Household Income Volatility in the UK, 2009-2016

Silvia Avram
(University of Essex)

Mike Brewer
(University of Essex)

Nick Buck
(University of Essex)

Paul Fisher
(University of Essex)

Laura Fumagalli
(University of Essex)

Paper prepared for the 35th IARIW General Conference

Copenhagen, Denmark, August 20-25, 2018

Session 5 (Plenary): Mobility, Opportunity and Longitudinal Data

Time: Thursday, August 23, 2018 [9:00-12:30]

Household Income Volatility in the UK, 2009-2016[1]

Silvia Avram, Mike Brewer, Nick Buck, Paul Fisher and Laura Fumagalli[2]

Draft of July 2018 prepared for IARIW conference, August 2018.

Preliminary and incomplete; do not cite.

Abstract: We study the volatility of individual- and household-level income in the UK between 2009 and 2016 using data from a large longitudinal household panel survey. The volatility of earnings for the working-age has fallen in this period, partly due to a fall in the prevalence of large negative earnings shocks. For older aged individuals, we also find a large fall in the volatility of private income, mainly as a result of a fall in large positive income shocks. The tax-benefit system significantly reduces volatility, especially for retired and low income households; this effect has diminished over the period, and taxes and benefits have become less well correlated with earnings, limiting their ability to counteract swings in labour income, perhaps due to cuts to working age benefits, but by nothing like enough to outweigh the first impact.

Key words: income volatility, income risk, taxes, transfers, insurance, recession, austerity, longitudinal data

[1] "Understanding Society: the UK Household Longitudinal Study" is funded primarily by the UK Economic and Social Research Council with additional funding from a consortium of government departments. The authors are grateful for funding from the UK Economic and Social Research Council (award no. ES/K005146/1) for "Understanding Society: the UK Household Longitudinal Study, phase 3." and for the Research Centre on Micro-Social Change (MiSoC) at the University of Essex, grant number ES/L009153/1.
[2] All authors are from the University of Essex. Correspondence to: savram@essex.ac.uk.

Introduction

A very large number of studies have documented rising earnings and income volatility in the US (Gottschalk and Moffitt, 2009, Bania and Leete, 2009, Dahl et al., 2011, Dynan et al., 2012, Ziliak et al., 2011, Gosselin and Zimmerman, 2008, Shin and Solon, 2011) and in other developed countries (Chauvel and Hartung, 2014, Van Kerm, 2003, OECD, 2011, Daly and Valletta, 2008). Volatility, defined as the extent to which individuals and households experience sizable income swings, has usually been interpreted as a proxy for risk and insecurity (Dynan et al., 2012, Jensen and Shore, 2015).

The amount of risk and uncertainty faced by individuals and households is an important economic question both for understanding individual economic behaviours and for the welfare consequences. Under the usual assumption that individuals prefer certain over uncertain outcomes, an increase in risk and uncertainty has negative consequences for individual welfare. Many households have zero or only very limited assets that can be used to smooth consumption when faced with a negative income shock. Consistent with this fact, a large literature has documented the sensitivity of consumption to income swings, a phenomenon that has been termed as the 'excess sensitivity of consumption to current income'(Flavin, 1985). In addition, income instability may negatively impact on aspects of individual well-being other than consumption. Previous studies have documented an association between income instability and clinical depression (Prause et al., 2009), poor health (Halliday, 2007), food insecurity (Dahl et al., 2014, Leete and Bania, 2010), mortgage delinquency (Diaz-Serrano, 2005), reduced educational achievement of children (Hardy, 2014), poorer child health outcomes (Wolf and Morrissey, 2017), and problem behaviour in adolescents and children (Gennetian et al., 2015, Hill et al., 2013).

The vast majority of the literature on income volatility and instability focuses on the US. Whereas the exact estimates differ somewhat depending on data and methodology, most indicate a substantial rise in the volatility of (male) earnings (Gottschalk and Moffitt, 2009) between the 1970s and the 1990s, as well as increased volatility of household incomes (Dahl et al., 2011, Dynan et al., 2012, Gosselin and Zimmerman, 2008, Ziliak et al., 2011). Low paid or low skilled workers experience higher levels of income volatility and have been affected by increases in volatility to a greater extent (Hill et al., 2017, Hannagan and Morduch, 2015).

The evidence on trends in earnings and income instability in the UK is much sparser and less conclusive. Early work using pseudo-panels and administrative data suggests that earnings have become more unstable in the later part of the 1980s (Blundell and Preston, 1998, Dickens, 2000). More recent work using panel data is inconclusive. Jenkins and Cappellari (2014) show that the volatility of earnings has been relatively constant during the 1990s and early 2000s whereas labour market volatility (i.e. including individuals with zero earnings) has fallen primarily due to stronger employment attachment. Blundell and Etheridge (2010) study the transitory variance (another measure of instability) of earnings and disposable income and conclude that while the former is flat, the latter is u-shaped falling in the early 1990s and rising subsequently. Finally, Jenkins (2011) examines trends both in income volatility and in the transitory variance of earnings and disposable income between 1991 and 2003. He concludes that there is no trend in the instability of earnings, especially among prime-aged male earners. Whereas volatility measures of disposable income also indicated no trend, the transitory variance measure fell slightly over the period. Bartels and Bönke (2013) find that the transitory variance of male earnings increased significantly between 1993 and 2004 but that of household net income remained flat. Finally, Daly and Valetta (2008) find that the transitory variance of male earnings increased in the early 1990s and fell subsequently.

The volatility of earnings is usually higher than that of household disposable income[3] suggesting that other sources of income, in particular taxes and benefits, play an important role in mitigating earnings shocks. Many transfer programs are explicitly designed to cushion incomes in case of adverse shocks to earnings generated by unemployment, ill-health or child birth. While not explicitly designed as insurance, progressive taxation also plays an important role in the intertemporal smoothing of incomes (Knieser and Ziliak, 2002, Varian, 1980). Several studies have suggested that part of the increase in income volatility in the US can be explained by welfare reforms that reduced the coverage and the generosity of US income support programs (Hardy and Ziliak, 2014, Hardy, 2016, Bania and Leete, 2009). Similarly, Jenkins(2011) suggests that different levels and trends in income volatility in the US and UK can be partly explained by the much stronger British safety net. Finally, a significant body of evidence points to the direct consumption stabilization effect of many transfers (Bronchetti, 2012, Gruber, 2000, Browning and Crossley, 2001, Gundersen and Ziliak, 2003).

Typically, low skilled individuals and low income households are found to experience higher levels of income volatility than the rest of the population (Hannagan and Morduch, 2015, Hardy and Ziliak, 2014, Hill et al., 2017). In addition, they are more likely to experience consumption volatility as a direct result of income volatility, consistent with their lower assets and stronger borrowing constraints (Hannagan and Morduch, 2015). As a result, income smoothing effected by the tax-benefit system is disproportionately important for the poor (Morduch and Siwicki, 2017).

[3] Jenkins finds the opposite result JENKINS, S. 2011. Has the Instability of Personal Incomes Been Increasing? *National Institute Economic Review,* 218**,** R33-R43.

We contribute to the literature on income instability by examining the levels and trends of volatility in the UK between 2009 and 2016. As far as we know, this is the first paper to examine the extent to which British incomes have become more or less unstable during and after the Great Recession. Following the literature, we start by examining individual earnings. We then look at household disposable income volatility and how labour income, income from non-labour but private sources, income from transfers and taxes affect it. We are especially interested in examining the role of transfers and taxes in mitigating income volatility during this period given that is has often been described as one of 'austerity' and benefit cuts. Because low income households typically experience higher levels of volatility and are more reliant on welfare transfers for protection against shocks, we examine volatility levels and trends and the role of taxes and transfers by income level. Finally, to gain a better understanding of how the tax-benefit system reduces volatility and how this has changed over the period we study, we examine volatility trends by income source for households affected either by a labour market or a demographic shock.

To estimate our volatility measures, we use data from the UK Household Longitudinal Study (UKHLS) known as "Understanding Society". UKHLS began in 2009 with a sample of approximately 40,000 households, and seeks to interview all household members annually (see Knies, 2017 for more details). We use data from the first 7 waves. The (lengthy) Annex to this paper aims to provide readers with information about the dataset that is relevant for research using income data, as well as providing a background to the analyses undertaken in this paper.

Our findings are as follows. First, in line with previous work, we find that volatility of household labour income is always significantly below that of individual earnings for those of working age, suggesting that the labour income of other household members provide some insurance against swings in own earnings. As expected, taxes and transfers reduce volatility significantly but almost all the impact is due to social security cash benefits or income-dependent refundable tax credits, which reduce around a quarter of the volatility of household private income for the working age (and 40% for those aged 60 or over). High-income households see less volatility than low-income households, but the tax and transfer system is very important in reducing volatility for the bottom quintile. Looking across the period, our main finding is of a decline in volatility. For the working-age, this is driven by a falling volatility of individuals' own earnings, and for those aged 60 or over, by a falling volatility of private unearned income. On the other hand, taxes and benefits became less well correlated with earnings, and became a less important component of disposable income, both of which limit their ability to counteract swings in labour income. This is not enough, though, to outweigh the first impact, and so overall the volatility of disposable incomes fell. Results are valid to most choices of alternative samples, except when we restrict to the very select group of individuals whose households provided valid, non-imputed, responses to all components of income in all waves.

The rest of the paper proceeds as follows. Section 1 describes the UK economic and policy context during the period we study. Section 2 reviews the data and our measure of volatility. The main results are discussed in Section 3. Section 4 concludes. A substantial annex provides full details on the construction of the income variables and the underlying data source, the UK Household Longitudinal Survey.

1. The UK economic and policy context after the Great Recession

The period we study in this paper, 2009-2016, includes most of the strongest economic downturn in the post-war era -the Great Recession of 2008-2012-, as well as the subsequent economic recovery (2012-2016). The Great Recession was atypical in that the fall in output has been passed through to earnings rather than employment. In 2011-12, employment was just 2 pp lower than its pre-recession peak. In contrast, median earnings were 8% lower than before the recession (Cribb et al., 2017). Since 2012, earnings have recovered but remain 2-3% lower compared to 2007/8 whereas the employment rate grew and is now 1.5pp higher (Cribb et al., 2018). Employment and earnings growth since 2012 has been strongest for low income households (Cribb et al., 2018).

The evolution of incomes depends not only on labour market changes but also on the tax-benefit system. Taxes and benefits significantly cushioned the fall in earnings during the recession. At the 10th percentile, the fall in earnings was around 12% but the fall in earnings plus transfers (benefits and tax-credits) was only around 4% (Cribb et al., 2017). Pensioner incomes were especially protected during the recession.

Since 2007/08, a series of tax-benefit reforms have taken place. First, before 2010 in-work benefits to low paid individuals and their families have been made considerably more generous. Subsequently, higher earner families have lost entitlement to working age benefits while out of work benefits have been made considerably less generous and harder to access. For example, among families with no earners, benefit income fell around 6% or £620 per year (Cribb et al., 2018). Thus, some of the gains in employment income experienced by low income households have been off-set by benefit cuts. Overall, these policy reforms continue a long-running trend in which out of work benefits have been reduced and partially replaced by in-work benefits (tax credits).

Given the economic and policy context, what should we expect regarding income volatility? Some authors have suggested that volatility rises during recession and falls during periods of economic growth (Gottschalk and Moffitt, 2009, Jenkins, 2011). However, the evidence that economic downturns increase volatility is weak (for a study that finds the opposite result, see Carey and Shore, 2013). Since the Great Recession affected earnings more than employment, it might be expected that any negative effects would be spread more widely, thus limiting the extent of the shock to any one household. Since low income household generally experience higher volatility, strong growth of employment and earnings in this group might be expected to reduce average volatility. However, during this period the UK also experienced an increase in temporary forms of employment (including zero-hours contracts) and especially in self-employment (Hudson-Sharp and Runge, 2017). In addition to increasing labour market income volatility, unstable and insecure work may also make it harder to claim the correct benefits (Ben-Ishai, 2015).

It is not entirely clear to what extent policy reforms affected income volatility. Cuts to benefits are most likely to affect low income families who have higher levels of volatility. On the other hand, the expansion of the tax credits that occurred during the late 1990s and 2000s produced a system that is well positioned to respond to falls in earnings albeit not to falls in employment (Cribb et al., 2017). Cuts to out of work benefits might not affect volatility so much as the level of income. Finally, changes in the administration of benefits that make it harder for potential recipients to access them may be more important than changes in the rules. Yet, the evidence on this point remains anecdotal.

2. Data and methodology

To estimate our volatility measures, we use data from Understanding Society, the UK Household Longitudinal Study (UKHLS). UKHLS interviews a sample of approximately 40000 households and their members yearly and collects a wealth of information, including demographic, labour market and detailed income data. The study started in 2009 and we use data from the first seven waves. The Annex to this paper includes much more detail, including a comparison of the estimated distribution of income in UKHLS with that in the official dataset for the UK, HBAI.

We follow the large literature on earnings and income volatility and define volatility as the standard deviation of the arcpercentage change in income:

$$V_t = \sqrt{Var[100 * \frac{Y_{it} - Y_{it-1}}{(Y_{it} + Y_{it-1})/2}]}$$

where Y_{it} is income at time t for individual i . We divide the change in income by the mean of the two years rather than by income in the first year (Y_{t-1}) because this has been shown to minimize the influence of outliers and allows for the inclusion of observations where income is zero in either year (Ziliak et al., 2011, Jenkins and Cappellari, 2014). When income is zero in both years, we set the arcpercentage change equal to zero as well (as this implies no change). The arcpercentage change is a symmetric measure that can take values between -200% and 200%.

The standard deviation of the arcpercentage change has the advantage that is simple to compute, requires information on incomes only in two adjacent years and is defined at the individual level. The disadvantage is that it lacks a theoretical foundation. Many income changes, both positive and negative, will be predictable or even voluntary. From a theoretical perspective, the distinction between anticipated and unpredictable changes is important. The earnings dynamics literature has developed variance decomposition methods that attempt to capture these differences. However, as Shin and Solon (2011) point out, results are often sensitive to the actual parametric specification of these models. While not originating in an income dynamics model, our measure of volatility has been shown to be closely related to the variance of transitory shocks in more complex models[4] (Ziliak et al., 2011).

[4] More specifically, Ziliak et al. (2011) show that "volatility" includes changes in income stemming from changes to the time factor loadings of the permanent variance component (i.e. the 'prices' of unobserved skill) and changes to the time factor loadings and shocks of the transitory variance component.

We calculate volatility measures for five income concepts: individual labour earnings (including self-employment), household labour earnings, household private income (defined as the sum of household labour earnings and all other non-labour private sources of income such as investment income, private pensions, alimony etc.), household total gross income (defined as the sum of household private income and all state benefits received by members of the household, including tax credits), and household disposable income (defined as total gross income minus income taxes and national insurance contributions). An overview of our income concepts can be found in Table 1.

Table 1: Definition of income concepts

Income	Definition
Individual labour earnings	Gross monthly labour income: sum of usual gross earnings, self-employment income and earnings from second jobs
Household labour earnings	Sum of total personal monthly income from labour income received by all household members
Household private income	Sum of household labour earnings, private benefit income received by all household members, pension income received by all household members, investment income received by all household members and miscellaneous income received by all household members
Household total gross income	Sum of household private income and social benefit income received by all household members
Household disposable income	Net household income; taxes deducted only on earnings

Source: Authors' compilation

These income concepts have been constructed using the UKLHS derived income variables. The derived income variables summarize and aggregate the detailed income information collected by the survey, including earnings, pensions, benefits and other income[5]. UKHLS also imputes missing values due to item and individual[6] non-response in these derived variables but not in their components[7]. We make use of these imputed values throughout our main analysis but present some robustness checks in Section 5. Information on income taxes and national social insurance contributions is not collected directly by the survey. However, the UKHLS makes available net income estimates derived from imputations based on gross incomes and household and individual characteristics. The imputations seek to replicate as close as possible the methodology developed by the Department for Work and Pensions (DWP) for computing Households Below Average Income (HBAI) estimates (see the Annexes to this paper for details).

[5] See the User Guide on more info on the derivation of these variables: https://www.understandingsociety.ac.uk/documentation/mainstage/user-guide
[6] Missing values for non-respondent households are not imputed.
[7] The detailed imputed income components are unavailable.

All the derived income variables refer to current monthly values. We use monthly Before Housing Costs HBAI CPI values to deflate all incomes to average 2016 levels. Volatility estimates are presented by year of issue. In most cases, this is the year the interview actually took place. However, due to field work constraints, a small number of households are interviewed in the subsequent year.

All incomes are equivalised using the 'modified OECD' scale[8]. This essentially means that demographic changes (such as for example the birth of a child) will appear as income shocks (including in the labour market income estimates) even though income may have remained unchanged. Finally, to avoid unusual arcpercentage change measures, we set all negative incomes to zero; this affects only between 0.18% and 0.07% of observations (depending on the income concept).

Our sample consists of all individuals aged 25 and over who have valid information on all of our five income concepts in at least two consecutive years. Given the divergent evolution of median incomes for the working age and pensioner households, we carry out our analysis separately for individuals aged 25-69 and individuals aged 60+. We do not include individuals younger than 25 in our analyses as many of them are students or apprentices and their larger than average volatility of earnings or income does not necessarily translate into economic instability or insecurity. We are left with 32,239 working age individuals (and approx. 112,000 observations) and 16,901 individuals aged 60 or over (and approx. 62,000 observations). Note that individuals who are working age in one year may move into the 60+ group in subsequent years. We use longitudinal weights throughout to account for selective attrition. However, unweighted results are very similar (available from the authors).

[8] The 'modified OECD' scale assigns a weight of 1 to the first person, 0.5 to subsequent adults and 0.3 to children (defined as aged 13 or under).

3. Volatility of UK incomes after the Great Recession

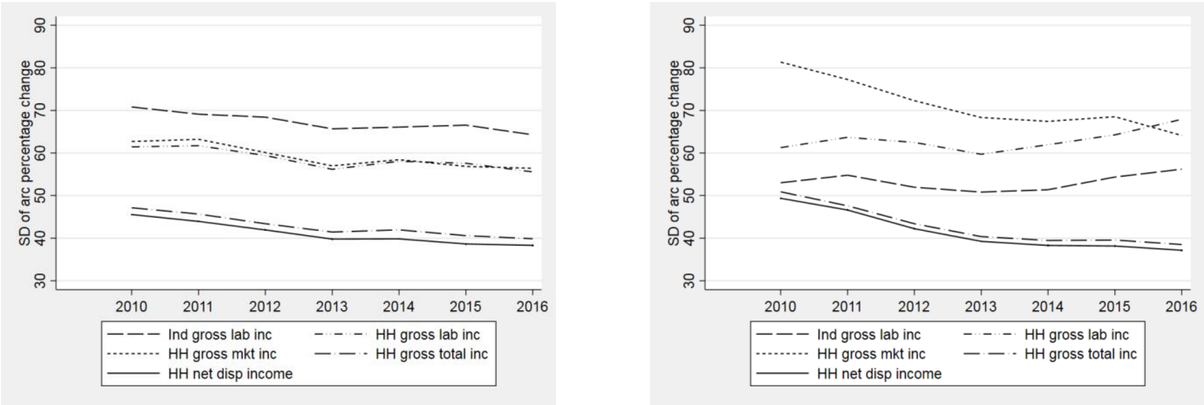
Fig 1 shows the volatility of individual and household incomes between 2009 and 2016 separately for the working age and 60+. For the working age, the volatility of household labour income is always significantly below that of individual earnings suggesting labour and wage shocks are not positively correlated within the household. In multiple earner households, the labour income of other household members provide some insurance against swings in own earnings. Non labour private income sources have a negligible effect on volatility. As expected, taxes and transfers reduce volatility significantly with most of the effect attributable to transfers: around a quarter of the volatility of household private income is reduced by transfers, as opposed to less than 1% in the case of taxes. Transfers play an even larger role in reducing volatility for the 60+ group: almost 40% of the volatility of household private income is reduced by transfers, reflecting the important of state pensions for those who retire.

For the working age, the volatility of individual earnings fell somewhat from around 71 in 2010 to 64 in 2016.[9] This downward trend in the volatility of individual earnings is mirrored almost exactly in the volatility of the other income sources, including disposable income. The absolute difference between the volatility of household private income and household disposable income remained constant during the period meaning that in relative terms, taxes and benefits reduced volatility more in 2016 (32%) compared to 2010 (27%).

Looking at individuals aged 60 or older, there is a striking downward trend in the volatility of household private income which decreases from 81 points to 64. Further analysis has shown that the trend is mostly due to private pension income. It is not clear what lies behind this; the downward trend is still present if we omit data from wave 1, but is less pronounced. As in the

[9] The volatility measure shares the same units as the arcpercentage change, which runs from -200 to 200.

case of the working age, the absolute change in income volatility brought about by taxes and benefits is constant throughout the period.



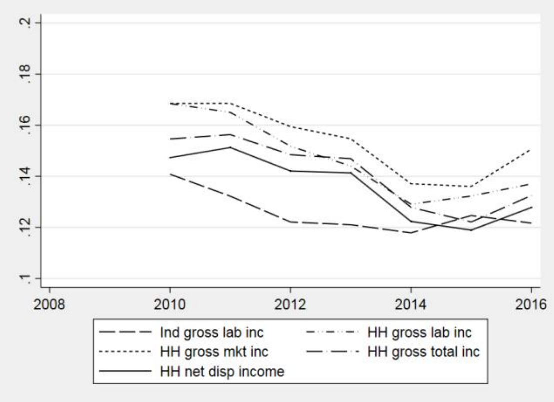
(a) Working age (25-59) (b) Older age (60+)

Fig 1: Volatility of individual and household incomes, 2009-2016

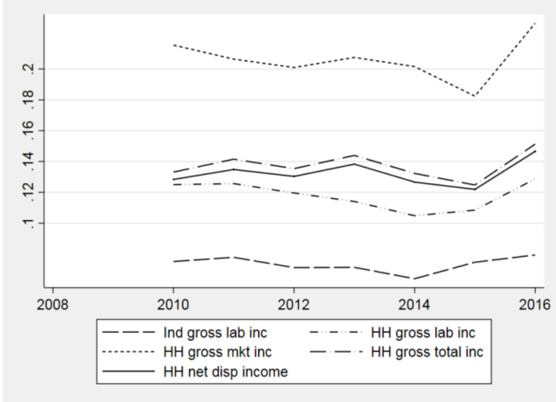
3.1 Trends in the incidence of large shocks

The standard deviation is sensitive to potential outliers. It can also hide changes to the distribution of income shocks. We thus examine the share of changes which are 'large' i.e. where the arceprcentage change is greater than 30% in absolute terms Figure 2 plots the share of large negative shocks by income type, separately for the working-age and 60+.

The share of large negative shocks to individual earnings falls significantly for the working age from around 14% to 12%. The fall is even steeper when looking at household labour income from around 17 to 14%. It thus appears that the reduction in the volatility of earnings observed in the previous graph is at least partly due to a fall in the share of large negative shocks. We can also see that the tax-benefit system reduces the prevalence of large negative shocks by around 2 ppts.



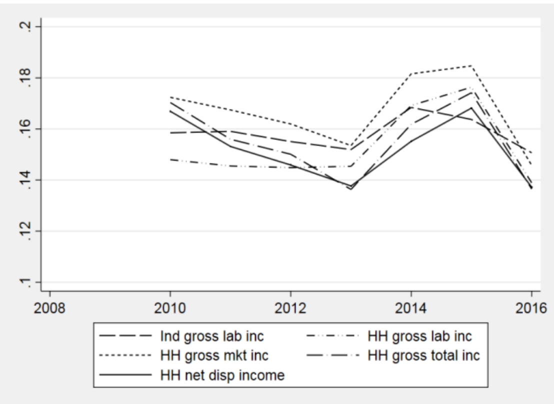
(a) Working age (25-59)



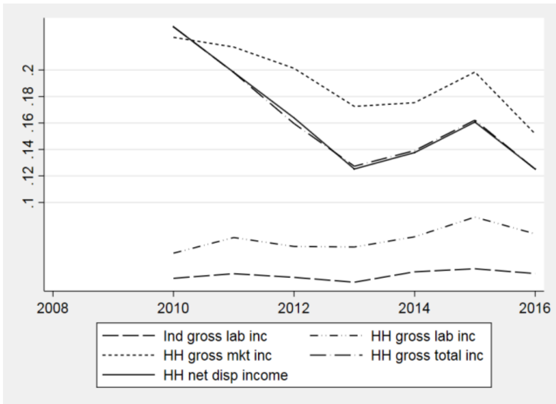
(b) Older age(60+)

Fig 2: Proportion of negative income shocks larger than -30%

The proportion of large negative shocks is more stable among individuals aged 60+. The prevalence of large negative shocks to household private shocks falls initially before rising at the end of the period. Transfers play a larger role in mitigating large negative income shocks for the older age group (about 6-7 pp) compared to the working age.



(a) Working age (25-59)



(b) Older age (60+)

Fig 3: Proportion of positive income shocks larger than 30%

The proportion of income changes larger than +30% is shown in Figure 3. Among working age individuals, the share of positive shocks in individual and household labour income was relatively stable from 2010 to 2013, increased by 3-4 pp between 2013 and 2015 and declined 2pp between 2015 and 2016. A similar trend is noticeable in the case of household disposable income except there is a 2pp decline between 2010 and 2013 rather than stability. The proportion of large positive shocks among individuals aged 60+ follows the same pattern of decline, followed by an increase and another subsequent decline. However, overall the declines are much steeper. Looking at household private income, the share of large positive shocks decreases from around 22% in 2010 to 16% in 2016. Using household disposable income instead, the proportion of large positive income shocks falls by a remarkable 10pp.

3.2 Volatility of low and high income households

We next examine volatility levels and trends by income level. Figure 4 shows income volatility by quintile of household private income in the initial year, pooling all individuals in the sample. We obtain the familiar finding that income volatility is higher among low-income households. The volatility of household private income ranges between 95 and 110 in the first quantile and 40 in the fourth and fifth quintiles.

As expected, the tax-benefit system is most important for reducing volatility at the bottom. Whereas households in the fifth quintile experience almost no reduction in volatility due to taxes and benefits, volatility is reduced by approx. 40% in the first quintile and by around 20-25% in the second. Note also that the volatility is increased when moving from individual earnings to household gross labour income and again to household private income in the first quantile. This pattern suggests that households in the first quantile have fewer earners and that non-labour private income sources are relatively more important for this group.

