ISER Working Paper Series

A Comparison of Earnings Measures from Longitudinal and Cross-sectional Surveys: Evidence from the UK

Marco Francesconi
University of Essex
Institute for Fiscal Studies

Holly Sutherland
Institute for Social and Economic Research
University of Essex

Francesca Zantomio
Institute for Social and Economic Research
University of Essex

No. 2009-14
April 2009

INSTITUTE FOR SOCIAL
& ECONOMIC RESEARCH

www.iser.essex.ac.uk

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

A Comparison of Earnings Measures from Longitudinal and Cross-sectional Surveys: Evidence from the UK

Non-technical summary

Official UK income and poverty statistics come mainly from special income surveys, such as the Family Resources Survey (FRS). Information from alternative sources, such as the British Household Panel Survey (BHPS) is also used to complement the picture and describe how the income distribution and poverty status change over time. Such use of survey data however is challenged by the absence of systematic comparisons between BHPS data and other official income data, and therefore by our knowledge about differences or similarities across alternative data sources. This paper is a first step in that direction. It compares the BHPS earnings data with those collected in the FRS, using several earnings measures, which account for various key aspects of the two surveys, and contrasting two different points in time (1995/96 and 2003/04).

We find that the 1995/96 comparisons deliver results that are typically closer between the two surveys than the 2003/04 comparisons. The process that seems to drive most of the differences across the two surveys in our comparative work has to do with the reduced capability of BHPS data to capture characteristics of the current population as the panel becomes older.

A Comparison of Earnings Measures from Longitudinal and Cross-sectional Surveys: Evidence from the UK

Marco Francesconi
University of Essex and Institute for Fiscal Studies

Holly Sutherland
Institute for Social and Economic Research, University of Essex

Francesca Zantomio
Institute for Social and Economic Research, University of Essex

April 2009

Abstract

This paper compares earnings data from the BHPS with those collected in the FRS, contrasting two different points in time (1995/96 and 2003/04), allowing us to assess the possible extent of differential attrition in the BHPS data. We perform non-parametric tests of equality at the centre of the distributions and over the whole earnings distributions. We then apply multivariate regression methods to establish whether the earnings data yield different results in relation to three typical uses of earnings data. The two surveys have fairly similar earnings data in the first comparison year, while sizable differences emerge in the later comparison. This finding suggests the important role played by attrition and 'vintage' effects.

Keywords: Survey comparison; Current and usual earnings; Attrition and panel data; Item nonresponse.

Corresponding author: Marco Francesconi, Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK (mfranc@essex.ac.uk)

Acknowledgments

We thank Richard Berthoud, Mark Bryan, Nick Buck, Maria Iacovou, Stephen Jenkins, Peter Lynn, Heather Laurie, and seminar participants at the Institute for Social and Economic Research (Essex), for comments, discussions and suggestions on previous drafts of the paper. We also thank Judi Egerton for excellent technical support. Data from the Family Resources Survey (FRS) are made available by the UK Department of Work and Pensions (DWP) through the UK Data Archive. Material from the FRS is Crown Copyright and is used by permission. The BHPS data used were originally collected by the ESRC Research Centre on Micro-social Change at the University of Essex (now incorporated within the Institute for Social and Economic Research) made available through the ESRC UK Data Archive. Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here.

1. Introduction

Official UK income distribution estimates come mainly from special income surveys, such as the Family Resources Survey (FRS). Information from alternative sources, such as the British Household Panel Survey (BHPS), are also used to complement such estimates and draw a longitudinal picture of income mobility and poverty duration (e.g., the Households Below Average Income (HBAI) series published by the Department for Work Pensions). This use however is limited by the absence of systematic comparisons between BHPS data and other official income data, and thus by our inability to draw informed inferences from differences or similarities across alternative sources. This paper is a first step in that direction. It compares the BHPS earnings data with those collected in the FRS, using several *earnings* measures, which account for various key aspects of the two surveys, and contrasting two different points in time (1995/96 and 2003/04), allowing us to assess how the measures have changed over time.

The reason we analyse labour income is because it constitutes a dominant fraction of total income, especially among individuals below retirement age. Furthermore, of the almost 2000 studies that have been produced using BHPS data up to February 2009, about 30% have used either “labour income” or “earnings” or both in their title or in their keyword list (see the BHPS publications database at <http://www.iser.essex.ac.uk/iser/research-library/ bhps-publications-search>). Despite this impressive record of academic publications, the BHPS earnings information has never been formally evaluated against other data sources.

Our paper makes two contributions. First, we compare earnings collected from BHPS respondents to those obtained from respondents of the Family Resources Survey (FRS) at two separate points in time, 1995 (wave 5 of the BHPS, and fiscal year 1995/96 for the FRS) and 2003 (wave 13 and year 2003/04 for BHPS and FRS respectively). Second, we investigate whether the two data sources on earnings deliver different results in relation to a number of substantive issues which are relevant to policy makers and social analysts.

For the purposes of our exercise, the FRS seems to be an appropriate comparator. Like the BHPS, it is a household survey and not an administrative survey of employers (such as the Annual Survey of Hours and Earnings) or an individual-based survey (such as the Labour Force Survey). In addition, the earnings information in the FRS and BHPS questionnaires is collected in similar ways (e.g., question wording and routing; see Sections 2 and 3, and Appendix Table 1). Comparing earnings data from FRS and from BHPS, therefore, is likely to be meaningful. The FRS also offers a relatively large sample coverage (as opposed to other household surveys, such as the General Household Survey) and is routinely used to derive official income and poverty statistics.

Using two separate time periods will allow us to see whether the earnings data in the two surveys have become more or less similar over time. During this same time, the BHPS has presumably suffered from attrition bias and benefited from major sample replenishments (for example, through the Scottish and Wales booster samples in 1999 and the Northern Ireland sample in 2001). The choice of 1995/96 as our first comparison year is motivated by the fact that this was the second year in which FRS data were collected, thus allowing for any initial problems in that survey to be corrected. The second comparison year (2003/04) is relatively recent and allows for complex cross-survey comparisons over time.

Our comparative analysis starts with an analysis of means and moves on to consider the whole earnings distributions. For robustness, we look at different earnings measures (e.g., current and usual pay) and make comparisons both using and not using sample weights, imputed cases and other survey-specific features. For each period separately, we perform this analysis both for the entire population and for subsamples stratified along key socio-economic characteristics (e.g., sex, age, education and marital status).

In relation to the second of our contributions, we estimate the probability that a worker has earnings below a specific earnings cut-off (set at 60% of median earnings) or above another cut-off (top earnings decile). Another exercise in which earnings are the dependent variable (although continuously rather than in a discrete fashion as before) involves the estimation of wage equations where earnings are regressed on a set of individual attributes that are expected to affect wage determination. For this comparison, we use least squares and quantile regressions. Finally, we estimate models that explain the likelihood that a worker contributes to an occupational pension scheme, in which labour income is one of the key determinants, that is, an independent variable. The idea is to check whether the wage effects in such models differ by survey and over time, holding a set of standard covariates constant.

Interpreting our results (regardless of whether we detect similarities or differences between the two surveys) is not trivial and yet very important. The key criterion underlying our interpretation is related to the inherent trade-off between *measure homogeneity* (that is, the maximal similarity of the definition of earnings measures from the two surveys) and *sample size*. This trade-off affects the statistical power of the tests in all comparisons and, hence, inference.

This criterion is also related to other survey-specific characteristics, which may render our comparison more difficult. For example, the FRS nonresponse rate has increased from 30% to 36% between the 1995/96 and 2003/04 surveys (Kirri et al., 2005). Despite lower nonresponse rates, the BHPS is affected by attrition problems. For instance, only 77% of the respondents in the first wave (1991) were successfully re-interviewed in the fifth wave (1995) and this proportion further declined to 65% in the 2003 survey. In addition, given its longitudinal nature and in absence of ad-hoc booster samples, the BHPS has a lower chance than the FRS to keep up with the changing composition of the UK population (e.g., foreign migrants and specific household types). As the panel grows older, this 'vintage effect' is likely to become more prominent.

This study extends recent research that investigates issues of quality concerning income variables in a cross-survey comparative perspective to earnings. For example, Micklewright and Schnepf (2007) compare income distributions from single-question surveys of income, such as the ONS Omnibus survey and the British Social Attitudes surveys, with those from the Family Resources Survey and the Family Expenditure Survey, which ask income information in much greater detail. Another recent example of comparative data analyses in an international context is the work by Brown et al. (2007), which builds on earlier UK studies on income data reliability (e.g., Atkinson and Micklewright, 1983; Böheim and Jenkins, 2006). This in turn is related to other data reliability exercises involving large-scale surveys as well as retrospective information and panel data (e.g., Morgenstern and Barratt, 1974; Mathiowetz and Duncan, 1988; Biemer et al., 1991; Elias, 1996; Dex and McCulloch, 1998; Dayal et al., 2000; Francesconi, 2005).

