Does Repeated Measurement Improve Income Data Quality?

Paul Fisher (Institute for Social and Economic Research, University of Essex, UK)

Paper prepared for the 34th IARIW General Conference

Dresden, Germany, August 21-27, 2016

Session 2G: Income Distribution Issues

Time: Monday, August 22, 2016 [Afternoon]

Does repeated measurement improve income data quality?*

Paul Fisher†

Institute for Social and Economic Research, University of Essex

May 2016

Abstract

This paper provides evidence that measurement error in income data systematically reduces across waves of a panel survey and particularly across the initial waves. We exploit a unique feature of the survey design of the new UK Understanding Society Study - that, accounting for attrition, random samples of households are responding at different waves of the panel in a given calendar year - as a quasi-experiment to estimate the effect of repeated interviewing on reported income. Our OLS estimates indicate that the effect of being interviewed for a second time is to increase the mean of reported monthly income by £124 (7.8 percent). Dependent interviewing a common recall device used in household panel surveys takes effect only after a first survey interview. It can explain approximately one third of the observed increase in reported income, with the remaining share attributed to changes in respondent reporting behaviour (panel conditioning). The results have implications for the reliability of any analysis based on repeated survey measures of incomes and also for the comparability with income data from cross-sectional surveys.

Keywords: income, measurement error, household survey, panel conditioning

*I would like to thank Jonathan Burton; Mike Brewer; Nick Buck; Emanuele Ciani; Tom Crossley; Annette Jackle; Steve Pudney and seminar participants at the 'Improving the measurement of household finances' workshop, University of Essex, for their useful comments. Financial support received from the ESRC is also gratefully acknowledged. 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

Economists, other social scientists and policy makers wanting to measure material living standards commonly rely on data from surveys that measure the income of the same individual at multiple points in time eg. large-scale household panel surveys or panel data collected as part of field experiments. Whilst it is known that income is under-reported in household surveys, particularly state transfers (Meyer and Sullivan (2003, 2011); Lynn et al. (2012a); Brewer, Etheridge, and O'Dea (forthcoming)) but also self-employment income (Hurst, Li, and Pugsley (2014)), less is known about whether measurement error is constant across waves of a given panel. If this is not the case, then estimates of distributional changes and income transitions will confound true changes with the effects of changing measurement error and will therefore be biased. In this paper we provide evidence from a quasi-experiment that the quality of measured income systematically differs across the early waves of a leading panel survey and as such analysis of reported incomes from the initial waves of data will suffer from bias.

There are two reasons to think that reported income may not be comparable across waves of a panel, particularly in the initial waves. First, panel conditioning (PC) effects may be present. PC refers to the idea that respondents learn from their previous interview experience and their willingness to reveal personal information at future interviews depends on this learning. PC could take place through building trust with the interviewer or data holders; improved comprehension of the questionnaire; or giving strategic responses to reduce the length of the interview. Crossley et al. (forthcoming) show that repeated interviewing can also lead to real changes in (savings) behaviour. Second, dependent interviewing (DI) - a tool that reminds survey respondents of their reports at the previous interview - whilst helping to reduce spurious change between waves, will lead to differences in data quality between the baseline interview and subsequent interviews where DI takes effect.[1]

[1]Jenkins (2011) notes that when DI was introduced in the British Household Panel Survey there were "no obvious discontinuities in income series". A third reason is non-random attrition but we address this

Only a few existing studies have examined the stability of measurement error in income across waves of a panel and this likely reflects that it is often not possible to link survey data to longitudinal administrative income records.[2] David and Bollinger (2005), as part of a small scale validation study, find that false negative reporting of US food stamps is highly correlated across wave one and wave two of the Survey of Income and Program Participation suggesting that respondents have a latent tendency to cooperate (or not cooperate) with the survey. Das, Toepoel, and van Soest (2011) discuss an alternative methodological approach to estimating PC, based on comparing responses from first-time responders in refreshment samples to more experienced panel members and making assumptions on the attrition process.[3] 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 to answer questions about their income. Similarly, Frick et al. (2006) find that experienced panel members report higher income in the German Socio-Economic Panel and they conclude that the differences are driven by changes in response behaviour of new panel members and not attrition. Despite these incidental findings and centrality of income to economics, we know of no study that has performed a systematic analysis of how measurement error in household income and its components evolve across waves of a panel.

In this study we provide causal evidence on the comparability of reported income across the initial waves of a large general purpose panel survey: the UK Household Longitudinal Study. The novelty of our approach is that it does not require data linkage or refreshment samples but exploits two features of the survey design as a quasi-experiment to separate changes in reported income due to panel conditioning and dependent interviewing from real income changes that evolve over time. First, we remove the time effects by exploiting the fact that the fieldwork period for adjacent waves overlaps by one year, giving us random

directly in the identification section.

[2]Hurst, Li, and Pugsley (2014) estimate that the self-employed under-report income by about 25 percent in the Panel Study of Income Dynamics. Because of under-reporting to tax authorities, linked administrative records would not help in estimating how measurement error changes across waves for the self-employed.

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

samples of individuals being interviewed at different waves of the panel but in the same calendar year (accounting for attrition). Second, we use the fact that UKHLS uses reactive DI meaning that we can observe exactly which individuals would have failed to report an income source in the absence of DI.

Our approach offers several 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. (2012a); David and Bollinger (2005)). First, the large sample sizes available mean we can precisely estimate the effects of interest even when they are small in magnitude. Second, our data is 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 enabling 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 and that the improvements are strongest for unearned income sources (largely pensions and state benefits) and take effect in the initial waves of the panel. Being interviewed for a second time, relative to the first, increases reported monthly income by £124 or 7.8 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, 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, which is also backed up with an examination of refusal rates on the income variables which fall off most sharply between waves one and two.

Finally, on the broader implications of our results: we present evidence suggestive of similar effects in another leading panel survey (British Household Panel Survey) and more generally they could be expected to extend to other sensitive areas of data collection. 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.

The paper proceeds as follows: the next section describes the data and compares our estimates of the UK income distribution to those from official cross-sectional sources. Section 3 discuss the identification strategy. The empirical results are presented in section 4 and section 5 concludes.

2 Data

This paper makes use of data from the UK Household Longitudinal Study (UKHLS) that began in 2009. UKHLS is a large general purpose social survey that collects information on a range of outcomes including: income, education, aspirations, health, happiness, household organisation, housing, tenancy, geography, time-use, relationships and objectively measured bio-markers. It is the main UK longitudinal data source of income and it will replace the former BHPS as the data source for official UK Government statistics on poverty dynamics. Of relevance to this paper are the large sample sizes (billed as 'the largest panel survey in the world') and the questions on individual income receipt collecting information on earnings, investment income, and unearned income sources (discussed below).

UKHLS is an individual level survey and all adult members (age 16 or over) in households selected to be a part of the panel in wave one form permanent sample members and are interviewed annually, aswell as non-permanent (temporary) sample members who may have become co-resident with a sample member.[4] UKHLS has a somewhat complex sample design consisting of multiple sub-samples. The identification strategy of this paper exploits features of the data collection for the main 'General Population Sample Great Britain'

[4]Children born to women who are permanent sample members become permanent sample members. A household questionnaire is also completed by one member of the household and each adult is asked to complete an additional self-completion questionnaire. Household members aged 10-15 years are asked to complete a short self-completion youth questionnaire.

sub-sample and 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.[5]

The survey design of UKHLS is unusual in that, due to the large sample sizes, the fieldwork takes place over a 2-year period with an overlap between waves for the GPGB sample that forms the focus of our paper. Identification of our main results exploits the overlap in waves and the random allocation of participants across the survey period. In particular, the sample selection procedure for the GPGB sample is a proportionately stratified, equal probability (clustered) sample. In wave 1, the selected sampling units (postcode sectors[6]) were randomly allocated across the 24 interview months of the survey. Households are then selected from each postcode sector using systematic random sampling. All persons resident at a household form permanent sample members. To quote Lynn (2009) 'Each monthly sample will therefore be a representative random sample of the total population, as will any amalgamation of months, such as quarters and years.' Each monthly sample is issued to the field again 12 months from the initial allocation and every 12 months subsequently.[7]

2.1 Income data and dependent interviewing

Although a general purpose survey, a specific goal of the study is to provide reliable data for the measurement of income dynamics and so a sizeable amount of questionnaire time is devoted to income data collection from each sample member. Sources covered by the survey are earnings from main and second jobs, self-employment income, social security benefits, state and private pensions, private transfers and investment income.

The main income concept we work with in our analysis is current monthly gross individ-

[5]Response rates were lower for: those under 30, in urban locations, being a renter and expecting to move home at wave 1 (Lynn et al. (2012b)).

[6]Postcodes are codes referring to a group of UK postal address. All UK postcodes are stored in a national database known as the Postcode Address File, from which primary sampling units are selected.

[7]It is the issue date that is randomly assigned and not the actual date of interview. The two may differ by some months in order to maximise the chances of an individual response.

ual income before taxes, deductions and national insurance contributions. Data collection occurs for 3 distinct subcomponents which we analyse separately: i.) earnings, ii.) unearned income (39 sources) and iii.) investment income. In our analysis, we further decompose ii) into social security benefits, pensions and other unearned income sources.

The data collection of ii) makes use of dependent interviewing which is a tool that reminds a survey participant of a source they reported receipt of at the previous wave when failing to report it at the present wave.[8] The aim of DI is to reduce spurious change in reports between waves and thus improve data quality. The survey records whether a DI reminder was given for each relevant income stream and it is therefore possible to remove the effects of DI by setting to zero any source for which an individual received a dependent interviewing reminder (i.e. in the absence of DI, the source would have gone unreported). Appendix B contains a list of the income questions from which the income variables are constructed, alongside the full DI question.

2.2 Missing Data and false negative reporting of unearned income

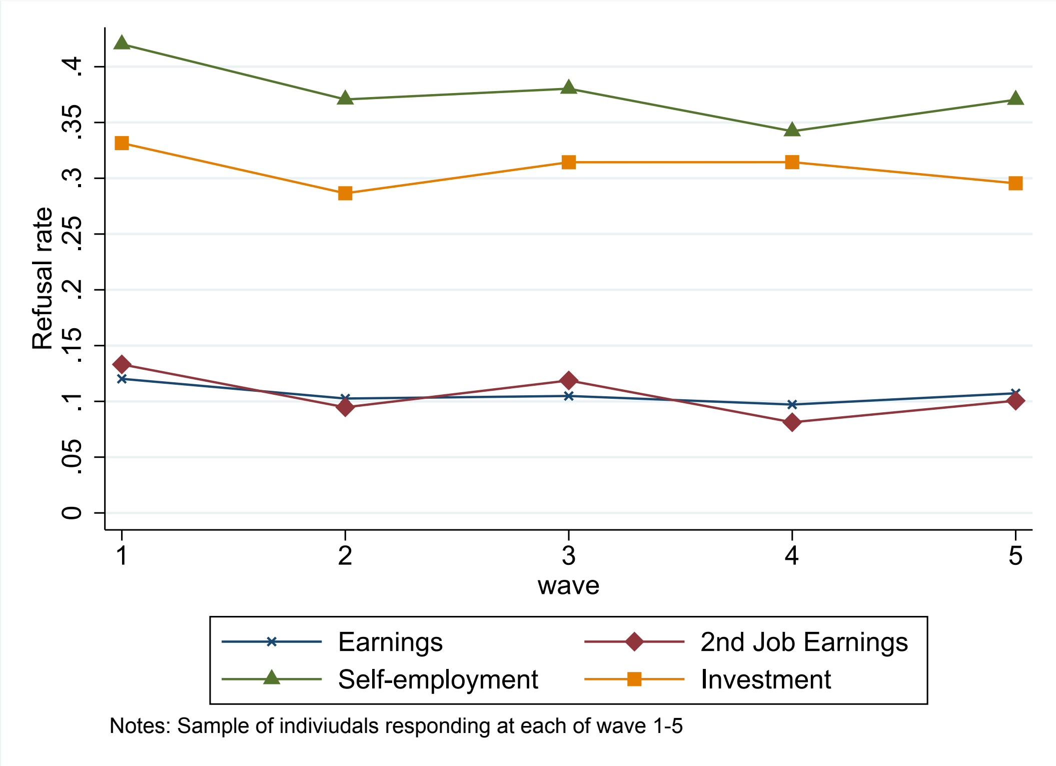
Income is a sensitive area of questioning with associated issues around privacy and trust[9] and consequently item non-response rates for income variables are typically high in comparison to other variables collected in surveys. As part of the standard UKHLS data release, missing values are filled by imputation. If respondents' trust concerns lessen with repeated interviewing, then item non-response rates would fall as a panel ages and the survey would provide better coverage of income sources over time. It is therefore important to document item non-response rates across waves of the panel.