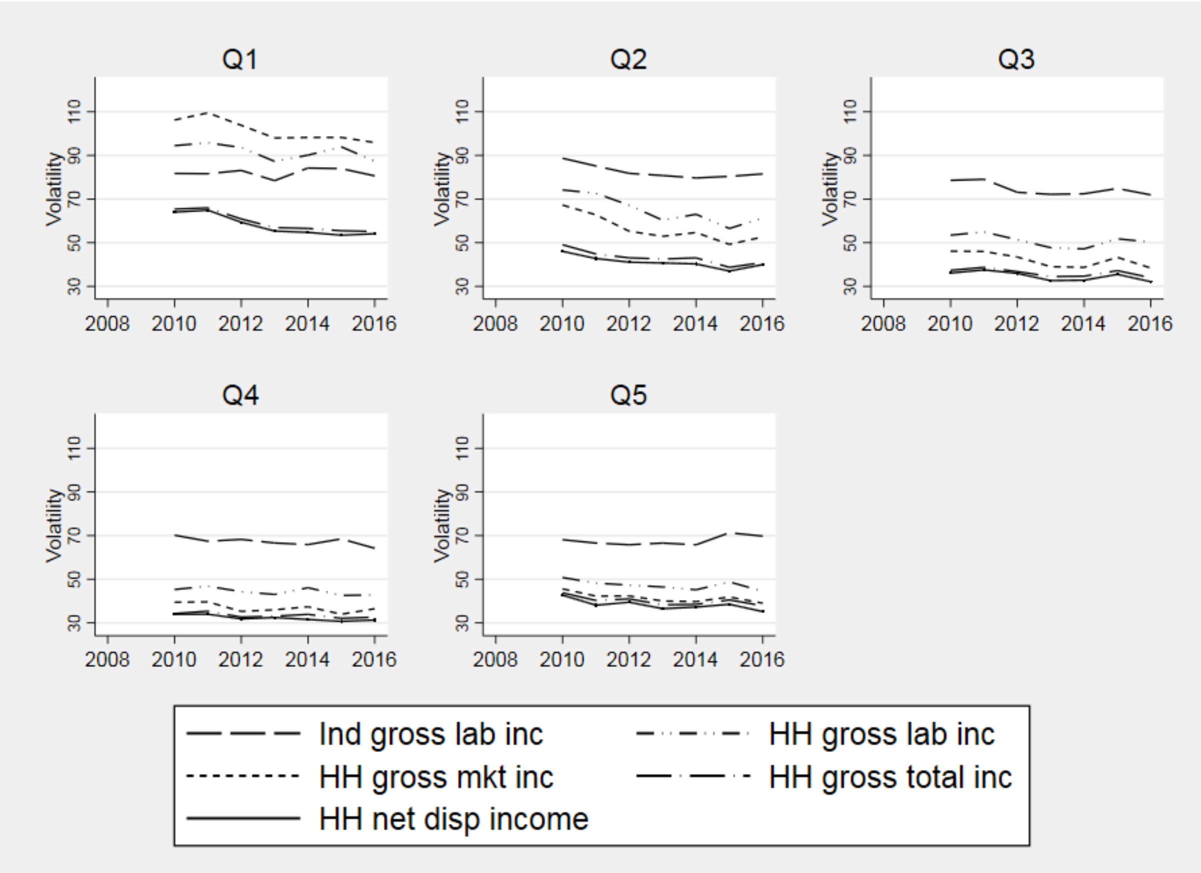


Fig 4: Income volatility by quintile of household private income

The volatility of household gross labour income and household private income fell during the period in the first two quantiles. As a result, the volatility of disposable income has also fallen albeit to a lesser extent in the second quintile. Volatility trends are relatively flat in the upper three quantiles.

3.3 Decomposing trends in the volatility of household disposable income

An important issue for understanding changes in the volatility of disposable income is the extent to which shocks to various income sources are correlated or not. To gain insight, we decompose the variance of disposable income into the sum of the component variances and its co-variances. We decompose disposable income as the sum of own earnings (I1), earnings of other household members (I2), non-labour private income (I3), transfers (I4) and taxes (I5). The variance of changes in disposable income can be written as the sum of the variances of five weighted income components and the corresponding covariances, where the weights are the shares of the income components in disposable income. Formally, we have

$$Var(Y_t) = \sum_{i=1}^{J} Var(s_{it}I_{it}) + 2\sum_{i=1}^{J-1}\sum_{k=1}^{i-1} Cov(s_{it}I_{it,}\, s_{ik}I_{kt})$$

where $Var(Y_t)$ is the variance of the arcpercentage change in disposable income in year *t*, $Var(s_{it}I_{it})$ is the variance of the arcpercentage change in the income component *i* in year *t* weighted by its share in total disposable income, $Cov(s_{it}I_{it}, s_{ik}I_{kt})$is the covariance of the weighted changes in income *i* and *k* in year *t* and *J* is the number of income components which in our case is five. This implies that changes to the variance of disposable income changes can arise from three sources (see (Hardy and Ziliak, 2014) for the full decomposition formula): i) changes to variances of the constituent income sources , ii) changes to their co-variances and iii) changes in the shares. Table 1 below shows the evolution of all three between 2010 and 2016.

Table 1: Decomposition of changes to the volatility of household disposable income

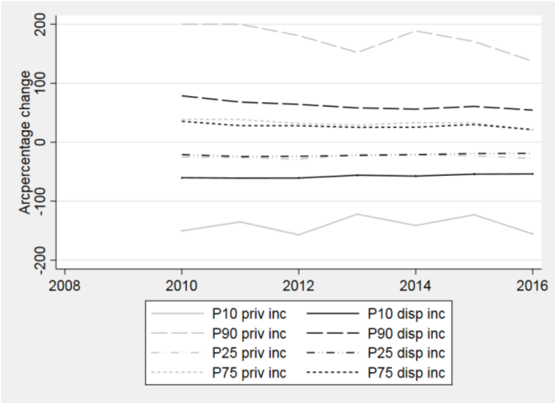
	2010	2011	2012	2013	2014	2015	2016
V (own earnings)	5178.9	5025.62	4823.94	4618.03	4676.22	4989.27	4737.95
V(earnings of others)	4876.42	5012.34	4831.25	4542.16	4642.48	4795.01	4502.63
V(non-labour private income)	10958.67	10821	10196.59	9724.47	9712	9643.52	9482.01
V(benefits)	6417.37	6024.64	5763.39	5801.87	5698.93	6129.15	5737.73
V(taxes)	3453.05	3343.44	3349.11	3358.29	3314.72	3137.04	3053.82
C(own earnings, earnings of others)	82.77	26.28	24.4	-38.31	1.04	-20.27	22.46
C(own earnings, non-labour private income)	-480.6	-381.62	-474.43	-440.86	-386.95	-375.94	-289.74
C(own earnings, benefits)	-835.49	-722.76	-721.11	-622.84	-552.56	-644.5	-507.56
C(own earnings, taxes)	-281.32	-249.6	-253.4	-243.03	-284.71	-210.37	-108.89
C(earnings of others, non-labour private income)	-217.63	-93.91	-161.14	-176.83	-151.12	-151.06	-73.18
C(earnings of others, benefits)	-569.5	-470.48	-403.18	-366.06	-427.22	-548.26	-414.23
C(earnings of others, taxes)	-263.14	-273.3	-262.71	-243.23	-304.3	-244.38	-201.79
C(non-labour income, benefits)	924.9	691.96	636.35	821.02	750.02	816.77	576.38
C(non-labour income, taxes)	-1.63	84.27	171.49	16.15	-75.17	-43.63	-44.57
C(benefits, taxes)	-109.38	-82.99	-47.43	-55.05	-71.06	-5.35	-37.33
S(own earnings)	34.54	35.97	35.84	37.16	38.13	37.74	38.08
S(earnings of others)	41.73	41.76	40.47	40.79	41.98	42.15	42.9
S(non-labour private income)	16.94	16.86	18.38	18.09	17.94	18.81	18.51
S(benefits)	21.99	21.99	22.14	21.58	20.5	19.99	19.44
S(taxes)	0.55	0.57	0.62	0.56	0.48	0.32	0.34

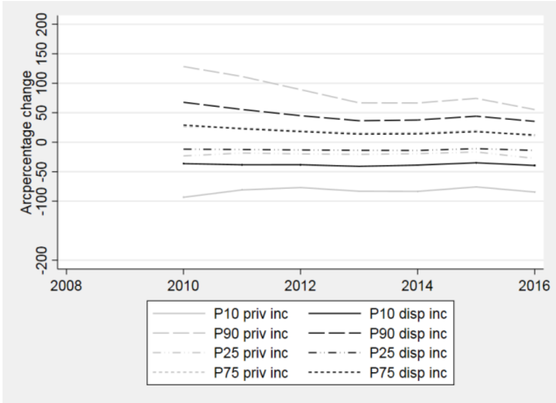
Source: Authors' calculations based on UKHLS.

The variances of arcpercentage changes fell for all five income components between 2010 and 2016. During the seven year period, the volatility of non-labour income fell by almost 13% and that of benefit and tax incomes by around 11%. The correlation between changes to own earnings and changes to the earnings of other household is small but positive in the first year but declines to zero subsequently. As expected, changes to earnings are negatively correlated with changes to non-labour private income, benefits and taxes (taxes are entered negatively signed). However, these correlations are getting weaker: in particular, the covariance between earnings and benefits and that between earnings and taxes fell significantly over the period. Finally, the share of earnings in household disposable income rose slightly, as did the share of non-labour private income, whereas the share of benefits and of taxes fell. Overall then , the changes in the volatility of household disposable income were the result of two conflicting trends. On the one hand, declining earnings and non-labour income volatility reduced the instability of household disposable incomes. On the other hand, taxes and benefits became less well correlated with earnings, and became a less important component of disposable income, both of which limit their ability to counteract swings in labour income. This is not enough, though, to outweigh the first impact, and so overall the volatility of disposable incomes fell.

3.4 Labour market and household dynamics and the stabilizing effect of taxes and benefits

In this subsection, we review the ability of the tax-benefit system to mitigate volatility stemming from labour market and family transitions. We do so by looking at the distribution of shocks to household private and disposable incomes for those households where at least one member was affected by a labour market or a household composition shock. We first examine labour market exits. By construction, individuals who exit employment entirely have an arcpercentage change of individual labour earnings of -200%. Figure 5 though shows selected quantiles of the change in their household private and disposable incomes. Because their labour market exits are likely of a different nature, we examine working-age and 60+ individuals separately.





(a) Working age (25-59)

(b) Older age(60+)

Fig 5: Quantiles of household private income and household disposable income changes of labour market leavers

As expected, labour market exits are associated with smaller shocks to household incomes in the older age group. However, even among the working age, some labour market exits are associated with large positive shocks, which can only happen if other sources of income increase at the same time, or there is a change in household composition. Consistent with the general trends in volatility documented earlier, there is a large fall in the extent to which old-age labour market exits are associated with positive household private income shocks: the 90^{th} percentile decreases from around 128% in 2010 to 55% in 2016. For both groups, the effect of taxes and transfers in cushioning the size of the shock of exiting employment is concentrated in the tails of the distribution: at the 10^{th} percentile, shocks to disposable income are 60% smaller compared to household private income among the working age (and around 50-55% smaller among the older age). At the 90^{th} percentile, shocks are 60-70% smaller among the working aged and 35-50% smaller among those 60+. In contrast, there is virtually no difference in the size of shocks at the 25^{th} and at the 75^{th} percentiles for either group.

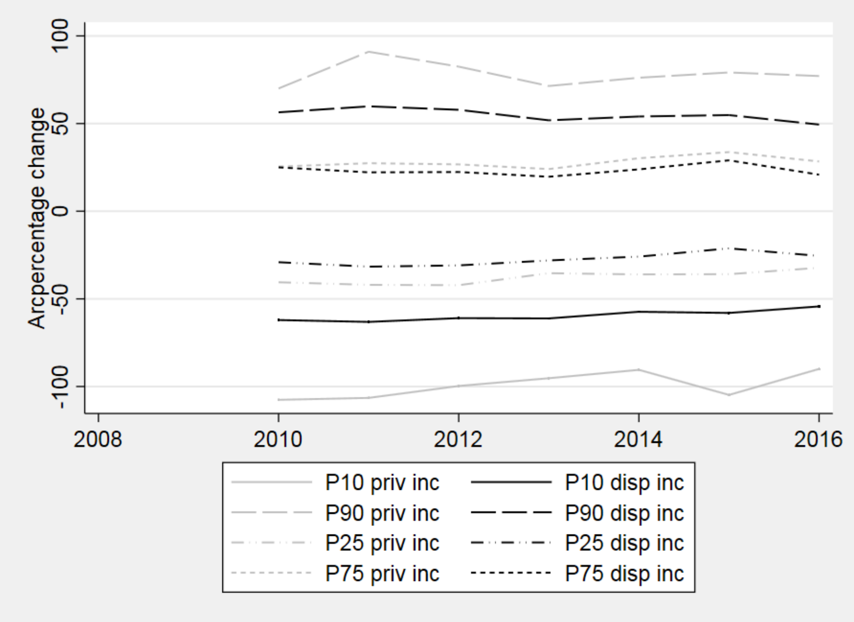


Fig 6: Quantiles of household private income and household disposable income changes of individuals in households that experienced a compositional change

We next examine the distribution of shocks to household income when households experience a compositional change. We define a change to occur every time the list of household members (including children) changes from one year to the next. Thus, changes include both instances when a household member leaves and instances when a new member joins. The results are in Figure 6. As in the case of labour market exits, taxes and benefits reduce the prevalence of large shocks, especially negative shocks. At the 10th percentile, changes to household disposable income are around 40% smaller than changes to household private income. At the 25th percentile, the reduction is only around 25 %.

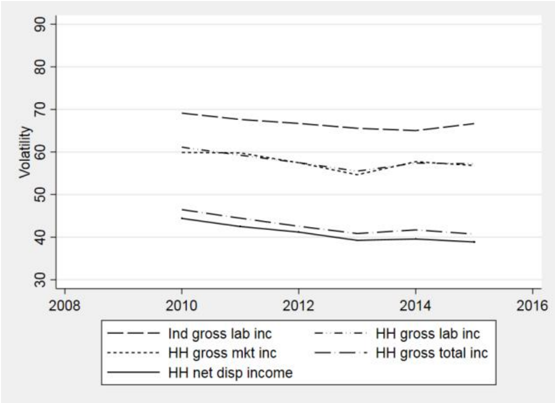
3.5 Robustness and sensitivity checks

In this section, we present results from alternative specifications as a sensitivity check on our main findings.

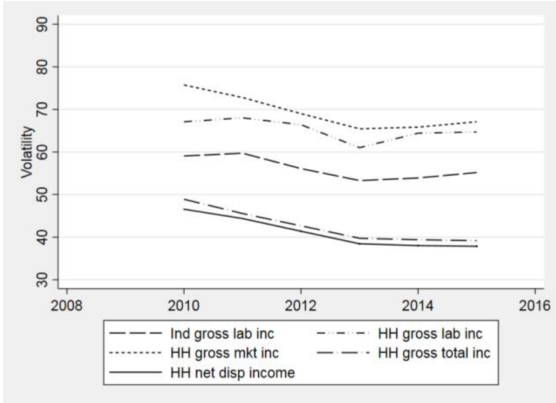
3.5.1 Using balanced samples

We first present volatility trends by income source when the sample is restricted to be '*balanced'*. This provides a check that differential attrition is not affecting our results. Because most of the income concepts we use are household level but households themselves are not a longitudinal unit of observation, it is not immediately clear how '*balanced*' should be defined. We therefore include all individuals for whom an individual interview exists in all 7 waves regardless of whether any components of household income are imputed (which would be due to non-response from the individual herself or other household members). [10]We review the impact of income imputation on results later on.

[10] We have also experimented with a more relaxed definition, where we include all individuals for whom income data (collected or imputed) exists in all seven waves. Results (available from the authors upon request) are unchanged.



(a) Working age (25-59)



(b) Older age (60+)

Fig 7: Volatility of individual and household incomes, 2009-2016, balanced panel

Using a balanced panel makes virtually no difference for the estimation of income volatility. In the case of the working age, the level of volatility is very similar and trends are identical. In the case of older age individuals, the differences in the level of volatility are more noticeable. As might be expected, using a balanced panel yields lower estimates of volatility for all income concepts. The differences range between 1 and 6 points on average, depending on the income concept. The fall in the volatility of household private and disposable income is generally less steep when using the balanced panel compared to the full sample.

We next review the sensitivity of our results to income imputations. Our income concepts are aggregations of individual income sources. As a result, rather than being binary, the imputation flag indicator records the proportion of income that has been imputed. It ranges from 0 to 1. Figure 8 shows volatility trends in our five income concepts change when we restrict our sample to observations where imputed income accounts for i) less than 50% ii) less than 20% , iii) 0% (i.e. there is no imputed income) and iv) 0% in all waves. Note that the fourth specification is very restrictive as it requires valid income information in all waves not only of the individual but also of all the other members of her household. In fact, less than 2000 individuals (out of almost 61000) satisfy this condition (See Table 2).

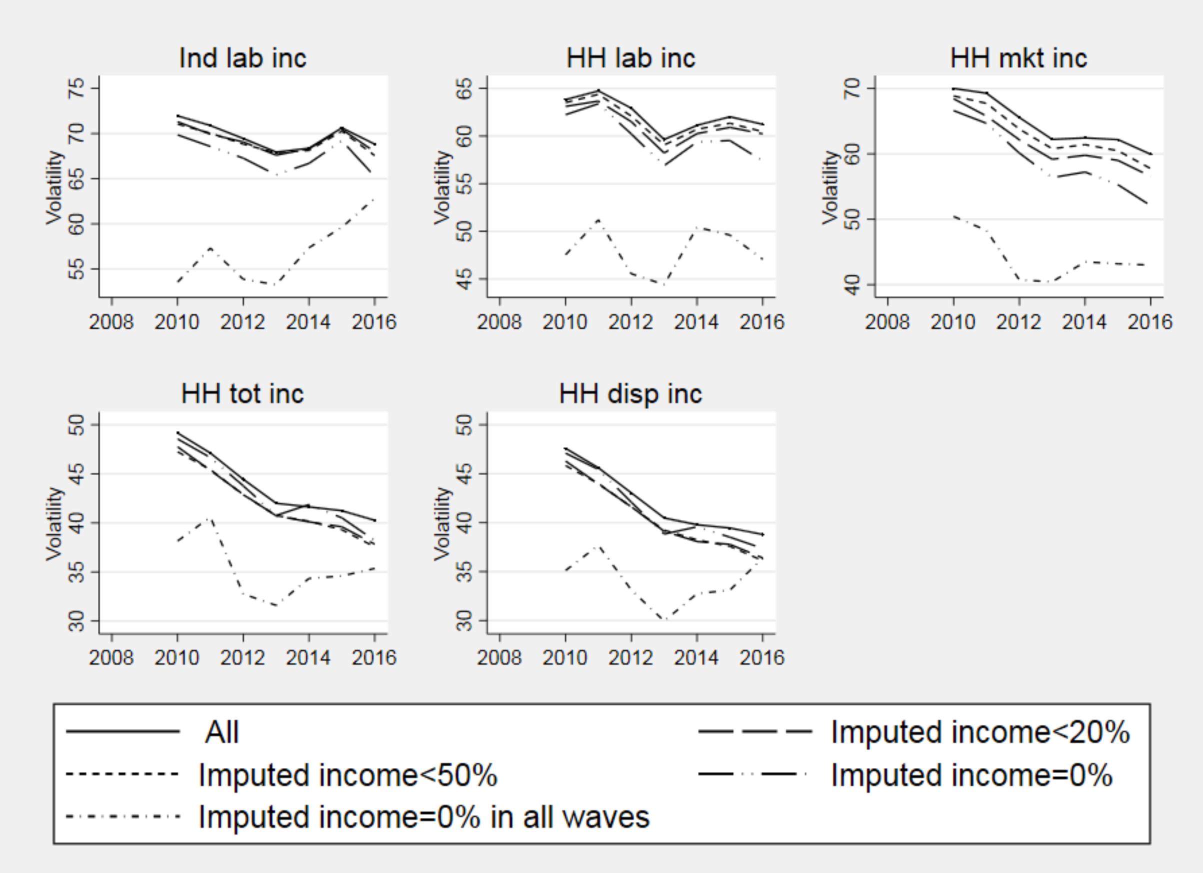


Fig 8: Impact of income imputations on the volatility of individual and household income 2009-2016

Volatility levels and trends are remarkably similar for all income concepts, irrespective of the restrictions we impose on the amount of imputed income. The only exception is the fourth specification where we restrict our sample to individuals in households where all members supplied valid income data in all waves. The level of volatility is much lower for this subsample and the volatility of individual labour income is increasing rather than falling. However, this is a very small and selected sample compared to the rest.

Table 2: Number of individuals and observations, various alternative specifications

Sample	# individuals	# observations
Balanced sample	25,093	162,035
Imputed income <20%	52,778	166,325
Imputed income<50%	56,715	196,863
Imputed income=0%	36,949	91,691
Imputed income=0% in all waves	1,827	9,854

Source: Authors' calculations based on UKHLS.

4. Discussion and conclusions

Using individual and household longitudinal data, we examine the volatility of earnings, households disposable income and intermediate income concepts between 2009 and 2016. We find that the volatility of individual earnings declined significantly among the working age. The decline is concentrated in households in the first and second quintiles of the private market income distribution and is at least partly attributable to a fall in the prevalence of large negative shocks to earnings. These findings are consistent with the strong employment and earnings growth experienced by low income households between 2012 and 2016. We also find a large decline in the volatility of household private market income among individuals aged 60 and older driven by a fall in the share of large positive shocks. It is not clear what explains this trend.

Consistent with the existing body of evidence, we find that the tax-benefit system plays a significant role in reducing the volatility of labour and other private income, although this is dominated by the transfer system (that is social security benefits, means-tested safety net benefits, and income-related refundable tax credits) rather than taxes. The reduction in volatility is substantially higher for older individuals (around 40%) compared for working age individuals (around 25%) and for low income households (around 40%) compared to high income households.

Contrary to what might be expected based on the benefit cuts introduced after 2010, we find that the reduction in income volatility attributable to the tax-benefit system is unchanged throughout the period. This is despite benefits becoming less negatively correlated with earnings and their share in disposable income falling. Finally, we have shown that the tax-benefit system reduces volatility mainly by reducing the prevalence of large shocks (both positive and negative) and that its ability to cushion large shocks has remained constant throughout the period we study.

References:

BANIA, N. & LEETE, L. 2009. Monthly Household Income Volatility in the U.S., 1991/92 vs. 2002/03. *Economics Bulletin,* 29**,** 2100-2112.

BARTELS, C. & BÖNKE, T. 2013. Can Households and Welfare States Mitigate Rising Earnnings Instability? *Review of Income and Wealth,* 59**,** 250-282.

BEN-ISHAI, L. 2015. Volatile Job Schedules and Access to Public Benefits. Washington DC.: Center for Law and Social Policy.

BLUNDELL, R. & ETHERIDGE, B. 2010. Consumption, Income and Earnings Inequality in Britain. *Review of Economic Dynamics,* 13**,** 76-102.

BLUNDELL, R. & PRESTON, I. 1998. Consumption Inequality and Income Uncertainty. *The Quarterly Journal of Economics,* 113**,** 603-640.

BRONCHETTI, E. T. 2012. Workers' Compensation and Consumption Smoothing. *Journal of Public Economics,* 96**,** 495-508.

BROWNING, M. & CROSSLEY, T. F. 2001. Unemployment Insurance Benefit Levels and Consumption Changes. *Journal of Public Economics,* 80**,** 1-23.

CAREY, C. & SHORE, S. H. 2013. From the Peaks to the Valleys: Cross-State Evidence on Income Volatility and the Business Cycle. *Review of Economics and Statistics,* 95**,** 459-562.

CHAUVEL, L. & HARTUNG, A. 2014. Dynamics of Income Volatility in the US and in Europe, 1971-2007: The Increasing Lower Middle Class Instability. *33-rd IARIW General Conference.* Rotterdam.

CRIBB, J., HOOD, A. & JOYCE, R. 2017. Recessions, Income Inequality and the Role of the Tax-Benefit System. London: Institute for Fiscal Studies.

CRIBB, J., NORRIS KEILLER, A. & WATERS, T. 2018. Living Standards, Poverty and Inequality in the UK: 2018. London: The Institute for Fiscal Studies.

DAHL, M., DELEIRE, T. & MOK, S. 2014. Food Insufficiency and Income Volatility in US Households: The Effects of Imputed Income in the Survey of Income and Program Participation. *Applied Economic Perspectives and Policy,* 36**,** 416-437.

DAHL, M., DELEIRE, T. & SCHWABISH, J. A. 2011. Estimates of Year-to-Year Volatility in Earnings and Household Incomes from Administrative, Survey, and Matched Data. *The Journal of Human Resources,* 46**,** 750-774.

DALY, M. C. & VALLETTA, R. G. 2008. Cross-National Trends in Earnings Inequality and Instability. *Economics Letters,* 99**,** 215-219.

DIAZ-SERRANO, L. 2005. Income Volatility and Residential Mortgage Delinquency Across the EU. *Journal of Housing Economics,* 14**,** 153-177.

DICKENS, R. 2000. The Evolution of Individual Male Earnings in Great Britain: 1975-1995. *The Economic Journal,* 110**,** 27-49.

DYNAN, K., ELMENDORF, D. & SICHEL, D. 2012. The Evolution of Household Income Volatility. *The B.E. Journal of Economic Analysis and Policy,* 12**,** 1-40.

FLAVIN, M. 1985. Excess Sensitivity of Consumption to Current Income: Liquidity Constraints or Myopia? *Canadian Journal of Economics/ Revue canadienne d'Economique,* 18**,** 117-136.

GENNETIAN, L. A., WOLF, S., HILL, H. D. & MORRIS, P. 2015. Intrayear Household Income Dynamics and Adolescent School Behavior. *Demography,* 52**,** 455-583.

GOSSELIN, P. & ZIMMERMAN, S. 2008. Trends in Income Volatility and Risk, 1970-2004. *Urban Institute Working Paper*.

GOTTSCHALK, P. & MOFFITT, R. 2009. The Rising Instability of U.S. Earnings. *Journal of Economic Perspectives,* 23**,** 3-24.

