I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14405

Reconciling Reports: Modelling Employment Earnings and Measurement Errors Using Linked Survey and Administrative Data

Stephen P. Jenkins
Fernando Rios-Avila

MAY 2021

I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 14405

Reconciling Reports: Modelling Employment Earnings and Measurement Errors Using Linked Survey and Administrative Data

Stephen P. Jenkins
LSE and IZA

Fernando Rios-Avila
The Levy Institute

MAY 2021

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Reconciling Reports: Modelling Employment Earnings and Measurement Errors Using Linked Survey and Administrative Data*

We contribute new UK evidence about measurement errors and employment earnings to a field dominated by findings about the USA. We develop and apply new econometric models for linked survey and administrative data that generalize those of Kapteyn and Ypma (Journal of Labor Economics, 2007). Our models incorporate mean-reverting measurement error in administrative data in addition to linkage mismatch and mean-reverting survey measurement error and 'reference period' error, while also allowing error distributions to vary across individuals. Annualised survey earnings underestimate true annual earnings on average. Mean-reversion in survey measurement errors is absent. Both earnings sources underestimate true earnings inequality. The survey earning measure is more reliable than the administrative data earnings measure, but hybrid earnings predictors based on both sources are distinctly more reliable than either source-specific measure. The models with heterogeneous measurement error distributions indicate how data quality may be improved. For example, for survey quality, our results highlight how respondents showing payslips to interviewers have smaller survey error variances. For administrative data, our results suggest that greater error variances are associated with non-standard jobs, private sector jobs, and employers without good payroll systems.

JEL Classification: C81, C83, D31

Keywords: measurement error, earnings, survey data, administrative data, finite mixture models

Corresponding author:
Stephen P. Jenkins
Department of Social Policy
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

E-mail: s.jenkins@lse.ac.uk

* Thanks to the FRS team at the UK Department of Work and Pensions (DWP) for facilitating this project and for helpfully responding to queries. The FRS-P14 Linked Dataset is available from the DWP through a special secure-data contract. Thanks to Dean Hyslop and Erik Meijer for helpful discussions at an early stage, and seminar participants at Leeds and PSE, and the Social Policy Reading Group for helpful comments.

1. Introduction

There are many studies of measurement error in household survey data on labour earnings which assume that linked administrative data provide a benchmark earnings measure that is error-free. A few, more recent, studies allow for measurement error in the administrative data as well. Our paper belongs to this second generation of research. We contribute new UK evidence about measurement errors in employment earnings to a field dominated by findings about the USA. We develop and fit new econometric models that allow for various types of measurement error in both administrative and survey data while also allowing error distributions to vary with observed characteristics. Our findings have relevance for improving data quality and assessments of earnings inequality levels.

1.1 Why learning about measurement error is important

Information about the quality of survey data on employment earnings is important for several reasons. First, there is substantial interest in earnings inequality levels, trends, and cross-national differences. Notwithstanding relatively rare studies using administrative data (e.g. Kopczuk et al. 2010), household survey data are the source for most studies of earnings inequality levels and trends. Two classic papers among many on US inequality trends using Current Population Survey (CPS) data are Levy and Murnane (1992) and Autor et al. (2008). Cross-national comparisons of earnings inequality are almost invariably undertaken using household survey data, illustrated by Gottschalk and Smeeding's (1997) review and the many subsequent articles based on Luxembourg Income Study data. Ascertaining the true level of earnings inequality is problematic when earnings are affected by measurement error. Observed earnings inequality is an upwardly biased estimate of true earnings inequality if measurement errors are classical. However, if high earners tend to under-report and low earners tend to over-report (errors are mean-reverting), measured inequality may be less than true inequality because mean reversion acts to offset the effect of classical measurement error (Gottschalk and Huynh 2010). We find no mean reversion in our UK data on average, and estimate how much lower 'true' earnings inequality is by comparison with observed inequality.

Second, and related, the accuracy of employment earnings measures has an impact on the accuracy of measures of household income. Household income inequality and poverty levels and trends are key social indicators. In financial year 2018/19, wage and salary income made up around 60% of total gross household income for the median UK household rising to around 80% at the median among households containing a working-age adult (Department for Work and Pensions 2020, Tables 9.3, 9.4). See Blundell et al. (2018) for a detailed analysis of the importance of labour earnings and other factors for levels and trends in household income inequality in the UK and the USA.

Third, information about the quality of survey measures of earnings is important for the design of data collection. If administrative data are much more reliable than survey data, there are pay-offs to introducing methods that allow survey responses to be substituted by linked administrative data responses, not only for survey quality but also because respondent burden is reduced when whole question blocks can be skipped. For example, the Canadian Survey of Labour and Income Dynamics (SLID) offered respondents the choice of reporting earnings and other income data to the interviewer or, instead, providing consent for administrative record data about these sources to be linked to the respondent's survey responses (Michaud et al. 1995). We show that our survey data on earnings are more reliable than are the linked administrative data, a finding that should give pause to data substitution strategies on data quality grounds.

Fourth, and related, instead of data substitution, statistical agencies might combine the information about earnings in the two data sources and produce hybrid measures of 'true' earnings that can be publicly released without the usual confidentiality concerns associated with administrative data (Abowd and Stinson 2013, Meijer, Rohwedder, and Wansbeek 2012). Implementing this strategy is conditional on having estimates of the model relating true earnings to observed earnings and measurement errors. Meijer et al. (2012) provide general methods for optimal prediction for mixture factor models and illustrate them using Kapteyn and Ypma's (2007) models. We apply Meijer et al.'s methods to our more general models, derive hybrid earnings measures from our linked data using model estimates, and illustrate how hybrid earnings measures might be used.

Fifth, household survey data are a fundamental ingredient of much contemporary empirical research because, by contrast with administrative data, they contain detailed information about the characteristics of individuals and their households in addition to their earnings. Labour economists have been particularly interested in the bias in estimates of regression model parameters that are introduced when error-ridden earnings measures are used as either a dependent variable or as an explanatory variable (cf. Bound et al. 2001, Kapteyn and Ypma 2007). We provide illustrations of this issue using our model and hybrid earnings predictors.

1.2 Earlier research: first- and second-generation studies

The extensive first-generation literature on measurement error in labour earnings assumes that administrative data measures are error-free. It finds substantial survey measurement error and that survey errors are mean-reverting. Bound and Kreuger (1991) and Bollinger (1998) compared labour earnings measures from the March Supplement to the CPS with earnings records held by the Social Security Administration (SSA). Duncan and Hill (1985), Bound et al. (1994), and Pischke (1995) used the Panel Study of Income Dynamics (PSID) Validation Study in which earnings measures derived using the PSID questionnaire for workers at a manufacturing establishment were linked with payroll records on earnings. Gottschalk and Huynh (2010) and Kim and Tamborini (2014) compared earnings responses in the Survey of Income and Program Participation (SIPP) with linked SSA ('DER') earnings records. Gottschalk and Huynh report that 'the SIPP yields lower estimates of the variance of log earnings than the DER, even within demographic cells. … [M]ean reversion more than offsets the additional variance of measurement error in reported earnings' (2010: 307). Outside the USA, there are studies for Austria by Angel et al. (2019) and for Denmark by Kristensen and Westergaard-Nielsen (2007), both of which find mean-reverting errors in their survey earnings measures.

The few second-generation measurement error studies are distinguished by not assuming that administrative record data on earnings represent the truth. Errors may arise because of mismatch when linking survey respondents or in the compilation of the administrative data. Kapteyn and Ypma (2007), using a cross-sectional Swedish linked dataset for a sample of individuals aged 50+, were the first to incorporate administrative data error, focusing on linkage mismatch. They found no evidence for mean reversion in survey measurement error, a sharp contrast with first-generation study findings. Jenkins and Rios-Avila (2020) fit Kapteyn and Ypma's model with linkage mismatch to UK linked data for a sample covering the full age range (the same data as used in this paper), and also find no evidence for mean reversion in survey measurement error.

Three further papers, each using longitudinal data, have accounted for administrative data error in earnings in different ways. Abowd and Stinson (2013) use SSA DER administrative data linked to multiple SIPP panels at the person-job spell level. They allowed for measurement error in both the administrative and survey data, fitting linear mixed models in which there is a common cross-source error variance as well as separate error variances for DER and SIPP earnings, and first-order autocorrelation in each of the three errors. DER variances and autocorrelations are generally larger than their SIPP counterparts (Abowd and

Stinson 2013, Table 5).[1] Abowd and Stinson do not model linkage mismatch. They also do not allow for mean reversion – usually specified as a correlation between true earnings and measurement error (see below) – because there is no concept of true earnings in their model.

True earnings concepts are used in the other two studies. Using linked data for New Zealand, Hyslop and Townsend (2020) fit models of earnings dynamics with persistent and transitory components, finding that 'survey errors are mean-reverting when administrative reports are assumed correct, but not when this assumption is relaxed' (2020, 457). They report error variances that are greater for the survey data than the administrative data (2020, Table 6), but they do not model mismatch error explicitly. Bollinger et al. (2018) model not only observed earnings and errors among respondents but also earnings non-response, exploiting the two-year panel structure of the CPS. Their models allow for measurement error in the survey and administrative data (SSA DER earnings) as well as linkage mismatch. Bollinger et al. report that there is at most weak evidence for mean reversion in survey measurement error (what they label the 'common man' hypothesis), and they estimate the probability of mismatch to be around 3% (2018, Table 15).

1.3 What this paper contributes

This paper provides new evidence for the UK taking careful account of UK-specific data features using new models of earnings and errors, with more extensive post-estimation analysis than previous research.

We fit our models to the FRS-P14 Linked Dataset, a new source for the UK, which we discuss in Section 2. Most earnings measurement error analysis has been conducted using US data and it is of interest to know whether these findings are special or generic. The existence or absence of mean-reverting errors is one example. Differences in data sources across countries inevitably mean there will be some differences in the relative importance of different types of error and these may confound cross-national comparisons of earnings distributions. Relatedly, we pay particular attention to a potential confounding factor that is UK-specific. Our administrative data contain a measure of annual earnings. Our survey data provide a 'current' earnings measure referring to jobs in progress at the time of the interview and we derive a measure of annual earnings ('annualised' earnings) by combining information about an earnings amount and the reference period to which it refers. The

[1] Abowd and Stinson (2013) provide point estimates but no standard errors for their variance components, a consequence of their Restricted Maximum Likelihood (REML) estimation method. We provide point estimates and SEs for all variance components.

difference between annualised survey and true annual earnings is what we call reference period error.

In Section 3, we review how differences between survey and administrative data earnings measures may arise, and this discussion motivates our econometric model specifications that follow. We generalize the models of Kapteyn and Ypma (2007) to allow for (i) measurement error in the administrative data and (ii) measurement error distributions differing across individuals with different observed characteristics.[2] We propose mixture factor models with up to nine latent classes characterized by combinations of error-ridden or error-free administrative and survey earnings measures.

Incorporating administrative data error means that our study has all the key features of a second-generation study. By contrast with other measurement error studies which relate the mean of true or observed earnings to observed characteristics, we fit models in which all mixture distribution and mean-reversion parameters potentially vary with observed characteristics. This provides us with a succinct but informative way to examine how error distributions vary across types of respondent, job, and employer. We have not seen this approach used before. Section 4 presents and discusses estimates of four models without covariates, and Section 5 extends the analysis to models with covariates.

We find that measurement errors are pervasive but quite different in nature for the two data sources. We estimate the probability of survey error occurrence to be around 93% and the probability of administrative data error around 47%. However, the standard deviation (SD) of survey data errors is markedly smaller than the SD of administrative data errors. In addition, there is mismatch error with an estimated probability of around 6%.

Reference period error has a low prevalence (around 10%) but, where it occurs, annualized survey earnings under-estimate true annual earnings, and under-estimation is associated with being aged 60+, working part-time, and working in an unstable job. We find negligible mean reversion in survey measurement errors. There is also little mean reversion in administrative data on average, though we find a positive correlation between error and true earnings among some groups such as private sector employees (which we argue reflects reference period error not otherwise accounted for).

The models with heterogeneous error distributions point to ways to improve data quality. For example, regarding survey quality, our estimates highlight how respondents

[2] Our earlier note (Jenkins and Rios-Avila 2020) used the same dataset as this paper, but did not consider the more general models considered in this paper (i.e., with administrative data errors and heterogeneous error distributions).

showing payslips to interviewers have a survey error SD around half the size of the SD for respondents who do not. For administrative data, our results point to areas in which to target quality improvement drives: greater error variances are associated with non-standard jobs, private sector jobs, and employers without good payroll systems.

In Section 6, we present post-estimation analysis. First, we show the combinations of survey and administrative earnings values that are associated with the memberships of the different latent classes. For example, although linkage mismatch has a low estimated probability, it has a large adverse impact on linked administrative data reliability. Second, we extend the analysis of Meijer et al. (2012) to our more complex mixture factor models and derive 'hybrid' earnings predictors that combine the information from both survey and administrative data sources, showing that they have very high reliability. Third, we show the extent to which observed earnings inequality over-estimates the inequality of model-based true earnings and our benchmark hybrid earnings predictor – by around 11% according to the standard deviation of log earnings inequality index. Fourth, we illustrate the extent of bias arising in regression in which the error-ridden observed earnings measures are used as the dependent variable or as an explanatory variable.

Section 7 contains a summary and conclusions. Supplementary materials are contained in Appendices A–E.

2. The Linked FRS-P14 Dataset

We use the UK Linked FRS-P14 Dataset on employment earnings, created by linking records for respondents to the Family Resources Survey (FRS) for financial/tax year 2011/12 to P14 administrative record data for the same year held by the UK tax authorities, Her Majesty's Revenue and Customs (HMRC).

2.1 Administrative record data on earnings (P14)

The P14 label arises because the dataset is compiled from P14 forms – employers' end-of-tax-year returns to HMRC about wages and salaries paid to employees and taxes and National Insurance contributions withheld. Thus, the UK's P14 forms are similar to the W-2 forms returned by US employers to the SSA.

Our administrative measure of earnings for each linked respondent i, r_i, is the logarithm of total gross earnings per year (the sum across all earnings spells recorded in

2011/12).[3]

2.2 Survey data on earnings (FRS)

The FRS is the UK's main income survey with an annual sample of around 20,000 private households. It is a continuous household survey with monthly subsamples combined into financial year samples: our data include responses from interviews undertaken in the 12 months between April 2011 and March 2012. The FRS attempts a face-to-face computer-assisted personal interview with every individual aged 16+ years in a household. See DWP (2013) for the documentation of the 2011/12 FRS.

The FRS is the source for the DWP's annual income distribution report focusing on low-income prevalence, *Households Below Average Income* (*HBAI*, DWP 2020), and other leading UK income distribution series such as those published by the Institute for Fiscal Studies (Bourquin et al. 2020). Thus, the FRS and *HBAI* play the same role in the UK as the Annual Social and Economic Supplement to the CPS (CPS/ASEC) and the Census Bureau's annual P-60 reports and associated data series do for the USA.

FRS questions about gross earnings (i.e. earnings prior to deductions) refer to jobs in progress at the interview date – hence the 'current' earnings label. The interviewer first asks each employed respondent 'What was the gross wage/salary – i.e. the total, before any deductions?'. A follow-up question asks about the reference period to which that amount refers. Around 70% of our sample (discussed below) report '1 calendar month'. The next most prevalent report is '1 week' (17%), then '4 weeks' (7%), '1 year / 52 weeks / 12 months' (4%), and '2 weeks' and 'other' (each 2%). Three other response options receive few responses.

The FRS data producers convert the gross earnings responses for each job to weekly amounts pro rata – the originally-reported amounts are not released – which we converted to annual amounts (pounds per year). Earnings amounts are not top-coded.

FRS interviews are undertaken throughout the financial year and so respondents' earnings reference periods do not refer to specific calendar dates that are common to all. By contrast, in the CPS/ASEC, respondents provide information about earnings over the previous calendar year and this is also the reference period for SSA administrative data.

Hence, for the UK, there is an issue of non-comparability between the annualised and

[3] P14 earnings spells cannot be linked to the jobs in the survey data and no individual characteristics besides sex are recorded.

genuinely annual measures that linked data analysis must address and we do this using a model-based approach that is explained in the next section.[4] Kapteyn and Ypma (2007: 538) briefly mention reference period recall error as a source of 'contamination' and they model this in addition to mean-reverting survey measurement error. We bring this component more to the fore, renaming it 'reference period' error. Moreover, in another new departure, we allow reference period errors to be correlated with true earnings.

In sum, our survey measure of earnings for each respondent i, s_i, is the logarithm of total gross earnings (the annualized sum across all jobs reported). Fewer than 5% of our sample report earnings for more than one job, and our preliminary analysis indicated that measurement error distribution parameters did not differ between single-job and multiple-job holders. Hence, we focus on the annual earnings total and do not examine multiple job holding further (except as a correlate of mean true earnings – see Section 4).

2.3 Linked FRS-P14 data

DWP statisticians extracted P14 records for 2011/12 FRS respondents who consented to record linkage: at the end of each FRS individual interview, the respondent is asked if it would be 'okay to pass their name and address to the Department for Work and Pensions'.[5] The statisticians linked records deterministically using match keys constructed from information about first name, last name, postcode, sex, and date of birth.[6] The DWP passed us two datasets. One contained FRS personal identifiers and P14 earnings data and the second

---

[4] In Kristensen and Westergaard's (2007) Danish study, the reference period is one year for both administrative and survey data, referring to the year from November in the former and the year prior to the survey in the latter. However, the authors argue that the date of interview differences should not be over-emphasised, and do not consider them further. The UK's current earnings measures bear some similarities to the earnings measures collected in the CPS Outgoing Rotation Group (ORG) interviews. ORG earnings measures refer to jobs held in the week prior to the interview, and gross hourly pay information is collected for hourly-paid employees and weekly pay information for all other employees. We are unaware of any study that has compared ORG earnings measures with linked administrative earnings data to study measurement error. Hirsch and Schumacher (2004) analyse the 'match bias' that arise because of CPS imputation procedures for missing ORG earnings responses. Böheim and Jenkins (2006) compare current and annual measures in the British Household Panel Survey, finding similar estimates of many distributional summary statistics (inequality, poverty, etc.) However, their analysis was about household income, not individual earnings from employment, and no administrative data were linked in.

[5] This follows a preamble given by the interviewer that states that the Department 'would like to add the records they already hold on your benefits, tax and employment to your answers to this survey. Adding everyone's records in this way will help us with further research to get a more accurate picture of people's living standards'. The consent question is not put to the small number of partially-responding adults for whom some limited proxy information is collected.

[6] The DWP does not use National Insurance numbers (NINOs), the UK analogue to US Social Security numbers, for linking because respondent-reported NINOs are unreliable (Jenkins et al. 2008).

dataset (multiple files) was a secure-data version of the FRS.[7] We merged the P14 and secure-data FRS datasets using FRS personal identifiers as the match key.

In the 2011/12 secure-data FRS, there are 13,851 employed respondents with at least one job, of whom 9,014 men and women (65%) gave their consent to data linkage. We successfully linked FRS and P14 records for 6,432 men and women (71% of the employees consenting to data linkage). Post-linkage, we dropped a small number of observations who declared themselves in the FRS to be 'self-employed' (N = 23) and then also observations for whom either FRS or P14 earnings were equal to zero (N = 18), giving us 6,391 employees (2,794 men; 3,599 women). Finally, we followed common practice[8] and dropped observations with imputed or otherwise edited values for the gross earnings or reference period for any FRS job reported (N = 420), leaving our principal estimation sample of 5,971 individuals (2,595 men; 3,376 women). The age range of respondents in this sample is 16–84 years, with the vast majority (84%) aged 25–59 years, 6% aged 16–24 years, and 10% aged 60+ years.

We also undertook analysis using a subsample of 3,564 individuals aged 25–59 years who declared themselves to be in full-time work and not participating in any form of education.[9] This sample yields similar estimates to those based on our main sample (we cite relevant Appendices later). For brevity, we focus on estimates from the main sample. In any case, there are good reasons for focusing on the main sample's estimates. First, with the main sample we can directly compare measurement error distributions for part-time workers and older workers with working-age full-timers (Section 4). Second, it is the earnings of all individuals that are relevant to household income inequality (there may be multiple workers per household; part-time work contributes to household income; and there are earnings from individuals outside the standard 'working age' age range).

The representativeness of our estimation samples is a potential issue. Although the FRS provides nationally representative estimates (when the sample weights are used), consent to data linkage and record linkage success among consenters may be selective

[7] The secure-data files do not contain the various confidentiality adjustments used for the public-use FRS file (e.g. respondent age is in years rather than banded), and they also include a file allowing us to identify which earnings records had been edited or imputed. The secure-data and public-use FRS cannot be linked because they use different identifier variables.
[8] Cf. Bound and Krueger (1991), Bollinger (1998), and Kim and Tamborini (2014). Gottschalk and Huynh (2010) and Abowd and Stinson (2013) report results based on samples excluding and including imputed observations. In preliminary analysis, we re-ran our regressions including imputed or otherwise edited observations, and estimates changed hardly at all.
[9] Whether an FRS respondent is working full- versus part-time work status is decided by the respondent. There is no official UK survey definition of full-time work.

processes. Following Bollinger et al. (2019, Appendix A.4), we investigated this issue by constructing inverse-probability weights. We regressed the probability of the binary outcome ‘consented to data linkage and successful linkage’ on a large number of individual characteristics using a probit model applied to the FRS sample of employed respondents, and derived weights equal to the inverse of the predicted probabilities. We then multiplied these weights by the FRS individual sample weight to create a new composite weight. We find that unweighted and composite-weighted estimates of corresponding measurement error models are very similar (see Appendices A–E). Hence, for brevity, we report only unweighted estimates in the main text.

2.4 Linked FRS-P14 data: distributions of earnings and earnings differences

We now describe the distributions of FRS and P14 earnings and the individual-level differences between them. For brevity we refer to earnings rather than log earnings.

Figure 1 shows that the distributions of FRS earnings (s) and P14 earnings (r) are quite similar. Each has greater concentration around the mean than a normal distribution with the same mean and standard deviation and is slightly asymmetric. P14 earnings have a slightly lower mean than FRS earnings, 9.75 compared to 9.77, and greater inequality (SD), 0.842 compared to 0.813. The greater inequality is inconsistent with a model in which P14 earnings represent the truth and FRS earnings contain only classical measurement error (Kapteyn and Ypma 2007, 524).

<Figure 1 near here>

Figure 2 shows the distribution of the differences between FRS and P14 earnings, $d_i = s_i – r_i$. There is a large spike at zero, with most differences tightly clustered around this value. There are close similarities with the corresponding graphs shown by Bound and Krueger (1991, Figures 2 and 3, CPS data) and Kim and Taborian (2014, Figure 1B, SIPP data) for the USA, Hyslop and Townsend (2020, Figure 1C) for New Zealand, and Kapteyn and Ypma (2007, Figure 3) for Sweden.

<Figure 2 near here>

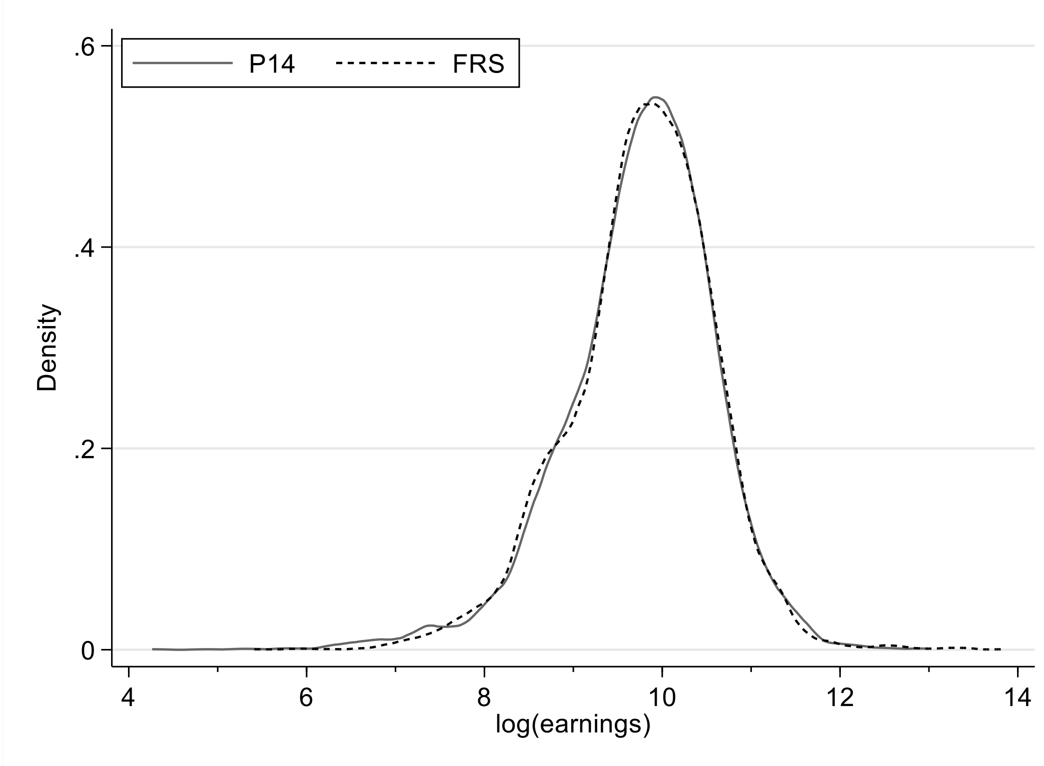
Figure 3 shows a scatterplot of FRS earnings against P14 earnings with a linear regression line superimposed. Were the basic classical measurement error model to apply, the regression line would have a slope coefficient of one; a slope of less than one is indicative of mean-reverting error. In Figure 3, the slope is 0.793 (SE 0.007) and significantly less than 1.

