0.1881	0.1915	-0.0034	0.0038	21447	20139
50-59 years old	0.1571	0.1575	-0.0004	0.0036	21447	20139
60-69 years old	0.1432	0.1458	-0.0026	0.0034	21447	20139
70+	0.1301	0.1329	-0.0027	0.0033	21447	20139
ethnicity:						
white	0.9136	0.9046	0.0090**	0.0029	20646	19339
mixed	0.0104	0.0108	-0.0004	0.0010	20646	19339
indian and chinese	0.0217	0.0270	-0.0052***	0.0015	20646	19339
other asian	0.0239	0.0260	-0.0021	0.0016	20646	19339
african or black caribean	0.0231	0.0226	0.0005	0.0015	20646	19339
other	0.0074	0.0090	-0.0016	0.0009	20646	19339
highest qualification:						
Degree	0.2025	0.2081	-0.0055	0.0040	21419	20104
Other higher degree	0.1128	0.1125	0.0003	0.0031	21419	20104
A-level	0.1894	0.1911	-0.0017	0.0039	21419	20104
GCSE	0.2121	0.2115	0.0006	0.0040	21419	20104
other	0.1111	0.1045	0.0066*	0.0030	21419	20104
no qualification	0.1720	0.1723	-0.0003	0.0037	21419	20104
marital status:						
married or civil partnership	0.5087	0.5084	0.0003	0.0049	21443	20132
cohabiting	0.1258	0.1244	0.0013	0.0032	21443	20132
single and never married	0.2139	0.2189	-0.0050	0.0040	21443	20132
divorced or separated	0.0886	0.0846	0.0040	0.0028	21443	20132
widowed	0.0630	0.0637	-0.0007	0.0024	21443	20132
long-standing illness or impairment	0.3701	0.3694	0.0007	0.0047	21414	20088
SF-12 Physical Component Summary	49.2970	49.4700	-0.1731	0.1174	20164	18642
SF-12 Mental Component Summary	50.6627	50.5925	0.0702	0.1020	20164	18642
housing tenure:						
owned	0.3048	0.2982	0.0066	0.0045	21394	20098
mortgage	0.3928	0.3810	0.0118*	0.0048	21394	20098
rent	0.1360	0.1425	-0.0065	0.0034	21394	20098
social housing	0.1628	0.1751	-0.0123***	0.0037	21394	20098
# bedrooms	2.9238	2.9173	0.0064	0.0099	21430	20117
# people	2.7510	2.7841	-0.0330*	0.0134	21447	20139
# children	1.7237	1.7310	-0.0072	0.0151	6733	6590

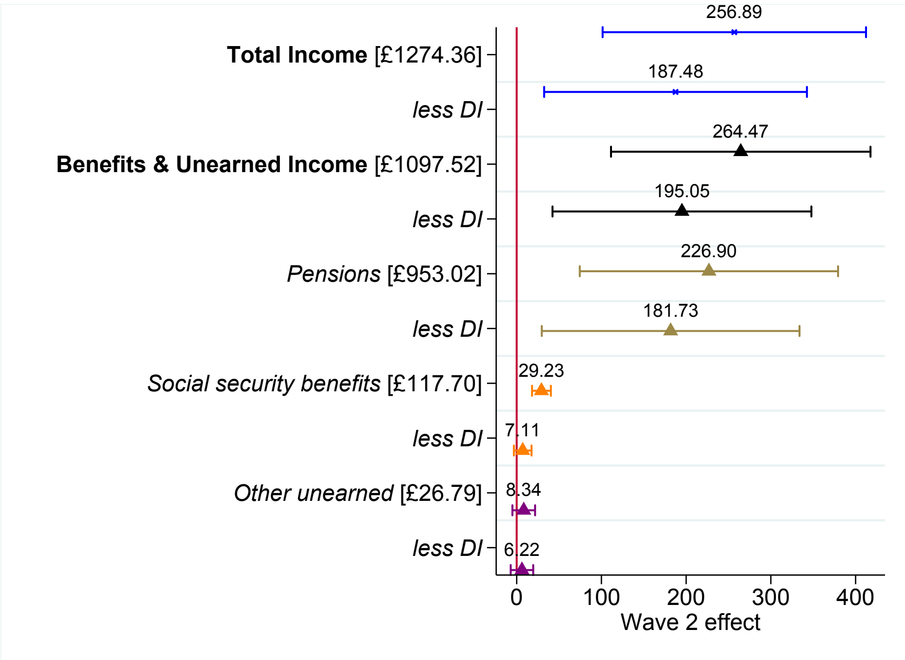
Notes: *p < 0.05, **p < 0.01, ***p < 0.001.

Figure A1: Effect of second interview on reported income (pensioner sample)



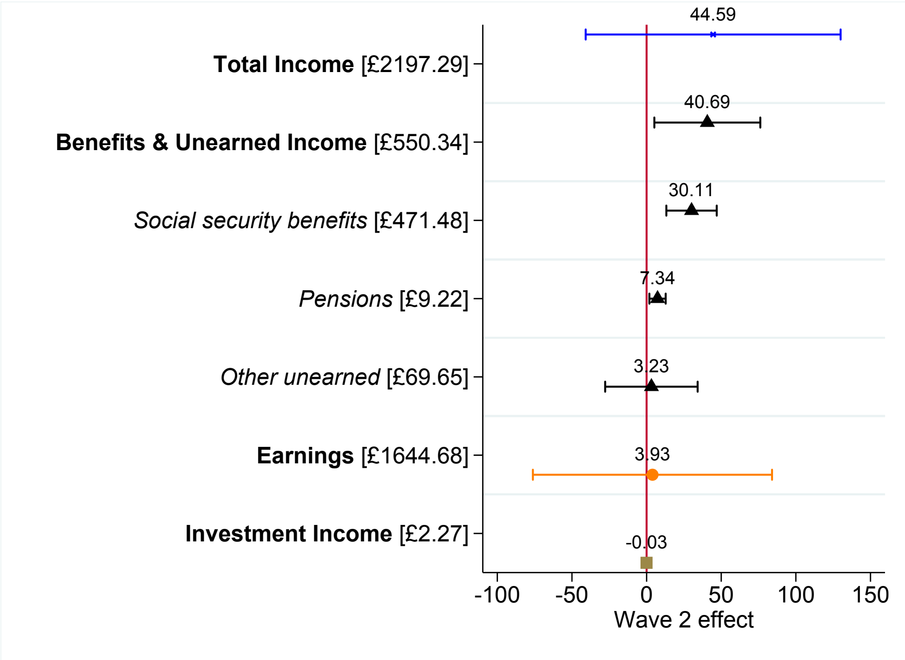
Notes: see figure 3 notes. Sample restricted to respondents of state pension age.