GRUBER, J. 2000. Cash Welfare As a Consumption Smoothing Mechanism for Divorced Mothers. *Journal of Public Economics,* 75**,** 157-182.
GUNDERSEN, C. & ZILIAK, J. P. 2003. The Role of Food Stamps in Consumption Stabilization. *Journal of Human Resources,* 38**,** 1051-1079.
HALLIDAY, T. J. 2007. Income Volatility and Health. *IZA Discussion Paper Series,* 3234.
HANNAGAN, A. & MORDUCH, J. 2015. Income Gains and Monthy-to-Month Income Volatility: Household Evidence from the US Fianancial Diaries. *NYU Wagner Research Paper,* 2659883.
HARDY, B. L. 2014. Childhood Income Volatility and Adult Outcomes. *Demography,* 51**,** 1641-1665.
HARDY, B. L. 2016. Income Instability and the Response of the Safety Net. *Contemporary Economic Policy,* 35**,** 312-330.
HARDY, B. L. & ZILIAK, J. P. 2014. Decomposing Trends in Income Volatility: The "Wild Ride" at the Top and Bottom. *Economic Inquiry,* 52**,** 459-476.
HILL, H. D., MORRIS, P., GENNETIAN, L. A., WOLF, S. & TUBBS, C. 2013. The Consequences of Income Instability for Children's Well-being. *Child Development Perspectives,* 7**,** 85-90.
HILL, H. D., ROMICH, J., MATTINGLY, M. J., SHAMSUDDIN, S. & WETHING, H. 2017. An Introduction to Household Economic Instability and Social Policy. *Social Service Review,* 91**,** 371-389.
HUDSON-SHARP, N. & RUNGE, J. 2017. International Trends in Insecure Work: A Report for the Trades Union Congress. London: National Institute of Economic and Social Research.
JENKINS, S. 2011. Has the Instability of Personal Incomes Been Increasing? *National Institute Economic Review,* 218**,** R33-R43.
JENKINS, S. & CAPPELLARI, L. 2014. Earnings and Labour Market Volatility in Britain with a Transatlantic Comparison. *Labour Economics,* 30**,** 201-211.
JENSEN, S. T. & SHORE, S. H. 2015. Changes in the Distribution of Earnings Volatility. *Journal of Human Resources,* 50**,** 811-836.
KNIES, G. 2017. Understanding Society. The UK Household Longitudinal Study Waved 1-7 User Guide. Institute for Social and Economic Research.
KNIESER, T. J. & ZILIAK, J. P. 2002. Explicit Versus. Implicit Income Insurance. *The Journal of Risk and Uncertainty,* 25**,** 5-20.
LEETE, L. & BANIA, N. 2010. The Effect of Income Shocks on Food Insufficiency. *Review of Economics of the Household,* 8**,** 505-526.
MORDUCH, J. & SIWICKI, J. 2017. In and Out of Poverty: Episodic Poverty and Income Volatility in the US Financial Diaries. *Social Service Review,* 91**,** 390-421.
OECD 2011. Earnings volatility: Causes and Consequences. *In:* OECD (ed.) *OECD Employment Outlook 201.* Paris: OECD Publishing.
PRAUSE, J., DOOLEY, D. & HUH, J. 2009. Income Volatility and Psychological Depression. *American Journal of Community Psychology,* 43**,** 57-70.
SHIN, D. & SOLON, G. 2011. Trends in Men's Earnings Volatility: What does the Panel Study of Income dynamics Show? *Journal of Public Economics,* 95**,** 973-982.
VAN KERM, P. 2003. An Anatomy of Household Income Volatility in European Countries. *CHER Working Paper.* Differdange: CEPS/INSTEAD.
VARIAN, H. R. 1980. Redistributive Taxation and Social Insurance. *Journal of Public Economics,* 14**,** 49-68.
WOLF, S. & MORRISSEY, T. 2017. Economic Instability , Food Insecurity, and Child Health in the Wake of the Great Recession. *Social Service Review,* 91**,** 534-570.

ZILIAK, J. P., HARDY, B. & BOLLINGER, C. 2011. Earnings Volatility in America: Evidence from matched CPS. *Labour Economics*, 18**,** 742-754.

Annex: the UK Household Longitudinal Study (UKHLS) and its income data

The aim of this annex is to provide readers with information about Understanding Society, the UK Household Longitudinal Study (UKHLS), that is relevant for research using income data, and to provide a background to the analyses undertaken in this paper. It provides a brief overview of the overall UKHLS study design and implementation of the survey, drawing particular attention to features which are particularly relevant for research using income data. Further information on the aspects is available in a number of other sources including the UKHLS User Manual (Knies, 2017) and the UKHLS Quality Profile (Lynn and Knies, 2016). It also discusses in more detail the approach to income data collection, statistical adjustment for missing data through imputation and the computation of derived new net income measures. In order to assess the quality and representativeness of the data it also provides some comparisons with other UK income survey sources.

1. UKHLS Study Design

1.1. Overview: the household panel design

The design of the UKHLS follows the standard household panel design, describe briefly below. It built on the success of the on the success of the British Household Panel Survey (BHPS)[1]. The BHPS has been heavily used by government departments and by researchers within and outside the UK. It has been accessed by more than 5000 users and generates more than 150 publications per year. The ambition of the UKHLS has been to support a much wider range of research than the BHPS, based on a much larger sample size and a wider range of data collection. It started data collection in 2009. Sample members are interviewed annually. Data collection for a single wave is scheduled across 24 months. As of July 2018, seven waves of data are currently available to researchers, with an eight to be released this autumn. The study incorporates the BHPS sample for which 25 waves of data are now available.

The household panel design involves selecting a sample of all individuals found in an initial representative sample of households and following those individuals over time and also collecting data from other individuals with whom they form households, as well as including new births to sample members. The design means that there is repeated measures longitudinal data about individuals but also data from other members of households in which they reside over time, allowing high quality household context measures to be used in analysis. It is important to stress here that the longitudinal elements, just as in the cohort studies, are the individual people. It is not a longitudinal study of households, since arguably households have no coherent existence over time, and focusing analysis only on households whose composition does not change between waves leads to severe biases (see Duncan and Hill 1985). Rather, it is a study of individuals in their changing household contexts and this context is very important for analysis of many life domains (Giles 2001).

The UKHLS forms part of an international network of such studies including the Panel Study of Income Dynamics (Hill 2001), the German Socio-Economic Panel Study (Wagner et al 1993), the Household, Income, and Labour Dynamics in Australia Survey (Wooden et al 2002), the Swiss

[1] This annex is modelled loosely on Jenkins 2011, which discusses BHPS income data

Household Panel (Budowski et al 2001), the Survey of Labour and Income Dynamics in Canada (Webber 1994) and other active household panels in South Africa, Israel, Korea, China. The household panel design was established in the Panel Study of Income Dynamics (PSID) in the USA in the late 1960s. This design has proved extremely powerful in understanding the dynamics of populations and the determinants of behaviour and outcomes at household and individual level.

A key feature of the household panel design is that the initial sample is representative of the whole population of all ages, and with appropriate following rules it will remain representative of the population as it evolves over time, with the exception of new immigrants to the population. Research from studies with a household panel design supports direct inferences about the whole population.

The design is particularly relevant for the analysis of income and economic well-being more generally since it allows the collection of high quality longitudinal measures of household income in order to analyse the transitions in economic circumstances which members of the population experience, for example the persistence of poverty or the factors which affect income mobility. Data from BHPS and UKHLS based on this design is regularly used in UK government publications on the dynamics of low income.

In addition to the main survey, there is a separate survey, the Innovation Panel (IP), which is fielded in the year before the main survey. It tests varying measurement issues, and its instruments are somewhat different from the main survey, though they contain the same main measures. The IP has been used for experimental work on improving measurement of income and related measures (see, for example, Gaia (2017)).

1.2. Sample Design

The *Understanding Society* main survey sample consists of a new large General Population Sample (GPS) plus three other components: the Ethnic Minority Boost Sample (EMBS), the former BHPS sample and the Immigrant and Ethnic Minority Boost Sample (IEMBS). The design of the first two components is described in more detail in an *Understanding Society* working paper, see Lynn (2009). The design of the IEMBS is described in Lynn, Nandi et al. (2016). The GPS is based upon two separate samples of residential addresses in England, Scotland and Wales and in Northern Ireland. The England, Scotland and Wales sample is a proportionately stratified (equal probability), clustered sample of addresses selected from the Postcode Address File. Northern Ireland has an unclustered systematic random sample of addresses selected from the Land and Property Services Agency list of domestic addresses.

1.2.1. General Population Sample

The general population sample is a stratified, clustered, equal probability sample of residential addresses drawn to a uniform design throughout the whole of the UK (including north of the Caledonian Canal). The Northern Ireland sample is not clustered. Within Great Britain, the Primary Sampling Units (PSUs) are postal sectors stratified by nine regions of England plus Scotland and Wales), population density and minority ethnic density. 2,640 postal sectors were selected systematically, with probability proportional to size (number of addresses). Within each sampled sector, 18 addresses were selected systematically, resulting in an equal-probability sample of a total of 47,520 addresses in Great Britain. In Northern Ireland, 2,400 addresses were selected systematically from the Land and Property Services Agency list of domestic properties, thus making a total of 49,920 selected addresses in the UK. Since constraints of survey capacity meant that fieldwork needed to be spread over a two year period, the overall sample was divided into 24 monthly sub-samples, each independently representative of the UK population. This means that differences over time within a wave can be compared using nationally representative samples, and annual or quarterly subsets can be independently analysed.

1.2.2. Ethnic Minority Boost Sample

The goal for the ethnic minority boost sample was to provide samples of at least 1,000 adults in each of the five largest ethnic minority groups: Indian, Pakistani, Bangladeshi, Caribbean and African. Such a sample would support group-specific analyses of these ethnic groups (Berthoud et al 2009). While the sampling targets are defined in terms of numbers of adults, the sample is of households.

The sampling approach first identifies geographic areas with at least 5% density of ethnic minority groups. Because the 2001 Census was becoming outdated, the density estimates were adjusted using more recent survey estimates. These high density sectors were 36 per cent of the total sectors and accounted for 85% of all members of minorities. Further sub-sampling of the high density areas was done to increase the efficiency of the yield. Thus, a higher sampling fraction was used for areas anticipated to yield three or more households while successively smaller fractions were used for areas expected to yield two, one or zero ethnic minority households. The initial step was identifying postal sectors with relatively high proportions of relevant ethnic minority groups, based upon 2001 Census data and more recent Annual Population Survey data. The set of 3,145 sectors constituted approximately 35% of the sectors in Great Britain and covered between 82% and 93% of the population of the five ethnic minority groups.

At selected addresses, households were screened for the presence of a member of a minority ethnic group. The screening question was, “Do you come from or have parents or grandparents who come from any of the following ethnic groups?” The response categories are Indian, mixed Indian, Pakistan, Bangladeshi, Sri Lankan, Caribbean/West Indian, mixed Caribbean/West Indian, North African, Black African, African Asian, Chinese, Far Eastern, Turkish, or Middle Eastern/Iranian, or other. At the screening stage, all households with the smaller ethnic groups were selected and there is some de-selection of larger ethnic minority groups, e.g. Indians.

Following the first six months of data collection the procedures were reviewed and modified. One change was to increase the number of addresses issued in areas estimated to be high in Bangladeshi, the smallest of the five main ethnic groups.

The screening question also identified persons in the following categories in addition to the five target groups: Chinese, other Far Eastern, Sri Lankan, and Middle Eastern. While it is useful to be able to identify members of these ethnic groups, the number of cases is well below 1,000. White minorities were not selected in the screening but can be identified by survey questions in the general sample.

The overall sampling fractions combine a) the probability of sampling the sector, b) the fraction of addresses selected within the sector, and c) the probability of a household being retained following the application of the random selection mechanism described above.

1.2.3. Former BHPS Sample

Understanding Society incorporates the BHPS sample members into the overall sample design beginning in Wave 2. The extensive longitudinal data of the BHPS has great scientific value, including the opportunity for early longitudinal analyses of *Understanding Society*. The BHPS was a random sample of Great Britain, excluding the Scottish Highlands and Islands. In its first wave in 1991, it achieved a sample of 5,500 households. Boost samples of Scotland and Wales were added in 1999 and of Northern Ireland in 2001. These modifications were motivated by interest in analyses in these countries, related to political changes associated with devolution in the UK.

In planning the timing of fieldwork for the BHPS sample, it was necessary to balance fully integrating the sample into Understanding Society as against creating a discontinuity in the BHPS series. After consultation, it was decided that it was most important to ensure the integration of BHPS into the new study (Laurie 2010). So instead of having its fieldwork concentrated between September and December, as was the practice up to 2008, fieldwork is distributed evenly over the 12 months of the first year of data collection beginning in January 2010, as part of wave 2 of *Understanding Society*. This does introduce a one-off longer gap between interviews for the BHPS sample. From wave 2 onwards the BHPS sample has the same questionnaire as the *Understanding Society* general population sample.

1.2.4. Immigrant and Ethnic Minority Boost Sample

This sample was introduced at Wave 6. It includes people who were born outside the United Kingdom (“immigrants”) and members of five ethnic minority groups: Indian, Pakistani, Bangladeshi, Caribbean, and African. Some people, of course, fall into both categories. This sample therefore provides coverage for the first time of people who have entered the UK since Wave 1 of the Study (“new immigrants”), while also boosting the numbers of immigrants who arrived earlier and of ethnic minorities who either arrived earlier or were born in the UK. The IEMBS was designed to provide around 2,000 adult immigrant respondents and around 2,500 from the target ethnic minority groups.

The sample was identified through in-person doorstep screening of a set of addresses that were sampled from the Postcode Address File following a stratified multi-stage design in which the strata were defined by small area level indicators from the 2011 population census of the distribution of ethnic groups and immigrants. Five strata were created. Sampling was restricted to four strata, the fifth consisting of the sectors with the very lowest proportions of immigrants and ethnic minorities. Sampling fractions varied between the four strata, with the highest sampling fraction applied to a stratum with the highest proportions of Africans. In each sampled stratum, a number of postcode sectors were selected with probability proportional to the predicted number of eligible households. In each sampled sector, a number of addresses were selected such that the predicted number of eligible households in the sample did not vary between sectors within a stratum (so the number of selected addresses was larger in sectors with a lower predicted proportion of eligible households). A screened household was eligible for interview if it contained at least one person who was born outside the UK and/or a member of a relevant ethnic minority group, even if that person was a child.

The “boost” samples do not therefore provide complete population coverage of the relevant subgroups but are instead designed to be used in combination with the other samples, as described above. The sample of “new immigrants” is estimated to provide around 74% population coverage.

1.2.5. Household and family definitions

The UKHLS definition of a household and the unit to which household income refers is the same as that use in UK government surveys and follows UK statistical practice. A household is 'one person living alone or a group of people who either share living accommodation or share one meal a day and who have the address as their only or main residence'. The definition also requires six months' continuous residence, implying that students will be included at their term time address, unless living at a hall of residence.

UKHLS does not work with a family definition in the way that PSID does for example as its primary unit of analysis. However, UKHLS data does provide indicators for units below the household level, consistent with that used in the British tax and benefit system known as benefit units. These are defined as a single person or a couple living together and would include dependent children of adult members of the benefit unit. Dependent children are aged under 16 or under 19 and in full-time education, excluding higher education, and not married or with a child. These units are used for the assessment of means tested state benefits. A household may combine several benefit units, for example where a non-dependent child lives with his or her parents.

1.3. Following respondents over time and eligibility for interview

The composition of the household, the first stage of sampling, determines the rules for following individual respondents over time. The individuals found at selected households in the first wave are designated as Original Sample Members (OSM). We attempt to retain OSM respondents as part of the sample as long as they live in the UK. Individuals joining the household of an OSM after the sample selection/first interview are temporary sample members (TSM). However, births to an OSM are also classified as OSMs. We attempt to interview TSM participants in successive waves as long as they live in the household of an OSM. In sum, TSMs are not followed for interviews when they leave the household, but OSMs are.

The following rules mimic the demographic processes by which the population is reproduced, including births and deaths, partnership formations and dissolutions, and emigration. They provide a natural sampling method over time, which represents the evolving pattern of households and families in the UK. The one exception is that there is no direct way in which the following rules capture immigrants into the UK. Apart from immigration, the sample remains representative of the UK population as it changes over time, subject to weighting for attrition. Whether and how to sample new immigrants remains an issue to be decided in the future development of the study.

In general, longitudinal analysis of individuals will focus on OSM respondents only since TSMs will drop out when they no longer live with OSMs. However, data from TSM respondents will be used to compute household measures when they are co-resident with OSMs. This will particularly apply to household income. So typically analysis of income mobility will use both OSM and TSM data to construct household income measures, but will only include OSM individuals in the analysis and will treat the household income as an attribute of those individuals.

1.4. Data Collection

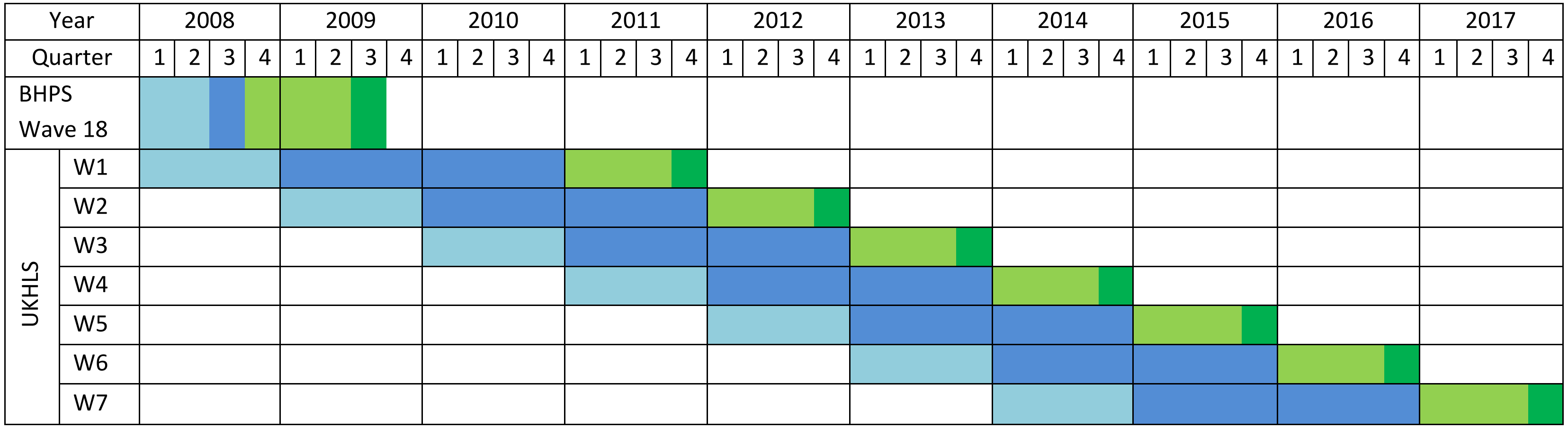
As a result of fieldwork capacity issues given the very large sample size, particularly in the early years, data collection for each wave of UKHLS covers a 24 month period, and individual waves overlap so that sample members are interviewed at annual intervals. The sample is issued in 24 monthly sample and sample members remain in the same month of issue at each wave, so that the interval between interviews should be approximately 12 months. The field period for each month sample lasts for around four months in order to track movers and undertake refusal conversion, so there may be some variation in the interval between individual interviews.

There are some specific exceptions to the 24 month fieldwork period. The Northern Ireland and the former BHPS sample components are issued over the first 12 months of the wave. It should also be noted that the BHPS sample was first interviewed as part of UKHLS at wave 2, i.e. in 2010. In the autumn of 2008 and early in 2009 BHPS wave 18 was conducted. Table 1 shows the timing of the sample issue for the data included at the latest (2017) release.

Most of the data collection is conducted face-to-face via computer aided personal interview (CAPI). There are also self-completion instruments for youth and adults. The youth instruments are administered on paper. The adult self-completion questionnaire shifted from paper to computer administered self- interview (CASI) in Wave 3. From Wave 3 onwards, there was also a telephone mop-up at the end of the fieldwork period for each sample month. From Wave 7 onwards, some proportion of the sample take part in computer assisted online interviews (CAWI).

Income data is almost entirely collected from the adult individual schedule, and there are no household level measures of total income collected. This is on the basis that household level measures of income are unlikely to be particularly accurate. Housing cost measures are collected in the household schedule and there is some information on housing benefit income associated with this.

Table 1: Timing of data collection start



Notes: Northern Ireland (from Wave 1 onwards), BHPS (from Wave 2 onwards) and IEMB (from Wave 6 onward) samples interviewed in year 1 of each wave only.

development
data collection
data processing & documentation
data release

1.5. RESPONSE OUTCOMES

Response outcomes can be considered in four different ways. Firstly, there is the question of whether a household containing sample members provides any response. This is consistent with the way in which cross-sectional household surveys report response. Secondly, and most important for longitudinal analysis is the question of whether individual sample members continue to respond at successive waves or drop out, so-called attrition. Thirdly there is the question of whether all eligible members of participating households provide a response. This is important for the construction of household level measures, including household income. Finally there is the issue of response to individual questions for sample members who do provide some response to the questionnaire, so-called item response or non-response. This section briefly discusses the last three of these types of response. The first, household level response is discussed in Lynn and Knies, (2016). Lynn and Borkowska (2018) also provides some analysis of overall sample representativeness.

1.5.1. ATTRITION

Wave on wave retention rates for full adult respondents are shown in table 2 below. It shows for part of the sample the percentage of those interviewed at the first wave who are interviewed the next wave. The retention rates are clearly less than in the early waves of BHPS, reflecting the much greater difficulty of achieving very high response rates in recent years compared with the early 1990s. A similar decline is reflected in other major UK (cross-sectional) surveys. The lower response of the ethnic minority boost sample should also be noted. The drop in response in wave 6 is associated with a change in fieldwork organisation.

Table 2: Wave-on-wave retention rates: adult respondents

	General Population Sample	BHPS	Ethnic Minority Boost	Immigrant and Ethnic Minority Boost	All
Wave 1 - wave 2	76.23		63.49		74.56
Wave 2 - wave 3	80.53	86.34	71.3		80.83
Wave 3 - wave 4	85.29	86.47	76.54		84.72
Wave 4 - wave 5	87.85	88.44	79.65		87.19
Wave 5 - wave 6	83.41	86.01	74.27		83.09
Wave 6 - wave 7	86.74	87.68	79.25	63.35	83.79

Table 3 shows retention rates for all enumerated individuals and effectively indicates whether households containing sample members have been productive at each wave. It should be noted that for both respondents and enumerated individuals there are significant numbers who drop out for one wave but are re-interviewed at the subsequent wave so the cumulative pattern of attrition is not simply the product of the individual wave attrition rates.

Table 3: Wave-on-wave retention rates: enumerated individuals

	General Population Sample	BHPS	Ethnic Minority Boost	Immigrant and Ethnic Minority Boost	All
Wave 1 - wave 2	77.29		67.7		75.72
Wave 2 - wave 3	83.03	88.59	78.33		83.52
Wave 3 - wave 4	87.03	88.28	81.35		86.59
Wave 4 - wave 5	85.86	88.26	78.05		85.36
Wave 5 - wave 6	82.4	86.11	76.91		82.47
Wave 6 - wave 7	88.3	89.35	84.04	67.16	85.41

1.5.2. WITHIN-HOUSEHOLD RESPONSE

Ideally we want all adult members of participating households to respond to the full questionnaire. For income analysis in particular this is desirable in order to be able to compute total income from reports of all members. Unfortunately, it is difficult to obtain complete response for all households because some members are unable or unwilling to respond. Where the household member has not refused to be involved in the study, proxy data may be collected. This contains some limited information about personal income. Where the member has refused to be involved or no-one can provide proxy information, information is restricted to that collected on the household membership roster.

Table 4 shows the extent of within household non-response in UKHLS at wave 7, the latest wave at the time of writing. Differences across wave are relatively small. There is some difference between samples, with lower complete household response in the ethnic minority and immigrant samples. The BHPS has a somewhat higher percentage with complete household, though this is not as high as it was when the BHPS was being conducted on its own.

Table 4: Percentage distribution of households by within household response, Wave 7

	General Population Sample	BHPS	Ethnic Minority Boost	Immigrant and Ethnic Minority Boost	All
Complete response: all eligible adults interviewed	77.82	80.44	66.38	70.42	76.91
All eligible adults interview or proxy	9.37	8.63	15.24	6.71	9.45
At least one within household refusal	12.8	10.93	18.37	22.88	13.64

There are two approaches to the analysis of household income in the presence of incomplete household response. The first is restrict analysis to complete only, and possibly reweight for the biases which may result from the selection. The second is to impute incomes for the proxy and within-household non-respondents and create a total household income based on these imputations. UKHLS data production processes support the second of these approaches, as discussed below in the section on imputation. As indicated above, the incomplete households have different characteristics from those in which all members respond, and it is recommend that if the analysis of household income is restricted to complete households then results should be reweighted. A specific weight for this purpose is not presently provided.

1.5.3. ITEM NON-RESPONSE

Average rates of item non-response are relatively low with around 1% of items missing on average in the individual interview administered schedule. However income questions are often seen by respondents as sensitive or in some instances may be asking for information the respondent does not necessarily know without referring to documents. As a result item non response rates tend to substantially higher than the average level. This is shown in table 5 below (data drawn from Lynn and Knies, 2016) which shows money related questions. There is some tendency for the rates of item non-response to drop as participants continue in the survey for longer. Rates are broadly comparable with those experienced in the BHPS. The very high item non response for self-employed profit may be noted. This is found in many other surveys.

Table 5: Item non response rates for some money related questions

	wave				
Description	1	2	3	4	5
Gross pay last payment	15.1	11.7	11.8	11	11.7
Usual pay	8.8	9.1	11.1	9.5	11.2
Self-employed net profit in last yearly account	46.9	41.8	41	38.5	39.3
Net amount of last rent payment	6	4.7	4.8	4.3	4.2
Value of property: home owners	8.9	7.5	7.6	6.3	6

1.6. WEIGHTING AND USE OF THE DIFFERENT SAMPLES

UKHLS has a complex sample design and is used in various ways by data analysts. Consequently the weighting strategy is also complex. *UKHLS* provides weights for the household and individual levels, including all individuals enumerated in respondent households, and eligible individuals that respond or do not respond to different instruments, e.g. the adult questionnaire, the self-completion instrument, for responding to different combinations of study waves, and for the diverse sample components.

In general, weights are the product of a design weight to convey the probability of selection, adjustment for non-response, and sometimes post-stratification, to make the distribution a closer match to the population distribution.

Units in the major sample components have different probabilities of selection. For example, the members of different ethnic minority groups in the boost sample have different probabilities of selection. In addition, the countries in the former BHPS sample have different sampling fractions, including boost samples for Scotland, Wales and Northern Ireland. Different weights may also be used for analyses which combine the sampling components. For example, when combining the general population component with the former BHPS, the weights adjust for the fact that the BHPS sample does not contain immigrants for its period of fieldwork.

The development of weights also takes the time pattern of response into consideration. For example, weights for complete longitudinal responses are produced. These take into account differential probabilities of attrition after wave 1. They would include those for Waves 1 and 2 or Waves 1, 2, and 3. Cross-sectional weights and weights for single year samples waves are also produced. This brief summary of the weighting strategy can be supplemented by Lynn and Kaminska (2010). The User Manual (Knies, 2017) provides more detail on the specific weights that have been produced.

Longitudinal income analysis typically uses one of the adult individual longitudinal weights, though if the concern is with the whole population including children, the enumerated individual weights are used. If a population level measure is required, for example a relative poverty line based on 60% of median equivalised income, the enumerated individual weights would be used to compute median income. Jenkins (2011, 91-92) provides a further discussion of the use of weights in income analysis.

2. Income Data in the UKHLS

In this section we present an overview of the income data collected in UKHLS, including the approach to collection and the target measures, and the way we derive measures from data collected in the questionnaire

In the following three sections we discuss aspects of this in more detail, looking firstly at the imputation of missing income data, then at the approach to the estimation of net or disposable household after taxes, national insurance and other deductions. Finally we provide some comparisons between UKHLS income measures and measures based on the FRS and HBAI.

2.1. Overview – goals of income collection

Given the importance of longitudinal income analysis within the household panel study research agenda a significant proportion of questionnaire time is devoted to collection of income measures. This is still less than would be devoted in surveys which specialise more on income, such as household budget surveys, so some compromises have to be made.

Good measure of individual and household income is based on asking about each of the separate sources of income received, rather than asking for a global figure, which respondents will often not know and will report with significant error. The approach of asking for individual income sources is reinforced by the need to support analysis of separate income sources. There are significant research agendas around using data on earnings, state benefit receipts, pension and other sources of income.

The collection of income from different sources also permits the construction of a range of different income measures. The UKHLS data focuses on two key measures 'gross' household income and 'net' household income after deduction of taxes etc. (see section 4). However other measures can be computed, for example 'original' or 'market' income before the addition of state benefits.