<Figure 3 near here>

Separate regressions by sex show the slope of the regression line is 0.693 (SE 0.013) for men and 0.830 (SE 0.010) for women. Also, separate regressions by age yield a slope coefficient of 0.742 (SE 0.031) for workers aged 60+ years and of 0.794 (SE 0.007) for workers aged less than 60 years. These estimated slopes are smaller than Bollinger (1998) reported for US workers in the late 1970s: 0.91 for men, 0.97 for women.[10]

Conclusions about mean reversion are contingent on the assumed model. We investigate mean reversion further, and differences in it by age and sex, using our non-classical measurement error models.

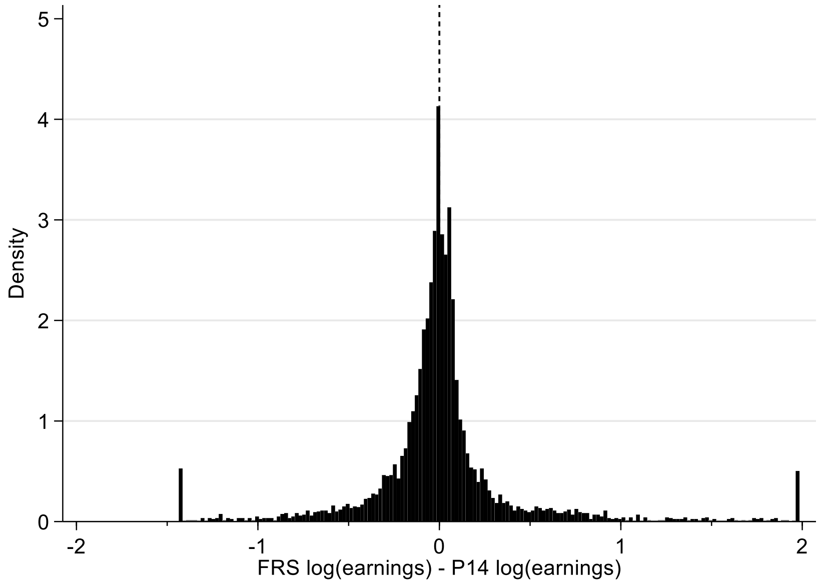
Figure 1. Distributions of FRS and P14 log(earnings)



Notes. Kernel density estimates (Epanenikov kernel, 'optimal' bandwidth). Summary statistics for (s, r): mean (9.77, 9.75); p5 (8.37, 8.32); p10 (8.69, 8.69); p50 (9.83, 9.83); p90 (10.71, 10.70); p95 (10.95, 10.96); standard deviation (0.81, 0.84). Sample N = 5, 971.

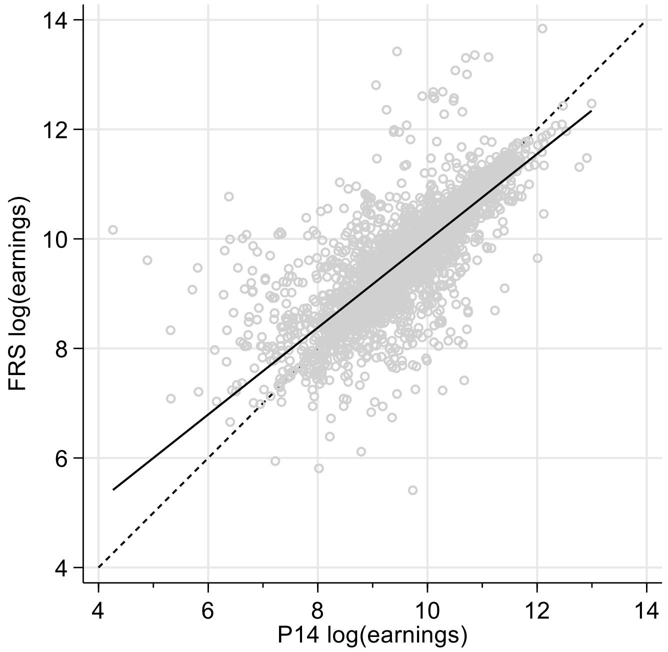
[10] Using non-parametric regression, Bollinger (1998) also found that deviations from linearity (with slope 1) were most prevalent for low earning men. When we applied the same methods, we found less clear evidence of non-linearity. It was apparent only in the lowest few percent of the distribution, and slopes were very imprecisely estimated.

Figure 2. Distribution of difference between FRS and P14 earnings (*s* – *r*)



Notes. Histogram with bin width = 0.02. Earnings differences are bottom-coded at *p*1 (–1.44) and top-coded at *p*99 (1.97) for purposes of presentation. Summary statistics for *s* – *r* (without bottom- or top-coding): mean, 0.016; standard deviation, 0.496; *p*5, –0.579; *p*10, –0.315; *p*50, –0.005; *p*90, 0.331; *p*95, 0.714. Sample *N* = 5,971.

Figure 3. The relationship between FRS and P14 earnings



Notes. Scatterplot shows all (r_i, s_i) observations. A linear regression of FRS earnings on P14 earnings has slope coefficient 0.793 (SE 0.003), shown by the solid line. Sample *N* = 5, 971.

3. Reconciling survey and administrative data reports: econometric models and their motivation

Differences between earnings measures from the two data sources arise because of measurement errors and because of differences in definition. It is the former that have received most attention to date; the latter are also important for our UK data. In this section, we review the main types of cross-source differences to motivate the econometric models that we then introduce.

3.1 How differences between survey and administrative data earnings measures may arise

Survey measurement error arise for multiple reasons. Moore et al. (2000) point out that 'deliberate prevarication – in particular, deliberate underreporting – is probably the most commonly-assumed cause of income survey response errors. However, close consideration reveals an abundance of areas of potential difficulty without invoking motivated misreporting at all' (2000, 349). They go on to review various cognitive factors: respondent misunderstanding of the concept asked about, faulty retrieval by respondents because of faulty recall or low salience of some items, and various types of sensitivities to questions about money. See also Bound et al. (2001, section 5). To illustrate the points about the potential for misunderstanding, faulty recall, or low salience, note that the FRS gross earnings measure is intended to be more comprehensive than referring to regular wage and salary earnings alone. FRS interviewer instructions refer to gross wage/salary as including overtime, bonuses, commission, tips, or other payments. Errors may also enter survey responses through interviewer key entry errors (e.g. mis-keying numbers and subsequent data processing), notwithstanding the FRS's long-standing use of experienced interviewers and computer-assisted interview scripts that incorporate checks.

Moore et al.'s (2000, 342–345) review indicates there is significant variance in survey measurement errors for earnings but no clear evidence about bias (summarized by the mean of the measurement error distribution). Our models provide new evidence about both aspects. Moore et al. (2000, 353–4) also discuss US studies about how record use may reduce differences between survey and administrative reports. Using our models, we quantify the extent to which interviewer-recorded consultation of a payslip during the FRS interview is associated with a reduction in measurement errors of various kinds.

Question sensitivity is closely related to issues of social desirability bias. Bound et al.

write that '[i]t is widely believed and well documented that … questions [about socially and personally sensitive topics] elicit patterns of underreporting (for socially undesirable behavior and attitudes) as well as overreporting (for socially desirable behaviors and attitudes)' (2001, 3746). This is the standard explanation for mean reversion in survey measurement error. As Angel et al. put it, '[r]espondents at the lower tail of the wage distribution overreport as they feel ashamed of their actual economic conditions, whereas respondents at the upper tail of the distribution underreport since they do not want to disclose their high wages to an (unknown) interviewer' (2019, 1414). Our models allow for mean reversion in survey measurement error.

Administrative data may also contain errors of various kinds. (See Abowd and Stinson 2013 for a review.) Kapteyn and Ypma (2007) focus on mismatch error, the situation arising when a survey respondent is linked to the wrong individual in the administrative data, in which case, the individual's linked administrative data earnings measure is a draw from the complete P14 earnings distribution. Kapteyn and Ypma (2007) and Meijer et al. (2012) show that even a small mismatch fraction has serious consequences for reliability. Our models allow for mismatch error alongside measurement error in administrative data per se.

Measurement errors in P14 data may arise through employer mistakes, for example by mis-keying of entries into the payroll software that generates the year-end P14 return to HMRC. We would expect large mistakes to be noticed and fixed because HMRC has procedures by which employers can submit corrected returns, but smaller mistakes may be over-looked or simply ignored. Such mistakes may also be more prevalent in small businesses without suitable payroll software, and more likely to occur with the pay of workers who are not on full-time or permanent contracts if records for these staff are of poorer quality. P14 errors may also arise if a respondent has earnings for a second or third job which are reported in the FRS but not recorded in a P14 return. (There has to be a P14 record for at least one job because, otherwise, the individual would not appear in the FRS-P14 Linked file.) This could arise, for example, if the employee were paid informally, 'under the counter', for those additional jobs.[11] Overall, these factors lead to positive error variance but

[11] The DWP informed us that 'some employers didn't submit records for people [for whom] were there were no NI [National Insurance] or tax liabilities – but these were largely people working for small employers who did not operate electronic payroll. We suspect these people would not appear in PAYE [Pay As You Earn withholding] at all and would effectively be 'cash in hand' employees. In reality we don't believe this was a massive issue and that most employment records were captured.' (DWP FRS Team, email 2020-02-25). This statement is not inconsistent with the possibility that some earnings may not have been recorded in P14 files. Another potential issue is that NI contributions do not have to be paid if a worker earns less than the Lower Earnings Limit, which was £102 per week (£5,304 per year) in 2011/12. However, we understand that P14 records are typically submitted for such workers even if there is no liability. Moreover, there are no

the nature of bias will depend on whether mistakes average out. Our models provide evidence about the prevalence of administrative error and its bias and dispersion.

Abowd and Stinson state that processing error in survey and administrative data is 'very different from the [process] typically postulated for self-reported data as it is unrelated to the actual amount or the person reporting. More research is needed to determine the extent of this error and quantify its specific impact' (2013, 1461). We investigate these issues, specifically the prevalence, mean and variance of administrative error relative to survey measurement error. Also, the quality of an employer's payroll processes and monitoring of it may differ by type of employer and hence also by type of employee and earnings level. We investigate whether there is prima facie evidence for this (see section 4). Although there is no administrative data equivalent to social desirability bias – suggesting mean reversion in administrative measurement error is unlikely – our models allow us to check this hypothesis.

Definitional differences between the FRS and P14 measures are important. We referred to reference period differences earlier, explaining that the P14 measure refers to annual earnings for financial year 2011/12, whereas the FRS measure is an annualised measure derived from the current earnings of jobs in progress at the time of the FRS interview.

If a respondent stays in the same job(s) throughout the tax year or longer, receiving the same pay, no reconciliation of reports is required on reference period grounds: the survey's annualized current earnings measure equals the administrative data's annual measure by construction.[12] However, the annualized FRS measure may be greater than the P14 annual earnings measure if there are spells of unemployment or of lower pay either before or after the reported reference period. That is, the interview response captures earnings in good times but misses the shortfall of earnings in bad times, but the P14 records both, in effect averaging them. Conversely, the annualized FRS earnings measure may be smaller than its P14 counterpart if the respondent has a higher-paid job outside the reported reference period whether through job change or promotion, or if end-of-year bonuses are not reported in current earnings. It is unclear ex ante whether survey reference period error – the difference between survey earnings and true earnings due to reference period differences – leads to under- or over-estimation of annual earnings on average. In our models, the former

discontinuities in the distribution of FRS earnings around the Lower Earnings Limit (log earnings of around 8.58): see Figure 1. This suggests that any non-capture by P14 records of survey-reported earnings for second and third jobs is for reasons other than NI contribution ones.

[12] Subject to the complicating caveat that the 'year' round the survey interview date corresponds to the tax year for relatively few respondents.

(latter) case corresponds to a negative (positive) mean of the reference period error distribution.

In addition, we expect higher-paying and full-time jobs to be more stable than lower-paying and part-time jobs (see e.g. Golden 2016), and also that shorter pay reference periods are more prevalent among workers with lower-paying jobs.[13] Hence, reference period error may be negatively correlated with true earnings. Our models incorporate this possibility, and we also investigate how the distribution of reference period error varies with job stability. We also discuss (section 5) how some features of reference period error may induce a correlation between administrative data error and true earnings.

Another potential definitional difference is that the FRS collects earnings information for three jobs at most. Additional earnings from a fourth or other employment would be missed by the survey but may be captured in the worker's P14 earnings record compiled from all employments. This is likely to be of negligible importance because the prevalence of multiple job holding is very low (see earlier).

In sum, to reconcile the earnings reports contained in FRS and P14 data, we need to take account of measurement errors in both sources, linkage mismatch error, and reference period error. In addition, our discussion has indicated that survey measurement errors are likely to vary with respondent characteristics whereas P14 measurement errors are likely to vary with the characteristics of respondents' jobs and their employers. We now present econometric models incorporating these various features.

3.2 Econometric models

We propose mixture factor models that generalize Kapteyn and Ypma's (2007) models.

The key idea is that true annual earnings for an individual i, ξ_i, are unobserved but there are two observed earnings measures available: s_i from the FRS and r_i from the P14 data. Each measure is subject to error for the reasons discussed earlier, though not all individuals experience all types of error. For some, their FRS earnings measure is error-ridden and their P14 measure is not; for others it is vice versa; or both earnings measures are error-ridden. We can classify individuals into groups (latent classes) according to which types of error their earnings measures contain. Observed earnings are a combination ('mixture') of the distributions for the latent classes. We can identify the various error components by having

[13] In our FRS data, and on average, annualized earnings for workers reporting earnings using an annual reference period are greater than those for workers reporting using a monthly reference period, and these are in turn greater than for those reporting using a weekly reference period.

repeated measures on earnings and making assumptions about the structure of the measurement errors.

More specifically, we assume that the distribution of P14 earnings is a mixture of the distributions for three types of observation, as set out in eq. (1). We distinguish between individuals who are correctly linked with an FRS respondent with probability π_r and individuals who are incorrectly linked with probability $1-\pi_r$. Among the correctly linked observations, P14 earnings are either equal to true earnings, ξ_i, with probability π_υ (type $R1$), or are measured with error with probability $1-\pi_\upsilon$ (type $R2$). For each type $R2$ observation, we suppose that P14 earnings contain measurement error and this may be correlated with true earnings with the correlation summarized by parameter ρ_r. (There is mean reversion if $\rho_r < 0$; mean affirmation if $\rho_r > 0$.) In the third case (type $R3$), linkage mismatch, P14 earnings represent the earnings not of the FRS respondent as intended but of someone else in the P14 dataset, ζ_i.

$$r_i = \begin{cases} \xi_i \quad \text{with probability } \pi_r\pi_\upsilon & \text{(type } R1\text{)} \\ \xi_i + \rho_r(\xi_i - \mu_\xi) + \upsilon_i \;\; \text{with probability } \pi_r(1-\pi_\upsilon) & \text{(type } R2\text{)} \\ \zeta_i \quad \text{with probability } (1-\pi_r) & \text{(type } R3\text{)} \end{cases} \qquad (1)$$

FRS observations are also a mixture of three types, as summarized by eq. (2). In the first case (type $S1$), s_i equals true earnings with probability π_s. In the second case (type $S2$), s_i contains mean-reverting error with probability $(1-\pi_s)(1-\pi_\omega)$, with ρ_s summarizing the correlation between error and true earnings. Third, there are observations subject to reference period error (ω_i) in addition to survey measurement error (type $S3$), with probability $(1-\pi_s)\pi_\omega$.

$$s_i = \begin{cases} \xi_i \quad \text{with probability } \pi_s & \text{(type } S1\text{)} \\ \xi_i + \rho_s(\xi_i - \mu_\xi) + \eta_i \;\; \text{with probability } (1-\pi_s)(1-\pi_\omega) & \text{(type } S2\text{)} \\ \xi_i + \rho_s(\xi_i - \mu_\xi) + \eta_i + \omega_i \;\; \text{with probability } (1-\pi_s)\pi_\omega. & \text{(type } S3\text{)} \end{cases} \qquad (2)$$

In sum, observations in the linked dataset are a mixture of nine types depending on the combination of FRS and P14 observation types. For example, group 1 contains observations with the combination ($R1$, $S1$), group 2 contains observations with the combination ($R1$, $S2$), etc. Table 1 summarises the groups and their probabilities.

<Table 1 near here>

Table 1. Groups (latent classes) in mixture factor model of FRS and P14 earnings

Group, j	Description	Types	Probability, $\pi_j = \ldots$
1	No error in P14 or in FRS earnings	$R1$, $S1$	$\pi_r \pi_\upsilon \pi_s$
2	No error in P14 earnings; error in FRS earnings	$R1$, $S2$	$\pi_r \pi_\upsilon (1-\pi_s)(1-\pi_\omega)$
3	No error in P14 earnings; error and reference period error in FRS earnings	$R1$, $S3$	$\pi_r \pi_\upsilon (1-\pi_s)\pi_\omega$
4	Error in P14 earnings; no error in FRS earnings	$R2$, $S1$	$\pi_r (1-\pi_\upsilon)\pi_s$
5	Error in P14 earnings; measurement error in FRS earnings	$R2$, $S2$	$\pi_r (1-\pi_\upsilon)(1-\pi_s)(1-\pi_\omega)$
6	Error in P14 earnings; measurement error and reference period error in FRS earnings	$R2$, $S3$	$\pi_r (1-\pi_\upsilon)(1-\pi_s)\pi_\omega$
7	Mismatched P14 earnings; no error in FRS earnings	$R3$, $S1$	$(1-\pi_r)\pi_s$
8	Mismatched P14 earnings; measurement error in FRS earnings	$R3$, $S2$	$(1-\pi_r)(1-\pi_s)(1-\pi_\omega)$
9	Mismatched P14 earnings; measurement error and reference period error in FRS earnings	$R3$, $S3$	$(1-\pi_r)(1-\pi_s)\pi_\omega$

Notes. π_s: probability FRS survey data are error-free. π_ω: probability of survey reference period error. $1-\pi_r$: probability of linkage mismatch. $1-\pi_\upsilon$: probability P14 administrative data contain measurement error.

We assume that true earnings (ξ_i), mismatched earnings (ζ_i), and errors (υ_i, η_i, ω_i) are each normally distributed with the exception that true earnings and reference period errors (ω_i) are bivariate normal. The distributions are identically distributed and mutually independent (assumptions we relax shortly). Thus, the distributions of the 'factors' may be written as:

$$\begin{pmatrix} \xi_i \\ \omega_i \end{pmatrix} = BVN\left(\begin{pmatrix} \mu_\xi \\ \mu_\omega \end{pmatrix}, \begin{pmatrix} \sigma_\xi^2 & \rho_{\xi\omega}\sigma_\xi\sigma_\omega \\ \rho_{\xi\omega}\sigma_\xi\sigma_\omega & \sigma_\omega^2 \end{pmatrix} \right), \quad (3)$$

$$\zeta_i \sim N(\mu_\zeta,\ \sigma_\zeta^2),\ \eta_i \sim N(\mu_\eta,\ \sigma_\eta^2),\ \text{and}\ \upsilon_i \sim N(\mu_\upsilon,\ \sigma_\upsilon^2),$$

where 'μ' and 'σ' denote mean and SD, respectively, and $\rho_{\xi\omega}$ is the correlation between true earnings and reference period error. N(.) is the univariate normal distribution; BVN(.) is the bivariate normal distribution. We do not restrict means to equal zero for reasons discussed earlier.

We allow distributions to vary with observed characteristics by writing transformations of model parameters as linear indices of characteristics, i.e.,

$$G(\gamma_i) = \alpha_\gamma + \beta_\gamma X_i. \quad (4)$$

For each model parameter with generic label γ_i, α_γ is a constant, and X_i is a vector of observed characteristics for individual i. Transformation function G(.) is the identity function for means (μ), the logarithmic function for SDs (σ), the logistic function for probabilities (π), and Fisher's Z transformation for correlations (ρ).[14]

We assume normality to fit models by maximum likelihood and because it facilitates post-estimation derivations. The assumption is ubiquitous, employed for example, by Kapteyn and Ypma (2007), Abowd and Stinson (2013), and Bollinger et al. (2018). These assumptions do not constrain observed earnings measures to be normally distributed since s and r are mixtures of normal distributions. The more substantive but untestable assumption is that true earnings are lognormally distributed, though flexibility is gained by conditioning on characteristics.

How does our model compare with others? Kapteyn and Ypma's (2007) Full model is the special case of our general model in which there are no administrative data errors ($\pi_\upsilon = 1$), $\rho_{\xi\omega} = 0$, and so there are six latent classes rather than nine (types 4–6 in Table 1 are not present). Kapteyn and Ypma also consider a Full model in which mean true earnings (μ_ξ) depend on observed characteristics but they do not allow other model parameters to vary with characteristics; we do.

Our model shares features of Abowd and Stinson's (2013) model such as measurement error in both survey and administrative data and allowing mean earnings to vary with characteristics. Their model does not incorporate other parameter heterogeneity or mismatch error but, because their dataset provides more repeated measures than ours (across time and across jobs as well as across person), they can fit additional cross-source correlations in addition to our single individual random effect (our true earnings factor ξ_i corresponds to their 'common' earnings random effect, c). Our model also shares features of Bollinger et al.'s (2018) mixture factor model (see their eqq. 1–3). Both their models and our models incorporate mean-reverting survey measurement error and administrative mismatch. However, Bollinger et al. do not allow for parameter heterogeneity other than in the mean of true earnings. Instead, their focus is on modelling non-response per se, exploiting their longitudinal data. The classical measurement error model augmented with mean-reverting survey error assumes (i) there are no administrative data measurement errors or mismatch ($\pi_\upsilon = \pi_r = 1$), i.e. the administrative data represent the truth; (ii) the survey data contain no reference period error ($\pi_\omega = 0$); and (iii) there is a positive probability that survey earnings equal true earnings ($\pi_s > 0$).

[14] Reversion to the mean in the models with a heterogeneous mean earnings function refers to reversion to the mean among individuals with the same observed characteristics.

We focus below on four variants of our general model:

- the *Extended* model is the general model;
- the *Constrained Extended* model is the Extended model with constraint $\rho_{\xi\omega} = 0$;
- the *Full* model is Kapteyn and Ypma's (2007) Full model but also allowing $\rho_{\xi\omega} \neq 0$; and
- the *Constrained Full* model is Kapteyn and Ypma's (2007) Full model (with constraint $\rho_{\xi\omega} = 0$).[15]

Comparisons of estimates of the two Extended models with the two Full models highlight the effects of neglecting measurement errors in the administrative data.

3.3 Identification and estimation

Our mixture factor models are identified by the assumptions about the relationships between the two observed measures and true earnings and the non-normal error structure arising from the mixture of normal distributions (Kapteyn and Ypma, 2007, 532). See also Yakowitz and Spragins (1968).

The first latent class (group 1) plays an important role in identification. Group 1 contains the observations for whom survey earnings equal administrative earnings and thence also true earnings. Kapteyn and Ypma (2007) call these observations 'completely labelled', borrowing the term from Redner and Walker (1984). Parameters μ_ξ and σ_ξ^2 are identified by the completely labelled observations and, in turn, this contributes to identification of the other parameters. See Appendix A.

Hence, to fit our models, we must define when an observation's survey and administrative earnings measures are sufficiently close to count as 'equal'. Kapteyn and Ypma (2007) assumed that observations were completely labelled if survey and administrative earnings differed by less than SEK 1000 per year, which translates to a relatively large fraction of their sample, 14.8%. We are reluctant to use a completely labelled fraction that is so large because of the relevance of reference period error in the UK context. We assume that observations with $|r_i - s_i| < 0.005$ are completely labelled, which is 3.4% of our main sample. We have repeated analyses using a completely-labelled fraction more than twice as large, 7.7% (observations with $|r_i - s_i| < 0.010$), and we find that conclusions are

[15] We also estimated versions of the Constrained Full model with additional constraints on model parameters, i.e. all of the models considered by Kapteyn and Ypma (2007). For brevity, and because the Constrained Full model invariably fitted better, we do not report estimates from these additionally constrained models.

robust. See Appendices B–E.[16]

We fit our models by maximum likelihood. Given the assumptions made earlier, the sample likelihood function is a finite mixture of latent class distributions. Appendix A provides expressions for probability density functions and the log-likelihood function. We report cluster-robust standard errors using the FRS household as the cluster: the 5,971 individuals in our main sample live in 4,874 households. For further details of the estimation method and our Stata programs for estimation and post-estimation summary statistics and prediction, see Jenkins and Rios-Avila (2021*b*).

We report marginal means to facilitate interpretation: for a covariate X_k, we report sample average values of the parameter (in its natural metric) estimated at different levels for X_k while holding the values of other covariates in the parameter's equation (if present) at their sample values.