Figure 1 plots trends in refusal rates seperately for earnings from main and second

[8] Survey methodologists refer to this as reactive DI, in contrast to proactive DI where all respondents are reminded of their previous wave response before answering the present wave question.

[9] Perhaps best evidenced by recent discussions in the media around the panama papers leak. To quote the Guardian newspaper 'earnings......being up there with their sexual orientations and religious beliefs, as matters strictly between them and their gods' (http://www.theguardian.com/news/2016/apr/05/david-cameron-tax-affairs-should-be-no-private-matter).

Figure 1: Income refusal rates by wave



job, self-employment profit and income from investments, for a sample of respondents who completed a full-interview at each of waves one (2009-10) to five (2014-15).[10] The refusal rates fall across all sources as the panel ages with the biggest drop occurring 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, non-response for self-employment profit falls from 42.0 per cent to 37.1 per cent and the earnings refusal rate from 12.0 per cent to 10.3 per cent. These patterns are consistent with a panel conditioning effect where second-time respondents change their reporting behaviour due to prior experience with the panel. The fall in the refusal rates implies improvements in income data quality over time as the quantity of observations that have to be estimated by imputation would fall; although, in contrast to unearned income below, if imputation works well we would not expect it to lead to big shifts in mean measurement error.

[10] Refusals are counted as 'refusals' + 'don't knows' where a 'don't know' could be a polite refusal.

For unearned income, respondents are sequentially presented with lists of transfer payments and are asked to indicate which they receive. The refusal rates rates for unearned income are low and show the same falling pattern over time. For the balanced waves 1-5 sample and the primary list of unearned sources they are: 0.02, 0.02, 0.01, 0.001, and 00.3. For a respondent giving a refusal, it is assumed the source is not received and so refusals contribute to false negative reporting, which can be extensive in household surveys (see for example Meyer and Sullivan (2011)).

The reported distribution of unearned income will be particularly sensitive to improvements in respondents willingness to reveal personal information, as sources previously not counted in income totals (the false negatives) get covered by the survey and become included. In contrast to the items covered in figure 1, imputation would not dampen the effects of differences in false negative reporting across waves of a panel, as it is not possible to identify false negative reporting.[11]

In our main analysis, we replace missing values with the standard longitudinal imputes released by the data providers. As imputation may hide some of the effects of changing item non-response on estimates of the income distribution, we separately present results when missing values are imputed to zero. Finally, there can be two other types of missing income data that could be sources of income differences across waves: i.) missing an individual interview (unit non-response) and ii.) missing an individual interview but agrees that a proxy answers a shorter interview on their behalf. We address these issues directly in the identification section.

2.3 Comparison of UKHLS to the cross-sectional 'gold standard'

We compare cross-sectional estimates of selected quantiles of the income distribution from UKHLS and the source for UK official statistics on the income distribution, the Family Resource Survey (FRS), in order to: a) confirm the quality of UKHLS income data against

[11] Refusals are not an explicit option given to respondents and so explicitly refusing may indeed feel confrontational or unhelpful and so giving a false negative may be preferred. False negatives are of course not possible on earnings variables.

the cross-sectional 'gold standard' and b) assess changes in the UKHLS income distribution over time relative to a baseline survey, which would be indicative of changes in the reporting behaviour of respondents.[12] While we cannot directly attribute b) to changes in reporting behaviour, where reporting changes do occur, they should show-up therefore making the comparison meaningful.

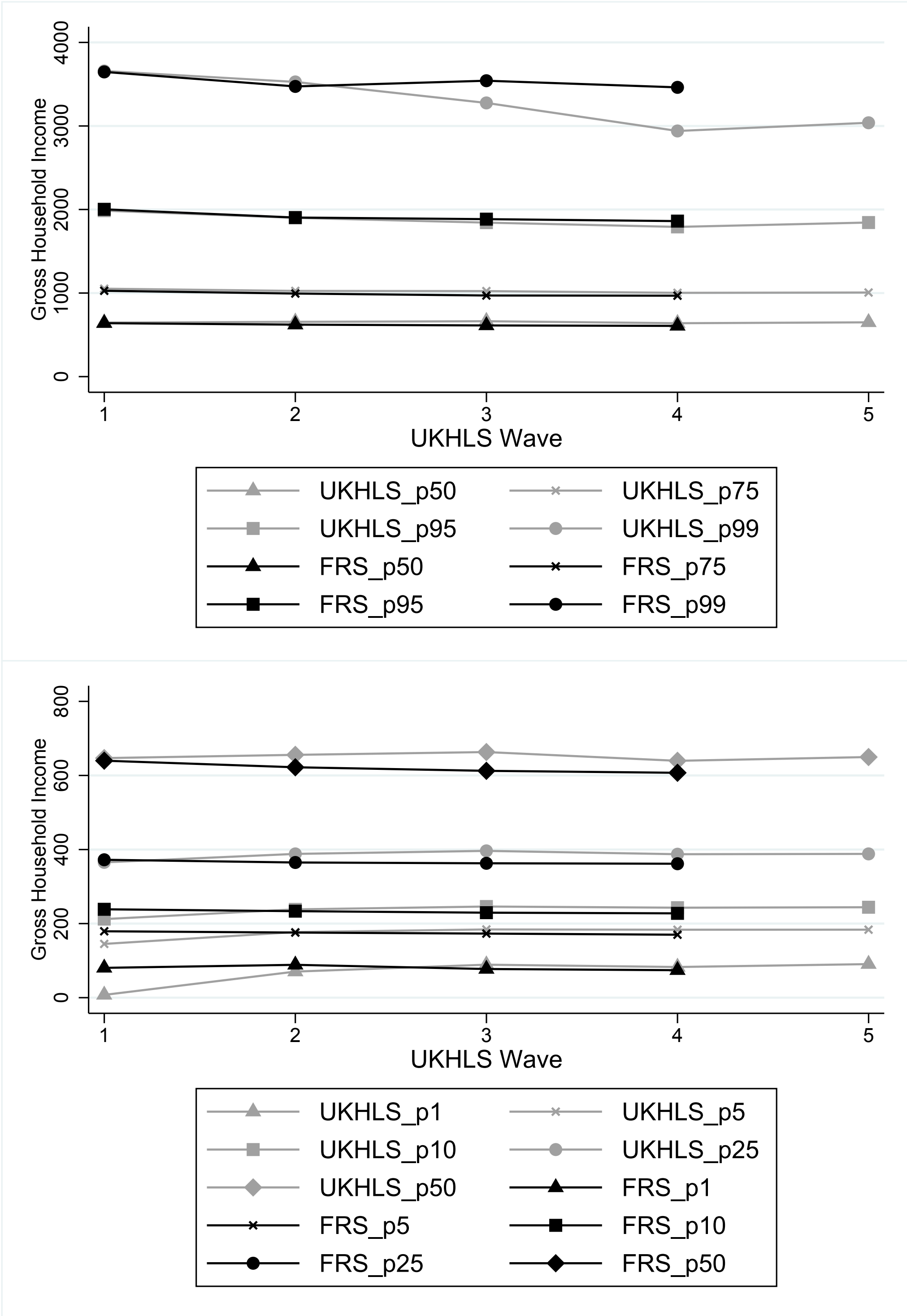
Estimates of selected quantiles of the income distribution from the two surveys are shown in figure 2. All figures are expressed in 2010 terms using the 'all items rpi excluding council tax' monthly price index, which is used in official UK income statistics, and produced by the Office for National Statistics. The top half of the figure refers to quantiles at the median and above and the lower panel to quantiles at the median and below. The FRS corresponds to a financial year (April to April) and a UKHLS wave to two calendar years. To account for differences in the fieldwork period of the two surveys, we pool two consecutive FRS data sets when comparing to a single UKHLS wave. The small remaining differences in coverage of the two surveys should cause only minor differences in the corresponding distributions.

The top half of the panel shows the median, 75th, 95th and 99th percentiles. The two distributions line-up remarkably closely and moreover the difference between the two surveys remains small and stable over time. The one exception is for the 99th percentile where the two surveys are diverging from wave 3 onwards and this likely reflects the known difficulties of measuring the very highest incomes in household surveys (see for example Bricker et al. (2015)).[13] Turning to the lower panel, the story is somewhat different. Whilst there is a clear similarity between the estimates the difference between them is changing over-time, most strongly between waves 1 and 2 and to a lesser extent between waves 2 and 3. For example, at wave 1 the FRS gives higher estimates of incomes for the 1st, 5th,

[12]The FRS is a purpose built income survey that collects information from a random sample of approximately 20,000 households each year. Our analysis is based on the 'Households Below Average Income' data sets which are produced by the Department for Work and Pensions as the basis for official UK statistics on the income distribution.

[13]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.

Figure 2: Selected quantiles of UKHLS/FRS (2009/10-2013/14)



Notes: Each FRS datapoint calculated from pooling 2 consecutive FRS surveys and using the HBAI datasets.

10th and 25th percentiles but by wave 3 the pattern has reversed. Given the strongest divergence occurs between wave one and two, we concentrate the most detailed part of our analysis on reporting behaviour at these waves.

3 Identification strategy

We are first interested to know whether respondents to the survey show any difference in income reporting behaviour between the first and second interviews and we would like to decompose the difference into PC and DI. We then wish to extend our analysis to other waves with a view to assessing how measurement error evolves over later waves of the panel. For presentational purposes, we focus on differences in the mean as a summary measure but our approach could be extended to other distributional measures. A naive comparison of the wave one and wave two income distributions would confound changes in reporting with real changes in individual incomes (time effect) and the compositional differences across the waves of the panel (attrition effect). We separate out the reporting effect from the time and attrition effects using the fact that the fieldwork period for adjacent survey waves overlaps by one year and that any subset of months forms a representative sample of the GB population, once attrition differences are accounted for.

Specifically, we construct a sample of individuals allocated for interview in 2010 but who are responding at different waves of the panel and attribute any difference between the mean of the two groups to response behaviour. Using a treatment/control terminology, the wave one 2010 sub-sample forms the control group (interviewed for the first time) and a different group of individuals responding to wave two in 2010 (interviewed for the second time) forms the treated group. With non-random attrition, there will be compositional differences between the groups that would bias our comparison. We remove the differences by restricting the analysis to a balanced sample of respondents who completed a full-interview at both waves one and two, under the assumption that the wave 2 interview outcome is statistically independent of the wave one survey year allocation, an assumption we return to below.[14] We then perform ordinary least squares regressions of our income components of interest (earnings, benefits and unearned income and investment income) on an indicator for being a wave 2 respondent, with the coefficient on this indicator giving our estimate of the wave 2 reporting difference on the mean.