The next section describes the BHPS and FRS data sources, paying attention to differences in sample design, data collection and statistical adjustment procedures. These are important statistical features which could help us explain possible differences arising from our comparative exercises. Section 3 discusses our sample selection procedures and presents the measures of earnings used in the analysis. Section 4 formulates the key statistical hypotheses underlying our comparative analysis, and illustrates the empirical strategy. Section 5 reports the main results from our non-parametric comparisons, while Section 6 uses the earnings measures in three different parametric exercises. In the first two, earnings are on the left-hand side of our regressions (in some cases as a dichotomous variable, in other cases as a continuous variable), whereas in the third evaluation, they are on the right-hand side as an explanatory variable. Section 7 concludes.

2. Data

2.1 Sample design

The BHPS is a longitudinal survey designed to be representative of all individuals resident in Britain at multiple time points corresponding to the waves of yearly data collection, starting from 1991. All individuals in wave-1 respondent households become part of the longitudinal sample as Original Sample Members (OSMs) and remain sample members at all subsequent waves until they die. Two categories of new permanent sample members could join at all waves subsequent to wave 1. These are babies born to (or adopted by) an OSM (i.e., OSM by virtue of descent), and parents of OSMs, who have joined the OSM household. Other non-OSM individuals are also eligible to be interviewed as long as they live in the same household as an OSM. The initial BHPS coverage included only Great Britain south of the Caledonian Canal, but later booster samples over-represented Scotland and Wales from the 1999 (wave 9) survey, and covered Northern Ireland from 2001 onwards.

The FRS is instead a cross-sectional annual survey that started in 1992. Originally meant to be representative of all private households in Great Britain south of the Caledonian Canal, from 2001/02 it also included the Scottish Islands and the area north of the Caledonian Canal, and from 2002/03 was extended to include Northern Ireland as well.

2.2 Data collection

Both surveys use face-to-face in-home interviewing as the main mode of data collection and both questionnaires involve household and individual blocks. The BHPS fieldwork is conducted mostly during the Autumn, whereas the FRS interviews are spread throughout the fiscal year, from April to March of the following year. All FRS interviews are carried out on a Computer Assisted Personal Interviewing (CAPI) mode, while BHPS interviews were conducted by Paper and Pen Interviewing (PAPI) up to wave 8 (1998), and by CAPI afterwards. In the BHPS case, because of the longitudinal nature of the data, special attempts have been made to match respondents to the same interviewers over time and a number of strategies were implemented to maintain high panel and unit response rates.

2.3 Adjustment procedures

Missing data on a range of income variables have been imputed in all waves of the BHPS using both hot-deck imputation routines and, for monetary amount variables, regression-based imputation techniques. Since all imputed values are flagged, users are in a position to decide whether to use them or not. The FRS uses mostly hot-deck imputations, but also algorithms and case-by-case “mop-up” imputation methods. Imputations are recorded in "transact" databases, in such a way that it is always possible to reproduce the original unimputed data.

Besides longitudinal weights, the BHPS data provide cross-sectional weights, for use with single wave analyses such as ours. In the first wave (1991), weights adjust for unequal selection probabilities of addresses, non-response at the household level, nonresponse of individuals within responding households, and rescaling so that the weighted sample size equals the unweighted (interviewed) sample size. For all the subsequent waves, cross-sectional weights account for new entrants and adjust for within-household nonresponse.

The FRS weights attempt to correct for differential nonresponse while scaling up sample numbers to the overall population, using the ratio of population to sample counts for subgroups defined on variables reflecting differential response rates. We ought to emphasize that the weighting procedures in the FRS are not entirely identical to those used for the cross-sectional weights in the BHPS. In particular, the FRS weights only aim to calibrate the sample to the current population characteristics, while the BHPS cross-sectional weights try to account for the inclusion of new entrants who, by definition, do not have a wave-1 weight.

Appendix Table 1 summarises these salient features for both surveys, and provides elements of comparison on important aspects other than those just described.

3. Samples, Measures of Earnings, and Other Variables

3.1 Sample Selection

Our empirical study is based on samples of employees currently in employment. These samples also include employees who report having worked the week prior to interview or having a job they were away from in the week prior to interview. Question wording and routing on such variables is very similar between the two surveys. A difference, however, is that while the FRS collects information on up to three jobs (with the first job being either the 'most remunerative' job or the job in which respondents spend most of their time), the BHPS covers only the main job. We, therefore, shall restrict our FRS analysis to first jobs only.

Employees aged 18 or less are excluded, because the FRS collects employment information only from individuals aged at least 19, or aged 16 to 18 if not in full time education or married. Such restrictions lead us to a sample of 21,638 employees in the 1995/96 FRS and to a sample of 4,635 employees in the corresponding fifth wave of the BHPS. The 2003/04 FRS sample, instead, is made up of 24,885 individuals, and the 2003 BHPS sample from its thirteenth wave contains 8,042 employees.

3.2 Earnings Measures

We primarily analyse two earnings measures, although for robustness purposes we also consider several variants of such two measures. Our baseline measures aim at achieving two opposite goals. At one end, we identify a measure that should be collected as accurately as possible in both surveys and thus, ideally, reflects maximal *homogeneity* between our two data sources. At the other end, we have a measure that trades off homogeneity for maximal *sample size*, a trade-off that analysts often have to face. We expect to obtain more similar results with the first measure, but at the potential cost of lower statistical power, as a consequence of smaller sample sizes and greater variance. With the second measure, instead, we expect to observe larger differentials by source but to gain greater power and achieve more reliable inferences (see also the discussion in Section 4).

The first measure is given by 'current gross earnings' based on the last pay received. To enhance homogeneity, we restrict our samples to individuals who consulted their payslip, exclude workers with imputed amounts, exclude workers from the 2003 BHPS booster samples (which oversample the Welsh and Scottish populations and include Northern Ireland) and, consequently, exclude individuals living in Northern Ireland from the 2003/04 FRS samples.

For the FRS, we use the variable GRWAGE. In the 1995/96 survey, this comes from the question: "*What was the gross wage/salary as shown on your payslip*?", whereas the 2003/04 question is: "*What was the gross wage/salary - i.e. the total, before any deductions but excluding any tax credit payments?*". The equivalent information for the BHPS is given by the variable PAYGL, which refers to the question: "*The last time you were paid, what was your gross pay – that is including any overtime, bonuses, commission, tips or tax refund, but before any deductions for tax, national insurance or pension contributions, union dues and so on*?". Payslip consultation is embedded in the 1995/96 FRS question, while for the 2003/04 FRS and both BHPS years this information is available from a separate question. Our baseline current earnings measures are constructed only for workers whose payslip was consulted.

As mentioned earlier, we explore several departures from this baseline definition. Besides its relevance as a sensitivity check, analysing each variant separately allows us to assess how critical each restriction is for the homogeneity/sample-size trade-off. First, we include imputed cases. This only affects the FRS sample (0.22% and 4.9% of GRWAGE observations in 1995/96 and 2003/04 respectively), since the PAYGL variable in the BHPS is not imputed. Second, respondents whose payslip was not consulted are included. Third, we add workers from Northern Ireland to the 2003/04 FRS sample as well as employees from the booster samples in the 2003 BHPS sample. Fourth, we perform our comparisons on weighted data, using cross-sectional weights. Fifth, we consider current earnings measures that are computed without the AEI indexing.

Our second measure, which is intended to deliver larger sample sizes, is given by 'usual gross earnings'. This is a measure favoured by many analysts of BHPS and FRS earnings data (e.g., Davies and Joshi, 1998; Ermisch and Francesconi, 2000; Sutherland and Piachaud, 2001; Stewart, 2004; Blundell *et al*., 2008). At the opposite extreme to the current earnings measure, the baseline version of this measure includes imputed cases, employees whose last payslip was not consulted and, for the 2003 surveys, workers either from the BHPS all booster samples or, in the FRS case, living in Northern Ireland.

Both FRS and BHPS questionnaires are similar in the way they elicit information on usual earnings. They first ask about the last pay, then ask whether such a pay corresponds to the usual pay, and, only for cases where this does not, they explicitly inquire about usual pay. Therefore, when last pay is the same as usual pay, information on the former is used to construct our usual pay measure. In both the 1995/96 and 2003/04 FRS surveys, if the last pay is reported not to be the usual pay, then we use the variable UGROSS, which records: "*What do/did you usually receive before all deductions*?". When, instead, the last pay coincides with usual pay, then GRWAGE is used. In the 1995/96 FRS the last pay information is collected only if the last payslip is consulted. For employees whose payslip is not consulted, we instead derive their gross pay from the last net pay and related deductions which, among others, include national insurance, pensions and superannuations, and medical insurance. The BHPS usual pay information is obtained from the derived variable PAYGU. This includes imputed data and is derived from the PAYGL variable described above if this corresponds to usual pay, and otherwise from the PAYU variable, which refers to the question: "*How much are you usually paid*?". If any of such pieces of information is reported after deductions, PAYGU is constructed for the BHPS official data release as a gross amount using information about marital status, partner's activity, and pension scheme membership.