We checked that our maximization algorithms converged to global rather than local maxima. Our strategy was to fit successively more complex models using starting values from the less complex models. We also used relatively parsimonious covariate specifications to avoid problems. In the few cases where problems arose, we addressed them by using multiple sets of initial values to ensure we found a global maximum.

3.4 Post-estimation calculations: reliability and combined-data predictors of true earnings

It is of substantial interest to know how well survey and administrative data earnings measures relate to (unobserved) true earnings. We report two estimates of reliability for each model. The first, Reliability1, is analogous to the reliability statistic often reported for the classical measurement error model with mean-reversion (see Bound and Krueger 1991, 8–9). It is equal to the slope coefficient from a (hypothetical) regression of true earnings on the observed earnings measure:

$$\text{Reliability1}(r) = \frac{\text{cov}(\xi_i, r_i)}{\text{var}(r_i)}, \text{Reliability1}(s) = \frac{\text{cov}(\xi_i, s_i)}{\text{var}(s_i)} \tag{5}$$

where cov(.,.) means covariance and var(.) variance. Larger values of Reliability1 correspond to greater reliability. It is possible for Reliability1 to be greater than one if there is mean reversion.

[16] Jenkins and Rios-Avila (2020) fit Kapteyn-Ypma Full models without covariates using completely labelled sample fractions ranging between 0.25% ($|r_i - s_i| = 0$) and 17% ($|r_i - s_i| < 0.025$). Estimates were generally robust. The most noticeable effect of increasing the completely-labelled fraction over this range was a small increase in the proportion for whom there is no error in survey earnings (π_s) combined with a small decrease in the proportion with reference period error (π_ω).

The second statistic, Reliability2, proposed by Meijer et al. (2012), is the squared correlation between true earnings and an observed earnings measure:

$$\text{Reliability2}(r) = \frac{[\text{cov}(\xi_i, r_i)]^2}{\text{var}(\xi_i)\text{var}(r_i)}, \ \text{Reliability2}(s) = \frac{[\text{cov}(\xi_i, s_i)]^2}{\text{var}(\xi_i)\text{var}(s_i)}. \quad (6)$$

Reliability2 lies between 0 and 1.

We derive reliability statistics using analytical expressions for the relevant variances and covariances that are implied by each of our models.[17] Because the reliability statistics are model-specific, they should not be compared across models.[18] In other words, reliability statistics describe how close a given measure is to the true measure assuming the data generating process is described by the model under consideration.

In Section 6, we use model estimates to derive a measure of true earnings that combines the information from the survey and administrative measures. Meijer et al. (2012) show, using estimates from the Kapteyn and Ypma's (2007) Full model, that their true earnings predictors perform remarkably well in statistical terms (low mean squared error and high reliability) compared to each of the observed earnings measures. We have extended Meijer et al.'s prediction methods to apply to our more general models: see Jenkins and Rios-Avila (2021*b*) for details. In Section 6. we show earnings predictors from our more general models also perform remarkably well and illustrate how these 'hybrid' earnings measures can be used in regression analyses using earnings as an outcome variable or as an explanatory variable.

4. Estimates for models without covariates

Table 2 presents estimates of four models without covariates, ranging from the most general Extended model to the Constrained Full model. Goodness of fit summary statistics are reported near the bottom of the table. Goodness of fit is substantially better for the Extended models allowing for P14 measurement error than for the two models that do not, whether

[17] See Jenkins and Rios-Avila (2021*b*) for details. In Section 6 of this paper, we also calculate reliabilities using simulation methods so that we can compare them with the reliabilities of predicted hybrid earnings variables (which we derive using simulation because there are no closed form expressions for them).

[18] Abowd and Stinson argue that, in the absence of a genuine audit study, 'defining truth with respect to an observed quantity requires a researcher to place priors on which source of data is the most reliable' (2013, 1451) and they propose reliability statistics that differ according to the differing weights that researchers place on the survey and administrative data. Our modelling approach makes stronger assumptions than Abowd and Stinson's, not least because their dataset is richer than ours and all other datasets we are aware of. The advantage of doing so is that 'true' earnings are characterized by our model and we can discuss reliability. Also, we are sceptical that a genuine audit study is possible in practice.

assessed using the log pseudo-likelihood or AIC or BIC values (which adjust for differences in numbers of parameters estimated). For example, the BIC is 17,758 for the Extended model, 17,750 for the Constrained Extended model, 18,174 for the Full model, and 18,173 for the Constrained Full model. Of the two Extended models, we prefer the constrained one because, not only does it have the lower BIC but also the lower AIC and smaller log pseudo-likelihood in absolute value, and $\hat{\rho}_{\xi\omega}$ is insignificantly different from zero. Of the two Full models, we prefer the unconstrained one. Although it has a slightly larger BIC, its AIC and absolute log pseudo-likelihood values are smaller, and $\hat{\rho}_{\xi\omega}$ differs significantly from zero.

<Table 2 near here>

The Extended and Full models provide different perspectives on the role of reference period error. The probability an FRS earnings measure contains reference period error (π_ω) is estimated to be around 26% according to the two Full models, but around half as large (11%) according to the two Extended models. Correspondingly, the estimated probabilities of membership of latent classes 2 (type $R1,S2$) and 3 (type $R1,S3$), $\hat{\pi}_2$ and $\hat{\pi}_3$, are larger for the Full models. The distribution of reference period errors according to the two Full models has mean $\hat{\mu}_\omega = -0.12$, SD $\hat{\sigma}_\omega = 0.65$, and $\hat{\rho}_{\xi\omega} = -0.097$. The distribution of reference period error according to the Extended models has a more negative mean, $\hat{\mu}_\omega = -0.27$, greater dispersion, $\hat{\sigma}_\omega = 1.0$, and $\hat{\rho}_{\xi\omega} \approx 0$.

Put differently, all four models indicate that annualized FRS earnings under-estimate true annual earnings on average ($\hat{\mu}_\omega < 0$), but the under-estimation is greater according to our preferred (Constrained Extended) model and is combined with a greater dispersion in reference period errors.

<Table 2 near here>

Table 2. Four models without covariates: parameter estimates

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	9.8082***	9.8077***	9.8078***	9.8105***
	(0.0119)	(0.0112)	(0.0105)	(0.0104)
σ_ξ	0.7236***	0.7243***	0.7606***	0.7565***
	(0.0100)	(0.0093)	(0.0100)	(0.0100)
μ_ζ	8.1350***	8.0941***	8.4843***	8.6211***
	(0.1903)	(0.1687)	(0.1720)	(0.1347)
σ_ζ	1.2432***	1.2302***	1.2776***	1.2881***
	(0.0925)	(0.0862)	(0.0726)	(0.0622)
μ_ω	–0.2652***	–0.2656***	–0.1190***	–0.1239***
	(0.0479)	(0.0478)	(0.0234)	(0.0235)
σ_ω	0.9990***	1.0081***	0.6642***	0.6369***
	(0.1261)	(0.1052)	(0.0650)	(0.0708)
μ_η	–0.0102***	–0.0101***	–0.0094***	–0.0091***
	(0.0029)	(0.0029)	(0.0024)	(0.0025)
σ_η	0.0939***	0.0937***	0.1147***	0.1142***
	(0.0070)	(0.0066)	(0.0054)	(0.0059)
μ_υ	–0.0361	–0.0349		
	(0.0367)	(0.0335)		
σ_υ	0.3638***	0.3631***		
	(0.0262)	(0.0257)		
ρ_s	0.0072	0.0073	–0.0175***	–0.0192***
	(0.0040)	(0.0040)	(0.0043)	(0.0042)
ρ_r	0.0939	0.0908		
	(0.0692)	(0.0609)		
$\rho_{\xi\omega}$	0.0265		–0.0967**	
	(0.0697)		(0.0340)	
π_s	0.0516***	0.0517***	0.0361***	0.0365***
	(0.0053)	(0.0051)	(0.0025)	(0.0025)
π_ω	0.1093***	0.1093***	0.2649***	0.2606***
	(0.0219)	(0.0207)	(0.0201)	(0.0229)
π_r	0.9702***	0.9710***	0.9461***	0.9362***
	(0.0054)	(0.0056)	(0.0077)	(0.0070)
π_υ	0.6829***	0.6810***		
	(0.0511)	(0.0464)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0342***	0.0342***	0.0342***	0.0342***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)
π_2 (π_2)	0.5597***	0.5586***	0.6704***	0.6669***
	(0.0359)	(0.0338)	(0.0209)	(0.0235)
π_3 (π_3)	0.0687***	0.0685***	0.2415***	0.2351***
	(0.0184)	(0.0172)	(0.0178)	(0.0200)
π_4	0.0159***	0.0160***		
	(0.0039)	(0.0036)		

π_5	0.2599***	0.2617***		
	(0.0455)	(0.0410)		
π_6	0.0319***	0.0321***		
	(0.0039)	(0.0040)		
π_7 (π_4)	0.0015***	0.0015***	0.0019***	0.0023***
	(0.0004)	(0.0004)	(0.0003)	(0.0003)
π_8 (π_5)	0.0252***	0.0245***	0.0382***	0.0455***
	(0.0047)	(0.0049)	(0.0052)	(0.0045)
π_9 (π_6)	0.0031***	0.0030***	0.0138***	0.0160***
	(0.0007)	(0.0006)	(0.0025)	(0.0027)
Log pseudo-likelihood	–8805.1	–8805.2	–9030.4	–9034.3
AIC	17644.1	17642.3	18086.7	18092.6
BIC	17757.9	17749.5	18173.8	18173.0
Reliability1 (r)	0.7405	0.7416	0.7552	0.7377
Reliability1 (s)	0.8101	0.8100	0.8395	0.8401
Reliability2 (r)	0.7398	0.7410	0.7144	0.6906
Reliability2 (s)	0.8188	0.8156	0.8072	0.8245

Notes. Cluster-robust standard errors in parentheses (cluster is household). Sample N = 5,971 individuals within 4,874 households. Models based on a completely-labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$: see main text). Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Latent class definitions are shown in Table 1 and discussed in the main text. [an04]

All four models provide similar estimates for other model parameters. Hence for brevity we focus our discussion on the estimates of the Constrained Extended model.

True earnings We estimate mean true earnings to be $\hat{\mu}_\xi$ = 9.81, with SD $\hat{\sigma}_\xi$ = 0.72. This distribution has a higher mean (by around 5 log percentage points) and smaller dispersion than both FRS and P14 earnings (SD = 0.81, 0.84 respectively): see the notes to Figure 1. We discuss the over-estimation of true earnings inequality by the FRS and P14 measures in greater detail in Section 7.

Survey and administrative data measurement errors Measurement errors for survey and administrative data have means ($\hat{\mu}_\eta$, $\hat{\mu}_\upsilon$) that are both approximately zero but, perhaps surprisingly, the dispersion of P14 measurement error ($\hat{\sigma}_\upsilon$) according to the Extended Model is four times greater than the dispersion of FRS measurement error ($\hat{\sigma}_\eta$): 0.36 compared with 0.09. This is consistent with administrative processing errors leading to a small number of influential outlier entries on P14 forms. The probability that P14 data have no measurement error, $\hat{\pi}_\upsilon$ = 0.68, according to the two Extended models.[19] This may seem relatively low, but it is substantially greater than the estimated probability that FRS earnings are error-free, $\hat{\pi}_s$ =

[19] Similarly, for the USA, Abowd and Stinson (2013, Table 5) report larger model estimates for random effects variances for administrative (DER) data than for survey (SIPP) data.

5%. Mean reversion is absent in both FRS and P14 measurement errors: $\hat{\rho}_s$ and $\hat{\rho}_r$ are both approximately zero and imprecisely estimated. The result for $\hat{\rho}_s$ is consistent with the findings of other second-generation studies (Section 1); the result for $\hat{\rho}_r$ is consistent with our expectations.

Mismatch error and mismatch earnings The additional complication regarding P14 data is mismatch error, with estimated probability $1 - \hat{\pi}_r \approx 3\%$. As we show shortly, even this small amount has important consequences for the relative reliabilities of survey and administrative data. Mismatched observations are draws from the full P14 earnings distribution. According to the Constrained Extended model, $\hat{\mu}_\zeta$ = 8.09 and $\hat{\sigma}_\zeta$ = 1.23, i.e. this distribution has a smaller mean and greater dispersion than observed P14 earnings. Kapteyn and Ypma (2007) report similar differences (Tables 4 and C2). However, their estimate of σ_ζ/σ_r is larger than ours (1.88 compared with 1.46), a difference that we attribute to differences in sample composition. We use a sample whose ages span the full range; Kapteyn and Ypma's sample is more homogenous, containing only men and women aged 50+, leading to a relatively small σ_r.

The probability that the P14 data contain no measurement or mismatch errors is two-thirds, $\hat{\pi}_1 + \hat{\pi}_2 + \hat{\pi}_3 \approx 66\%$, according to the Constrained Extended model, a sharp contrast to the 95% implied by the Full models. The difference is due almost entirely to the Extended Models' incorporation of P14 measurement error rather than mismatch: the sum $\hat{\pi}_4 + \hat{\pi}_5 + \hat{\pi}_6$ accounts for the difference between the 66% and the 95%.

Reliability Table 2's bottom rows show reliability estimates. Regardless of model and for each reliability statistic, the conclusion is clear: the FRS data are more reliable than the linked P14 data. For example, according to the Constrained Extended model, Reliability2 is 0.74 for P14 data but 0.82 for FRS data. The explanation is that, although FRS data contain measurement error, so too do the P14 data which are also subject to mismatch error.

Meijer et al. (2012, Table 6) report Reliability2 statistics for Kapteyn and Ypma's (2007) estimates of what we label the Constrained Full model: 0.69 for their survey data and 0.47 for their administrative data. Meijer et al. (2012) note that this large difference might be because their model incorporates mismatch but not administrative data error. They comment that, were it to be incorporated, 'it is not immediately clear whether [errors] would be positively or negatively correlated with true earnings, and thus whether the reliability of register earnings as a measure of true earnings would be higher or even lower than estimated' (2012, 200). Our Extended models with administrative data error produce higher reliability estimates for the administrative data compared to the Swedish case, reflecting a combination of a point estimate $\hat{\rho}_r > 0$ (albeit not statistically significant) and using a data for respondents spanning the full age range.

5. Estimates for the Constrained Extended model with covariates

Models with covariates allow us to investigate factors associated with differences in measurement error distributions. We parameterise survey measurement errors using characteristics relating to individuals and administrative errors using characteristics relating to their jobs and employers, but do so parsimoniously for model fitting reasons. We derived all covariates from the FRS because there are no covariates in the P14 data (with one exception discussed below). Like other researchers, we ignore potential measurement errors in the covariates.[20]

Means for all covariates are in Appendix Table A3. To summarize: the sample is 57% female and the average age is 44 years. Just under a half have educational qualifications to A-level or higher (the minimum qualification for university entrance in the UK). Just over 70% are employed full-time and 70% are living with a spouse. Only 4% report having more than one employment. Around 70% work in the public sector in their main job. Around two-thirds showed a current or recent payslip to the interviewer when reporting earnings. Employers did not supply payslips in around one-tenth of cases. More than three-quarters (77%) are paid monthly. For 61% of the sample, the earnings spells reported in the P14 data cover the full 2011/12 financial year.[21]

Table 3 summarises goodness of fit statistics for four models (we discuss covariate specifications shortly). Every model fits substantially better – has substantially smaller AIC and BIC values – than its counterpart without covariates. Of the four models, we prefer the Constrained Extended one. It has the smallest AIC and BIC values and the log pseudo-likelihood is almost identical to that for the Extended model. The only difference between the

[20] The potentially most problematic variable is educational qualifications, for which around 9% of observations are missing. We assume these observations have educational qualifications below A-level standard. In preliminary analysis, we allocated the missing observations to an additional separate educational qualifications category. The coefficients were similar to the below-A-level category and estimates of other model parameters were no different.

[21] Appendix Table A3 also reports weighted means for this sample, and for the subsample of individuals aged 25–59 in full-time employment and not in education. Means for the main sample and subsample are quite similar; the most notable difference being that the proportion male is higher in the subsample (around 55% rather than around 43%). Weighting makes hardly any difference to estimates of means for either sample.

Extended and Constrained Extended model is that the former allows a non-zero correlation between reference period error and true earnings but we find $\hat{\rho}_{\xi\omega}$ = –0.050 (SE 0.069), i.e., of the expected sign but insignificantly different from zero. As for the models without covariates, goodness of fit for the two Extended models is substantially better than for the two Full models. The important lesson is that linked data models of earnings and measurement error should incorporate measurement error in the administrative data.

Table 3. Four models of log(earnings) with covariates: goodness of fit statistics and reliabilities

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo-likelihood	–6186.4	–6186.7	–6509.9	–6551.6
AIC	12474.8	12473.4	13097.9	13179.2
BIC	12816.2	12808.1	13359.0	13433.6
Reliability1 (r)	0.7279	0.7268	0.7560	0.7240
Reliability1 (s)	0.8260	0.8265	0.9159	0.8954
Reliability2 (r)	0.6935	0.6916	0.6930	0.6560
Reliability2 (s)	0.8413	0.8449	0.8837	0.8908

Notes. Sample N = 5,971 individuals within 4,874 households. Models based on a completely-labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$: see main text). [an22]

Table 3's bottom rows show reliability statistics. Regardless of model and for each reliability statistic, the conclusion is clear and the same as derived from the models without covariates: the FRS data are more reliable that the linked P14 data. For example, according to the Constrained Extended model, Reliability2 is 0.84 for FRS earnings but only 0.69 for P14 earnings.

We focus on the Constrained Extended model henceforth, with estimates shown in Table 4. In the top panels of the table are estimates of marginal means (and their standard errors), by covariate, with estimates for factor means (μ) and their standard errors shown on the left-hand side of the table and estimates for factor SDs (σ) on the right-hand side. The bottom panels of the table display estimates of the error probabilities (left-hand side) and latent class membership probabilities (right-hand side).

<Table 4 near here>

Table 4. Constrained Extended model with covariates: estimates of marginal means (MMs) and probabilities

		MM	(SE)		MM	(SE)
Male	μ_ξ	9.9462^{***}	(0.0113)	σ_ξ	0.5116^{***}	(0.0103)
Female		9.6876^{***}	(0.0095)		0.4747^{***}	(0.0077)
Education: less than A-level		9.6149^{***}	(0.0093)		0.4346^{***}	(0.0078)
Education: A-level or more		10.0071^{***}	(0.0118)		0.5540^{***}	(0.0098)
Full-time employee		10.0429^{***}	(0.0080)		0.4406^{***}	(0.0067)
Part-time employee		9.2097^{***}	(0.0181)		0.6167^{***}	(0.0149)
Married, cohabiting		9.8263^{***}	(0.0091)			
Single, divorced, separated, widowed		9.7377^{***}	(0.0125)			
Has 1 job		9.7954^{***}	(0.0081)			
Has 2+ jobs		9.9186^{***}	(0.0359)			
Age = 25 years		9.5195^{***}	(0.0147)		0.4188^{***}	(0.0115)
Age = 35 years		9.7722^{***}	(0.0097)		0.4757^{***}	(0.0078)
Age = 45 years		9.8993^{***}	(0.0101)		0.5115^{***}	(0.0085)
Age = 55 years		9.9011^{***}	(0.0107)		0.5207^{***}	(0.0094)
	μ_ζ	8.9563^{***}	(0.0954)	σ_ζ	1.2759^{***}	(0.0709)
Payslip(s) not shown to interviewer	μ_η	-0.0453^{***}	(0.0059)	σ_η	0.1286^{***}	(0.0095)
Payslip(s) shown (all jobs)		–0.0003	(0.0036)		0.0744^{***}	(0.0043)
Aged < 60 years		-0.0068^{*}	(0.0031)		0.0824^{***}	(0.0042)
Aged 60+ years		-0.0966^{***}	(0.0205)		0.1892^{***}	(0.0311)
Reference period: not 'other'	μ_ω	–0.0716	(0.0619)	σ_ω	0.9934^{***}	(0.1098)
Reference period: other		-0.5990^{***}	(0.1552)		0.6010^{*}	(0.2834)
Job spells do not span year		-0.2851^{***}	(0.0746)		0.8828^{***}	(0.0973)
Job spells all span year		0.0440	(0.0902)		1.0473^{***}	(0.1426)
Aged < 60 years		–0.0217	(0.0625)		0.9978^{***}	(0.1085)
Aged 60+ years		-0.6231^{***}	(0.1389)		0.8650^{**}	(0.2840)
Full-time employee		0.0390	(0.0780)		1.0423^{***}	(0.1008)
Part-time employee		-0.3789^{***}	(0.1031)		0.8421^{***}	(0.2123)
Payslip provided by employer	μ_υ	0.0166	(0.0099)	σ_υ	0.2284^{***}	(0.0198)
Payslip not provided by employer		-0.1415^{***}	(0.0427)		0.4445^{***}	(0.0500)
Full-time employee		0.0034	(0.0089)		0.1764^{***}	(0.0189)
Part-time employee		–0.0030	(0.0248)		0.4294^{***}	(0.0324)
Private sector employee		0.0258^{*}	(0.0114)		0.2655^{***}	(0.0189)
Public sector employee		-0.0530^{***}	(0.0131)		0.2186^{***}	(0.0316)
Male	ρ_s	0.0014	(0.0093)	ρ_r	0.0409^{*}	(0.0189)
Female		0.0023	(0.0073)		0.2464^{*}	(0.0994)
Aged < 60 years		0.0065	(0.0058)			
Aged 60+ years		–0.0389	(0.0371)			
Private sector employee					0.1103^{***}	(0.0235)
Public sector employee					–0.0513	(0.0320)
Probabilities	π_s	0.0685^{***}	(0.0065)	π_1	0.0342^{***}	(0.0024)
	π_ω	0.0847^{***}	(0.0119)	π_2	0.4259^{***}	(0.0317)
	π_r	0.9381^{***}	(0.0063)	π_3	0.0394^{***}	(0.0055)
	π_υ	0.5325^{***}	(0.0329)	π_4	0.0300^{***}	(0.0045)
				π_5	0.3739^{***}	(0.0228)
				π_6	0.0346^{***}	(0.0059)
				π_7	0.0042^{***}	(0.0007)
				π_8	0.0528^{***}	(0.0053)
				π_9	0.0049^{***}	(0.0008)

Notes. As for Table 3. Cluster-robust standard errors in parentheses (cluster is household). Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an22]

True earnings We model the mean of true log earnings (μ_ξ) as varying by sex, age, age-squared, educational qualifications, marital status, whether working part- or full-time, and number of jobs. We employ the same covariates in the SD equation (σ_ξ), except that we exclude marital status and number of jobs as they were never significant in preliminary analyses. Earlier research has modelled the mean as a function of characteristics but our heteroskedastic error specification allows for greater flexibility in distributional shape while remaining feasible to fit.

Differences in true earnings are as we expect. Men have not only a higher log earnings mean than women but also more dispersed earnings at each age. The estimates imply average earnings levels of £23,791 per year for men and £16,471 for women (in 2011/12 prices).[22] Individuals with at least university entry-level educational qualifications (1+ A-level exam grade passes or higher) earn more than less-qualified individuals and their SD is also greater. Hence predicted average earnings for the more educated group are substantially greater: £25,863 per year compared with only £18,582 per year. Individuals working full-time earn more than double than those working part-time on average: £25,336 per year compared with only £12,087 per year. True earnings are greater for married individuals than single people, and for those with more than one job. Average true earnings and their dispersion increase with age up to the mid-forties but then both flatten off: predicted average earnings are £14,871 per year at age 25, rising to £22,747 per year at age 45 and £22,849 at age 55.

Mismatched earnings We do not model the mean and SD of the mismatched earnings distribution as functions of characteristics because it makes no sense to model the relationship between some person *j* in the P14 distribution using the characteristics of some other person *i* from the linked dataset.[23] As for the models without covariates, and again as expected, the relatively large size of $\hat{\sigma}_\zeta$ stands out.

Survey measurement error We allow the distribution of survey measurement error (η) to differ by whether the individual showed a current or recent payslip to the FRS interviewer, and whether the respondent was aged 60+ years or not (age is related to the cognitive factors cited earlier). Mean measurement error is much the same, $\hat{\mu}_\eta = 0$, regardless of whether a current or recent payslip was shown to the interviewer, but $\hat{\sigma}_\eta$ is almost half as large for those

[22] Given lognormality, expected true earnings equals $\exp(\hat{\mu}_\xi + \hat{\sigma}_\xi^2/2)$.

[23] Some relationship may exist if (*j*, *i*) pairs share some characteristics. This might arise if observations were correctly matched according to a linkage variable such as sex (with the incorrect linkage arising from mismatch according to other linkage variables such as postcode).

who do show one compared to those who do not (0.07 versus 0.13). Thus, within-interview validation of responses can substantially reduce survey measurement error dispersion. Our finding provides strong support for the FRS's long-standing procedure of seeking such validations and suggests that efforts by data collectors to raise payslip prevalence further would improve survey data quality.

Mean survey measurement error is the same for workers aged 60+ or less than 60 years, zero, but $\hat{\sigma}_{\eta}$ is more than twice as large among the older worker group, 0.19 compared to 0.08. Potential explanations are that there are differences in cognitive function (accurate recall is harder for older workers) or that older workers have more variable working arrangements (making recall for a specific reference period more difficult).