We estimate separately the share of the treatment effect which is due to PC and DI, for income sources where the survey made use of DI (39 state benefits and unearned income sources). DI reminders were only given if a respondent failed to report an unearned income source which they received at the previous interview. We therefore estimate the reporting effect net of DI by setting to zero any income source for which an individual received a DI reminder and then re-estimating our main coefficient of interest. The initial reports of all income sources were collected prior to the DI reminders being given, meaning that there is no concern that a first DI reminder would affect the reporting of subsequent income sources during an interview.

There are three violations of our identification strategy that would lead to compositional differences between the treatment and control groups and so bias our results. First, population changes across 2009/10 leading to differences in the make-up of the household population across the initial wave of the panel. Second, the ageing effect of the panel resulting in individuals responding for the second time in 2010 being one year older than those responding for the first time. Third, a time effect in attrition making the decision to attrite statistically dependent on the initial survey year allocation. To credibly address these concerns we include a wide range of control variables in our regression models. including controls for age. The fact that UKHLS is a general purpose survey covering a large number of topics - including determinants of income such as labour market behaviours, retirement status, demographics and household composition - makes this strategy persuasive. Conditional on the controls, it is assumed that allocation to the treatment and control groups is randomly assigned. Given that the control variables are potentially also subject

[14]This implies that new entrants to the survey at wave 2 who are being interviewed for the first time are excluded from the analysis, alongside individuals who had a proxy interview at either wave one or two.

to panel conditioning, we focus on controls with low item non-response rates and that we judge unlikely to be sensitive areas of questioning. The full list controls is given in the footnote to figure 3. Formally, we estimate:

$$Y_i = \alpha + \beta_1 wave\_2_i + \beta_3 X_i + \epsilon_i \tag{1}$$

where Y_i is the income component of interest, $wave\_2_i$ an indicator variable taking the value 1 for wave 2 respondents, X_i a vector of controls and ϵ_i an error term with $E[\epsilon_i|wave\_2, X] = 0$. β_1 is our coefficient of interest, estimates of which are presented in the results. We report standard errors robust to heteroskedasticity. We also tried clustering standard errors at the level of the Primary Sampling Unit but it made little difference to the estimated standard errors.

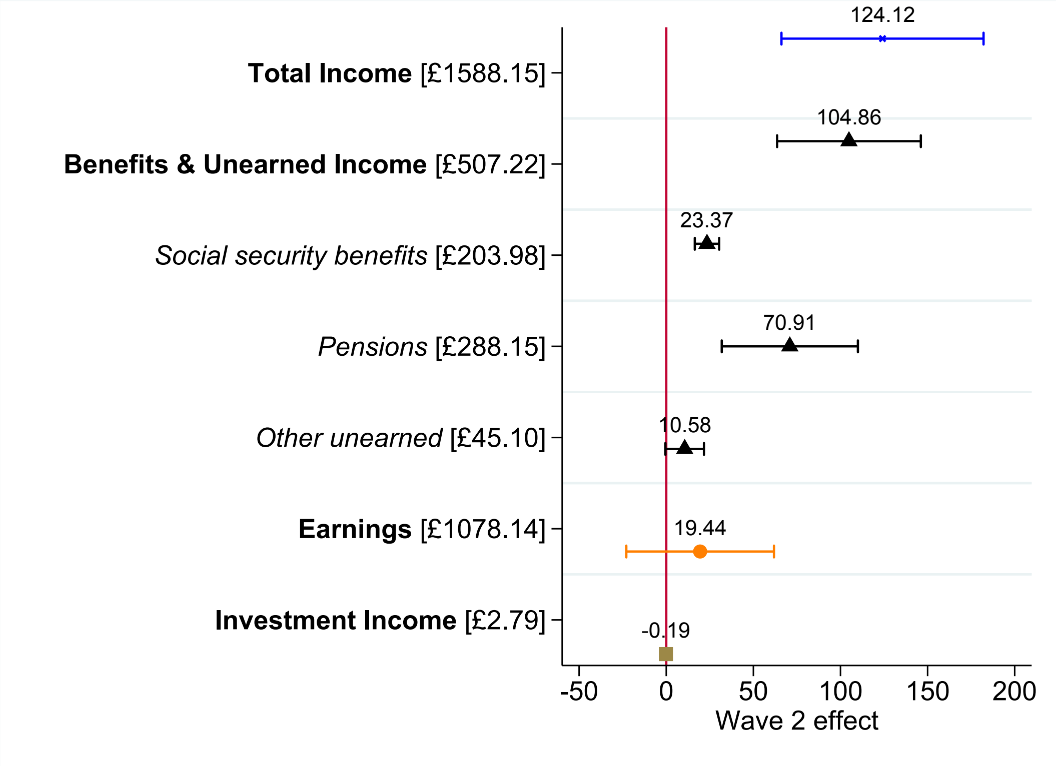
Table 1 presents evidence on the validity of our identification strategy by comparing sample means of our treatment and control groups in their baseline (wave one) interview to assess differences in composition. The large sample sizes make it possible to detect even small differences where they occur. On demographics, the two subsamples are balanced in terms of sex, age, ethnicity (half a percentage point differences in the share of Indian and Chinese), qualifications and marital status (although nearly 1ppt difference in the share that are single and never married). For household composition, we see no differences in the mean number employed, of working age, or number of couples or single parents in a household. There are small differences in the mean number of people, children, and the age of the youngest child. Larger statistically significant differences are seen for living in social housing (1.6ppt) and having children (1.9 ppt) but not for the shares that own, rent or mortgage their home or number of bedrooms. Overall we interpret the comparison as indicating that the samples are well balanced and where there are small differences, as stated above, they are accounted for in our regression models.

Table 1: Comparison of baseline (wave one) characteristics by year of interview

	2010	2009	Mean Diff	SE	N_2010	N_2009
sex	0.4292	0.4281	-0.0011	0.0057	14731	14912
age	48.4870	48.4485	-0.0385	0.2068	14731	14912
ethnicity:						
white	0.9233	0.9283	0.0050	0.0030	14720	14899
mixed	0.0092	0.0095	0.0003	0.0011	14720	14899
indian and chinese	0.0221	0.0171	-0.0050**	0.0016	14720	14899
other asian	0.0195	0.0203	0.0008	0.0016	14720	14899
african or black caribean	0.0186	0.0193	0.0007	0.0016	14720	14899
other	0.0073	0.0054	-0.0018*	0.0009	14720	14899
highest qualification:						
Degree	0.2119	0.2102	-0.0018	0.0047	14726	14906
Other higher degree	0.1203	0.1200	-0.0003	0.0038	14726	14906
A-level	0.1861	0.1829	-0.0032	0.0045	14726	14906
GCSE	0.2124	0.2076	-0.0048	0.0047	14726	14906
other	0.1051	0.1121	0.0070	0.0036	14726	14906
no qualification	0.1641	0.1672	0.0030	0.0043	14726	14906
marital status:						
married or civil partnership	0.5294	0.5315	0.0020	0.0058	14725	14911
cohabiting	0.1229	0.1241	0.0013	0.0038	14725	14911
single and never married	0.1883	0.1784	-0.0099*	0.0045	14725	14911
divorced or separated	0.0932	0.0976	0.0043	0.0034	14725	14911
widowed	0.0662	0.0684	0.0022	0.0029	14725	14911
economic status:						
self-employed	0.0744	0.0665	-0.0079**	0.0030	14729	14911
employed	0.4627	0.4787	0.0160**	0.0058	14729	14911
unemployed	0.0562	0.0496	-0.0067*	0.0026	14729	14911
retired	0.2433	0.2417	-0.0016	0.0050	14729	14911
student	0.0473	0.0493	0.0020	0.0025	14729	14911
long-term sick or disabled	0.0356	0.0388	0.0033	0.0022	14729	14911
other	0.0806	0.0754	-0.0052	0.0031	14729	14911
usual weekly hours worked	35.6694	35.8480	0.1786	0.2312	7067	7430
long-standing illness or impairment	0.3884	0.3910	0.0026	0.0057	14712	14901
SF-12 Physical Component Summary	49.4445	49.1819	-0.2627*	0.1336	14652	14849
SF-12 Mental Component Summary	50.7113	50.8040	0.0928	0.1155	14652	14849
household						
tenure:						
owned	0.3174	0.3270	0.0096	0.0054	14712	14879
mortgage	0.3881	0.3961	0.0079	0.0057	14712	14879
rent	0.1192	0.1167	-0.0024	0.0038	14712	14879
social housing	0.1722	0.1565	-0.0157***	0.0043	14712	14879
# bedrooms	2.9264	2.9256	-0.0008	0.0115	14722	14902
# people	2.7206	2.6762	-0.0444**	0.0155	14731	14912
any children	0.3269	0.3076	-0.0193***	0.0054	14731	14903
# children	1.7317	1.7315	-0.0002	0.0178	4815	4584
age of youngest child	6.6287	6.3709	-0.2578*	0.1017	4815	4584
# employed	1.2321	1.2543	0.0222	0.0121	14731	14912
# working age	1.7032	1.7004	-0.0028	0.0143	14731	14912
# couples	0.7184	0.7209	0.0025	0.0055	14731	14912
# single parents	0.0650	0.0612	-0.0039	0.0029	14731	14912

Notes: *p < 0.05, **p < 0.01, ***p < 0.001. Balanced sample of respondents who complete a full-interview at wave 1 and wave 2.

Figure 3: Effect of second interview on reported income



Notes: The point estimates presented correspond to β_1 from equation 1 with the full-set of controls. Confidence intervals are calculated using robust standard errors. The controls are dummy variables for (number of categories in parenthesis): sex, age (7), 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).

4 Results

4.1 Differences in reporting at waves 1 and 2

4.1.1 Panel conditioning and dependent interviewing effects

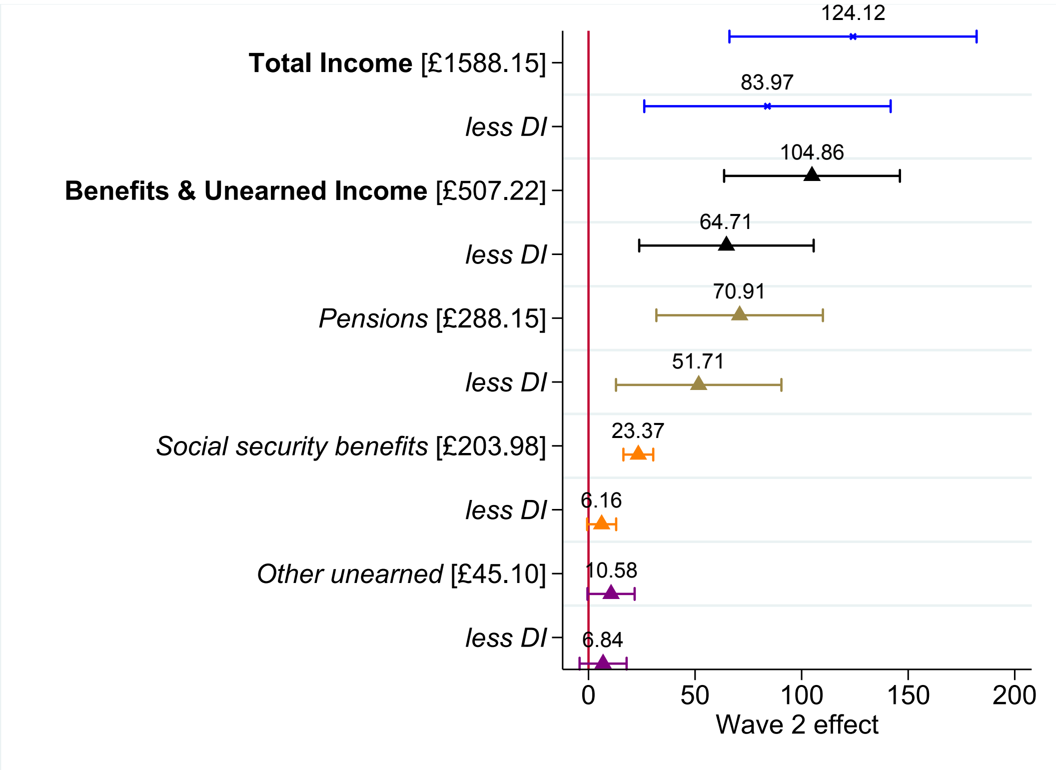
Figure 3 presents estimates of β_1 from equation (1). They can be interpreted as the causal effect (DI + PC) of being interviewed for a second time relative to a first on reported income. The figure shows results from models estimated separately for total: income, benefits and unearned income (and separately for the subcategories social security benefits, pensions, other earned income), earnings, and investment income. Means from the baseline (wave 1) interview are reported in square brackets.