Again, to check for robustness and assess differential results across surveys, we consider five alternative variants of this baseline measure by separately dropping imputed observations, dropping BHPS non-OSMs, dropping FRS workers living in Northern Ireland and BHPS workers from booster samples, using weighted data, and removing the AEI indexing.

Notice that, because of the fieldwork timing differences between the two surveys, for both types of earnings measure we use the seasonally unadjusted Average Earnings Index (AEI) to deflate all monetary amounts to the month of September of the relevant year (either 1995 or 2003). Moreover, to allow for cross-time comparisons, all earnings figures are expressed in constant (1995) prices. Finally, we present the BHPS figures in weekly amounts using information from a period code question (the FRS data being already provided in weekly terms).

3.3 Other variables

Part of our cross-survey comparisons are based on multivariate analysis in which we use individual demographic and socio-economic characteristics as standard control variables. These include age (grouped in three classes: <25, 25-50, and >50), gender, marital status (four categories: married or cohabiting, never married, and single/separated/divorced), country of residence (England, Wales, Scotland and Northern Ireland), whether the individual is working full time (i.e., 30 or more hours per week) or part time (less than 30 hours per week), whether he/she has dependent coresident children, and whether he/she left full time education before or after age 18. The definitions of such variables are broadly similar in both surveys and are thus unlikely to induce additional differential biases in our comparative exercises. For the sake of brevity, summary statistics on these variables are not reported.

4. Statistical Hypotheses and Empirical Strategy

As described in Section 3, the way in which our earnings measures are defined adheres to two (possibly conflicting) statistical objectives that are salient in our comparative analysis (Biemer et al. 1991; Rudas, 2008)**.** At one extreme, we have *measure homogeneity* while, at the other extreme, *sample size*. Our earnings variables have been constructed so that cross-survey homogeneity is arguably highest in the case of the baseline current earnings definition. Sample size, instead, is likely to be maximal in both surveys when we use baseline usual earnings measures.

Most of the statistical analyses performed in this study will be based on hypothesis testing of the type:

$$H_0: \quad s_{BHPS} = s_{FRS}, \tag{H0}$$

where s denotes any statistic of interest, such as unconditional and conditional means, the whole distribution of earnings, and relevant estimates in multivariate analyses.

In all our cross-survey comparisons, we expect similarity in results from (H0) to be proportional to measure homogeneity. That is, we expect to be less likely to reject the null (H0) when current earnings definitions are used than when usual earnings definitions are used. However, because usual earnings can be defined on larger samples, it is likely that the statistical power of any test (H0) performed on such measures is greater, and thus the probability of a Type II error smaller (Mood et al. 1974; Billingsley, 1995). Therefore, a greater statistical power associated with usual earnings measures may offset our earlier expectation, making us less likely to reject (H0) with usual earnings definitions. Which of such two opposite tendencies actually dominates is an empirical issue which will be assessed in the next two sections.

The variants of each baseline definition (also described in the previous section) relax each definition's specific criterion (that is, homogeneity on the one hand, and sample size on the other) at the cost of either employing less homogenous measures or constructing measures on smaller sample sizes. Analysing these measures separately therefore can give us some evidence of the sensitivity of the results found with our two baseline definitions and may also reveal whether measure homogeneity or sample size matters most to gain statistical power in this application.

As anticipated in the Introduction, the FRS suffers from sizable nonresponse rates, which have increased between the two survey points considered in this study. On the other hand, the BHPS (but not the FRS) data are likely to suffer from attrition, whereby employees with certain (observed and unobserved) characteristics or in specific occupations and industries are less likely to be followed over time (Madow *et al*., 1983; Kalton and Kasprzyk, 1986; Little and Rubin, 1987; Rubin, 1996). This issue may add one layer of complications to our cross-survey comparisons. Another layer might come from the BHPS 'vintage effect', that is, the effect driven by undetected trends in the composition of the UK population. If — as a result of panel attrition or vintage effect — the BHPS becomes an increasingly more select sample of the British population, and thus of British workers, we might reject the null more often than we would expect in the 2003/04 comparisons as opposed to the 1995/96 comparisons. Rejecting the null on these grounds, however, does not necessarily invalidate the inference relevant for our analysis (Robins et al., 1995; Chen and Little, 1999); nonetheless, it emphasizes the importance of the non-random component of the missing data process in the BHPS, which data collectors and analysts must face.

5. Results

5.1 Comparison of unconditional means

Our first evaluation exercise is based on pair-wise non-parametric comparisons of *mean* earnings obtained from our two data sources. The mean is a widely used measure of centre of distributions, but, in finite samples, it is known to suffer from the presence of outliers. We repeated the whole analysis using the *median* as an alternative (more robust) measure of centre and obtained the same qualitative results as those found with the mean. For ease of presentation we focus only on the mean-based results.

Figures 1 and 2 report means and 95% confidence bands of current and usual earnings, respectively. Each figure contains six panels, which report the baseline definition of earnings (panel A) and its five variants (panels B to F), and each panel shows data from both comparison years. For each measure, Table 1 complements Figures 1 and 2, providing information on sample sizes by source and year as well as *p*-values from equality-of-mean tests.

We emphasize three findings. First, there appear to be stronger similarities between surveys in the 1995/96 comparisons than in the later 2003/04 comparisons, regardless of whether we use current or usual pay and, generally, irrespective of the earnings definition. For example, using baseline definitions in panel A, the BHPS current earnings mean of £300.72 for 1995 is not statistically significantly different from the corresponding FRS figure of £294.10 (with a *p*-value for the mean equality test of 0.258 in Table 1). Similar evidence emerges if we focus on the baseline definition of usual earnings (with a *p*-value of 0.149). The story, however, is different if our mean comparisons are computed on the 2003/04 surveys. For either current or usual baseline earnings measure, the hypothesis that the means are equal in the two surveys is rejected at conventional levels of statistical significance.

Second, we cannot find substantially different results along the homogeneity/sample size trade-off. With the less homogenous usual earnings measures, we detect only one statistically significant difference among the 1995/96 BHPS-FRS means comparisons (i.e., when the AEI indexing is not applied). Similarly, not consulting payslips in the 1995/96 current earnings exercise for BHPS leads to significant cross-survey differentials. In the 2003/04 exercise, instead, both measures detect differences in three out of six comparisons.

Third, it is not the case that the baseline definition of current earnings results in greater between-survey homogeneity than the other earnings definitions; similarly, the baseline definition of usual earnings does not result in greater between-survey mean difference than the other, more selective, usual earnings measures.

Table 1. Sample sizes and goodness-of-fit tests for mean earnings, by survey, year and earnings definition

	1995/96			2003/04		
	Sample size		Test of	Sample size		Test of
Earnings definition	BHPS	FRS	equality	BHPS	FRS	equality
Current earnings						
Baseline	1,450	12,323	0.258	1,226	10,403	**0.026**
Imputed cases included	1,450	12,332	0.261	1,226	10,443	**0.026**
Weighted data	1,450	12,323	0.070	1,217	10,403	**0.040**
AEI indexing removed	1,450	12,323	0.506	1,226	10,403	0.191
Payslip not consulted included	4,045	12,323	**0.000**	3,864	22,190	0.565
Booster samples and NI included				1,707	10,933	0.103
Usual earnings						
Baseline	4,426	21,497	0.149	7,574	24,850	**0.014**
Imputed cases dropped	3,454	17,672	0.439	6,046	21,742	**0.033**
Weighted data	4,424	21,497	0.529	7,465	24,850	0.728
AEI indexing removed	4,426	21,497	**0.046**	7,574	24,850	**0.000**
Non-OSMs dropped (BHPS only)	3,867	21,497	0.251	3,367	24,850	0.310
Booster samples and NI dropped				4,329	23,319	0.233

Note. In the 'Test of equality' column, the figures are *p*-values from *t*-tests of equality between the BHPS means and the corresponding FRS means. In bold are the cases in which the cross-survey difference is statistically significant at the 5% level. All the underlying earnings measures are deflated using the Retail Price Index (available from the Office for National Statistics) and are expressed in constant prices with April 1995/March 1996 set equal to 100.

Taken together, these results suggest that the main driver of BHPS-FRS mean differentials in gross earnings may be a BHPS attrition/vintage effect, that is, the later 2003/04 comparisons fare generally slightly worse than the earlier 1995/96 comparisons, regardless of the earnings definition (current versus usual) and irrespective of measure homogeneity. This, in turn, points to the importance of differential attrition in all types of earnings data collection which can affect the BHPS samples, given their longitudinal nature. Interestingly, analyses based on measures of usual earnings do not typically lead to greater departures between sources (as opposed to those based on more homogeneous current earnings definitions), especially if weighted data are used.

5.2 Comparison of means by characteristics

Similarities or differences at the aggregate level may only reveal part of the story. We thus repeat our analysis after stratifying workers in each sample along a number of individual demographic and socio-economic characteristics. The results of this analysis are in Table 2, which shows *p*-values of equality tests. For the sake of brevity, we only report the results obtained with baseline definitions of current and usual earnings. Since these measures are at the opposite extreme of the homogeneity/sample size trade-off, they are likely to identify more extreme cross-survey differences.