Mean reversion in survey measurement error We allow ρ_s to vary by sex and age, prompted by the differences across groups in slope coefficients from regressions of FRS earnings on P14 earnings (Section 2). However, we find no statistically significant associations: point estimates equal zero for all groups. Thus, our research for the UK confirms the finding of other second-generation studies that there is no mean reversion in survey measurement error once one controls for administrative data error.

Reference period error We model the distribution of reference period error (ω) as depending on age, whether the reported reference period for the main job was 'other', part- versus full-time work status, and a measure of job stability. In preliminary analyses, we found no clear differences reference period error distributions when we differentiated between the full range of reported reference period options; hence, we simply report contrasts between workers reporting one of more standard options (weekly, monthly, annual, or their weekly equivalents) and workers reporting the 'other' option. We estimate this latter group (2% of the sample) to have a lower mean, $\hat{\mu}_{\omega} = -0.61$, which may be compared to –0.06 (not significantly different from zero) for the former group. This means that, among employees subject to reference period error, annualised FRS earnings under-estimate P14 annual earnings in the small group with 'other' survey reference periods. However, the dispersion of the 'other' group's reference period error is also markedly smaller: $\hat{\sigma}_{\omega} = 0.60$ compared to 0.99. There are substantial differences in reference period error by age. Mean error for respondents aged 60+ is substantially lower – under-estimation of annual P14 earnings by annualized FRS earnings is greater – than for younger workers: $\hat{\mu}_{\omega} = -0.62$ compared to zero. Differences in $\hat{\sigma}_{\omega}$ by age are small. Annualized survey earnings for part-time workers under-estimate P14 annual earnings, but $\hat{\sigma}_{\omega}$ is smaller for this group by comparison with full-time workers.

Reference period error distributions also differ by whether all job spells recorded in the P14 data for the respondent span the 2011/12 financial year. This is a measure of job stability – we expect those with more stable jobs to have less reference period error (see earlier). This is what we find: $\hat{\mu}_\omega$ is noticeably smaller for individuals with less stable jobs compared to those with stable jobs, –0.29 compared to zero, and $\hat{\sigma}_\omega$ is slightly smaller for those with less stable jobs.[24]

In sum, we find that among employees with reference period error, those with 'non-standard' job features have annualized FRS earnings that under-estimate annual P14 earnings and noisier.

Administrative data measurement error When modelling differences in the distribution of P14 measurement error (υ), we focused on covariates describing features of respondents' jobs or employers: whether the employer provided a payslip at all (for any job; according to the respondent), whether the employee worked in the public or private sector (main job), and whether worked full- or part-time. Not providing a payslip is an indicator of an employer's lack of capacity to provide accurate P14 returns. We find that such lack is associated with a lower error mean, $\hat{\mu}_\upsilon$ = –0.14 compared with zero, and $\hat{\sigma}_\upsilon$ is almost twice as large, 0.44 compared with 0.23. Thus, among observations with P14 measurement error, reported P14 earnings of those whose employer does not provide a payslip under-estimate true earnings on average but there is also substantial dispersion. We also hypothesize that P14 returns are less accurate for part-time workers compared to full-time workers (part-time workers may have more variable hours and hence pay, and so harder for employers to maintain accuracy in their year-end P14 returns), and for private sector employers compared to public sector employers (we expect full compliance with P14 reporting rules in the public sector but potentially not in the private sector, and the public sector is more likely to have appropriate payroll software).[25] It turns out that $\hat{\mu}_\upsilon$ is much the same, zero, across these groups. But there are differences in dispersion consistent with our hypotheses: $\hat{\sigma}_\upsilon$ is slightly

[24] The FRS provides a measure of how long a respondent reports having held the current job. This provides another perspective on job stability but provides no information about tenure after the FRS interview date. In preliminary analyses, we used the job tenure variable as a covariate in the reference period error equations using a range of specifications (continuous, categorical) but found that these models did not converge and/or estimates were statistically insignificant, especially when the length of P14 job spell variable was also included. We therefore dropped the tenure variable.

[25] In preliminary analysis, we also experimented with specifications in which the measurement error parameters varied with firm (and establishment) size and respondent's occupation (using SOC2010 codes for the main job). Models including these variables did not converge and/or estimates were statistically insignificant, especially when the public/private sector variable was also included, and so we dropped them.

smaller for public sector workers than private sector workers (0.22 compared to 0.27), and markedly larger for part-time workers than full-time workers (0.43 compared to 0.18).

Administrative data mean reversion We examined whether the correlation between P14 errors and true earnings differed by sex and whether the main job was in the public or private sector. It turns out that the $\hat{\rho}_r \approx 0$ finding reported for Constrained Extended models without covariates (Table 2) disguises some differences across groups. For women, $\hat{\rho}_r$ = 0.23 compared to zero for men, and $\hat{\rho}_r$ = 0.11 for workers in the private sector compared to zero for public-sector workers. Our explanation for these mean-affirming errors is that they reflect reference period errors for some groups that are not adequately described by our specification in equation (2). For example, poorly-paid part-time workers (mostly women) may be reporting earnings in the FRS but their employers not recording these (occasional?) earnings on year-end P14 forms (hence apparent P14 under-reporting below the mean). Annual bonuses are more likely to be paid to high-paid private sector employees and we conjecture that they are not well-captured by the FRS current pay measures but are included in P14 annual earnings – hence apparent P14 over-reporting above the mean.

Error and latent class probabilities There are some differences in error probabilities for the Constrained Extended model with covariates compared to their counterparts in the model without covariates. The former model is a mixture of conditional densities, not of unconditional densities and, as within-class earnings differences are more accurately characterized, there are also corresponding changes in the class probabilities to best fit the data. These class probabilities depend on the error probabilities (Table 1). The probability that FRS earnings are error-free increases slightly ($\hat{\pi}_s$ = 0.07, compared to 0.05) and the probability of reference period error is a little smaller ($\hat{\pi}_\omega$ = 0.08, compared with 0.11). More noticeably, the mismatch error probability is larger (1–$\hat{\pi}_r$ = 0.06 compared with 0.03), and the probability of administrative data measurement error is also larger (1–$\hat{\pi}_\upsilon$ = 0.47 compared with 0.32).

The most noticeable consequences are that the model with covariates estimates a smaller probability of error-free P14 earnings being combined with FRS earnings with measurement error (type $R1$,$S2$): $\hat{\pi}_2$ = 0.43 compared to 0.59 according to the no-covariate model. And there is a larger probability of error-ridden P14 earnings being combined with error-ridden FRS earnings (type $R2$,$S2$): $\hat{\pi}_5$ = 0.37 compared to 0.26 according to the no-covariate model.

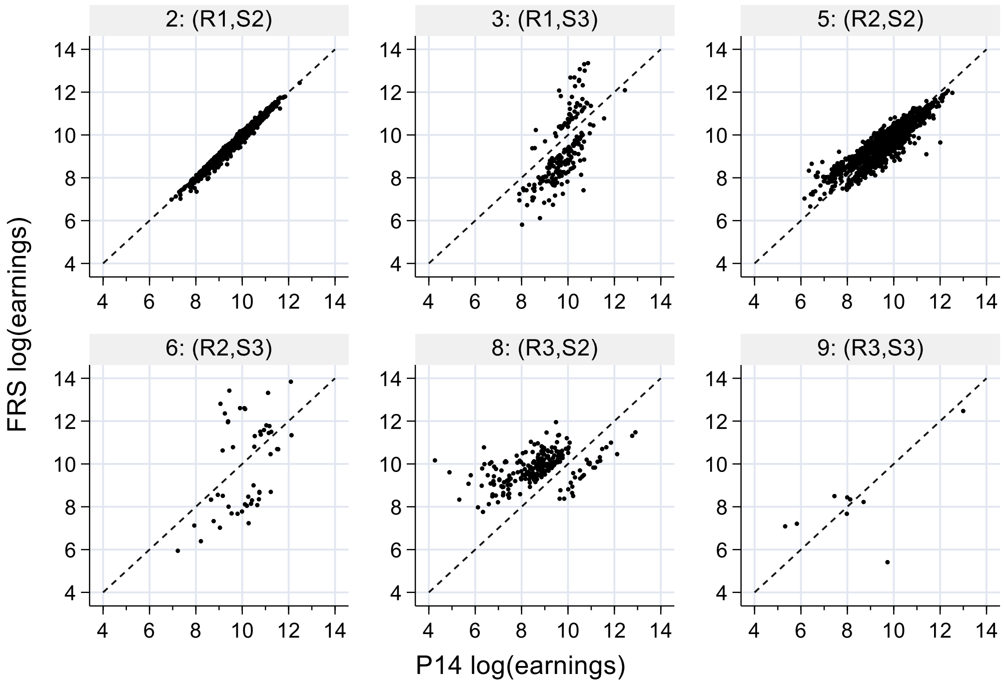
6. Further implications of the estimates

We investigate the implications of the model estimates in further detail in this section. We begin by allocating observations to latent classes: model estimates tell us the fraction of individuals in each group but not the combinations of survey and administrative earnings associated with class memberships and these are interesting to know – we get a picture of the size of the various different errors in each case. Following Kapteyn and Ypma (2007), Meijer et al. (2012), and Jenkins and Rios-Avila (2021*a*), we allocate an individual to the class for which the posterior probability of membership is the largest, with posterior probabilities calculated using the Constrained Extended model with covariates.

Figure 4 summarizes the results. Not shown are class 1 (completely labelled observations), or classes 4 and 7 (no observations were predicted to belong to them). The largest class is 2 with error-free P14 earnings and FRS measurement error (a combination of observation types *R*1 and *S*2). Figure 4 shows its members are located close to the 45° line – survey measurement errors are not large. The second largest class is 5 with measurement error in both data sources (*R*2,*S*2), and P14 and FRS observations are relatively close to the 45° line. For both this class and class 8 (type *R*3,*S*2), low P14 earnings tend to be associated with higher FRS earnings (above the 45° line). These are the observations contributing to the mean reversion observed in the raw data (Figure 3). Reference period error is present in classes 3, 6, and 9 (*S*3 cases). The majority of members of these classes lie below the 45° line but there is also a substantial range in FRS earnings associated with a given P14 earnings level. This is the visual counterpart of the $\hat{\mu}_{\omega} < 0$ and relatively large $\hat{\sigma}_{\omega}$ shown in Tables 2 and 4. Mismatched observations belong to classes 8 and 9 (*R*3 cases). Most of these observations lie above the 45° line: FRS earnings are greater than (incorrect) P14 earnings on average, as the model estimates indicated.

<Figure 4 near here>

Figure 4. Predicted latent class memberships and observed FRS and P14 earnings



Notes. For the sample of (s_i, r_i) pairs shown in Figure 3, the charts show predicted latent class memberships (classes as defined in Table 1). The observations for class 1 ($R1$, $S1$), not shown, lie on the 45° line by assumption. No observations are predicted to belong to classes 4 ($R2$, $S1$) or 7 ($R3$, $S1$). Predictions based on Constrained Extended model with covariates (Table 4) and formulae given by Jenkins and Rios-Avila (2021*b*).

Now we generate 'hybrid' earnings predictions combining information from FRS and P14 data and model estimates. We consider the seven types of predictor developed by Meijer et al. (2012), extended to our mixture models (see Jenkins and Rios-Avila 2021*b*). Table 5 summarizes the precision of the predictors calculated using estimates from the Constrained Model with covariates. We calculate all statistics using simulation methods (as Meijer et al. 2012 did).[26] In addition to reliabilities, we show estimates of Mean Squared Error (MSE) and its Bias and Variance components.

There are two striking conclusions. First, every hybrid earnings measure is more precise than either of the two observed earnings measures, whether judged in terms of reliability or MSE. Second, the fourth and sixth of the hybrid measures perform particularly well, i.e. the Weighted (conditional) unbiased and Two-stage unbiased predictors. Both have Reliability2 estimates of around 0.96 and MSE of around 0.02, with the fourth predictor performing slightly better. We use the fourth predictor as the hybrid earnings measure in

[26] This explains why the reliability estimates for FRS and P14 earnings differ slightly from those shown in Table 3 (derived using analytical formulae). Closed form expressions do not exist for all the predictors.

what follows.

Table 5. Precision of earnings predictors

	Reliability1	Reliability2	MSE	Bias	Variance
P14 data (*r*)	0.7262	0.6911	0.2177	–0.0519	0.2150
FRS data (*s*)	0.8270	0.8455	0.1016	–0.0211	0.1011
Hybrid earnings predictors					
1. Weighted (unconditional)	0.8897	0.8574	0.0831	–0.0295	0.0822
2. Weighted (unconditional) unbiased	0.8728	0.8504	0.0895	–0.0327	0.0885
3. Weighted (conditional)	0.9699	0.9461	0.0289	0.0002	0.0289
4. Weighted (conditional) unbiased	0.9855	0.9640	0.0191	0.0001	0.0191
5. Two-stage	0.9643	0.9415	0.0315	–0.0006	0.0315
6. Two-stage, unbiased	0.9730	0.9546	0.0243	0.0009	0.0243
7. System-wide linear	1.0002	0.9027	0.0513	0.0001	0.0513

Notes. Predictors 1–7 derived using estimates from Constrained Extended model (Table 4) and the formulae reported by Jenkins and Rios-Avila (2021*b*). All statistics computed using simulation (1,000 repetitions). MSE: E(predictor – ξ)^2 = Bias^2 + Variance. [an22_postest]

Table 6 compares the distributions of true, hybrid earnings, and observed earnings. Column 1 refers to true earnings, with estimates calculated using the Constrained Extended model without covariates to derive an unconditional mean and SD (Table 2 estimates). Column 2 is based on the hybrid earnings measure just discussed and the other two columns refer to FRS and P14 earnings.

Table 6. Distributions of predicted and observed earnings: summary statistics

Log earnings	True (ξ_i, mixture model without covariates) (1)	Predicted (P14 and FRS data) (2)	FRS data (3)	P14 data (4)
Mean	9.81	9.80	9.77	9.75
*p*10	8.88	8.79	8.69	8.69
*p*50	9.81	9.86	9.83	9.84
*p*90	10.74	10.67	10.71	10.70
*p*50 – *p*10	0.93	1.07	1.14	1.14
*p*90 – *p*50	0.93	0.81	0.88	0.86
*p*90 – *p*10	1.86	1.87	2.01	2.00
Standard deviation	0.73	0.73	0.81	0.84

Notes. *p*XX refers to the XXth percentile of the log earnings distribution. Statistics rounded to 2 d.p. Column 1 estimates are derived from the parameter estimates of the Constrained Extended model without covariates (Table 2), with percentiles calculated using the normality assumption. Column 2 is based on distribution of earnings predicted using the 'weighted (conditional) unbiased' predictor and estimates from the Constrained Extended model with covariates (see Table 5). [an22_postest]

The headline finding is that the FRS and P14 measures over-estimate true earnings inequality whether we use the mixture model estimate or hybrid earnings measure. Taking the SD of log earnings as the inequality measure, the upward bias in the FRS measure is about 11% and about 15% in the P14 measure. Upward biases are smaller, around 8%, if we summarize inequality using the difference between the 90th and 10th percentiles of the log earnings distribution. The bias is larger if inequality is summarized by the *p*90/*p*10 ratio for earnings levels, by around 18% for FRS earnings (*p*90/*p*10 is 6.50 for true earnings, 7.75 for FRS earnings, and 7.67 for P14 earnings). This metric also allows us to benchmark the bias. For example, *p*90/*p*10 for annual earnings for UK employees on all adult rates retaining the same job for a year rose from 7.06 in 2006 to 7.55 in 2011, an increase of 7%.[27] In other words, the estimated bias of 18% in cross-sectional inequality is greater than the change in earnings inequality in the UK in the period spanning the Great Recession.

This over-estimation result contrasts with the finding of Gottschalk and Huynh (2010) who reported no bias in survey earnings inequality measured in terms of the SD: mean reversion offset the effect of survey measurement error. Our analysis indicates that, once you model administrative measurement error and mismatch error (which Gottschalk and Huynh did not do), there is no mean reversion component to survey measurement error to provide an offsetting effect (Section 4).

The hybrid earnings distribution has the same mean and SD as the modelled true distribution but also incorporates (by construction) features of the FRS and P14 earnings distributions. In particular, the hybrid distribution has greater lower tail inequality (*p*50–*p*10) than upper tail inequality (*p*90–*p*50), as the observed earnings distributions do. In the estimated true distribution, the two inequalities are equal by assumption.

We now discuss two further exercises to illustrate the usefulness of a hybrid earnings variable. First, we use it as a dependent variable in an earnings regression equation and compare estimates with the corresponding mixture model estimates and estimates arising if FRS and P14 earnings are used as dependent variables. We use the same heteroskedastic specification as before. Table 7, column 1, reports the estimates from the Constrained Extended model. (Here we show fitted coefficients rather than the marginal means implied by them, as displayed in Table 4.) Column 2 shows the coefficient estimates when the hybrid earnings variable is the dependent variable and reassuringly each of them is remarkably close to their mixture model counterparts.

---

[27] Source: Annual Survey of Hours and Employment, Table 1.7a, various years. These ASHE data are the only statistics available for annual earnings in the UK. ASHE is an employer-based survey.

Table 7. Regressions using different log earnings dependent variables

Dependent variable: log earnings	True (ξ_i, mixture model)	Predicted true (P14 and FRS data)	FRS data	P14 data
	(1)	(2)	(3)	(4)
Mean				
Female	–0.2586***	–0.2606***	–0.2376***	–0.2611***
	(0.0133)	(0.0125)	(0.0159)	(0.0174)
Age (years)	0.0629***	0.0642***	0.0785***	0.0694***
	(0.0039)	(0.0035)	(0.0040)	(0.0049)
Age squared	–0.0006***	–0.0006***	–0.0008***	–0.0007***
	(0.0000)	(0.0000)	(0.0000)	(0.0001)
Education: A-level or more	0.3922***	0.4008***	0.4043***	0.4234***
	(0.0139)	(0.0131)	(0.0155)	(0.0175)
Part-time employee	–0.8332***	–0.8351***	–0.8963***	–0.8177***
	(0.0194)	(0.0162)	(0.0196)	(0.0205)
Has 2+ jobs	0.1232***	0.1271***	0.1120***	0.1123**
	(0.0363)	(0.0354)	(0.0434)	(0.0412)
Single, widowed, divorced, separated	–0.0885***	–0.0889***	–0.0968***	–0.1026***
	(0.0142)	(0.0135)	(0.0165)	(0.0188)
Constant	8.5649***	8.5284***	8.2616***	8.3246***
	(0.0802)	(0.0723)	(0.0841)	(0.1031)
Log of standard deviation				
Female	–0.0749**	–0.0680**	0.0041	–0.1657***
	(0.0252)	(0.0235)	(0.0353)	(0.0378)
Age (years)	0.0292***	0.0290***	0.0184**	–0.0107
	(0.0066)	(0.0057)	(0.0071)	(0.0089)
Age squared	–0.0003***	–0.0003***	–0.0002	0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Education: A-level or more	0.2426***	0.2302***	0.1765***	0.1199***
	(0.0243)	(0.0225)	(0.0315)	(0.0360)
Part-time employee	0.3363***	0.3196***	0.2345***	0.2271***
	(0.0283)	(0.0262)	(0.0363)	(0.0362)
Constant	–1.6148***	–1.6074***	–1.2299***	–0.3301
	(0.1417)	(0.1227)	(0.1520)	(0.1872)

Notes. The top panel shows estimates for mean earnings; the bottom panel shows estimates for the log of the standard deviation of earnings. Column 1 reports parameter estimates from the constrained Extended model used to derive the marginal mean estimates shown in Table 4. Column 2 reports parameter estimates derived using 'weighted (conditional) unbiased' predicted earnings (see Table 5). Log pseudo-likelihood for mixture model is shown in Table 3. Columns 3 and 4 report parameter estimates derived using observed FRS and P14 earnings, respectively. Columns 2–4 fit heteroskedastic regression models by maximum likelihood. Cluster–robust SEs in parentheses (cluster is household). Sample described in notes to Tables 3 and 4. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an22_postest]

When FRS and P14 earnings are the dependent variables (columns 3 and 4), coefficients in the mean component of the models are both quite close to the benchmarks in columns 1 and 2. In the log-SD component of the model, more substantial differences from the benchmarks appear. Overall, the estimates for FRS earnings are closer to them than are the estimates for P14 earnings.

What explains this? Kapteyn and Ypma (2007) derive coefficient bias expressions for the special case in which the log earnings regression has only one explanatory variable and true earnings are described by the (Constrained) Full model. Their analysis indicates that the survey earnings estimator of the slope coefficient is consistent if there is no survey mean reversion, and less biased than the corresponding coefficient from administrative data if there is mismatch (and no mean reversion). Although we have multiple explanatory variables, these conditions otherwise correspond to our mixture model estimates, and our models also incorporate administrative data errors that are likely to worsen the relative performance of the P14 data estimates.[28]

Our second exercise uses the hybrid earnings variable as an explanatory variable and compares estimates with situations in which FRS or P14 earnings are used instead. Our illustration models the probability that an individual is a member of an employer pension scheme, using the subsample of individuals for which there is an eligible pension scheme. Explanatory variables in addition to earnings are sex, a quadratic in age, whether working part-time, occupational category (SOC2010 groups). Table 8 shows the estimates, reported in the form of average marginal effects. According to the benchmark with hybrid earnings (column 1), membership probabilities are greater for women than men, part-time workers, and for those working in more professional and white-collar occupations, and higher earners. As in the first exercise, estimates from the model using FRS earnings are closer to those for the benchmark model than are the estimates from the model using P14 earnings. (There are also differences in statistical significance.[29]) The coefficient on the earnings variable highlights the differences across the models. For the model with hybrid earnings, an increase in log earnings of 0.5 (under half the gap between $p50$ and $p10$) is associated with an increase of almost 9 percentage points (0.5×0.1744) in the probability of belonging to an employer pension scheme. When FRS earnings is the explanatory variable, the estimated increase is 7.5 percentage points but, when we use P14 earnings, the estimated increase is markedly smaller, only 5.6 percentage points.

[28] Kapteyn and Ypma's (2007, Table C7) earnings regressions with different dependent variables provide less clear-cut evidence about whether the survey earnings estimates are closer to the true benchmarks than the administrative data estimates. Their regressions do not incorporate a heteroskedasticity component.
[29] Our comparisons of SEs do not take account of the fact that hybrid earnings is a generated variable.

Table 8. Probit regressions for employer pension scheme membership using different log earnings explanatory variables: average marginal effects

Log earnings:	Predicted (P14 and FRS data)	FRS data	P14 data
	(1)	(2)	(3)
Female	0.0446^{**}	0.0352^{*}	0.0278
	(0.0156)	(0.0157)	(0.0157)
Part-time employee	0.0726^{***}	0.0628^{***}	0.0330
	(0.0172)	(0.0171)	(0.0177)
Age (years)	0.0035^{***}	0.0038^{***}	0.0038^{***}
	(0.0007)	(0.0007)	(0.0007)
SOC2010 category†			
Professional	0.1924^{***}	0.1923^{***}	0.1928^{***}
	(0.0295)	(0.0293)	(0.0288)
Associate professional & technical	0.1158^{***}	0.1147^{***}	0.1051^{***}
	(0.0319)	(0.0317)	(0.0314)
Administrative & secretarial	0.1290^{***}	0.1232^{***}	0.1029^{**}
	(0.0325)	(0.0324)	(0.0323)
Skilled trades	0.0561	0.0468	0.0311
	(0.0390)	(0.0390)	(0.0394)
Caring, leisure, other services	0.0910^{**}	0.0789^{*}	0.0625
	(0.0353)	(0.0350)	(0.0353)
Sales & customer services	-0.0981^{*}	-0.1077^{*}	-0.1413^{***}
	(0.0422)	(0.0420)	(0.0420)
Process, plant, machine operatives	0.0439	0.0416	0.0075
	(0.0396)	(0.0399)	(0.0400)
Elementary	0.0336	0.0214	–0.0171
	(0.0387)	(0.0387)	(0.0389)
Log earnings	0.1744^{***}	0.1511^{***}	0.1134^{***}
	(0.0155)	(0.0146)	(0.0125)

Notes. Outcome variable refers to main job. Table shows average marginal effects (discrete changes for categorical variables). Mean of dependent variable = 0.76. †: Reference SOC2010 category is 'managers, directors, senior officials'. Cluster–robust SEs in parentheses (cluster is household). Sample N = 3,784 individuals in jobs with the possibility of employer pension membership. Statistical significance indicators: $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$. [an22_postest]

The attenuation in the coefficients on the variables containing errors is unsurprising; more interesting is why the attenuation is greater for the P14 earnings variable. Again, Kapteyn and Ypma's (2007) analysis is suggestive. They consider a linear model in which log earnings is the only explanatory variable and show that the relative biases of the survey and administrative data earnings cases are complicated functions of the mixture model parameters. However, if there is no survey mean reversion, and making some plausible assumptions about factor means and SDs, they find that 'survey data exhibit less bias if the reliability ratio of the survey data is greater than the proportion of perfect administrative data. In the more general case where $\mu_\xi \neq \mu_\zeta$ and $\sigma_\zeta^2 \geq \sigma_\xi^2$, the balance tips a little more in favor of the survey data' (2007, 531). Kapteyn and Ypma (2007) did not consider measurement error in their administrative data but its presence is likely to increase the gap between the reliability of the survey and administrative earnings variables.