Figure A2: Panel Conditioning effect of second interview on reported income (pensioner sample)



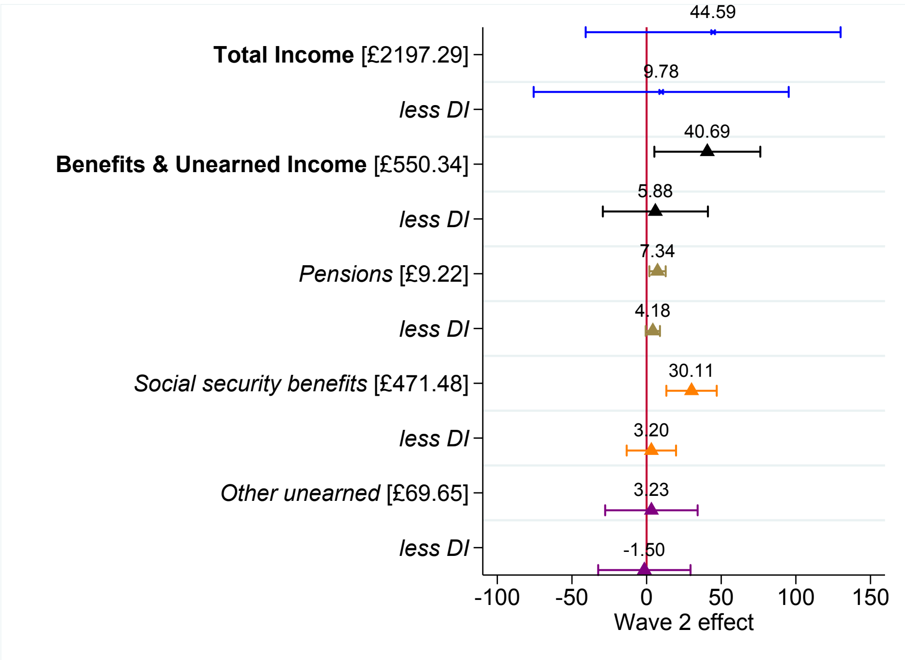
Notes: see figure 3 notes. Sample restricted to respondents of state pension age.

Figure A3: Effect of second interview on reported income (working age with children sample)



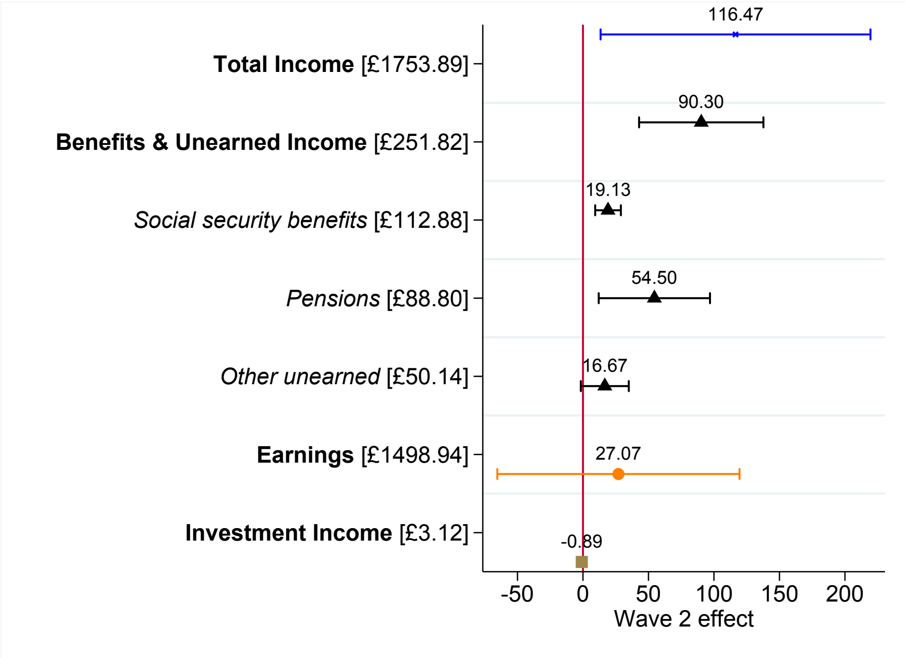
Notes: see figure 3 notes. Sample restricted to respondents less than state pension age and with children.

Figure A4: Panel Conditioning effect of second interview on reported income (working age with children sample)



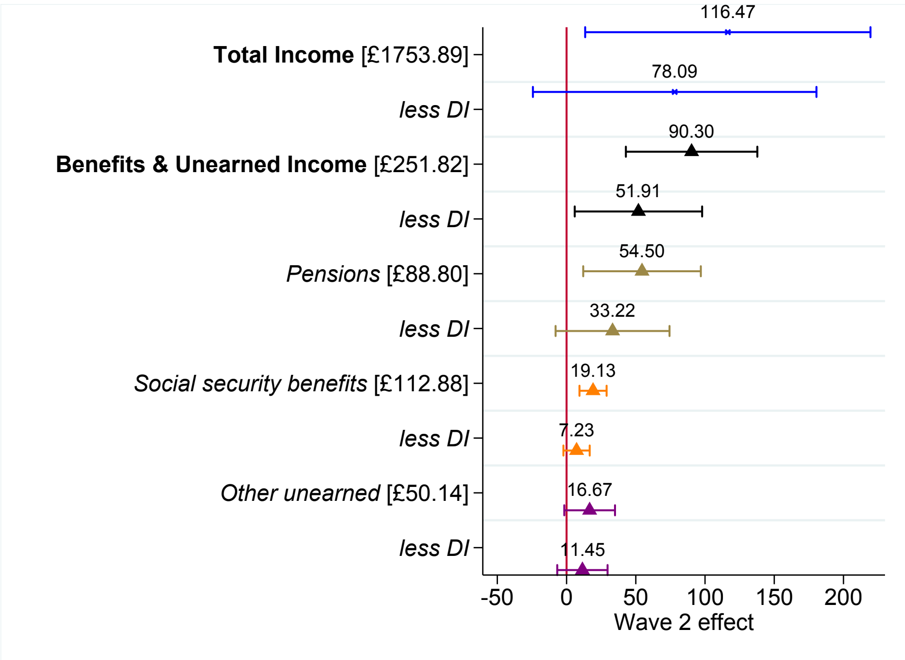
Notes: see figure 3 notes. Sample restricted to respondents less than state pension age and with children.

