

**Understanding Society
Working Paper Series**

No. 2022 – 01

March 2022

Assessing earnings and income data from a short web survey

Thomas F. Crossley¹, Paul Fisher^{2,#}, Omar Hussein²

¹European University Institute

²Institute for Social and Economic Research, University of Essex

[#]Corresponding author.

Non-technical summary

We study the data quality and measurement error properties of earnings and income data collected in the Understanding Society COVID-19 study. Like many short web surveys fielded during the pandemic, earnings and income data were collected with single or short question sets, rather than the detailed question sets recommended for surveys that focus on earnings and income.

Earnings and income data is also collected with single questions or short question sets in surveys that prioritize other content domains. Examples include the US General Social Survey, the Gallup Healthways Wellbeing Index, the British Social Attitudes Survey, the British Crime Survey, the European Social Survey, and the Health Survey for England. Given the importance of earnings and income data to research and policymaking, it is imperative to understand the quality of the data collected in such cases.

We find that measures of earnings and income in the COVID-19 Study are noisier than those from the main annual Understanding Society survey, and that there is evidence of systematic under-reporting for household totals. However, we find no evidence that measurement errors in the COVID-19 Study are substantively correlated with true values. We conclude that the COVID-19 Study collected useful data on earnings and income, and therefore, that simple collection of useful earnings or income data is feasible.

Acknowledgements: Author order is alphabetical. The *Understanding Society* COVID-19 Study was funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). Fieldwork for the survey was carried out by Ipsos MORI and Kantar. *Understanding Society* is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research at the University of Essex. The research data are distributed by the UK Data Service. The authors thank Hamish Low for useful comments.

Assessing earnings and income data from a short web survey

Thomas F. Crossley* Paul Fisher[†] Omar Hussein[‡]

March 4, 2022

Abstract

In short web-surveys, or in surveys that prioritize other content domains, earnings and income are often elicited with single questions or small question sets. This contrasts with the detailed question sets recommended for surveys that focus on earnings and income. We evaluate earnings and income data collected with a short question set in a series of recent web-surveys: the *Understanding Society* COVID-19 Study. The fact that many COVID-19 Study respondents also contemporaneously answered the main annual *Understanding Society* survey provides individual- and household-level validation data. We find that measures of earnings and income in the COVID-19 Study are noisier than those from the main annual *Understanding Society* survey, and that there is evidence of systematic under-reporting for household totals. However, we find no evidence that measurement errors in the COVID-19 Study are substantively correlated with true values. We conclude that the COVID-19 Study collected useful data on earnings and income, and therefore, that simple collection of useful earnings or income data is feasible.

Keywords: Validation, measurement error, data quality, COVID-19

JEL codes: C81, C83, I32

Declaration of interests: None

*European University Institute and *Understanding Society* (email: thomas.crossley@EUI.eu).

[†]ISER, University of Essex and *Understanding Society* (Corresponding Author. email: pfishe@essex.ac.uk. Postal address: University of Essex, Wivenhoe Park, Colchester, Essex, CO4 3SQ, UK).

[‡]ISER, University of Essex and *Understanding Society* (email: om20319@essex.ac.uk).

1 Introduction

During times of crisis, the demand for timely data may rise, and social science surveys must be developed and fielded at speed. In particular, earnings and income can be tracked with a short question set in web surveys. Many such surveys were fielded during the COVID-19 pandemic and continue to be analysed by economists, see for example, Belot et al. (2020); Adams-Prassl et al. (2020); Bundervoet et al. (2022); Institute for Social and Economic Research (2020); Crossley et al. (2021). These short surveys are important because administrative data sources (such as tax returns) may be available only at annual frequency, and official household income surveys may also lack timeliness, or may have been disrupted eg. Ward and Edwards (2021). Earnings and income data is also collected with single questions or short question sets in surveys that prioritize other content domains. Examples include the US General Social Survey, the Gallup Healthways Wellbeing Index, the European Social Survey, and the Health Survey for England. Data from these surveys are used extensively in economic research.

The best practise approach to collecting earnings and income data in a household survey is with a detailed set of questions that ask about each household member about each individual income source, and then, post fieldwork, to aggregate over sources and then over individuals in the household (United Nations (2011)). In the crisis context, or where a survey prioritizes other content domains, practise may deviate from this ideal. Demands on interview space push data collectors towards fewer questions that ask about earnings and income totals, rather than going source by source. Where field work periods are compressed, it may not be possible to interview all household members, and one individual may need to report household totals. Given the importance of earnings and income data to research and policymaking, it is imperative to understand the quality of the data collected in such cases.

In this paper, we study the quality of earnings and income data collected with short question sets in a series of web surveys fielded during the COVID-19 pandemic - the *Understanding Society* COVID-19 Study. We take a novel approach to data validation. We link individuals who were interviewed twice in a short-time period - with a long set of detailed earnings and income questions in the main annual *Understanding Society* survey, and in the COVID-19 Study. As the gap between interviews is very close, we can use the more detailed survey to learn about the error properties of the data from the shorter one. We exploit the individual-level linkage between the two surveys to describe the differences in reports at an individual level. We then interpret those differences through the lens of an estimated model of measurement error. That model does not impose that the validation source (here, the *Understanding Society* Main Study) is error free, but it does allow

us to compare the quality of the two sources.

We have three main findings. First, short question sets in the COVID-19 Study produce very reliable data on individual earnings; second, household earnings and income is systematically under-reported in the short question sets; and third, measurement error in the COVID-19 Study earnings and income measure is largely uncorrelated with true values. The latter means that, reassuringly, that an instrumental variables approach can correct for measurement error in a regression context.

A growing literature has looked at the misreporting of earnings and income in household surveys. An assumption of early validation studies was that the validation source, typically administrative data, was error free (see for example Bound et al. (1994); Pischke (1995)). However, more recent papers have done away with the assumption that validation sources are error free, allow for errors of reporting or of linkage in administrative data (Abowd and Stinson (2013); Bingley and Martinello (2017); Wilhelm (2018); Kapteyn and Ypma (2007)).

Whilst the validation literature has assessed data from income-focussed surveys with detailed earnings and income question sets, there is little evidence on the performance of short question sets. This is a surprising fact given their widespread use. A rare exception is Micklewright and Schnepf (2010) who performs aggregate validations by comparing responses to a single income question in one survey to responses from a more detailed set of questions in another. The authors conclude that distributions compare less well for household income than for individual income.

Our approach offers several advantages over the existing literature. An important way our work differs from Micklewright and Schnepf (2010) is that we observe responses to both the detailed and short earnings and income question sets for the same individual. This means that we can directly estimate measurement error models and are not limited to simple aggregate comparisons of distributions from different surveys. For example, we can quantify the bias of OLS and IV estimators estimated on data from the short question sets. A further strength is that we do not assume our validation source – the more detailed survey – to be error free.

In the broader validation literature, early studies typically focussed on special samples (for example the early PSID validation studies that used data from a single firm (Bound et al., 1994; Pischke, 1995)). Latter studies matched large national surveys to social security records or other administrative sources (Gottschalk and Huynh (2010); Meyer and Mittag (2019)) but then had to deal with the possibility of linkage error (Kapteyn and Ypma (2007); Jenkins and Rios-Avila (2020)). Unusually, we are able to work with a large nationally representative sample, but we do not need to worry about linkage error because the COVID-19 Study was launched directly from

our validation source (the *Understanding Society* Main Study).

The rest of the paper proceeds as follows. Section 2 describes our data sources and lays out our research design. Section 3 provides a descriptive comparison of our two data sources before presenting our estimated measurement error models. Section 4 concludes.

2 Data and methods

2.1 *Understanding Society* and the *Understanding Society* COVID-19 Study

We link the *Understanding Society*: the UK Household Longitudinal Study (henceforth Main Study) with the *Understanding Society* COVID-19 study (henceforth COVID-19 Study). The former is a long-standing general purpose panel survey from the UK, and it is our validation source for information on earnings and incomes. The latter is our source of income data collected with a short set of questions, and it is a series of short web surveys conducted during the pandemic and completed by the Main Study participants. It is a prime example of a survey collecting earnings and income data with a short question set in the face of pressures on interview length.

Our analysis is based on waves two to six of the COVID-19 Study and wave eleven of the Main Study.¹ The COVID-19 waves were fielded in the last weeks of April, May, June, July, September, and November 2020, and of March 2021. We work with the waves fielded from May onwards. The Main Study is a mixed-mode survey, collecting data from participants annually by face-to-face or web interview since 2009, but switched to web mode (with telephone follow-up) from the onset of the pandemic (Burton et al., 2020). It is one of the largest household panel studies in the world. We take the Main study data from wave 11 for which the fieldwork was conducted across 2019-2020. The COVID-19 Study employs shorter and more frequent web surveys to record the experiences and behaviour of Main Study participants during the COVID-19 pandemic. Each such web survey is designed to take about 20 minutes to complete, in contrast to the 45 minutes for the more detailed and extensive questions of the Main Study. Both studies contain a mix of repeating and rotating content. Income and earnings variables are collected at almost all waves.

All individual members of the Main Study who were aged sixteen or over in April 2020, and who belonged to active households, were invited to participate in the COVID-19 Study.² In mid-April, potential respondents were sent a pre-notification letter introducing the study and offering a small incentive for each web survey they completed. The fieldwork period for each web survey

¹We do not include wave one of the COVID-19 Study because the income questions were only asked from wave two. However, we provide some analysis of the wave one earnings data in Supplemental Appendix B.

²An active household is one that participated in at least one of the last two waves of the main study.

lasted seven days, whereas in the Main Study the fieldwork extends for up to nearly six months. Additional information on the *Understanding Society* COVID-19 Study can be found in Institute for Social and Economic Research (2020) and Institute for Social and Economic Research (2021b), and for the Main Study in Institute for Social and Economic Research (2021a).

The detailed earnings and income questions of the Main Study follow international best practise (United Nations, 2011). Main Study income data have been shown to be of high quality. Earnings and income distributions from the Main Study match closely those from other high-quality sources (eg. Fisher et al. (2019), Fisher (2019)), and Main Study income data are the source for official UK statistics on poverty dynamics. While the earnings and income questions of the COVID-19 Study have been extensively used in academic research, they have been difficult to validate against external sources as no benchmark for the pandemic period yet exists. Further relevant details of the earnings and income questions asked in each survey are discussed below.