Overall, these two exercises illustrate the feasibility of using hybrid earnings variables in regression analysis. The exercises also show that using such a (highly reliable) variable leads to estimates that differ from those derived using observed survey and administrative earnings measures.

7. Summary and Conclusions

Much research has argued that survey data on earnings suffer from measurement error; little research has drawn attention to measurement errors in administrative data on earnings. Our paper demonstrates the importance of both types of error in the UK context. We have also taken account of the complication that UK surveys provide 'annualised' measures rather than genuinely annual measures as in the P14 data (reference period error), and shown how models with covariates are informative.

We estimate that more than 90% of FRS earnings are subject to measurement error whereas the probability of P14 measurement error is only around one-third. However, the variance of P14 errors is larger than the variance of FRS errors. The reliability of linked P14 data is further diminished by a small but consequential fraction of records that are mismatched. For the FRS, too, measurement error is not the only problem: there is a probability of reference period error of around one tenth and when it occurs the FRS annualised current earnings measure underestimates annual earnings on average and is noisier. Overall, FRS earnings are more reliable than P14 earnings.

The combination of errors means that observed earnings, whether from the FRS or P14 data, substantially overestimate true earnings inequality. There is insufficient mean reversion to offset the greater noise arising from measurement error.

Our models with covariates highlight factors associated with poorer data quality. For example, survey measurement error variance is greater for workers not showing a payslip to validate oral responses, and for older workers. P14 measurement error variance is greater for respondents whose employers do not provide payslips and among employees working part-time. On the one hand, there is a positive take-away for labour economists: restricting survey samples to full-time earners of 'standard' working age reduces the variance of measurement error and annualised current earnings measures would be less biased measures of annual earnings.[30] On the other hand, such sample selections are inappropriate for derivation of household income measures that are the basis of national statistics on inequality and poverty. Our findings point to potential areas in which to target initiatives to improve data quality, for example, encouraging greater use of payslips and other records in survey interviews, and improving employers' payroll systems especially for employees in part-time and less stable jobs.

Another positive take-away is that it is possible to derive hybrid earnings measures based on both survey and linked administrative data that have substantially greater reliability than either source used separately. These hybrid measures also provide a route for data producers to provide researchers with higher quality earnings data without releasing the confidential administrative data on which they are based.

Our research has confirmed the importance of taking account of errors in administrative data (as found in the small number of other second-generation studies) and we have shown the usefulness of models incorporating parameter heterogeneity. Nonetheless, there are issues that could be addressed in further research. For example, we have addressed the issue of reference period error but the finding of mean-affirming P14 errors for some groups suggests that our specification could be revisited. We assume (as other researchers have) that true earnings are normally distributed. Clearly, one could use more flexible functional forms to characterize marginal distributions of earnings and errors (e.g. Generalized Beta of the Second Kind). However, we need bivariate counterparts to these distributions to specify cross-factor correlations and the likelihood function and we are not aware of any. Our models could be extended further with longitudinal linked data. The distribution of latent true earnings already depends on time (because the mean and SD depend on age), but extensions could, for example, introduce autocorrelated measurement errors in the survey and administrative data (Abowd and Stinson 2013) and time-varying error means and SDs.

Updating our analysis to a later year to assess whether the nature of measurement error has changed would be valuable in any case. Since 2011/12, the UK has changed the technology used to administer the withholding system for income tax and National Insurance. Starting April 2012, HMRC began to phase in a system of Real Time Information (RTI) and,

[30] This is confirmed by our estimates for the subsample of full-time workers aged 25–59: see Jenkins and Rios-Avila (2021*c*, Appendices B–E).

since April 2014, employers must communicate to HMRC information about tax and other deductions under PAYE every time an employee is paid. Year-end P14 forms no longer exist and HMRC's administrative data on employee earnings are now based on RTI (Office for National Statistics 2019). Moreover, taking advantage of legislative changes, the FRS no longer asks respondents for consent to data linkage; instead, they are informed prior to interview that responses will be linked to administrative data for statistical and research purposes. In addition, data linkage also incorporates probabilistic matching. As a result, linkage rates are expected to be close to 90% (Burke and Matejic 2018) which is substantially greater than the case for 2011/12 that we have analysed. Although these changes are likely to reduce issues related to selectivity of consent, it is unclear how the move to RTI has changed the probability of linkage mismatch and the intrinsic quality of administrative data earnings measures. Also, reference period issues continue, albeit in different form, because of the 'calendarisation' methodology the RTI data use (Office for National Statistics 2019). In sum, the modelling approach taken in this paper will remain applicable when linked RTI data become available to researchers.

References

Abowd, J. and Stinson, M. (2013). Estimating measurement error in annual job earnings: a comparison of survey and administrative data. Review of Economics and Statistics, 95 (2), 1451–1467.

Angel, S., Disslbacher, F., and Humer, S. (2019). What did you really earn last year? Explaining measurement error in survey income data. Journal of the Royal Statistical Society, Series A, 182 (4), 1411–1437.

Autor, D. H., Katz, L. F., and Kearney, M. S, (2008). Trends in U.S. wage inequality: revising the revisionists. Review of Economics and Statistics, 90 (2), 300–323.

Blundell, R., Joyce, R., Norris-Keillor, A., and Ziliak, J. (2018). Income inequality and the labour market in Britain and the US. Journal of Public Economics, 162 (June), 48–62.

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 (4), 733–758.

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., and Ziliak, J. P. (2018). The good, the bad and the ugly: measurement error, non-response and administrative mismatch in the

CPS. Working Paper, Gatton College of Business, University of Kentucky.

Bollinger, C. R., Hirsch, B. T., Hokayem, C. M., and Ziliak, J. P. (2019). Trouble in the tails? What we know about earnings nonresponse 30 years after Lillard, Smith, and Welch', Journal of Political Economy, 127 (5), 2143–2185.

Bourquin, P., Joyce, R., and Norris-Keillor, A. (2020). Living Standards, Poverty and Inequality in the UK: 2020. Report 170. London: Institute for Fiscal Studies.

Burke, D. and Matejic, P. (2018). Family Resources Survey and related series – update and developments. Family Finance Surveys User Conference presentation.

Department for Work and Pensions (2013). Family Resources Survey United Kingdom, 2011/12.

Department for Work and Pensions (2020). Households Below Average Income. London: Department for Work and Pensions.

Golden, L. (2016). Still falling short on hours and pay. Part-time work becoming new normal. Washington DC: Employment Policy Institute.

Gottschalk, P. and Smeeding, T. M. (1997). Cross-national comparisons of earnings and income inequality. Journal of Economic Literature, 35 (2), 633–687.

Hirsch, B. T. and Schumacher, E. J. (2004). Match bias in wage gap estimates due to earnings imputation. Journal of Labor Economics, 22 (3), 689–722.

Hyslop, D. R. and Townsend, W. (2020). Earnings dynamics and measurement error in matched survey and administrative data. Journal of Business and Economic Statistics, 38 (2), 457–469.

Jenkins, S. P., Lynn, P., Jäckle, A., and Sala, E. (2008). The feasibility of linking household survey and administrative record data: new evidence for Britain. International Journal of Social Research Methodology, 11 (1), 29–43.

Jenkins, S. P. and Rios-Avila, F. (2020). Modelling errors in survey and administrative data on labour earnings: sensitivity to the fraction assumed to have error-free earnings. Economics Letters, 192, 109253.

Jenkins, S. P. and Rios-Avila, F. (2021*a*). Measurement error in earnings data: replication of Meijer, Rohwedder, and Wansbeek's mixture model approach to combining survey and register data, Journal of Applied Econometrics, online first.

Jenkins, S. P. and Rios-Avila, F. (2021*b*). Finite mixture models for linked survey and administrative data: estimation and post-estimation. IZA Discussion Paper, forthcoming.

Kapteyn, A. and Ypma, J. Y. (2007). Measurement error and misclassification: a comparison of survey and administrative data. Journal of Labor Economics, 25 (3), 513–551.

Kim, C. and Tamborini, C. R. (2014). Response error in earnings: an analysis of the Survey of Income and Program Participation matched with administrative data, Sociological Methods and Research, 43 (1), 39–72.

Kopczuk, W., Saez, E., and Song, J. (2010). Earnings inequality and mobility in the United States: evidence from Social Security data since 1937, Quarterly Journal of Economics, 125 (1), 91–128.

Kristensen, N. and Westergaard-Nielsen, N. (2007). A large-scale validation study of measurement errors in longitudinal survey data. Journal of Economic and Social Measurement, 32 (2), 65–92.

Levy, F. and Murnane, R. J. (1992). U.S. earnings level and earnings inequality: a review of recent trends and proposed explanations. Journal of Economic Literature, 30 (3), 1333–1381.

Meijer, E., Rohwedder, S., and Wansbeek, T. (2012). Measurement error in earnings data: using a mixture model approach to combine survey and register data. Journal of Business and Economic Statistics, 30, 191–201.

Michaud, S., Dolson, D., Adams, D., and Renaud, M. (1995). Combining administrative and survey data to reduce respondent burden in longitudinal surveys. Proceedings of the Survey Research Methods Section, American Statistical Association.

Office for National Statistics (2019). New methods for monthly earnings and employment estimates from Pay As You Earn Real Time Information (PAYE RTI) data: December 2019.

Redner, R. A. and Walker, H. F. (1984). Mixture densities, maximum likelihood and the EM algorithm. SIAM Review, 26 (2), 195–239.

Yakowitz, S. J. and Spragins, J. D. (1968). On the identifiability of finite mixtures. Annals of Mathematical Statistics, 39 (1), 209–214.

APPENDICES

to

Reconciling reports: modelling employment earnings and measurement errors using linked survey and administrative data

Stephen P. Jenkins
(LSE)

Fernando Rios-Avila
(The Levy Institute)

APPENDICES

Appendix A provides the expressions for the mixture model log-likelihood function and for the earnings distribution for each of the latent classes, explains the inverse-probability weighting scheme we developed to address potential sample selection issues, and provides summary statistics for covariates for the various samples.

Appendices B–E contain additional estimates. The tables in the main text refer to estimates derived from the unweighted sample of 'all individuals' and a completely labelled fraction defined using the criterion $|r_i – s_i| \leq 0.005$. Appendices B–E provide additional estimates for this sample (using weighted data), and also for (i) the sample of individuals aged 20–59 years, working full-time, and not in education; and (ii) a completely labelled fraction defined using the criterion $\delta = |r_i – s_i| \leq 0.010$ (weighted and unweighted estimates).

The 'constrained Extended' model is the Extended model with constraint $\rho_{\xi\omega} = 0$, i.e. 'Model 5' in Jenkins and Rios-Avila (2021*b*).

Appendix A

- Table A1. Mixture model log-likelihood function and latent class earnings distributions
- Table A2. Inverse-probability weights
- Table A3. Covariate means, by sample

Appendix B (all individuals; $|r_i – s_i| \leq 0.005$)

- Table B1. Weighted estimates for four models of log(earnings) without covariates [an14w]
- Table B2. Goodness of fit statistics for four models of log(earnings) with covariates: weighted estimates [an22w]
- Table B3. Weighted estimates of constrained Extended Model of log(earnings) with covariates [an22w]

Appendix C (individuals aged 20–59 years, working full-time, not in education; $|r_i – s_i| \leq$ 0.005)

- Figure C1. Distributions of FRS and P14 log(earnings)
- Figure C2. Distribution of difference between FRS and P14 earnings ($s – r$)
- Figure C3. The relationship between FRS and P14 earnings
- Table C1. Unweighted estimates of four models of log(earnings) without covariates [an27]

- Table C2. Weighted estimates of four models of log(earnings) without covariates [an27w]
- Table C3. Goodness of fit statistics for four models of log(earnings) with covariates: unweighted estimates [an32]
- Table C4. Unweighted estimates of Constrained Extended Model of log(earnings) with covariates [an32]
- Table C5. Goodness of fit statistics for four models of log(earnings) with covariates: weighted estimates [an32w]
- Table C6. Weighted estimates of Constrained Extended Model of log(earnings) with covariates [an32w]

Appendix D (all individuals; $|r_i – s_i| \leq 0.010$)

- Table D1. Unweighted estimates of four models of log(earnings) without covariates [an05]
- Table D2. Weighted estimates of four models of log(earnings) without covariates [an05w]
- Table D3. Goodness of fit statistics for four models of log(earnings) with covariates: unweighted estimates [an25]
- Table D4. Unweighted estimates of Constrained Extended Model of log(earnings) with covariates [an25]
- Table D5. Goodness of fit statistics for four models of log(earnings) with covariates: weighted estimates [an25w]
- Table D6. Weighted estimates of Constrained Extended Model of log(earnings) with covariates [an25w]

Appendix E (individuals aged 20–59 years, working full-time, and not in education; $|r_i – s_i| \leq 0.010$)

- Table E1. Unweighted estimates of four models of log(earnings) without covariates [an28]
- Table E2. Weighted estimates of four models of log(earnings) without covariates [an28w]
- Table E3. Goodness of fit statistics for four models of log(earnings) with covariates: unweighted estimates [an33]
- Table E4. Unweighted estimates of Constrained Extended Model of log(earnings) with covariates [an33]
- Table E5. Goodness of fit statistics for four models of log(earnings) with covariates: weighted estimates [an33w]
- Table E6. Weighted estimates of Constrained Extended Model of log(earnings) with covariates [an33w]

Appendix A1. Likelihood function for the mixture factor model

The likelihood function for our finite mixture factor model has the same structure as that of Kapteyn and Ypma's (2007, Appendix B) except that we have three additional distributions in the mixture.

Define group 1 to be the 'completely labelled' group for all observations for which r_i and s_i are counted as being equal, for each i = 1, …, n_1. The remaining observations, $i = n_1+1$, …, N, belong to one of the other eight distributions. The sample log-likelihood is:

$$\mathcal{L}(\theta, \Pi\ ) = \sum_{i=1}^{n_1} \log\big(\pi_1 f_1(r_i, s_i|\theta)\big) + \sum_{i=n_1+1}^{N} \log\left(\sum_{j=2}^{8} \pi_j f_j(r_i, s_i|\theta)\right) \qquad \text{(A1)}$$

where θ is the vector of parameters characterizing the distributions for each latent class, and $\Pi' = \{\pi_r, \pi_s, \pi_\upsilon, \pi_\omega\}$ is the vector of parameters underlying the latent class probabilities, π_j, with j = 1, …, 9. The expressions for the latent class distributions $f_j(r_i,s_i|\,\theta)$, are shown in Table A1 overleaf.

Table A1. Expressions for distributions of survey and administrative earnings, by latent class (group)

Group, j	Label	Probability, π_j	Group (latent class) distributions, $f_j(r_i, s_i \vert \theta)$
1	R1,S1	$\pi_\nu \pi_s$	$N\left(\begin{pmatrix}\mu_\xi \\ \mu_\xi\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & 1 \\ 1 & \sigma_\xi^2\end{pmatrix}\right)$
2	R1,S2	$\pi_r \pi_\nu (1-\pi_s)(1-\pi_\omega)$	$N\left(\begin{pmatrix}\mu_\xi \\ \mu_\xi + \mu_\eta\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & (1+\rho_s)\sigma_\xi^2 \\ (1+\rho_s)\sigma_\xi^2 & (1+\rho_s)^2\sigma_\xi^2 + \sigma_\eta^2\end{pmatrix}\right)$
3	R1,S3	$\pi_r \pi_\nu (1-\pi_s)\pi_\omega$	$N\left(\begin{pmatrix}\mu_\xi \\ \mu_\xi + \mu_\eta + \mu_\omega\end{pmatrix}, \begin{pmatrix}\sigma_\xi^2 & (1+\rho_s)\sigma_\xi^2 + \rho_\omega\sigma_\xi\sigma_\omega \\ (1+\rho_s)\sigma_\xi^2 + \rho_\omega\sigma_\xi\sigma_\omega & (1+\rho_s)^2\sigma_\xi^2 + \sigma_\eta^2 + \sigma_\omega^2 + 2\rho_\omega\sigma_\xi\sigma_\omega\end{pmatrix}\right)$
4	R2,S1	$\pi_r (1-\pi_\nu)\pi_s$	$N\left(\begin{pmatrix}\mu_\xi + \mu_\nu \\ \mu_\xi\end{pmatrix}, \begin{pmatrix}(1+\rho_r)^2\sigma_\xi^2 + \sigma_\nu^2 & (1+\rho_r)\sigma_\xi^2 \\ (1+\rho_r)\sigma_\xi^2 & \sigma_\xi^2\end{pmatrix}\right)$
5	R2,S2	$\pi_r (1-\pi_\nu)(1-\pi_s)(1-\pi_\omega)$	$N\left(\begin{pmatrix}\mu_\xi + \mu_\nu \\ \mu_\xi + \mu_\eta\end{pmatrix}, \begin{pmatrix}(1+\rho_r)^2\sigma_\xi^2 + \sigma_\nu^2 & (1+\rho_r)(1+\rho_s)\sigma_\xi^2 \\ (1+\rho_r)(1+\rho_s)\sigma_\xi^2 & (1+\rho_s)^2\sigma_\xi^2 + \sigma_\eta^2\end{pmatrix}\right)$
6	R2,S3	$\pi_r (1-\pi_\nu)(1-\pi_s)\pi_\omega$	$N\left(\begin{pmatrix}\mu_\xi + \mu_\nu \\ \mu_\xi + \mu_\eta + \mu_\omega\end{pmatrix}, \begin{pmatrix}(1+\rho_r)\sigma_\xi^2 + \sigma_\nu^2 & (1+\rho_r)(1+\rho_s)\sigma_\xi^2 + (1+\rho_r)\rho_{\xi\omega}\sigma_\xi\sigma_\omega \\ (1+\rho_r)(1+\rho_s)\sigma_\xi^2 + (1+\rho_r)\rho_{\xi\omega}\sigma_\xi\sigma_\omega & (1+\rho_s)^2\sigma_\xi^2 + \sigma_\eta^2 + \sigma_\omega^2 + 2\rho_{\xi\omega}\sigma_\xi\sigma_\omega\end{pmatrix}\right)$
7	R3,S1	$(1-\pi_r)\pi_s$	$N\left(\begin{pmatrix}\mu_\zeta \\ \mu_\xi\end{pmatrix}, \begin{pmatrix}\sigma_\zeta^2 & 0 \\ 0 & \sigma_\varepsilon^2\end{pmatrix}\right)$
8	R3,S2	$(1-\pi_r)(1-\pi_s)(1-\pi_\omega)$	$N\left(\begin{pmatrix}\mu_\zeta \\ \mu_\xi + \mu_\eta\end{pmatrix}, \begin{pmatrix}\sigma_\zeta^2 & 0 \\ 0 & (1+\rho_s)^2\sigma_\varepsilon^2 + \sigma_\eta^2\end{pmatrix}\right)$
9	R3,S3	$(1-\pi_r)(1-\pi_s)\pi_\omega$	$N\left(\begin{pmatrix}\mu_\zeta \\ \mu_\xi + \mu_\eta + \mu_\omega\end{pmatrix}, \begin{pmatrix}\sigma_\zeta^2 & 0 \\ 0 & (1+\rho_s)^2\sigma_\xi^2 + \sigma_\eta^2 + \sigma_\omega^2 + 2\rho_{\xi\omega}\sigma_\xi\sigma_\omega\end{pmatrix}\right)$

Note: $N(.,.)$ is the bivariate normal distribution with specified means and (co)variances.

Appendix A2: Inverse probability weights

This appendix describes how we constructed a set of composite sample weights in order to address sample selection issues. Our analysis samples are of employed FRS respondents who both consented to data linkage and whose FRS and P14 records were successfully linked. These samples may be unrepresentative of all employees.[31]

In many other countries outside the UK, including the USA, securing respondent consent is not an issue when constructing linked datasets, because the legal frameworks differ. (The UK's legal framework has recently changed, but linked data for this recent period are not publicly available.) However, other countries face similar issues to ours regarding linkage rates per se. For example, US studies such as Bound and Krueger (1991), Bollinger (1998), and Gottschalk and Huynh (2010) rely on respondents volunteering their Social Security Number and those that do not (or provide unusable data) are dropped from the analysis sample. Around 50% of respondents were dropped in the first two studies, based on the same CPS data, and around 30% in the third study, based on the SIPP. In Bollinger et al.'s (2018) CPS-based study, around 10% of respondents were lost at the linking stage. In Abowd and Stinson's study based on multiple SIPP panels, between 17% and 25% of SIPP respondents did not also appear in the administrative data (2013, Appendix Table C1).

Following Bollinger et al. (2019, Appendix A.4), we investigated selection issues by constructing inverse-probability weights. We regressed the probability of the binary outcome, 'consent to data linkage and successful linkage', on a large number of individual characteristics using a probit model applied to the FRS sample of employed respondents, and derived weights equal to the inverse of the predicted probabilities. The estimates from the regression are shown below in Table A2; we discuss them shortly.

We multiplied the inverse probability weights by the FRS individual sample weight to create a new composite weight. When we compared the unweighted and composite-weighted estimates of our leading measurement error models, we found that the estimates were very similar: see Appendix B–E. Hence, for brevity, we report only unweighted estimates in the main text. If we were to trim the small number of very large weights – trimming weights is a common practice in survey statistics – we would expect the unweighted and weighted estimates to be even more similar.

[31] There may also be selection into employment; we do not consider that here. Jenkins et al. (2006) find, using a multivariate probit model with incidental truncations applied to a UK general social survey, that unobservable factors associated with employment and consent to DWP record linkage were not significantly correlated.

The inverse probability weighting approach is reliant on the assumption that unobserved differences between individuals are irrelevant, an assumption that we cannot check. As Bollinger et al. (2019) point out, the assumption is more tenable for the linkage success component of the binary outcome than the consent to linkage component, because the former process is more of a technical issue related to record matching and hence more likely to be unrelated to respondent characteristics, observed or unobserved.

The probit model estimates of the correlates of the probability of the binary outcome, consent to data linkage and successful linkage are shown in Table A2. The first pair of columns shows the probit regression coefficients and their standard errors; the second pair shows the corresponding marginal effects (MEs) and their standard errors; and the final column shows covariate means. When choosing the set of covariates and covariate specifications, we drew on UK studies of consent to DWP record linkage by Jenkins et al. (2006) and McKay (2012). The sample mean of the outcome variable is 0.46. Given this reference point, we take an ME estimate of greater than 5 percentage points (ppt) as being substantively relevant.

The estimates indicate selection on observables. Consent and linkage rates are associated with age, being around 10–15 percentage points (ppt) higher for individuals aged 50+ (around 30% of the sample) compared to younger individuals. (The rate is even higher for those aged 75+, but there are very few such individuals in employment.) Consent and linkage rates are not associated with the respondent's sex, whether s/he has an activity-limiting health problem, whether there are 4+ adults in the household, or s/he has pays child support to a former partner via the DWP or the Child Support Agency. Other contact with the DWP, as a cash benefit recipient, is associated with a small increase in the consent and linkage rate. Variations in the rate by educational qualification level and marital status are also relatively small. Having dependent children younger than 11 years is associated with a rate some 7 ppt greater than that for childless individuals.

The largest variations are by ethnic group, housing tenure, and especially region of residence. Asian, Black, and other groups, which together comprise around 7% of the sample, have consent and linkage rates that around 10 ppt lower than White respondents. Respondents living in privately-rented accommodation (16% of the sample), have rates around 16 ppt lower than those in other tenures. Consent and linkage rates vary substantially more by region. Respondents in the North East and Yorkshire and Humberside have rates around 15 ppt higher than the reference region, London, whereas rates are some 11 ppt lower in Northern Ireland.

In sum, our inverse-probability weights mainly adjust for differences in consent and linkage rates by age, housing tenure, ethnic minority group, and region, with the largest variation in weights by age and region. This may help explain why our unweighted and weighted estimates vary little. Our models with covariates include basic controls for age. And the covariates that we do find associated with variations in model parameters, for example, prevalence of payslip presentation to the interviewer, are likely uncorrelated with region.

Additional references

Jenkins, S. P., Cappellari, L., Lynn, P., Jäckle, A., and Sala, E. (2006). Patterns of consent: evidence from a general household survey', Journal of the Royal Statistical Society, Series A, 169 (4), 701–722.

McKay, S. (2012). Evaluating approaches to Family Resources Survey data linking. Working Paper 110. London: Department for Work and Pensions.