Figure A5: Effect of second interview on reported income (working age no children sample)



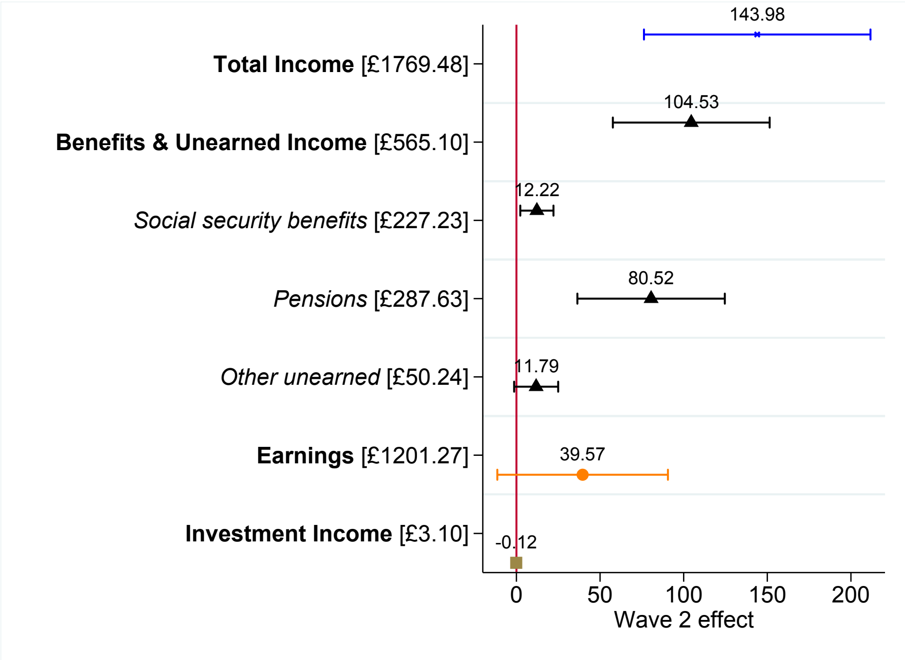
Notes: see figure 3 notes. Sample restricted to respondents less than state pension age and without children.

Figure A6: Panel Conditioning effect of second interview on reported income (working age no children sample)



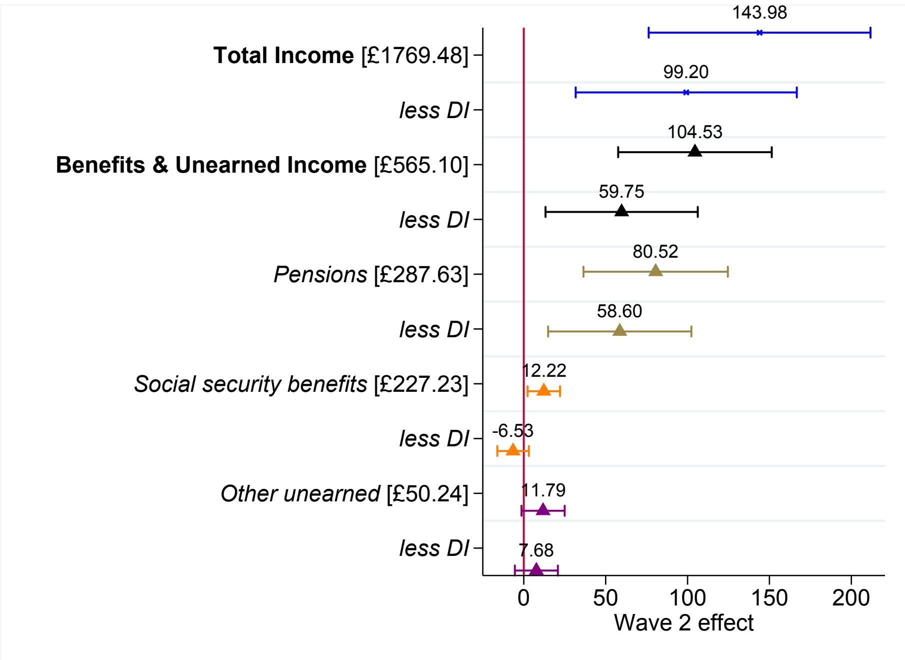
Notes: see figure 3 notes. Sample restricted to respondents less than state pension age and without children.