2.2 Research design

Our research design is based on the fact that the *Understanding Society* Main Study continued to run throughout the pandemic (though with a reduced set of survey modes). Study participants are issued to the field for annual main surveys interview throughout the year, with batches issued at the beginning of each month. This means for each COVID-19 Study wave, approximately one twelfth of the sample will have been invited to a main survey interview within a few weeks of the invitation to the COVID-19 Study interview for that wave. The resulting pattern of interviews is shown in Figure 1. Time is recorded on the horizontal axis and number of observations (or interviews), by day, are record on the vertical axis. The number of COVID-19 Study interviews is plotted in red, while the number of Main Study interviews is plotted in blue. The top panel of Figure 1 shows the period from late May 2020 through to the end of the calendar year used in our analysis. This captures five COVID-19 Study Waves (Waves two through six). Because each COVID-19 Study wave occurred in the last week of the month, and each monthly batch of Main Survey invitations are released in the first week of the month, there are repeated Main Study interview peaks shortly after each COVID-19 Study interview peak.

The utility of this for our research questions is that, when a study participant completes a COVID-19 Survey interview and Main Study interview in quick succession, they respond to (for a similar period) both the short set of earnings and income questions in the COVID-19 Study, and the full set of earnings and income questions in the Main Study (these are described in detail in the next subsection). We construct our sample by looking for Main Study interviews by the same

participant in a window of plus or minus fourteen days around their COVID-19 Study interview.³ This is shown in the second panel of Figure 1, (which is a blow-up of one section of the first panel, corresponding to the May COVID-19 Study). We construct our sample by matching COVID-19 Survey interviews to main survey interviews in this window, indicated in the figure by the vertical dashed lines. Each match gives us two sets of earnings and income measures to compare. As the Main Survey is annual, each study participant will provide at most one match in our data period, but we pool across COVID-19 Surveys to get a substantial sample of 1915 individuals. Because of the timing of fieldwork for COVID-19 and Main Surveys, most matches will involve a Main Study interview shortly *after* the COVID-19 Survey interview. This is useful, as one concern might be that completing the longer set of Main Survey questions first would improve the responses that are given in the COVID-19 Survey. To further guard against this possibility, below we report a robustness check in which we construct a sample from matches found only in the three weeks *after* the COVID-19 Survey interview.

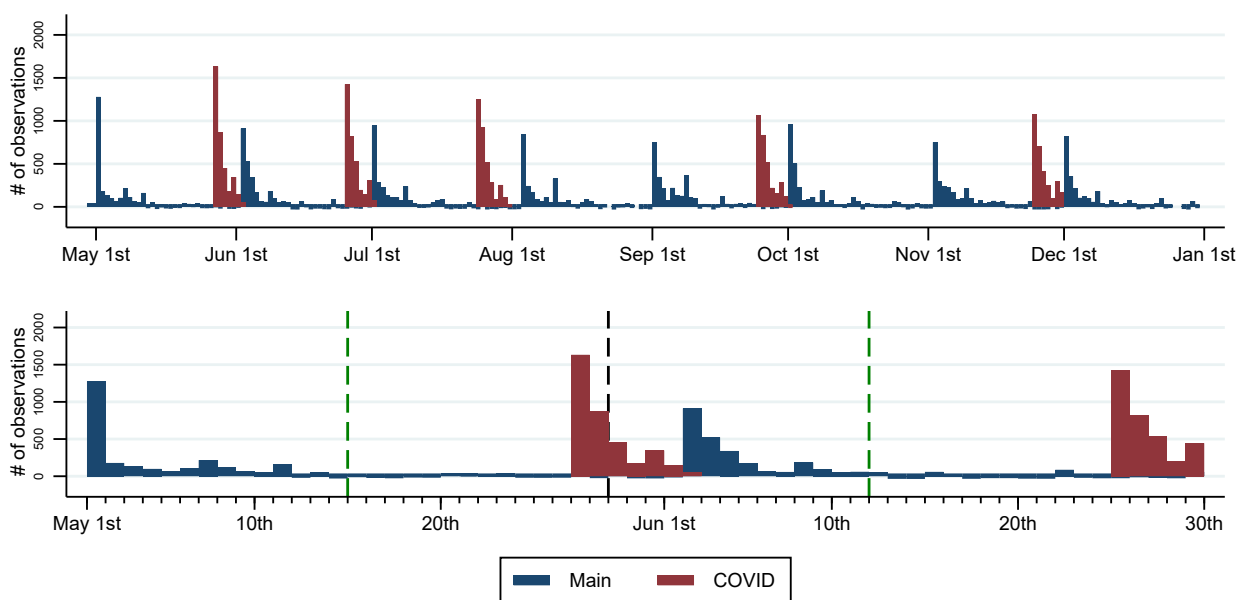


Figure 1: Interview dates: COVID-19 and Main study

Notes: Time is recorded on the horizontal axis and number of interviews completed on the vertical axis. The top panel covers the full sample period and the bottom panel focuses on the months May and June.

The Main Survey data are our validation data. The main survey data will of course contain measurement error due to misreporting. However, we note that recent validation studies (eg.

³In the rare cases where two Covid-19 interviews fall within 14 days of a Main study interview, we keep the nearest match.

Abowd and Stinson (2013); Bingley and Martinello (2017); Wilhelm (2018)) have documented measurement error in administrative income data, and we deal with measurement error in the validation sample using a similar econometric approach. An advantage of our design is that there is no problem of linkage error. Kapteyn and Ypma (2007) and Jenkins and Rios-Avila (2020) have shown that this can be a problem in validation studies involving linkage to external records.

We begin with a descriptive comparison of the two measures. We first examine the distributions of individual earnings, household earnings and household income in the two data sources. For many questions, such as whether the short- and long- measures give similar impressions of poverty or inequality, these comparisons of distributions are sufficient.

We then examine the distribution of individual *differences* across those two data sources for each measure. These differences will comprise measurement error from both sources, and so cannot be interpreted directly as measurement error in the single-question, Covid-19 Survey measures. Nevertheless, the magnitudes of differences are a useful starting point.

To go further, we examine the data through the lens of a standard measurement error model. The model closely follows Bingley and Martinello (2017).

Let y be the variable of interest (individual earnings, household earnings, or household income), with mean μ_y and variance σ_y^2 (this is the unobserved true value). Denote the measure in the COVID-19 Survey by y_c , and the measure in the main survey by y_m . We assume that y_c is linearly related to the true value y .

$$y_c = \mu_y + k_c + (1 + \rho_c)(y - \mu_y) + \epsilon_c \quad (1)$$

The parameter k_c captures under- or over-reporting at the mean ($y = \mu_y$), while ρ_c allows the error in y_c to be systematically related to the true value, y . The error term ϵ_c has mean zero, variance $\sigma_{\epsilon_c}^2$, and is independent of y . Note that if $\rho_c = k_c = 0$, the measurement error in y_c is classical. Consider estimating a bivariate regression with $x = \beta y + u$, where the usual regression assumptions hold, but the independent variable y is replaced by measure y_c . With this measurement error model, it is well known that (Bound et al., 2001):

$$\text{plim } \hat{\beta}^{OLS} = \beta \frac{(1 + \rho_c)\sigma_y^2}{(1 + \rho_c)^2\sigma_y^2 + \sigma_{\epsilon_c}^2} \quad (2)$$

and if an instrument z is available that is correlated with y but uncorrelated with ϵ_c

$$\text{plim } \hat{\beta}^{IV} = \beta \frac{1}{(1 + \rho_c)} \quad (3)$$

Finally, if the regression of interest is instead $y = x\alpha + v$, with y as the dependent variable, and measured by y_c :

$$\text{plim } \hat{\alpha}^{OLS} = \alpha(1 + \rho_c) \quad (4)$$

Thus, we can summarize the quality of y_c by the parameters ρ_c , k_c and σ_c^2 . The attenuation factors in Equations (2) through (4) suggest that an earnings or income measure is more useful the closer it's measurement error is to classical (and in particular, the closer ρ_c is to zero), and the smaller the variance of the classical measurement error component (σ_c^2).

To recover these parameters, we require two things. First, we assume that measurement error in the Main Study measures is classical:

$$y_m = y + \epsilon_m \quad (5)$$

where ϵ_m has mean zero, variance σ_m^2 , and is independent of y . Two lines of evidence support this assumption. First, as described in Section 2.1, income and earnings data from the Main Study has been validated against other high-quality sources. Second, more recent validation studies of survey earnings and income measures have failed to reject that the measurement error in those measures is classical. For example, Bingley and Martinello (2017) cannot reject that measurement error in earnings data from the Survey of Health Ageing and Retirement in Europe (SHARE) is classical. SHARE collects earnings and income with a detailed approach similar to that taken by the *Understanding Society* Main Study. The second requirement is that we have an instrumental variable that is correlated with y but uncorrelated with the measurement errors ϵ_c and ϵ_m . Then μ_y is identified by the mean of y_m , and given μ_y , k_c is identified by the mean of y_c . We can combine 1 and 5 to give:

$$y_c = \mu_y + k_c + (1 + \rho_c)(y_m - \mu_y) + \epsilon_c - (1 + \rho_c)\epsilon_m \quad (6)$$

where $(1 + \rho_c)$ can be estimated by IV. Given $(1 + \rho_c)$ the 2nd moments of y_m and y_c identify σ_y^2 , σ_m^2 and σ_c^2 . Following Bingley and Martinello (2017), we estimate $(\mu_y, k_c, \rho_c, \sigma_y^2, \sigma_m^2, \sigma_c^2)$ by Generalized Method of Moments (GMM).

We make use of well-measured Main Study variables as instruments. Our three household-level instruments are the number of cars the household owns or has access too; the number of

rooms in the home; and council tax liability. Council tax is a local tax and is assigned by allocating each residential property to one of eight bands based on the property value. While the first two variables are reported by a single respondent for the whole household, the latter is linked (by the data producers) to official council tax information, where the household postcode is the linking variable. For individual earnings, we use either an average of 2 lags of individual earnings or a small set of common wage predictors: age; gender; and the highest educational qualification achieved.