Appendix Table A2. Probability of the binary outcome 'consent and successful data linkage': probit regression coefficients (Coeff.) and marginal effects (ME)

	Coeff.		(SE)	ME		(SE)	Mean
Female	0.0246		(0.0197)	0.0093		(0.0075)	0.55
Age (ref.: 16–19 years)							
20–24	–0.0772		(0.1141)	–0.0289		(0.0429)	0.06
25–29	–0.0729		(0.1140)	–0.0273		(0.0429)	0.10
30–34	–0.0357		(0.1136)	–0.0134		(0.0428)	0.12
35–39	0.0148		(0.1135)	0.0056		(0.0428)	0.12
40–44	0.1123		(0.1117)	0.0427		(0.0422)	0.14
45–49	0.1720		(0.1120)	0.0657		(0.0423)	0.14
50–54	0.2667	*	(0.1131)	0.1021	*	(0.0427)	0.12
55–59	0.3129	**	(0.1141)	0.1198	**	(0.0431)	0.10
60–64	0.3484	**	(0.1182)	0.1334	**	(0.0446)	0.06
65–69	0.3838	**	(0.1388)	0.1468	**	(0.0525)	0.02
70–74	0.4321	*	(0.1803)	0.1651	*	(0.0678)	0.01
75+	0.6430	*	(0.2953)	0.2421	*	(0.1051)	0.00
Educational qualifications (ref.: A-level(s) or higher)							0.50
Some ed. quals. below A-levels	0.0688	**	(0.0252)	0.0261	**	(0.0096)	0.39
No formal quals.	0.1176		(0.0930)	0.0447		(0.0354)	0.02
Missing ed. quals.	–0.0727		(0.0447)	–0.0274		(0.0168)	0.09
Marital status is single, widowed, divorced, or separated	0.0857	**	(0.0277)	0.0325	**	(0.0105)	0.30
Has 1+ dependent children aged …							
0–4 years	0.1674	***	(0.0371)	0.0634	***	(0.0140)	0.16
5–10 years	0.2002	***	(0.0363)	0.0761	***	(0.0138)	0.17
11–15 years	0.0905	*	(0.0382)	0.0344	*	(0.0145)	0.15
16–18 years	0.1312	**	(0.0487)	0.0499	**	(0.0185)	0.08
4+ adults in household	0.0528		(0.0555)	0.0201		(0.0211)	0.07
Health limits activities	0.0597		(0.0391)	0.0227		(0.0149)	0.09
Region (ref.: London)							0.08
North East	0.4194	***	(0.0767)	0.1607	***	(0.0292)	0.04
North West	0.3060	***	(0.0588)	0.1171	***	(0.0223)	0.10
Yorkshire & Humberside	0.3718	***	(0.0614)	0.1425	***	(0.0233)	0.08
East Midlands	0.2409	***	(0.0643)	0.0919	***	(0.0245)	0.07
West Midlands	0.2956	***	(0.0625)	0.1131	***	(0.0238)	0.07
East of England	0.2368	***	(0.0598)	0.0903	***	(0.0227)	0.09
South East	0.1061		(0.0566)	0.0401		(0.0213)	0.12
South West	0.1775	**	(0.0638)	0.0674	**	(0.0242)	0.07
Wales	0.1301		(0.0765)	0.0493		(0.0291)	0.04
Scotland	0.2741	***	(0.0539)	0.1047	***	(0.0204)	0.16
Northern Ireland	–0.2997	***	(0.0646)	–0.1062	***	(0.0227)	0.08
Ethnic group (ref.: white)							0.84
Mixed or multiple	–0.0721		(0.1191)	–0.0274		(0.0451)	0.01
Asian or Asian British	–0.2432	***	(0.0643)	–0.0911	***	(0.0235)	0.04
Black, African, Caribbean, Black British	–0.2285	**	(0.0841)	–0.0858	**	(0.0309)	0.02
Other ethnic group	–0.3553	**	(0.1192)	–0.1313	**	(0.0420)	0.01
Receiving 1+ cash benefits	0.1137	*	(0.0476)	0.0432	*	(0.0181)	0.07
Paying child support via DWP or Child Support Agency	–0.0572		(0.1457)	–0.0216		(0.0549)	0.01
Housing tenure (ref: social housing)							0.11
Renting privately	–0.4165	***	(0.0502)	–0.1576	***	(0.0190)	0.16
Owned: outright or with mortgage	–0.0848		(0.0440)	–0.0327		(0.0170)	0.72
Other: e.g. rent-free	–0.1966		(0.1317)	–0.0756		(0.0503)	0.01
Constant	–0.4109	**	(0.1285)				

Probit regression. Cluster-robust standard errors in parentheses (cluster is household). Number of individuals: 13,787. Number of households = 10,443. Log-pseudolikelihood = –9112.8. Pseudo R^2 = 0.0430. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. ME: marginal effects (for categorical variables: discrete change relative to reference category). Mean of dependent variable: 0.465.

Appendix Table A3: Means of covariates, by sample

Characteristics	All individuals		Aged 25–59, in full-time work, not in education	
	Unweighted	Weighted	Unweighted	Weighted
	(1)	(2)	(3)	(4)
Male	0.43	0.46	0.55	0.56
Female	0.57	0.54	0.45	0.44
Age (years)	44	41	43	42
Aged < 60 years	0.90	0.92	1.00	1.00
Aged 60+ years	0.10	0.08	0.00	0.00
Education: less than A-level	0.53	0.50	0.49	0.47
Education: A-level or more	0.47	0.50	0.51	0.53
Employed full-time	0.71	0.74	1.00	1.00
Employed part-time	0.29	0.26	0.00	0.00
Married, cohabiting	0.70	0.70	0.73	0.73
Single, divorced, separated, widowed	0.30	0.30	0.27	0.27
Has 1 job	0.96	0.96	0.98	0.98
Has 2+ jobs	0.04	0.04	0.02	0.02
Payslip(s) shown (all jobs)	0.65	0.66	0.67	0.68
Payslip not provided by employer (any job)	0.10	0.10	0.09	0.09
Pay reference period:				
1 week	0.17	0.16	0.16	0.16
4 weeks or 1 calendar month	0.77	0.77	0.77	0.78
1 year, 12 months, or 52 weeks	0.04	0.04	0.05	0.05
Other	0.02	0.02	0.01	0.01
Job spells in P14 data do not span 2011/12	0.39	0.39	0.32	0.33
Job spells in P14 data all span 2011/12	0.61	0.61	0.68	0.67
Private sector employee	0.69	0.72	0.69	0.74
Public sector employee	0.31	0.28	0.31	0.26

Notes. Weighted estimates calculated using the composite weight described in Appendix A2. Sample in columns (1), (2): all individuals, unweighted *N* individuals = 5,971 within 4,874 households. Sample in columns (3), (4): individuals aged 25–59 years, in full-time work, and not in any form of education, unweighted *N* individuals = 3,564 within 3,151 households.

Appendix B: All individuals ($|r_i - s_i| \le 0.005$)

Table B1. Weighted estimates for four models of log(earnings) without covariates: all individuals ($|r_i - s_i| \le 0.005$)

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	9.9053***	9.9051***	9.9043***	9.9072***
	(0.0134)	(0.0133)	(0.0132)	(0.0130)
σ_ξ	0.7060***	0.7065***	0.7310***	0.7257***
	(0.0104)	(0.0103)	(0.0118)	(0.0111)
μ_ζ	8.5652***	8.5092***	8.4563***	8.7279***
	(0.2700)	(0.2773)	(0.2803)	(0.1559)
σ_ζ	1.3757***	1.3652***	1.3134***	1.3542***
	(0.1331)	(0.1402)	(0.1563)	(0.1052)
μ_ω	0.1835	0.1720	0.0667	0.0668
	(0.1691)	(0.1485)	(0.0553)	(0.0632)
σ_ω	1.2836***	1.2833***	0.8171***	0.8014***
	(0.1225)	(0.1215)	(0.1182)	(0.1384)
μ_η	–0.0094*	–0.0094*	–0.0125***	–0.0125***
	(0.0043)	(0.0044)	(0.0032)	(0.0032)
σ_η	0.0929***	0.0929***	0.1268***	0.1271***
	(0.0121)	(0.0121)	(0.0063)	(0.0069)
μ_υ	–0.0162	–0.0162		
	(0.0289)	(0.0289)		
σ_υ	0.3263***	0.3261***		
	(0.0512)	(0.0514)		
ρ_s	0.0003	0.0004	–0.0252***	–0.0269***
	(0.0062)	(0.0062)	(0.0055)	(0.0053)
ρ_r	0.0288	0.0287		
	(0.0352)	(0.0352)		
$\rho_{\xi\omega}$	0.0604		–0.1242**	
	(0.1234)		(0.0473)	
π_s	0.0518***	0.0517***	0.0344***	0.0347***
	(0.0073)	(0.0073)	(0.0029)	(0.0029)
π_ω	0.0613***	0.0629***	0.1870***	0.1765***
	(0.0111)	(0.0113)	(0.0232)	(0.0251)
π_r	0.9677***	0.9692***	0.9639***	0.9527***
	(0.0096)	(0.0094)	(0.0082)	(0.0070)
π_υ	0.6613***	0.6612***		
	(0.0711)	(0.0713)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0331***	0.0331***	0.0331***	0.0331***
	(0.0028)	(0.0028)	(0.0028)	(0.0028)
π_2 (π_2)	0.5696***	0.5695***	0.7567***	0.7574***
	(0.0696)	(0.0699)	(0.0208)	(0.0232)
π_3 (π_3)	0.0372***	0.0382***	0.1740***	0.1623***

	(0.0080)	(0.0081)	(0.0221)	(0.0233)
π_4	0.0170**	0.0170**		
	(0.0055)	(0.0056)		
π_5	0.2918***	0.2918***		
	(0.0573)	(0.0574)		
π_6	0.0190***	0.0196***		
	(0.0052)	(0.0054)		
π_7 (π_4)	0.0017*	0.0016*	0.0012***	0.0016***
	(0.0007)	(0.0007)	(0.0003)	(0.0003)
π_8 (π_5)	0.0287***	0.0274***	0.0284***	0.0376***
	(0.0085)	(0.0083)	(0.0067)	(0.0059)
π_9 (π_6)	0.0019**	0.0018**	0.0065***	0.0080***
	(0.0006)	(0.0006)	(0.0015)	(0.0015)
Log(pseudo-likelihood)	–6177.3	–6177.6	–6371.4	–6375.2
AIC	12388.6	12387.2	12768.9	12774.3
BIC	12498.7	12490.7	12853.0	12852.0
Reliability1 (*r*)	0.7567	0.7592	0.7919	0.7708
Reliability1 (*s*)	0.8214	0.8219	0.8202	0.8204
Reliability2 (*r*)	0.7395	0.7430	0.7633	0.7343
Reliability2 (*s*)	0.8269	0.8222	0.7797	0.7991

Notes. Cluster-robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 5,971; unweighted number of households = 4,874. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Weighted estimates, with weights equal to product of FRS individual weight and inverse-probability of having consented to data linkage and been successfully linked (see text). Weights normalized so that sum equals sample number of individuals. The corresponding unweighted estimates are shown in Table 2 of the main text. [an14w]

Table B2. Goodness of fit statistics for four models of log(earnings) with covariates: all individuals ($|r_i - s_i| \leq 0.005$), weighted estimates

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo-likelihood	–6261.7	–6262.1	–6631.6	–6636.0
AIC	12625.5	12624.2	13341.3	13347.9
BIC	12966.9	12959.0	13602.4	13602.3
Reliability1 (r)	0.7347	0.7333	0.7078	0.6940
Reliability1 (s)	0.8059	0.8069	0.8012	0.8053
Reliability2 (r)	0.7020	0.6999	0.6450	0.6256
Reliability2 (s)	0.8197	0.8243	0.7807	0.7984

Notes. Cluster-robust standard errors in parentheses (cluster is household). Sample: all individuals (number of individuals = 5,971 within 4,874 households). Model based on a completely-labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$: see main text). Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an22w]

Table B3. Weighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, all individuals ($|r_i - s_i| \leq 0.005$)

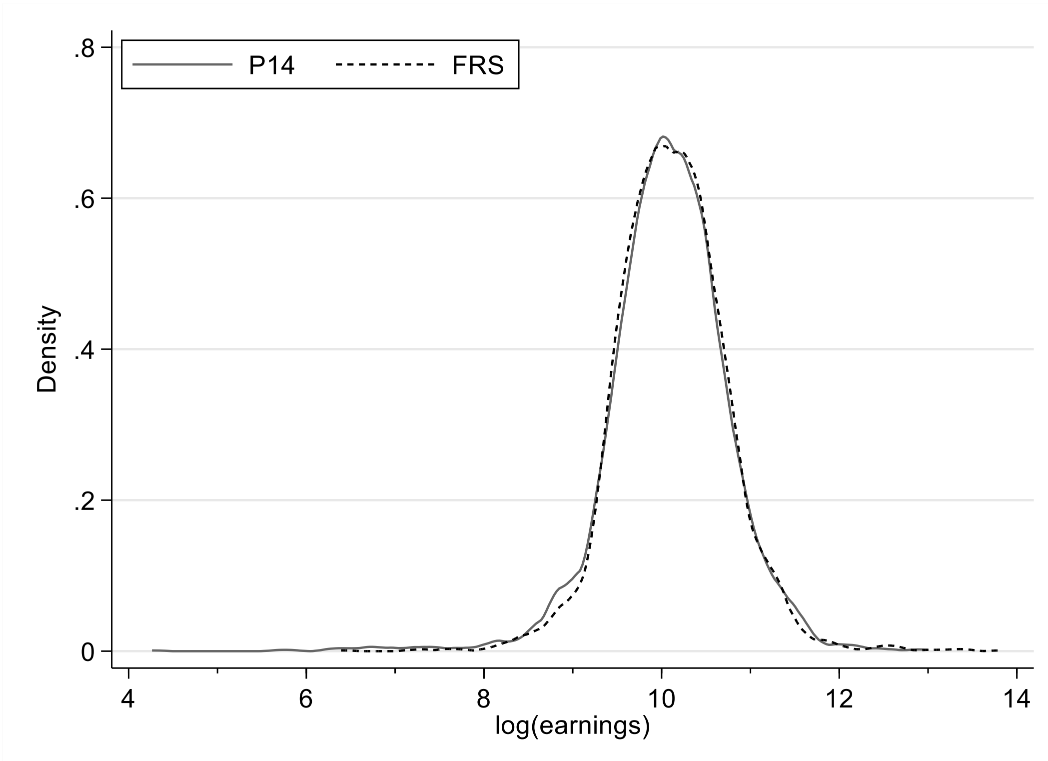
		MM	(SE)		MM	(SE)
Male	μ_ξ	9.9414***	(0.0147)	σ_ξ	0.5185***	(0.0128)
Female		9.7128***	(0.0110)		0.4692***	(0.0086)
Education: less than A-level		9.6322***	(0.0117)		0.4368***	(0.0103)
Education: A-level or more		10.0018***	(0.0142)		0.5470***	(0.0108)
Full-time employee		10.0413***	(0.0099)		0.4456***	(0.0080)
Part-time employee		9.1748***	(0.0221)		0.6277***	(0.0179)
Married, cohabiting		9.8474***	(0.0108)			
Single, divorced, separated, widowed		9.7466***	(0.0165)			
Has 1 job		9.8131***	(0.0098)			
Has 2+ jobs		9.9252***	(0.0466)			
Age = 25 years		9.5643***	(0.0171)		0.4112***	(0.0127)
Age = 35 years		9.8309***	(0.0119)		0.4797***	(0.0089)
Age = 45 years		9.9574***	(0.0123)		0.5263***	(0.0100)
Age = 55 years		9.9439***	(0.0137)		0.5430***	(0.0124)
	μ_ζ	8.9747***	(0.1131)	σ_ζ	1.2290***	(0.0869)
Payslip(s) not shown to interviewer	μ_η	–0.0455***	(0.0078)	σ_η	0.1375***	(0.0103)
Payslip(s) shown (all jobs)		–0.0033	(0.0042)		0.0764***	(0.0052)
Aged < 60 years		–0.0111**	(0.0043)		0.0895***	(0.0058)
Aged 60+ years		–0.0992***	(0.0191)		0.1938***	(0.0261)
Reference period: not 'other'	μ_ω	0.0792	(0.1061)	σ_ω	1.0626***	(0.1119)
Reference period: other		–0.5399***	(0.1044)		0.4011***	(0.2809)
Job spells do not span year		–0.1855*	(0.0854)		0.8729***	(0.1049)
Job spells all span year		0.2266	(0.1599)		1.1584***	(0.1458)
Aged < 60 years		0.1128	(0.1055)		1.0546***	(0.1077)
Aged 60+ years		–0.5128**	(0.1821)		0.9694**	(0.3683)
Full-time employee		0.1806	(0.1266)		1.0928***	(0.1056)
Part-time employee		–0.2650	(0.1626)		0.9163***	(0.2440)
Payslip provided by employer	μ_υ	0.0064	(0.0134)	σ_υ	0.2284***	(0.0308)
Payslip not provided by employer		–0.1977*	(0.0845)		0.4796***	(0.0804)
Full-time employee		–0.0118	(0.0174)		0.1896***	(0.0336)
Part-time employee		–0.0164	(0.0304)		0.4382***	(0.0402)
Private sector employee		0.0086	(0.0187)		0.2727***	(0.0329)
Public sector employee		–0.0691***	(0.0173)		0.2082***	(0.0413)
Male	ρ_s	0.0061	(0.0113)	ρ_r	0.0496*	(0.0234)
Female		–0.0045	(0.0085)		0.2647	(0.2140)
Aged < 60 years		0.0056	(0.0078)			
Aged 60+ years		–0.0635	(0.0380)			
Private sector employee					0.1174**	(0.0388)
Public sector employee					–0.0523	(0.0393)
Probabilities	π_s	0.0646***	(0.0089)	π_1	0.0332***	(0.0026)
	π_ω	0.0856***	(0.0126)	π_2	0.4399***	(0.0545)
	π_r	0.9391***	(0.0099)	π_3	0.0412***	(0.0064)
	π_υ	0.5477***	(0.0567)	π_4	0.0274***	(0.0067)
				π_5	0.3633***	(0.0388)
				π_6	0.0340***	(0.0072)
				π_7	0.0039***	(0.0011)

π_8	0.0521^{***}	(0.0081)
π_9	0.0049^{***}	(0.0011)

Notes. Cluster-robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 5,971; unweighted number of households = 4,874. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Weighted estimates, with weights equal to product of FRS individual weight and inverse-probability of having consented to data linkage and been successfully linked (see text). Weights normalized so that sum equals sample number of individuals. Log(pseudo-likelihood) = –6262.1. AIC = 12624.2. BIC = 12958.9. Reliability1 (r) = 0.7333. Reliability1 (s) = 0.8069. Reliability2 (r) = 0.6999. Reliability1 (s) = 0.8243. [an22w]

Appendix C: Individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$)

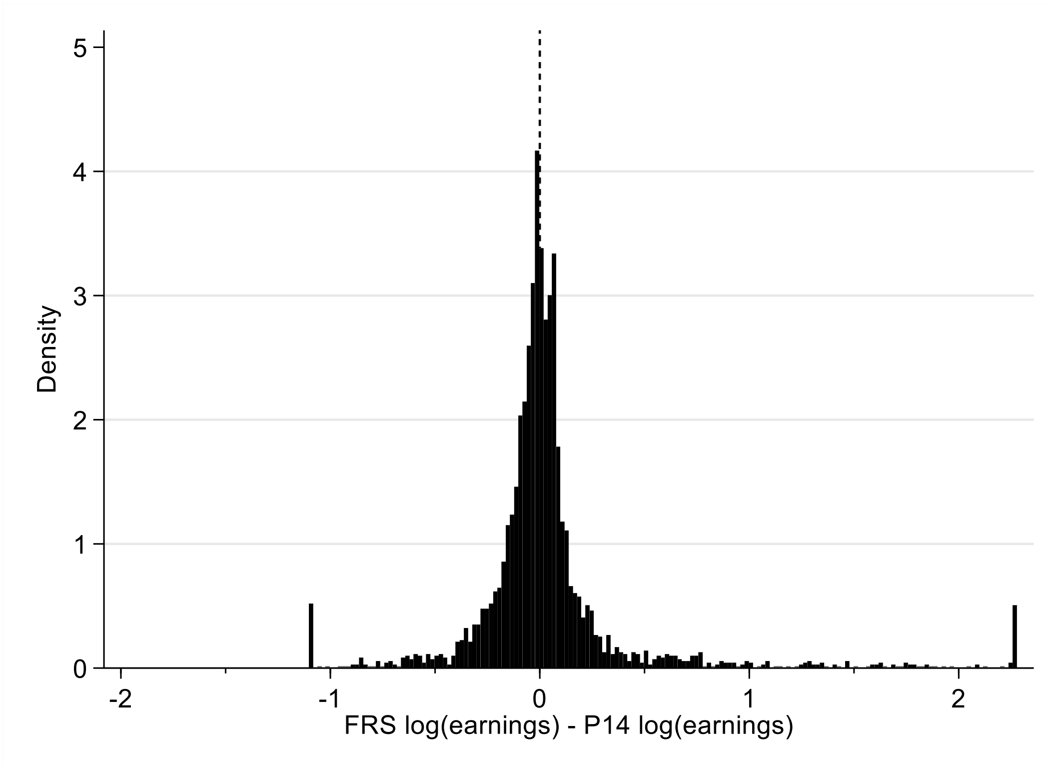
Figure C1. Distributions of FRS and P14 log(earnings)



Notes. Kernel density estimates (Epanenikov kernel, 'optimal' bandwidth). Summary statistics for (s, r): mean (10.12, 10.08); p5 (9.22, 9.00); p10 (9.42, 9.35); p50 (10.10, 10.09); p90 (10.83, 10.85); p95 (11.12, 11.13); standard deviation (0.62, 0.69). Sample: all individuals in subsample (N = 3, 564).

Appendix C: Individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$)

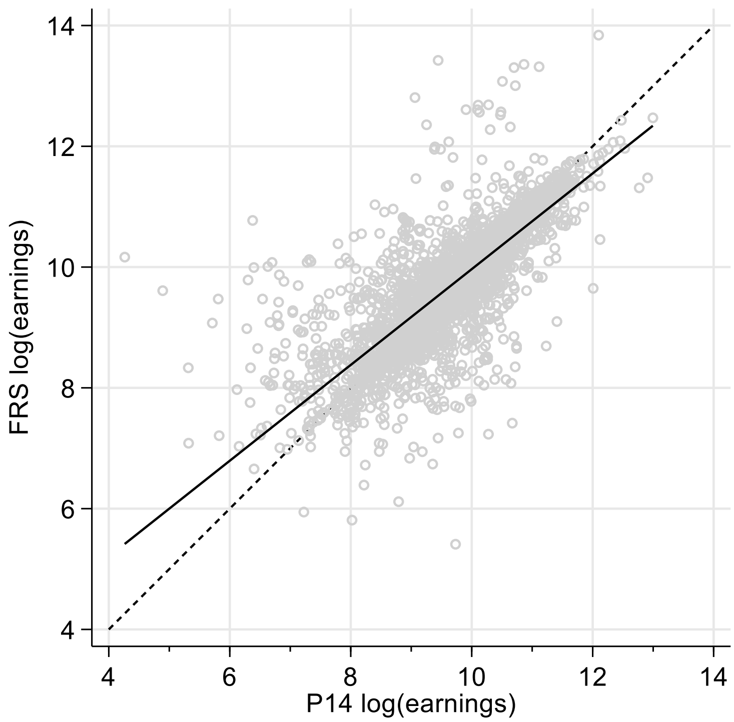
Figure C2. Distribution of difference between FRS and P14 earnings ($s - r$)



Notes. Histogram with bin width = 0.02. Earnings differences are bottom-coded at $p1$ (–1.10) and top-coded at $p99$ (2.26) for purposes of presentation. Summary statistics for $s - r$ (without bottom- or top-coding): mean, 0.01; standard deviation, 0.468; $p5$, –0.364; $p10$, –0.235; $p50$, –0.004; $p90$, 0.255; $p95$, 0.640. Sample: all individuals in subsample (N = 3, 564).

Appendix C: Individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$)

Figure C3. The relationship between FRS and P14 earnings



Notes. The solid line represents a linear regression with a slope coefficient of 0.673 (SE 0.010). Sample: all individuals in subsample (N = 3, 564).