Figure A7: Models without controls: replication of figure 3



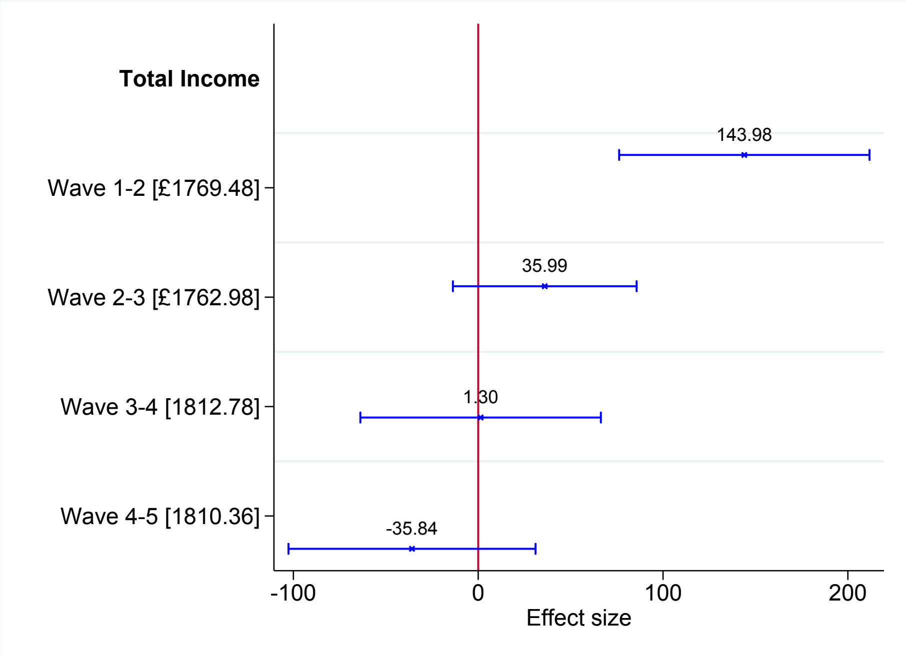
Notes: see figure 3 notes.

Figure A8: Models without controls: replication of figure 4



Notes: see figure 3 notes.

Figure A9: Models without controls: replication of figure 5



Notes: see figure 3 notes.

Table A2: Models without controls: replication of table 1

	Receipt		***Amount***	
	DI + PC	PC	DI + PC	PC
Pensions				
State pension [23.62, £528.24]	1.391***	0.538*	3.712	4.399
	(0.212)	(0.221)	(4.591)	(4.630)
Private pension [5.00, £455.50]	1.961***	0.907***	71.568	91.720
	(0.264)	(0.255)	(62.191)	(70.688)
Employer pension [14.84, £822.87	1.641***	0.516	262.097*	259.576*
	(0.350)	(0.348)	(121.416)	(126.692)
Spouse employer pension [2.76, £649.47]	0.575**	0.224	-81.634	-43.900
	(0.195)	(0.190)	(221.910)	(223.093)
Family benefits				
Working Tax Credit [6.33, £254.43]	0.982***	0.438	-7.198	-9.411
	(0.287)	(0.282)	(8.760)	(8.861)
Child benefit [24.04, £145.75]	-0.020	-0.477	-2.264	-2.453
	(0.425)	(0.425)	(1.432)	(1.433)
Child tax credit [17.68, £297.48]	0.663	-0.137	-12.973	-10.693
	(0.406)	(0.403)	(7.424)	(7.513)
Disability benefits				
Incapacity benefit [2.71, £409.44]	0.595**	0.253	2.683	1.511
	(0.199)	(0.194)	(10.296)	(10.541)
Disability living allowance [4.96, £305.65]	1.175***	0.542*	-13.080	-16.884*
	(0.267)	(0.261)	(8.220)	(8.275)
Low income benefits				
Income support [4.81, £355.62]	0.006	-0.237	-5.201	-4.880
	(0.250)	(0.247)	(12.487)	(12.707)
Housing benefit [10.57, £372.02]	0.366	-0.759*	-5.387	-4.940
	(0.362)	(0.354)	(7.162)	(7.346)
Council tax benefit [12.10, £101.06]	1.318***	-0.245	2.219	1.549
	(0.390)	(0.380)	(4.471)	(4.629)

Table A3: Benefit receipt (thousands): Comparison to administrative sources

	UKHLS		*Administrative*		*Ratio*	
	Wave 1	*Wave 2*	*Wave 1*	*Wave 2*	*Wave 1*	*Wave 2*
Individual benefits:						
State pension	20602	22712	22961	23156	0.897	0.981
Incapacity Benefit	2711	2775	3840	3387	0.706	0.819
Disability living allowance	4534	5574	6283	6401	0.722	0.871
Household benefits:						
Housing benefit	7516	8920	9345	9730	0.804	0.917
Council tax benefit	8627	10324	11376	11679	0.758	0.884
Family level benefits:						
Tax credits	10340	9696	12560	11980	0.823	0.809
Child benefit	14256	14610	15612	15727	0.913	0.929
Income support†	6510	6805	10188	12134	0.639	0.561

Sources: Department for Work and Pensions (https://www.gov.uk/government/publications/benefit-expenditure-and-caseload-tables-2013); De Agostini and Sutherland (2014): EUROMOD UK country report; HM Revenue & Customs Statistics (https://www.gov.uk/government/statistics/child-and-working-tax-credits-statistics-finalised-annual-awards-2015-to-2016)

† Includes Job Seeker's Allowance and Employment and Support Allowance (ESA). ESA receipt was strongly increasing during our observation window. It was introduced in 2008 to replace Income Support for the disabled.

Table A4: The effect of reporting changes on inequality and mobility estimates

	Control (g=1)	*Treated (g=2)*	*difference*
Cross-sectional measures (t=2012)			
Gini	0.336	0.319	-0.0168
			(0.00962)
Standard deviation of logs	0.585	0.579	-0.00642
			(0.0123)
Atkinson	0.553	0.318	-0.235
			(0.150)
Ratios:			
90-10	3.790	3.720	-0.0703
			(0.0744)
90-50	1.976	1.903	-0.0734**
			(0.0284)
10-50	0.521	0.512	-0.00987
			(0.00788)
Share below 60% median hh income	0.153	0.159	0.00558
			(0.00575)
Observations	11,244	10,553	
Transitions (between 2012 and 2013)			
Exit poverty (share)	0.416	0.402	-0.014
Enter poverty (share)	0.076	0.071	-0.005
Mean income change (£ per week)	-10.325	-11.688	-1.363
Mean income change (log)	-0.016	-0.002	0.014*
Changed decile (share)	0.617	0.624	0.007
Observations	9861	9360	

Notes: Household income is net of taxes, deflated and equivalised using the 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 A5: Interviewer experience: number of interviews completed

Percentile	*Wave 1 (2010)*	*Wave 2 (2010)*
1	2	4
5	11	16
10	22	27
25	47	52
50	85	94
75	139	148
90	199	207
95	237	246
99	324	330
Mean	100.00	108.09
sd	71.2	73.4

Notes: Sample is defined as in the methodology section.