In UK income research and analysis it is much more common to use a measure of current income than annual income, which is more often used in other countries. Current income measurement is based on taking the last receipt from regular payments, calibrated to a standard time metric (in the UKHLS case the month). For employee earnings a usual amount from that employer is used if the last receipt was unusual in some way. Where income is received more normally on an annual basis (e.g. income from savings and investments) it is asked on this basis and converted back to monthly. It should be noted that some of the income patterns which justify annual measures (e.g. 13^{th} month salary payments and annual bonuses) are rather uncommon in the UK.

In BHPS both annual and current measures were produced. Greater pressure on questionnaire space in the UKHLS meant that it was not possible retain the questions required for both measures and the choice was made to prioritise current income, so that currently there are no annual measures of household income in UKHLS data.

The UKHLS collects detailed information each wave on personal income. All individuals aged 16 or more are asked to report:

- wages from main job, including gross amount and amount received net of taxes and other deductions,
- self-employment earnings: this is asked net or gross as the respondent is best able to answer and also whether taxes and national insurance has been deducted,
- second job earnings gross of any deductions,
- interest and dividends,
- pensions (National Insurance/state retirement pension, pension from a previous employer, pension from a spouse's previous employer, private pension/annuity, widow's or war widow's pension, widowed mother's allowance or widowed pension),
- benefits (severe disablement allowance, disability living allowance, war disablement pension, attendance allowance, carer's allowance, incapacity benefit, income support, job seeker's allowance, national insurance credits, child benefit, child tax credit, working tax credit, maternity allowance, housing benefit, council tax benefit, foster allowance/guardian allowance/rent rebate, rate rebate, employment and support allowance, respond to work credit, sickness and accident insurance, in-work credit for lone parents and pension credit, any other benefit) and
- other income sources (educational grant, trade union and friendly society payment, maintenance or alimony, payments from a family member not living together, amount for rent from boarders or lodgers, rent from any other property and any other income source not asked about separately).

These personal income variables can be summed to obtain the total personal income. Total household income can be computed from the personal total incomes of all household members.

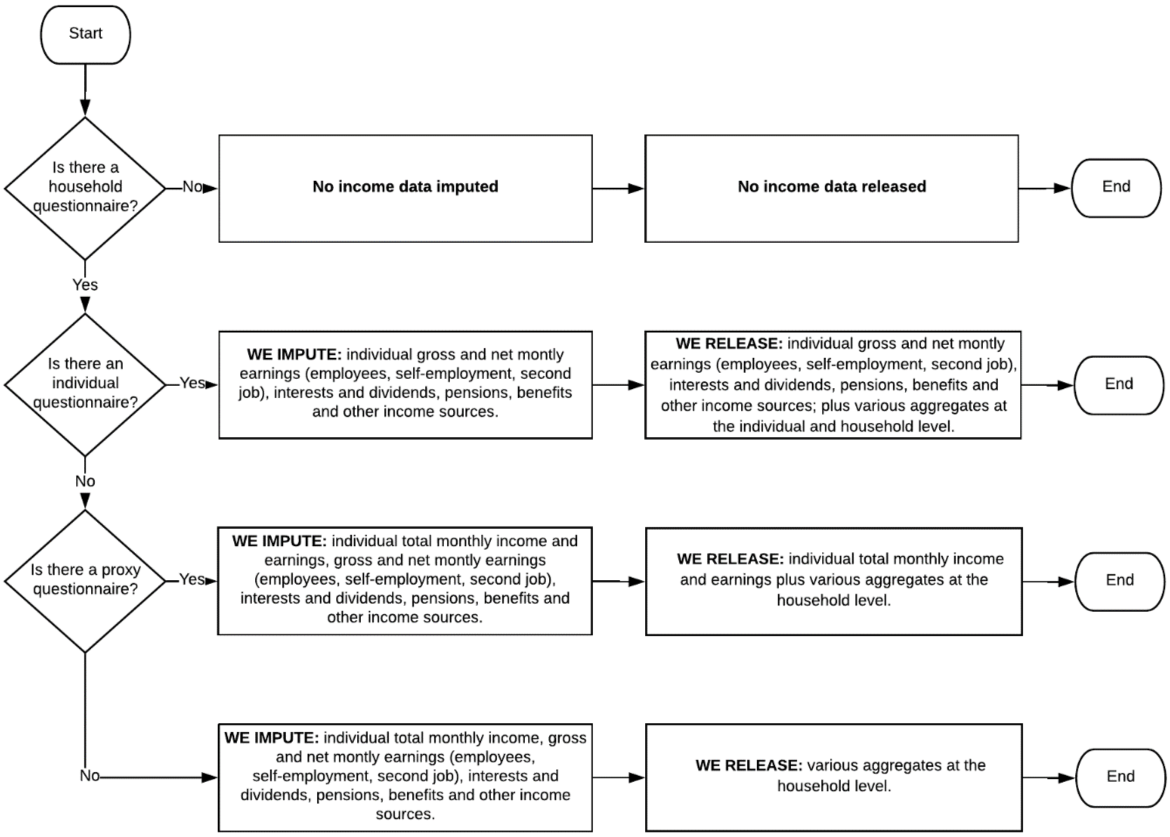
2.2. WHAT WE DERIVE AND IMPUTE

A range of derived variables are produced. Since in general the questionnaire will ask about last amount received and the period that this covers, there will be a derived variable containing the monthly amount associated with each of the income questions. In addition to the basic individual source derived variables we include a set of measures of personal income and household income. We provide these derived variables as well as net income estimates, discussed below, as part of the released data files. We also provide derived variables relating to housing costs so that measures of income after housing costs can be computed.

Information about income can be missing due to unit or item non response. We have unit non response when the whole individual questionnaire is missing for a sample member. This is the case of sample members in non-responding households (where the household questionnaire is also missing), or the case of non-respondents in responding households. We have item non response where the individual questionnaire is available, but some items are missing. This is the case of proxy and individual respondents in responding households who failed to provide a valid answer to some questions.[2] In our case, for example, information on income may be missing even when the individual or the proxy questionnaires are available. As indicated above, the level of item non response on income questions is rather high. Given the number of income sources that individuals receive a complete case analysis based only on respondent with no missing data on income sources would tend to drop a rather high proportion of cases. Thus in common with most other income surveys and other household panel studies we impute for item non-response.

Figure 1 shows how missing income information arises, as well as the type of income information we impute and release. Table 6 shows the income variables available in Understanding society. The top panel presents the single income components. The middle and the bottom panel present income aggregates at the individual and the household level, respectively. For each income variable we indicate whether any imputation has been released, the flag for the imputed cases, and the type of respondents (i.e., individual respondents, proxy respondents, individual non-respondents in responding households) each variable is available for.

Figure 1. What is imputed and released.



[2] Proxy respondents are cases where some information about the individual is reported by somebody else in the household. Proxy interviews are much shorter than individual interviews. For example, information about income is collected in much less detail.

Table 6. Income components, income aggregates and their availability for type of respondent

Variable name	Label	Data file	Imputation	Imputation flag	Available for: Individual respondents	Available for: Proxy respondents	Available for: Individual non respondents in responding households
Individual income variables							
payg_dv	gross pay per month in current job: last payment	indresp	No	x	Yes	No	No
payn_dv	net pay per month in current job: last payment	indresp	No	x	Yes	No	No
payu_dv	usual pay per month if different from last	indresp	No	x	Yes	No	No
paygu_dv	usual gross pay per month: current job	indresp	Yes	paygu_if	Yes	No	No
paynu_dv	usual net pay per month: current job	indresp	Yes	paynu_if	Yes	No	No
seearngrs_dv	self employment earnings - gross	indresp	Yes	seearngrs_if	Yes	No	No
seearnnet_dv	self employment earnings - net	indresp	Yes (indirectly)	seearngrs_if	Yes	No	No
fiyrinvinc_dv	income from savings and investments, annual	indresp	Yes	fiyrinvinc_if	Yes	Yes	No
j2pay_dv	pay in second job	indresp	Yes	j2pay_if	Yes	No	No
frmnth_dv	monthly income received from benefit/other miscellaneous income source	income	No	x	Yes	No	No
frmnthimp_dv	Total income from benefit/other miscellaneous income source, including imputed	income	Yes	frmnthimp_if	Yes	No	No
Individual income aggregates							
fimnlabnet_dv	amount income component 1: net labour income	indresp	Yes	x	Yes	Yes	No
j2paynet_dv	amount income component 1c: net earnings second job	indresp	Yes	x	Yes	Yes	No
fimnmisc_dv	amount income component 2: miscellaneous income	indresp	Yes	x	Yes	Yes	No
fimnprben_dv	amount income component 3: private benefit income	indresp	Yes	x	Yes	Yes	No
fimninvnet_dv	amount income component 5: investment income	indresp	Yes	x	Yes	Yes	No
fimnpen_dv	amount income component 6: pension income	indresp	Yes	x	Yes	Yes	No
fimnsben_dv	amount income component 7: social benefit income	indresp	Yes	x	Yes	Yes	No
fimnnet_dv	total net personal income - no deductions	indresp	Yes	x	Yes	Yes	No
fimnlabgrs_dv	total monthly labour income gross	indresp	Yes	fimnlabgrs_if	Yes	Yes	No
fimngrs_dv	total monthly personal income gross	indresp	Yes	fimngrs_if	Yes	Yes	No
fibenothr_dv	Total income from benefits and other sources	indresp	Yes	fibenothr_if	Yes	No	No
Household income aggregates							
fihhmnnet1_dv	total household net income - no deductions	indresp	Yes (indirectly)	x	Yes	Yes	Yes
fihhmnlabgrs_dv	total gross household labour income: month before interview	indresp	Yes	x	Yes	Yes	Yes
fihhmnlabnet_dv	total net household labour income: month before interview	indresp	Yes	x	Yes	Yes	Yes
fihhmngrs_dv	gross household income: month before interview	indresp	Yes	fihhmngrs_if	Yes	Yes	Yes
fihhmnmisc_dv	total household miscellaneous income: month before interview	indresp	Yes	x	Yes	Yes	Yes
fihhmnprben_dv	total household private benefit income: month before interview	indresp	Yes	x	Yes	Yes	Yes
fihhmninv_dv	total household investment income: month before interview	indresp	Yes	x	Yes	Yes	Yes
fihhmnpen_dv	total household pension income: month before interview	indresp	Yes	x	Yes	Yes	Yes
fihhmnsben_dv	total household social benefit income: month before interview	indresp	Yes	x	Yes	Yes	Yes

Note: The variables frmmth_dv and frmmthimp_dv collects information on the income sources benefit and other income sources listed above

3. IMPUTATION OF MISSING DATA

3.1. A TWO-PHASE PROCESS

The imputation of missing income data is done in two steps. The first step is the “initialisation”: missing values are imputed using cross-sectional imputation (deterministic imputation, stochastic imputation, stochastic imputation via chained equations, and hot deck) and longitudinal imputation (carryover methods). This permits to obtain a rectangular dataset where all individuals in responding households have a reported or imputed value for each income source. In the second step, the values imputed in the first stage are replaced using the longitudinal imputation method by Little and Su (1989). The Little and Su method is widely used in other major longitudinal surveys, such as the Australian household panel- HILDA-, the Longitudinal Study of Australian Children –LSAC-, the German Socio Economic Panel –SOEP- and the Swiss Household Panel –SHP- (see: Frick and Grabka, 2004, 2007; Hayes and Watson, 2009; Starick and Watson, 2009; Lipps, 2010; Mullan et al., 2015). The imputation process is summarised in Figure 2.

For individual respondents in responding households, the initialisation is carried out by year-wave using a range of parametric and semi-parametric methods. Parametric methods are: linear regression (for continuous variables), interval regression (for continuous censored variables), logistic regression (for binary variables), ordered logistic regression (for ordered variables), multinomial logistic regression (for non-ordered categorical variables). The semi-parametric method used is predictive mean matching imputation (PMM). The income variables and their predictors are generally imputed jointly, using chained equations (ICE). A summary of the characteristics of the initialisation process for responding individuals is reported in table 7.

Figure 2. Imputation summary

Note: pink boxes indicate cross sectional imputation, while blue boxes indicate longitudinal imputation

Table 7. initialisation for responding individuals

Variable name	Label	Imputation method	Predictors (regressors)
paygu_dv	usual gross pay per month: current job	linear regression, through chained equation	Education, region, sex, age (+square), number of children (+square),marital status, ethnicity, health, whether the job is permanent or temporary, job hours (log), size of company, soc classification (check), managerial duties, earnings second job
seearngrs_dv	self employment earnings - gross	linear regression, through chained equation	Education, region, sex, age (+square), number of children (+square),marital status, ethnicity, health, whether the job is permanent or temporary, job hours (log), size of company, soc classification (check), managerial duties, earnings second job
j2pay_dv	pay in second job	linear regression, through chained equation	Education, region, sex, age (+square), number of children (+square),marital status, ethnicity, health, whether the job is permanent or temporary, job hours (log), size of company, soc classification (check), managerial duties, earnings from first job (either from self employment of as an employee)
fiyrinvinc_dv	income from savings and investments, annual	Whether pidp has investment income: predictive mean matching	Education, region, sex, age (+square), number of children (+square),marital status, ethnicity, whether is an employee, whether is a self employed, financial situation
		Amount of investment income when positive (in logs): interval regression	Education,, region, sex, age (+square), number of children (+square),marital status, ethnicity, whether has a job, whether the first job is temporary or permanent, job hours, company size, soc code, earnings from first and second job, whether has a second job, financial situation
frmnth_dv	monthly income received from benefit/other miscellaneous income sources	Predictive mean matching or interval regression when missing is only partial	Education, region, sex, age (+square), number of chidren (+square), marital status, ethnicity, health, financial situation, sum of earnings and investment income

For proxy respondents and non-respondents in responding households, the initialisation is carried out using longitudinal carryover methods and hot-deck. Carryover methods are used to impute employment status. Income sources are imputed by hot-deck. Missing values in the variables defining the categories for the hot deck –other than employment status- are set equal to their median. Note that the proxy questionnaire only collects information on total personal income and earnings reported in bands. There is no income information for individual non respondents. Therefore, neither proxy respondents nor individual non respondents have any reported income sources. This means that all income sources need to be imputed. The hot deck method permits to impute all income sources by taking them from the same donor. For proxy respondents, this makes sure imputed income sources are coherent with the reported bounds. A summary of the characteristics of the initialisation process for proxy respondents and non respondents in responding households is reported in table 8.

Table 8. initialisation for proxy respondents and individual non-respondents in responding households

Variable name	Label	Proxy respondents: Imputation method	Proxy respondents: Perdictors (categories)	Non respondents: Imputation method	Non respondents: Perdictors (categories)
paygu_dv	usual gross pay per month: current job	Hot deck	Total income bands, total earnings bands, year, wave, sex, age, education employment status, sample origin, marital status, whether a parent, housing tenure, health, number of durables	Hot deck	Year, wave, sex, age, sample origin, marital status,whether a parent, housing tenure, number of durables.
seearngrs_dv	self employment earnings - gross				
j2pay_dv	pay in second job				
fiyrinvinc_dv	income from savings and investments, annual				
frmnthimp_dv	Total income from benefit/other miscellaneous income				

Note: When a donor matching the full set of categories is not found, categories are made coarser, or, ultimately, removed from the list of predictors.

3.2. IMPUTATION PROCEDURES

Imputation by chained equations (ICE)

It is a multivariate stochastic imputation method used to impute a set of variables jointly.[3] ICE allows for interdependence between the imputed variables by estimating each variable sequentially (see van Buuren et al., 1999, and Ragunathan et al., 2001). The ICE method has been used in major household panel surveys such as the ECHP, as well in combination with the Little and Su method (Westermeier and Grabka, 2016).

ICE starts by considering the following recursive (triangular) system of imputation equations:

$$\begin{cases} Y_1 = \alpha_{10} + X\beta_1 + u_1 \\ Y_2 = \alpha_{20} + X\beta_2 + \alpha_{21}Y_1 + u_2 \\ Y_3 = \alpha_{30} + X\beta_3 + \alpha_{31}Y_1 + \alpha_{32}Y_2 + u_3 \\ \vdots \\ Y_k = \alpha_{k0} + X\beta_k + \alpha_{k1}Y_1 + \alpha_{k2}Y_2 + \cdots + \alpha_{kk-1}Y_{k-1} + u_k \end{cases}$$

[3] For more details about stochastic imputation, see Rubin (1987), Schafer (1997), and Kenward and Carpenter (2007).

$Y_1,\ Y_2, \dots, Y_k$ are the income and auxiliary variables to be imputed, ordered from the one with the smallest percentage of missing values, Y_1, to the one with the largest percentage of missing values Y_k. X is a set of auxiliary variables observed for all individuals, $\alpha_{10}, \dots, \alpha_{ko}, \dots, \alpha_{kk-1}$ and $\beta_{10}, \dots, \beta_{ko}, \dots, \beta_{kk-1}$ are parameters, and $u_1,\ u_2, \dots, u_k$ are random errors. This recursive system allows us to impute each variable separately and sequentially in the following steps:

i. The first equation is estimated, and the missing values for Y_1 are imputed.
ii. The second equation is estimated by replacing the missing values of Y_1 with those imputed in the previous step. The missing values of Y_2 are imputed.
iii. The above steps are repeated sequentially for each of the remaining equations, until all missing values in $Y_1,\ Y_2, \dots, Y_k$ have been imputed.[4]

This sequential estimation is consistent only if the recursive system is valid. This is not necessarily a valid assumption. To address this problem, ICE uses the imputed values produced using the above recursive system as starting values in an iterative imputation process. In other words, the starting values are used to begin a new cycle of imputations where each equation is estimated sequentially, using as explanatory variables both X and the imputed variables $\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_k$, except the one used as dependent variable. At the end of this new cycle, a new set of imputed variables is produced and used to begin a further cycle of imputations. These cycles of imputations are repeated until convergence.

Predictive mean matching (PMM)

Predictive mean matching is a semi-parametric imputation method. For a given variable, PMM replaces missing values with observed values from a donor, i.e. a respondent with non-missing information on the variable of interest (see also Little, 1988). This is done in four steps:

i. Regression models for the variable to be imputed are estimated
ii. Fitted values are produced
iii. Records with missing information (recipients) are matched to donors based on the fitted values computed in ii)
iv. Missing values are replaced with observed values from donors.

Hotdeck (HD)

Hot deck is a non parametric imputation method. For individuals with missing information, the hotdeck method identifies suitable donors within imputation classes. Characteristics reported in the data associated with the missing information are used to define imputation classes (see table 3). Once a suitable donor is identified, information on all income sources are carried over from the donor.

[4] Imputed values are drawn from the posterior predictive distribution of the variable to be imputed, conditional on the observed data. Note that not all the X and the Y predictors need to be included in every equation (see table 7 for the list of the predictors used in the imputation of our income data).

Population Carryover (PC)

PC is a longitudinal imputation method. It uses data from adjoining waves to replace missing wave information. With only one adjoining wave of non-missing data, the information is carried-over with probability one. When two waves of adjoining information are available, the information carried-over is chosen based on proportions reported in the non-missing population

Little and Su (LS)

The Little and Su method is a longitudinal non parametric imputation method. The LS method imputes missing values using a multiplicative model (see Little and Su 1989). The final imputation is the product of 3 terms: a trend effect across waves (column effect), the recipient's departure from the trend (row effect), and a residual effect donated from another respondent with complete information for the corresponding income component.

In Understanding Society, we initialise the Little and Su by using a combination of cross sectional and longitudinal methods to impute all income sources for all individuals in a responding household (see previous sections). Then, the LS method is applied as follows.

First, the data are put in wide form. Table 9 shows a simple hypothetical case of earnings data after the initialisation but before going through the LS imputation.

Table 9. Example of earnings data after the initialisation phase

	Earnings		
pidp	w1	w2	w3
1	135	130	200
2	200	200 *	200 *
3	200	480	210
4	380 *	400	420
5	350	370	300 *
6	235	243	342
7	400 *	450	470
8	0	790	790
9	0 *	790	790
10	360	450	600
11	675	235	700
12	345	690	800
13	.	230 *	0
14	0	230 *	0

Note: *indicates imputed data. The zeros indicate cases where the individual was not applicable for earning data, for example due to unemployment.

Note cases of interest 9, 13, and 14 which correspond to an imputed inapplicable, a temporary sample member (ineligible at wave 1), and a respondent applicable only in one wave.

Second, the column effects are calculated. In our case, column effects are cross-sectional mean earnings at wave x as a proportion of mean earnings across all waves. For example, a column effect of 0.7 at wave 1 means that wave 1 mean income is 70% of overall mean income.

In understanding Society, column effects are calculated based on cases that are applicable in all waves (either reported or in the initial imputes). In our case, pids 8, 9, 14 are excluded as inapplicable in some waves, while 13 is excluded as it is ineligible at wave 1.

Third, row effects are calculated. Row effects are mean earnings for each individual where each reported value is scaled by the individual's column effect. When calculating the row effects, we make use of the imputes obtained in the initialisation phase. Waves in which a respondent is inapplicable are not included in the calculation of the row effects. For example, in the case of pids 8 and 9, only the –rescaled – non-zero values at waves 2 and 3 are used. Table 10 shows row and column effects for the data in table 9.

Table 5. Example of earning data: row and column effects.

	Earnings			
pidp	w1	w2	w3	Row effect
1	135	130	200	153.8
2	200	200 *	200 *	202.2
3	200	480	210	300.4
4	380 *	400	420	402.8
5	350	370	300 *	346.1
6	235	243	342	271.7
7	400 *	450	470	442
8	0	790	790	749.9
9	0 *	790	790	749.9
10	360	450	600	464.9
11	675	235	700	540.2
12	345	690	800	599.4
13	.	230 *	0	234.7
14	0	230 *	0	234.7
Sum	3280	3648	4242	
Column Effect	0.8809311	0.97977	1.1393	

Note: the mean of row effects is 3723.3

Fourth, data are sorted by row effects.

Fifth, imputed data from the initialisation process are substituted with LS imputes. This is done as follows. For each pidp with imputed income data, we identify the closest donor, i.e. the individual with the closest row effect. Then, we replace all imputed data with the corresponding donor value multiplied by the ratio of the recipient-donor row effects. Note that a single donor is used for each pidp with earnings missing at least one wave. Donors should therefore be applicable and non-imputed in all waves for which they are donating (see cases 13, 14, 8, 9 who do not qualify as donors). Table 11 shows LS imputes for the data in table 9.

Some cases do not get a LS impute. They are:

i. The non applicable cases
ii. Individuals applicable at only one wave (cases 13, 14 and 9).
iii. Cases for which the row effect of the donor and the row effect of the recipient differ by more than 10%.
iv. Negative incomes/self employed incomes
v. Some benefit income

Table 11. Example of earning data: row and column effects, plus LS imputes.

	Earnings						
pidp	w1		w2		w3		Row effect
pidp	w1		w2		w3		
1	135		130		200		153.8
2	200		170.912	*	262.94	*	202.2
13	.		230	*	0		234.7
14	0		230	*	0		234.7
6	235		243		342		271.7
3	200		480		210		300.4
5	350		370		241.92	*	346.1
4	407.31	*	400		420		402.8
7	342.27	*	450		470		442
10	360		450		600		464.9
11	675		235		700		540.2
12	345		690		800		599.4
8	0		790		790		749.9
9	0	*	790		790		749.9
Column Effect	0.8809311		0.97977		1.1393		

Note: data are sorted by row effect.

3.3. Effect of imputation

This section presents descriptive statistics on the effect of the imputation on the main income variables. Table 12 shows the share of income data by decile for the main income components, i.e., monthly gross earnings from the first job (for employees and self employed, respectively), monthly gross earnings from second job, annual income from interests and dividends. Table 8 reports the same information for the main income aggregates, i.e. total individual income from benefits and other sources, total individual gross labour income, total individual gross income, and total household gross income. Finally, for the main income components, table 9 compares descriptive statistics for the imputed and the non imputed data.

Table 12: Share of missing data by wave and decile (main income components)

	Wave 2			Wave 3			Wave 4			Wave 5			Wave 6			Wave 7		
Decile	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**
Monthly gross earnings from first job (employees)																		
1	0.08	0.27	2475	0.08	0.27	2322	0.07	0.26	2152	0.08	0.27	2024	0.11	0.32	2013	0.09	0.29	1933
2	0.08	0.27	2484	0.08	0.27	2148	0.07	0.26	2102	0.07	0.26	2017	0.11	0.32	2034	0.09	0.28	1931
3	0.08	0.27	2548	0.08	0.27	2234	0.07	0.26	2071	0.07	0.26	2018	0.11	0.31	2064	0.08	0.27	1932
4	0.08	0.28	2390	0.07	0.25	2544	0.05	0.22	2161	0.07	0.25	2021	0.11	0.32	1960	0.07	0.26	2064
5	0.07	0.26	2521	0.08	0.27	1924	0.05	0.23	2054	0.06	0.23	2173	0.11	0.32	1994	0.09	0.29	1882
6	0.07	0.26	2442	0.06	0.25	2245	0.06	0.23	2173	0.06	0.24	1892	0.12	0.32	2015	0.07	0.25	2076
7	0.08	0.27	2485	0.08	0.27	2248	0.08	0.27	2043	0.07	0.26	1993	0.10	0.31	2011	0.09	0.28	1742
8	0.07	0.26	2501	0.07	0.26	2220	0.07	0.25	2179	0.07	0.26	2067	0.11	0.31	2034	0.08	0.28	1896
9	0.08	0.26	2421	0.07	0.26	2301	0.06	0.24	2064	0.07	0.25	1972	0.10	0.30	2002	0.08	0.27	1973
10	0.07	0.26	2474	0.07	0.25	2158	0.06	0.24	2081	0.06	0.25	2019	0.12	0.33	2003	0.07	0.25	1891
Monthly gross earnings from first job (self employed)																		
1	0.30	0.46	361	0.31	0.46	343	0.27	0.44	329	0.24	0.43	321	0.34	0.48	336	0.39	0.49	322
2	0.33	0.47	363	0.25	0.43	342	0.26	0.44	329	0.32	0.47	320	0.43	0.50	336	0.37	0.48	321
3	0.42	0.49	364	0.39	0.49	342	0.38	0.49	329	0.37	0.48	320	0.43	0.50	336	0.44	0.50	321
4	0.42	0.49	355	0.39	0.49	342	0.43	0.50	329	0.42	0.49	320	0.49	0.50	336	0.45	0.50	322
5	0.43	0.50	360	0.42	0.49	342	0.40	0.49	331	0.43	0.50	320	0.52	0.50	344	0.46	0.50	323
6	0.42	0.49	361	0.43	0.50	342	0.37	0.48	327	0.43	0.50	327	0.52	0.50	328	0.54	0.50	319
7	0.40	0.49	361	0.41	0.49	342	0.39	0.49	329	0.43	0.50	314	0.49	0.50	337	0.42	0.49	322
8	0.43	0.50	360	0.46	0.50	342	0.43	0.50	329	0.38	0.49	319	0.45	0.50	335	0.43	0.50	321
9	0.36	0.48	361	0.35	0.48	342	0.36	0.48	329	0.37	0.48	320	0.43	0.50	336	0.42	0.49	322
10	0.31	0.46	360	0.33	0.47	342	0.31	0.46	329	0.31	0.46	320	0.36	0.48	336	0.38	0.49	320
Monthly gross earnings from second job (self employed)																		
1	0.07	0.25	275	0.10	0.31	230	0.07	0.26	243	0.10	0.30	193	0.14	0.34	273	0.16	0.37	207
2	0.10	0.29	240	0.11	0.32	179	0.12	0.33	189	0.07	0.25	182	0.21	0.41	193	0.15	0.35	172
3	0.11	0.31	315	0.08	0.27	258	0.07	0.26	261	0.08	0.28	203	0.12	0.32	291	0.18	0.39	204
4	0.14	0.35	166	0.15	0.36	197	0.09	0.28	203	0.09	0.29	181	0.17	0.38	196	0.19	0.39	175
5	0.08	0.27	372	0.14	0.35	210	0.08	0.27	215	0.12	0.32	212	0.17	0.38	257	0.15	0.36	210
6	0.19	0.39	144	0.17	0.37	155	0.18	0.39	164	0.11	0.32	144	0.27	0.44	190	0.12	0.33	208
7	0.14	0.34	237	0.11	0.31	274	0.08	0.27	288	0.10	0.30	177	0.17	0.37	287	0.19	0.40	150
8	0.08	0.26	265	0.14	0.35	129	0.07	0.25	138	0.05	0.22	206	0.15	0.36	177	0.14	0.35	197
9	0.18	0.39	228	0.15	0.36	204	0.13	0.34	220	0.10	0.31	163	0.14	0.35	257	0.15	0.35	198
10	0.13	0.33	248	0.11	0.31	204	0.10	0.30	204	0.07	0.26	181	0.22	0.42	208	0.18	0.38	173
Annual income from interests and dividends																		
1	0.20	0.40	1530	0.19	0.39	1826	0.17	0.37	1922	0.15	0.36	1752	0.25	0.43	1567	0.29	0.46	1187
2	0.15	0.36	1683	0.24	0.42	926	0.26	0.44	940	0.24	0.43	841	0.38	0.49	1162	0.22	0.41	1186
3	0.33	0.47	1375	0.27	0.45	1875	0.28	0.45	1933	0.24	0.43	1591	0.28	0.45	1404	0.24	0.43	1425
4	0.26	0.44	1529	0.42	0.49	776	0.45	0.50	927	0.25	0.43	1624	0.30	0.46	1743	0.25	0.43	1491
5	0.17	0.38	1529	0.28	0.45	1416	0.26	0.44	1441	0.36	0.48	644	0.51	0.50	953	0.39	0.49	715
6	0.30	0.46	1815	0.23	0.42	1380	0.24	0.43	1419	0.32	0.47	1152	0.30	0.46	1352	0.30	0.46	1216
7	0.40	0.49	1244	0.38	0.49	1423	0.34	0.48	1514	0.31	0.46	1268	0.37	0.48	1378	0.30	0.46	1089
8	0.29	0.46	1534	0.33	0.47	1186	0.33	0.47	1378	0.26	0.44	1286	0.36	0.48	1360	0.31	0.46	1180
9	0.27	0.45	1675	0.29	0.46	1370	0.28	0.45	1439	0.27	0.44	1298	0.32	0.47	1353	0.29	0.45	1189
10	0.28	0.45	1378	0.30	0.46	1329	0.29	0.46	1390	0.27	0.44	1217	0.27	0.45	1363	0.25	0.43	1183

Note: For earnings from first job, the share of missing cases is computed for applicable cases only. For earnings from second job and earnings from interests and dividends, the share of missing cases is computed for positive cases only. IEMB and wave 1 are excluded.