Appendix C: Individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$)

Table C1. Unweighted estimates for four models of log(earnings) without covariates: individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$)

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	10.1206***	10.1205***	10.1288***	10.1287***
	(0.0102)	(0.0102)	(0.0100)	(0.0100)
σ_ξ	0.5451***	0.5454***	0.5604***	0.5606***
	(0.0083)	(0.0083)	(0.0090)	(0.0088)
μ_ζ	8.9196***	8.9094***	9.3915***	9.3850***
	(0.1682)	(0.1689)	(0.1290)	(0.1188)
σ_ζ	1.3400***	1.3402***	1.3124***	1.3146***
	(0.1079)	(0.1086)	(0.0879)	(0.0852)
μ_ω	0.0446	0.0391	0.0115	0.0114
	(0.1108)	(0.1091)	(0.0590)	(0.0584)
σ_ω	1.2708***	1.2754***	0.9094***	0.9088***
	(0.1092)	(0.1057)	(0.1306)	(0.1304)
μ_η	–0.0044	–0.0044	–0.0118***	–0.0118***
	(0.0035)	(0.0035)	(0.0027)	(0.0027)
σ_η	0.0863***	0.0863***	0.1229***	0.1229***
	(0.0066)	(0.0066)	(0.0044)	(0.0044)
μ_υ	0.0148	0.0148		
	(0.0187)	(0.0186)		
σ_υ	0.2821***	0.2820***		
	(0.0379)	(0.0379)		
ρ_s	–0.0029	–0.0028	–0.0469***	–0.0468***
	(0.0082)	(0.0082)	(0.0056)	(0.0056)
ρ_r	0.0551	0.0544		
	(0.0284)	(0.0282)		
$\rho_{\xi\omega}$	0.0494		0.0138	
	(0.0949)		(0.0744)	
π_s	0.0566***	0.0566***	0.0396***	0.0396***
	(0.0065)	(0.0065)	(0.0034)	(0.0034)
π_ω	0.0489***	0.0492***	0.0985***	0.0994***
	(0.0072)	(0.0074)	(0.0225)	(0.0212)
π_r	0.9539***	0.9544***	0.9195***	0.9202***
	(0.0082)	(0.0081)	(0.0121)	(0.0105)
π_υ	0.6751***	0.6748***		
	(0.0488)	(0.0487)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0364***	0.0364***	0.0364***	0.0364***
	(0.0031)	(0.0031)	(0.0031)	(0.0031)
π_2 (π_2)	0.5779***	0.5777***	0.7961***	0.7960***
	(0.0490)	(0.0489)	(0.0138)	(0.0138)
π_3 (π_3)	0.0297***	0.0299***	0.0870***	0.0879***

	(0.0046)	(0.0047)	(0.0208)	(0.0196)
π_4	0.0175***	0.0176***		
	(0.0042)	(0.0042)		
π_5	0.2781***	0.2784***		
	(0.0386)	(0.0385)		
π_6	0.0143***	0.0144***		
	(0.0032)	(0.0033)		
π_7 (π_4)	0.0026***	0.0026***	0.0032***	0.0032***
	(0.0007)	(0.0007)	(0.0006)	(0.0005)
π_8 (π_5)	0.0413***	0.0409***	0.0697***	0.0690***
	(0.0072)	(0.0071)	(0.0118)	(0.0103)
π_9 (π_6)	0.0021***	0.0021***	0.0076***	0.0076***
	(0.0005)	(0.0005)	(0.0011)	(0.0011)
Log(pseudo-likelihood)	–3541.3	–3541.4	–3693.0	–3693.0
AIC	7116.5	7114.8	7411.9	7409.9
BIC	7221.5	7213.7	7492.2	7484.1
Reliability1 (*r*)	0.6202	0.6210	0.6174	0.6184
Reliability1 (*s*)	0.7834	0.7837	0.7900	0.7897
Reliability2 (*r*)	0.6022	0.6031	0.5676	0.5691
Reliability2 (*s*)	0.7854	0.7816	0.7561	0.7542

Notes. Cluster-robust standard errors in parentheses (cluster is household). Number of individuals = 3,564. Number of households = 3,151. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an27]

Table C2. Weighted estimates for four models of log(earnings) without covariates: individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	10.1290***	10.1288***	10.1339***	10.1336***
	(0.0135)	(0.0135)	(0.0124)	(0.0124)
σ_ξ	0.5511***	0.5519***	0.5638***	0.5647***
	(0.0104)	(0.0104)	(0.0101)	(0.0100)
μ_ζ	9.0078***	8.9776***	9.5042***	9.4803***
	(0.1947)	(0.1980)	(0.1185)	(0.1149)
σ_ζ	1.3361***	1.3374***	1.2597***	1.2695***
	(0.1424)	(0.1438)	(0.0991)	(0.0999)
μ_ω	0.2668	0.2485	0.1902	0.1799
	(0.1947)	(0.1811)	(0.1554)	(0.1390)
σ_ω	1.3178***	1.3261***	1.0867***	1.0823***
	(0.1317)	(0.1326)	(0.1818)	(0.1801)
μ_η	–0.0075	–0.0075	–0.0141***	–0.0141***
	(0.0053)	(0.0053)	(0.0035)	(0.0035)
σ_η	0.0887***	0.0886***	0.1287***	0.1286***
	(0.0134)	(0.0133)	(0.0047)	(0.0048)
μ_υ	0.0019	0.0020		
	(0.0339)	(0.0336)		
σ_υ	0.2790***	0.2789***		
	(0.0582)	(0.0579)		
ρ_s	–0.0003	–0.0002	–0.0420***	–0.0417***
	(0.0120)	(0.0119)	(0.0079)	(0.0079)
ρ_r	0.0553	0.0535		
	(0.0404)	(0.0398)		
$\rho_{\xi\omega}$	0.1206		0.0804	
	(0.1255)		(0.1074)	
π_s	0.0539***	0.0539***	0.0376***	0.0375***
	(0.0096)	(0.0096)	(0.0036)	(0.0036)
π_ω	0.0565***	0.0576***	0.0848***	0.0886***
	(0.0096)	(0.0101)	(0.0198)	(0.0193)
π_r	0.9547***	0.9561***	0.9111***	0.9146***
	(0.0120)	(0.0117)	(0.0142)	(0.0129)
π_υ	0.6663***	0.6656***		
	(0.0942)	(0.0937)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0343***	0.0343***	0.0343***	0.0343***
	(0.0033)	(0.0033)	(0.0033)	(0.0033)
π_2 (π_2)	0.5678***	0.5674***	0.8024***	0.8024***
	(0.0919)	(0.0915)	(0.0124)	(0.0126)
π_3 (π_3)	0.0340***	0.0347***	0.0744***	0.0780***
	(0.0073)	(0.0075)	(0.0183)	(0.0178)
π_4	0.0172*	0.0172*		

	(0.0074)	(0.0074)		
π_5	0.2844***	0.2851***		
	(0.0747)	(0.0743)		
π_6	0.0170**	0.0174**		
	(0.0057)	(0.0059)		
π_7 (π_4)	0.0024*	0.0024*	0.0033***	0.0032***
	(0.0010)	(0.0010)	(0.0007)	(0.0006)
π_8 (π_5)	0.0405***	0.0392***	0.0783***	0.0749***
	(0.0105)	(0.0102)	(0.0137)	(0.0124)
π_9 (π_6)	0.0024**	0.0024**	0.0073***	0.0073***
	(0.0008)	(0.0008)	(0.0012)	(0.0011)
Log(pseudo-likelihood)	–3761.3	–3762.2	–3907.0	–3907.6
AIC	7556.6	7556.4	7840.0	7839.1
BIC	7661.6	7655.3	7920.3	7913.3
Reliability1 (*r*)	0.6404	0.6428	0.6256	0.6305
Reliability1 (*s*)	0.7399	0.7412	0.7439	0.7427
Reliability2 (*r*)	0.6226	0.6256	0.5700	0.5766
Reliability2 (*s*)	0.7511	0.7411	0.7232	0.7129

Notes. Cluster-robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 3,564. Unweighted number of households = 3,151. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Weighted estimates, with weights equal to product of FRS individual weight and inverse-probability of having consented to data linkage and been successfully linked (see text). Weights normalized so that sum equals sample number of individuals. [an27w]

Table C3. Goodness of fit statistics for four models of log(earnings) with covariates: individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$), unweighted estimates

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo-likelihood	–2891.0	–2891.0	–3064.8	–3064.8
AIC	5857.9	5856.0	6185.6	6183.7
BIC	6092.7	6084.6	6358.6388	6350.5
Reliability1 (*r*)	0.5922	0.5925	0.6073	0.6067
Reliability1 (*s*)	0.7625	0.7622	0.7710	0.7714
Reliability2 (*r*)	0.5571	0.5576	0.5549	0.5541
Reliability2 (*s*)	0.7653	0.7636	0.7355	0.7373

Notes. Cluster-robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 3,564. Unweighted number of households = 3,151. Model based on a completely-labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$: see main text). Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an32]

Table C4. Unweighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, individuals aged 25–59, working full-time, not in education, ($|r_i - s_i| \leq 0.005$)

		MM	(SE)		MM	(SE)
Male	μ_ξ	10.2471***	(0.0117)	σ_ξ	0.4754***	(0.0107)
Female		9.9671***	(0.0121)		0.4392***	(0.0100)
Education: less than A-level		9.9010***	(0.0112)		0.4116***	(0.0094)
Education: A-level or more		10.3329***	(0.0134)		0.5058***	(0.0115)
Married, cohabiting		10.1401***	(0.0105)			
Single, divorced, separated, widowed		10.0662***	(0.0156)			
Has 1 job		10.1193***	(0.0092)			
Has 2+ jobs		10.1545***	(0.0546)			
Age = 25 years		9.8234***	(0.0244)		0.3621***	(0.0184)
Age = 35 years		10.0769***	(0.0117)		0.4443***	(0.0096)
Age = 45 years		10.1909***	(0.0123)		0.4853***	(0.0104)
Age = 55 years		10.1653***	(0.0152)		0.4719***	(0.0138)
	μ_ζ	9.2463***	(0.1151)	σ_ζ	1.3078***	(0.0863)
Payslip(s) not shown to interviewer	μ_η	–0.0330***	(0.0072)	σ_η	0.1030***	(0.0135)
Payslip(s) shown (all jobs)		0.0097*	(0.0042)		0.0626***	(0.0046)
Reference period: not ‘other’	μ_ω	#		σ_ω	#	
Reference period: other		#			#	
Job spells do not span year		–0.1073	(0.1097)		1.0369***	(0.1367)
Job spells all span year		0.1182	(0.1259)		1.2097***	(0.1431)
Payslip provided by employer	μ_υ	0.0281**	(0.0086)	σ_υ	0.1576***	(0.0233)
Payslip not provided by employer		–0.0803	(0.0410)		0.3114***	(0.0543)
Private sector employee		0.0455***	(0.0094)		0.1812***	(0.0220)
Public sector employee		–0.0459***	(0.0120)		0.1452***	(0.0364)
Male	ρ_s	0.0034	(0.0114)	ρ_r	0.0192	(0.0240)
Female		–0.0015	(0.0098)		0.1398	(0.0781)
Private sector employee					0.0732**	(0.0228)
Public sector employee					–0.0771	(0.0492)
Probabilities	π_s	0.0778***	(0.0125)	π_1	0.0362***	(0.0031)
	π_ω	0.0664***	(0.0121)	π_2	0.4006***	(0.0629)
	π_r	0.9321***	(0.0078)	π_3	0.0285***	(0.0044)
	π_υ	0.4992***	(0.0664)	π_4	0.0363***	(0.0101)
				π_5	0.4019***	(0.0442)
				π_6	0.0286***	(0.0077)
				π_7	0.0053***	(0.0012)
				π_8	0.0585***	(0.0064)
				π_9	0.0042***	(0.0009)

Notes. Unweighted number of individuals = 3,564. Unweighted number of households = 3,151. Log(pseudo-likelihood) = –2891.0. AIC = 5856.0. BIC = 6084.6. Reliability1 (r) = 0.5925. Reliability1 (s) = 0.7622. Reliability2 (r) = 0.5576. Reliability1 (s) = 0.7636. #: Excluded due to numerical instability. [an32]

Table C5. Goodness of fit statistics for four models of log(earnings) with covariates: individuals aged 25–59, in full-time work, not in education ($|r_i - s_i| \leq 0.005$), weighted estimates

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo–likelihood	–3126.9	–3126.9	–3290.3	–3290.5
AIC	6329.8	6327.9	6636.6	6635.1
BIC	6564.6	6556.5	6809.6111	6801.9
Reliability1 (r)	0.6140	0.6135	0.6270	0.6250
Reliability1 (s)	0.7029	0.7035	0.7116	0.7134
Reliability2 (r)	0.5843	0.5837	0.5740	0.5713
Reliability2 (s)	0.7068	0.7092	0.6803	0.6870

Notes. Cluster–robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 3,564. Unweighted number of households = 3,151. Model based on a completely–labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$: see main text). Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an32w]

Table C6. Weighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, individuals aged 25–59, working full–time, not in education, ($|r_i - s_i| \le 0.005$)

		MM	(SE)	MM	(SE)
Male	μ_ξ	10.2445***	(0.0158)	0.4888***	(0.0131)
Female		9.9836***	(0.0140)	0.4389***	(0.0108)
Education: less than A–level		9.9017***	(0.0148)	0.4197***	(0.0126)
Education: A–level or more		10.3327***	(0.0162)	0.5106***	(0.0127)
Married, cohabiting		10.1504***	(0.0126)		
Single, divorced, separated, widowed		10.0740***	(0.0221)		
Has 1 job		10.1296***	(0.0114)		
Has 2+ jobs		10.1515***	(0.0731)		
Age = 25 years		9.8631***	(0.0299)	0.3625***	(0.0194)
Age = 35 years		10.1154***	(0.0146)	0.4524***	(0.0113)
Age = 45 years		10.2192***	(0.0149)	0.5029***	(0.0120)
Age = 55 years		10.1746***	(0.0225)	0.4979***	(0.0194)
	μ_ζ	9.2341***	(0.1420)	1.2909***	(0.1109)
Payslip(s) not shown to interviewer	μ_η	–0.0386***	(0.0107)	0.1212***	(0.0130)
Payslip(s) shown (all jobs)		0.0033	(0.0049)	0.0668***	(0.0059)
Reference period: not ‘other’	μ_ω	#		#	
Reference period: other		#		#	
Job spells do not span year		–0.0630	(0.1069)	0.9769***	(0.1402)
Job spells all span year		0.4334	(0.2405)	1.3443***	(0.1486)
Payslip provided by employer	μ_υ	0.0197	(0.0116)	0.1649***	(0.0335)
Payslip not provided by employer		–0.1184	(0.0713)	0.3416*	(0.1329)
Private sector employee		0.0344*	(0.0138)	0.1943***	(0.0371)
Public sector employee		–0.0660***	(0.0182)	0.1409*	(0.0548)
Male	ρ_s	0.0141	(0.0130)	0.0383	(0.0267)
Female		0.0019	(0.0126)	0.1057	(0.1007)
Private sector employee				0.0825**	(0.0291)
Public sector employee				–0.0623	(0.0670)
Probabilities	π_s	0.0683***	(0.0140)	0.0342***	(0.0033)
	π_ω	0.0735***	(0.0146)	0.4320***	(0.0876)
	π_r	0.9377***	(0.0111)	0.0343***	(0.0065)
	π_υ	0.5337***	(0.0921)	0.0299***	(0.0114)
				0.3774***	(0.0636)
				0.0300***	(0.0099)
				0.0043***	(0.0015)
				0.0538***	(0.0090)
				0.0043***	(0.0012)

Notes. Unweighted number of individuals = 3,564. Unweighted number of households = 3,151. Log(pseudo–likelihood) = –3126.9. AIC = 6327.9. BIC = 6556.5. Reliability1 (r) = 0.6135. Reliability1 (s) = 0.7035. Reliability2 (r) = 0.5837. Reliability1 (s) = 0.7092. #: Excluded due to numerical instability. [an32w]

Appendix D: All individuals ($|r_i - s_i| \leq 0.010$)

Table D1. Unweighted estimates for four models of log(earnings) without covariates: all individuals ($|r_i - s_i| \leq 0.010$)

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	9.8178***	9.8162***	9.8085***	9.8112***
	(0.0121)	(0.0120)	(0.0105)	(0.0103)
σ_ξ	0.7190***	0.7199***	0.7586***	0.7542***
	(0.0082)	(0.0081)	(0.0096)	(0.0093)
μ_ζ	8.3048***	8.1466***	8.4134***	8.5716***
	(0.4049)	(0.2684)	(0.1610)	(0.1170)
σ_ζ	1.3056***	1.2726***	1.2524***	1.2752***
	(0.0986)	(0.0994)	(0.0775)	(0.0655)
μ_ω	–0.2477***	–0.2467***	–0.1295***	–0.1354***
	(0.0386)	(0.0409)	(0.0254)	(0.0254)
σ_ω	0.8820***	0.9105***	0.7066***	0.6814***
	(0.1538)	(0.1329)	(0.0550)	(0.0556)
μ_η	–0.0149***	–0.0149***	–0.0107***	–0.0105***
	(0.0029)	(0.0032)	(0.0026)	(0.0027)
σ_η	0.1134***	0.1131***	0.1257***	0.1255***
	(0.0072)	(0.0087)	(0.0047)	(0.0048)
μ_υ	–0.1181	–0.1121		
	(0.0794)	(0.0874)		
σ_υ	0.3795***	0.3809***		
	(0.0400)	(0.0373)		
ρ_s	0.0038	0.0044	–0.0211***	–0.0231***
	(0.0067)	(0.0067)	(0.0044)	(0.0043)
ρ_r	0.2359	0.2232		
	(0.1300)	(0.1442)		
$\rho_{\xi\omega}$	0.0676		–0.1017**	
	(0.0847)		(0.0350)	
π_s	0.1012***	0.1015***	0.0813***	0.0821***
	(0.0090)	(0.0105)	(0.0036)	(0.0037)
π_ω	0.1461***	0.1453***	0.2548***	0.2483***
	(0.0323)	(0.0363)	(0.0173)	(0.0183)
π_r	0.9724***	0.9754***	0.9482***	0.9381***
	(0.0077)	(0.0048)	(0.0069)	(0.0059)
π_υ	0.7832***	0.7783***		
	(0.0629)	(0.0723)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0770***	0.0771***	0.0771***	0.0770***
	(0.0034)	(0.0034)	(0.0034)	(0.0034)
π_2 (π_2)	0.5845***	0.5830***	0.6492***	0.6472***
	(0.0332)	(0.0401)	(0.0163)	(0.0170)
π_3 (π_3)	0.1000***	0.0991**	0.2220***	0.2139***

	(0.0299)	(0.0344)	(0.0151)	(0.0157)
π_4	0.0213**	0.0220*		
	(0.0080)	(0.0092)		
π_5	0.1618**	0.1661**		
	(0.0516)	(0.0588)		
π_6	0.0277***	0.0282***		
	(0.0037)	(0.0038)		
π_7 (π_4)	0.0028***	0.0025***	0.0042***	0.0051***
	(0.0007)	(0.0006)	(0.0006)	(0.0006)
π_8 (π_5)	0.0212***	0.0189***	0.0355***	0.0427***
	(0.0055)	(0.0037)	(0.0048)	(0.0041)
π_9 (π_6)	0.0036*	0.0032**	0.0121***	0.0141***
	(0.0016)	(0.0010)	(0.0018)	(0.0018)
Log(pseudo–likelihood)	–9786.5	–9787.1	–9994.7	–9998.7
AIC	19606.9	19606.3	20015.48	20021.4
BIC	19720.7	19713.4	20102.5	20101.7
Reliability1 (*r*)	0.7554	0.7586	0.7553	0.7375
Reliability1 (*s*)	0.8034	0.8016	0.8321	0.8327
Reliability2 (*r*)	0.7721	0.7766	0.7162	0.6918
Reliability2 (*s*)	0.8149	0.8048	0.7975	0.8151

Notes. Cluster–robust standard errors in parentheses (cluster is household). Number of individuals = 5,971. Number of households = 4,874. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an05]

Table D2. Weighted estimates for four models of log(earnings) without covariates: all individuals ($|r_i - s_i| \leq 0.010$)

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	9.8330***	9.8323***	9.8199***	9.8219***
	(0.0121)	(0.0120)	(0.0125)	(0.0123)
σ_ξ	0.7146***	0.7160***	0.7513***	0.7475***
	(0.0096)	(0.0095)	(0.0113)	(0.0107)
μ_ζ	8.2440***	8.0356***	8.3836***	8.5435***
	(0.3083)	(0.3291)	(0.2100)	(0.1432)
σ_ζ	1.3639***	1.3080***	1.2144***	1.2518***
	(0.1312)	(0.1422)	(0.1108)	(0.0837)
μ_ω	–0.1420	–0.1453*	–0.0578	–0.0587
	(0.0744)	(0.0684)	(0.0374)	(0.0394)
σ_ω	1.0046***	1.0132***	0.7929***	0.7777***
	(0.1268)	(0.1172)	(0.0960)	(0.1005)
μ_η	–0.0178***	–0.0180***	–0.0122***	–0.0122***
	(0.0034)	(0.0034)	(0.0033)	(0.0034)
σ_η	0.1220***	0.1223***	0.1329***	0.1330***
	(0.0055)	(0.0056)	(0.0059)	(0.0061)
μ_υ	–0.1756**	–0.1785**		
	(0.0556)	(0.0555)		
σ_υ	0.4170***	0.4115***		
	(0.0537)	(0.0509)		
ρ_s	–0.0018	–0.0013	–0.0266***	–0.0281***
	(0.0066)	(0.0066)	(0.0054)	(0.0053)
ρ_r	0.2499***	0.2531***		
	(0.0757)	(0.0745)		
$\rho_{\xi\omega}$	0.0909		–0.0823*	
	(0.0672)		(0.0392)	
π_s	0.0944***	0.0939***	0.0774***	0.0781***
	(0.0060)	(0.0060)	(0.0040)	(0.0040)
π_ω	0.1298***	0.1341***	0.2327***	0.2252***
	(0.0208)	(0.0199)	(0.0211)	(0.0218)
π_r	0.9797***	0.9826***	0.9500***	0.9412***
	(0.0066)	(0.0057)	(0.0086)	(0.0076)
π_υ	0.7945***	0.7969***		
	(0.0311)	(0.0309)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0735***	0.0735***	0.0735***	0.0735***
	(0.0037)	(0.0037)	(0.0037)	(0.0037)
π_2 (π_2)	0.6134***	0.6144***	0.6725***	0.6723***
	(0.0222)	(0.0223)	(0.0195)	(0.0202)
π_3 (π_3)	0.0915***	0.0952***	0.2040***	0.1954***
	(0.0173)	(0.0168)	(0.0187)	(0.0189)
π_4	0.0190***	0.0187***		

	(0.0037)	(0.0037)		
π_5	0.1587***	0.1566***		
	(0.0260)	(0.0256)		
π_6	0.0237***	0.0243***		
	(0.0031)	(0.0032)		
π_7 (π_4)	0.0019**	0.0016**	0.0039***	0.0046***
	(0.0006)	(0.0006)	(0.0007)	(0.0007)
π_8 (π_5)	0.0160**	0.0136**	0.0354***	0.0420***
	(0.0051)	(0.0044)	(0.0062)	(0.0055)
π_9 (π_6)	0.0024*	0.0021**	0.0107***	0.0122***
	(0.0009)	(0.0008)	(0.0020)	(0.0020)
Log(pseudo–likelihood)	–9744.6	–9746.1	–9984.8	–9987.2
AIC	19523.2	19524.2	19995.7	19998.3
BIC	19637.0	19631.3	20082.7	20078.7
Reliability1 (r)	0.7694	0.7749	0.7573	0.7422
Reliability1 (s)	0.7858	0.7853	0.8079	0.8091
Reliability2 (r)	0.7925	0.8006	0.7194	0.6985
Reliability2 (s)	0.7963	0.7843	0.7730	0.7882

Notes. Cluster–robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 5,971. Unweighted number of households = 4,874. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Weighted estimates, with weights equal to product of FRS individual weight and inverse–probability of having consented to data linkage and been successfully linked (see text). Weights normalized so that sum equals sample number of individuals. [an05w]

Table D3. Goodness of fit statistics for four models of log(earnings) with covariates: all individuals ($|r_i - s_i| \leq 0.010$), unweighted estimates

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo–likelihood	–7185.3	–7185.8	–7529.5	–7525.6
AIC	14472.6	14471.5	15133.0	15127.2
BIC	14814.0	14806.2	15380.7	15381.6
Reliability1 (r)	0.7280	0.7264	0.7212	0.7008
Reliability1 (s)	0.8230	0.8238	0.8328	0.8289
Reliability2 (r)	0.6962	0.6936	0.6648	0.6372
Reliability2 (s)	0.8378	0.8424	0.7991	0.8169

Notes. Cluster–robust standard errors in parentheses (cluster is household). Sample: all individuals (number of individuals = 5,971 within 4,874 households). Models based on a completely–labelled fraction of 7.74% (observations with $|r_i - s_i| < 0.010$: see main text). [an25]

Table D4. Unweighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, all individuals, ($|r_i - s_i| \leq 0.010$)