2.3 Earnings and income measures

The Main Study questionnaire has been better optimised around earnings and income data collection where the different components of income are collected in distinct survey modules (employee's; self-employment; second jobs; unearned income and state benefits; and household finances) with the aim of maximising response and data quality. In contrast, the short length of the COVID-19 interviews necessitates a compromise, and so the questions sit within a general 'employment module' that asks about various aspects of employment, earnings and income in one place.

Other features of the Main Study would suggest it produces data of higher quality than the COVID-19 study. First, in-interview respondent help notes were much more limited in the COVID-19 interviews, as the latter were optimised for completion by smartphone or similar device. While 67% of interviews were by smartphone in the COVID-19 study, only 33% of interviews were on a similar device in the main study. Second, to improve recall, respondents are encouraged to check relevant documents like payslips, in the Main Study. Third, the Main Study uses in-interview tools to improve reporting. Dependent interviewing reminds survey respondents of their reports at the previous interview, with the aim of reducing spurious change between waves. Non-response is reduced with follow-ups that prompt for reports where a respondent initially refused to answer a question. Fourth, while respondents are familiar with the reoccurring structure and questions of the main study, having participated in previous interviews, the COVID-19 Study is new.⁴

We construct three earnings and income variables for each study: individual earnings in main job, total household earnings, and total household income. For the Main Study, we work with the publicly available derived household earnings and income variables from the data producers. For the COVID-19 Study, we use the amounts reported on the short set of earnings and income

⁴Both surveys make use of in-interview soft-checks that, for example, notify a respondent if they have reported a gross amount less than a net amount.

questions. While for the Main Study, household totals are arrived at by summing over an individual's own income sources and then across individuals, in the COVID-19 study, each individual reports on household totals directly. We always work with the variables that are asked net of relevant taxes.⁵ The reference periods for each study broadly align. The main study asks about income around the time of the interview, with some exceptions described below, the COVID-19 study always asks 'now'. The respondent chooses the exact period code to report on in both surveys (week, two week, month or year in the COVID-study with a slightly larger set of options in the Main Study). The majority of respondents choose a monthly period code. For example, in the COVID-19 Study, this option was chosen by 80% of respondents for individual earnings; 74% for household earnings and 67% for household income. We always report amounts converted to monthly equivalents.

Our definition of individual earnings is employee pay in the main job. Both surveys record this amount with a similar question. We exclude self-employees from our definition, as the reference periods do not match across the surveys.⁶

Our definition of household earnings includes employee pay from all jobs, including self-employment profit and earnings from any second jobs. For the Main Study, this is calculated as the sum of reported main job, profits and second job earnings across all household members. For the COVID-19 study, we take this amount from the question asking about total earnings of the household ie. including second jobs and self-employee profit.⁷

Our definition of household income covers all sources of individual incomes including household earnings as above plus social security benefits, state and private pensions, private transfers and investment income. While the global household income figure is collected directly in the COVID-19 study, the Main Study asks separately about each of 41 sources of benefits and unearned income, where the reference period is the 'last payment received'. Investment income is asked about for the 'last 12 months'. The exact wording of the earnings and income questions for both studies are provided in the Supplemental Appendix.

We never work with imputed earnings or income data in our analysis, even though they are supplied by the data providers. Instead, observations with item-missing data on a source are removed for that source. Our overall distributional comparisons apply survey weights to correct

⁵Some minor conversion from gross to net does take place in the Main Study ie. second jobs is reported gross, but the data producers translate the amount to a net equivalent by tax simulation.

⁶In the Main Study, self-employees receive a distinct set of questions about profits on last years accounts, whereas in the COVID-19 study they receive the same pay question as employees asking about pay 'now'.

⁷This does mean that our household earnings comparisons will suffer from the definitional difference for the self-employed stated above. However, the self-employed correspond to only 6.8% of our sample (COVID-19).

for item non-response, survey non-response and survey design. Weights match each sample to Wave 11 main survey totals defined by gender, age and education.

3 Results

3.1 Item missingness

One benefit of quick income questions is that they may suffer less from item-missingness when compared to a detailed set of questions that go source by source and household member by household member. For detailed question sets, non-response cumulates over items, or where some household members decline an interview so that household totals cannot be constructed – we refer to the latter as ‘incomplete households’. This is exactly what we see in Table 1 where item-missingness is less prevalent in the COVID-19 Study, relative to the Main Study. The difference is largest for earnings at the household, rather than the individual level, and for multiple, rather than single adult households.

Overall missing rates for individual earnings, household earnings and household income, respectively, are 13%, 38%, and 49% in the Main Study and 10%, 13% and 24% in the COVID-19 Study. We count an observation as item missing where an individual declines to answer the corresponding survey question or a question it is routed on; or where an individual resides in a household that is ‘incomplete’ for the household concepts in the Main Study. The right half of Table 1 shows that the difference in missingness holds even when we move to individuals from complete households in the Main Study. As expected, the differences are smaller in this subsample (Main Study: 10%, 12%, and 23%; COVID-19 study: 13%, 22%, 38%). Our reported missing data numbers are comparable to those seen in similar household surveys. Note, the sample sizes for individual earnings marginally differ across the studies as they include only those reporting paid work in the corresponding study. Also, the sample sizes for household earnings are smaller than those for household income, as the former is restricted to individuals of working age.

Table 1 also shows that the difference in item missing rates are larger for couples than for singles, and the missingness gap is largest for the household concepts and particularly large for household income. This reflects the fact that household income consists of more subcomponents than household earnings, and so there is a greater chance that at least one subcomponent is missing. For example, the missing rate for individuals in multiple adult households is 41 percent in the Main Study, compared to only 23 percent for the COVID-19 study (these numbers are for complete households).

	All respondents				Respondents in complete hhs.			
	(1) Main %	(2) N	(3) COVID %	(4) N	(5) Main %	(6) N	(7) COVID %	(8) N
All individuals								
Ind. earnings	13	(926)	10	(960)	13	(737)	10	(770)
HH. earnings	38	(1438)	13	(1438)	22	(1144)	12	(1144)
HH. income	49	(1959)	24	(1959)	38	(1594)	23	(1594)
Single adult hhs.								
Ind. earnings	6	(140)	7	(145)	6	(140)	7	(145)
HH. earnings	8	(225)	8	(225)	8	(225)	8	(225)
HH. income	29	(337)	20	(337)	29	(337)	20	(337)
Multiple adult hhs.								
Ind. earnings	14	(786)	11	(815)	14	(597)	11	(625)
HH. earnings	44	(1213)	14	(1213)	26	(919)	13	(919)
HH. income	53	(1622)	25	(1622)	41	(1257)	23	(1257)

Table 1: Prevalence of item-missing data (percent)

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview. In the case of multiple matches the nearest match is kept. A ‘complete household’ is one in which all adult members complete a main study individual interview and households not meeting this condition are recorded as ‘incomplete households’. ‘Single adult households’ refers to individuals living in households with no other adults age 16 or over as reported in the main survey; and ‘Multiple adult households’ refer to individuals in households with more than one adult aged 16 or over. Individual earnings are conditional on reporting paid employment as an employee in the relevant survey (COVID-19 or main). Individual and household earnings are reported for the sample of working age (age<66).

3.2 Descriptive comparison

We now compare the earnings and income distributions from each survey. Figure 2 shows the cumulative distributions (CDF) for individual earnings, household earnings and household income, while Table 2 shows selected percentiles as well as the Gini coefficient. We drop from the analysis samples cases which suffer item-missing for a given concept, and then apply calibration weighting to adjust for selection into the analysis samples. Weights are constructed to match each survey-specific earnings and income concepts to Wave 11 main survey totals defined by gender, age and education.⁸ All sources are expressed in levels.

Looking at the distributions, the most striking feature is a high degree of agreement for all concepts. Individual earnings are most similar, as the CDFs for the two surveys overlap, and there is very little difference in the selected percentiles or Gini in Table 2. The comparison differs for household earnings and household income, where the surveys disagree somewhat. Here, the COVID-19 Study estimates are consistently lower than the equivalents from the Main Study, and more notably so for household income. For example, the estimated medians are about 90 percent of the main figures for household earnings and 87 percent for household income. Also,

⁸In robustness checks, we also performed an unweighted analysis and the main conclusions of the paper all hold.

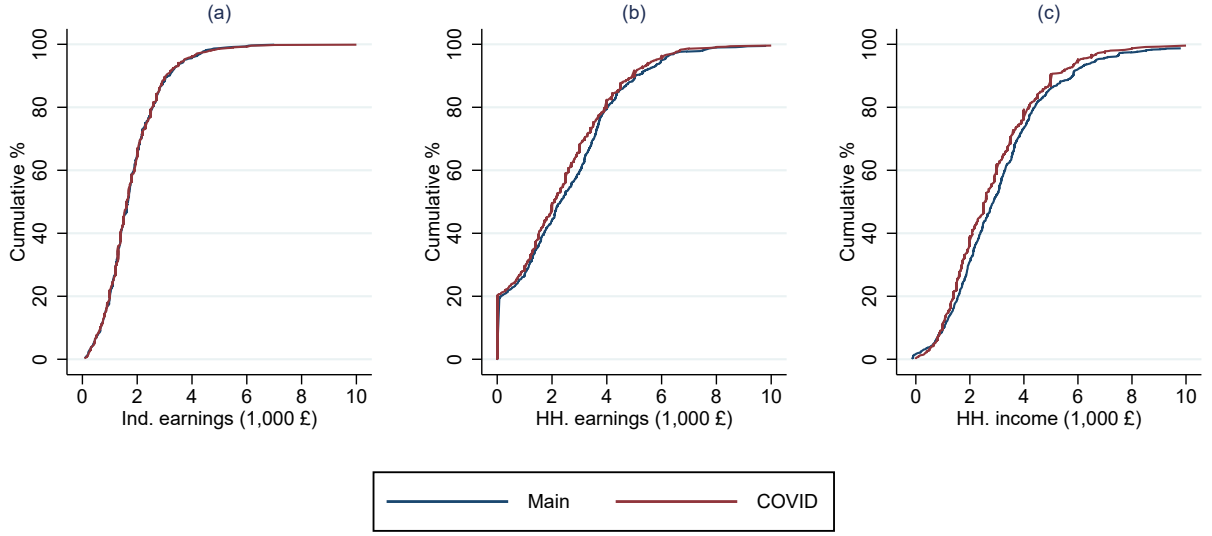


Figure 2: Income and earnings CDFs

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in the relevant survey (COVID-19 or main). Individual and household earnings are reported for the sample of working age (age < 66). Observations with item-missing data on an income source are removed for that source. Weights correct for item non-response, survey non-response and survey design. Weights match each (income source and survey specific) sample to Wave 11 main survey totals defined by gender, age and education. Sample sizes as in panel A of Table 2.

the COVID-19 study gives a slightly higher estimate of the Gini for household earnings (45.2 to 45.7) and household income (33.4 and 34.6). The similarity in individual earnings measures, and apparent difference of household earnings and income, is not completely surprising. First, both surveys ask similar individual earnings questions and second, individual earnings can be arrived at without respondents needing to aggregate sources or have knowledge of other household members finances. A different interpretation is that the lower item-non response rate of the COVID-19 study leads to differing compositions across the surveys, and consequently distributional estimates differ. The bottom panel of Table 2 presents evidence against the latter interpretation by focussing on the sample of individuals that report a given source in both surveys.