For the 1995/96 comparisons, the FRS means are always not significantly different from their BHPS counterparts along all characteristics when usual earnings measures are used, whereas they are significantly different in the case of women when we use current earnings (*p*-value=0.029). The story is reversed in the case of the 2003/04 comparisons. In this case, two out of eight cross-survey comparisons deliver statistically different means with current earnings (i.e., for women and single individuals), while four of the eight comparisons are different when we consider usual earnings.

Despite these differences between usual and current earnings measures, the results of this exercise are similar to those found with unconditional means: that is, there is again evidence of a BHPS vintage effect, whereby greater cross-survey differences are detected in 2003/04 than in 1995/96. As before, this suggests that differential attrition might play a role.

Table 2. Tests of equality in mean earnings in the BHPS and FRS surveys, by year, earnings definition, and individual characteristic

	1995/96		2003/04	
Individual characteristic	Current	Usual	Current	Usual
Male	0.263	0.160	0.306	**0.024**
Female	**0.029**	0.153	**0.000**	0.466
Aged less than 25	0.341	0.861	0.052	**0.000**
Aged 25-50	0.086	0.770	0.181	**0.004**
Aged more than 50	0.254	0.076	0.651	0.142
Single	0.094	0.913	**0.045**	0.988
Couple	0.673	0.144	0.255	**0.000**
Other	0.663	0.928	0.145	0.058

Note. The figures are *p*-values from *t*-tests of equality between the BHPS means and the corresponding FRS means conditional on each characteristic. In bold are the cases in which the cross-survey difference is statistically significant at the 5% level.

5.3 Comparison of distributions

A further nonparametric exercise is to compare the pay measures in the two surveys along their entire distributions. Figure 1 presents the kernel density estimates of current and usual earnings using baseline definitions. Eye-balling these graphs, we can easily see that the BHPS and FRS distributions are fairly close to each other in 1995/96 (panels A and B), and especially in the case of usual earnings. By 2003/04, however, the two distributions are quite different, with the BHPS densities being typically below the FRS densities at low levels of earnings and above them in the middle part or in the right tail (panels C and D).

This evidence is complemented by Kolmogorov-Smirnov (KS) tests of equality of the distributions, which are reported in Table 3. The table shows *p*-values of such tests for both unconditional distributions (given in the first row) and distributions conditional on specific socio-economic characteristics (given in the remaining rows of the table). As suggested by Figure 1, we cannot reject the null hypothesis that the 1995/96 FRS and BHPS (unconditional) usual earnings distributions are identical at conventional level. This result emerges also when we compare current earnings in the same year, although in such a case the *p*-value is only 0.056. Similar estimates are found with Mann-Whitney two-sample statistics, which test the hypothesis that our two independent samples are from populations with the same distribution and, differently from the KS test, are more powerful against differences in the tails of distributions (Conover, 1999).

Table 3. Tests of equality in earnings distributions in the BHPS and FRS surveys, by year, earnings definition, and individual characteristic

	1995/96		2003/04	
	Current	Usual	Current	Usual
Unconditional	**0.056**	0.385	**0.000**	**0.000**
Male	**0.011**	0.380	**0.002**	**0.000**
Female	**0.016**	0.058	**0.000**	**0.000**
Aged less than 25	0.543	0.347	**0.028**	**0.000**
Aged 25-50	**0.038**	0.505	**0.000**	**0.018**
Aged more than 50	0.761	0.252	0.443	0.477
Single	0.176	0.608	**0.001**	**0.002**
Couple	0.092	0.733	**0.001**	**0.005**
Other	0.933	0.924	0.078	0.053

Note. Figures are *p*-values from Kolmogorov-Smirnov tests of equality between the BHPS distributions and the corresponding FRS distributions. In bold are the cases in which the cross-survey difference is statistically significant at the 5% level.

After stratifying workers by socio-economic characteristics, we find again that the two surveys generate statistically identical usual earnings distributions in all the 1995/96 comparisons. But this is not the case when we employ our current earnings measure. For this measure, in three out of the eight conditional distributions (for male and female workers, and for employees aged 25-50) we can reject the hypothesis of equality. For the 2003/04 comparisons, instead, BHPS and FRS earnings data are significantly different regardless of whether we look at unconditional or conditional distributions (the only two exceptions being observed among employees aged more than 50 and among workers whose marital status is neither married nor single) and irrespective of the earnings definition.

Therefore, similarly to what emerged from the analysis of means, differences in results by survey do not appear to be sentivive to the homogeneity/sample size trade-off, with the greater homogeneity for current-type earnings measures being presumably offset by their lower statistical power. Cross-survey differences instead seem again to be driven by the BHPS attrition/vintage effect, with a sizable deterioration of all comparisons between the 2003/04 distributions and their 1995/96 counterparts. This, in turn, is likely to be explained (at least in part) by differential attrition.

6. Does the Difference Matter for Substantive Analysis?

It is important to establish whether the differences and similarities discussed in the previous section matter for substantive analysis. For this purpose, we perform three exercises. The first exercise looks separately at the probability that an individual's earnings falls below 60% of median earnings and at the probability that an individual's earnings are in the top decile group of the earnings distribution. In the second exercise, earnings are again the dependent variable, but are used as a continuous variable. In this case, we estimate OLS wage equations and also perform quantile regressions. The third exercise considers the probability of participating to an occupational pension scheme for which earnings are one of the (key) explanatory variables. In all exercises, we use simple statistical specifications containing a small number of covariates which allow us to limit sample selection issues as much as possible. Our goal, in fact, is not to come up with a definitive model of, say, the risk of low earnings or of wage determination, but to provide convincing cross-survey comparisons of earnings data.

After pooling FRS and BHPS data, we interact each covariate with a sample indicator (which takes value one if the worker is from the BHPS, and zero otherwise). Such interaction terms are the focus of our attention. If they are not significantly different from zero, then we take this as evidence that the FRS and BHPS earnings data deliver equivalent results in terms of the outcome variable under analysis. If, instead, they are statistically different from zero, then there is evidence of a discrepancy between the two data sources.

Specifically, in the first two sets of cross-survey comparisons, and for each of the two survey years, we separately estimate

$$y_i = \mathbf{X}_i'\alpha + \mathbf{X}_i' \times \mathrm{I}(\mathrm{BHPS}_i = 1)\beta + \varepsilon_i, \qquad (1)$$

where y_i is the earnings outcomes of interest (e.g., the probability of having low earnings or earnings in the top decile, or the logarithm of monthly earnings) for worker *i*, the term $\mathrm{I}(\mathrm{BHPS}_i = 1)$ is a function indicating that worker *i* belongs to the BHPS, $\mathbf{X}_i$ is a vector of year-specific worker's characteristics, and ε_i is an i.i.d. error term, with $E(\varepsilon_i \mid \mathbf{X}_i, \mathrm{BHPS}_i) = 0$, where $E(\cdot)$ is the mathematical expectation operator. Our interest is in the parameter vector β. For models of the probability of having low or high earnings, we present results from probit regressions. (Similar results were obtained with logit regression and linear probability models, which for convenience are not reported.) In the case of (log) earnings functions, we also estimate (1) using quantile regression methods (Koenker, 2005); in such cases, there is a vector of β coefficients at each estimated percentile. For both exercises, our analysis is mainly based on the two extreme measures of earnings (baseline current and baseline usual), although we also briefly discuss the results obtained with the other intermediate (non-baseline) earnings measures. Moreover, for all exercises, we use same set of covariates, $\mathbf{X}$, which include dummy variables for sex, age (3 dummy variables), whether or not an individual left school before age 18, whether he/she has a child, whether he/she is employed full time (as opposed to part time), marital status (3 dummy variables) and country (4 dummy variables). The reference individual is a woman aged more than 50, who did not leave school before age 18, is childless, has a partner, has a part-time job and lives in England.

In the third set of evaluations, our multivariate probit analysis takes the form (see Maddala (1983), pages 22-27)

$$\Pr(p_i = 1 \mid \mathbf{X}) = \Phi\left[\mathbf{X}_i'\alpha + \mathbf{X}_i' \times \mathrm{I}(\mathrm{BHPS}_i = 1)\gamma + \beta_1 w_i + \beta_2 w_i \times \mathrm{I}(\mathrm{BHPS}_i = 1)\right], \qquad (2)$$

where p_i is a binary variable that takes value one if worker *i* contributes to an occupational pension scheme and zero otherwise, w_i is the (continuous) earnings measure of interest (e.g., current earnings or usual earnings), and Φ is the cumulative distribution function of the standard normal distribution. The parameters of interest are β_1 and, especially, β_2, which along with the other parameters have been estimated by maximum likelihood. Estimations of (2) with linear probability models and logistic regressions yielded the same results as those shown and discussed later, and are therefore not reported.