Table A6: Effect of changing interviewer on reported total income

	(1)	(2)
Wave 2	141.71***	126.07**
	(33.542)	(41.346)
Different interviewer		5.56
[36.70]		(39.568)
Different interviewer X Wave 2		45.77
		(74.202)
Interviewer:		
No. Interviews completed		-0.29
[100.00]		(0.262)
Female		4.21
[54.55]		(32.577)
Age		2.59
[57.68]		(1.859)
Observations	29528	29528

Notes: Robust 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).

Table A7: Effect of confidentiality concerns on item non-response

	(1) *Earnings*	(2) *2nd Job*	(3) *Self-employment*	(4) *Investment*	(5) *Pensions*
Respondent queries:					
purpose [0.43]	0.04***	0.04*	-0.01	0.05***	0.00
	(0.006)	(0.020)	(0.024)	(0.009)	(0.002)
interview length [0.29]	-0.00	0.01	0.01	0.02*	0.00
	(0.006)	(0.022)	(0.025)	(0.010)	(0.002)
panel design [0.03]	-0.01	-0.03	-0.12*	-0.09***	0.00
	(0.016)	(0.056)	(0.054)	(0.023)	(0.007)
confidentiality [0.19]	0.10***	0.09**	0.10***	0.11***	0.01**
	(0.008)	(0.028)	(0.028)	(0.011)	(0.003)
incentive/payment [0.05]	0.00	0.01	-0.01	0.02	-0.00
	(0.013)	(0.046)	(0.052)	(0.022)	(0.005)
other query [0.03]	0.04*	0.07	0.01	0.08***	0.00
	(0.018)	(0.065)	(0.074)	(0.025)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	19015	1795	2197	12862	10937

Notes: Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.
Sample of wave 1 respondents.
For a list of control variables, see the notes to figure 3.

A2. Effects of imputation

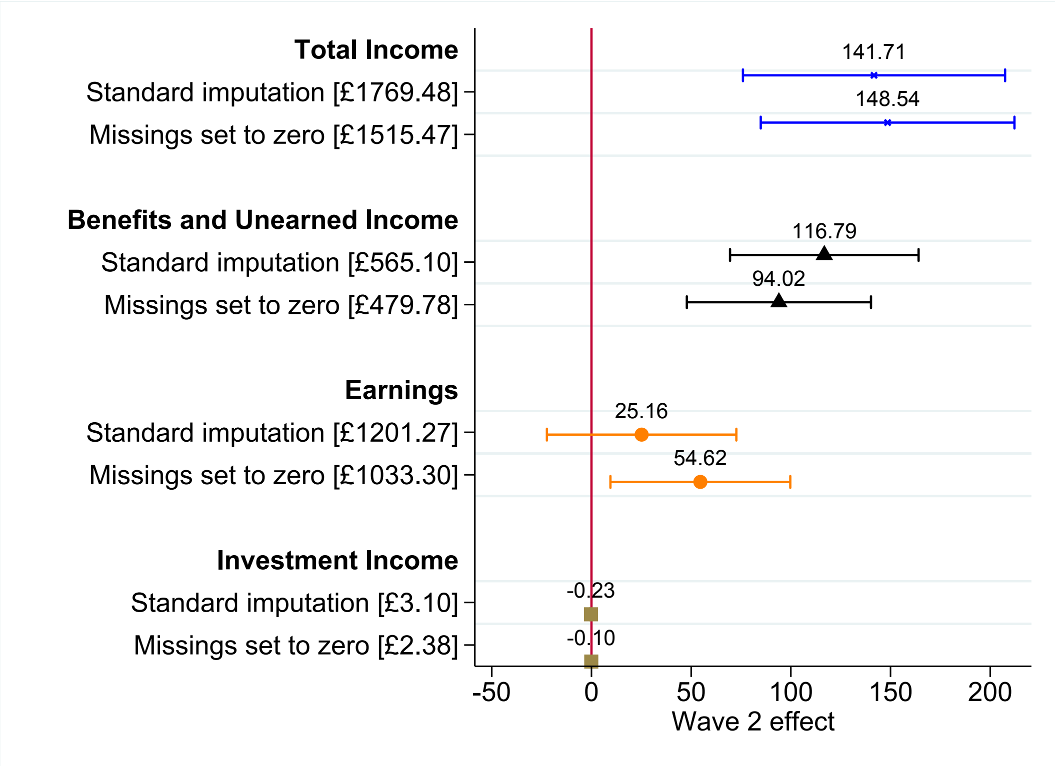
An important question is whether imputation of missing data masks the extent of reporting differences across waves. To explore this issue, we estimate equation (3) for income and its sub-components, where we set to zero any source which has missing income information. In this way, changes in the extent of missingness are not corrected by imputation.

Figure A10 presents the estimates (labelled imputes to zero) alongside our main results (labelled standard imputes). Estimates of the effect on total income are stable but this hides offsetting changes within its components. For earnings, the estimated effects with zero imputes are larger relative to those using standard imputes and they have become statistically significant. This implies that imputes produced as standard in most household surveys help in reducing some of the differences due to differential reporting behaviour across the early waves of the panel.

In contrast, for benefits and unearned income, we observe that the effects have slightly fallen in magnitude when using zero values for missing amounts. This results as our main effects for benefits and unearned income are driven by changes in reported receipt, but some of these are now not binding (the ones with a missing amount that is now set to zero).