Table 13: Share of missing data by wave and decile (main income aggregates)

	Wave 2			Wave 3			Wave 4			Wave 5			Wave 6			Wave 7		
Decile	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**	**Mean**	**sd**	**N**
Total individual income from benefits and other sources																		
1	0.10	0.29	3443	0.12	0.32	2980	0.12	0.32	2684	0.11	0.31	2473	0.15	0.36	2507	0.13	0.34	2383
2	0.10	0.27	2931	0.12	0.31	2797	0.11	0.29	2673	0.10	0.28	2490	0.14	0.33	2507	0.13	0.32	2336
3	0.11	0.28	3131	0.14	0.32	2875	0.13	0.31	2683	0.13	0.31	2465	0.14	0.32	2506	0.12	0.31	2359
4	0.13	0.30	3182	0.17	0.33	2883	0.15	0.32	2673	0.14	0.31	2465	0.16	0.34	2506	0.15	0.33	2360
5	0.14	0.30	3158	0.16	0.32	2881	0.15	0.31	2680	0.13	0.30	2476	0.16	0.33	2509	0.14	0.31	2359
6	0.14	0.29	3164	0.17	0.31	2883	0.15	0.30	2677	0.14	0.29	2468	0.17	0.32	2504	0.17	0.32	2359
7	0.16	0.29	3169	0.16	0.29	2883	0.16	0.29	2677	0.15	0.29	2473	0.18	0.32	2507	0.16	0.30	2362
8	0.16	0.28	3168	0.18	0.30	2883	0.16	0.28	2678	0.15	0.29	2472	0.19	0.32	2506	0.17	0.30	2357
9	0.18	0.30	3167	0.21	0.31	2883	0.19	0.31	2678	0.18	0.30	2473	0.20	0.32	2507	0.18	0.30	2359
10	0.22	0.34	3168	0.22	0.34	2883	0.20	0.32	2678	0.19	0.32	2472	0.19	0.32	2506	0.18	0.32	2359
Total individual gross labour income																		
1	0.20	0.40	3099	0.18	0.38	2856	0.19	0.39	2701	0.21	0.40	2611	0.25	0.43	2582	0.24	0.43	2445
2	0.19	0.39	3093	0.19	0.39	2788	0.21	0.41	2700	0.22	0.41	2613	0.25	0.43	2581	0.23	0.42	2438
3	0.20	0.40	3131	0.21	0.40	2822	0.20	0.40	2714	0.21	0.41	2801	0.25	0.43	2581	0.20	0.40	2442
4	0.19	0.39	3077	0.18	0.38	2898	0.19	0.39	2685	0.21	0.40	2369	0.22	0.41	2642	0.20	0.40	2441
5	0.17	0.37	3329	0.17	0.38	2748	0.16	0.37	2700	0.18	0.39	2598	0.22	0.41	2699	0.17	0.37	2515
6	0.18	0.38	2972	0.18	0.38	2827	0.17	0.38	2700	0.18	0.38	2598	0.23	0.42	2406	0.18	0.38	2387
7	0.18	0.39	3033	0.20	0.40	2844	0.20	0.40	2700	0.17	0.37	2599	0.21	0.40	2577	0.16	0.37	2423
8	0.17	0.37	3101	0.18	0.38	2790	0.16	0.37	2756	0.18	0.38	2602	0.20	0.40	2582	0.18	0.38	2441
9	0.17	0.37	3037	0.17	0.37	2822	0.17	0.38	2665	0.17	0.37	2594	0.21	0.40	2580	0.16	0.37	2442
10	0.20	0.40	3087	0.20	0.40	2821	0.19	0.39	2678	0.20	0.40	2598	0.24	0.42	2581	0.19	0.39	2441
Total individual gross income																		
1	0.19	0.39	5459	0.22	0.41	4975	0.22	0.41	4717	0.23	0.42	4492	0.21	0.41	4452	0.19	0.39	4191
2	0.21	0.39	5460	0.24	0.41	4973	0.24	0.41	4715	0.25	0.42	4489	0.27	0.43	4447	0.26	0.42	4191
3	0.23	0.39	5458	0.25	0.39	4985	0.25	0.40	4716	0.24	0.40	4490	0.28	0.42	4449	0.26	0.41	4189
4	0.21	0.36	5458	0.23	0.37	4963	0.22	0.37	4715	0.23	0.38	4492	0.25	0.39	4450	0.23	0.38	4197
5	0.20	0.35	5459	0.22	0.36	4974	0.20	0.35	4860	0.21	0.37	4489	0.23	0.38	4448	0.22	0.37	4185
6	0.20	0.36	5645	0.23	0.37	4981	0.21	0.36	4572	0.20	0.36	4544	0.24	0.39	4467	0.21	0.37	4189
7	0.22	0.37	5294	0.22	0.37	4967	0.22	0.38	4715	0.21	0.37	4437	0.24	0.39	4431	0.21	0.37	4190
8	0.23	0.39	5437	0.23	0.39	4974	0.22	0.38	4716	0.21	0.38	4491	0.23	0.39	4451	0.21	0.38	4191
9	0.22	0.39	5459	0.22	0.39	4974	0.22	0.39	4732	0.22	0.39	4489	0.25	0.41	4447	0.22	0.39	4193
10	0.25	0.42	5458	0.26	0.42	4973	0.24	0.41	4699	0.23	0.41	4490	0.26	0.42	4449	0.23	0.40	4186
Total household gross income																		
1	0.20	0.33	5460	0.21	0.33	4974	0.20	0.33	4717	0.19	0.32	4491	0.26	0.37	4450	0.23	0.35	4193
2	0.17	0.28	5459	0.19	0.29	4976	0.17	0.29	4715	0.17	0.28	4490	0.21	0.32	4450	0.19	0.30	4188
3	0.18	0.29	5459	0.19	0.29	4973	0.17	0.27	4717	0.17	0.27	4490	0.21	0.31	4448	0.19	0.30	4190
4	0.19	0.29	5458	0.19	0.29	4973	0.18	0.28	4714	0.18	0.28	4493	0.21	0.30	4449	0.18	0.29	4190
5	0.20	0.30	5458	0.20	0.30	4975	0.20	0.29	4716	0.19	0.29	4488	0.22	0.32	4450	0.21	0.31	4190
6	0.21	0.31	5459	0.21	0.31	4976	0.18	0.29	4717	0.18	0.29	4491	0.23	0.31	4448	0.20	0.30	4191
7	0.20	0.30	5458	0.21	0.30	4971	0.18	0.29	4715	0.19	0.30	4491	0.22	0.32	4450	0.21	0.30	4191
8	0.21	0.31	5460	0.22	0.31	4975	0.20	0.30	4716	0.19	0.30	4490	0.23	0.32	4451	0.21	0.31	4192
9	0.21	0.31	5458	0.21	0.31	4975	0.20	0.30	4715	0.18	0.29	4490	0.24	0.33	4448	0.23	0.31	4187
10	0.22	0.33	5458	0.22	0.32	4971	0.20	0.31	4715	0.18	0.31	4489	0.23	0.33	4447	0.22	0.32	4190

Note: For total individual income from benefits and other sources and for total household labour income, the share of missing cases is computed for positive cases only. IEMB and wave 1 are excluded.

Table 14: descriptive statistics of main income components, by imputation status and wave

Imputation status	mean	sd	min	max	p25	p50	p75	iqr	N	wave
Monthly gross earnings from first job (employees)										
Not imputed	1831.31	1459.09	0.08	15000.00	868.48	1500.00	2408.33	1539.85	22830	2
Imputed	1749.09	1348.71	1.00	12697.15	833.23	1450.12	2300.29	1467.06	1911	2
Not imputed	1863.08	1471.87	0.08	15000.00	881.00	1541.29	2500.00	1619.00	20682	3
Imputed	1772.50	1294.24	0.98	10003.92	840.10	1500.08	2400.05	1559.95	1662	3
Not imputed	1911.43	1523.72	0.08	15000.00	907.00	1591.00	2500.00	1593.00	19714	4
Imputed	1807.94	1303.00	27.14	9184.67	833.03	1583.25	2475.53	1642.50	1366	4
Not imputed	1948.09	1558.21	0.83	15000.00	925.47	1600.00	2500.33	1574.86	18818	5
Imputed	1870.78	1365.85	1.00	9421.15	866.65	1581.34	2507.36	1640.71	1378	5
Not imputed	2013.81	1610.19	0.08	15000.00	980.00	1650.00	2600.00	1620.00	16647	6
Imputed	1921.16	1496.26	34.66	13386.03	902.78	1575.36	2500.38	1597.60	1654	6
Not imputed	2053.99	1635.16	0.08	15000.00	1000.00	1680.49	2666.67	1666.67	16521	7
Imputed	1859.36	1381.26	24.35	12448.89	874.93	1584.66	2500.34	1625.41	1268	7
Monthly gross earnings from first job (self employed)										
Not imputed	1785.51	2672.53	-15473.92	15000.00	368.33	1002.79	2132.90	1764.57	2231	2
Imputed	1650.10	2331.80	-4298.95	15000.00	431.86	1015.12	1970.53	1538.67	1375	2
Not imputed	1772.01	2594.26	-8038.33	15000.00	390.00	966.06	2123.31	1733.31	2140	3
Imputed	1733.92	2326.33	-6297.96	15000.00	511.55	1104.89	2023.65	1512.09	1281	3
Not imputed	1737.45	2550.18	-8851.81	15000.00	382.31	923.10	2100.74	1718.43	2108	4
Imputed	1612.07	1988.89	-3957.16	15000.00	510.51	1003.24	1993.52	1483.01	1182	4
Not imputed	1754.30	2560.54	-17888.89	15000.00	386.45	914.09	2122.37	1735.92	2022	5
Imputed	1651.24	2281.84	-17536.53	15000.00	520.05	1004.15	1975.52	1455.47	1179	5
Not imputed	2058.11	3019.90	-4970.85	15000.00	418.07	1000.00	2333.33	1915.26	1745	6
Imputed	1799.46	2481.75	-1686.18	15000.00	520.73	1013.68	2024.17	1503.44	1236	6
Not imputed	1898.20	3038.83	-42903.73	15000.00	390.00	985.34	2153.93	1763.93	1697	7
Imputed	1870.02	2637.39	-1712.15	15000.00	483.80	1051.67	2110.27	1626.47	1202	7
Monthly gross earnings from second job										
Not imputed	501.45	1723.40	1.00	31000.00	80.00	200.00	400.00	320.00	2208	2
Imputed	625.11	1935.98	5.08	25075.36	100.00	202.33	500.00	400.00	282	2
Not imputed	500.74	1519.58	1.00	30000.00	90.00	200.00	450.00	360.00	1789	3
Imputed	387.09	742.73	0.97	10000.00	99.90	200.23	450.91	351.01	251	3
Not imputed	550.21	2693.91	1.00	75000.00	90.00	200.00	450.00	360.00	1918	4
Imputed	605.26	2362.44	4.01	29999.99	85.00	220.89	497.90	412.90	207	4
Not imputed	510.00	1693.53	1.00	33000.00	86.00	200.00	450.00	364.00	1679	5
Imputed	382.42	799.21	5.99	7474.12	81.62	189.85	378.64	297.02	163	5
Not imputed	530.36	1777.41	1.00	34000.00	80.00	200.00	475.00	395.00	1816	6
Imputed	403.78	629.88	10.00	6016.53	100.00	206.53	410.30	310.30	344	6
Not imputed	694.29	2568.83	1.00	42000.00	100.00	200.00	500.00	400.00	1487	7
Imputed	626.77	2521.84	3.00	29999.99	91.59	200.00	502.91	411.31	280	7
Annual income from interests and dividends										
Not imputed	1116.28	5233.05	1.00	180000.00	25.00	100.00	500.00	475.00	11302	2
Imputed	950.51	3495.94	0.11	106183.17	49.83	150.10	501.65	451.81	3990	2
Not imputed	1259.14	6483.29	1.00	180000.00	30.00	100.00	500.00	470.00	9667	3
Imputed	1090.51	5187.42	0.22	176999.73	44.99	150.23	500.93	455.94	3840	3
Not imputed	1248.67	5564.38	1.00	180000.00	30.00	120.00	500.00	470.00	10273	4
Imputed	1094.13	4333.88	0.05	146616.05	49.90	151.18	548.05	498.15	4030	4
Not imputed	1313.29	6431.45	1.00	180000.00	30.00	130.00	500.00	470.00	9415	5
Imputed	1003.64	3237.97	0.02	50741.13	49.97	199.85	600.73	550.76	3258	5
Not imputed	1546.22	6899.77	1.00	180000.00	40.00	190.00	600.00	560.00	8945	6
Imputed	942.10	3679.58	0.06	132377.27	49.61	179.97	500.52	450.91	4102	6
Not imputed	1608.17	7241.80	1.00	180000.00	37.00	150.00	600.00	563.00	8294	7
Imputed	1089.45	4011.64	0.02	95313.11	42.82	180.12	600.44	557.62	3142	7

Note: For earnings from first job, the statistics are computed for applicable cases only. For earnings from second job and earnings from interests and dividends, the share of missing cases is computed for positive cases only. IEMB and wave 1 are excluded.

4. DERIVATION OF NET HOUSEHOLD INCOME VARIABLES

Most analysis of standards of living, income dynamics and poverty and low income tend to use net or disposable income after taxes and other major deductions from income. This in effect is the income that people have available for consumption or saving. UKHLS has generally followed the approach used by the Department for Work Pensions (DWP) for their Households Below Average Income (HBAI)data sets in defining net income, subject to the limitations imposed by the UKHLS questionnaire[5]. There are some deductions from individual income and some incomes sources which are not available to us from the questionnaire. We work closely with staff from the DWP to check the consistency of the approach and to validate income derived variable estimates we produce. The DWP use *Understanding Society* data for the longitudinal component of UK statistics on income dynamics (DWP, 2018). The approach broadly follows that used Jenkins and his colleagues in constructing BHPS net incomes. Some differences are noted below.

In the UK there is also a distinction between incomes before and after housing costs. The UKHLS data contains housing cost measures, including imputed values where there is item non response.

In the discussion below we have included data set variable names for ease of reference to the documentation and data.

4.1.1. NET INCOME ESTIMATES

Individual income estimates are included in the individual level data files, ***w*_indresp** and the household-level income measures are included in the household level data files, ***w*_hhresp**.

At the individual level, the total estimated net monthly income is ***w*_fimnnet_dv** where "net" refers to net of taxes on earnings and national insurance contributions. It is constructed from the income components described below. The gross monthly income, ***w*_fimngrs_dv** is also estimated from individual income components described below except the earnings components are gross, that is, before taxes and National Insurance contributions are deducted. The associated imputation flag for both variables is ***w*_fimngrs_if**.

At the household level, total household gross income is included in the variable ***w*_fihhmngrs_dv**. This comprises imputed income from proxy and within-household non-respondents. The extent of imputation is indicated by the variable ***w*_fihhmngrs_if**. The calculation of housing costs (see below) implies that there is housing benefit implicitly reported in the rent information which has not been reported in the individual questionnaire. ***w*_fihhmngrs1_dv** includes an adjustment for this.

In addition to the summary variables described above the individual level data files also include estimates of the different income components, following the structure used by HBAI. These are as follows:

Component 1: Labour income (***w*_fimnlabnet_dv**)

[5] This means that deductions from earnings include contributions to occupational pension schemes. This would not be the case in all net income definitions.

This is the sum of three earnings components: net usual pay[6] (***w*_paynu_dv**); net self-employment income (***w*_seearnnet_dv**); net pay in second job (***w_j2paynet_dv***), which is gross pay in second job (***w_j2pay_dv***), less estimated tax and national insurance.

Component 2: Miscellaneous income (***w_fimnmisc_dv***)

This includes receipts reported in the income data file where ***w*_ficode** equals [24] "educational grant (not student loan or tuition fee loan)", [27] "payments from a family member not living here", or [38] "any other regular payment (not asked in Wave 1)". This is assumed to be reported net of tax.

Component 3: private benefit income (***w_fimnprben_dv***)

This includes receipts reported in the income data file where ***w*_ficode** equals [25] "trade union / friendly society payment", [26] "maintenance or alimony", or [35] "sickness and accident insurance". This is assumed to be reported net of tax.

Component 5: investment income (***w_fimninvnet_dv***)

This includes receipts reported in income record where ***w*_ficode** equals [4] "a private pension / annuity", [28] "rent from boarders or lodgers (not family members) living here", or [29] "rent from any other property". To this is added the monthly income from savings and investments, estimated as the annual income from savings and investments **(*w*_fiyrinvinc_dv),** divided by 12. All these sources are assumed to be reported net except for rent from other property which is assumed reported gross, and a tax liability is deducted.

Component 6: pension income (***w_fimnpen_dv***)

This includes receipts reported in the income data file where ***w*_ficode** equals [2] "a pension from a previous employer", or [3] "a pension from a spouse's previous employer". This is assumed to be reported net of tax.

Component 7: social benefit income (***w_fimnsben_dv***)

This includes receipts reported in income record where ***w*_ficode** equals [1] "state retirement (old age) pension", [5] "a widow's or war widow's pension", [6] "a widowed mother's allowance / widowed parent's allowance", [7] "pension credit (includes guarantee credit & saving credit)", [8] "severe disablement allowance", [9] "industrial injury disablement allowance", [10] "disability living allowance", [11] "attendance allowance", [12] "carer's allowance (formerly invalid care allowance)", [13] "war disablement pension", [14] "incapacity benefit", [15] "income support", [16] "job seeker's allowance", [18] "child benefit (including lone-parent child benefit payments)", [19] "child tax credit", [20] "working tax credit (includes disabled person's tax credit)", [21] "maternity allowance", [22] "housing benefit", [23] "council tax benefit (offset against council tax)", [30] "foster allowance / guardian allowance", [31] "rent rebate (NI only)", [32] "rate rebate (NI only – offset against rates)", [33] "employment and support allowance", [34] "return to work credit", [36] "in-work credit for lone parents", [37] "other disability related benefit or payment", [39] "income from any other state benefit (not asked in Wave 1). This is assumed to be reported net of tax.

At the household level, the following income variables are available:

[6] In UKHLS we include the value of net earnings as reported by the respondent where this is available and estimate net earnings using the rules of the tax and national insurance system as well as deducting occupational pension contributions only when net earnings are not reported. This is distinct from BHPS practice where deductions from gross earnings are estimated for all cases.

w_fihhmnnet1_dv is the net household monthly income. It is the sum of net monthly incomes from all household members (including proxies and within household non-respondents, see **w_fimnnet_dv**).

Local taxation liability is also estimated for Great Britain, though not currently Northern Ireland and the variable **w_fihhmnnet3_dv** is equal to **w_hhnetinc1** less council tax liability. Council tax liability, for most people, is equal to their estimated council tax, **w_ficountax_dv** (see below). Some people receive council tax benefit to help them pay their council tax. For these people council tax liability equals their estimated council tax minus their council tax benefit.

As indicated in the previous section income components are imputed for all proxy and withn household non-respondents. Hence net income estimates are included for all households. Users may decide to drop estimates based on such imputed data, but there is a very strong case for adjusting results to take account of the consequent sample selection.

4.1.2. HOUSING COSTS ESTIMATES

We provide derived variables for total housing costs including imputations in order to allow computation of income after housing cost measures. Derived housing cost variables cover renters and those paying mortgages.

For renters, w_rentg_dv is the computed monthly gross rent including any housing benefit received. It is equal to w_rent_dv where no housing benefit is received. Missing values are imputed, and where the participant reports 100% housing benefit the value is set equal to housing benefit reported in the individual questionnaire and a value imputed if not reported there. The variable w_rentg_if is an imputation flag for w_rentg_dv. In addition **w_hbadjust_dv** is an adjustment for total household income where housing benefit is implicitly reported in the difference between gross and net rent, but is not reported in the individual questionnaire.

For those paying mortgages **w_xpmg_dv** is monthly total mortgage payments including imputation for missing data on **w_xpmg**. The variable **w_xpmg_if** is the imputation flag for this variable. Most definitions of housing costs for purposes of measuring income after housing costs seek to exclude repayments of capital included in mortgage payments and only include interest payments. **w_xpmgint_dv** is the estimated interest within **w_xpmg_dv**. For short period mortgages it is based on data on current interest rates times the outstanding principal and for mortgages with more than two years to run based on a standard repayment mortgage formula.

In the imputation of rent and mortgage payment it is assumed that variations over time are small and where other reports at the same address are available, missing values are set equal to the median of these reports. Where no report at that address is available a single value is imputed on the basis of characteristics of the accommodation and household and applied to all relevant waves.

5. Comparison of Cross-sectional estimates with the Family Resources Survey/Households Below Average Income

Comparisons to a cross-sectional gold-standard

Currently there is no longitudinal counterpart with which to validate the UKHLS net income series.[7] Instead, we compare UKHLS estimates of the income distribution to those from a cross-sectional gold-standard. This is a useful quality check as longitudinal measures of change, such as income mobility and poverty transitions, are essentially formed from the difference between two cross-sectional estimates. Much of the comparisons that follow replicate those reported in Jenkins (2011) for the BHPS.

As the UKHLS net income series aims to replicate the Households Below Average Income (HBAI) series, our cross-sectional counterpart is the HBAI.[8] Moreover, the HBAI is the data source for official UK statistics on the income distribution. It is based on a specialist income survey (the Family Resources Survey) that 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. It is considered to be of high quality.

As the HBAI corresponds to a financial year (April to March) and a UKHLS wave to two calendar years, we pool two consecutive HBAI data sets when comparing to a single UKHLS wave. All figures are expressed in the 2016-17 prices using a bespoke monthly CPI price index produced by the Office for National Statistics. Household net income is equivalised using the modified-OECD scale[9] and the UKHLS amounts converted to a weekly equivalent. All figures are weighted to the UK adult population using the relevant weights. Due to a known issue with the UKHLS wave one income data (see Fisher (2016)), we exclude it from our validations. To maintain cross-sectional representativeness, we also remove the IEMB subsample that was added at wave six.

Estimates of selected quantiles of the income distribution are presented in figure 3. Reassuringly, the estimates from the two surveys line up closely and show a similar time series pattern. However, we do see systematic differences for the lower half of the distribution, although they are relatively small in magnitude. UKHLS, relative to HBAI, overestimates percentiles 1, 5, 10 and 25 by a small amount that is stable over time. The higher percentiles line up very closely indeed with the exception of the very richest households (p99) where the difference is fluctuating over time. The latter is consistent with the known difficulties in measuring the incomes of the very rich for which household surveys are not well suited.