Figure 3 presents a related analysis by plotting the rank in one survey against the rank in the other, for each of our earnings and income measures. Income ranks are used in various empirical literatures such as intergenerational mobility (e.g. Chetty et al. (2014)), and are unaffected by monotonic transformation (such as from levels to logarithms), motivating the comparison. While

Percentile	Ind. Earnings			HH. Earnings			HH. Income		
	(1) Main	(2) COVID	(3) Ratio (2)/(1) x100	(4) Main	(5) COVID	(6) Ratio (5)/(4) x100	(7) Main	(8) COVID	(9) Ratio (7)/(8) x100
Panel A: All individuals									
5	403	430	107	0	0	.	667	624	94
10	662	650	98	0	0	.	1054	867	82
25	1127	1150	102	798	867	109	1795	1500	84
50	1660	1624	98	2315	2100	91	2877	2500	87
75	2300	2343	102	3707	3500	94	4094	3870	95
90	3182	3059	96	5056	5000	99	5808	5000	86
95	3737	3716	99	6000	6000	100	6686	6344	95
Gini	30.80	31.50		45.20	45.70		33.40	34.60	
Weighted	Yes	Yes	-	Yes	Yes	-	Yes	Yes	-
N	644	694		890	1003		984	1233	
Panel B: No item-missing data in both surveys (source specific)									
5	457	450	98	0	0	.	779	700	90
10	700	700	100	0	0	.	1117	1000	90
25	1127	1144	102	763	700	92	1833	1540	84
50	1643	1600	97	2279	2100	92	2906	2600	89
75	2300	2337	102	3690	3500	95	4094	3980	97
90	3053	3033	99	5025	5000	100	5796	5000	86
95	3700	3716	100	5955	5800	97	6686	6000	90
Gini	30.20	30.80		45	46.60		32.30	33.20	
Weighted	Yes	Yes	-	Yes	Yes	-	Yes	Yes	-
N	616	616		838	838		846	846	

Table 2: Income and earnings distributions

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview, where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches, the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in the relevant survey (COVID-19 or main). Household earnings refers to working-age individuals (age<66). Observations with item-missing data on an income source are removed for that source. Panel B further restricts the sample to individuals with non-missing data on a source in both surveys. Weights correct for item non-response, survey non-response and survey design. Individual and household earnings are reported for the sample of working age (age<66). Weights match each (income source and survey specific) sample to Wave 11 main survey totals defined by gender, age and education.

most points in the rank-rank plots are close to the 45 degree line - indicating a high degree of agreement between the surveys - there are noticeable deviations, which are strongest for the household measures. Spearman's rank correlation is high for individual earnings (0.94); and lower for household earnings (0.8) and household income (0.79).

Figure 4 examines the distribution of the *differences* between the surveys (COVID-19 minus Main). To focus on percentage differences, we take the natural logarithm of each income source before differencing. We present CDFs for the differences and also the absolute differences. Several features are notable. First, there is a large mass of zeros for all three measures. Second, many of the differences fall both above and below zero, but more fall below than above for the household

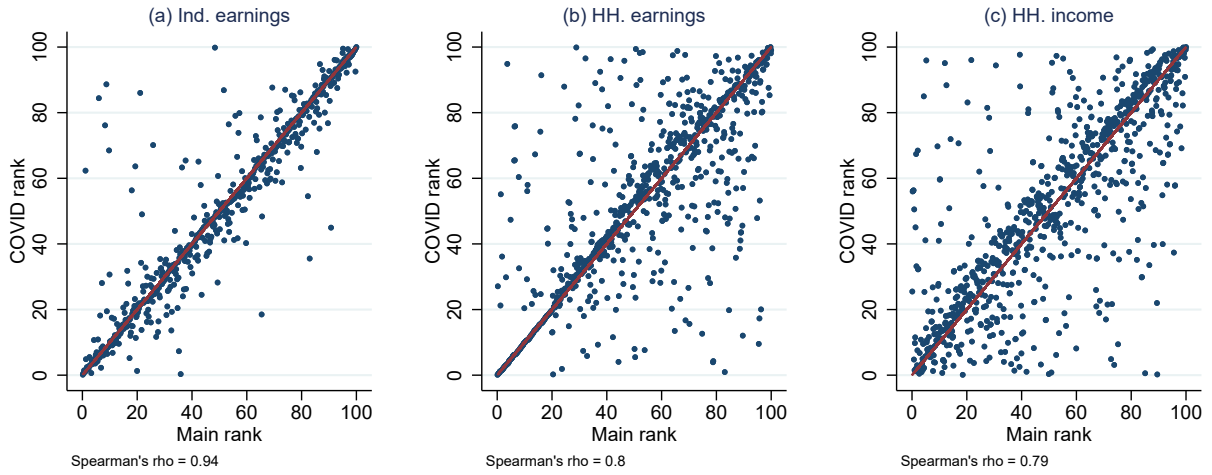


Figure 3: COVID-19 study rank vs. Main study rank

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Household earnings refers to working-age individuals (age < 66). Observations with item-missing data on an income source are removed for that source. Individual and household earnings are reported for the sample of working age (age < 66). Ranks are the percentile rank of each source in the corresponding survey. Sample sizes are 616 (individual earnings), 838 (household earnings) and 846 (household income).

measures (pointing to possible under-reporting in the COVID-19 Study). Individual earnings has the smallest mass below zero at 35 percent, but the same figure for the household measures is around 55 percent. For each of our measures, around 35 percent of differences are greater than zero. Third, the absolute differences show fewer and small differences for individual earnings compared to the household measures.

The evidence of this section is that both surveys produce similar data, although the short set of questions of the COVID-19 survey give lower estimates of household earnings and income, and marginally higher estimates of inequality as measured by the Gini coefficient. We also see that the same individuals tend to report lower income in the COVID-19 study than in the main study, although again only for the household measures. A standard interpretation of the findings is that the short question sets of the COVID-19 study tend to under-record household earnings and income, as the main study data is collected with gold standard interviewing methods. In the next section, we allow for the possibility of errors in both surveys and estimate directly their

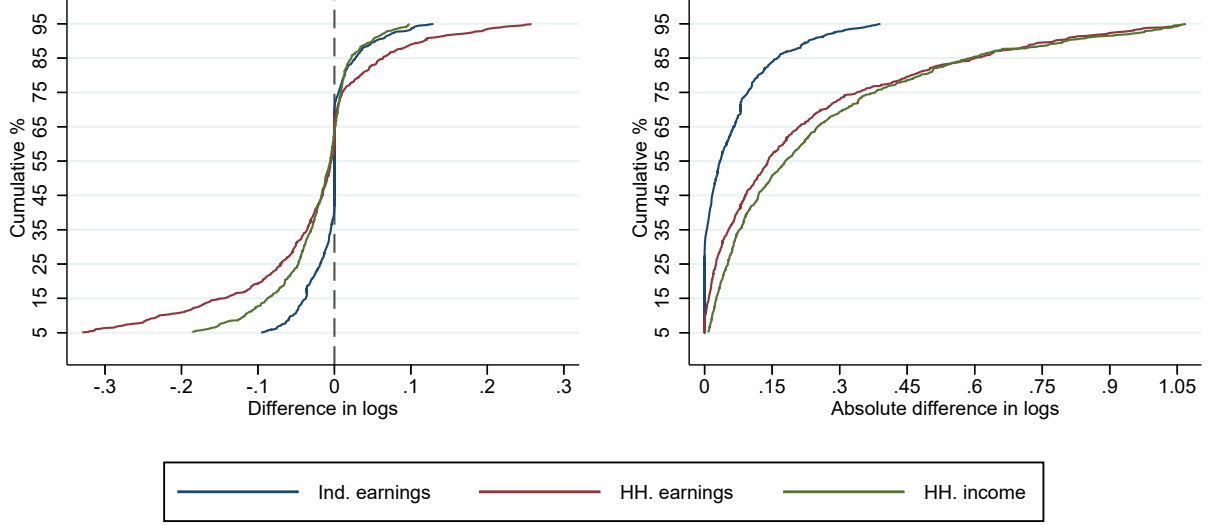


Figure 4: CDFs of differences in reported earnings and income (COVID-19 - Main)

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Individual and household earnings are reported for the sample of working age (age<66). Observations with item-missing data on an income source are removed for that source. Differences are constructed as: $\log(\text{COVID-19 study amount}) - \log(\text{main study amount})$.

measurement error properties.

3.3 GMM estimates of a measurement error model

We now turn to estimates of the measurement error model outlined in Section 2.2. For individual log earnings, we compare two sets of instruments. The first is an average of lagged earnings (from previous waves of the Main Study). The second is a set of human capital variables: gender, age, education. The results are presented in Table 3. Both sets of instruments are strongly correlated with reported earnings in the Main Study, as indicated by the First Stage F-statistic. Results are very similar for both instrument sets.