		MM	(SE)		MM	(SE)
Male	μ_ξ	9.9474***	(0.0115)	σ_ξ	0.5100***	(0.0107)
Female		9.6897***	(0.0098)		0.4739***	(0.0078)
Education: less than A–level		9.6168***	(0.0096)		0.4333***	(0.0080)
Education: A–level or more		10.0086***	(0.0120)		0.5531***	(0.0101)
Full–time employee		10.0443***	(0.0082)		0.4394***	(0.0070)
Part–time employee		9.2124***	(0.0187)		0.6158***	(0.0151)
Married, cohabiting		9.8280***	(0.0094)			
Single, divorced, separated, widowed		9.7395***	(0.0127)			
Has 1 job		9.7971***	(0.0084)			
Has 2+ jobs		9.9197***	(0.0361)			
Age = 25 years		9.5210***	(0.0149)		0.4165***	(0.0118)
Age = 35 years		9.7732***	(0.0099)		0.4741***	(0.0081)
Age = 45 years		9.9006***	(0.0103)		0.5106***	(0.0087)
Age = 55 years		9.9033***	(0.0110)		0.5203***	(0.0096)
	μ_ζ	8.9699***	(0.0965)	σ_ζ	1.2706***	(0.0713)
Payslip(s) not shown to interviewer	μ_η	–0.0527***	(0.0067)	σ_η	0.1415***	(0.0099)
Payslip(s) shown (all jobs)		–0.0018	(0.0044)		0.0849***	(0.0044)
Aged < 60 years		–0.0087*	(0.0040)		0.0926***	(0.0046)
Aged 60+ years		–0.1153***	(0.0211)		0.2108***	(0.0287)
Reference period: not ‘other’	μ_ω	–0.0694	(0.0610)	σ_ω	0.9820***	(0.1185)
Reference period: other		–0.5851***	(0.1564)		0.6008*	(0.2771)
Job spells do not span year		–0.2786***	(0.0757)		0.8733***	(0.1042)
Job spells all span year		0.0438	(0.0900)		1.0354***	(0.1506)
Aged < 60 years		–0.0212	(0.0614)		0.9829***	(0.1152)
Aged 60+ years		–0.6034***	(0.1416)		0.8906***	(0.3363)
Full–time employee		0.0389	(0.0789)		1.0295**	(0.1086)
Part–time employee		–0.3703***	(0.1106)		0.8356***	(0.2265)
Payslip provided by employer	μ_υ	0.0138	(0.0109)	σ_υ	0.2202***	(0.0260)
Payslip not provided by employer		–0.1447**	(0.0441)		0.4268***	(0.0580)
Full–time employee		0.0013	(0.0099)		0.1645***	(0.0256)
Part–time employee		–0.0077	(0.0261)		0.4285***	(0.0384)
Private sector employee		0.0242	(0.0124)		0.2622***	(0.0229)
Public sector employee		–0.0586***	(0.0142)		0.1976***	(0.0440)
Male	ρ_s	0.0032	(0.0141)	ρ_r	0.0586*	(0.0229)
Female		0.0015	(0.0086)		0.2800**	(0.1011)
Aged < 60 years		0.0084	(0.0088)			
Aged 60+ years		–0.0517	(0.0442)			
Private sector employee					0.1295***	(0.0264)
Public sector employee					–0.0322	(0.0348)
Probabilities	π_s	0.1525***	(0.0134)	π_1	0.0774***	(0.0035)
	π_ω	0.0955***	(0.0156)	π_2	0.3889***	(0.0375)
	π_r	0.9372***	(0.0068)	π_3	0.0410***	(0.0063)
	π_υ	0.5413***	(0.0381)	π_4	0.0655***	(0.0106)
				π_5	0.3296***	(0.0206)
				π_6	0.0348***	(0.0069)
				π_7	0.0096***	(0.0016)

π_8	0.0481***	(0.0048)
π_9	0.0051***	(0.0010)

Notes. Sample: all individuals (number of individuals = 5,971 within 4,874 households). Log(pseudo–likelihood) = –7185.8. AIC = 14471.5. BIC = 14806.2. Reliability1 (r) = 0.7264. Reliability1 (s) = 0.8238. Reliability2 (r) = 0.6936. Reliability1 (s) = 0.8424. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an25]

Table D5. **Goodness of fit statistics for four models of log(earnings) with covariates: all individuals ($|r_i - s_i| \leq 0.010$), weighted estimates**

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo–likelihood	–7206.5	–7207.0	–7542.1	–7545.6
AIC	14514.9	14514.0	15162.1	15167.1
BIC	14856.4	14848.8	15423.2	15421.5
Reliability1 (r)	0.7349	0.7332	0.7066	0.6953
Reliability1 (s)	0.8032	0.8044	0.8017	0.8054
Reliability2 (r)	0.7045	0.7019	0.6432	0.6272
Reliability2 (s)	0.8162	0.8218	0.7810	0.7966

Notes. Cluster–robust standard errors in parentheses (cluster is household). Sample: all individuals (unweighted number of individuals = 5,971 within unweighted 4,874 households). Models based on a completely–labelled fraction of 7.74% (observations with $|r_i - s_i| < 0.010$: see main text). [an25w]

Table D6. Weighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, all individuals, ($|r_i - s_i| \leq 0.010$)

		MM	(SE)		MM	(SE)
Male	μ_ξ	9.9423***	(0.0151)	σ_ξ	0.5173***	(0.0130)
Female		9.7142***	(0.0112)		0.4685***	(0.0089)
Education: less than A–level		9.6336***	(0.0120)		0.4357***	(0.0104)
Education: A–level or more		10.0028***	(0.0144)		0.5463***	(0.0111)
Full–time employee		10.0422***	(0.0101)		0.4447***	(0.0083)
Part–time employee		9.1766***	(0.0224)		0.6266***	(0.0180)
Married, cohabiting		9.8485***	(0.0111)			
Single, divorced, separated, widowed		9.7480***	(0.0167)			
Has 1 job		9.8143***	(0.0100)			
Has 2+ jobs		9.9266***	(0.0469)			
Age = 25 years		9.5652***	(0.0173)		0.4096***	(0.0130)
Age = 35 years		9.8317***	(0.0121)		0.4784***	(0.0092)
Age = 45 years		9.9585***	(0.0125)		0.5255***	(0.0102)
Age = 55 years		9.9454***	(0.0139)		0.5428***	(0.0126)
	μ_ζ	8.9867***	(0.1159)	σ_ζ	1.2230***	(0.0875)
Payslip(s) not shown to interviewer	μ_η	–0.0511***	(0.0085)	σ_η	0.1465***	(0.0097)
Payslip(s) shown (all jobs)		–0.0050	(0.0052)		0.0868***	(0.0054)
Aged < 60 years		–0.0131*	(0.0051)		0.0990***	(0.0057)
Aged 60+ years		–0.1144***	(0.0207)		0.2091***	(0.0265)
Reference period: not ‘other’	μ_ω	0.0797	(0.1048)	σ_ω	1.0526***	(0.1157)
Reference period: other		–0.5327***	(0.1049)		0.3919***	(0.2753)
Job spells do not span year		–0.1782*	(0.0854)		0.8636***	(0.1058)
Job spells all span year		0.2229	(0.1593)		1.1483***	(0.1501)
Aged < 60 years		0.1124	(0.1041)		1.0417***	(0.1110)
Aged 60+ years		–0.5004**	(0.1804)		0.9990*	(0.4102)
Full–time employee		0.1774	(0.1266)		1.0803***	(0.1113)
Part–time employee		–0.2533	(0.1675)		0.9141***	(0.2436)
Payslip provided by employer	μ_υ	0.0045	(0.0146)	σ_υ	0.2210***	(0.0383)
Payslip not provided by employer		–0.2038*	(0.0929)		0.4596***	(0.0887)
Full–time employee		–0.0141	(0.0194)		0.1783***	(0.0409)
Part–time employee		–0.0189	(0.0322)		0.4391***	(0.0484)
Private sector employee		0.0072	(0.0211)		0.2696***	(0.0389)
Public sector employee		–0.0738***	(0.0182)		0.1864***	(0.0545)
Male	ρ_s	0.0081	(0.0166)	ρ_r	0.0664*	(0.0271)
Female		–0.0061	(0.0105)		0.3183	(0.2391)
Aged < 60 years		0.0067	(0.0116)			
Aged 60+ years		–0.0768	(0.0436)			
Private sector employee					0.1375**	(0.0432)
Public sector employee					–0.0318	(0.0420)
Probabilities	π_s	0.1411***	(0.0199)	π_1	0.0738***	(0.0038)
	π_ω	0.0959***	(0.0158)	π_2	0.4064***	(0.0642)
	π_r	0.9383***	(0.0108)	π_3	0.0431***	(0.0071)
	π_υ	0.5578***	(0.0664)	π_4	0.0585***	(0.0163)
				π_5	0.3222***	(0.0368)
				π_6	0.0342***	(0.0083)
				π_7	0.0087***	(0.0025)

π_8	0.0479***	(0.0075)
π_9	0.0051***	(0.0013)

Notes. Sample: all individuals (unweighted number of individuals = 5,971 within unweighted 4,874 households). Log(pseudo–likelihood) = –7207.0. AIC = 14514.0. BIC = 14848.7. Reliability1 (r) = 0.7332. Reliability1 (s) = 0.8044. Reliability2 (r) = 0.7019. Reliability1 (s) = 0.8218. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an25w]

Appendix E: Individuals aged 25–59, in full–time work, not in education ($|r_i - s_i| \leq 0.010$)

Table E1. Unweighted estimates for four models of log(earnings) without covariates: individuals aged 25–59, in full–time work, not in education ($|r_i - s_i| \leq 0.010$)

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	10.1247***	10.1246***	10.1289***	10.1288***
	(0.0108)	(0.0108)	(0.0100)	(0.0100)
σ_ξ	0.5480***	0.5482***	0.5601***	0.5604***
	(0.0086)	(0.0085)	(0.0087)	(0.0086)
μ_ζ	8.6991***	8.6839***	9.3966***	9.3873***
	(0.3737)	(0.3745)	(0.1169)	(0.1126)
σ_ζ	1.3110***	1.3081***	1.3082***	1.3114***
	(0.1445)	(0.1474)	(0.0848)	(0.0835)
μ_ω	0.0293	0.0238	0.0170	0.0168
	(0.1138)	(0.1122)	(0.0664)	(0.0654)
σ_ω	1.2770***	1.2821***	0.9655***	0.9650***
	(0.1104)	(0.1062)	(0.1260)	(0.1267)
μ_η	–0.0081	–0.0081	–0.0124***	–0.0124***
	(0.0044)	(0.0044)	(0.0029)	(0.0029)
σ_η	0.1040***	0.1039***	0.1303***	0.1303***
	(0.0117)	(0.0116)	(0.0041)	(0.0041)
μ_υ	–0.0081	–0.0080		
	(0.0407)	(0.0402)		
σ_υ	0.3498***	0.3496***		
	(0.0888)	(0.0872)		
ρ_s	–0.0148	–0.0147	–0.0514***	–0.0513***
	(0.0125)	(0.0123)	(0.0062)	(0.0062)
ρ_r	0.0869	0.0859		
	(0.0668)	(0.0656)		
$\rho_{\xi\omega}$	0.0524		0.0246	
	(0.0942)		(0.0753)	
π_s	0.1116***	0.1116***	0.0890***	0.0889***
	(0.0134)	(0.0132)	(0.0051)	(0.0050)
π_ω	0.0516***	0.0519***	0.0913***	0.0926***
	(0.0077)	(0.0078)	(0.0186)	(0.0183)
π_r	0.9633***	0.9638***	0.9186***	0.9197***
	(0.0127)	(0.0125)	(0.0106)	(0.0097)
π_υ	0.7614***	0.7611***		
	(0.0723)	(0.0713)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0819***	0.0819***	0.0818***	0.0818***
	(0.0046)	(0.0046)	(0.0046)	(0.0046)
π_2 (π_2)	0.6179***	0.6178***	0.7604***	0.7603***
	(0.0735)	(0.0725)	(0.0119)	(0.0120)
π_3 (π_3)	0.0336***	0.0338***	0.0764***	0.0776***

	(0.0070)	(0.0070)	(0.0162)	(0.0160)
π_4	0.0257*	0.0257*		
	(0.0103)	(0.0102)		
π_5	0.1937***	0.1940***		
	(0.0543)	(0.0535)		
π_6	0.0105***	0.0106***		
	(0.0031)	(0.0031)		
π_7 (π_4)	0.0041*	0.0040*	0.0072***	0.0071***
	(0.0018)	(0.0018)	(0.0011)	(0.0010)
π_8 (π_5)	0.0309**	0.0305**	0.0674***	0.0664***
	(0.0104)	(0.0102)	(0.0097)	(0.0090)
π_9 (π_6)	0.0017**	0.0017**	0.0068***	0.0068***
	(0.0006)	(0.0006)	(0.0010)	(0.0010)
Log(pseudo–likelihood)	–4173.2	–4173.4	–4300.1	–4300.2
AIC	8380.5	8378.8	8626.3	8624.4
BIC	8485.5	8477.7	8706.6	8698.5
Reliability1 (*r*)	0.6355	0.6363	0.6163	0.6178
Reliability1 (*s*)	0.7838	0.7843	0.7893	0.7888
Reliability2 (*r*)	0.6249	0.6259	0.5661	0.5682
Reliability2 (*s*)	0.7779	0.7740	0.7551	0.7519

Notes. Cluster–robust standard errors in parentheses (cluster is household). Number of individuals = 3,564. Number of households = 3,151. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. [an28]

Table E2. Weighted estimates for four models of log(earnings) without covariates: individuals aged 25–59, in full–time work, not in education ($|r_i - s_i| \leq 0.010$)

Parameters	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
μ_ξ	10.1343***	10.1341***	10.1341***	10.1338***
	(0.0127)	(0.0127)	(0.0124)	(0.0124)
σ_ξ	0.5520***	0.5527***	0.5643***	0.5652***
	(0.0101)	(0.0100)	(0.0100)	(0.0099)
μ_ζ	8.8194***	8.7744***	9.4918***	9.4688***
	(0.3013)	(0.3073)	(0.1135)	(0.1114)
σ_ζ	1.3651***	1.3611***	1.2617***	1.2707***
	(0.1459)	(0.1502)	(0.0970)	(0.0983)
μ_ω	0.2363	0.2215	0.2033	0.1919
	(0.1938)	(0.1822)	(0.1597)	(0.1434)
σ_ω	1.3113***	1.3175***	1.1183***	1.1147***
	(0.1361)	(0.1370)	(0.1708)	(0.1698)
μ_η	–0.0123*	–0.0124*	–0.0145***	–0.0145***
	(0.0049)	(0.0049)	(0.0037)	(0.0037)
σ_η	0.1115***	0.1116***	0.1349***	0.1348***
	(0.0126)	(0.0125)	(0.0047)	(0.0047)
μ_υ	–0.0438	–0.0441		
	(0.0630)	(0.0633)		
σ_υ	0.3564***	0.3569***		
	(0.0829)	(0.0812)		
ρ_s	–0.0110	–0.0108	–0.0457***	–0.0454***
	(0.0129)	(0.0128)	(0.0086)	(0.0086)
ρ_r	0.1286	0.1275		
	(0.0957)	(0.0973)		
$\rho_{\xi\omega}$	0.1164		0.0863	
	(0.1222)		(0.1072)	
π_s	0.1013***	0.1012***	0.0839***	0.0836***
	(0.0122)	(0.0121)	(0.0054)	(0.0053)
π_ω	0.0612***	0.0623***	0.0835***	0.0871***
	(0.0106)	(0.0112)	(0.0175)	(0.0174)
π_r	0.9651***	0.9665***	0.9125***	0.9158***
	(0.0117)	(0.0113)	(0.0128)	(0.0119)
π_υ	0.7832***	0.7836***		
	(0.0729)	(0.0727)		
Class probabilities Extended (Full)				
π_1 (π_1)	0.0766***	0.0766***	0.0765***	0.0766***
	(0.0048)	(0.0048)	(0.0048)	(0.0048)
π_2 (π_2)	0.6377***	0.6383***	0.7661***	0.7661***
	(0.0718)	(0.0711)	(0.0120)	(0.0121)
π_3 (π_3)	0.0416***	0.0424***	0.0698***	0.0731***
	(0.0096)	(0.0100)	(0.0154)	(0.0153)
π_4	0.0212*	0.0212*	0.0073***	

	(0.0092)	(0.0092)	(0.0012)	
π_5	0.1765**	0.1763**	0.0735***	
	(0.0562)	(0.0562)	(0.0116)	
π_6	0.0115**	0.0117**	0.0067***	
	(0.0037)	(0.0038)	(0.0011)	
π_7 (π_4)	0.0035*	0.0034*		0.0070***
	(0.0015)	(0.0015)		(0.0012)
π_8 (π_5)	0.0294**	0.0282**		0.0705***
	(0.0097)	(0.0094)		(0.0108)
π_9 (π_6)	0.0019**	0.0019**		0.0067***
	(0.0007)	(0.0006)		(0.0010)
Log(pseudo–likelihood)	–4361.6	–4362.5	–4486.8	–4487.4
AIC	8757.3	8757.1	8999.6	8998.8
BIC	8862.3	8855.9	9079.9	9072.9
Reliability1 (*r*)	0.6561	0.6587	0.6278	0.6323
Reliability1 (*s*)	0.7354	0.7367	0.7442	0.7430
Reliability2 (*r*)	0.6509	0.6542	0.5728	0.5790
Reliability2 (*s*)	0.7393	0.7296	0.7228	0.7121

Notes. Cluster–robust standard errors in parentheses (cluster is household). Unweighted number of individuals = 3,564. Unweighted number of households = 3,151. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Weighted estimates, with weights equal to product of FRS individual weight and inverse–probability of having consented to data linkage and been successfully linked (see text). Weights normalized so that sum equals sample number of individuals. [an28w]

Table E3. Goodness of fit statistics for four models of log(earnings) with covariates: individuals aged 25–59, in full–time work, not in education ($|r_i - s_i| \leq 0.010$), unweighted estimates

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo–likelihood	–3409.8	–3409.8	–3673.6	–3673.6
AIC	6895.6	6893.7	7403.2	7401.2
BIC	7130.4	7122.3	7576.2	7568.0
Reliability1 (r)	0.6012	0.6014	0.6091	0.6088
Reliability1 (s)	0.7604	0.7601	0.7719	0.7721
Reliability2 (r)	0.5660	0.5664	0.5574	0.5570
Reliability2 (s)	0.7432	0.7416	0.7346	0.7357

Notes. Cluster–robust standard errors in parentheses (cluster is household). Sample: individuals aged 25–59, in full–time work, not in education (unweighted number of individuals = 3,564 within unweighted 3,151 households). Model based on a completely–labelled fraction of 7.74% (observations with $|r_i - s_i| < 0.010$: see main text). [an33]

Table E4. Unweighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, individuals aged 25–59, working full–time, not in education, ($|r_i - s_i| \leq 0.010$)

		MM	(SE)		MM	(SE)
Male	μ_ξ	10.2530***	(0.0115)	σ_ξ	0.4823***	(0.0117)
Female		9.9739***	(0.0117)		0.4447***	(0.0098)
Education: less than A–level		9.9033***	(0.0108)		0.4142***	(0.0095)
Education: A–level or more		10.3430***	(0.0130)		0.5158***	(0.0125)
Married, cohabiting		10.1457***	(0.0101)			
Single, divorced, separated, widowed		10.0745***	(0.0154)			
Has 1 job		10.1260***	(0.0087)			
Has 2+ jobs		10.1464***	(0.0563)			
Age = 25 years		9.8214***	(0.0241)		0.3602***	(0.0171)
Age = 35 years		10.0796***	(0.0112)		0.4487***	(0.0095)
Age = 45 years		10.1980***	(0.0119)		0.4938***	(0.0111)
Age = 55 years		10.1765***	(0.0151)		0.4800***	(0.0148)
	μ_ζ	9.2610***	(0.0986)	σ_ζ	1.2696***	(0.0788)
Payslip(s) not shown to interviewer	μ_η	–0.0601***	(0.0112)	σ_η	0.2053***	(0.0126)
Payslip(s) shown (all jobs)		0.0050	(0.0062)		0.1473***	(0.0079)
Reference period: not ‘other’	μ_ω	#		σ_ω	#	
Reference period: other		#			#	
Job spells do not span year		–0.1177	(0.1327)		1.1265***	(0.1411)
Job spells all span year		0.1484	(0.1495)		1.3139***	(0.1308)
Payslip provided by employer	μ_υ	0.0098**	(0.0031)	σ_υ	0.0539***	(0.0037)
Payslip not provided by employer		–0.0035	(0.0138)		0.0703***	(0.0153)
Private sector employee		0.0348***	(0.0043)		0.0697***	(0.0048)
Public sector employee		–0.0549***	(0.0020)		0.0204***	(0.0020)
Male	ρ_s	–0.0588*	(0.0295)	ρ_r	0.0356**	(0.0133)
Female		–0.0505*	(0.0220)		0.0289	(0.0241)
Private sector employee					0.0541**	(0.0177)
Public sector employee					–0.0112*	(0.0045)
Probabilities	π_s	0.3927***	(0.0253)	π_1	0.0825***	(0.0046)
	π_ω	0.0818***	(0.0110)	π_2	0.1171***	(0.0136)
	π_r	0.9248***	(0.0069)	π_3	0.0104***	(0.0018)
	π_υ	0.2270***	(0.0187)	π_4	0.2807***	(0.0240)
				π_5	0.3986***	(0.0123)
				π_6	0.0355***	(0.0050)
				π_7	0.0295***	(0.0031)
				π_8	0.0420***	(0.0046)
				π_9	0.0037***	(0.0006)

Notes. Sample: individuals aged 25–59, in full–time work, not in education (unweighted number of individuals 3,564 within unweighted 3,151 households). Log(pseudo–likelihood) = –3409.8. AIC = 6893.7. BIC = 7122.3. Reliability1 (r) = 0.6014. Reliability1 (s) = 0.7601. Reliability2 (r) = 0.5664. Reliability1 (s) = 0.7416. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. #: Excluded due to numerical instability. [an33]

Table E5. Goodness of fit statistics for four models of log(earnings) with covariates: individuals aged 25–59, in full–time work, not in education ($|r_i - s_i| \leq 0.010$), weighted estimates

	Extended model with $\rho_{\xi\omega} \neq 0$ (1)	Constrained Extended model ($\rho_{\xi\omega} = 0$) (2)	Full model with $\rho_{\xi\omega} \neq 0$ (3)	Constrained Full model ($\rho_{\xi\omega} = 0$) (4)
Log pseudo–likelihood	–3684.0	–3684.0	–3817.7	–3817.9
AIC	7443.9	7441.9	7691.4	7689.8
BIC	7678.7	7670.5	7864.4	7856.6
Reliability1 (r)	0.6372	0.6374	0.6284	0.6268
Reliability1 (s)	0.7129	0.7126	0.7130	0.7147
Reliability2 (r)	0.6193	0.6196	0.5761	0.5739
Reliability2 (s)	0.7103	0.7091	0.6807	0.6867

Notes. Cluster–robust standard errors in parentheses (cluster is household). Sample: individuals aged 25–59, in full–time work, not in education (unweighted number of individuals = 3,564 within unweighted 3,151 households). Model based on a completely–labelled fraction of 3.43% (observations with $|r_i - s_i| < 0.005$: see main text). [an33w]

Table E6. Weighted estimates of constrained Extended model with covariates: marginal means (MMs) and probabilities, individuals aged 25–59, working full–time, not in education, ($|r_i - s_i| \leq 0.010$)

		MM	(SE)		MM	(SE)
Male	μ_ξ	10.2485***	(0.0160)	σ_ξ	0.4932***	(0.0136)
Female		9.9892***	(0.0143)		0.4445***	(0.0110)
Education: less than A–level		9.9056***	(0.0148)		0.4242***	(0.0131)
Education: A–level or more		10.3380***	(0.0166)		0.5158***	(0.0132)
Married, cohabiting		10.1554***	(0.0129)			
Single, divorced, separated, widowed		10.0778***	(0.0224)			
Has 1 job		10.1342***	(0.0118)			
Has 2+ jobs		10.1555***	(0.0753)			
Age = 25 years		9.8656***	(0.0300)		0.3653***	(0.0185)
Age = 35 years		10.1211***	(0.0149)		0.4573***	(0.0118)
Age = 45 years		10.2250***	(0.0152)		0.5086***	(0.0131)
Age = 55 years		10.1774***	(0.0226)		0.5028***	(0.0198)
	μ_ζ	9.0066***	(0.2215)	σ_ζ	1.4276***	(0.1777)
Payslip(s) not shown to interviewer	μ_η	–0.0411***	(0.0096)	σ_η	0.1395***	(0.0133)
Payslip(s) shown (all jobs)		–0.0005	(0.0058)		0.0914***	(0.0119)
Reference period: not 'other'	μ_ω	#		σ_ω	#	
Reference period: other		#			#	
Job spells do not span year		–0.0990	(0.1285)		1.0660***	(0.1392)
Job spells all span year		0.4637	(0.2595)		1.4094***	(0.1398)
Payslip provided by employer	μ_υ	–0.0075	(0.0423)	σ_υ	0.2572***	(0.0689)
Payslip not provided by employer		–0.4969*	(0.2380)		0.7082***	(0.1268)
Private sector employee		–0.0076	(0.0484)		0.2934***	(0.0654)
Public sector employee		–0.1665*	(0.0785)		0.3038**	(0.0965)
Male	ρ_s	–0.0049	(0.0207)	ρ_r	0.0666	(0.0533)
Female		–0.0154	(0.0168)		–0.0364	(0.3524)
Private sector employee					0.1234*	(0.0554)
Public sector employee					–0.1246	(0.1372)
Probabilities	π_s	0.1088***	(0.0152)	π_1	0.0763***	(0.0048)
	π_ω	0.0659***	(0.0118)	π_2	0.5839***	(0.0859)
	π_r	0.9614***	(0.0117)	π_3	0.0412***	(0.0073)
	π_υ	0.7295***	(0.0848)	π_4	0.0283***	(0.0123)
				π_5	0.2164***	(0.0615)
				π_6	0.0153***	(0.0060)
				π_7	0.0042***	(0.0018)
				π_8	0.0322***	(0.0092)
				π_9	0.0023***	(0.0009)

Notes. Sample: individuals aged 25–59, in full–time work, not in education (unweighted number of individuals 3,564 within unweighted 3,151 households). Log(pseudo–likelihood) = –3684.0. AIC = 7441.9. BIC = 7670.5. Reliability1 (r) = 0.6374. Reliability1 (s) = 0.7126. Reliability2 (r) = 0.6196. Reliability1 (s) = 0.7091. Statistical significance indicators: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. #: Excluded due to numerical instability. [an33w]