6.1 At the Extremes of the Pay Distribution

6.1.1 Low Earnings

We perform our analysis on (1) for the case in which the 60% cut-off line is computed on the FRS earnings distribution separately from the case in which the cut-off line is determined on the BHPS distribution. Because this exercise is repeated for the two measures of earnings (baseline current and baseline usual) and for each of the two years under analysis, we have eight different cut-offs, which are reported in parentheses in Table 4. The 1995/96 current earnings cut-off values are £162.58 and £151.47 for the BHPS and FRS samples respectively; the corresponding 2003/04 values are £177.70 and £164.39. The usual earnings cut-off values are slightly lower and their absolute between-survey differences are also smaller. Table 4 also presents the β cross-survey comparison estimates from each of the eight regressions, but for convenience the other estimates are not shown. (Complete results are available from the authors upon request.)

Table 4. Probability of being below 60% of median earnings, by survey period and type of earnings measure†

	1995/96		2003/04	
	Current	Usual	Current	Usual
FRS low earnings line	*(£151.47)*	*(£136.18)*	*(£164.39)*	*(£150.79)*
Male	-0.079**	-0.041**	-0.103**	-0.043**
Left school before age 18	0.019	0.009	0.038	0.000
Has children	-0.024	0.012	-0.006	0.000
Wales	-0.099*	-0.044	-0.129**	0.021
Scotland	-0.021	0.011	-0.026	0.033
Northern Ireland				0.012
Working full time	-0.013	0.009	-0.015	-0.003
Aged below 25	-0.011	-0.038	-0.051	-0.065**
Aged 25-50	0.032	-0.035	-0.049	-0.049**
Single	0.000	0.067**	0.028	0.032
Widowed, separated, divorced	0.006	0.001	0.001	-0.023
Joint significance†	**0.011**	**0.034**	**0.000**	**0.000**
BHPS low earnings line	*(£162.58)*	*(£135.50)*	*(£177.70)*	*(£155.60)*
Male	-0.066*	-0.041**	-0.102**	-0.042**
Left school before age 18	0.029	0.012	0.031	0.004
Has children	-0.026	0.013	-0.033	-0.002
Wales	-0.095	-0.043	-0.170**	0.03
Scotland	-0.034	0.014	-0.059	0.029
Northern Ireland				0.018
Working full time	-0.008	0.007	-0.024	-0.009
Aged below 25	-0.023	-0.039	-0.026	-0.064**
Aged 25-50	0.023	-0.035*	-0.038	-0.050**
Single	-0.016	0.070**	-0.002	0.029
Widowed, separated, divorced	-0.004	0.000	0.005	-0.01
Joint significance†	0.085	**0.032**	**0.000**	**0.000**

Note. Figures are marginal effects on the β coefficients (see equation (1)) of the BHPS interaction terms obtained from probit regressions.
† Figures are the p-values of the tests of joint significance of the β interaction terms (in bold are the cases in which the BHPS-FRS difference of the β interaction terms is jointly statistically significant at conventional levels). Notice that the 2003/04 estimates for usual earnings are insensitive to whether the line is drawn with FRS or BHPS data. This is because the two lines are virtually at the same level (£155.21 on FRS and £155.19 on BHPS) and, thus, define identical dependent variables.
** statistically significant at the 0.01 level; * statistically significant at the 0.05 level.

In general, the determinants of the probability that a worker has earnings below 60% of median earnings differ depending on whether we use BHPS or FRS data. This is true regardless of the choice of earnings measure, sample year, and cut-off line. The only exception emerges for the 1995/96 comparison when current earnings and BHPS cut-off line are used, for which we cannot reject the hypothesis that the sets of estimates obtained from the two surveys are statistically identical at conventional levels of significance (*p*-value=0.085).

The variables along which we detect the largest departures between the two samples are sensitive to type of measure and survey year. But 'single', 'aged 25-50', 'Wales' and especially 'male' tend to pick up most of the cross-survey variation. For example, in the 1995/96 current earnings case with the FRS cut-off line, a male worker in the FRS sample is predicted to be 14 percentage points less likely to be below the earnings cut-off than the baseline woman (not shown). A male worker in the BHPS is a further 8 percentage points less likely. In the 2003/04 surveys, the corresponding FRS-BHPS differential for male workers goes up to 10 percentage points.

Despite these departures, the 2003/04 β estimates are not significantly greater than the corresponding 1995/96 estimates. If cross-survey differences arise primarily from the deteriorating quality of the BHPS data as the panel becomes older, then the time differences in the estimated β coefficients in Table 4 do not seem large enough to justify an interpretation of our results entirely based on differential attrition. Although differential attrition is likely to play a role (and appeared to be a credible explanation in our non-parametric analysis), there might be therefore other dimensions to consider, such as missing data issues on all covariates and not just the earnings variables, specification errors in (1), and unobserved heterogeneity affecting both (1) and the attrition process. Accounting for the effect of any such processes is beyond the scope of this paper, but we will come back to these issues in the conclusions.

5.1.2 Top Decile

Table 5 reports *p*-values of the tests of joint significance for the cross-survey comparisons on the probability that a worker is observed in the top decile of the earnings distribution by year and type of earnings measure. The β estimates associated with the covariates are not shown because of space concerns. Again, the eight cut-offs, which vary depending on year, earnings definition and survey, are in parentheses. With the exception of the 2003/04 current earnings figures, the year-specific differences in such values between the two surveys are small.

Table 5. Probability of being in the top decile group of earnings, by survey period and type of earnings measure

	1995/96		2003/04	
	Current	Usual	Current	Usual
FRS top decile	*(£516.84)*	*(£480.36)*	*(£585.94)*	*(£552.94)*
Joint significance	0.089	0.072	0.142	**0.000**
BHPS top decile	*(£529.33)*	*(£484.96)*	*(£625.85)*	*(£555.27)*
Joint significance	**0.046**	0.085	0.112	**0.001**

Note. In bold are the cases in which the BHPS-FRS difference of the β interaction terms is jointly statistically significant at conventional levels.

When the dependent variable is defined using current earnings measures, the two surveys do not produce significantly different estimates, except for the 1995/96 case if the top decile value is determined on the BHPS sample (*p*-value=0.046). If instead the dependent variable is based on usual earnings, we find that the cross-survey differences in 2003/04 are significantly greater than the corresponding differentials in 1995/96, regardless of whether the top decile value comes from the BHPS or the FRS sample. This suggests that differential attrition/vintage effect arguments could be relevant in this case. Interestingly, most of the significant cross-survey time differences are driven by education ('left school before age 18') and employment status ('working full time').

6.2 Wage Determination and Sensitivity Checks

We estimate equation (1) with log earnings as our dependent variable by ordinary least squares (OLS) and quantile regressions at the 10-th, 25-th, 50-th (median), 75-th and 90-th percentiles. Table 6 presents the *p*-values for the joint significance tests of the β interaction terms by measure type and year.

Irrespective of earnings measure and survey year, the OLS results reveal that the two surveys always produce estimates that are statistically different from each other. The results are slightly more mixed when we look at the whole set of quantile regressions, with 9 out of the 20 comparisons for which we cannot reject the null hypothesis of equal effects. In the case of current earnings, there is no systematic pattern of results by quantile. When we use usual earnings, however, we find that the cross-survey differences are greater in the 2003/04 comparisons than in the 1995/96 comparisons for all quantiles, except in the case of the 25-th percentile. This pattern can be (at least partly) explained by the presence of differential panel attrition. Most of the 2003/04 BHPS-FRS differentials are captured by the effects of age, employment status and, to a lesser extent, education.

Table 6. Joint significance tests on the β interaction terms in log earnings OLS and quantile regressions, by survey period and type of earnings measure

	OLS (mean)	0.10	0.25	0.50	0.75	0.90
Current						
1995/96	**0.021**	0.107	**0.000**	0.285	**0.026**	0.060
2003/04	**0.000**	**0.000**	**0.000**	**0.018**	0.140	0.170
Usual						
1995/96	**0.005**	0.542	**0.008**	0.298	0.309	0.074
2003/04	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**	**0.004**

Note. Figures are the *p*-values of the tests of joint significance of the β interaction terms (in bold are the cases in which the BHPS-FRS difference of the β interaction terms is jointly statistically significant at conventional levels).

To check for robustness, we repeated all the multivariate analyses presented so far (probability of having low earnings, probability of being in the top decile, and wage determination) using the other non-baseline earnings definitions. For the sake of brevity we cannot show such results but only summarise their two key implications. First, the pattern of results reported in Tables 4-6 and discussed above is broadly confirmed. In particular, even in the absence of a systematic relationship by type of measure and outcome, the cross-survey differences in 2003/04 tend to be greater than those in 1995/96. These differentials are typically accounted for by age and gender in the case of the probability of having low earnings, and by age, education, and employment status in the cases of the probability of being in the top decile and earnings equations. Second, and in line with what we found from the non-parametric analysis of Section 5, we cannot detect significant differences along the homogeneity/sample size trade-off. That is, there is no substantial gain in statistical power by reducing homogeneity i.e., by moving away from the baseline current earnings measure. Likewise, we do not reject the null hypothesis of equal effects less often when using more homogenous measures on smaller samples.