Second-time responders reported a total monthly income that was £124.12 or 7.8 percent higher than first-time reporters and this represents a causal effect of being interviewed for a second time. We then decompose the effect into its sub-components. The effect for earnings and investment income are small and highly insignificant, whereas, we see strong effects concentrated in benefits and unearned income. Decomposing this category further, reporting behaviour changes are positive in all of its sub-components with the strongest effects occurring for social security benefits and pensions, where we see a statistically significant increase in reported income of £23.37 (11.5 percent) and £70.91 (24.6 percent), respectively. 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 difference would give a highly misleading picture of changes in the income distribution across waves of the panel.

We would like to examine the extent to which the change in reporting behaviour of wave two respondents is due to panel conditioning. Figure 4 presents results from re-estimating equation (1) but by setting a reported wave 2 amount to zero where a dependent interviewing reminder was given ie. in the absence of the DI the source would have gone unreported. As expected, once the DI effects are removed the estimates fall in size but surprisingly they remain large and statistically significant. For example, the wave 2 effect on total income falls by around a third from £124.12 to £83.97, suggesting a considerable panel conditioning effect remains. That is, wave 2 respondents report a total income which is on average 5.3 percent higher per month and this effect represents a change in reporting behaviour not due to dependent interviewing.

We also explored the possibility of heterogenous treatment effects by estimating models separately for subsamples of: pensioners, working age with children and working age without children. 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 percent. Moreover, 85 percent of this reported increase is due to

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



Notes: see figure 3 notes.

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 37 percent of the wave 1 mean and 59 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' (8 percent of the wave 1 mean). The interested reader can find the full figures of results in appendix A.

In summary, we observe being interviewed for a second time, relative to a first, causally leads to respondents reporting a monthly income a considerable 7.8 percent higher. This finding lines up with the validation exercise reported in section 2.3 which compared the UKHLS income distribution to a cross-sectional gold standard. The effects we observe are largely driven by reporting of benefits and unearned income sources and in particular pensions. Around 1/3 of the total effect is attributed to the use of dependent interviewing, and the remaining 2/3 to panel conditioning effects. We return to the mechanism through which these panel conditioning effects operate in section 4.3.

4.1.2 Receipt vs. Amount

An important question is whether the differences in reporting of benefits and unearned income across waves 1 and 2 of the panel are due to changes in reporting of receipt or changes in the amounts reported. Table 2 explores this matter by presenting estimates of equation (1) separately for each of 12 of the most widely received benefits and unearned income sources in the data. These are organised in the table according to pensions, family benefits, disability benefits and low income benefits. Columns 1 and 2 refer to receipt, and 3 and 4 to the reported amounts. The odd columns show the total effect (PC + DI), and the even, the panel conditioning effects only.

We observe increases in reporting of all of the 12 unearned income sources as a result of being interviewed for a second time. All of the effects are statistically significant with the exception of income support and the magnitudes of the effects are non-trivial. For example, the effect for state pensions is to increase reporting by 1.33 percentage points or 5.4 percent of the wave 1 mean. Column (2) shows that a sizeable share of the observed pattern is attributable to panel conditioning and the panel conditioning effects are concentrated in disability benefits and pensions, although the effect for Working Tax Credit is also statistically significant. For example, the panel conditioning effect for the state pension is 0.48 percentage points or 1.94 percent of the wave 1 mean.

Interestingly, these results are in contrast with 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. One possibility is that a lack of statistical power in their study made it difficult to detect small reporting changes where they occurred.

Moving to the reported amounts in columns 3, we see that 11/12 of the estimated coefficients are statistically insignificant, with the exception of employer pensions. Column 4 confirms this finding when estimating the PC effect only. Put together, the results of this section tell us that it is the receipt of unearned sources that changes with panel experience

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

	Receipt		Amount	
	DI + PC	PC	DI + PC	PC
Pensions				
State pension [24.71, £471.75]	1.328***	0.481*	0.0236	0.425
	(0.219)	(0.228)	(4.366)	(4.412)
Private pension [6.07, £405.77]	1.834***	0.809**	22.46	43.79
	(0.260)	(0.251)	(55.48)	(61.90)
Employer pension [15.91, £762.73	1.511***	0.402	269.1**	286.9**
	(0.330)	(0.328)	(95.59)	(100.0)
Spouse employer pension [3.15, £523.14]	0.657***	0.311	218.1	252.1
	(0.171)	(0.167)	(166.2)	(185.0)
Family benefits				
Working Tax Credit [6.64, £198.72]	1.149***	0.603*	10.70	9.011
	(0.273)	(0.268)	(7.764)	(7.959)
Child benefit [23.52, £116.75]	0.796*	0.347	2.838	2.790
	(0.310)	(0.311)	(1.770)	(1.785)
Child tax credit [17.21, £235.77]	1.308***	0.508	1.843	3.345
	(0.332)	(0.332)	(5.056)	(5.128)
Disability benefits				
Incapacity benefit [2.99, £366.26]	0.787***	0.442*	6.118	6.401
	(0.189)	(0.184)	(9.249)	(9.580)
Disability living allowance [5.54, £266.22]	1.472***	0.842***	-26.57	-28.87*
	(0.251)	(0.244)	(13.67)	(14.06)
Low income benefits				
Income support [4.73, £295.73]	0.420	0.170	1.374	3.108
	(0.226)	(0.224)	(10.92)	(11.11)
Housing benefit [10.63, £305.51]	1.123***	-0.0305	4.315	3.321
	(0.283)	(0.281)	(6.709)	(6.999)
Council tax benefit [12.68, £83.28]	2.037***	0.456	6.146	5.928
	(0.326)	(0.321)	(5.157)	(5.546)

N = 30136
Notes: see figure 3 notes.

but with no indication of changes in the reported amounts. Combining the facts that i) state transfers are known to be under-reported in survey data and ii) item non-response rates drop-off at wave 2, suggests that the observed changes can be interpreted as improvements in data quality as the panel ages.

4.1.3 Effects of imputation

If the observed patterns in reporting behaviour are driven by differences across waves in the extent to which respondents are willing to reveal details of their income, then this raises the question of whether imputation of missing data masks the extent of reporting differences across waves. Indeed, section 2.2 documented falling item non-response rates

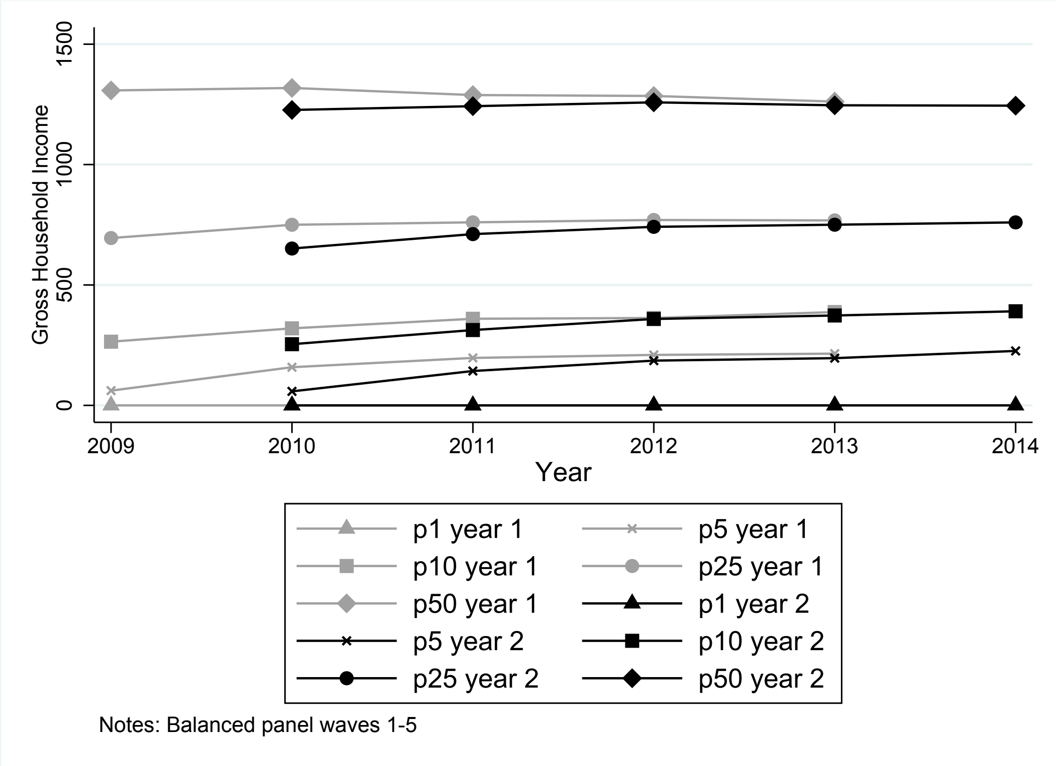
Figure 5: Effect sizes when setting missing amounts to zero

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.

across the waves of the panel, where the fall was most dramatic between waves 1-2. To explore this issue, we estimate equation 1 for income and its subcomponents, where we set to zero any source which has missing income information. In this way, changes in the extent of missingness would be reflected in our estimates.

Figure 5 presents the results (labelled imputes to zero) alongside the results from section 4.1 (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

Figure 6: Income trends by survey year



amounts. This can be explained by the observed effects for benefits and unearned income being driven by changes in reported receipt where some of the reporting increases are now not binding (where an amount is missing and now set to zero).