The estimates of ρ_c and k_c are very close to zero, indicating that measurement error in the COVID-19 earnings measure is not related to the true value, and that it does not suffer from systematic under-reporting. Note that the estimates do reject the null of no measurement error variance in the Main Study, so we focus on the full model that allows for classical measurement error in validation data (columns (2) and (4)). Comparing the estimates of measurement error variance (σ_{2c} and σ_{2m}) suggests that the variance of the measurement error in the COVID-19 earnings mea-

sure is about 50% larger than in the Main Study. However, estimates of classical measurement error variances for both studies are an order of magnitude smaller than the estimates of the true variance of earnings (σ_{2y}). This means that the implied attenuation factors (Equation (2)) are above 0.9 for both measures. Thus, by this summary statistic, the two measures of individual earnings are of very similar quality.

	Lagged earnings		Human capital	
	(1)	(2)	(3)	(4)
ρ_c	0.00000** (0.000)	0.00000 (0.000)	-0.00000*** (0.000)	-0.00000*** (0.000)
σ_y^2	0.318*** (0.024)	0.308*** (0.025)	0.370*** (0.028)	0.359*** (0.029)
σ_c^2	0.033** (0.014)	0.033** (0.014)	0.031*** (0.010)	0.024** (0.011)
σ_m^2		0.025** (0.012)		0.014* (0.007)
μ_y	7.433*** (0.026)	7.423*** (0.026)	7.384*** (0.025)	7.382*** (0.025)
k_c	-0.005 (0.009)	0.012 (0.012)	-0.008 (0.006)	-0.002 (0.007)
Observations	485	485	610	610
First stage F-stat	738.931	738.931	34.028	34.028
Attenuation factor				
Main		0.92		0.96
COVID	0.91	0.90	0.92	0.94

Table 3: GMM estimates for individual earnings

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Individual and household earnings are reported for the sample of working age (age<66). Attenuation factors are calculated from estimates according to Equation (2). Instruments: the average of lagged earnings in (1)-(2) and Education, gender, and age in (3)-(4).

Table 4 presents estimates of the same model, but for *household* earnings and income. Here there is greater scope for quality differences because the Main Study measure aggregates over the reports of multiple household members, and, in the case of household income, over responses to questions about different categories of income. Here we use a single instrument set containing a number of well-measured variables: council tax amount, the number of cars owned by the household, and the number of rooms in the home.

For household earnings and income, estimates of ρ_c are still very close to zero (albeit statistically significant at conventional levels) indicating that measurement error in the COVID-19 measures is only very weakly related to the true value. Estimates of k_c capture the significant under-reporting of 10-15 % show in Section 3.2. While this clearly matters for means, totals or the measurement of poverty, as noted in Section 2.2 it does not affect regression coefficients.

	HH. earnings		HH. income	
	(1)	(2)	(3)	(4)
ρ_c	-0.00001*** (0.000)	-0.00000 (0.000)	-0.00002*** (0.000)	-0.00001*** (0.000)
σ_y^2	0.420*** (0.038)	0.402*** (0.038)	0.379*** (0.028)	0.372*** (0.028)
σ_c^2	0.209* (0.120)	0.336*** (0.123)	0.310*** (0.103)	0.315*** (0.103)
σ_m^2		0.101*** (0.023)		0.143*** (0.043)
μ_y	7.883*** (0.027)	7.855*** (0.028)	7.933*** (0.023)	7.904*** (0.025)
k_c	-0.096*** (0.026)	-0.092*** (0.026)	-0.163*** (0.022)	-0.137*** (0.023)
Observations	636	636	835	835
First stage F-stat	31.972	31.972	64.901	64.901
Attenuation factor				
Main		0.80		0.73
COVID	0.67	0.54	0.55	0.54

Table 4: GMM estimates for household earnings and income

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Individual and household earnings are reported for the sample of working age (age < 66). Attenuation factors are calculated from estimates according to Equation (2). Instruments: Log(council tax amount), number of cars in the household, and number of rooms in the house.

Again, the estimates do reject the null of no measurement error variance in the Main Study measures, so we focus on estimates of the full model that allows for classical measurement error in validation data (columns (2) and (4)). Comparing the estimates of measurement error variance (σ_{2c} and σ_{2m}) suggests that the variance of the classical measurement error in the COVID-19 household earnings measure is three times larger than in the Main Study. For household income, the ratio is two to one. For both household earnings and income, the measures are noisier (in the sense that the ratio of the true variance to the measurement error variance is smaller) than for individual earnings. The consequence is that implied attenuation factors are substantially below one for both the Main Study and the COVID-19 study, and the measurement error in the COVID-19 Study implies substantially more attenuation than the Main Study. For household earnings, the attenuation factor (Equation (2) of the COVID-19 Study is 0.54, as opposed to 0.80 for the Main study. The corresponding numbers are 0.54 and 0.73 for household income.

A possible concern with our research design is that completing the Main Study interview *before* the COVID-19 survey might improve responses to the shorter COVID-19 question set. This would make our findings unrepresentative of what might be expected of short question sets in general. As noted in Section 2.2, for the majority of the observations in our analysis sample, the Main

Study interview followed the COVID-19 Study interview. Figure 5 plots the difference between Main Study COVID-19 Study (y-axis) against the difference in timing (Main Study date - COVID-study data, x-axis). When the Main Study interview followed the COVID-19 Study, this difference is positive, so such observations are to the right of the dashed vertical line at 0. We do this for individual earnings (Panel (a)), household earnings (Panel (b)) and household income (Panel (c)). In all three figures, large differences in the two measures are less frequent to the left of the dashed vertical line at 0. This suggests that completing the more detailed Main Study interview before the COVID-19 survey may have affected responses to the shorter COVID-19 question set.

Therefore, as a robustness check, we created a second analysis sample, comprising only cases where the Main Study interview followed the COVID-19 Study survey, within a 21-day window. We then re-estimated our measurement error models for household earnings and household income. The results are presented in Table 5. Comparing Table 5 and Table 4, the substantive results are very similar.



Figure 5: Differences in income reports and days between interviews

Notes: Sample of individuals completing a COVID-19 survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Individual and household earnings are reported for the sample of working age ($age < 66$). Differences are constructed as: $\log(\text{COVID-19 survey amount}) - \log(\text{main survey amount})$. The x-axis shows the number of days between the main wave 11 interview and COVID-19 interview where a positive value indicates the main interview took place after the COVID-19 interview. Each dot represents one observation.

	HH. earnings		HH. income	
	(1)	(2)	(3)	(4)
ρ_c	-0.00001*** (0.000)	-0.00000* (0.000)	-0.00002*** (0.000)	-0.00001** (0.000)
σ_y^2	0.424*** (0.039)	0.401*** (0.039)	0.367*** (0.026)	0.356*** (0.026)
σ_c^2	0.192 (0.127)	0.322** (0.130)	0.300*** (0.107)	0.304*** (0.107)
σ_m^2		0.109*** (0.025)		0.141*** (0.044)
μ_y	7.896*** (0.029)	7.869*** (0.030)	7.945*** (0.023)	7.919*** (0.025)
k_c	-0.108*** (0.026)	-0.105*** (0.026)	-0.161*** (0.022)	-0.135*** (0.024)
Observations	593	593	795	795
First stage F-stat	18.096	18.096	49.245	49.245
Attenuation factor				
Main		0.79		0.72
COVID	0.69	0.55	0.55	0.54

Table 5: GMM Estimates; Alternative Sample

Notes: Sample of individuals completing a COVID-19 survey interview within 21 days before their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Individual and household earnings are reported for the sample of working age (age < 66). Attenuation factors are calculated from estimates according to Equation (2). Instruments: log(council tax amount), number of cars in the household, and number of rooms in the house.

4 Conclusion

Many social science surveys were fielded during the COVID-19 pandemic that collected earnings and income data with single or short question sets, at odds with best practise methods for income data collection. We have presented evidence on the reliability of data collected with such questions, using the *Understanding Society* COVID-19 Study as a test case. Our evidence derives from a quirk of the data collection: a large subset of respondents separately answered the best practise questions in their annual *Understanding Society* interview around the time that they participated in the COVID-19 Study. The data collected in the annual interview can therefore act as a validation data source.

We find that the short questions produce data on individual earnings that is of comparable quality to the detailed questions in the Main Study annual interview. In contrast, for household earnings and income, the detailed set of questions produces measures that are less noisy, and the COVID-19 Study measures suffer from systematic under-reporting. This should be born in mind when, say, estimating poverty rates. However, there is little evidence that the measurement errors in the COVID-19 data are related to true values, suggesting that for regression analysis, an instrumental variables approach will be effective. We conclude that short question sets on earnings and income are useful content for short surveys or for longer surveys that prioritize content in other domains.

Lower respondent burden and lower cost are significant advantages of collecting earnings and income data with short question sets. In addition, as we have documented, short question sets suffer less from item non-response than do longer question sets. An implication of the latter point is that it may be useful to field such questions even in conjunction with longer question sets. How such combinations might be best designed and analysed is a question for future research.

Acknowledgements

Author order is alphabetical. *Understanding Society* COVID-19 Study is funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). Fieldwork for the survey is carried out by Ipsos MORI and Kantar. *Understanding Society* is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex. The research data are distributed by the UK Data Service. The authors thank Hamish Low for useful comments.