6.3 Occupational pension plan participation

In our last validation exercise, the earnings variable is on the right-hand side of a multivariate regression model as in equation (2). In this case, we estimate the probability that an employee contributes to an occupational pension scheme with a probit regression model and a specification that is commonly used in the statistical analysis of pension plan entitlements (Disney and Whitehouse, 1996; Lumsdaine and Mitchell, 1999; Ginn and Arber, 2000; Banks and Smith, 2006; Barr and Diamond, 2006).

The dependent variable takes value one if a worker belongs to an occupational pension plan, and zero otherwise. The BHPS records this information in the two variables JBPEN ("*Does your present employer run a pension scheme or superannuation scheme for which you are eligible*?") and JBPENM ("*Do you belong to your employer's pension scheme*?"), which are asked to all currently employed workers and have not changed between the fifth and the thirteenth waves. In the FRS, things are different. In the 1995/96 survey, the variable EMPPENS collects responses from the question: "*Thinking of your present job, do you currently belong to a pension or superannuation scheme run by your employer which will give you a pension when you retire*?", asked to all individuals who are currently employed. In the 2003/04 survey, the question recorded in the variable EMPPAY2 changes into: "*Are you (or your employer) paying contributions to any of the pension arrangements shown on this card? (1. A personal of private pension fund, or retirement annuity; 2. A company or occupational pension scheme run by my employer; 3. A stakeholder pension scheme fund; 4. None of these)*", and this is asked to all respondents aged 65 or less who are currently working or who have worked previously.

Beside (log) earnings, our regressions contain the same set of explanatory variables $\mathbf{X}$ used in the previous analysis when the earnings variable was on the left-hand side of the model (age, education, sex, region, employment status, and marital status).

Table 7 reports the results on the β_1 and β_2 coefficients of (2), that is, the average log wage effect and the differential wage effect of the BHPS survey respectively, separately for each year and for the two baseline earnings measures. To ease interpretation, they are expressed as marginal effects, while the estimates on the $\mathbf{X}$ variables are not shown for the sake of brevity. The table also shows *p*-values of the χ^2 tests for the joint significance of all survey interaction terms, including and excluding the earnings interactions.

Table 7. Comparing the effect of earnings on the probability of participating to an occupational pension plan across surveys, by survey period and type of earnings measure†

	1995/96		2003/04	
	Current	Usual	Current	Usual
Earnings (β_1)	0.375**	0.365**	0.285**	0.307**
	(0.010)	(0.007)	(0.009)	(0.006)
Earnings $\times$ BHPS (β_2)	0.007	0.004	0.022*	0.020**
	(0.010)	(0.006)	(0.010)	(0.005)
Joint significance of the β_2 and γ parameters†	0.066	0.332	**0.000**	**0.000**
	[11]	[11]	[11]	[12]
Joint significance of the γ parameters†	0.290	0.349	0.290	0.134
	[10]	[10]	[10]	[11]
Pseudo R^2	0.167	0.193	0.129	0.152
Mean of the dependent variable	0.571	0.495	0.564	0.495
Number of observations	13,747	25,810	11,604	32,217

Note. Figures are marginal effects on the β_1 and β_2 coefficients (see equation (2)) of the earnings and earnings-BHPS interaction terms obtained from probit regressions. Standard errors are in parentheses.
† In each cell, the top figure is the *p*-value of the chi-square test of significance of the estimated parameter, while the bottom figure in square bracket is the corresponding number of degrees of freedom.

At the bottom of the table, below the statistics on overall fit, we present the proportion of workers who participate in an occupational pension scheme and the size of each estimating sample. Such proportions vary slightly in relation to the earnings definition, which in turn affects the sample size. For the model with current earnings, 56-57% of workers in the pooled FRS-BHPS sample contribute to an occupational plan in either period, with the BHPS workers reporting a slightly greater propensity to contribute (61% versus 56% among FRS workers). The fraction of contributors is smaller (around 50%, both in the pooled sample and in each survey-specific sample) in the models with usual earnings, and these are also the models estimated on larger samples. Albeit not a formal test, the similarity of such proportions across the two surveys suggests that the above-mentioned changes in question routing and wording in the FRS questionnaires are unlikely to be the sources of substantial differences in our comparative exercise.

From the regression analysis, we detect no significant cross-survey difference from the 1995/96 comparisons, irrespective of the earnings definition. This is the case for the effect of earnings (which can be seen directly from β_2) as well as for the impact of all other covariates (as the tests on only the γ coefficients demonstrate). The 2003/04 comparisons, however, reveal a different story. As before, earnings increases are associated with a greater likelihood of pension contribution among all workers in both surveys. But, ceteris paribus, BHPS workers are now about 2 percentage points more likely to contribute to occupational pension schemes than their FRS counterparts.

We performed a number of robustness checks by repeating the whole analysis on the other (non-baseline) earnings definitions. The results (not shown) echo those reported in Table 7. In particular, β_2 is always small and, regardless of the earnings definition, never statistically different from zero in the early (1995/96) comparisons. In the 2003/04 comparisons instead, it almost invariably becomes larger and statistically significant, indicating a departure in the wage effect on the probability of occupational pension contribution between the two surveys.

7. Conclusions

This paper has performed a statistical comparison of the earnings data collected in the BHPS with the earnings data collected in the FRS. The cross-survey comparison is based on current earnings and usual earnings (as well as a number of variants of such baseline definitions) observed at two different points in time (1995/96 and 2003/04). The multiplicity of measures and survey periods gives us a broad range of dimensions over which our comparisons can be evaluated and a direct check of robustness. Two types of analysis have been performed. The first uses a set of non-parametric tests of equality at the centre (mean) of the distributions and of the whole earnings distributions. The second uses multivariate regressions to check whether the two data sources yield different results in relation to some illustrative exercises typical of uses of earnings data (such as the probability of being at the bottom or at the top of the earnings distribution, the estimation of earnings functions, and the probability of belonging to an occupational pension plan).

From the non-parametric exercises, we find that the 1995/96 comparisons deliver results that are typically closer between the two surveys than the 2003/04 comparisons. Changing measure homogeneity or sample size, as captured by our different earnings definitions, does not lead to different conclusions. The fact that cross-survey differences in means and along the entire earnings distributions tend to amplify over time therefore suggests that both attrition and vintage effects, which may reduce the representativeness of the BHPS panel (but not, or less so, of the FRS data), are likely to play a major role.

The regression analyses strongly uphold these BHPS vintage effects. Whether we look at the probability of having low earnings or at the probability of being in the top decile of the earnings distribution, or whether we consider earnings equations or models of occupational pension contribution with earnings as an explanatory variable, the latest (2003/04) cross-survey comparisons are almost always further apart than the earlier comparisons. Again, this indicates that the differences between the two surveys increase with time, which, for the BHPS, coincides with its age.

Where does this evidence point to? And what can be learnt for future data collections? In large-scale (representative), multi-purpose (general) household surveys, reliable income data are always hard to gather. The FRS procedure of retaining in the official release of the data individuals who provide valid information on a minimum number of income questions is likely to produce high quality data but may not be applicable to all survey designs. The less demanding BHPS procedures, however, do not seem to be inferior in terms of gaining similarly reliable information. Other attempts to increase data reliability (for instance, using only original sample members in the BHPS, restricting attention to employees whose payslips are consulted and whose earnings records are not imputed) affect the statistical power of our cross-survey comparisons only marginally and, overall, appear to induce relatively small differences. By and large, instead, the process that seems to drive most of the cross-survey differences in our comparative work has to do with the possible quality deterioration (in terms of population representativeness) of the BHPS data as the panel becomes older. This consistently points to differential attrition and vintage effects, and emphasizes the importance of modelling them in panel data research (as in Cappellari and Jenkins, 2008), especially when analysts are required to follow the same individuals over long time periods.

References

Atkinson, A.B. and Micklewright, J. (1983) On the reliability of income data in the Family Expenditure Survey 1970-1977. *Journal of the Royal Statistical Society*, Series A, **146**, 33-61.

Banks, J. and Smith, S. (2006) Retirement in the UK. *Oxford Review of Economic Policy*, **22**, 40-56.

Barr, N. and Diamond, P. (2006) The economics of pensions. *Oxford Review of Economic Policy*, **22**, 15-39.

Biemer, P. B., Groves, R. M., Lyberg, L. E.,Mathiowetz, N. A. and Sudman, S. (eds) (1991) *Measurement Errors in Surveys*. New York: Wiley.

Billingsley, P. (1995) *Probability and Measure*. New York: Wiley (third edition).

Blundell, R., Brewer, M. and Francesconi, M. (2008) Job changes and hours changes: understanding the path of labor supply adjustment. *Journal of Labor Economics*, **26**, 421-453.

Böheim, R. and Jenkins, S.P. (2006) A comparison of current and annual measures of income in the British Household Panel Survey. *Journal of Official Statistics*, **22**, 733-758.

Brown, G., Micklewright, J., Schnepf, S.V. and Waldmann, R. (2007) International surveys of educational achievement: how robust are the findings? *Journal of the Royal Statistical Society*, Series A, **170**, 623-646.