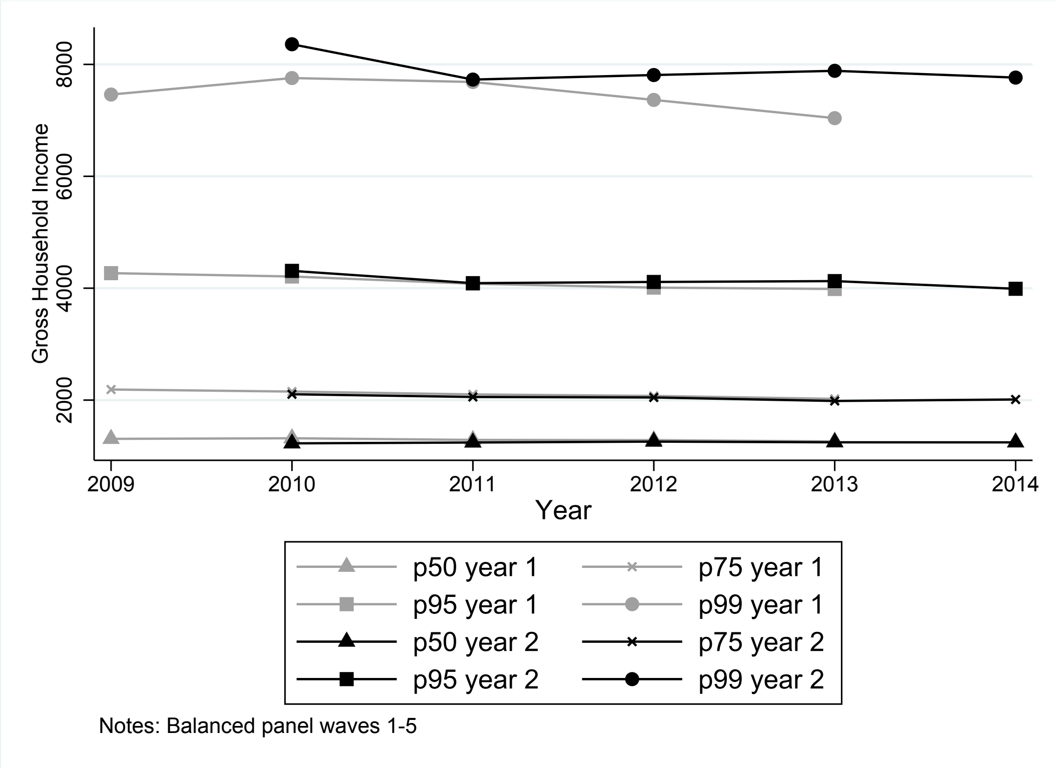
Overall, the results suggest imputation can work to correct data quality differences across the initial waves but only for amounts and not in case of false negative reporting where imputation cannot help.

4.2 Is reporting behaviour stable after the second interview?

We would like to know whether the reporting changes are focused only in the initial waves of the panel or extend further. The item non-response plots in section 2.2 show that while refusal rates on the income variables continued to fall as the panel aged, the largest drop-off was at wave 2.

To explore this matter further, figures 6 and 7 plot estimates of selected quantiles of gross income separately for the survey year 1 and survey year 2 subsamples with a view to

Figure 7: Income trends by survey year



investigating reporting differences between experienced and inexperienced panel members across the first five waves of the panel. Survey year 1 (2) respondents receive their first interview in 2009 (2010), their second interview in 2010 (2011) and so on. To remove any compositional differences in the series due to attrition, the two estimation samples are restricted to include only respondents who provide a full-interview at each of waves 1-5. In this way, each series provides estimates of the same population parameters and any difference in them can be attributed to the fact that the survey year 1 sample have one more year of experience of participating in the panel than do the survey year 2 sample.

Figure 6 shows the bottom half of the distribution showing the median, 1st, 5th, 10th, 25th and 50th percentiles. The two series line up closely as they should but with the exception of the early years of the panel where there are notable differences between the two. The survey year 1 sample gives higher estimates of income in both 2010 (waves 1 and 2) and to a lesser extent in 2011 (waves 2 and 3) across all of the quantiles in the figure but the differences almost completely disappear by 2012 (waves 3 and 4). This is suggestive of changes in reporting behaviour across waves, where prior exposure to a survey interview causes respondents to increase their reported income relative to the baseline interview.

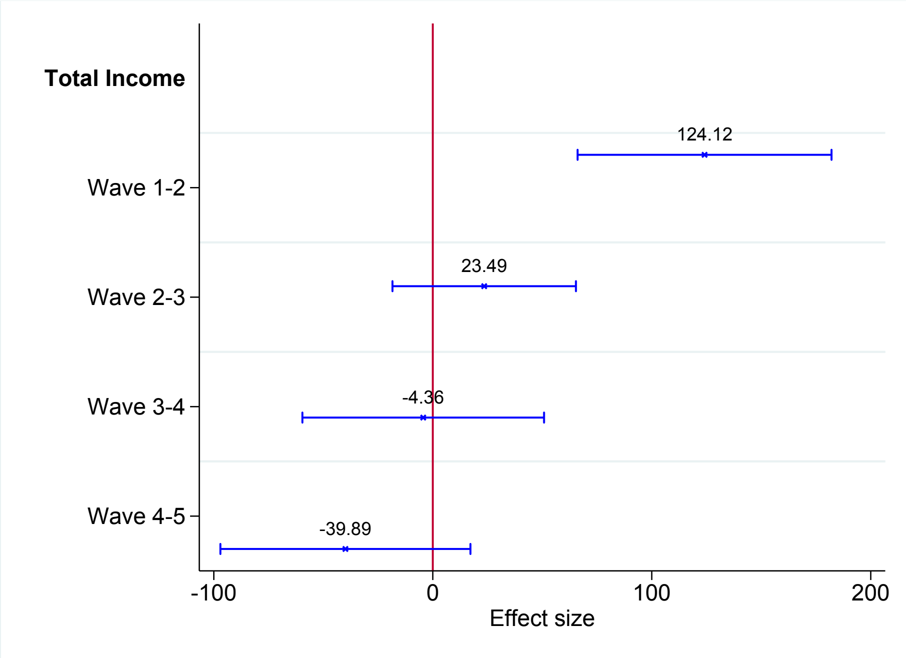
Figure 7 repeats the exercise for the upper half of the distribution showing the 50th, 75th, 95th and 99th percentiles. The estimates from the different subsamples line-up remarkably closely and so it would seem that any survey effects are confined to the bottom half of the distribution.

To provide formal estimates of mean changes in reporting differences across the later waves of the panel, we again use the over-lapping sample design of UKHLS as a quasi-experiment to provide estimates of income differences in time period t between wave w and wave w+1 (again restricting the analysis to a balanced sample of individuals completing a full interview in both wave w and w+1.) Figure 8 presents the results for the waves 1 and 2 (2010), waves 2 and 3 (2011), waves 3 and 4 (2012) and waves 4 and 5 (2013) pairs. Following the first wave pair, the estimated differences are small and statistically insignificant confirming that the changes in response behaviour occur for first-time respondents only. The result is consistent with the falls in item non-response documented in the earlier section and the graphical analysis above.

From the perspective of data users, it is reassuring that the effects are concentrated amongst first-time responders and that the main data quality improvements happen at the beginning of the life of the panel. So while analysis focussing on a short-panel from the early waves will suffer from bias, more longer-term longitudinal research based on multiple waves will be affected to lesser degree and researchers have the possibility of testing the sensitivity of their results to the inclusion of the wave 1 data. Distributional estimates based on the wave one cross-section will also suffer from underreporting problems (relative to later cross-sections), but insofar as wave one of a panel is a cross-section, analysis from cross-sectional surveys could be expected to suffer from similar amounts of under-reporting.

As to why these effects occur and why they are concentrated early on in the panel is the subject of the next section.

Figure 8: 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.

4.3 Explaining the panel conditioning effect

4.3.1 Respondent and interviewer behaviours

The observed effects raise the question of why respondents reported higher income (and lower item-non response) at the second interview given they faced the same questionnaire content as at wave 1. In this section, we explore explanations driven by the: a) interviewer and b) respondent. On a), a difference between a respondents first and second interiew is that i) respondents are familiar with the interviewers at wave 2, whereas at wave 1 they are a stranger and distinctly ii) interviewers accrue experience over the first wave of the panel and maybe able to better elicit responses at the wave 2 interviews. On b), respondents may i) have an improved comprehension of the complex interview when completing it for the second time or ii) have updated their beliefs about the trustworthiness of the data holders following a first successful interview and have less confidentiality concerns when sharing their sensitive (income) data. In the presence of even small doubts over confidentiality, respondents may be willing to misreport their income in a survey, even if the cost of lying is low (Hurst, Li, and Pugsley (2014)).[15]

While normally respondents are assigned the same interviewer at each wave, in 34.87 percent of cases in our estimation sample this was not achieved giving us variation with which to test the interviewer familiarity mechanism. Interviewer assignment may of course not be random and so a comparison of response behaviour between those that did and did not change interviewers would be misleading. To address this problem as best we can, we augment equation (1) with a dummy variable for whether a respondent had a different interviewer at wave 2 from wave 1 and its interaction with the wave 2 dummy. The dummy will capture time invariant differences betweent those that changed interviewer and not (eg. personality traits) and its interaction differences in reporting at wave 2 for those that changed interviewer. If familiarity with the interviewer is important at reducing the under-reporting of income then we would expect the interaction to be positive and significant and the main wave 2 effects to weaken. In that the interaction would be biased by any time varying factors associated with changing interviewer (eg. a location move associated with changing income and interviewer), the controls included in the regression model would work to reduce the bias. To test separately the interviewer experience mechanism, we add to the model a continuous variable for the number of interviews completed since the start of the survey, alongside controls for interviewer age and sex.

Table 3 presents the results with column 1 referring to our main specification for total income, and column 2 referring to models augmented with the interviewer controls. For total income in column 2, both the changed interviewer dummy and its interaction with the wave 2 dummy are highly insignificant indicating that the interviewer played a limited role in increasing the reporting of income. Moreover, our estimated main effects (wave 2 dummy) remain positive and statistically significant and are relatively stable. Turning to the other interviewer controls, again all are statistically insignificant including the

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

continuous interviewer experience variable. To strength evidence against the interviewer experience story, we point out that there are few differences in the 2010 wave 1 and 2 interviewer experience distributions (see appendix table A2). Overall, there is no evidence that interviewers played an important role in explaining our main results.

Table 3: Effect of changing interviewer on reported total income

	(1)	(2)
Wave 2	124.12***	115.48**
	(29.585)	(36.512)
Changed interviewer		5.69
[36.7]		(35.877)
Changed interviewer X Wave 2		27.16
		(66.601)
Interviewer:		
No. Interviews completed		-0.29
[65.13]		(0.234)
Female		1.62
[54.97]		(29.593)
Age		2.25
[57.80]		(1.642)
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).

In order to explore the respondent side of the relationship, we use information collected by the interviewers on respondent behaviours at the interview. First, to examine differences in respondent comprehension across waves, we exploit interviewer reports of how well the respondent understood the questions during the interview (on a 5 point scale). Second, to examine how confidentiality concerns may have lessened, we have available interviewer reports of whether a respondent was 'suspicious' about the study after the interview (3 point scale) and whether prior to the interviewer, the household respondent had questions about 'confidentiality' (binary variable).[16] As the later was recorded before the individual interviews took place, it should better reflect changes in the latent confidentiality concerns of the household, in contrast to the other two measures, which may reflect an interviewers interpretation of response behaviour during the interview. We estimate equation (1) for the 3 outcomes where we recoded the interviewer observations into binary indicators. We show that our results are insensitive to changes in the chosen thresholds (results available from the author on request). Full details of the questions and the construction of the interviewer observation variables are provided in appendix B.

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: 0.30, 0.12, 0.18, respectively.

Table 4 shows the results from our main specification for our three outcomes of interest. Sample means are reported in the footnote to the table. 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.

Having shown that repeated interviewing reduces the number of confidentiality queries, the final step in the causal chain is to show that confidentiality queries are related to respondent willingness to reveal income information. We do this using item non-response as a measure of willingness to reveal income and show that in a cross-section it is related to confidentiality concerns. We focus on the 'queries confidentiality' measure as it is measured

[16] The interviewer also recorded if the respondent had questions about the: purpose of the study, interview length, panel design, incentive/payment, other queries.