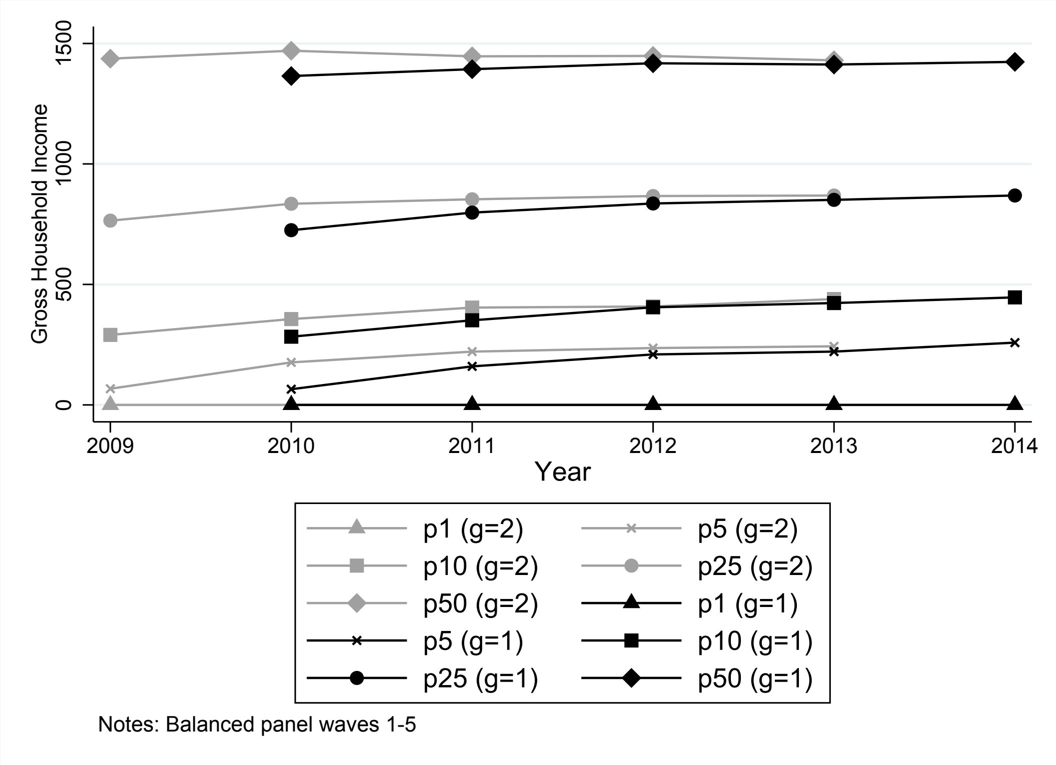
Overall, the results suggest imputation can work to correct data quality differences across the initial waves.

Figure A10: Effect sizes when setting missing amounts to zero



Notes: see figure 3 notes. 'Imputes to zero' sets missing amounts to zero. 'Standard imputation' replaces missing amounts with the imputes of the data providers.

Figure A11: Income trends by group

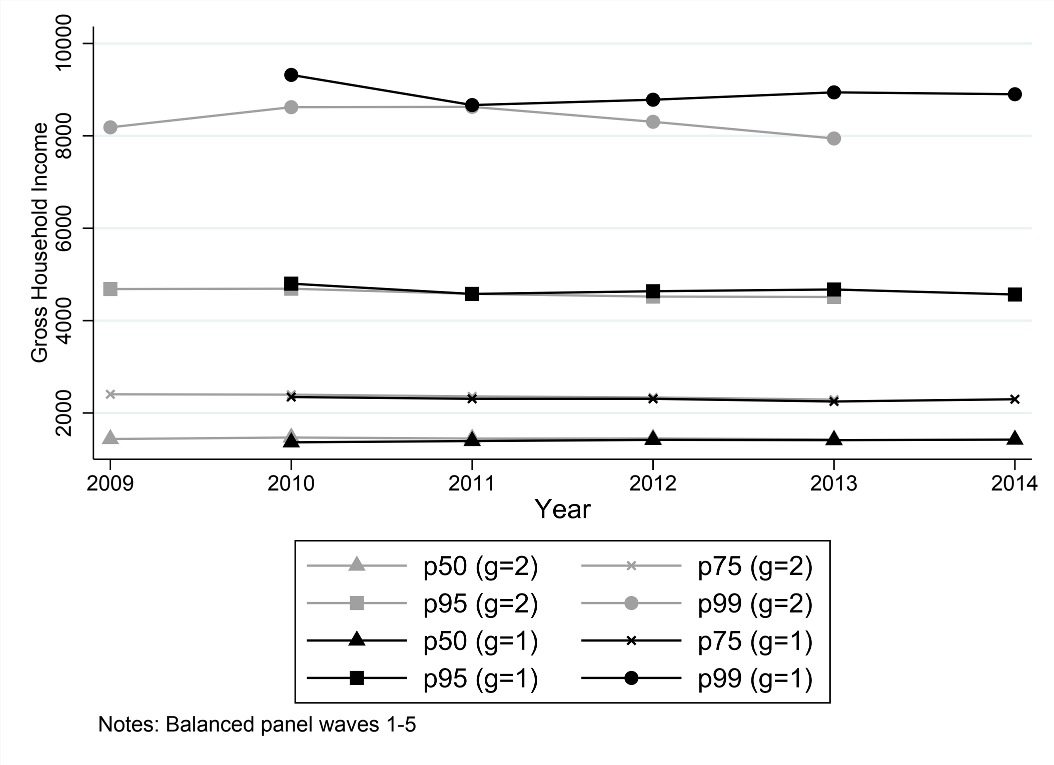


A3. Selected quantiles by year and group

Figure A11 shows the 1st, 5th, 10th, 25th and 50th percentiles. For each quantile, the $g = 2$ sample gives higher estimates of income in both 2010 (waves 1 and 2) and to a lesser extent in 2011 (waves 2 and 3) - but not from 2012 (waves 3 and 4) onwards. This suggests that the data quality improvements from repeated measurement occur early in the lifetime of the panel.

Figure A12 repeats the exercise for the upper half of the distribution showing the 50th, 75th, 95th and 99th percentiles. The estimates from the different subsamples line-up remarkably closely and so it would seem that any survey effects are confined to the bottom half of the distribution.