Cappellari, L. and Jenkins, S.P. (2008) Estimating low pay transition probabilities accounting for endogenous selection mechanisms. *Journal of the Royal Statistical Society*, Series C, **57**, 165-186.

Chen, H.Y. and Little, R. (1999) A test of missing completely at random for generalised estimating equations with missing data. *Biometrika*, **86**, 1-13.

Conover, W.J. (1999) *Practical Nonparametric Statistics*. New York: Wiley.

Davies, H. and Joshi, H. (1998) Gender and income inequality in the UK 1968-1990: the feminization of earnings or of poverty? *Journal of the Royal Statistical Society*, Series A, **161**, 33-61.

Dayal, N., Gomulka, J., Mitton, L., Sutherland, H. and Taylor, R. (2000) Enhancing Family Resources Survey income data with expenditure data from the Family Expenditure Survey: data comparisons. Microsimulation Research Note MU/RN/40, ISER, University of Essex. http://www.iser.essex.ac.uk/msu/publications/pdf/rn40.pdf.

Dex, S. and McCulloch, A. (1998) The reliability of retrospective unemployment history data. *Work, Employment & Society*, **12**, 497-509.

Disney, R. and Whitehouse, E. (1996) What are occupational pension plan entitlements worth in Britain? *Economica*, **63**, 213-238.

Elias, P. (1996) Who forgot they were unemployed? Unpublished. Institute for Employment Research, University of Warwick.

Ermisch, J. and Francesconi, M. (2000) Educational choice, families, and young people's earnings. *Journal of Human Resources*, **35**, 143-176.

Francesconi, M. (2005) An evaluation of the childhood family structure measures from the sixth wave of the British Household Panel Survey. *Journal of the Royal Statistical Society*, Series A, **168**, 539-566.

Ginn, J. and Arber, S. (2000) Personal pension take-up in the 1990s in relation to position in the labour market. *Journal of Social Policy*, **29**, 205-228.

Kalton, G. and Kasprzyk, D. (1986) The Treatment of Missing Survey Data. *Survey Methodology*, **12**, 1-16.

Kirri, C., Hill, C. and Whitfield, P. (2005) *Family Resources Survey: Annual Technical Report: 2003/04*, London: Office for National Statistics. Available at <http://www. statistics.gov.uk/StatBase/Product.asp?vlnk=9267>.

Koenker, R. (2005) *Quantile Regression*. Cambridge: Cambridge University Press.

Little, R. and Rubin, D. (1987) *Statistical Analysis with Missing Data*. New York: Wiley.

Lumsdaine, R.L. and Mitchell, O.S. (1999) New developments in the economic analysis of retirement. In *Handbook of Labor Economics*, volume 3C (eds. O.C. Ashenfelter and D. Card). Amsterdam: Elsevier.

Lynn, P. (ed.), with Buck, N., Burton, J., Laurie, H. and Uhrig, S.C.N. (2006) *Quality Profile: British Household Panel Survey: Waves 1 to 13, 1991-2003.* Institute for Social and Economic Research, University of Essex, Colchester. Available at <http://www.iser.essex.ac.uk/ulsc/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf>.

Maddala, G.S. (1983) *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

Madow, W., Olkin, I. and Rubin, D., eds. (1983*) Incomplete Data in Sample Surveys*, 3 volumes. New York: Academic Press.

Mathiowetz, N.A. and Duncan, G.J. (1988) Out of work, out of mind: response errors in retrospective reports of unemployment. *Journal of Business and Economic Statistics*, **6**, 221-29.

Micklewright, J. and Schnepf, S.V. (2007) How reliable are income data collected with a single question? Unpublished. Southampton Statistical Science Research Institute, Application and Policy Working Paper A07/08, University of Southampton.

Mood, A.M., Graybill, F.A. and Boes, D.C. (1974) *Introduction to the Theory of Statistics*. New York: McGraw-Hill (third edition).

Morgenstern, R.D. and Barratt, N.S. (1974) The retrospective bias in unemployment reporting by sex, race and age. *Journal of the American Statistical Association*, **69**, 355-357.

Robins, J.M., Rotnitzky, A. and Zhao, L.P. (1995) Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of the American Statistical Association*, **90**, 106-121.

Rubin, D. (1996) Multiple imputation after 18+ years, *Journal of the American Statistical Association*, **91**, 473-489.

Rudas, T., ed. (2008) *Handbook of Probability: Theory and Applications*. Los Angeles: Sage.

Stewart, M.B. (2004) The impact of the introduction of the U.K. Minimum Wage on the employment probabilities of low-wage workers. *Journal of the European Economic Association*, **2**, 67-97.

Sutherland, H. and Piachaud, D. (2001) Reducing child poverty in Britain: an assessment of government policy 1997-2001. *Economic Journal*, **111**, F85-F101.

Taylor, M.F. (ed.), with Brice, J., Buck, N. and Prentice-Lane, E. (2006) *British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices.* Institute for Social and Economic Research, University of Essex, Colchester. Available at <http://www.iser.essex.ac.uk/ulsc/bhps/doc/pdf_versions/ volumes/bhpsvola.pdf>.

Figure 1. Means and 95% confidence bands of current earnings, by survey, year and earnings definition

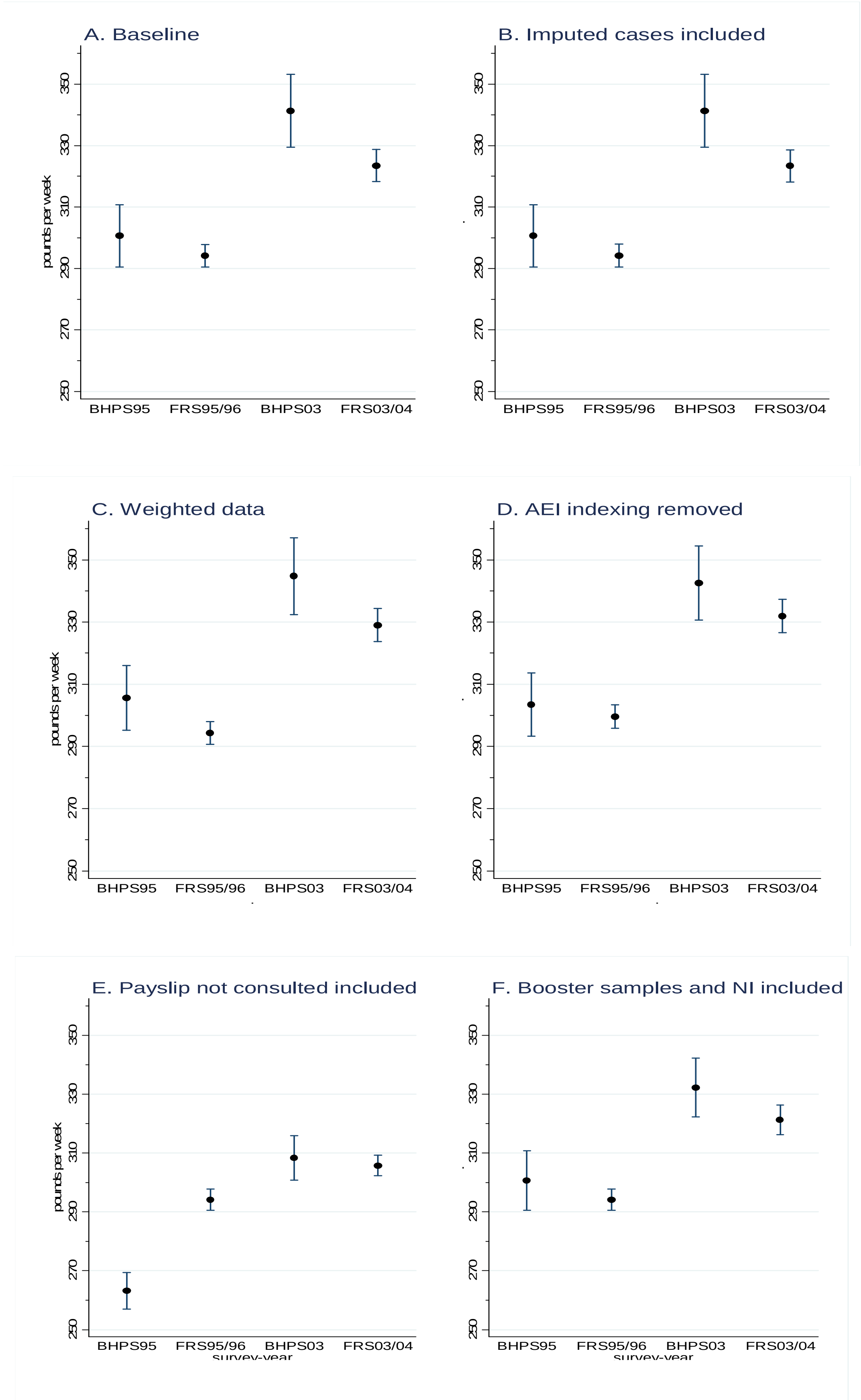


Figure 2. Means and 95% confidence bands of usual earnings, by survey, year and earnings definition