Table 5: Effect of confidentiality concerns on item non-response

	(1) Earnings	(2) 2nd Job	(3) Self-employment	(4) Investment	(5) Pensions
Respondent queries:					
purpose	0.04***	0.04*	-0.01	0.05***	0.00
	(0.006)	(0.020)	(0.024)	(0.009)	(0.002)
interview length	-0.00	0.02	0.01	0.02*	0.00
	(0.006)	(0.022)	(0.024)	(0.010)	(0.002)
panel design	-0.01	-0.02	-0.11*	-0.09***	0.00
	(0.016)	(0.057)	(0.052)	(0.023)	(0.006)
confidentiality	0.10***	0.09**	0.09***	0.11***	0.01**
	(0.008)	(0.028)	(0.028)	(0.011)	(0.003)
incentive/payment	0.00	0.01	-0.00	0.02	-0.00
	(0.013)	(0.044)	(0.051)	(0.022)	(0.005)
other query	0.04*	0.08	-0.01	0.08***	0.00
	(0.018)	(0.064)	(0.070)	(0.025)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes
Observations	19016	1795	2197	12863	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.

before the individual interviews and so there is no concern that it reflects item non-response during the interview. Table 5 presents the results and confirms that confidentiality concerns are predictors of item non-response in the wave 1 cross-section.

Put together, a plausible story that makes sense of these findings has to do with the fact 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, say with third party organisations. But following the first interview respondents learn that the data holders are reliable and that their data have not been shared. By the time the second interview comes, respondents have updated their beliefs about the trustworthiness of the survey organisation, and are so more open in revealing details of their personal finances.

4.3.2 An alternative explanation: fieldwork agency learning

Wave 1 of a panel survey is comparable to any wave of a repeated cross-section in terms of respondent behaviour, who are all new to the survey, but not in terms of interviewer/fieldwork agency behaviour (henceforth 'implementers'). Implementers of established cross-sectional surveys may have many years experience of conducting a survey, where as at wave 1 of a panel, implementers have no previous waves from which to draw on experience. Implementer learning therefore provides an appealing explanation for the observed panel conditioning effects. We point to two features of our analysis in order to argue against an implementer experience interpretation. To give further strength to this claim, we provide empirical evidence that implementer learning did not lead to improvements in data quality across wave 1 of the panel.

In order for interviewer experience 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. Second, substantial 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 (and interviewers) already had a full year of field experience (wave 1 2009) before the period of our analysis sample.

To provide direct evidence on implementer learning, we compare estimates of the income distribution from UKHLS wave 1 to the FRS and explore how the differences evolve over time. Specifically, we compare estimates of the income distribution from UKHLS wave 1 (2009 respondents) with the FRS 2009; and UKHLS wave 1 (2010 respondents) with the FRS 2010. If the problem is with implementers, rather than respondents, then the 2010 comparison should be more favourable. Table 6 shows the results from this comparison. Columns 2 and 3 shows estimated 2009 quantiles from the FRS and UKHLS, respectively, and column 4 shows their ratio, where a ratio greater than 1 indicates that UKHLS underestimates a quantile relative to FRS. Columns 5-7 repeats the analysis but for the 2010

Table 6: Comparison of UKHLS wave 1 to FRS by calendar year

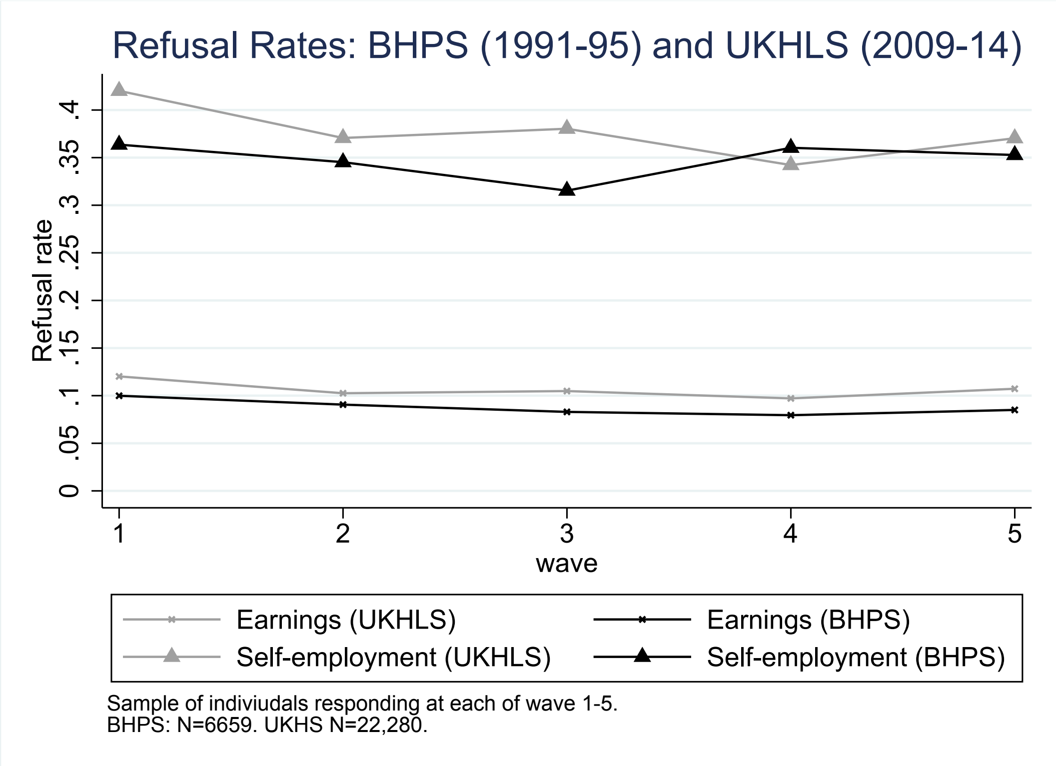
Percentile	FRS 2009	UKHLS 2009	Ratio (col 2/col 3)	FRS 2010	UKHLS 2010	Ratio (col 5/col 6)	Ratio of ratios (col 7/col 4)
5	182.88	148.42	1.23	176.06	137.88	1.28	1.04
10	242.50	218.88	1.11	234.90	209.36	1.12	1.01
15	291.20	273.33	1.07	280.51	259.88	1.08	1.01
20	335.77	327.17	1.03	324.01	308.02	1.05	1.02
25	378.67	378.96	1.00	366.06	358.72	1.02	1.02
30	425.31	429.94	0.99	413.74	413.38	1.00	1.01
35	476.60	488.20	0.98	464.39	461.97	1.01	1.03
40	531.56	545.31	0.97	514.72	518.36	0.99	1.02
45	592.33	602.23	0.98	568.10	575.13	0.99	1.00
50	652.92	664.10	0.98	628.87	633.98	0.99	1.01
55	715.58	731.41	0.98	689.40	706.85	0.98	1.00
60	784.84	806.25	0.97	758.85	768.96	0.99	1.01
65	865.89	884.49	0.98	832.81	845.78	0.98	1.01
70	949.64	973.64	0.98	912.83	928.51	0.98	1.01
75	1049.28	1079.53	0.97	1006.81	1037.49	0.97	1.00
80	1169.46	1208.55	0.97	1119.01	1156.70	0.97	1.00
85	1332.68	1369.94	0.97	1275.59	1317.01	0.97	1.00
90	1563.54	1609.03	0.97	1508.30	1564.63	0.96	0.99
95	2075.13	2023.11	1.03	1935.11	1966.40	0.98	0.96

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

(wave 1) calendar year.

In 2009, the UKHLS estimates typically match closely with the FRS ones but UKHLS misses income at the bottom of the distribution and most notably for percentiles 5, 10 and 15 where the ratios are 1.23, 1.11 and 1.07, respectively. Columns 5-7 repeat the analysis but for the 2010 year and a remarkably similar pattern emerges. In order to examine the stability of this comparison over-time, column 8 presents the ratio of ratios, where a value of less than 1 would indicate that UKHLS gets closer to the FRS in 2010 relative to 2009 and would be consistent with implementer learning. The ratio of ratios is always close to 1 indicating little change in the relative difference between the surveys over time. It reaches an absolute maximum of 1.04 for the first quantile, which if anything suggests that the coverage of UKHLS got worse relative to FRS in 2010 relative to 2009. We conclude that there is no evidence to suggest fieldwork agency learning lead to improvements in data quality across the first two years of the panel.

Figure 9: Comparison of UKHLS and BHPS item non-response rates by wave



4.4 Do the results generalise to other surveys?

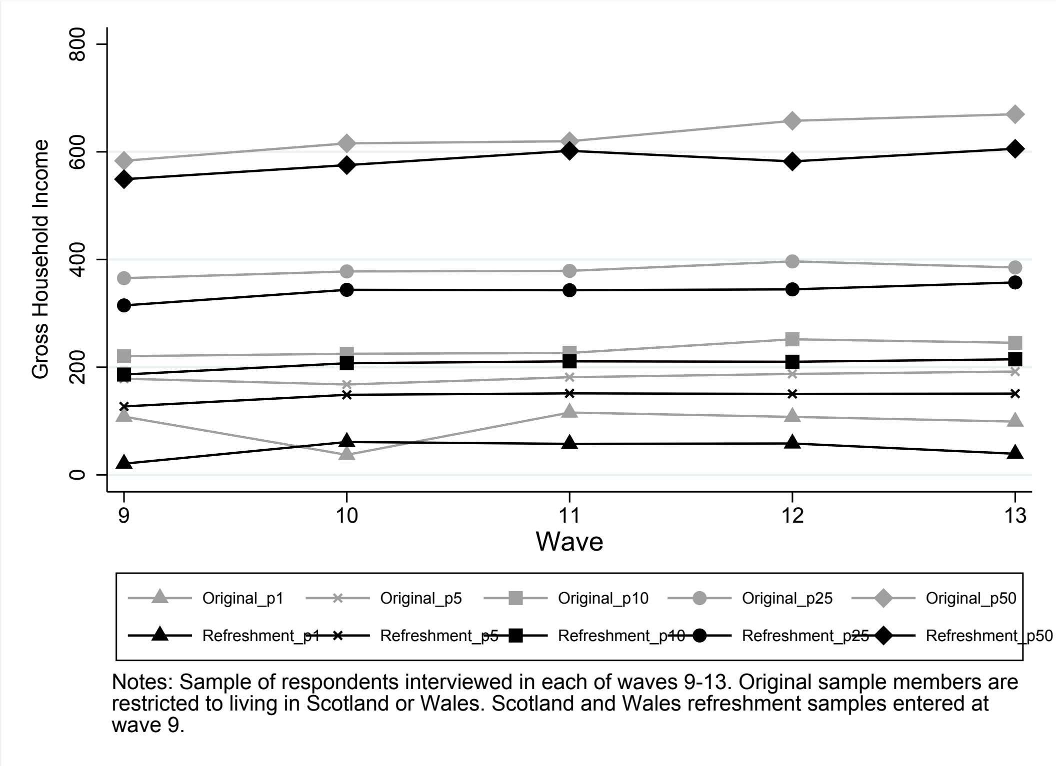
It is important to know whether the reporting pattern established is a peculiarity of UKHLS or a more general feature of income data collection. We turn to the predecessor of UKHLS - the British Household Panel Survey - for two reasons. First, some of the central income questions and features of its survey design (eg. survey instruments) are identical to those in UKHLS and if our main results are general, we would expect them to show up in a similar survey. Second, following the start of the survey in 1991, refreshment samples were added in 1999 and again in 2000 giving us new samples of first-time responders with which we can examine changes in reporting behaviour as their panel experience grows. A disadvantage relative to UKHLS is that the sample sizes are much smaller and so it makes it more difficult to observe effects where they occur. Moreover, dependent interviewing was not introduced in BHPS until 2006 and so the discussion in this section focuses on the effects of panel conditioning only.