References

- Abowd, John M and Martha H Stinson**, “Estimating measurement error in annual job earnings: A comparison of survey and administrative data,” *Review of Economics and Statistics*, 2013, 95 (5), 1451–1467.
- Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh**, “Inequality in the impact of the coronavirus shock: Evidence from real time surveys,” *Journal of Public Economics*, 2020, 189, 104245.
- Belot, Michèle, Syngjoo Choi, Julian C. Jamison, Nicholas W. Papageorge, Egon Tripodi, and Eline van den Broek-Altenburg**, “Six-Country Survey on Covid-19,” 5 2020. Mimeo, available at <https://www.egontripodi.com/papers/covidintro.pdf>.
- Bingley, Paul and Alessandro Martinello**, “Measurement error in income and schooling and the bias of linear estimators,” *Journal of labor economics*, 2017, 35 (4), 1117–1148.
- Bound, John, Charles Brown, and Nancy Mathiowetz**, “Measurement error in survey data,” in “Handbook of econometrics,” Vol. 5, Elsevier, 2001, pp. 3705–3843.
- , —, **Greg J Duncan, and Willard L Rodgers**, “Evidence on the validity of cross-sectional and longitudinal labor market data,” *Journal of Labor Economics*, 1994, 12 (3), 345–368.
- Bundervoet, Tom, Maria E. Dávalos, and Natalia Garcia**, “The Short-Term Impacts of COVID-19 on Households in Developing Countries: An Overview based on a Harmonized Dataset of High-Frequency Surveys,” *World Development*, 2022, p. 105844.
- Burton, Jonathan, Peter Lynn, and Michaela Benzeval**, “How Understanding Society: The UK household longitudinal study adapted to the COVID-19 pandemic,” in “Survey Research Methods,” Vol. 14 2020, pp. 235–239.
- Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez**, “Where is the land of opportunity? The geography of intergenerational mobility in the United States,” *The Quarterly Journal of Economics*, 2014, 129 (4), 1553–1623.
- Crossley, Thomas F, Paul Fisher, and Hamish Low**, “The heterogeneous and regressive consequences of COVID-19: Evidence from high quality panel data,” *Journal of Public Economics*, 2021, 193 (C).
- Fisher, Paul**, “Does Repeated Measurement Improve Income Data Quality?,” *Oxford Bulletin of Economics and Statistics*, 2019, 81 (5), 989–1011.
- , **Laura Fumagalli, Nick Buck, and Silvia Avram**, “Understanding Society and its income data,” 2019, (2019-08).
- Gottschalk, Peter and Minh Huynh**, “Are earnings inequality and mobility overstated? The impact of nonclassical measurement error,” *The Review of Economics and Statistics*, 2010, 92 (2), 302–315.
- Institute for Social and Economic Research**, *Understanding Society COVID-19 Study*, 2020. [data collection]. 1st Edition, UK Data Service. SN: 8644, 10.5255/UKDA-SN-8644-1, 2020.
- , *Understanding Society COVID-19 User Guide. Version 1.0*, University of Essex, Colchester, October 2021.

—, *Understanding Society: Waves 1-11, 2009-2020 and Harmonised BHPS: Waves 1-18, 1991-2009, UserGuide*, University of Essex, Colchester, 2021.

Jenkins, Stephen P and Fernando Rios-Avila, “Modelling errors in survey and administrative data on employment earnings: Sensitivity to the fraction assumed to have error-free earnings,” *Economics Letters*, 2020, 192, 109253.

Kapteyn, Arie and Jelmer Y Ypma, “Measurement error and misclassification: A comparison of survey and administrative data,” *Journal of Labor Economics*, 2007, 25 (3), 513–551.

Meyer, Bruce D and Nikolas Mittag, “Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net,” *American Economic Journal: Applied Economics*, 2019, 11 (2), 176–204.

Micklewright, John and Sylke V Schnepf, “How reliable are income data collected with a single question?,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2010, 173 (2), 409–429.

Pischke, Jörn-Steffen, “Measurement error and earnings dynamics: Some estimates from the PSID validation study,” *Journal of Business & Economic Statistics*, 1995, 13 (3), 305–314.

United Nations, *Canberra Group Handbook on Household Income Statistics*, second edition ed., New York and Geneva: United Nations, 2011.

Ward, Jason M and Kathryn Anne Edwards, “CPS Nonresponse During the COVID-19 Pandemic: Explanations, Extent, and Effects,” *Labour Economics*, 2021, 72, 102060.

Wilhelm, Daniel, “Testing for the presence of measurement error,” Technical Report, cemmap working paper 2018.

Supplemental Appendix

A. Questionnaire

A.1 Covid-19 study questions

semp [Employee or self-employed]

Universe: IF ff_prevsurv = 0 OR (ff_prevsurv = 1 AND sempchk is not 1) // Ask if has not completed any previous monthly surveys, or has completed previous surveys and has not confirmed previous employment status as still applying.

Source: UKHLS covid-19 survey

Text: Thinking about your situation now. Even if you did not do any paid work last week, are you currently employed or self-employed?

1. Yes, employed only
2. Yes, self-employed only
3. Both employed and self-employed
4. No

netpay_amount [Current earnings amount]

Universe: IF sempderived = 1, 2, 3 // Ask if currently in paid work or self-employment.

Source: UKHLS covid-19 survey

Scripting notes: Range [0 – 1,000,000].

Text: What is your usual **take-home pay/earnings** now? Take-home pay is after tax, National Insurance and pension contributions have been deducted. Please include all jobs and self-employment activities.

[Numeric textbox] Pounds

netpay_period [Current earnings period]

Universe: IF sempderived = 1, 2, 3 // Ask if currently in paid work or self-employment.

Source: UKHLS covid-19 survey

Scripting notes: Display on same page as netpay_amount.

Text: Per

1. Week
2. Two weeks
3. Month
4. Year

hhearners [Current household earners]

Universe: IF hhcompc > 0 OR hhcompd > 0 OR hhcompe > 0 // Ask if living with others aged 16-18, 19-69, or 70+.

Source: UKHLS covid-19 survey

Text: Thinking about the other people living with you at the moment, are any of them employed or self-employed (even if they did not do any paid work last week)?

1. Yes
2. No

hhearn_amount [Current household earnings amount]

Universe: IF (hhcompc > 0 OR hhcompd > 0 OR hhcompe > 0) AND hhearners = 1 // Ask if living with others aged 16-18, 19-69, or 70+, and at least one other earner in household.

Source: UKHLS covid-19 survey

Scripting notes: Range [0 – 2,500,000].

Soft check: IF (hhearn_period = 1 AND hhearn_amount < netpayweek) OR (hhearn_period = 2 AND hhearn_amount/2 < netpayweek) OR (hhearn_period = 3 AND hhearn_amount/4.33 < netpayweek) OR (hhearn_period = 4 AND hhearn_amount/52.14 < netpayweek) // Household earnings converted to weekly amount are less than weekly net pay.

Soft check text: “Your household earnings are lower than your net pay.”

Text: Thinking about everyone living with you at the moment, what is the **total take-home pay/earnings** of your household now? Please only include earnings from paid work or self-employment. If you are not sure, please tell us an approximate amount.

[Numeric textbox] Pounds

hhearn_period [Current household earnings period]

Universe: IF (hhcompc > 0 OR hhcompd > 0 OR hhcompe > 0) AND hhearners = 1 // Ask if living with others aged 16-18, 19-69, or 70+, and at least one other earner in household.

Source: UKHLS covid-19 survey

Scripting notes: Display on same page as hhearn_amount.

Text: Per

1. Week
2. Two weeks
3. Month
4. Year

hhincome_amount [Current household income amount]
--

Universe: Ask all.

Source: UKHLS covid-19 survey

Scripting notes: Range [0 – 2,500,000].

Soft check: IF (hhincome_period = 1 AND hhincome_amount < hhearnweek) OR (hhincome_period = 2 AND hhincome_amount/2 < hhearnweek) OR (hhincome_period = 3 AND hhincome_amount/4.33 < hhearnweek) OR (hhincome_period = 4 AND hhincome_amount/52.14 < hhearnweek) // Household income converted to weekly amount is less than weekly household earnings.

Soft check text: “Your household income should include earnings for all household members (after tax) added with other sources of income like benefits, pensions and earnings from investments on top. Normally it would be more than household earnings.”

Text: Many people have additional sources of income beyond earnings from paid work and self-employment. Thinking about everyone living with you at the moment, what is the **total take-home/after tax income** of your household now? Please include **all** sources of income, such as benefits, pensions and earnings from investments, as well as earnings from paid work or self-employment. If you are not sure, please tell us an approximate amount.

[Numeric textbox] Pounds

hhincome_period [Current household income period]
--

Universe: Ask all.

Source: UKHLS covid-19 survey

Scripting notes: Display on same page as hhincome_amount.

Text: Per

1. Week
2. Two weeks
3. Month
4. Year

A.2 Main study questions

Currentemployment_w11. **Jbhas.** *Did paid work last week*

Type	Don't Know	Refused	Inapplicable	Missing
choice	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

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

Options

1	Yes	Yes
2	No	No

Help

Include:

*Employment for any number of hours, including Saturday jobs, and casual work, e.g. baby sitting, running mail-order clubs etc. NB Since some informants may not consider casual work to be serious work, please be prepared to probe those (e.g. housewives with dependent children, full-time students) to whom you feel this may apply.

*Anyone who was paid a wage or salary by an employer while attending an educational establishment.

*Someone working for any number of hours in a spouse's business as long as they were paid, i.e. received (or will receive) an amount of money in remuneration or a share of the profits.

*Someone working unpaid in a spouse's business provided they work 15 hours or more per week.

*Anyone else working in a friend's or relative's business, as long as he or she received (or will receive) an amount of money in remuneration or a share of the profits.

*Students on 'sandwich courses' currently with an employer or those on employer placements should be coded as employed only if they receive regular payment from the employer.

Student nurses: Nurses being trained under the PROJECT 2000 scheme which is mainly based in colleges outside hospitals should be regarded as full-time students. Unless they have other employment outside their PROJECT 2000 course they should be coded "No" at JBHAS and JBOFF. For student nurses always check if the nursing training course is part of PROJECT 2000 and proceed accordingly. If the respondent is unsure treat as employed.

Government Training Schemes: If a respondent was on such a scheme last week, they should be coded as employed if they are 'employer based'. This is defined as: 1) Last week they were with an employer or on a project providing work experience or practical training or 2) If they are normally 'employer based' but were away last week because of illness or for some other reason (including attending a college based course). If respondents are not normally 'employer based' that is they usually attend a college or other training centre, they should be coded "No" at JBHAS and JBOFF with the following exceptions: 1) those who have employment outside the training scheme. Code here as employed and take details of this secondary employment. 2) those sent on training schemes by employers who continue to pay their wages or salaries. Treat as employed and code "Attending training course" at JBOFFY. Record details of their job with the employer who has seconded them to the training scheme.

Mixed Mode Alternatives

Web Interview

Text

Can we just check, did you do any paid work **last week** - that is in the seven days ending last Sunday - either as an employee or self-employed? Include employment for any number of hours, including casual work and Saturday jobs.

Help

Answer yes if:

*You were paid a wage or salary by an employer while attending an educational establishment.