Figure A12: Income trends by group



A4. Further evidence from the BHPS

Figure A13 shows results from the analysis of the BHPS Northern Ireland refreshment sample. Northern Ireland did not form part of the original BHPS sample and so our 'original' comparison sample consists of respondents living in England, Scotland or Wales.[20] 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 the higher percentiles, presented, there is no obvious reporting difference and this fact could reflect underlying differences in the composition of incomes.

[20]NI was not included in the HBAI data until 2002/03.

Figure A14: Comparison of UKHLS and BHPS item non-response rates by wave

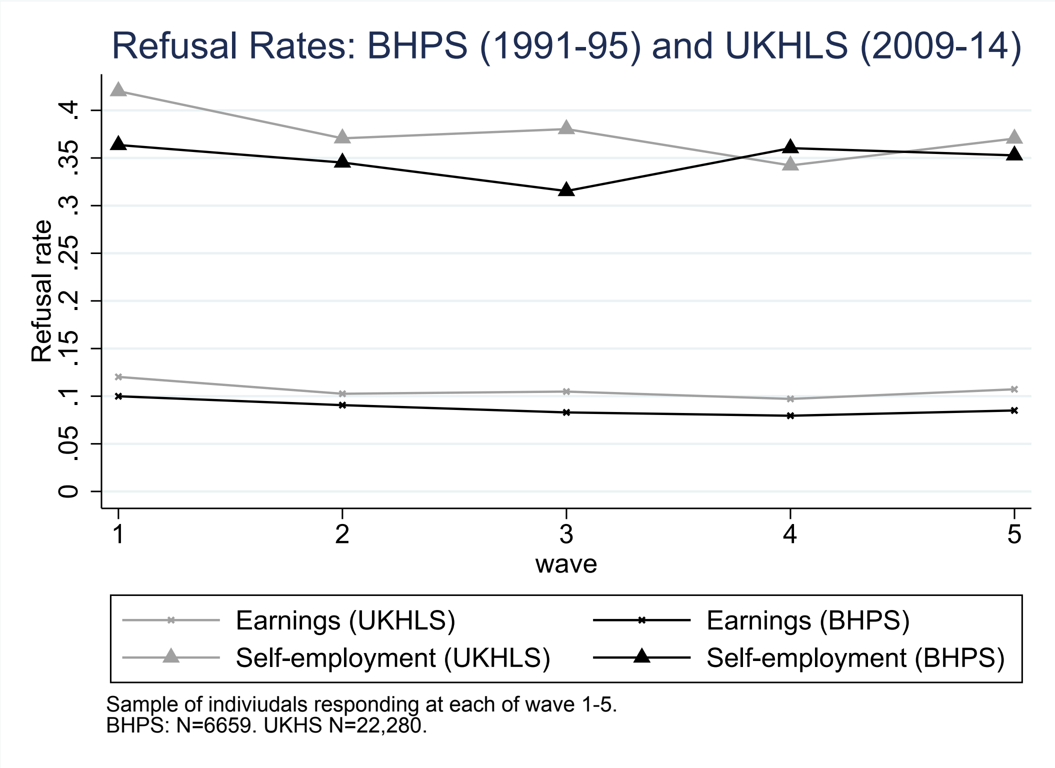
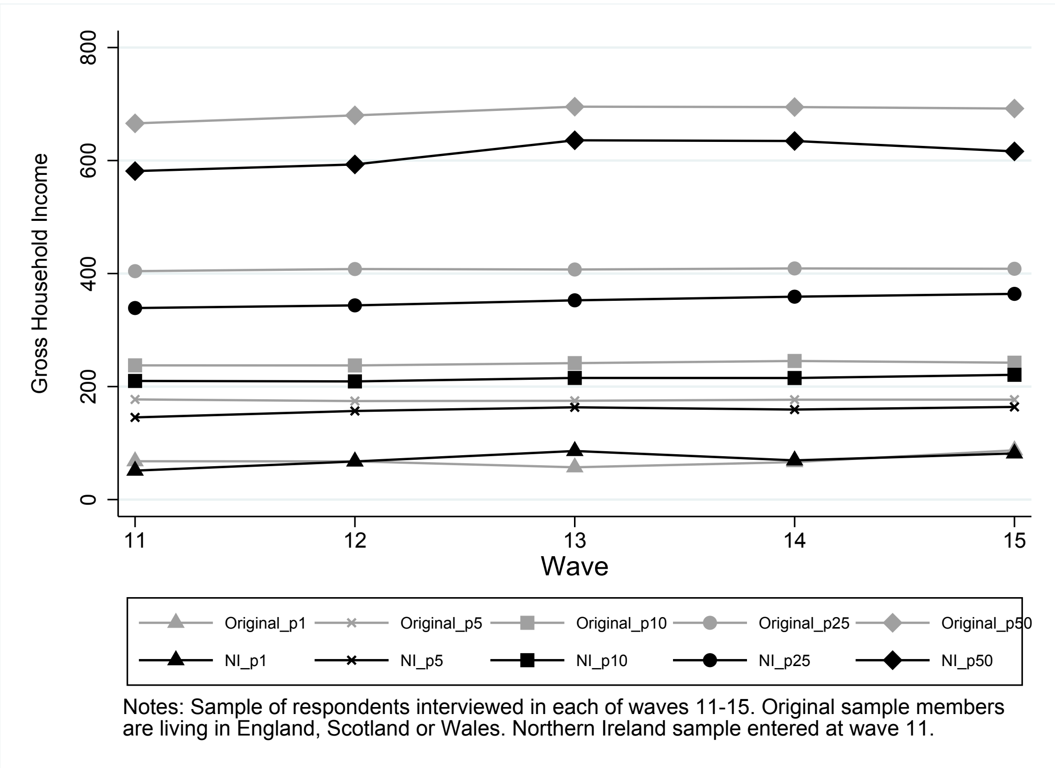


Figure A13: BHPS Northern Ireland refreshment sample (wave 11)



We compare trends in BHPS and UKHLS item non-response rates at the start of each panel.