We also compare estimates of inequality. Here, we trim the top and bottom 1 percent as measures of inequality maybe sensitive to outliers in the data. Figures 4-12 plot trends in percentile ratios (90-10, 90-50, 50-10 and 75-25), the standard UK poverty rate[10] and inequality indices (gini, theil, Atkinson, and GE2) for each data source. Looking across the measures, we see that measured inequality is similar in both data sets, although typically higher according to the HBAI. This reflects

[7] UKHLS respondents were asked to give consent to linked to administrative income sources. This will allow for longitudinal validations in the future.
[8] There are minor differences in definition across the two data sources. See previous section???
[9] The HBAI variable is re-normalized so that the scale rate equals one for a single-person household
[10] The share of individuals living in households with income 60% below the median

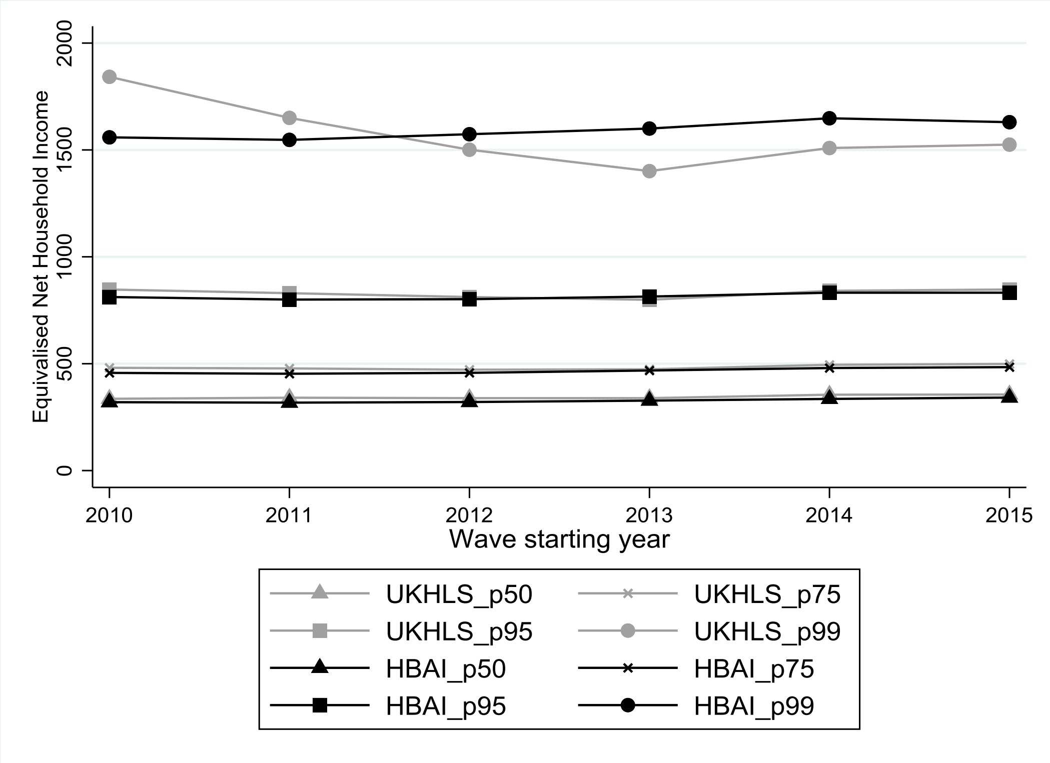
the differences at the bottom half of the distribution as above. The differences are typically larger for the percentile ratios and smaller for the other measures of inequality. We also observe a slight fall in inequality over the period in the UKHLS series but not in the HBAI one.

The comparisons so far suggest that the UKHLS measure of net income fares well at the population level. We now compare each survey at the level of population subgroups. Figure 13 examines the share of the population that falls into each of eight family types (pensioner couple, single female (male) pensioner, couple with (without) children, single with children, single female (male) without children). The series for the full population line up nicely but there are some differences. The family type breakdowns suggest that UKHLS over-estimates the proportion of pensioner couples at the population level (eg. 19.2% vs 16.9% in 2015) and conversely underestimates the share of couples without children (eg. 19.9 vs. 24.2 in 2015). Figures 14-18 replicate the family breakdowns but by income quintile. For example, for the bottom quintile (figure 14), UKHLS over-estimates the share of pensioners in particular single female pensioners (12.2 percent are seen in UKHLS vs. 10 percent in HBAI). Overestimation of single female pensioners in the bottom quintile was also found by Jenkins (2011) for the BHPS. The converse is that couples with children and without children are slightly under-represented in the bottom quintile (24.1 vs. 22 and 14.2 vs. 10.6, respectively).

We are also able to decompose household income into five subcomponents (figures 19-24: earnings, state benefits, occupation pensions, investments and other income). For this analysis, we work with (unequivalised) gross household income, as its subcomponents are directly comparable across data sets. Figures 19 and 20 show that the mean weekly income and earnings are close in both surveys. For example, for 2015 we see a mean total income of £860 in UKHLS and £842 in HBAI or a two percent difference. The corresponding figures for earnings are £623 and £629, less than a one percent difference. However, UKHLS captures more state benefit income relative to HBAI (£141 vs. £116 in 2015 or 22 percent). Therefore, UKHLS underestimates the share of earnings in total income relative to HBAI and overestimates the share of state benefit income (figure 25). Differences across the other sources, as a proportion of household income, are typically small. Figures 26-30 present the decompositions for each quintile separately and the broad pattern seen at the population level is seen in each quintile.

Put together, we conclude that the UKHLS provides reliable estimates of the income distribution.