*You worked some paid hours in your spouse's business i.e. received (or will receive) an amount of money in remuneration or a share of the profits.

*You worked unpaid in your spouse's business provided you work 15 hours or more per week.

*You worked in a friend's or relative's business, as long as you received (or will receive) an amount of money in remuneration or a share of the profits.

*You are a student on a 'sandwich course' currently with an employer, as long as you receive regular payment from your employer.

*You were on an 'employer based' Government Training Scheme.

Currentemployment_w11.Jboff. *No work last week but has paid job*

Type	Don't Know	Refused	Inapplicable	Missing
choice	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

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

Options

1	Yes	Yes
2	No	No
3	Waiting to take up job	Waiting to take up job

Help

Include any persons who were absent because of holiday, strike, sickness, maternity leave, furlough, unpaid leave, lay-off or similar reason, provided they have a job to return to with the same employer. Do not include those receiving redundancy payments who have no job to return to. Do not include people who have a job arranged but have not yet started work in it: such people should be coded "Waiting to take up job" at JBOFF. A job exists if there is a definite arrangement for work on a regular basis (i.e. every week or every month) whether this work is full-time or part-time.

Mixed Mode Alternatives

Web Interview

Help

*A job exists if there is a definite arrangement for work on a regular basis (i.e. every week or every month) whether this work is full-time or part-time.

*Answer yes if you were absent due to holiday, strike, sickness, maternity leave, furlough, unpaid leave, lay-off or similar reason, provided you have a job to return to with the same employer.

*Answer no if you are receiving redundancy payments and have no job to return to.

Currentemployment_w11.Jbsemp. *Employed or self-employed: current job*

Type	Don't Know	Refused	Inapplicable	Missing
choice	-1	-2	-8	-9

Source

BHPS

Version

1.0

Scripting Notes

If JBSEMPCHK = 1, code JBSEMP as per ff_JBSEMP. We need to do this because we use JBSEMP in subsequent routing.

Text

Are you an employee or self-employed?

Options

1	Employee	Employee
2	Self-employed	Self-employed

Help

Where there is doubt, try to find out how they are described for tax and National Insurance purposes. For self-employed, tax is not normally deducted at source but paid directly to the Inland Revenue. People working as child minders or odd-jobbing etc are usually classified as self-employed. If someone is self employed and works under contract to an employer, he or she may be treated as an employee for tax purposes (e.g. in the construction industry). Treat such people as self-employed. For all directors and managers who say that they are self-employed, check whether they work for a limited company. If they do, code them as employees for tax and NI purposes. An employee working through an agency (e.g. secretary, nurse etc) may either be an employee of that agency (i.e. the agency pays part of their NI contributions) or an employee with each different employer they go to.

Mixed Mode Alternatives

Web Interview

Help

If unsure whether you are an employee or you are self-employed, consider how you are described for tax and National Insurance purposes.

*Self-employed people usually pay tax directly to the Inland Revenue rather than having it deducted at source.

*People working as child minders or odd-jobbing are usually classified as self-employed.

*If you are self-employed and working under contract to an employer (e.g. in the construction industry), you may be treated as an employee for tax purposes, but please say self-employed here.

*If you are a director or manager working for a limited company, please select employee.

*Someone working through an agency may be an employee of that agency (the agency pays part of their NI contributions) or an employee of each place they work at.

Use

If (JBSEMPCHK = 1) Compute JBSEMP = ff_jbsemp;

Else: Ask ;

Modules

Module Currentemployment_w11. *Current Employment module*

Universe

If (JBHAS = 1 OR JBOFF = 1) //Has a job

Employees_w10. **Paygl.** *Gross pay at last payment*

Type	Decimals	Width	Range	Don't Know	Refused	Inapplicable	Missing
number	0	7	1..9999997	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

The last time you were paid, what was your gross pay - that is including any overtime, bonuses, commission, tips or tax refund but before any deductions for tax, National Insurance or pension contributions, union dues and so on?

Interviewer Instruction

IF DON'T KNOW/CAN'T REMEMBER PROBE FOR APPROXIMATE AMOUNT
ENTER TO NEAREST POUND
RESPONDENT TO CHECK PAYSIP IF POSSIBLE

Help

Include cases where respondent received Statutory Sick Pay or Statutory Maternity Pay (i.e. is currently on paid maternity leave). If it is possible to persuade the respondent to refer to his or her current or last payslip you should do so.

Mixed Mode Alternatives

Web Interview

text

The last time you were paid, what was your gross pay - that is including any overtime, bonuses, commission, tips or tax refund but before any deductions for tax, National Insurance or pension contributions, union dues and so on? Please provide amount to the nearest pound, check your payslip if possible.

Question Box Label

Amount in £

Help

Include Statutory Sick Pay or Statutory Maternity Pay if you are receiving these.

Delete

Interviewer Instruction

Use

Ask PayGL

Modules

Module Employees_w10. *Employees module*

Universe

`If (CURRENTEMPLOYMENT.JBSEMP = 1) //employee`

Employees_w10. **Paynl.** *Take home pay at last payment*

Type	Decimals	Width	Range	Don't Know	Refused	Inapplicable	Missing
number	0	7	0..9999997	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

And what was your take home pay last time, that is after any deductions were made for tax, National Insurance, pensions, union dues etc?

Interviewer Instruction

ENTER TO NEAREST POUND

IF NO DEDUCTIONS MADE PLEASE ENTER 0

IF DON'T KNOW/CAN'T REMEMBER PROBE FOR APPROXIMATE AMOUNT

RESPONDENT TO CHECK PAYSIP IF POSSIBLE

Help

Include cases where respondent received Statutory Sick Pay or Statutory Maternity Pay (i.e. is currently on paid maternity leave). If it is possible to persuade the respondent to refer to his or her current or last payslip you should do so.

Mixed Mode Alternatives

Web Interview

Text

And what was your take home pay last time, that is after any deductions were made for tax, National Insurance, pensions, union dues etc? If no deductions, enter 0. Otherwise, please provide an amount to the nearest pound. Check your payslip if possible.

Question Box Label

Amount in £

Help

Include Statutory Sick Pay or Statutory Maternity Pay if you are receiving these.

Delete

Interviewer Instruction

Use

Ask PayNL

Modules

Module Employees_w10. *Employees module*

Universe

If (CURRENTEMPLOYMENT.JBSEMP = 1) // employee

And If (PayGL <> REF) // Gross pay at last payment is not refused

Employees_w10. **Paynwc.** *Pay period: take home pay*

Type	Don't Know	Refused	Inapplicable	Missing
choice	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

How long a period did that cover?

Options

0	Per hour	Per hour
1	One week	One week
2	Two weeks	Two weeks
3	Three weeks	Three weeks
4	Four weeks	Four weeks
5	Calendar month	Calendar month
6	Two calendar months	Two calendar months
7	Eight times a year	Eight times a year
8	Nine times a year	Nine times a year
9	Ten times a year	Ten times a year
13	Three months/13 weeks	Three months/13 weeks
26	Six months/26 weeks	Six months/26 weeks
52	One year/12 months/52 weeks	1 yr/12 months/52 weeks
90	Less than once a week	Less than once a week
95	One off/lump sum	One off/lump sum
96	None of these	None of these

Help

Find out how often respondent gets paid. If respondent gets paid weekly, code 'one week'. If respondent is paid more regularly than once a week, code 'less than once a week'.

Mixed Mode Alternatives

Web Interview

Question Box Label

Period

Delete

Text, Help

Use

Ask PayNWC

Modules

Module Employees_w10. *Employees module*

Universe

```
If (CURRENTEMPLOYMENT.JBSEMP = 1) //employee
And If (PayGL <> REF) //Gross pay at last payment is not refused
And If (PayNL = 1..9999997) //Take home pay at last payment provided
```

Secondjobs_w7. **J2pay.** *Gross earnings from second jobs last month*

Type	Decimals	Width	Range	Don't Know	Refused	Inapplicable	Missing
number	0	5	0..99997	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

Before tax and other deductions, how much do you earn from your second and all other occasional jobs in a usual month?

Interviewer Instruction

IF NO USUAL PROBE FOR LAST AMOUNT RECEIVED
ENTER TO NEAREST £

Mixed Mode Alternatives

Web Interview

Text

Before tax and other deductions, how much do you earn from your second and all other occasional jobs in a usual month? Please provide an amount to the nearest pound. If no usual, please give the last amount received.

Question Box Label

Amount in £

Delete

Interviewer Instruction

Use

Ask J2PAY

Modules

Module Secondjobs_w7. *Second Jobs module*

Universe

If (J2HAS = 1) // Has a second job

Soft Check: If (J2PAY < 50 OR > 400) // Gross earning from second job, last month is less than 50 or greater than 400

Text

You have just entered your gross earnings from your second job are £[J2PAY] per month.

Module Gigeconomy_w11. *Gig Economy module*

Version

01 (Mon Jun 11 15:10:28 2018) produced Tue May 18 11:04:09 2021

Use

Use Gigeconomy_w11

Householdfinances_w7. **Fiyrdia.** *Amount received in interest/dividends*

Type	Decimals	Width	Range	Don't Know	Refused	Inapplicable	Missing
number	0	7	0..9999997	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

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

Interviewer Instruction

WRITE IN TO NEAREST £

PLEASE EXCLUDE RETURNS FROM PREMIUM BONDS

Help

Probe for an approximate amount. If 'don't know' there will be a series of questions to find out the range of the amount. Those who refuse will be routed past the follow-up questions.

Mixed Mode Alternatives

Web Interview

Text

In the past 12 months how much have you personally received in the way of dividends or interest from any savings and investments you may have? Please enter an amount to the nearest pound.

Question Box Label

Amount in £

Help

Please exclude returns from Premium Bonds

Delete

Interviewer Instruction, Help

Use

Ask FIYRDIA

Modules

Module Householdfinances_w7. Household Finances module

If (FIYRDIA = DK) // If Don't Know how much personally received in dividends or interest from any savings or investments

Benefits_w10.Frval. Last amount received

Type	Decimals	Width	Range	Don't Know	Refused	Inapplicable	Missing
number	0	7	0..9999997	-1	-2	-8	-9

Source

BHPS

Version

1.0

Text

How much was the last payment of *[FICODE]* you received (to nearest £)?