We focus on earnings from main job and self-employment profit as the questions are identical in both surveys. Figure A14 shows that following wave 1, we see a fall in item non-response rates in both surveys, although the fall appears to be stronger in the UKHLS sample.[21] Between waves one and two the earnings refusal rate fell from 12 to 10 percent and self-employment from 42 to 37 percent in UKHLS and from 10 to 9 percent and 36 to 35 percent in BHPS. Fitting a linear regression line through the data points confirms the negative trend in both surveys: for earnings the coefficient on the wave trend is -0.41 for BHPS and -0.31 for UKHLS; and for self-employment for BHPS -0.07 and -1.28 for UKHLS.

[21]We observe that the level of non-response is similar in both surveys. The rates are slightly higher in UKHLS. For example, at wave 1 earnings refusal rates are 10 and 12 percent; and self-employment refusal rates are 36 and 42 percent, for BHPS and UKHLS, respectively. Given that the BHPS started in the early 1990's and UKHLS some 18 years later, the differences may reflect the decline in data quality over time that has been observed in numerous surveys and across countries.

Appendix B (Data appendix)

This appendix provides details of the income questions asked in Understanding Society, including details on the use of DI.

B1. Benefits and unearned income module

The showcard of 9 broad payment types lists: Unemployment-related benefits or National Insurance Credits; Income Support; Sickness, disability or incapacity benefits; Any sort of pension including a private pension or the State pension; Child Benefit; Tax credits; Any other family related benefits or payment; Housing or Council Tax Benefit; Income from any other state benefit.

The 39 income sources are allocated as:

Pensions:

1) ni retirement/state retirement (old age) pension

2) pension, previous employer

3) pension from a spouse's previous employer

4) private pension/annuity

5) widow's or war widow's pension

Social security benefits:

6) widowed mother's allowance / widowed parent's allowance / bereavment allowance

7) pension credit (incl. guarantee credit saving credit)

8) severe disablement allowance

9) industrial injury disablement allowance

10) disability living allowance

11) attendance allowance

12) carer's allowance (was invalid care allowance)

13) war disablement pension

14) incapacity benefit

15) income support

16) job seeker's allowance

17) national insurance credits

18) child benefit (incl. lone-parent child benefit payments)

19) child tax credit

20) working tax credit (incl. disabled person's tax credit)

21) maternity allowance

22) housing benefit

23) council tax benefit

24) foster allowance / guardian allowance

25) rent rebate

26) rate rebate

27) employment and support allowance

28) return to work credit

29) in-work credit for lone parents

Other unearned income:

30) educational grant (not student loan or tuition fee loan)

31) trade union / friendly society payment

32) maintenance or alimony

33) payments from a family member not living here

34) rent from boarders or lodgers (not family members) living here

35) rent from any other property

36) sickness and accident insurance

Other:

37) other disability related benefit or payment

38) any other regular payment

39) income from any other state benefit

A scripting error at wave 1 meant that amounts were not collected for respondents who reported receipt of sources 37-39. The coverage of these sources is small so we deduct them from our income totals following wave 1 to ensure consistency of our totals across waves.

At the end of stage one above, respondents who fail to report a source at wave t but reported it at wave t-1 are asked:

Can I just check, according to our records, you have in the past received [x]. Are you currently receiving [x], either just yourself or jointly?

B2. Employee earnings

Q1) Can I just check, did you do any paid work last week - that is in the seven days ending last Sunday - either as an employee or self-employed?

Q2) Even though you weren't working did you have a job that you were away from last week?

Q3) Are you an employee or self-employed?

Q4) If an employee on Q3): The last time you were paid, what was your gross pay - that is including any overtime, bonuses, commission, tips or tax refund but before any deductions for tax, National Insurance or pension contributions, union dues and so on? How long a period did that cover?

Q5) Is this the amount you usually receive (before any statutory sick pay or statutory maternity, paternity or adoption pay)?

Q6) If 'no' to Q5): How much are you usually paid? How long a period did that cover?

Q7) If 'no' to Q5): And is that before or after any deductions for tax, National Insurance, union dues and so on or are there usually no deductions at all made from your salary?

B3. Self-employee earnings

Q6) If a self-employee on Q3): In this job/business are annual business accounts prepared for the Inland Revenue for tax purposes?

Q7) If yes to Q6): What was the amount of (your share of) the profit or loss figure shown on these accounts for this period? (And month/year accounts began and ended)

Q8) Does this figure relate to profit or loss?

Q9) Can i just check, is that figure before deduction of income tax?

Q10) Can i just check, is that figure before deduction of National Insurance?

Q11) If no to Q6): After paying for any materials, equipment or goods that you use(d) in your work, what was your weekly or monthly income, on average, from this job/business over the last 12 months?

Q12) Was that weekly or monthly income?

Q13) Can i just check, is that figure before deduction of income tax?

Q14) Can i just check, is that figure before deduction of National Insurance?

B4. Second job earnings

Q15) Do you currently earn any money from a second job, odd jobs, or from work that you might do from time to time, apart from any main job you have?

Q16) If yes to 15): Before tax and other deductions, how much do you earn from your second and all other occasional jobs in a usual month?

B5. Investment income

Q17) In the past 12 months how much have you personally received in the way of dividends or interest from any saving and investments you may have?

Where respondents cannot give an exact amount in 17) they are presented with a series of unfolding brackets where they can bound their annual investment income. For individuals reporting bounds, the data providers impute an amount.

B6. Interviewer observations

Misunderstood questions: In general, how would you describe the respondents understanding of the question?

1 Excellent

2 Good

3 Fair

4 Poor

5 Very poor

Responses 2-4 are coded as one and category 1 as zero.

Suspicious: Was the respondent suspicious about the study after the interview was completed?

1 No, not at all suspicious

2 Yes, somewhat suspicious

3 Yes, very suspicious

Responses 2-3 are coded as one and category 1 as zero.

Queries confidentiality: Did the household respondent query any of the following topics?

1 purpose (e.g. 'Whats the purpose? Whats all this about?')

2 interview length (e.g. 'How long will this take?')

3 panel design (e.g. 'Youll be coming back next year?')

4 confidentiality (e.g. 'Whos going to see the answers?')

5 incentive/payment (e.g. 'Whats in it for us/me?')

6 other query

A 0/1 indicator is constructed from the responses to item 4.