Figure 3. Selected quantiles of net income



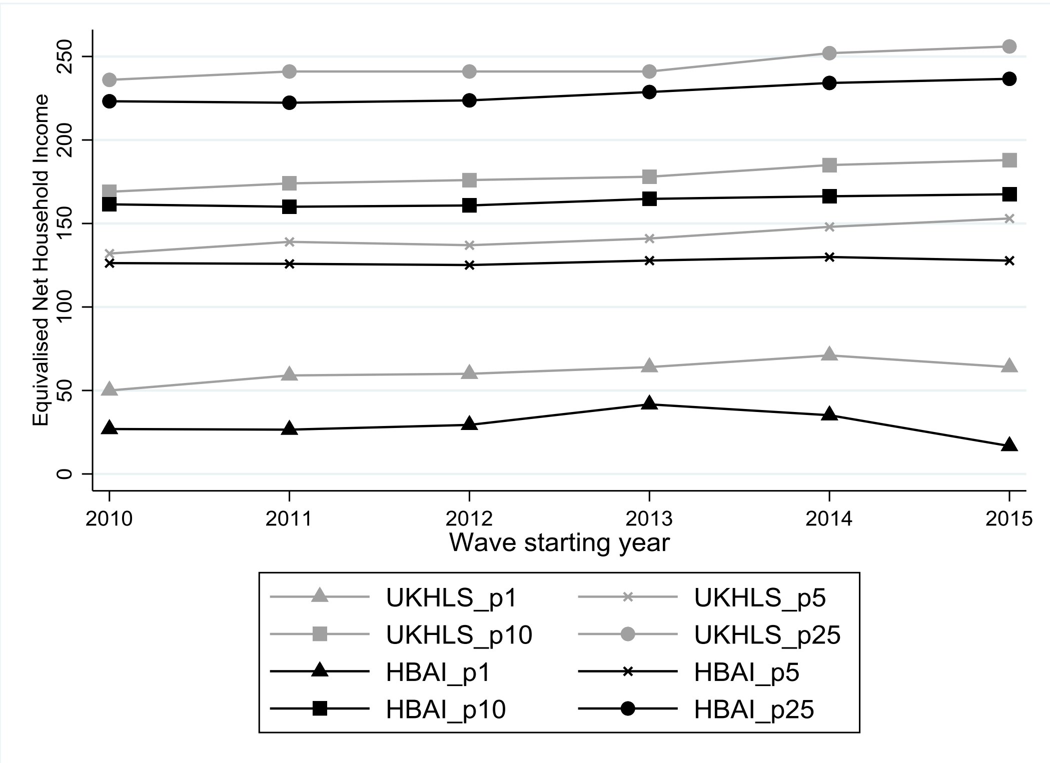


Figure 4. P90/P10

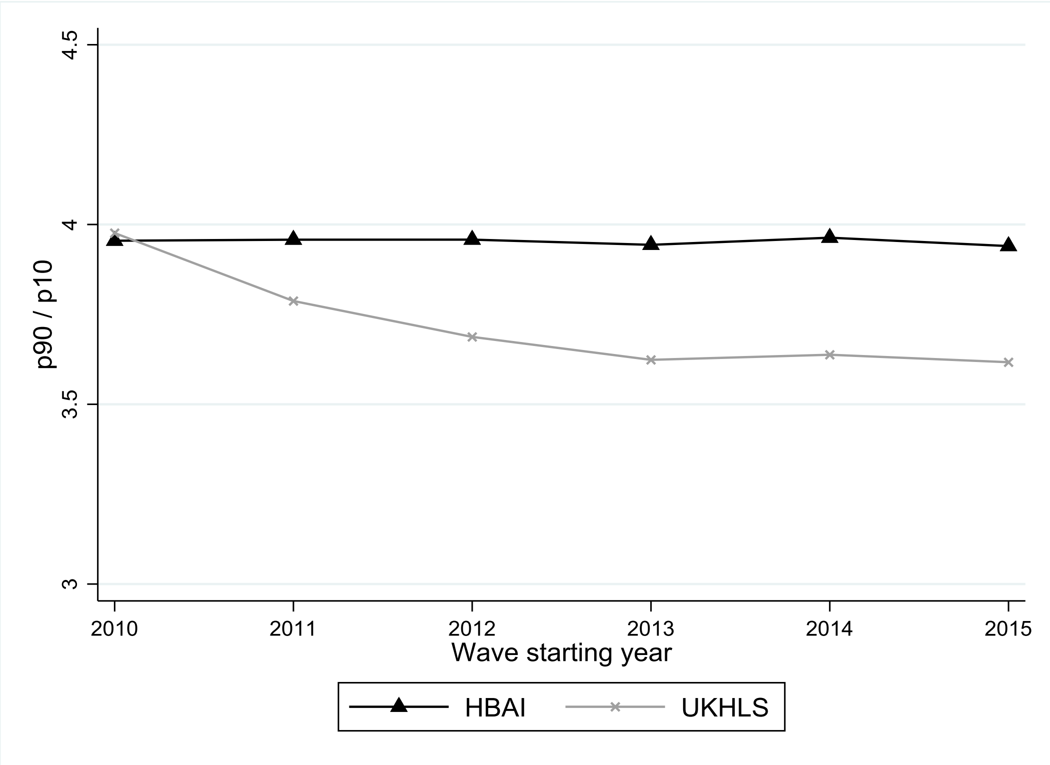


Figure 5. P90/P50

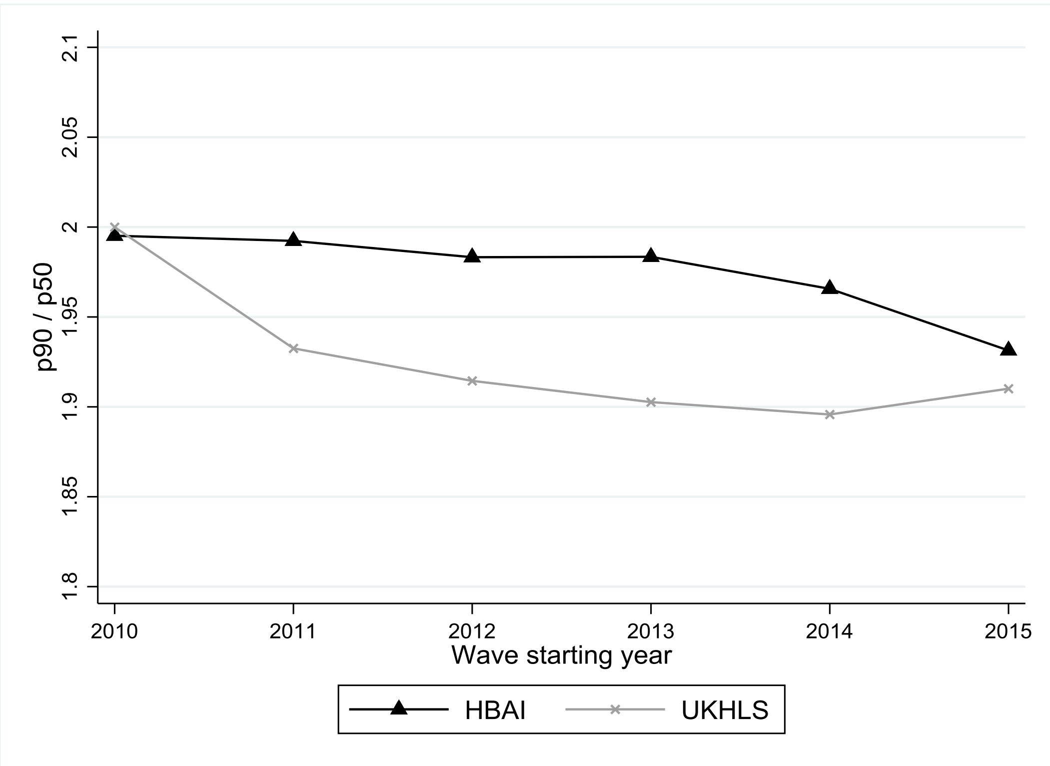


Figure 6. P50/P10

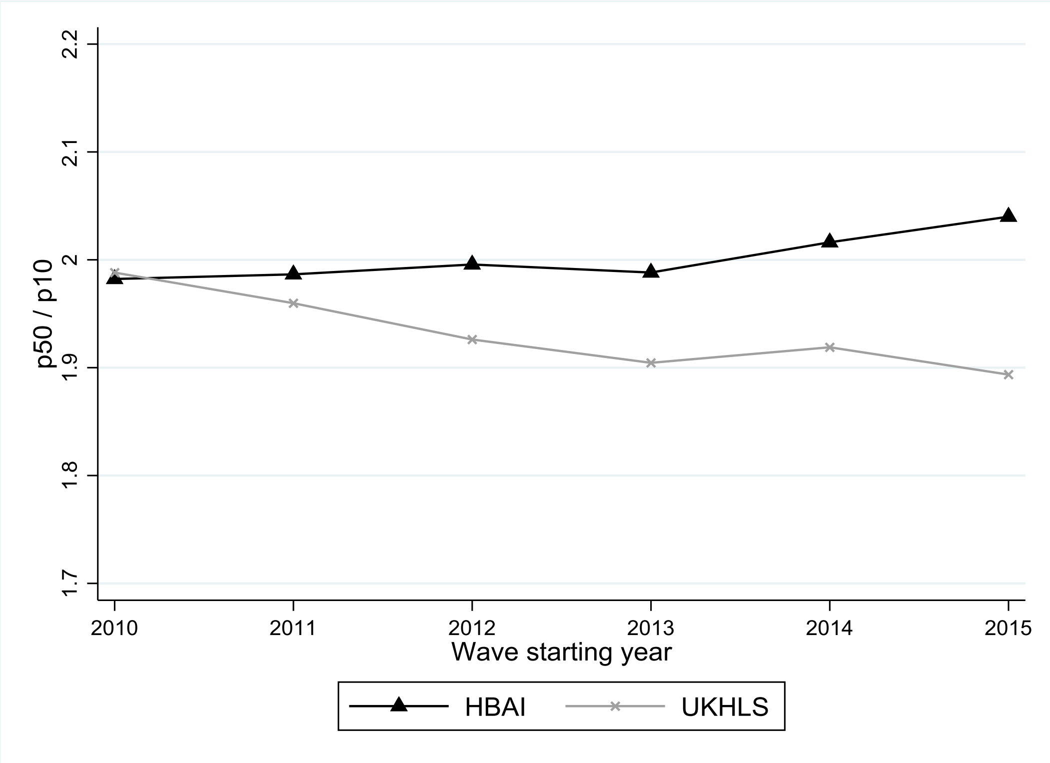


Figure 7. P75/P25

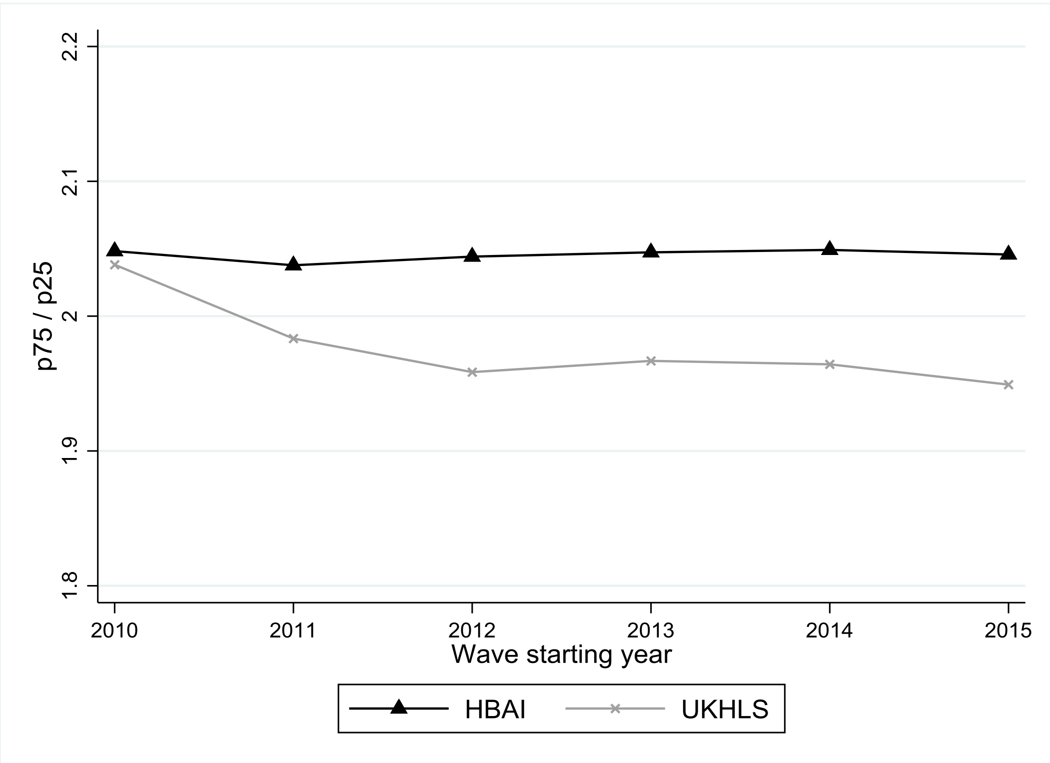


Figure 8. Share with income 60% below median

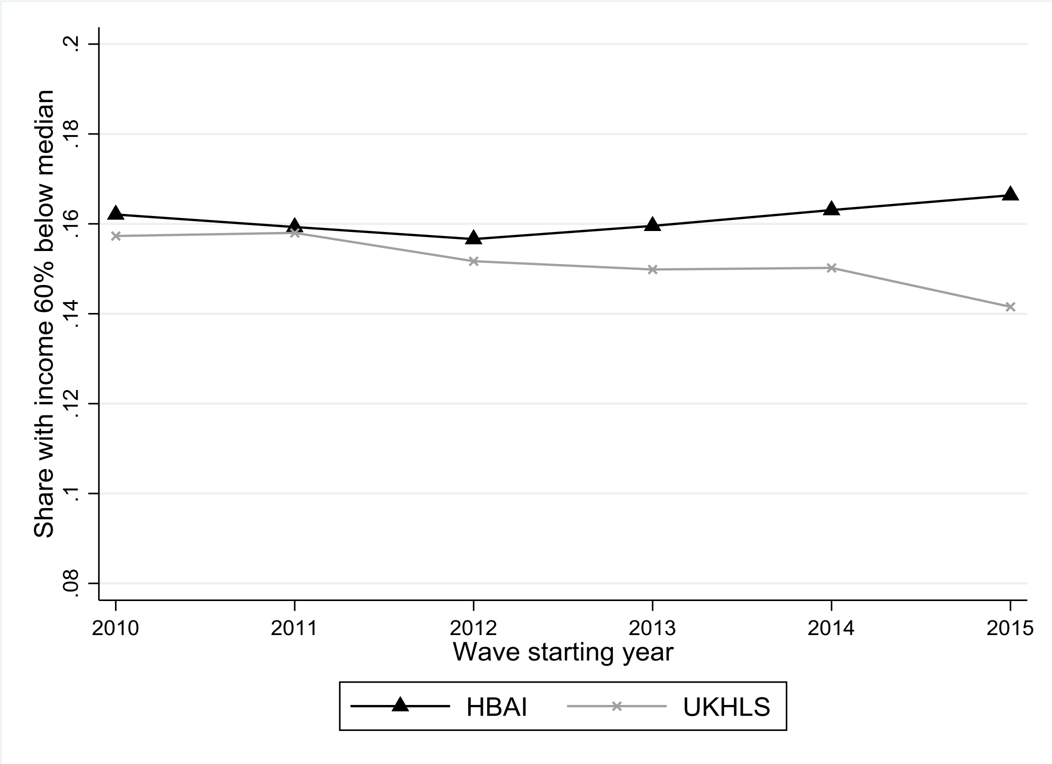


Figure 9. Gini

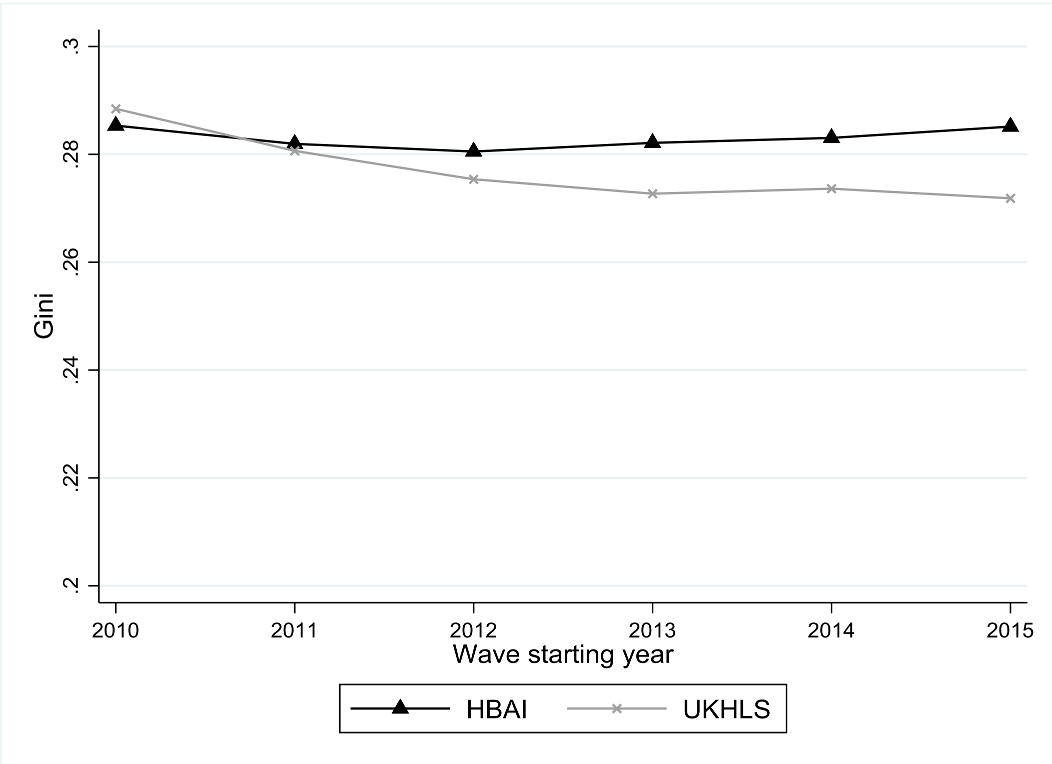


Figure 10. Theil

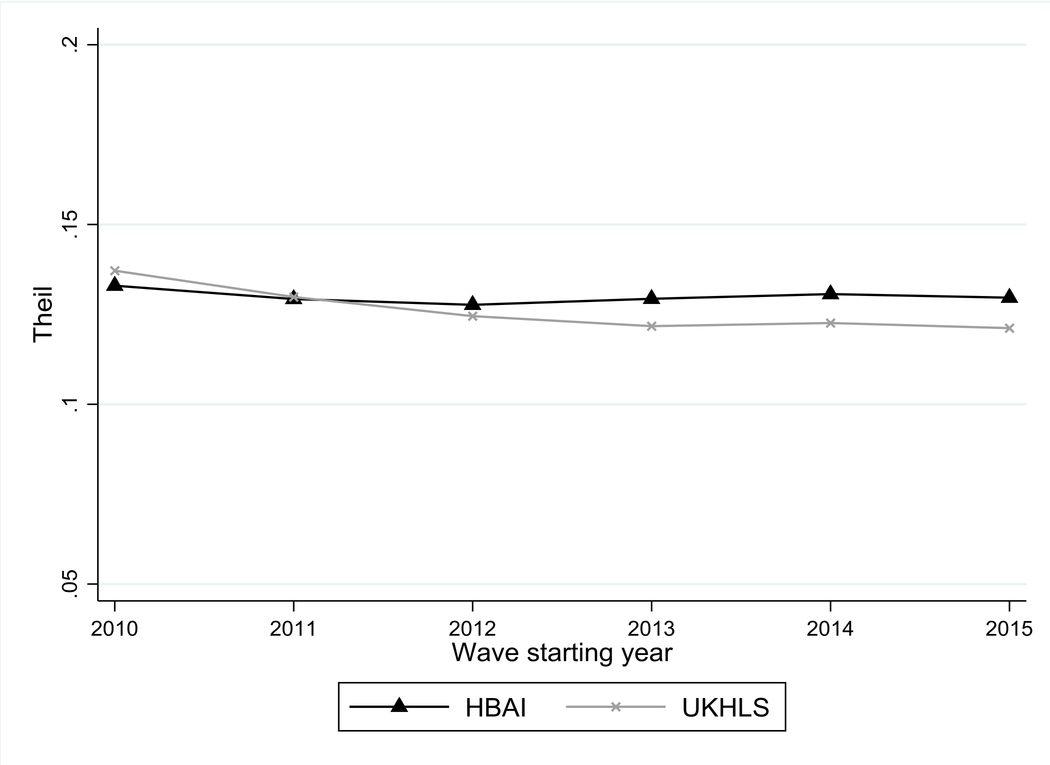


Figure 11. GE2

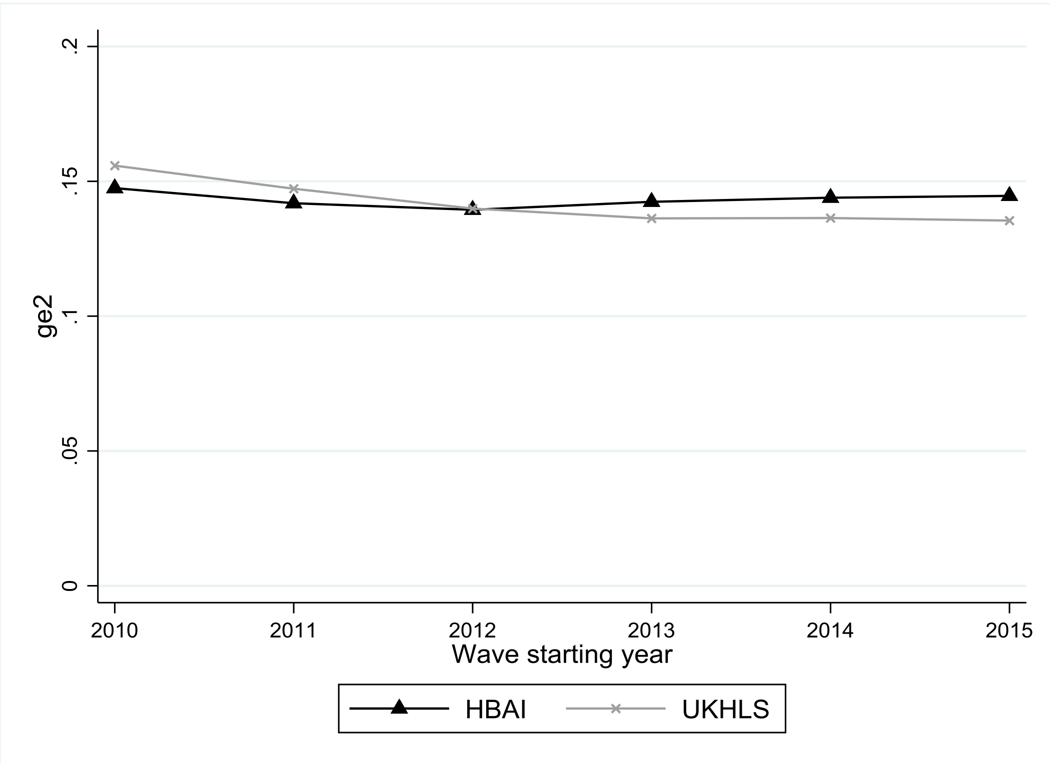


Figure 12. Atkinson

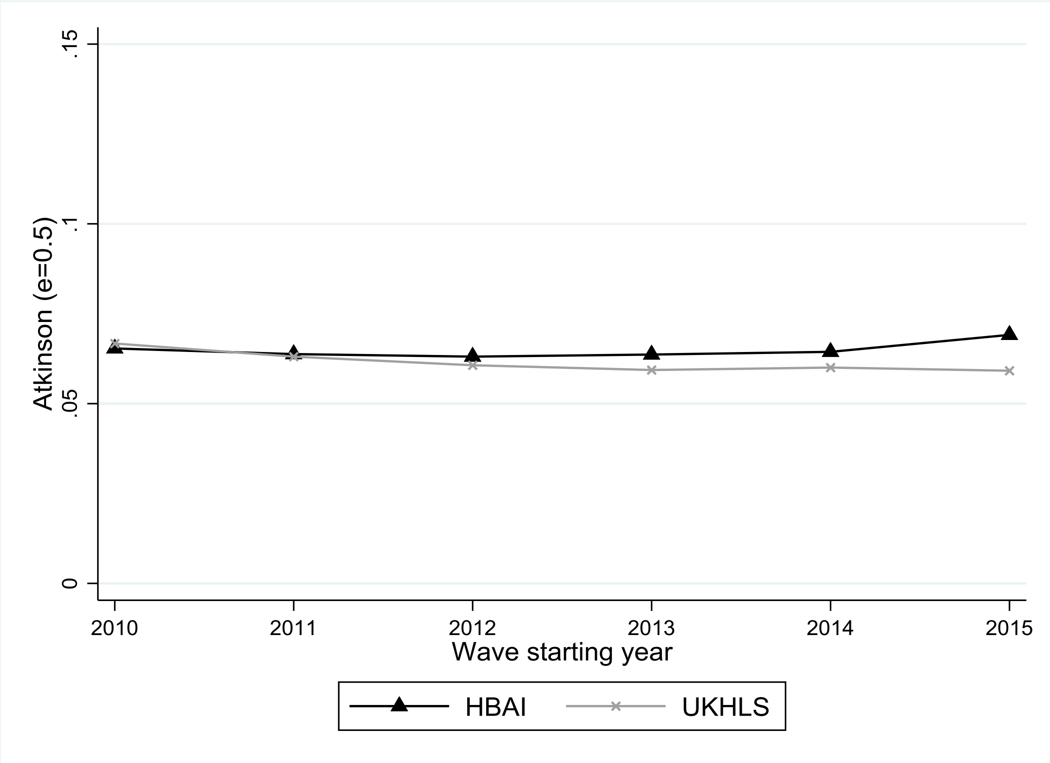


Figure 13. Benefit unit composition

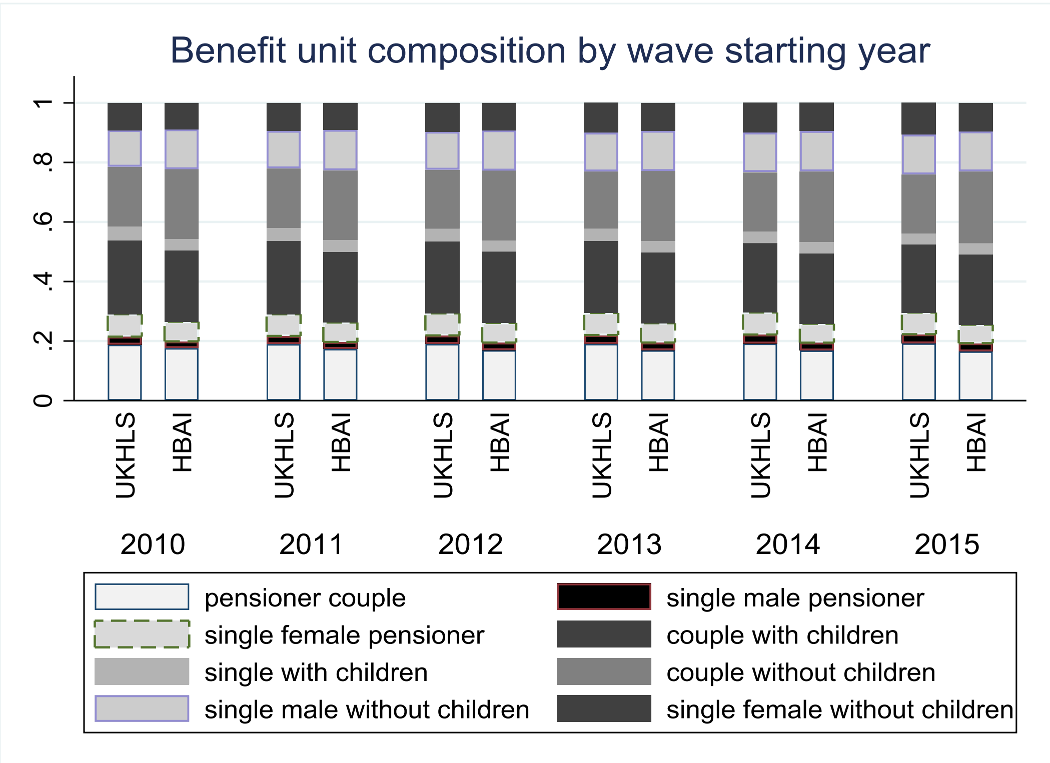


Figure 14. Benefit unit composition: lowest quintile

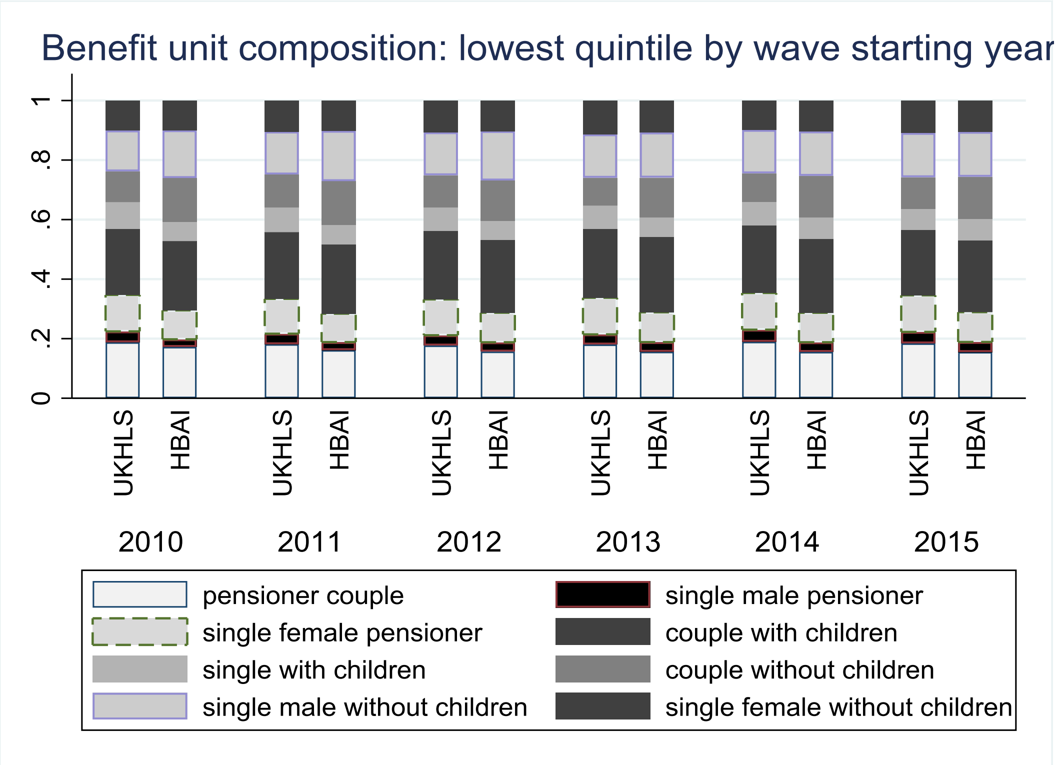


Figure 15. Benefit unit composition: second quintile

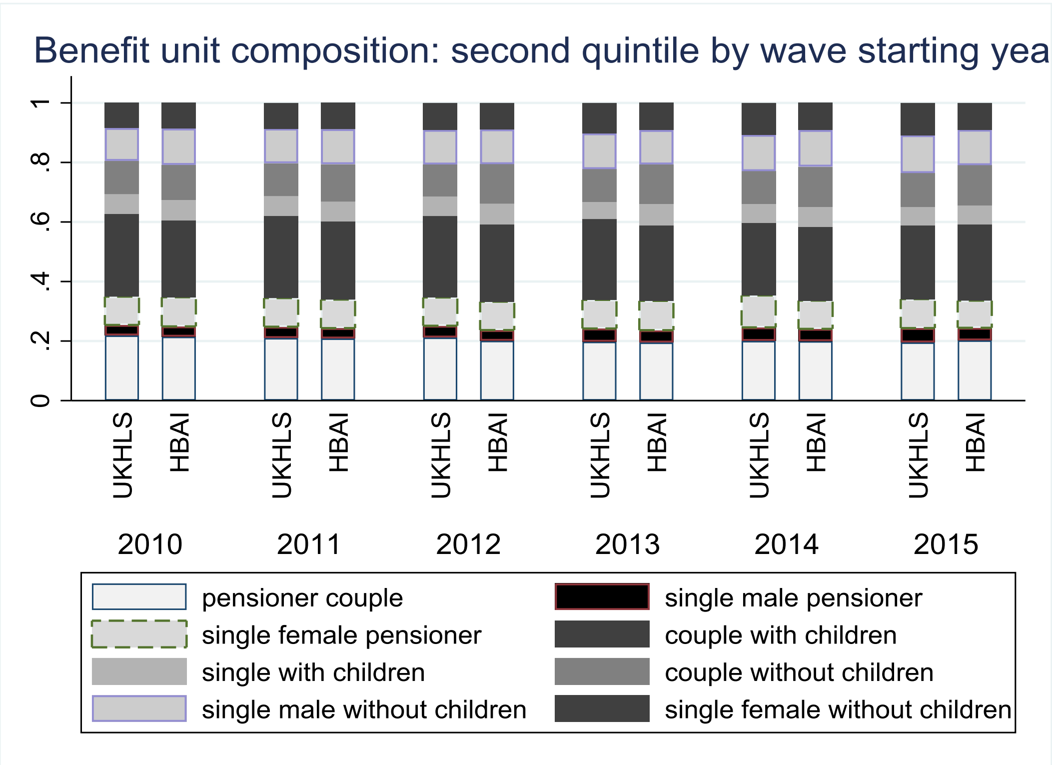


Figure 16. Benefit unit composition: third quintile

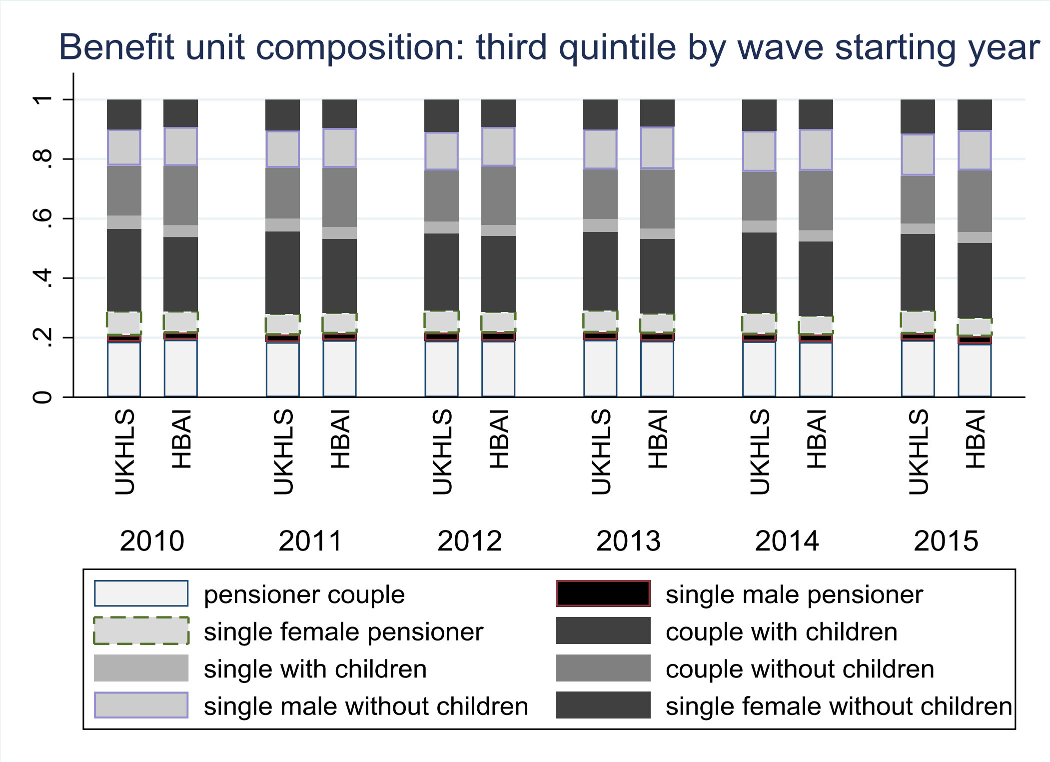


Figure 17. Benefit unit composition: fourth quintile

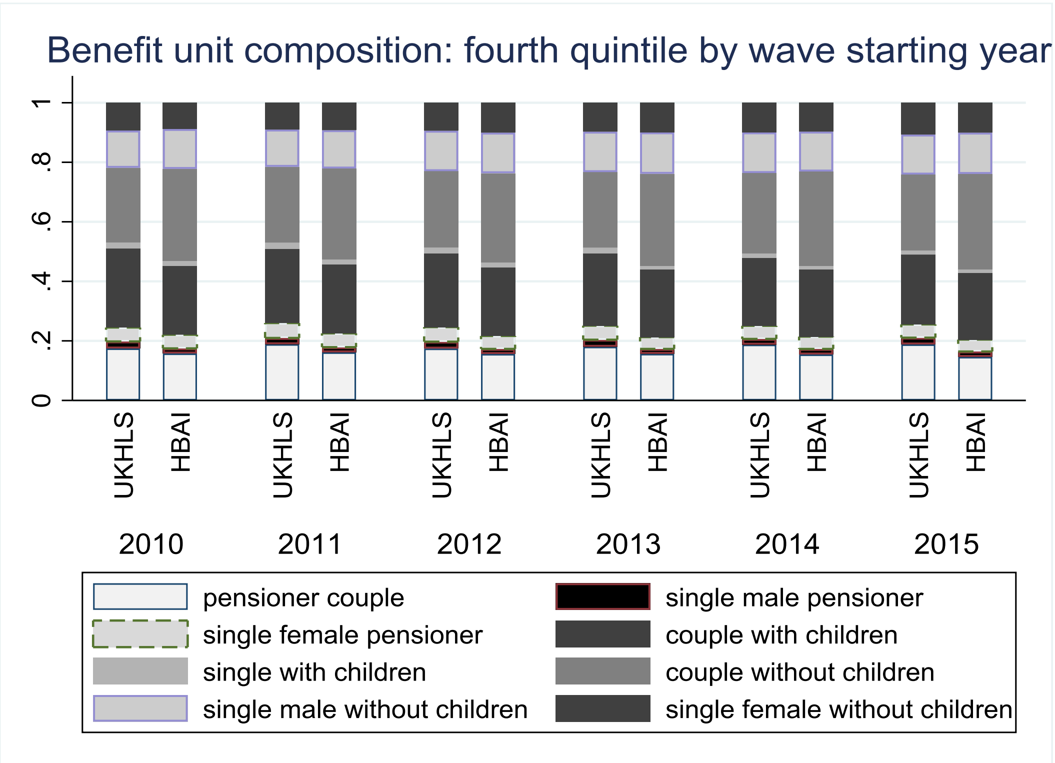


Figure 18. Benefit unit composition: top quintile

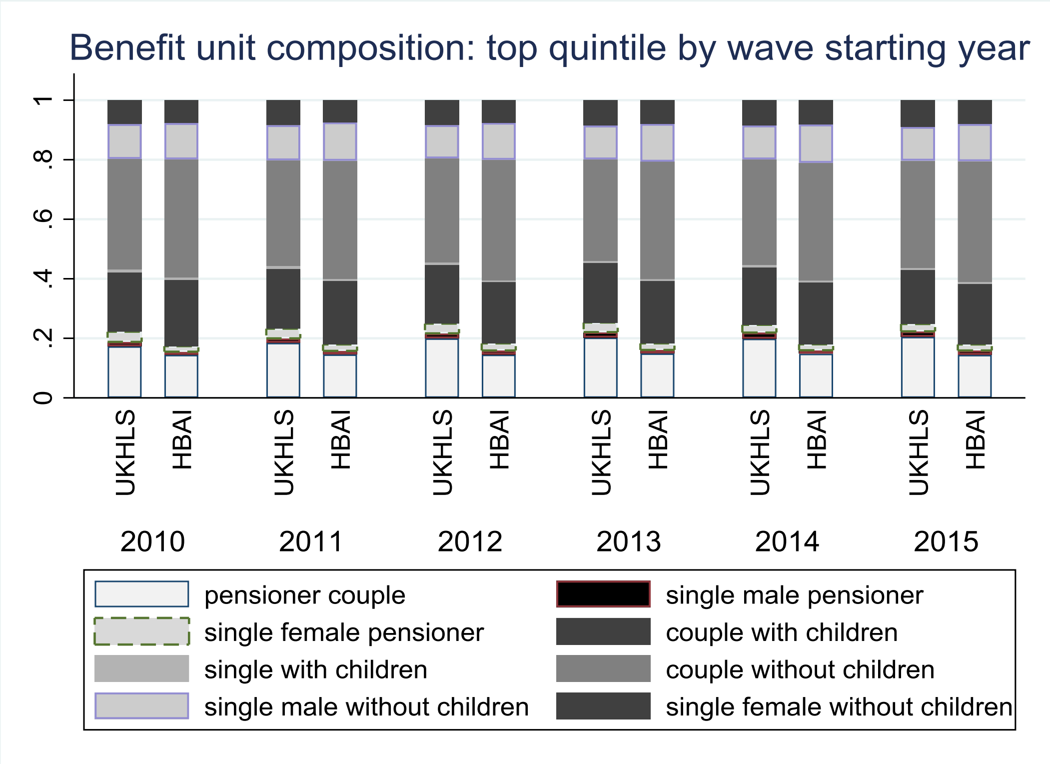


Figure 19. Mean gross household income

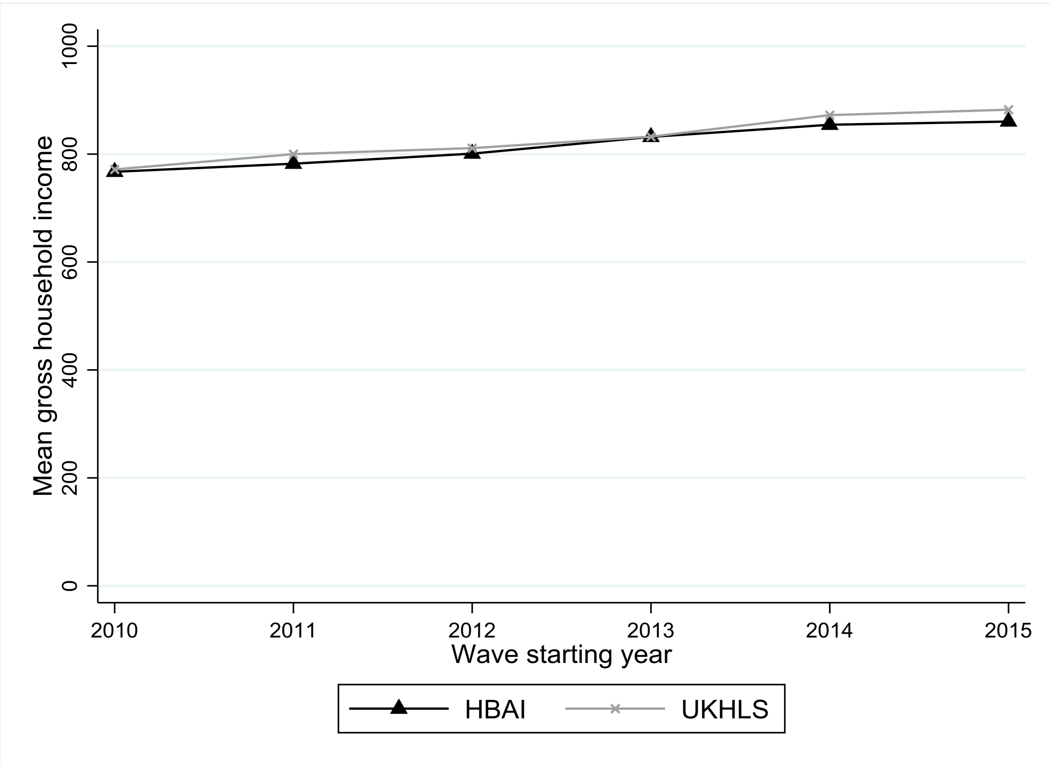


Figure 20. Mean gross household earnings

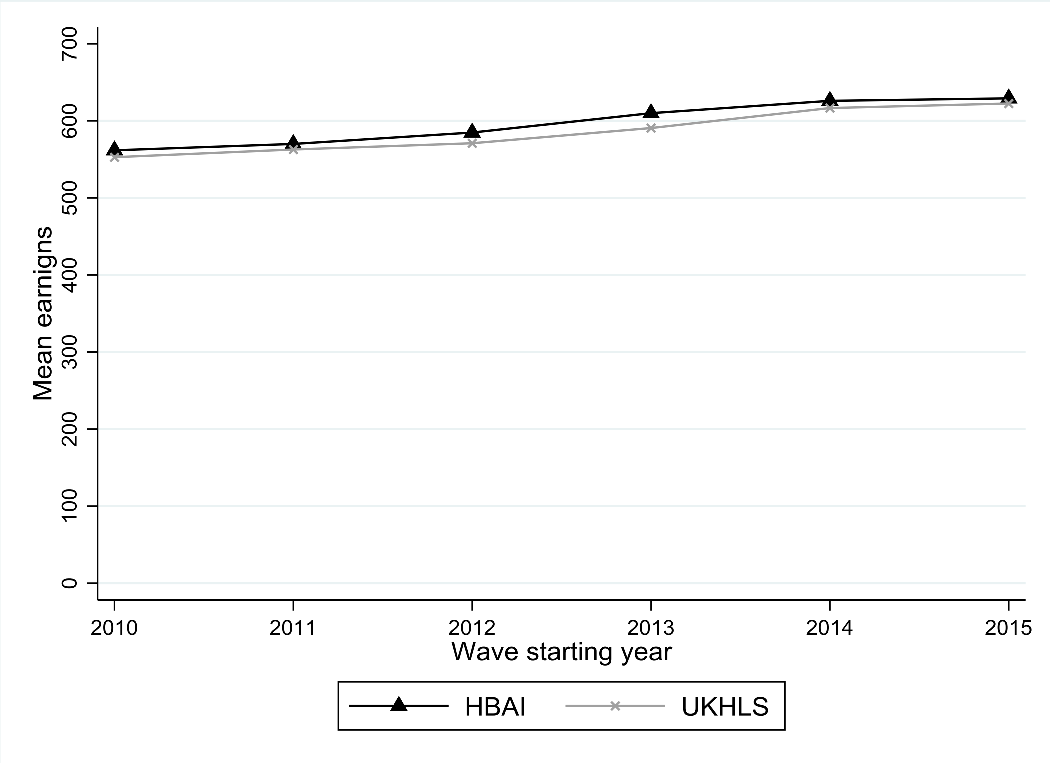


Figure 21. Mean state benefit income

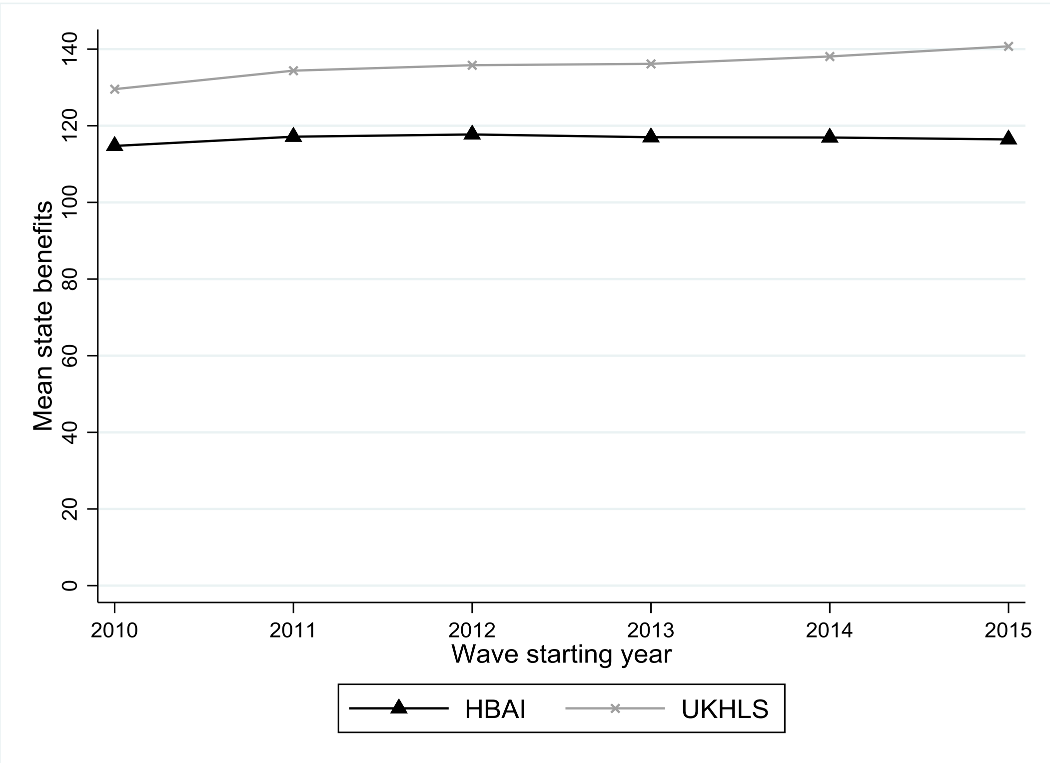


Figure 22. Mean occupation pension income

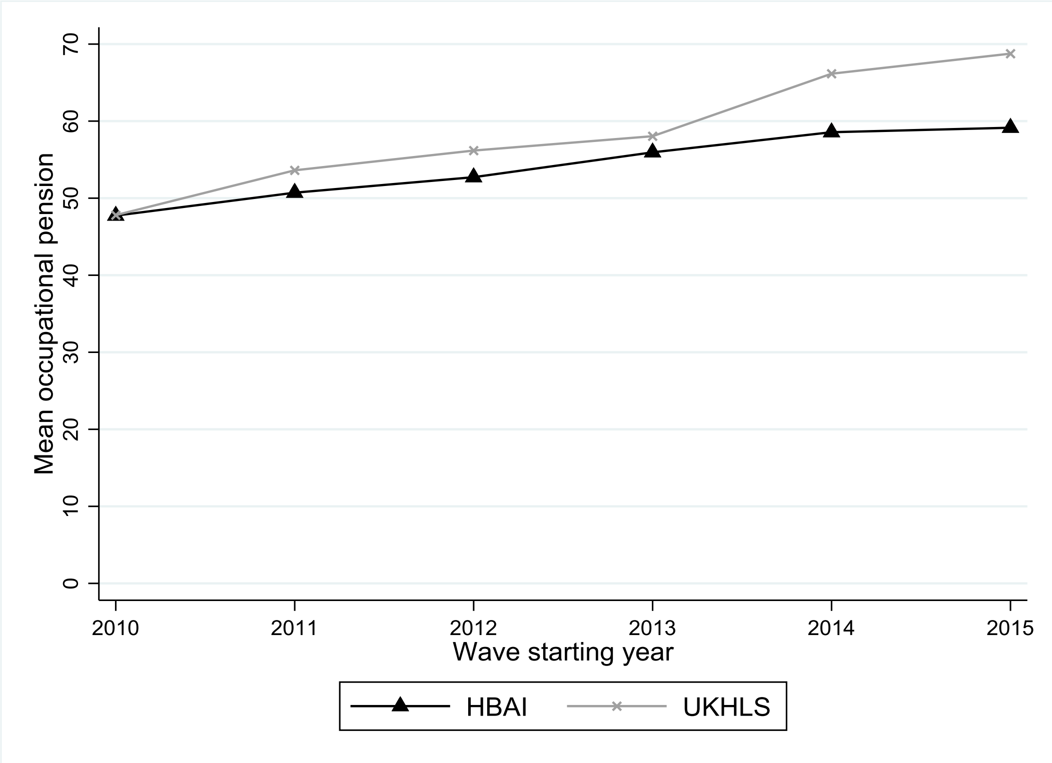


Figure 23. Mean of "other" income

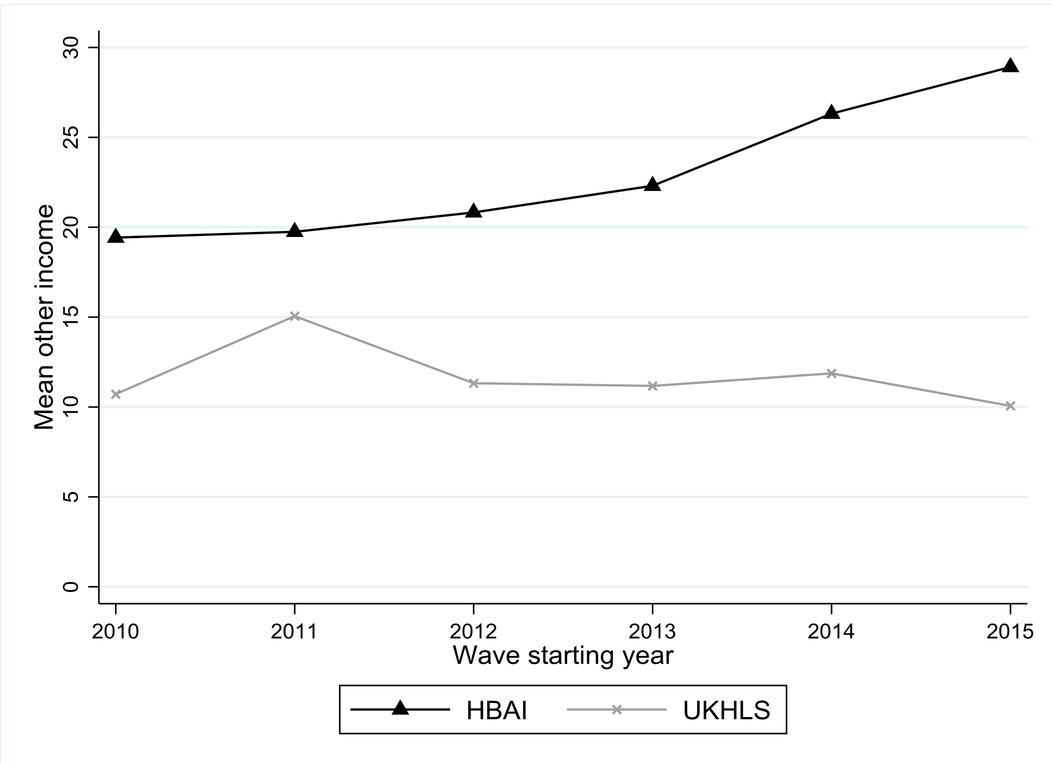


Figure 24. Mean investment income

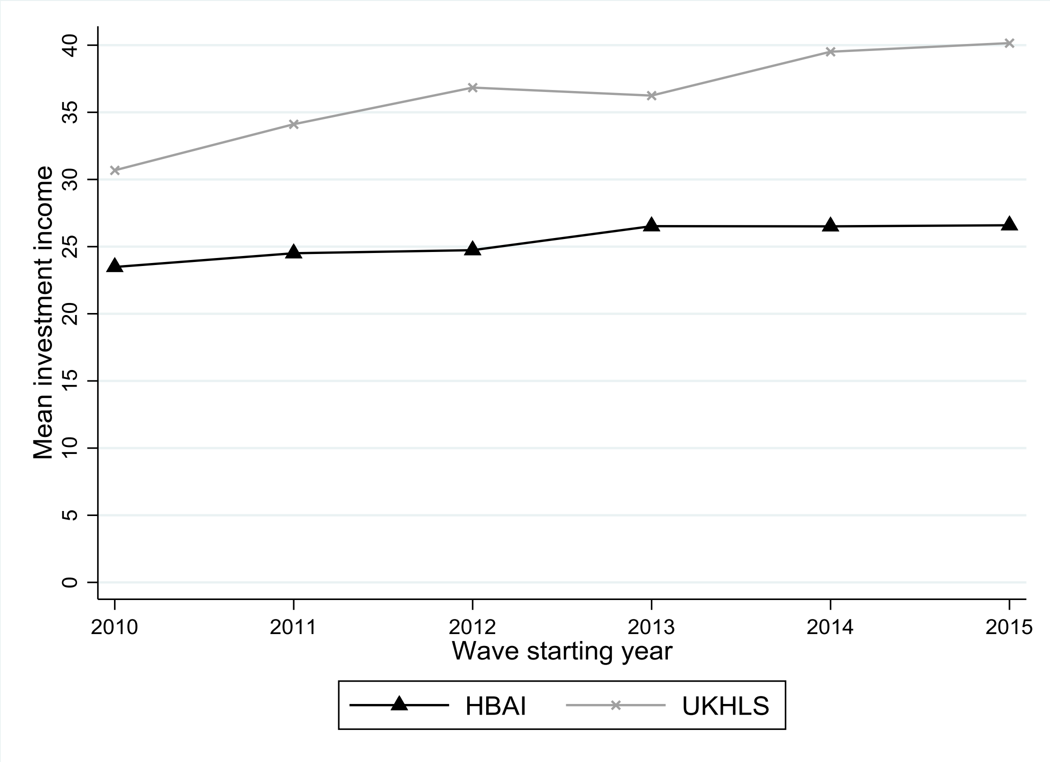


Figure 25. Decomposition of gross household income

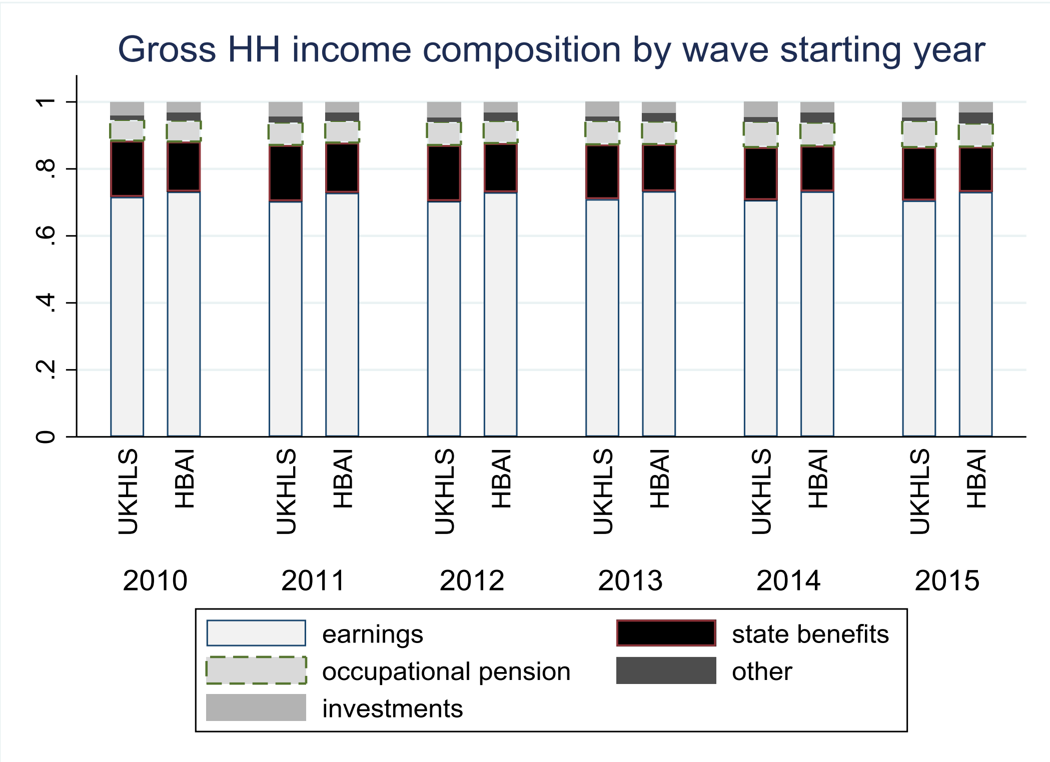


Figure 26. Decomposition of gross household income: lowest quintile

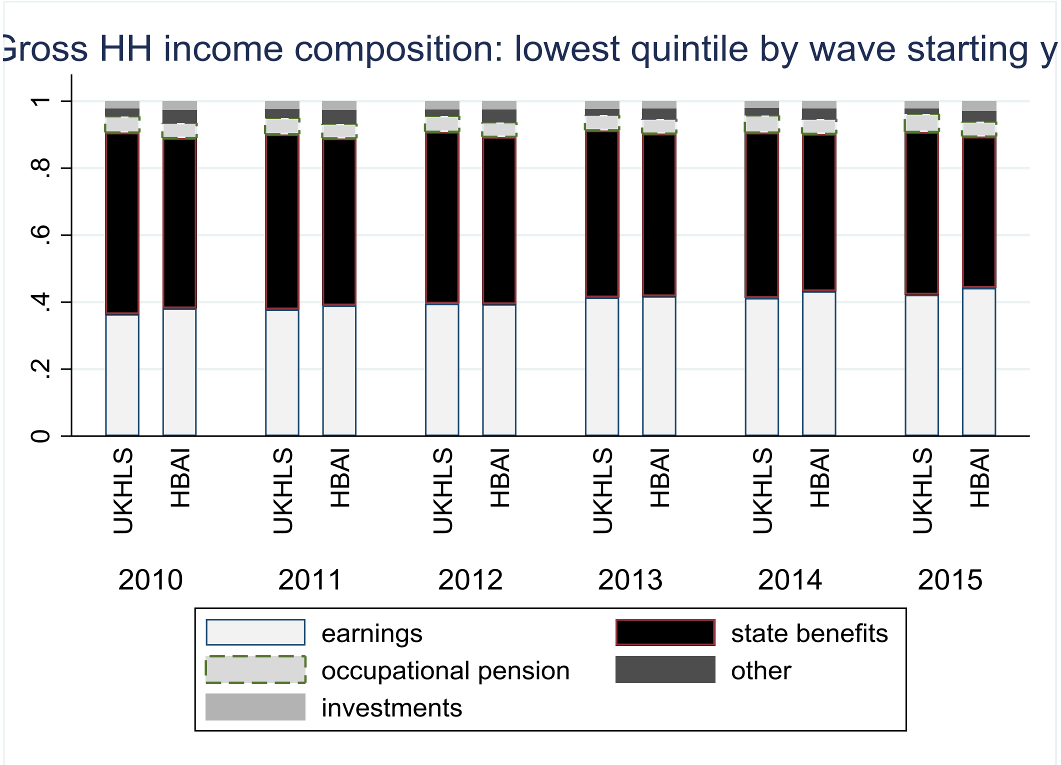


Figure 27. Decomposition of gross household income: second quintile

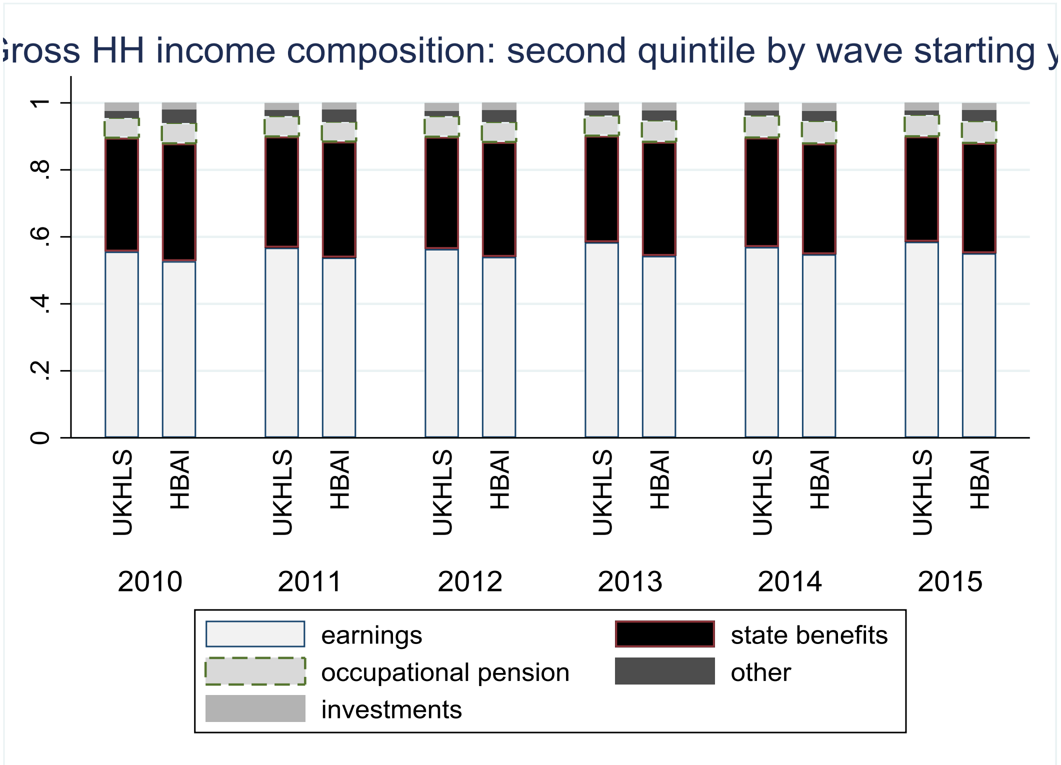


Figure 28. Decomposition of gross household income: third quintile

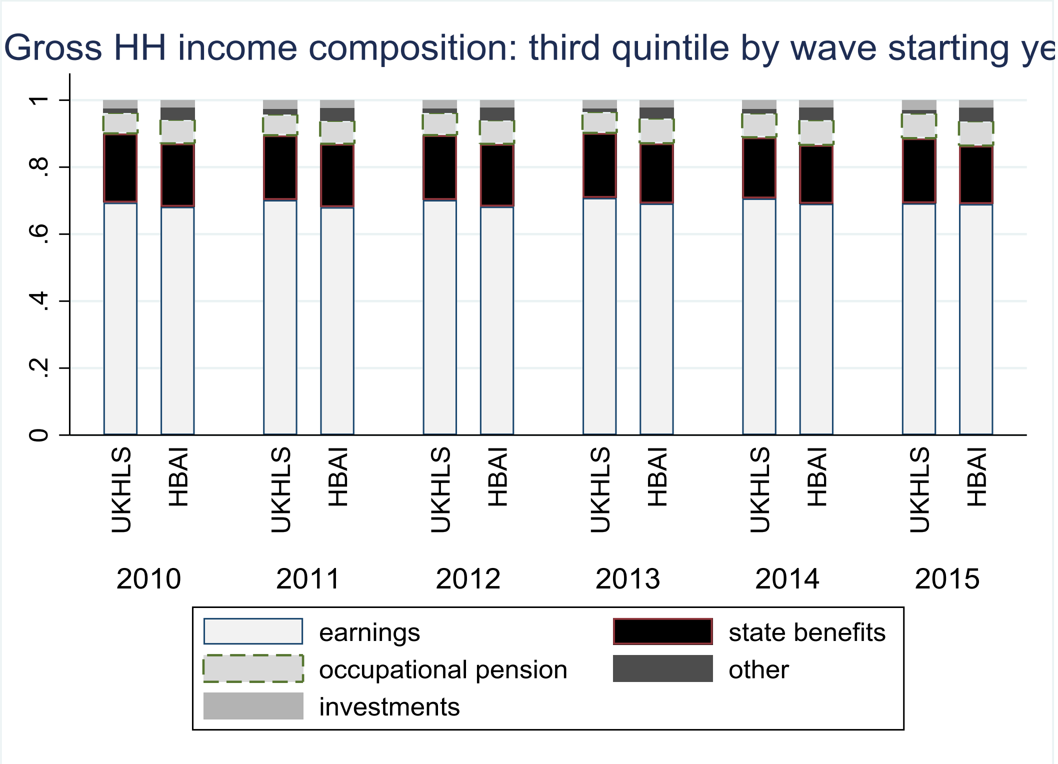


Figure 29. Decomposition of gross household income: fourth quintile

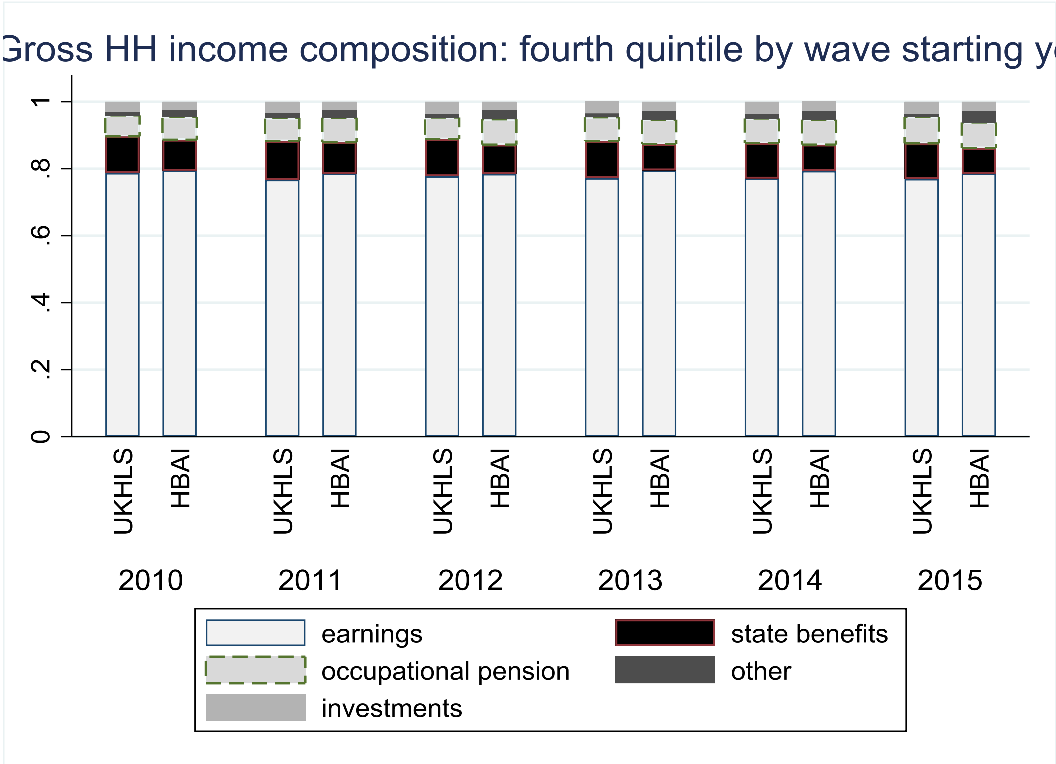
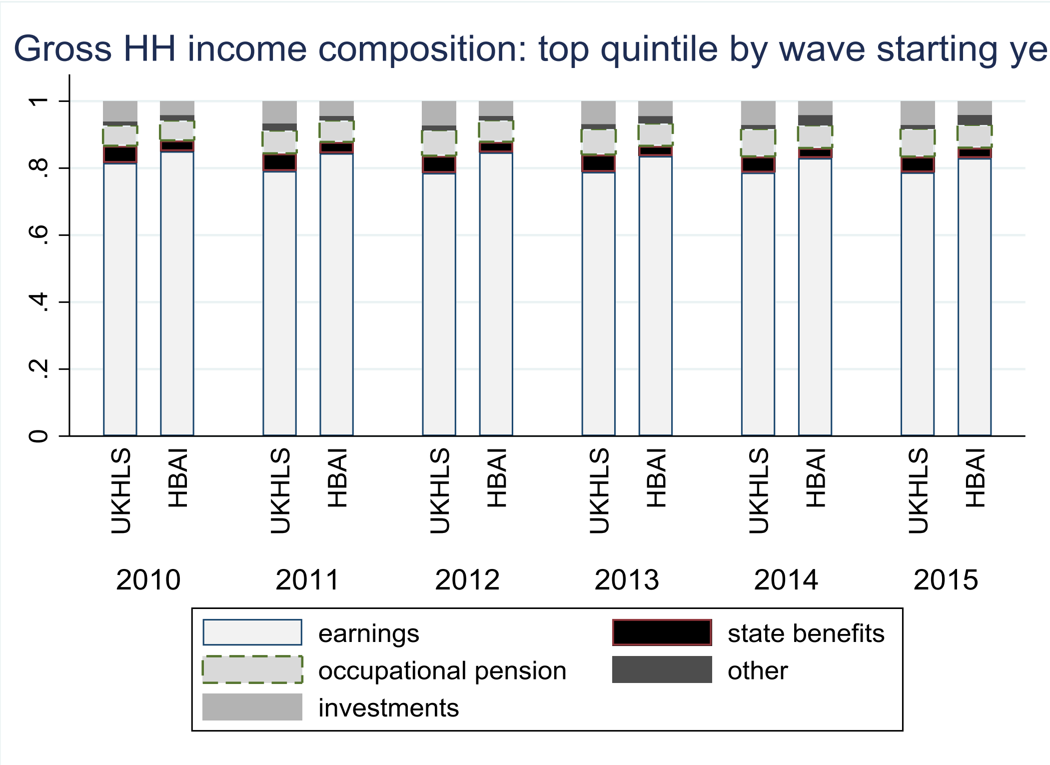


Figure 30. Decomposition of gross household income: top quintile



6. Annex References

Budowski, Monica, Robin Tillmann, Erwin Zimmermann, Boris Wernli, Annette Scherpenzeel & Alexis Gabadinho (2001), The Swiss Household Panel 1999-2003: Data for research on micro-social change, ZUMANachrichten 49, Jg. 25, November 2001.

Department for Work and Pensions (2018) *Income Dynamics: Income Movements and the Persistence of Low Incomes. https://www.gov.uk/government/statistics/income-dynamics-2015-to-2016*

Duncan, G., & Hill, M. (1985) Conceptions of Longitudinal Households: Fertile or Futile?, Journal of Economic and Social Measurement, 13, 361-375.

Fisher, Paul. (2016) Does repeated measurement improve income data quality?. No. 2016-11. ISER Working Paper Series, 2016

Frick, J.R. and Grabka M. M. (2004) "Missing Income Data in the German SOEP: Incidence, imputation and its Impact on the Income distribution". DIW discussion paper 376

Frick, J.R. and Grabka M. M. (2007) "Item Non-Response and Imputation of Annual Labor Income in Panel Surveys from a Cross-National Perspective". IZA DP No. 3043