Interviewer Instruction

IF 'DON'T KNOW/CAN'T REMEMBER' PROBE FOR APPROXIMATE AMOUNT

IF RESPONDENT CANNOT SEPARATE OUT THE AMOUNTS FOR A SPECIFIC STATE BENEFIT, RECORD TOTAL AMOUNT FOR ALL COMBINED STATE BENEFITS THEN INDICATE AT SUBSEQUENT STATE BENEFITS THAT THE SPECIFIC AMOUNT HAS BEEN INCLUDED AS PART OF A PAYMENT ALREADY RECORDED BY ENTERING A VALUE OF £0.

Mixed Mode Alternatives

Web Interview

Text

How much was the last payment of *[FICODE]* you received (to nearest £)? If you are unsure, please give an approximate amount. If you have already included this in another payment you have told us about, please enter 0.

Question Box Label

Amount in £

Delete

Interviewer Instruction

Use

Ask FRVAL

Modules

Module Benefits_w10. Unearned Income and State Benefits module

In Loops

!loop for each enumerated benefit source OR for each FRADDL = 1

If (FRVAL > 0) // Amount received stated

Benefit sources (corresponding to question 'frval' above)

(1) NI Retirement/State Retirement (Old Age) Pension (2) A Pension from a previous employer (3) A Pension from a spouse's previous employer (4) A Private Pension/Annuity (5) A Widow's or War Widow's Pension (6) A Widowed Mother's Allowance/ Widowed Parent's Allowance/ Bereavement Allowance (7) Pension Credit (includes Guarantee Credit & Saving Credit) (8) Severe Disablement Allowance (9) Industrial Injury Disablement Allowance (10) Disability Living Allowance (11) Attendance Allowance (12) Carer's Allowance (formerly Invalid Care Allowance) (13) War Disablement Pension (14) Incapacity Benefit (15) Income Support (16) Job Seeker's Allowance (17) Child Benefit (including Lone-Parent Child Benefit payments) (18) Child Tax Credit (19) Working Tax Credit (includes Disabled Person's Tax Credit) (20) Maternity Allowance (21) Housing Benefit (22) Council Tax Benefit (23) Educational Grant (not Student Loan or Tuition Fee Loan (24) Trade Union/ Friendly Society Payment (25) Maintenance or Alimony (26) Payments from a family member not living here (27) Rent from Boarders or Lodgers (not family members) living here with you (28) Rent from any other property (29) Foster Allowance / Guardian Allowance (30) Rent Rebate (31) Rate Rebate (32) Employment and Support Allowance (33) Return to Work Credit (34) Sickness and Accident Insurance (35) In-Work Credit for Lone Parents (36) Other Disability Related Benefit or Payment (37) Any other regular payment (38) Any other state benefit (39) Universal Credit (40) Personal Independence Payments.

B. Additional tables and figures

Background: Household income was asked from wave two of the COVID-19 study. Wave one also slightly differs from the other waves in terms of the household earnings question. All households were directly asked the household earnings total at wave one, whereas after wave one, households reporting no earners were routed out of the earnings total question. Furthermore, in interview soft checks were introduced from wave 3. The softchecks alerted respondents where $hhearnings < indearnings$ (earnings check) and $hhincome < hhearnings$ (income check).

B.1 Inconsistency checks

Wave	Earnings		Income	
	N	% inconsistent	N	% inconsistent
Panel A: Single adult households				
1	-	-	.	.
2	-	-	2078	3.75%
3	-	-	1997	3.76%
4	-	-	1800	4%
5	-	-	1654	3.99%
6	-	-	1550	4.65%
7	-	-	1557	4.17%
8	-	-	1544	4.53%
Total	-	-	12180	4.09%
Panel B: Multiple adult households				
1	6688	3.66%	.	.
2	5639	10.91%	9461	9.29%
3	5228	1.47%	8915	3.05%
4	4969	0.72%	7753	2.23%
5	4511	0.75%	7213	2.2%
6	3832	0.63%	6438	3.09%
7	3739	0.78%	6367	2.39%
8	4109	0.78%	6794	2.88%
Total	38715	2.82%	52941	3.83%

Table A.1: Inconsistent reporting by wave

Notes: Sample of all individuals completing a full Covid survey interview at a given wave and providing answers to the earnings and income questions. Household income was asked from wave two of the COVID-19 study only. Prior to wave 3 there were no in-interview soft checks for $hhearn < indearn$ (earnings check) and $hhincome < hhearn$ (income check). In the table, an inconsistent earnings report is defined as reporting household earnings*1.05 < net individual pay. An inconsistent income report is defined as reporting household income*1.05 < household earnings. The top left panel is missing as for single adult households, household earnings is equal to individual earnings.

B.2 Analysis of wave one and two

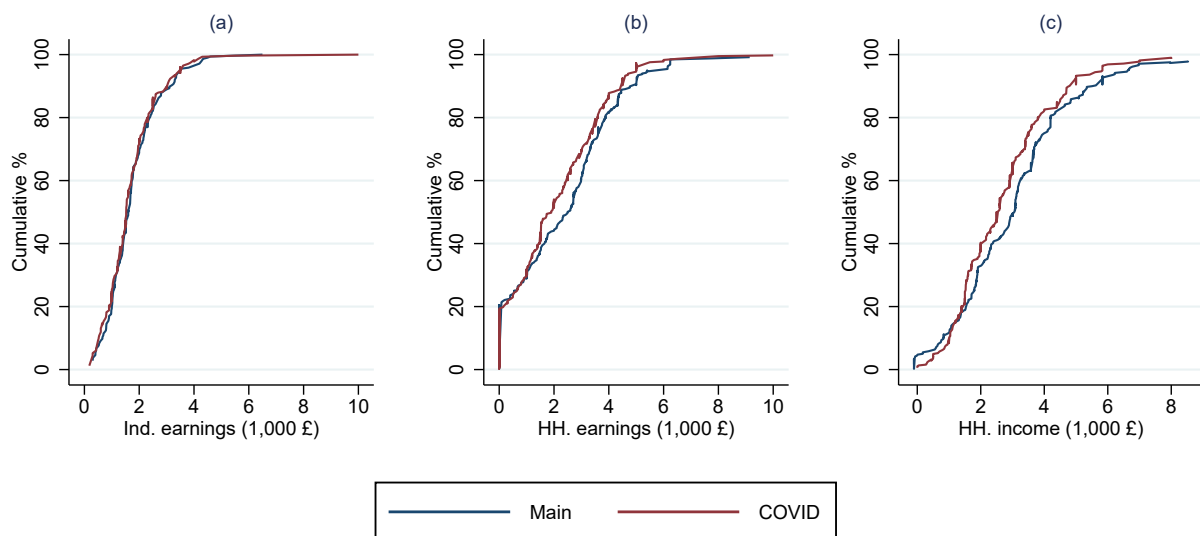


Figure A.1: Income and earnings CDFs, wave 1 and 2 sample

Notes: Sample of individuals completing a COVID survey interview within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). Household income was asked from wave two of the COVID-19 study only. All income sources are net of tax and national insurance contributions. Earnings are reported for the sample of working age (age < 66). Observations with item-missing data on an income source are removed for that source. Weights correct for item non-response, survey non-response and survey design. Weights match each (income source and survey specific) sample to Wave 11 main survey totals defined by gender, age and education. Sample sizes for the COVID (main) survey are as follows: 413 (377) individual earnings, 597 (539) for household earnings and 325 (567) for household income.

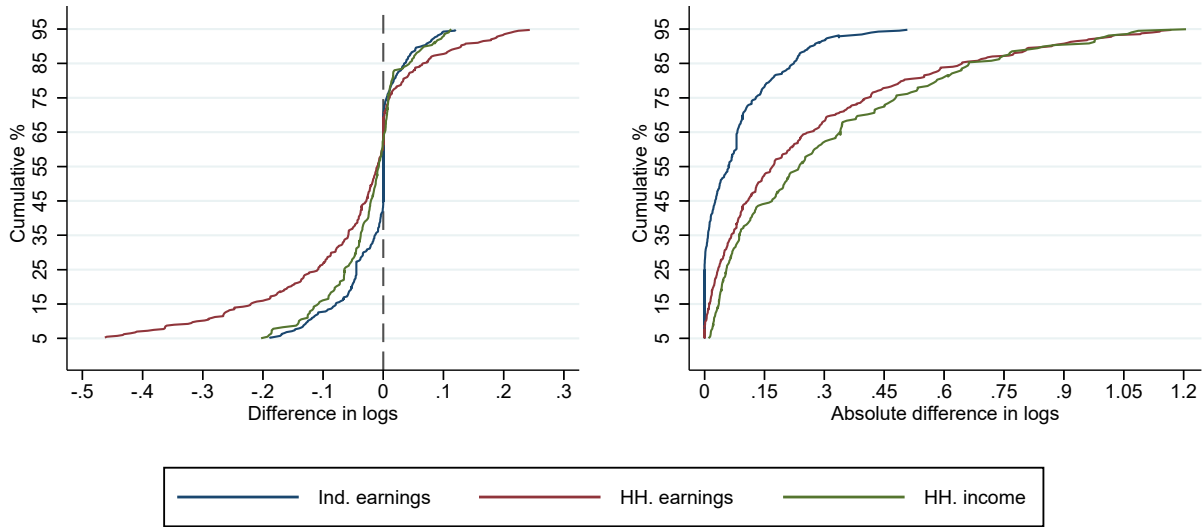


Figure A.2: CDFs of differences in reported earnings and income (COVID-19 - Main), wave 1 and 2 sample

Notes: Sample of individuals completing a COVID survey interview (wave one or two) within 14 days of their main survey interview where all adults in the household completed a main survey individual interview ('complete households'). In the case of multiple matches the nearest match is kept. Household income was asked from wave two of the COVID-19 study only. All income sources are net of tax and national insurance contributions. Individual earnings are conditional on reporting paid employment as an employee in both surveys. Individual and household earnings are reported for the sample of working age (age<66). Observations with item-missing data on an income source are removed for that source. Differences are constructed as: $\log(\text{COVID-19 study amount}) - \log(\text{main study amount})$.