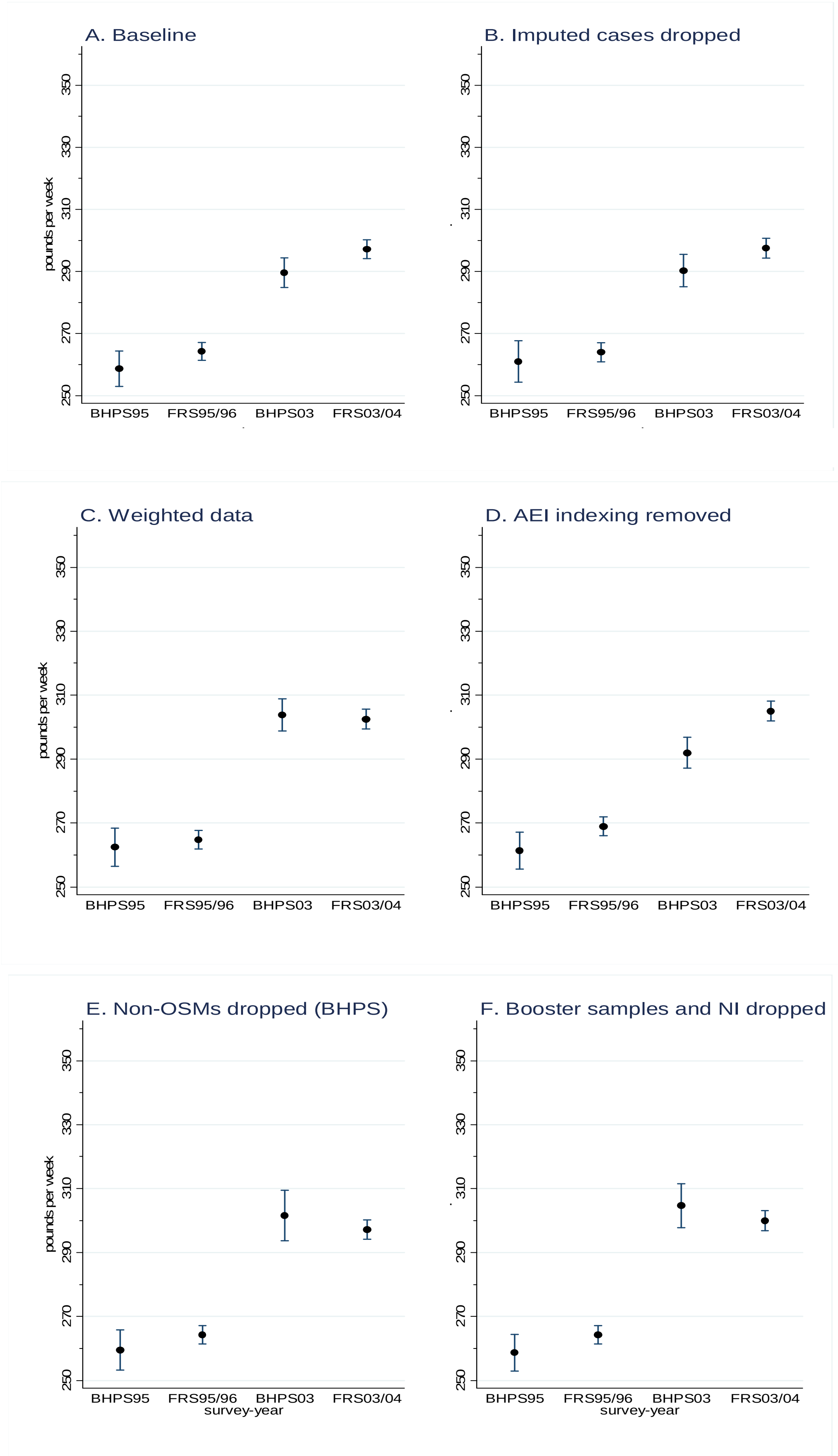
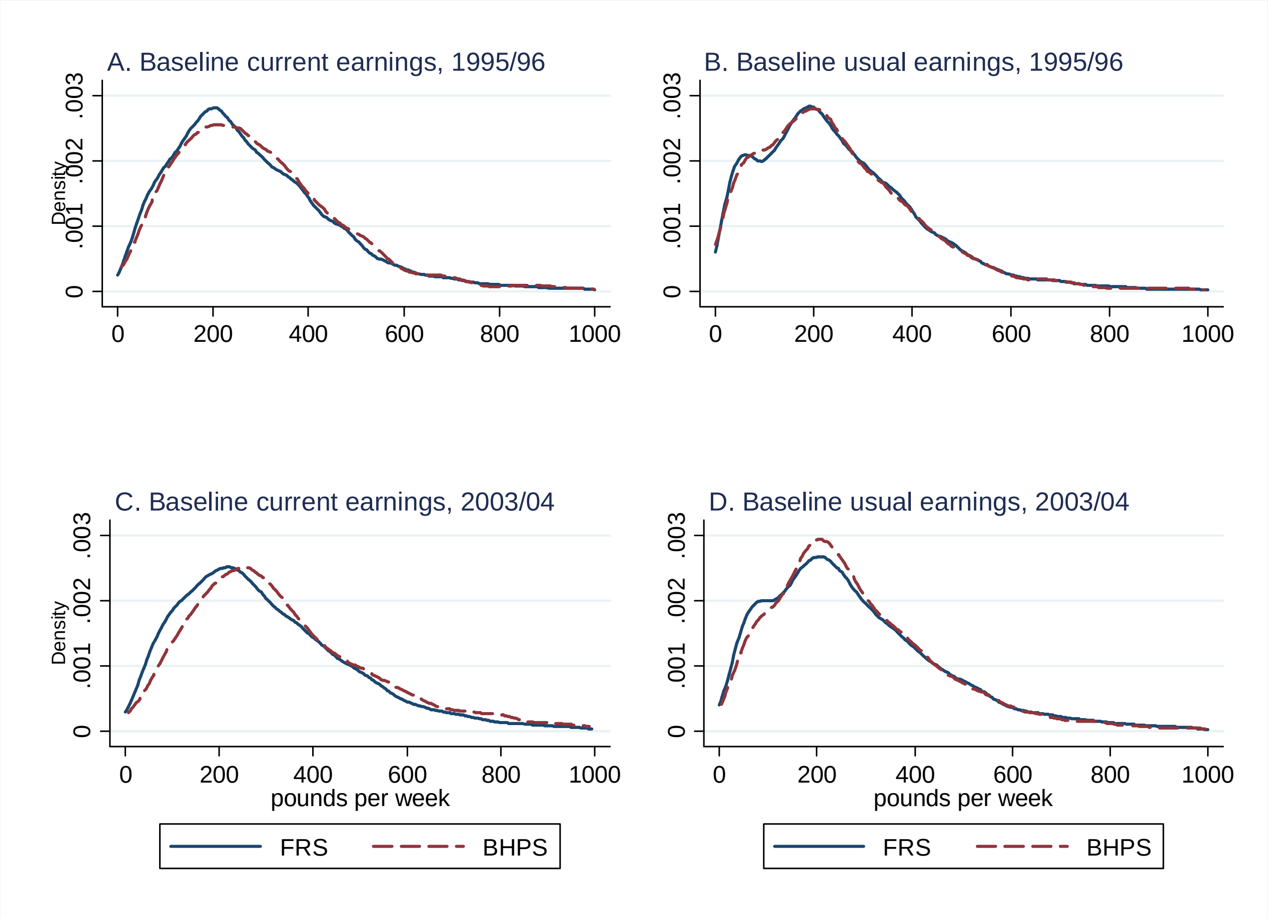


Figure 3. Estimated earnings density functions, by survey, year, and earnings definition



Appendix Table 1. Comparing FRS and BHPS along sample design and structure, data collection and adjustment procedures

	FRS	**BHPS**
Frame	Royal Mail's small users's PAF	Royal Mail's small users's PAF
Sample design	Stratified clustered probability (GB)	Equal probability clustered (wave one)
Stratification variables	Region Head socio-economic group Adult economic activity rate Male unemployment rate	Region Socio economic group profile Proportion of individuals of pensionable age Proportion of employed persons working in agriculture Proportion of persons living in single person non pensioners households
Later sampling	Random sample by region (NI) GB: 50% of PSU retained from previous year but new addresses chosen	W9: Scottish and Welsh booster samples W11: Extension in NI sample
Interview	Letter before interviewer call CAPI	Strategies to maintain the panel and unit response rates CAPI since wave 9
Questionnaire	Household and adults blocks 1 hour and 18 minutes, average length (per household) Rotated blocks	Household and individual questionnaire 45 minutes, average length (per individual)
Fieldwork	April – March	September – December (May)
Imputation	Hot decking Algorithms "Mop-up"	Hot decking Regression
Weighting	Correct for differential non-response Gross up sample estimates to the whole population	Unequal selection probability of addresses Household and individual non response Rescaling to unweighted sample size Cross-sectional weights ("fair shares approach")

Source: Lynn *et al*. (2006); Taylor *et al*. (2006); Kirri *et al*. (2005)