For the main BHPS sample we have no quasi-experiment with which to separate income reporting changes from real effects.[17] However, we can compare trends in BHPS item non-response rates at the start of the panel to the trends observed in UKHLS. If the first line-up with the latter, then we conclude that BHPS sees similar improvements in income data quality as UKHLS as the panel ages. Figure 9 plots the wave 1-5 refusal rates for a sample of respondents who completed a full-interview at each of waves one to five of the corresponding survey. We focus on earnings from main job and self-employment profit as the questions are identical in both surveys. First, we observe that the level of non-response is similar in both surveys but the rates are relatively 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 (see Meyer and Sullivan (2015)). Following wave 1, we see a fall in item non-response rates in both surveys but the fall appears to be stronger in the UKHLS sample. 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.

The BHPS was extended to include a refreshment sample of 3000 households from Scotland and Wales in 1999 (wave 9) and 2000 households from Northern Ireland in 2001 (wave 11). Figure 10 plots selected income quantiles for a sample of respondents interviewed in each of waves 9-13 and living in Scotland or Wales, separately by whether they form part of the refreshment sample or were an original sample member. If the panel conditioning effects we observe in UKHLS extend to the BHPS, then we would expect that the

[17] The cross-sectional gold standard files begin in 1994 meaning that we cannot compare them to the early waves of the BHPS.

Figure 10: BHPS Scotland and Wales refreshment samples (wave 9)



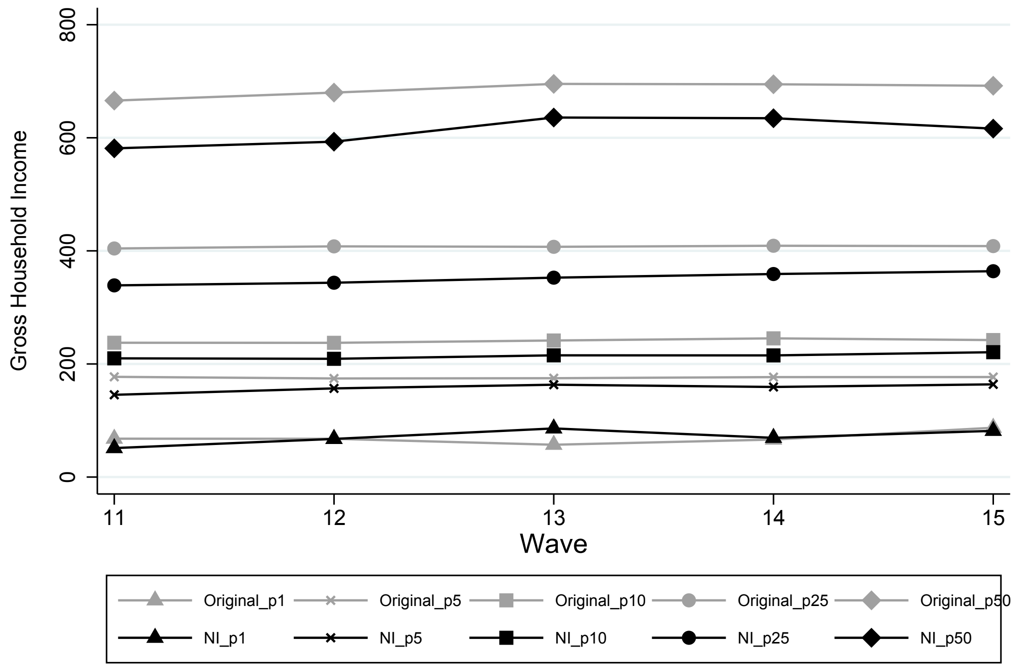
refreshment sample under-estimates the lower quantiles compared to the original sample, but that this amount lessens over time and particularly between waves 9 and 10 when the refreshment sample are being interviewed for the first and second time. The figure reveals trends consistent with that pattern. 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.[18]

Figure 11 shows results from a similar analysis for the 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.[19] 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

[18]The refreshment sample tends to give lower estimates compared to the original sample and this reflects compositional differences between the groups.

[19]NI not in HBAI until 2002/03

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



Notes: Sample of respondents interviewed in each of waves 11-15. Original sample members are living in England, Scotland or Wales. Northern Ireland sample entered at wave 11.

percentiles, presented, there is no obvious reporting difference and this fact could reflect underlying differences in the shape of the NI and UK income distributions.

5 Conclusions

We have shown 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 in 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. We also show that similar income data quality improvements are found in another leading panel survey, suggesting that our results may be generalisable to other surveys.

The use of income data from repeated survey measures is commonplace in social science research, including the use of large scale household panel surveys and purpose built panel surveys implemented as part of field experiments. Our results suggest that researchers working with data from the early years of panel data or with short panels be aware of this issue and should proceed with extreme caution when drawing firm conclusions from analysis based on such data. A more radical solution 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 source that may be under-reported.

Our work is suggestive that respondent confidentiality concerns play a role in the findings, 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

Brewer, Mike, Ben Etheridge, and Cormac O'Dea. forthcoming. "Why are households that report the lowest incomes so well-off?" *Economic Journal* .

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.).

Crossley, Thomas, Jochem de Bresser, Liam Delaney, and Joachim Winter. forthcoming. "Can survey participation alter household saving behavior?" *Economic Journal* .

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* .

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.

Jenkins, Stephen P. 2011. *Changing Fortunes: Income Mobility and Poverty Dynamics in Britain*. Oxford University Press.

Lynn, Peter. 2009. "Sample design for understanding society." Tech. rep., Understanding Society Working Paper Series: 2009-01.

Lynn, Peter, Annette Jäckle, Stephen P. Jenkins, and Emanuela Sala. 2012a. "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.

Lynn, Peter J, Jonathan Burton, Olena Kaminska, Gundi Knies, and Alita Nandi. 2012b. "An initial look at non-response and attrition in Understanding Society." Tech. rep., Understanding Society Working Paper Series: 2012-02.

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.

Meyer, Wallace KC Mok, Bruce D. and James X. Sullivan. 2015. “Household Surveys in Crisis.” *Journal of Economic Perspectives* 29 (4):199–226.

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

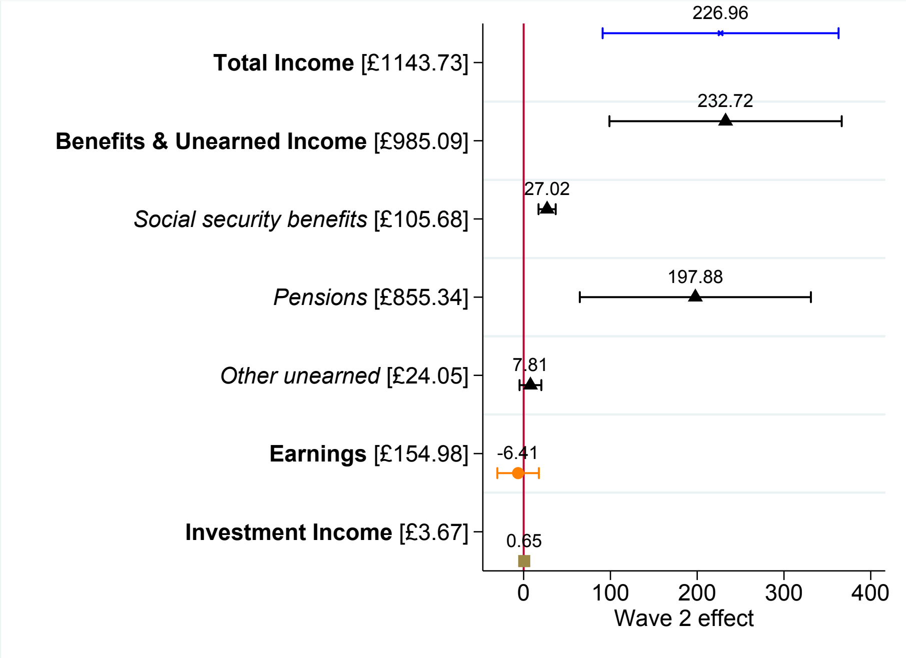
Appendix A

Table A1: Comparison of baseline (wave one) characteristics by year of interview

	2010	2009	Mean Diff	SE	N_2010	N_2009
age (bands):						
10-19 years old	0.0498	0.0494	-0.0004	0.0025	14731	14912
20-29 years old	0.1193	0.1210	0.0018	0.0038	14731	14912
30-39 years old	0.1657	0.1712	0.0055	0.0043	14731	14912
40-49 years old	0.1950	0.1900	-0.0049	0.0046	14731	14912
50-59 years old	0.1703	0.1685	-0.0018	0.0044	14731	14912
60-69 years old	0.1612	0.1616	0.0004	0.0043	14731	14912
70+	0.1388	0.1382	-0.0005	0.0040	14731	14912
has managerial duties:						
manager	0.2402	0.2535	0.0132	0.0071	7114	7476
foreman/supervisor	0.1333	0.1419	0.0087	0.0057	7114	7476
not manager/supervisor	0.6265	0.6046	-0.0219**	0.0081	7114	7476
work place size:						
1-2	0.0453	0.0374	-0.0079*	0.0033	7088	7454
3-9	0.1247	0.1238	-0.0009	0.0055	7088	7454
10-24	0.1552	0.1545	-0.0006	0.0060	7088	7454
25-49	0.1390	0.1468	0.0078	0.0058	7088	7454
50-99	0.1174	0.1118	-0.0056	0.0053	7088	7454
100-199	0.0993	0.1009	0.0016	0.0050	7088	7454
200-499	0.1206	0.1205	-0.0002	0.0054	7088	7454
500-999	0.0624	0.0694	0.0070	0.0041	7088	7454
1000+	0.1244	0.1237	-0.0007	0.0055	7088	7454
region:						
North East	0.0474	0.0484	0.0010	0.0025	14731	14912
North West	0.1244	0.1206	-0.0038	0.0038	14731	14912
Yorkshire and the Humber	0.0910	0.0841	-0.0069*	0.0033	14731	14912
East Midlands	0.0835	0.0868	0.0033	0.0032	14731	14912
West Midlands	0.0910	0.0889	-0.0020	0.0033	14731	14912
East of England	0.1037	0.0978	-0.0059	0.0035	14731	14912
London	0.0940	0.0805	-0.0134***	0.0033	14731	14912
South East	0.1383	0.1481	0.0097*	0.0041	14731	14912
South West	0.0924	0.1012	0.0088*	0.0034	14731	14912
Wales	0.0568	0.0568	0.0000	0.0027	14731	14912
Scotland	0.0777	0.0868	0.0091**	0.0032	14731	14912
month of interview:						
Jan	0.1085	0.1032	-0.0053	0.0036	14731	14912
Feb	0.0880	0.0832	-0.0049	0.0032	14731	14912
March	0.0936	0.0916	-0.0020	0.0034	14731	14912
April	0.0789	0.0815	0.0025	0.0032	14731	14912
May	0.0809	0.0750	-0.0059	0.0031	14731	14912
June	0.0823	0.0854	0.0031	0.0032	14731	14912
July	0.0791	0.0828	0.0037	0.0032	14731	14912
August	0.0828	0.0779	-0.0049	0.0032	14731	14912
September	0.0809	0.0880	0.0071*	0.0032	14731	14912
October	0.0822	0.0826	0.0004	0.0032	14731	14912
November	0.0940	0.0799	-0.0142***	0.0033	14731	14912
December	0.0487	0.0691	0.0203***	0.0027	14731	14912