Gaia, A. (ed). (2017). "Understanding Society Innovation Panel Wave 9: results from methodological experiments", Understanding Society Working Paper 2017-07, https://www.understandingsociety.ac.uk/research/publications/524448

Giles, Philip. (2001). An overview of the Survey of Labour and Income Dynamics (SLID). Canadian Studies in Population, 28, 363-375.

Hayes, C. and Watson, N. (2009) "HILDA Imputation Methods". HILDA project technical paper series No. 2/09, December 2009

Hill, M.S. (1992) The Panel Study of Income Dynamics: a User's Guide. Newbury Park CA: Sage Publications.

Jenkins, Stephen P. (2011) Changing fortunes: income mobility and poverty dynamics in Britain. OUP Oxford.

Kenward, M. and J. Carpenter (2007). "Multiple imputation: current perspectives." Statistical Methods in Medical Research 16 (3): 199-218

Knies, G. (2017) Understanding Society, The UK Household Longitudinal Study Waves 1-7 User Manual. Colchester, ISER University of Essex

Lipps, O. (2010) "Income imputation in the Swiss Household Panel 1999-2007". Fors Working Papers 2010-1

Little, R.J.A. (1988), ‘Missing Data Adjustments in Large Surveys’, Journal of Business and Economic Statistics, 6, 287-296.

Little and Su (1989) `Item non response in panel surveys' In D. Kasprzyk, G. Duncan and

M.P. Singh (eds) panel Surveys, New York, John Wiley.

Lynn, P. (2009). "Sample design for Understanding Society." Understanding Society Working Paper 2009-01.

Lynn, P. and G. Knies (2015). Understanding Society Quality Profile. Understanding Society Quality Profile.. Colchester, ISER University of Essex. 1.

Lynn, P., A. Nandi, et al. (2016). Design and implementation of a high quality probability sample of immigrants and ethnic minorities. Colchester, ISER University of Essex.

Lynn, P. and Borkowska, M. (2018) Some Indicators of Sample Representativeness and Attrition Bias for BHPS and Understanding Society. Understanding Society Working Paper 2018-01.

Mullan, K. Daraganova, G, Baker, K. (2015) “Imputing income in the Longitudinal Study of Australian Children”, LSAC Technical Paper No. 14

Ragunathan, E. T., J. M. Lepkowski, et al. (2001). "A Multivariate technique for multiply imputing missing values using a sequence of regression models." Survey Methodology 27(1): 85-95.

Rubin, D. B. (1987). Multiple imputation for nonresponse in surveys. New York, Wiley.

Schafer, J. (1997). Analysis of Incomplete Multivariate Data. London, Chapman & Hall.

Starick, R. and Watson, N. (2009) “Evaluation of Alternative Income Imputation Methods for the HILDA Survey”. HILDA project technical paper series No. 1/07, June 2007

van Buuren, S., H. C. Boshuizen, et al. (1999). "Multiple imputation of missing blood pressure covariates in survival analysis." Statistics in Medicine 18: 681-694

Wagner, G., Burkhauser, R., and Bheringer, F. (1993) 'The English language public use file of the German Socio-Economic Panel' Journal of Human Resources 28, 429-433.

Webber, M. (1994) 'Survey of labour and income dynamics: an overview' in Report1994, Cat. 75-201E. Ottawa: Statistics Canada.

Westermeier, C. and Grabka, M.M. (2016) “Longitudinal Wealth Data and Multiple Imputation. An Evaluation Study” Survey Research Methods. Vol. 10, No. 3, pp. 237-252

Wooden, M. ,Freidin, S., Watson, N. 2002. The Household, Income and Labour Dynamics in Australia (HILDA)Survey: Wave 1. Australian Economic Review, Volume 35, Issue 3, 339-348.