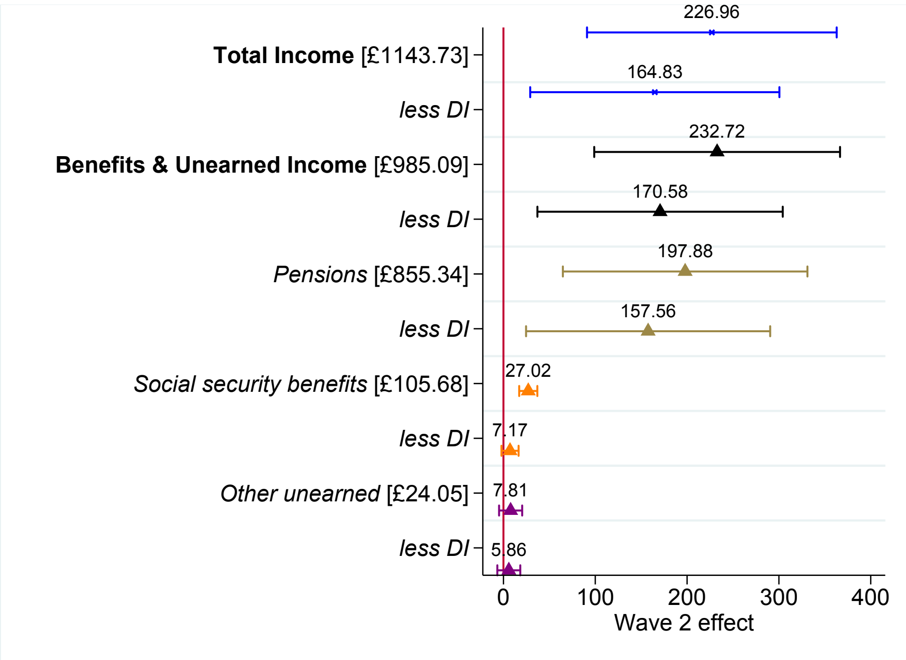
Notes: *p < 0.05, **p < 0.01, ***p < 0.001. Balanced sample of respondents who complete a full-interview at wave 1 and wave 2.

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



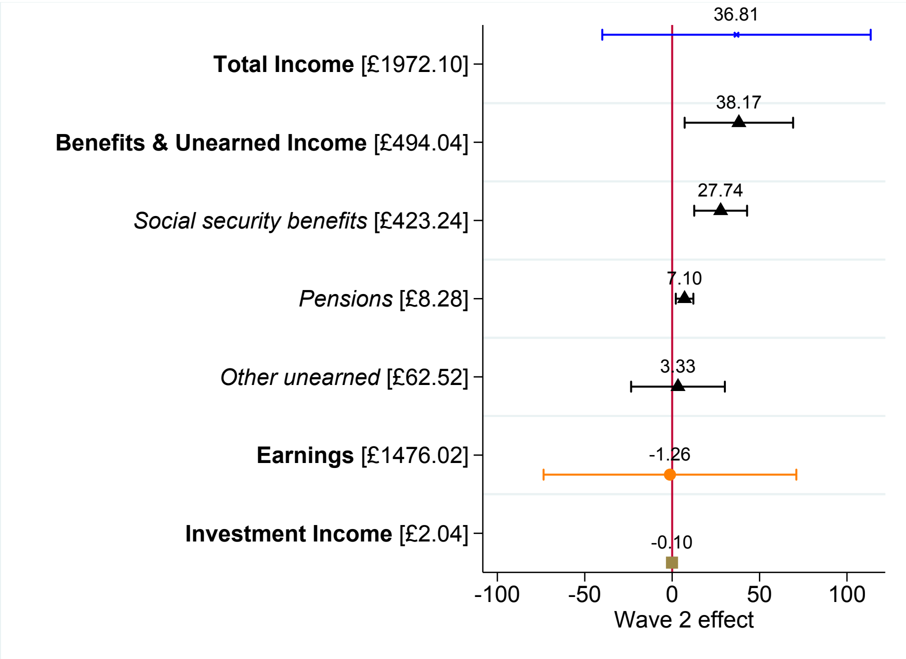
Notes: see figure 3 notes. Sample restricted to respondents of state pension age.

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



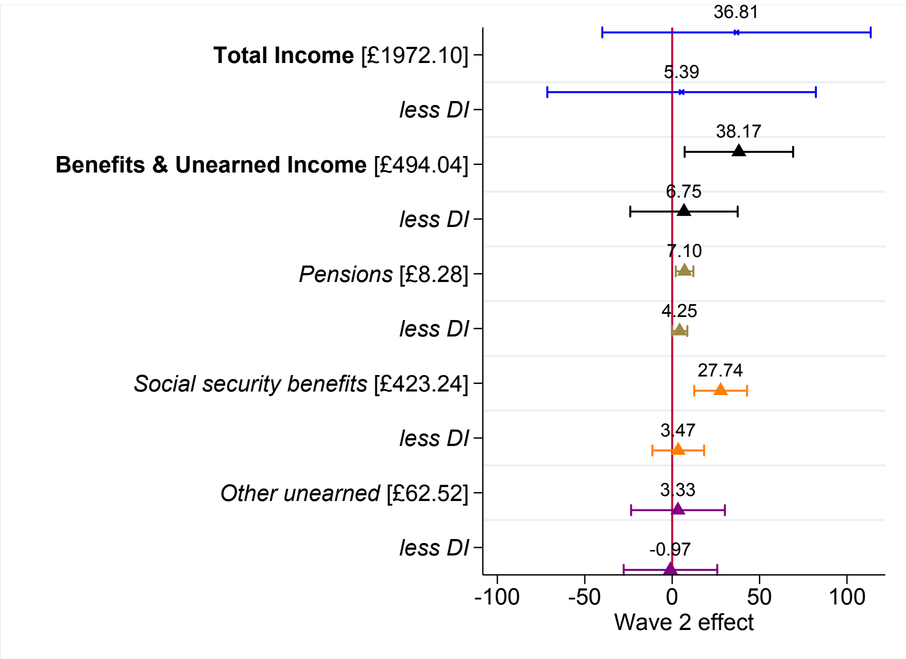
Notes: see figure 3 notes. Sample restricted to respondents of state pension age.

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



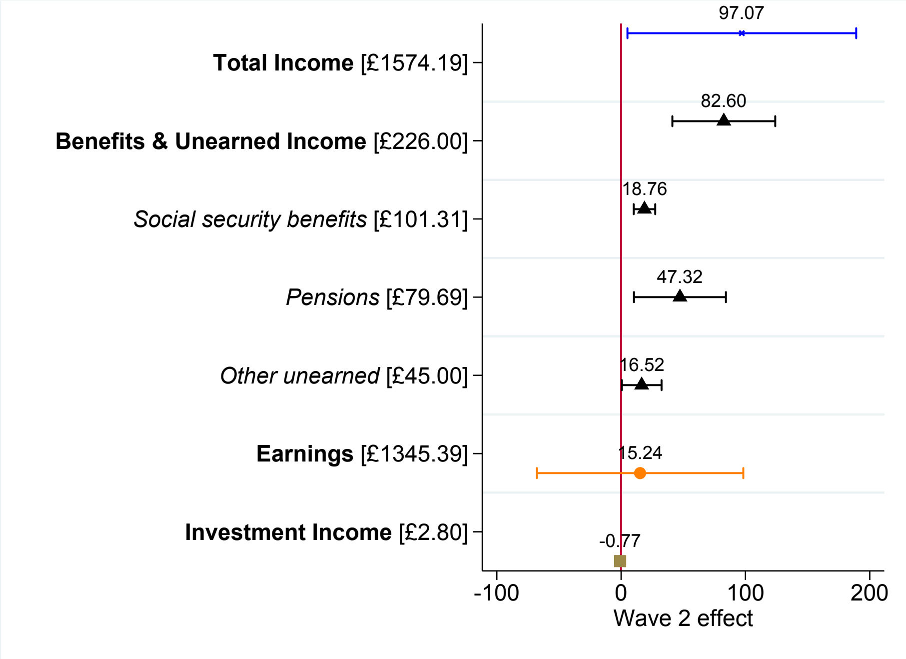
Notes: see figure 3 notes. Sample restricted to respondents less than state pension age and with children.

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



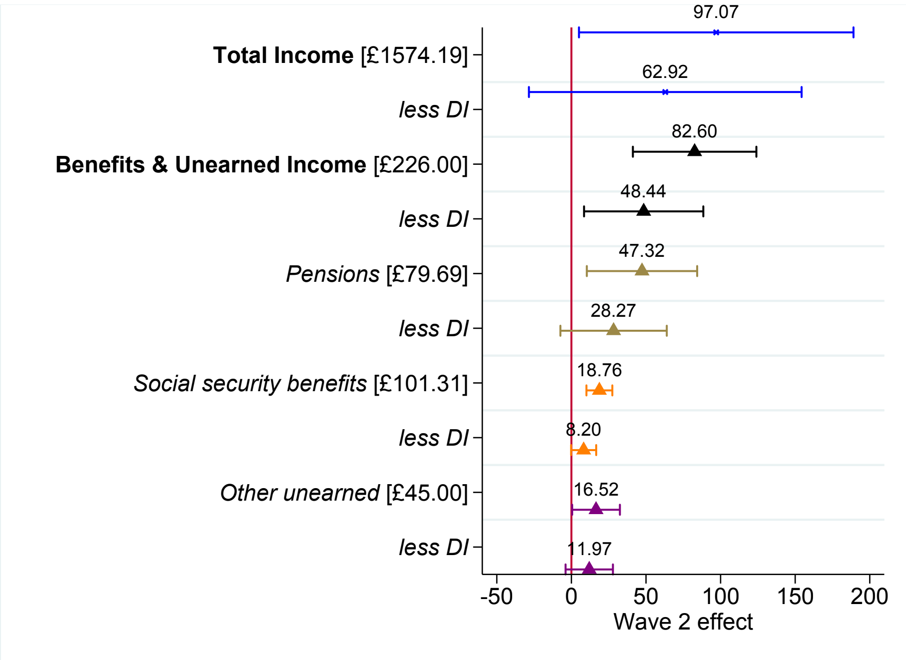
Notes: see figure 3 notes. Sample restricted to respondents less than state pension age and with children.

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.

Table A2: 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 identification section.

Appendix B (Data appendix)

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

B1. Benefits and unearned income

There are two stages to the collection of benefits and unearned income. The first stage sequentially presents a series of up to 6 showcards relating to: broad types of state benefit, unemployment benefits, disability benefits, pensions, family benefits and other income sources. For certain unearned income streams, respondents are asked directly whether they are received (eg. child tax credit for those responsible for children or receiving child benefit). In total 39 income sources are covered which are:

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

25) trade union / friendly society payment

26) maintenance or alimony

27) payments from a family member not living here

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

29) rent from any other property

30) foster allowance / guardian allowance

31) rent rebate

32) rate rebate

33) employment and support allowance

34) return to work credit

35) sickness and accident insurance

36) in-work credit for lone parents

37) other disability related benefit or payment

38) any other regular payment

39) income from any other state benefit

Once this stage is complete, respondents are then asked to report the amount received for each source and the period it covered.

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.

B1.B. Dependent interviewing

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?

Q5) How long a period did that cover?

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.